Programa de Doctorado en Estadística, Optimización y Matemática Aplicada

Estimación de fronteras de producción a través de la minimización del riesgo estructural y Support Vector Machines

Nadia María Guerrero Martínez

Director de la tesis

Dr. D. Juan Aparicio Baeza

Universidad Miguel Hernández de Elche

-2025-

La presente Tesis Doctoral, titulada "Estimación de fronteras de producción a través de la minimización del riesgo estructural y Support Vector Machines", se presenta bajo la modalidad de **tesis por compendio** de las siguientes **publicaciones**:

- Guerrero, N. M., Aparicio, J., & Valero-Carreras, D. (2022). Combining data envelopment analysis and machine learning. *Mathematics*, *10*(6), 909. https://doi.org/10.3390/math10060909

- Guerrero, N.M., Aparicio, J., Moragues, R., Valero-Carreras, D. (2024). Merging Data Envelopment Analysis and Structural Risk Minimization: Some Examples of Use of Multi-output Machine Learning Techniques on Real-World Data. In: Edalatpanah, S., Hosseinzadeh Lotfi, F., Kerstens, K., Wanke, P. (eds) Analytical Decision Making and Data Envelopment Analysis. Infosys Science Foundation Series(). Springer, Singapore. https://doi.org/10.1007/978-981-97-6972-8_1

- Guerrero, N. M., Moragues, R., Aparicio, J., & Valero-Carreras, D. (2024). Support Vector Frontiers with kernel splines. *Omega*, 103130. https://doi.org/10.1016/j.omega.2024.103130

UNIVERSITAS
Miguel Hernández

El Dr. D. *Juan Aparicio Baeza*, director de la tesis doctoral titulada “Estimación de fronteras de producción a través de la minimización del riesgo estructural y Support Vector Machines”

INFORMA:

Que Dña. *Nadia María Guerrero Martínez* ha realizado bajo mi supervisión el trabajo titulado **“Estimación de fronteras de producción a través de la minimización del riesgo estructural y Support Vector Machines”** conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, en Elche a 7 de abril de 2025.

Director de la tesis

Dr. D. *Juan Aparicio Baeza*

UNIVERSITAS
Miguel Hernández

El Dr. D. *Domingo Morales González*, Coordinador del Programa de Doctorado en **Estadística, Optimización y Matemática Aplicada**

INFORMA:

Que Dña. *Nadia María Guerrero Martínez* ha realizado bajo la supervisión de nuestro Programa de Doctorado el trabajo titulado **"Estimación de fronteras de producción a través de la minimización del riesgo estructural y Support Vector Machines"** conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, en Elche a 7 de abril de 2025.

Coordinador/a del Programa de Doctorado en Estadística,
Optimización y Matemática Aplicada
Prof. Dr. D. *Domingo Morales González*

La autora de esta tesis Dña. *Nadia María Guerrero Martínez* agradece la ayuda proporcionada por el proyecto de I+D+i *PID2022-136383NB-I00*, financiado por **MICIU/AEI /10.13039/501100011033** y por **FEDER, UE**, el proyecto *PID2019-105952GB-I00* financiado por el **Ministerio de Ciencia e Innovación/ Agencia Estatal de Investigación/10.13039/501100011033** y la **Universidad Miguel Hernández de Elche** a través de la beca predoctoral *01623/2020*.

UNIVERSITAS Miguel Hernández

1. ÍNDICE

2. RESUMEN GLOBAL DE LA TESIS

ESPAÑOL

Esta tesis doctoral explora la integración del **Análisis Envolvente de Datos (DEA)** con técnicas de **Aprendizaje Automático (Machine Learning, ML)**, particularmente la **Minimización del Riesgo Estructural (SRM)** y las **Máquinas de Vectores de Soporte (SVM)**, con el fin de mejorar la estimación de fronteras de eficiencia y reducir el problema del sobreajuste. Se introduce **Data Envelopment Analysis-based Machines (DEAM)**, un modelo inspirado en **Support Vector Regression (SVR)** que controla tanto el error empírico como el error de generalización mediante límites PAC, mostrando un mejor desempeño que DEA en términos de sesgo y error cuadrático medio. Posteriormente, DEAM se amplía a un entorno **multi-output**, lo que permite evaluar la eficiencia en procesos productivos con múltiples entradas y salidas, mejorando la capacidad de inferencia sobre la tecnología de producción. Finalmente, se desarrolla **Support Vector Frontiers with Kernel Splines (SVF-Splines)**, un método basado en SVR con **splines lineales** para estimar tecnologías de producción convexas en un solo paso, reduciendo significativamente la complejidad computacional y el error cuadrático medio en comparación con DEA y otros métodos basados en SVM. En conjunto, esta tesis representa un puente entre la eficiencia productiva y el aprendizaje automático, proporcionando modelos más robustos, precisos y computacionalmente eficientes para la estimación de tecnologías de producción en microeconomía e ingeniería.

3. GLOBAL SUMMARY OF THE THESIS

ENGLISH

This doctoral thesis explores the integration of **Data Envelopment Analysis (DEA)** with **Machine Learning (ML)** techniques, particularly **Structural Risk Minimization (SRM)** and **Support Vector Machines (SVM)**, to improve efficiency frontier estimation and mitigate overfitting. It introduces **Data Envelopment Analysis-based Machines (DEAM)**, a model inspired by **Support Vector Regression (SVR)** that controls both empirical and generalization errors using PAC bounds, demonstrating superior performance over DEA in terms of bias and mean squared error. DEAM is then extended to a **multi-output** setting, allowing efficiency evaluation in production processes with multiple inputs and outputs, enhancing inference on production technologies. Finally, **Support Vector Frontiers with Kernel Splines (SVF-Splines)** is developed, a method based on SVR with **linear splines** to estimate convex production technologies in a single step, significantly reducing computational complexity and mean squared error compared to DEA and other SVM-based methods. Overall, this thesis bridges production efficiency and machine learning, providing more robust, accurate, and computationally efficient models for estimating production technologies in microeconomics and engineering.

4. INTRODUCCIÓN GENERAL DE LA TESIS

El estudio de la eficiencia en procesos productivos ha sido un tema central en economía, ingeniería e investigación operativa. Dentro de este campo, el **Análisis Envolvente de Datos (DEA por sus siglas en inglés)** (Charnes et al. [2] y Banker et al. [1]) se ha consolidado como una de las metodologías no paramétricas más utilizadas para medir la eficiencia técnica de unidades de decisión (DMUs por sus siglas en inglés). Sin embargo, DEA ha sido criticado por su naturaleza puramente descriptiva y su tendencia al **sobreajuste (overfitting)** debido al principio de mínima extrapolación. Esto implica que la frontera de eficiencia estimada se ajusta excesivamente a la muestra observada, reduciendo su capacidad para generalizar más allá de los datos disponibles.

Por otro lado, en los últimos años, los avances en **Aprendizaje Automático (Machine Learning, ML)** han revolucionado diversas áreas de modelado estadístico y optimización. Técnicas como las **Máquinas de Vectores de Soporte (Support Vector Machines, SVM)** han demostrado ser particularmente útiles para la construcción de modelos con buen desempeño en términos de generalización. En este contexto, la presente tesis doctoral explora la intersección entre DEA y ML, con el objetivo de superar las limitaciones inferenciales del DEA tradicional mediante la integración de principios de **Minimización del Riesgo Estructural (SRM)** (Vapnik [7]) y el uso de **Máquinas de Vectores de Soporte para Regresión (Support Vector Regression, SVR)** (Vapnik [8]).

El trabajo se articula en tres artículos, cada uno de los cuales introduce mejoras progresivas en la estimación de fronteras de eficiencia, desde la adaptación inicial de DEA a un enfoque basado en aprendizaje automático hasta el desarrollo de métodos más eficientes y escalables. A continuación, se describe la relación entre estos artículos y su contribución al objetivo general de la tesis.

4.1. Primer Artículo: "Combining Data Envelopment Analysis and Machine Learning"

El primer artículo presenta un nuevo enfoque metodológico denominado **Data Envelopment Analysis-based Machines (DEAM)**, que integra DEA con técnicas de aprendizaje automático para mejorar la estimación de fronteras de eficiencia. En particular, se incorpora el principio de **Minimización del Riesgo Estructural (SRM)** para controlar simultáneamente el **error empírico y el error de generalización**.

El DEA tradicional solo minimiza el error empírico al construir una frontera que se ajusta exactamente a los datos observados, lo que conduce al problema de sobreajuste. En cambio, DEAM introduce una metodología inspirada en **Support Vector Regression (SVR)**, la cual permite obtener un estimador de frontera más robusto mediante la utilización de márgenes de tolerancia y la incorporación de límites PAC (Probably Approximately Correct) para la dimensión de Vapnik-Chervonenkis (VC) (Valiant [6] y Cristianini et al.[3]).

Los resultados de este estudio, obtenidos a través de simulaciones, muestran que **DEAM mejora significativamente la precisión en la estimación de fronteras de producción** en comparación con DEA. En particular, los valores de **sesgo y error cuadrático medio (MSE)** son consistentemente menores en DEAM, lo que indica que este método no solo se ajusta a los datos disponibles, sino que también es capaz de generalizar mejor a nuevos conjuntos de datos.

4.2. Segundo Artículo: "Merging Data Envelopment Analysis and Structural Risk Minimization: Some Examples of Use of Multi-Output Machine Learning Techniques on Real-World Data"

El segundo artículo extiende el modelo DEAM al contexto **multi-output**, es decir, escenarios en los que múltiples productos (outputs) son generados a partir de múltiples recursos (inputs). En la práctica, muchas aplicaciones económicas y de ingeniería requieren la estimación de tecnologías de producción con múltiples productos o servicios, lo que impone desafíos adicionales a los métodos tradicionales de eficiencia.

En este capítulo de libro, se presentan **ejemplos empíricos** en los que se evalúa el desempeño de DEAM en comparación con DEA en términos de varias medidas de eficiencia técnica. Se exploran distintas métricas utilizadas en la literatura, como la **medida radial orientada a los inputs y a los outputs**, la **función de distancia direccional (DDF)** y las **medidas de eficiencia de Russell y Aditivas Ponderadas**.

Adicionalmente, se incorpora un **procedimiento de validación cruzada** para seleccionar los hiperparámetros del modelo, lo que permite mejorar aún más su capacidad de predicción. Los resultados de los experimentos con datos reales demuestran que **DEAM en entornos multi-output mantiene su ventaja sobre DEA**, proporcionando estimaciones de eficiencia más precisas y menos susceptibles al sobreajuste.

Otro aspecto clave de este artículo es la diferenciación entre **eficiencia relativa y absoluta**. Mientras que DEA evalúa la eficiencia relativa con respecto a la muestra observada, DEAM tiene la capacidad de estimar la eficiencia absoluta en relación con la función de producción subyacente. Esto es particularmente útil en aplicaciones donde el objetivo no es solo clasificar las DMUs, sino también comprender la verdadera relación entre inputs y outputs en una industria o sector específico.

4.3. Tercer Artículo: "Support Vector Frontiers with Kernel Splines"

El tercer artículo introduce una nueva metodología basada en **Support Vector Regression (SVR)** llamada **Support Vector Frontiers (SVF)** (Valero-Carreras et al. [4, 5])**,** que adapta SVR para la estimación de tecnologías de producción a través de **fronteras escalonadas**. En particular, se propone **SVF-Splines**, una extensión que utiliza **splines lineales** como función de transformación para estimar tecnologías de producción convexas sin necesidad de convexificación en dos pasos, como ocurre en el Análisis Envolvente de Datos tradicional cuando se pasa del modelo “escalonado” FDH (Free Disposal Hull) al modelo “lineal a trozos” DEA.

El enfoque basado en **kernels de splines** tiene varias ventajas clave:

1. **Reduce la complejidad computacional** en comparación con SVF tradicional, permitiendo una estimación eficiente de tecnologías de producción convexas en un solo paso.
2. **Mejora la precisión de la estimación**, con reducciones de hasta **95% en el error cuadrático medio (MSE)** en comparación con DEA.
3. **Acelera los tiempos de ejecución**, siendo hasta **70 veces más rápido** que SVF en problemas de gran escala.

Este artículo también introduce una adaptación de medidas de eficiencia clásica a SVF-Splines, permitiendo la evaluación de unidades de decisión con métricas como la **medida radial, la función de distancia direccional y la medida aditiva ponderada**. La comparación empírica muestra que SVF-Splines no solo supera a DEA en términos de precisión, sino que también ofrece ventajas computacionales significativas, lo que lo convierte en un método más viable para aplicaciones en entornos reales.

4.4. Relación entre los artículos y contribución de la tesis

Los tres artículos forman un cuerpo de trabajo coherente que avanza en la integración de DEA con técnicas de aprendizaje automático para mejorar la estimación de fronteras de producción. En conjunto, la tesis hace las siguientes contribuciones clave:

1. **Superación del problema de sobreajuste en DEA** mediante la introducción de límites PAC y la minimización del error de generalización.
2. **Extensión del análisis de eficiencia basado en ML a escenarios multi-output multi-input**, permitiendo la evaluación de procesos productivos con múltiples entradas y salidas.
3. **Desarrollo de modelos computacionalmente más eficientes**, mediante el uso de splines y técnicas de SVR para estimar fronteras convexas sin necesidad de procedimientos en dos pasos.
4. **Diferenciación entre eficiencia relativa y absoluta**, proporcionando herramientas para la estimación de la verdadera función de producción subyacente en distintos sectores económicos.

En términos más amplios, esta tesis representa un **puente entre la teoría de eficiencia y el aprendizaje automático**, abriendo nuevas posibilidades para la estimación de tecnologías de producción en contextos de microeconomía, ingeniería e investigación operativa. Al proporcionar métodos más robustos, precisos y computacionalmente eficientes, las contribuciones de esta investigación podrían llegar a tener un alto potencial de impacto tanto en el ámbito académico como en aplicaciones prácticas en diversas industrias y sectores económicos.

5. MATERIALES Y MÉTODOS

Esta tesis doctoral se basa en un marco metodológico que combina enfoques del **Análisis Envolvente de Datos (DEA)** y el **Aprendizaje Automático (Machine Learning, ML)**, con énfasis en la **Minimización del Riesgo Estructural (SRM)** y las **Máquinas de Vectores de Soporte (SVM)**. La metodología utilizada en los tres artículos se fundamenta en técnicas matemáticas avanzadas, teoría estadística del aprendizaje y modelos de optimización, integrando herramientas de la investigación operativa, la inferencia estadística y el análisis de eficiencia productiva.

5.1. Fundamentos Teóricos y Antecedentes Metodológicos

5.1.1. Análisis Envolvente de Datos (DEA) y sus limitaciones

El **Análisis Envolvente de Datos (DEA)** es una técnica no paramétrica utilizada para la estimación de fronteras de producción y la medición de la eficiencia técnica en unidades de decisión (**DMUs**). DEA se basa en la construcción de un **conjunto de posibilidades de producción**, definido como el menor conjunto convexamente cerrado que contiene todas las observaciones y satisface la propiedad de **libre disponibilidad** de los recursos y los productos.

Matemáticamente, la frontera de producción en DEA se suele obtener resolviendo problemas de **programación lineal**, los cuales definen una envolvente convexa sobre los datos observados. Sin embargo, DEA ha sido criticado por ser una metodología puramente **descriptiva**, careciendo de fundamentos estadísticos para realizar inferencias sobre la tecnología de producción subyacente. Una de las principales limitaciones de DEA es su tendencia al **sobreajuste (overfitting)**, dado que construye la frontera más ajustada posible a los datos sin considerar la capacidad de generalización del modelo.

5.1.2. Teoría Estadística del Aprendizaje y Minimización del Riesgo Estructural (SRM)

Para abordar las limitaciones de DEA, esta investigación incorpora herramientas de la **teoría estadística del aprendizaje**, particularmente el principio de **Minimización del**

Riesgo Estructural (SRM) desarrollado por Vapnik. SRM establece que un modelo óptimo debe equilibrar dos fuentes de error:

- **Error empírico**: Error de ajuste sobre la muestra observada.
- **Error de generalización**: Error esperado en datos no observados.

En lugar de minimizar únicamente el error empírico, como lo hace DEA, la metodología propuesta introduce cotas **Probably Approximately Correct (PAC)** para el error de generalización, basadas en la **dimensión de Vapnik-Chervonenkis (VC dimension)** y la **dimensión fat-shattering**. Estas métricas permiten cuantificar la complejidad del modelo y establecer límites sobre la capacidad de generalización.

5.1.3. Máquinas de Vectores de Soporte (SVM) y Support Vector Regression (SVR)

El marco metodológico de la tesis también incorpora técnicas de **Máquinas de Vectores de Soporte (SVM)**, específicamente **Support Vector Regression (SVR)**. SVR es un método de regresión que maximiza el margen de tolerancia alrededor de la función de ajuste, asegurando que las predicciones sean robustas frente a la variabilidad de los datos.

Matemáticamente, SVR se formula como un problema de **optimización cuadrática restringida**, donde la función objetivo minimiza una combinación del error empírico y un término de regularización que controla la complejidad del modelo. En términos generales, la formulación básica de SVR es:

$$\underset{\mathrm{w},b,\xi}{Min} \quad \frac{1}{2}\|\mathrm{w}\|^2 + C\sum_{\mathbf{i}} \xi_i$$

$$s.a.$$

$$y_i - \left(\mathrm{w}\cdot \mathrm{x}_i + b\right) \leq \varepsilon + \xi_i$$

$$\left(\mathrm{w}\cdot \mathrm{x}_i + b\right) - y_i \leq \varepsilon + \xi_i,$$

donde ξ_i son términos de error y C es un hiperparámetro que establece un balance entre el margen y el error de ajuste (Vázquez y Walter [9]).

SVR se utiliza en la tesis como base para desarrollar **fronteras de eficiencia robustas**, evitando el sobreajuste característico de DEA. Además, se exploran transformaciones mediante **kernels de splines**, lo que permite extender SVR a problemas de estimación de fronteras convexas.

5.2. Métodos Aplicados para el Desarrollo de Nuevos Modelos

5.2.1. Desarrollo de Data Envelopment Analysis-based Machines (DEAM)

A partir de los fundamentos anteriores, se introduce el modelo **Data Envelopment Analysis-based Machines (DEAM)**, que reformula DEA bajo el principio de **Minimización del Riesgo Estructural (SRM)**. DEAM se modela como un problema de **optimización cuadrática**, incorporando restricciones de margen similares a SVR para estimar una frontera de producción con menor error de generalización.

Para evaluar la eficacia de DEAM, se realizan **simulaciones computacionales** comparando su desempeño con DEA en distintos escenarios de datos, analizando métricas como:

- **Sesgo de la estimación**
- **Error cuadrático medio (MSE)**
- **Capacidad de generalización en muestras no observadas**

5.2.2. Extensión a entornos Multi-Output

Se amplía la formulación de DEAM para abordar problemas con múltiples recursos y múltiples productos. En este contexto, se redefine la tecnología de producción como un **conjunto convexo de alta dimensión**, estimado mediante la intersección de múltiples hiperplanos. Se emplean distintas métricas de eficiencia:

- **Medidas radiales (input y output-oriented)**
- **Funciones de distancia direccional (DDF)**

- **Medidas aditivas ponderadas**

Para optimizar los hiperparámetros del modelo, se utiliza **validación cruzada (cross-validation)** con búsqueda en malla (**grid search**), permitiendo seleccionar los valores óptimos para la regularización y el margen de tolerancia.

5.2.3. Desarrollo de Support Vector Frontiers with Kernel Splines (SVF-Splines)

Finalmente, se propone **Support Vector Frontiers with Kernel Splines (SVF-Splines)**, un modelo basado en **SVR con transformación de splines** para estimar tecnologías de producción convexas.

A diferencia de DEA y DEAM, que generan **fronteras piecewise linear**, SVF-Splines utiliza un **kernel de splines lineales** para construir una frontera **suavizada y convexa** en un solo paso. Este modelo se formula como un problema de **optimización convexa**, minimizando una función objetivo de la forma:

$$\underset{w,b}{Min}\frac{1}{2}\|w\|^2+C\sum_i max\left(0,\left|y_i-\left(w\cdot x_i+b\right)\right|-\varepsilon\right)$$

Se evalúa el rendimiento de SVF-Splines en términos de **precisión y eficiencia computacional**, comparando su desempeño con DEA, DEAM y SVF tradicional. Los resultados muestran que SVF-Splines reduce **hasta un 95% el error cuadrático medio (MSE)** y **acelera el tiempo de cómputo hasta 70 veces** en comparación con DEA.

5.3. Validación y Evaluación Experimental

Para validar los modelos desarrollados, se implementan **dos estrategias de evaluación**:

1. **Simulaciones computacionales**: Se generan conjuntos de datos sintéticos con distintas estructuras de producción y ruido estadístico, permitiendo evaluar la capacidad de los modelos para recuperar la verdadera frontera de producción.

2. **Aplicaciones empíricas con datos reales**: Se emplean datasets provenientes de sectores económicos e industriales para comparar las estimaciones de DEA, DEAM y SVF-Splines en contextos prácticos.

Los resultados de la validación confirman que los modelos propuestos **superan a DEA en términos de precisión, robustez y capacidad de generalización**, estableciendo un nuevo marco metodológico para la medición de eficiencia en escenarios complejos.

6. DISCUSIÓN

Los resultados obtenidos en los tres artículos que conforman esta tesis doctoral han permitido analizar y mejorar la estimación de la eficiencia técnica mediante la integración del **Análisis Envolvente de Datos (DEA)** y técnicas avanzadas de **Aprendizaje Automático (Machine Learning, ML)**. En particular, la implementación de **Minimización del Riesgo Estructural (SRM)** y **Máquinas de Vectores de Soporte (SVM)** ha mostrado ser una estrategia efectiva para abordar las limitaciones del DEA clásico, especialmente en términos de **sobreajuste, capacidad de generalización y robustez estadística**.

Esta sección analiza los principales hallazgos en cada una de las contribuciones de la tesis, su impacto en la teoría de eficiencia y en la modelización de tecnologías productivas, y las posibles líneas de investigación futura que podrían ampliar o complementar los resultados obtenidos.

6.1. Comparación del Desempeño de DEAM frente a DEA

El primer artículo introduce **Data Envelopment Analysis-based Machines (DEAM)** como una extensión del DEA tradicional que incorpora principios de aprendizaje automático. A través de **simulaciones computacionales**, se ha demostrado que **DEAM supera a DEA en términos de precisión y robustez estadística**. La metodología aplicada en el desarrollo de DEAM permite estimar una frontera de producción **más confiable y con menor sesgo**, lo que representa un avance significativo en la modelización de eficiencia técnica.

Específicamente, los resultados muestran que:

- **DEAM presenta una frontera de eficiencia más estable y menos propensa al sobreajuste**, lo que permite mejorar la inferencia sobre la función de producción subyacente. Esto se debe a que, a diferencia de DEA, que minimiza exclusivamente el error empírico, DEAM emplea un control explícito sobre el error de generalización, evitando la excesiva dependencia de los datos observados.

- **La capacidad de generalización mejora significativamente al minimizar tanto el error empírico como el error de generalización**, contrastando con DEA, que únicamente minimiza el error empírico. Esto es especialmente relevante en contextos donde la cantidad de datos disponibles es limitada o donde los datos están sujetos a ruido y variabilidad.
- **Los valores de sesgo y error cuadrático medio (MSE) son menores en DEAM**, lo que indica que este método proporciona estimaciones más precisas y extrapolables a nuevas observaciones. Al controlar el margen de error en la estimación de la frontera de producción, DEAM reduce la incertidumbre asociada a la medición de eficiencia, permitiendo un análisis más sólido y confiable.

Estos hallazgos respaldan la hipótesis de que una reformulación de DEA bajo el paradigma de **Minimización del Riesgo Estructural (SRM)** puede mejorar la estimación de eficiencia sin comprometer la interpretación económica del modelo. Además, proporcionan un marco metodológico adaptable a diversos sectores productivos, en los cuales la eficiencia técnica es un factor clave para la optimización de recursos.

6.2. Evaluación de DEAM en un Entorno Multi-Output

El segundo artículo amplía la aplicación de DEAM al contexto **multi-output**, permitiendo la estimación de eficiencia en **escenarios más complejos con múltiples entradas y múltiples salidas**. Esta extensión es particularmente relevante para sectores donde los procesos productivos no pueden representarse de manera adecuada con un único producto o salida, como ocurre en la educación, la sanidad, la industria manufacturera y el sector bancario.

A partir de su validación en **datasets reales**, se han obtenido los siguientes resultados:

- **DEAM-MultiOutput mantiene su ventaja sobre DEA en contextos de alta dimensionalidad**, lo que permite evaluar tecnologías de producción más realistas. Esto es particularmente importante cuando la producción involucra múltiples bienes o servicios simultáneamente y donde la eficiencia técnica debe evaluarse de manera conjunta en todas las dimensiones de salida.
- **La incorporación de diferentes medidas de eficiencia (radiales, direccionales y aditivas ponderadas)** ofrece una perspectiva más flexible para la medición de eficiencia en distintos sectores económicos e industriales. La posibilidad de utilizar múltiples métricas permite analizar la eficiencia desde diferentes ángulos, proporcionando una comprensión más completa del desempeño de las unidades de decisión.
- **El uso de validación cruzada (cross-validation) mejora la selección de hiperparámetros**, optimizando el balance entre error empírico y generalización del modelo. La optimización de los parámetros de DEAM-MultiOutput garantiza que la estimación de la frontera de producción sea más robusta y menos sensible a las particularidades de la muestra observada.

En conjunto, esta extensión amplía el alcance de DEA y permite su aplicación en entornos donde la producción involucra **múltiples bienes y servicios**, superando la limitación de los modelos tradicionales que trabajan con una única salida. Además, los resultados obtenidos sugieren que DEAM-MultiOutput es una herramienta más adecuada para entornos productivos complejos, donde la eficiencia debe evaluarse en múltiples dimensiones simultáneamente.

6.3. Desempeño de SVF-Splines frente a DEA y DEAM

El tercer artículo introduce **Support Vector Frontiers with Kernel Splines (SVF-Splines)**, un modelo basado en **Support Vector Regression (SVR)** que utiliza **splines lineales** para estimar **fronteras de eficiencia convexas** sin necesidad de procedimientos de convexificación posteriores. Esta metodología permite obtener una representación **más**

precisa y natural de la frontera de producción, lo que contribuye a mejorar la calidad de las estimaciones de eficiencia.

Los resultados obtenidos muestran mejoras significativas respecto a los métodos anteriores:

- **SVF-Splines reduce el error cuadrático medio (MSE) hasta en un 95% en comparación con DEA y DEAM**, lo que confirma su superioridad en términos de precisión. Este resultado indica que el modelo es capaz de capturar mejor la relación entre recursos y productos, generando estimaciones más cercanas a la verdadera frontera de producción.
- **El tiempo de ejecución se reduce hasta 70 veces en entornos de gran escala**, debido a la eliminación del paso de convexificación requerido en DEA y otros métodos basados en SVR. La reducción en el tiempo de cómputo hace que SVF-Splines sea una alternativa viable para aplicaciones en **big data y análisis de alta dimensión**.
- **La implementación de kernels de splines genera fronteras suavizadas**, disminuyendo la sensibilidad del modelo al ruido en los datos y proporcionando estimaciones más robustas. Esto es especialmente beneficioso en entornos donde los datos contienen variabilidad inherente, como en el sector agrícola, el financiero y el energético.

Estos hallazgos sugieren que **SVF-Splines es una alternativa viable a los métodos tradicionales** para la estimación de eficiencia en entornos de alta complejidad computacional. Al combinar las ventajas de **DEA, SVM y técnicas de interpolación con splines**, este modelo representa un avance metodológico que podría tener aplicaciones en múltiples disciplinas.

7. CONCLUSIONES

Esta tesis doctoral ha explorado y desarrollado nuevas metodologías para la medición de eficiencia técnica, integrando el **Análisis Envolvente de Datos (DEA)** con herramientas avanzadas de **Aprendizaje Automático (Machine Learning, ML)**. La investigación ha abordado las limitaciones del DEA tradicional, principalmente el **sobreajuste, la falta de capacidad de generalización y la ausencia de inferencia estadística**, mediante la incorporación de técnicas derivadas de la **Minimización del Riesgo Estructural (SRM)** y **Máquinas de Vectores de Soporte (SVM)**. A través de los tres artículos que componen esta tesis, se han presentado nuevas metodologías y enfoques que han demostrado mejorar la precisión, robustez y aplicabilidad de los modelos de medición de eficiencia.

Los principales hallazgos de esta investigación pueden resumirse en los siguientes puntos clave:

1. **Desarrollo de Data Envelopment Analysis-based Machines (DEAM)**: Se introdujo una reformulación de DEA basada en **Support Vector Regression (SVR)**, que permite minimizar tanto el **error empírico como el error de generalización**, evitando el problema de sobreajuste característico de DEA. Los resultados de las simulaciones computacionales demostraron que **DEAM presenta menor sesgo y error cuadrático medio (MSE)** en comparación con DEA, lo que sugiere una mejora significativa en la capacidad de predicción del modelo.
2. **Extensión de DEAM a entornos Multi-Output**: Se amplió el marco de DEAM para incorporar escenarios en los que múltiples entradas generan múltiples salidas. Se evaluaron diferentes medidas de eficiencia técnica, incluyendo métricas **radiales, direccionales y aditivas ponderadas**, y se demostró que **DEAM-MultiOutput mantiene su ventaja sobre DEA en términos de precisión y robustez**. Además, se incorporó un mecanismo de **validación**

cruzada para la selección de hiperparámetros, optimizando el rendimiento del modelo en distintos entornos de producción.

3. **Desarrollo de Support Vector Frontiers with Kernel Splines (SVF-Splines)**: Se propuso un modelo basado en **SVR con transformación mediante splines**, lo que permitió estimar **tecnologías de producción convexas sin necesidad de procedimientos de convexificación posteriores**. Los experimentos indicaron que **SVF-Splines reduce hasta en un 95% el error cuadrático medio (MSE) en comparación con DEA**, además de disminuir **hasta 70 veces el tiempo de cómputo en entornos de gran escala**, lo que lo convierte en una alternativa viable para el análisis de eficiencia en **big data**.

En conjunto, estos avances representan un **nuevo marco metodológico para la medición de eficiencia técnica**, que combina las ventajas del **DEA clásico** con técnicas de aprendizaje automático para mejorar la precisión, la capacidad de generalización y la eficiencia computacional. Además, los resultados obtenidos abren nuevas oportunidades para la aplicación de modelos híbridos en diversos sectores económicos e industriales, donde la medición de eficiencia es un componente clave para la toma de decisiones estratégicas.

A pesar de los avances logrados en esta tesis, aún existen múltiples oportunidades para continuar investigando y expandiendo las metodologías propuestas. Algunas de las direcciones más relevantes incluyen:

7.1. Automatización de la Selección de Hiperparámetros en DEAM y SVF-Splines

En la presente investigación, la selección de hiperparámetros como el margen de tolerancia (ϵ) y el factor de regularización (C) en SVR se realizó mediante **validación cruzada y búsqueda en malla (grid search)**. Sin embargo, este proceso puede ser mejorado mediante el uso de **algoritmos de optimización automática**, como **Bayesian Optimization o Algoritmos Evolutivos**, los cuales podrían mejorar la adaptabilidad de los modelos y reducir el tiempo de entrenamiento.

7.2. Expansión de los Modelos a Entornos Dinámicos y Temporales

Los modelos desarrollados en esta tesis se basan en un enfoque **estático**, donde la eficiencia es evaluada en un único período de tiempo. Sin embargo, en muchos entornos productivos, la eficiencia evoluciona con el tiempo. La extensión de DEAM y SVF-Splines a un **marco dinámico** permitiría capturar los cambios en la eficiencia de las unidades de decisión, facilitando la toma de decisiones en todas aquellas empresas o instituciones públicas donde las técnicas fuesen aplicadas.

7.3. Aplicaciones en Sectores con Estructuras de Producción Complejas

Si bien los modelos han sido validados con datos empíricos de sectores industriales y económicos, aún queda por explorar su aplicación en **nuevos ámbitos con estructuras de producción más complejas**, como:

- **Sector financiero**: Evaluación de eficiencia en bancos y entidades crediticias.
- **Sector sanitario**: Medición de la eficiencia en hospitales y clínicas con múltiples indicadores de desempeño.

7.4. Integración con Métodos de Inteligencia Artificial Explicativa (XAI)

Uno de los desafíos en la implementación de modelos de aprendizaje automático es su **interpretabilidad y transparencia**. La combinación de DEAM y SVF-Splines con técnicas de **Explainable AI (XAI)** permitiría generar modelos más interpretables para la toma de decisiones, asegurando que los resultados obtenidos sean comprensibles para los responsables de política económica y estrategia empresarial.

7.5. Uso de Otras Funciones Kernel en SVF-Splines

Si bien SVF-Splines ha demostrado ser altamente eficiente utilizando **splines lineales**, sería interesante explorar el impacto de otros **kernels avanzados**, como **RBF (Radial Basis Function), polinomiales o kernels adaptativos**, en la precisión y flexibilidad del modelo. Esto podría mejorar la estimación de eficiencia en entornos donde la frontera de producción es difícilmente linealizable.

8. REFERENCIAS

[1] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, *30*(9), 1078-1092.

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

[3] Cristianini, N.; Shawe-Taylor, J. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods; Cambridge University Press: Cambridge, MA, USA, 2000.

[4] Valero-Carreras, D., Aparicio, J., & Guerrero, N. M. (2022). Multi-output support vector frontiers. *Computers & Operations Research*, *143*, 105765.

[5] Valero-Carreras, D., Aparicio, J., & Guerrero, N. M. (2021). Support vector frontiers: A new approach for estimating production functions through support vector machines. *Omega*, *104*, 102490.

[6] Valiant, L.G. A theory of the learnable. Commun. ACM 1984, 27, 1134–1142.

[7] Vapnik, V. (1991). Principles of risk minimization for learning theory. *Advances in neural information processing systems*, *4*.

[8] Vapnik, V. (1998). Statistical learning theory. *John Wiley & Sons google schola*, *2*, 831-842.

[9] Vazquez, E.;Walter, E. Multi-output suppport vector regression. IFAC Proc. Vol. 2003, 36, 1783–1788.

9. ANEXO. ARTÍCULOS QUE CONFORMAN LA TESIS

Guerrero, N. M., Aparicio, J., & Valero-Carreras, D. (2022). Combining data envelopment analysis and machine learning. Mathematics, 10(6), 909. https://doi.org/10.3390/math10060909

mathematics

Article

Combining Data Envelopment Analysis and Machine Learning

Nadia M. Guerrero, Juan Aparicio * and Daniel Valero-Carreras

Center of Operations Research (CIO), Miguel Hernandez University of Elche (UMH), 03202 Elche, Spain; nguerrero@umh.es (N.M.G.); dvalero@umh.es (D.V.-C.)
* Correspondence: j.aparicio@umh.es; Tel.: +34-966658517; Fax: +34-966658715

Abstract: Data Envelopment Analysis (DEA) is one of the most used non-parametric techniques for technical efficiency assessment. DEA is exclusively concerned about the minimization of the empirical error, satisfying, at the same time, some shape constraints (convexity and free disposability). Unfortunately, by construction, DEA is a descriptive methodology that is not concerned about preventing overfitting. In this paper, we introduce a new methodology that allows for estimating polyhedral technologies following the Structural Risk Minimization (SRM) principle. This technique is called Data Envelopment Analysis-based Machines (DEAM). Given that the new method controls the generalization error of the model, the corresponding estimate of the technology does not suffer from overfitting. Moreover, the notion of ε-insensitivity is also introduced, generating a new and more robust definition of technical efficiency. Additionally, we show that DEAM can be seen as a machine learning-type extension of DEA, satisfying the same microeconomic postulates except for minimal extrapolation. Finally, the performance of DEAM is evaluated through simulations. We conclude that the frontier estimator derived from DEAM is better than that associated with DEA. The bias and mean squared error obtained for DEAM are smaller in all the scenarios analyzed, regardless of the number of variables and DMUs.

Keywords: data envelopment analysis; PAC learning; support vector regression; machine learning; structural risk minimization

MSC: 90C08

check for updates

Citation: Guerrero, N.M.; Aparicio, J.; Valero-Carreras, D. Combining Data Envelopment Analysis and Machine Learning. *Mathematics* **2022**, *10*, 909. https://doi.org/10.3390/math10060909

Academic Editors: Kuo-Ping Lin, Chien-Chih Wang, Chieh-Liang Wu and Liang Dong

Received: 20 February 2022
Accepted: 9 March 2022
Published: 11 March 2022

1. Introduction

One of the most important issues in the field of statistical learning is the reliability of statistical inference methods. In this framework, a sophisticated theory, the so-called Generalization Theory, explains which factors must be controlled to achieve good generalization. Optimal generalization is achieved when the error generated on evaluating new data through an inference learning method is minimized. The Generalization Theory copes with those factors that allow for the minimization of the prediction or generalization error.

In terms of pattern classifiers, the generalization error is the probability of misclassifying a randomly chosen example that holds with high probability over randomly chosen training sets, and then, a good generalization is achieved when this is minimized. This aim is possible if an upper bound of the generalization error is found, and the parameters on which it depends are controlled in order to reduce it. These bounds are understood as Probably Approximately Correct (PAC) bounds, which specifically means that the probability of the bound failing is small (Probably) when the bound is achieved through the classifier that has a low error rate (Approximately Correct). The standard PAC learning model implements the idea of finding this classifier: it considers a fixed hypothesis (classifier) class together with a required accuracy and confidence, and takes into account the theory that characterizes when a function from this class can be learned from examples (training sample) in terms of a measure called the Vapnik–Chervonenkis dimension (VC dimension).

However, the statistical learning theory (Vapnik [1]) reveals that it is much more interesting not to preselect the class that will contain the target function to be learned. Instead, it is defined a set of hypothesis classes saved as a hierarchy, and the target function to be learned lies in one of them. The Structural Risk Minimization (SRM) copes with the problem of minimizing an upper bound on the expected risk over each of these hypothesis classes (Vapnik [2]). To implement the SRM in Support Vector Machines (SVM), one must consider the structures (classes) that control two factors that appear in the bound of the expected risk: the value of empirical risk and the complexity (the appropriate bound for the generalization error). Thus, under this principle, to select a learning algorithm, it is necessary to have the theoretical bound of the generalization error (PAC bounds) and to deal with the minimization of this bound together with the empirical risk.

In addition, standard regression methods are only concerned with the minimization of the empirical risk. This is the one based on the error produced by the regressors with respect to the observed dataset. This error is defined as the distance between the data to the approximation function; thus, it is a measure of the deviation of the data with respect to the regressors. It is characterized as a residual. The vertical distance is the most common way to measure the regression error, although it is not induced by a mathematical norm. In Support Vector Regression (SVR) (Vapnik [1]), for example, the residuals that participate in the empirical risk are measured through the vertical distance. Other distances, in this case based on a norm, have been used in order to establish these residuals, such as the l_1-*distance*, the l_2-*distance* and the l_∞-*distance* (Blanco et al. [3,4]).

The estimation of production functions and measures of efficiency and productivity have been the focus of a relatively large body of articles in the literature in both the economic and engineering contexts, as well as in operations research and statistics. In particular, Data Envelopment Analysis (DEA) (Charnes et al. [5] and Banker et al. [6]) is one of the existing techniques for estimating production functions and measuring efficiency. DEA relies on the construction of a polyhedral technology in the space of inputs and outputs that satisfies certain classical axioms of production theory (e.g., monotonicity and convexity). It is a non-parametric data-driven approach with many advantages from a benchmarking point of view. Additionally, the treatment of the multi-output multi-input framework is relatively straightforward with DEA, in comparison with other methods available. However, Data Envelopment Analysis has been criticized for its non-statistical nature, even being labeled as a pure descriptive tool of the data sample at a frontier level with little inferential power (its inferential power is exclusively based on the property of consistency and the increase in sample size instead of on the fundamentals of the method) (Esteve et al. [7]). DEA suffers from an overfitting problem because of the application of the minimal extrapolation principle, which places the estimator of the production function as close to the dataset as possible. This principle is also related to exclusively minimizing the empirical error (at a frontier level).

Regarding the literature related to this topic, some previous authors have tried to modify the standard DEA technique such that the new approaches work as inferential methods (with the focus on the DGP) rather than as mere descriptive tools. For example, Banker and Maindiratta [8] and Banker [9] associated DEA with maximum likelihood. Simar and Wilson [10–12] adapted bootstrapping to DEA. Kuosmanen and Johnson [13,14] introduced the Corrected Concave Nonparametric Least Squares. Unfortunately, despite the importance of machine learning techniques in the current literature, there have been few attempts to adapt DEA to the field of machine learning (see, for example, Esteve et al. [7], or Olesen and Ruggiero [15]). In this sense, our contribution could be seen as a new bridge between these two worlds: machine learning and efficiency measurement.

In this paper, our main objective is to propose, for the first time in the literature, a PAC bound in the context of the estimation of polyhedral technologies in microeconomics and engineering, enabling the possibility of controlling the generalization error of the estimation of the production frontier. Accordingly, we construct a model that controls the empirical error, together with the generalization error, through a PAC bound implementing the philosophy of Structural Risk Minimization by analogy with SVM. Our modeling has several implications:

(a) For the first time, a bound of the generalization error is implemented to determine the degree of technical inefficiency of a set of Decision Making Units (DMUs).
(b) We implement the minimization of the balance between the generalization error and the empirical error through a quadratic optimization model that will be called Data Envelopment Analysis-based Machines (DEAM), which has DEA as a particular case.
(c) Through a computational simulation experience, we show that DEAM outperform DEA regarding bias and mean squared error.
(d) We estimate production technologies using robust regression models that use the concept of margin. Due to that, the problem of efficiency measurement becomes a classification problem: to be efficient (being located within the margin) or not to be efficient (being located out of the margin).

Finally, we mention that the expected new insights gained by applying our approach (DEAM) are related to the determination of better estimates of production functions in engineering and microeconomics, in terms of bias and mean squared error. Additionally, these gains will also benefit the technical efficiency measures that can be derived from calculating the distance from a given observation to the production function estimate.

The rest of the paper is organized as follows. The following section provides the basic background. Next, in the third section, we introduce a new PAC for the class of piece-wise linear functions. In Section 4, a new approach called Data Envelopment Analysis-based Machines (DEAM) is defined and analyzed. Section 5 shows the main results associated with a computational experience for checking the new approach in comparison with DEA. Section 6 contains a discussion on the main results. Finally, the article ends with the conclusions section.

2. Background

In this section, we briefly introduce elemental notions of Support Vector Regression, Statistical Learning and Data Envelopment Analysis.

2.1. Support Vector Regression (SVR)

Machine learning (ML) is a methodology that studies computer processes that learn from experience and make improvements automatically. ML works with computer algorithms based on a learning sample (training data) and can make predictions about the behavior of future data. The study of this behavior is produced in two different scenarios: the scenario of supervised learning in which training data are vectors of predictors and responses, and the scenario of unsupervised learning, where no responses are considered in the data sample. In the first field, the objective of learning techniques is to determine the functional relationship between the predictors and the responses. In this case, the nature of the responses, if they come from a binary variable or are real values, determines the kind of problem to solve: a classification problem or a regression problem, respectively. In the second field, since there are no responses, the objective is to gain knowledge about the processes lying behind data generation, such as density estimation or clustering. Our paper largely focuses on the regression problem within supervised learning, bearing in mind that our data comprise inputs utilized by firms to produce outputs (real values).

Support Vector Machines (Vapnik [1,16]) is a technique that stands out in ML in the world of supervised learning. SVM represents an algorithm constructed on the foundations of statistical learning theory and is in line with the Structural Risk Minimization (SRM) method. SRM is implemented to construct support vector machines, where the objective is to control the value of empirical risk and the value of the VC dimension, which is the regularization term that appears when the generalization error must be minimized rather than minimizing only the empirical error (Vapnik [1,16]). In particular, the definition of the notion of the VC dimension is as follows:

Definition 1 (VC dimension). *Let H be a set of binary-valued functions. A set of points X is shattered by H if for all binary vectors b indexed by X; there is a function $f_b \in H$ performing b on X. The VC dimension, VC $dim(H)$, of the set H is the size of the largest shattered set.*

With regard to the classification problem (for the regression problem, the generalization error is defined in the same way, because the regression problem can be turned in to a classification problem, as we will go on to explain) in SVM, minimizing the generalization error consists of minimizing the probability of incorrectly classifying any new data that emerges from the unknown distribution that was generated by the learning sample. This aim is possible if a bound of the generalization error is found, and the parameters on which it is dependent are controlled to reduce the bound. These bounds are understood as Probably Approximately Correct (PAC) bounds, which were first proposed by Valiant [17]. The standard PAC learning implements the idea of finding this classifier: it considers a fixed hypothesis (classifier) class together with a required accuracy and confidence, and takes into account the theory that characterizes when a function from this class can be learned from examples (data). In the case of regression, the exercise involves converting the regression problem (estimation function) into a classification problem because bounds in the generalization error are precisely based on the VC dimension or when a margin is considered on the fat-shattering dimension (effective VC dimension).

Next, we show the definition of the fat-shattering dimension. Notice that bold will be utilized for denoting vectors, and non-bold for scalars.

Definition 2 (fat-shattering dimension). *Let F be a set of real-valued functions. A set of points X is γ-shattered by F if there are real numbers r_x indexed by $x \in X$ such that for all binary vectors b indexed by X, there is a function $f_b \in F$ such that*

$$f_b(\mathbf{x})\begin{cases} \geq r_x + \gamma & if\ b_x = 1 \\ \leq r_x - \gamma & otherwise \end{cases}.$$

The fat-shattering dimension of the set F, fat_F, is a function from the positive real numbers to the integers that maps a value γ to the largest γ-*shattered* set. The VC dimension corresponds to the largest shattered set, considering $\gamma = 0$, which is the concept first used by Vapnik to state a bound for the generalization error. This is the reason why the fat-shattering dimension is also known as the effective VC dimension.

To convert the regression problem into a classification problem, a threshold $\theta > 0$ that marks the limit needs to be set, such that a mistake will be considered to have been made if it is exceeded by the loss function when testing with new data in the model. The function that determines the distance between the real value of the output and the estimated value of said output through the model is called the loss function. Given a margin $\gamma > 0$, in the case of the training point, if the loss function exceeds the value $(\theta - \gamma)$, it will be considered as a mistake. Then, γ measures the discrepancy between the two losses: those measured on test data and those measured on training data. Under this re-interpretation of the regression problem, it is possible to use the dimension free bounds already constructed in the case of classification. In our case, we focus on the bound obtained by Shawe-Taylor and Cristianini [18], based on the fat-shattering dimension.

Theorem 1 (Shawe-Taylor and Cristianini [18]). *Let F be a sturdy class of real-valued functions with range $[-a, a]$ and fat-shattering dimension bounded by $fat_F(\gamma)$. Fix $\theta \in R$ with $\theta > 0$, and a scaling of the output range $\kappa \in R_+$. Consider a fixed but unknown probability distribution in the space $X \times R$. Then, with probability $1 - \rho$ over randomly drawn training sets S of size m for all γ with $\theta \geq \gamma > 0$, the probability that a training set filtered function $f \in F$ has an error larger than θ on a randomly chosen input is bounded by*

$$\varepsilon(m,d,\rho) = \frac{2}{m}\left(d\log_2\left(256m\left(\frac{c}{\gamma}\right)^2\right) \times \log_2\left(16em\left(\frac{c}{\gamma}\right)\right) + \log_2\left(\frac{16m^{1.5}a}{\rho\kappa}\right)\right) \quad (1)$$

where $c = \max\{a, D(S,f,\gamma)+\kappa\}$ *and*

$$d = \left\lceil fat_F(\gamma^-/16) + \left(\frac{16(D(S,f,\gamma)+\kappa)}{\gamma}\right)^2 \right\rceil, \quad (2)$$

provided $m \geq \frac{2}{\varepsilon}$.

In the statement of Theorem 1, $D(S,f,\gamma) = \sqrt{\sum_{(x,y)\in S} \xi((\boldsymbol{x},y),f,\gamma)^2} = \|\xi\|_2$ and $\xi((\boldsymbol{x},y),f,\gamma) = \max\{0, e(f)(\boldsymbol{x},y) - (\theta-\gamma)\}$, where $e(f)$ is the loss function that the analyst selects in order to measure how much f exceeds the error margin $(\theta-\gamma)$. In addition, the theorem introduces the concept of the fat-shattering dimension, $fat_F(\gamma)$, that is, the generalization of the VC dimension, which is sensitive to the size of the margin γ.

Theorem 1 is a general result, which in the case of each function class F, will be particularized: for each function class, the fat-shattering dimension is bounded in a different way, and consequently, the same happens with respect to the expected error proposed in (1). In the case of linear function classes, the fat-shattering dimension is bounded by Bartlett and Shawe-Taylor [19].

Theorem 2 (Bartlett and Shawe-Taylor [19]). *Suppose that X is a ball of radius r and center $\boldsymbol{0}_m$ in R^m, i.e., $X = \{\boldsymbol{x} \in R^m : \|\boldsymbol{x}\| \leq r\}$, and consider the set*

$$F = \{\boldsymbol{x} \to \boldsymbol{w}\cdot\boldsymbol{x} : \|\boldsymbol{w}\| \leq 1, \boldsymbol{x} \in X\},$$

Then

$$fat_F(\gamma) \leq \left(\frac{r}{\gamma}\right)^2.$$

The most general version of this theorem, in which $\|\boldsymbol{w}\|$ is not restricted to be at most 1, bounds the fat-shattering dimension of linear classifiers as $fat_F(\gamma) \leq \left(\frac{\|\boldsymbol{w}\|r}{\gamma}\right)^2$.

The following two previously published lemmas are significant for our purposes throughout this paper (see Bartlett and Shawe-Taylor [19]).

Lemma 1. *For every input set S γ-shattered by $F = \{\boldsymbol{x} \to \boldsymbol{w}\cdot\boldsymbol{x} : \boldsymbol{x} \in X\}$ (the linear hypothesis class) and for every subset $S_0 \subseteq S$, $\|\sum S_0 - \sum(S-S_0)\| \geq \frac{|S|\gamma}{\|\boldsymbol{w}\|}$ holds.*

Lemma 2. *For all $S \subseteq R^m_+$ with $\|\boldsymbol{x}\| \leq r$ for $\boldsymbol{x} \in S$, certain $S_0 \subseteq S$ satisfies that $\|\sum S_0 - \sum(S-S_0)\| \leq \sqrt{|S|}r$.*

Then, $\frac{|S|\gamma}{\|\boldsymbol{w}\|} \leq \|\sum S_0 - \sum(S-S_0)\| \leq \sqrt{|S|}r$. In particular, $\frac{|S|\gamma}{\|\boldsymbol{w}\|} \leq \sqrt{|S|}r$, and it is possible to conclude that all sets of inputs S *γ-shattered* by F are bounded. Therefore, the set *γ-shattered* by F with higher cardinality is also bounded, which is known as the fat-shattering dimension: $fat_F(\gamma) \leq \left(\frac{\|\boldsymbol{w}\|r}{\gamma}\right)^2$.

Now, if γ is fixed in such a way that $\theta \geq \gamma > 0$, and disregarding the logarithmic factors in (1), the only term to reduce the expected error is (2). This process can be performed by implementing the minimization of its bound, which in the case of linear functions, is as follows:

$$d = \left[fat_F(\gamma^-/16) + \left(\overbrace{\frac{16(D(S,f,\gamma)+\kappa)}{\gamma}}^{D} \right)^2 \right] \leq |w|^2 + CD^2. \quad (3)$$

This expected error bound meets the SRM objective: the minimization process leads to more than minimizing the empirical risk, i.e., $D^2 = \sum_{(x,y)\in S} \xi((x,y),f,\gamma)^2$. Instead, it minimizes the capacity of the estimation function to provide a suitable prediction when a new observation (out of sample) is introduced and that is given by the appearance of the regularization term, that is $\|w\|^2$, which bounds the fat-shattering dimension (PAC bound). The minimization of this bound corresponds to the objective of the regression problem associated with Support Vector Regression (SVR).

Support Vector Regression (SVR), as with any regression approach, attempts to construct a function that is capable of predicting the behavior of the response variable under the study. SVR sets out to predict the value of a continuous response variable $y \in R_+$ given a vector of covariables $x \in R^m_+$. Hence, SVR establishes a function $\hat{f} : R^m_+ \to R$ such that, given x, $\hat{f}(x)$ yields the response variable prediction. Under the SVR principle, the linear predictor $\hat{f}$ can be defined as $\hat{f}(x) = w^* \cdot x + b^*$, where $w^* \in R^m$ and $b^* \in R$ are optimal solutions of the optimization model below:

$$\begin{array}{lll} \underset{w,b,\xi'_i,\xi_i}{Min} & \|w\|^2 + C\sum_{i=1}^{n}(\xi'^2_i + \xi_i{}^2) & \\ & y_i - (w\cdot x_i + b) \leq \varepsilon + \xi'_i, & i = 1,\ldots,n \\ & (w\cdot x_i + b) - y_i \leq \varepsilon + \xi_i, & i = 1,\ldots,n \\ & \xi'_i, \xi_i \geq 0, & i = 1,\ldots,n \end{array} \quad (4)$$

In performing this methodology, the values of $C \in R_+$ and $\varepsilon \in R_+$ are obtained by a cross-validation process. The SVR yields an estimator $\hat{f}(x)$ of the response variable given x as well as lower and upper 'correcting' surfaces, defined as $\hat{f}(x) - \varepsilon$ and $\hat{f}(x) + \varepsilon$, where ε is a margin that enhances the estimator linked to SVR with robustness (see Figure 1). Additionally, observations below the surface $\hat{f}(x) - \varepsilon$ reveal an associated (empirical) error of $\xi_i > 0$ (with $\xi'_i = 0$), while observations above the surface $\hat{f}(x) + \varepsilon$ present an (empirical) error of $\xi'_i > 0$ (with $\xi_i = 0$). Observations between the surfaces $\hat{f}(x) - \varepsilon$ and $\hat{f}(x) + \varepsilon$ reveal an error of zero (with $\xi_i = \xi'_i = 0$). The objective function, however, represents the combination of regression and regularization involved in SVR, combining the empirical error term $\sum_{i=1}^{n}(\xi'^2_i + \xi_i{}^2)$ and the regularization term $\|w\|^2$ through a weight C, thus balancing both components (Vazquez and Walter [20]). Moreover, although hyperplanes are linear in shape, it must be highlighted that SVR is able to generate estimation functions that are not necessarily linear in the original (x,y) space, and that can be achieved by using a transformation function ϕ, a conversion arising from the covariable space, $\phi : R^m_+ \to Z$. Figure 1 shows the solution of the linear estimator achieved by an SVR model, as well as the graphical representation of the residuals (empirical error) for two points and the hyperplanes that define the margins (dashed lines).

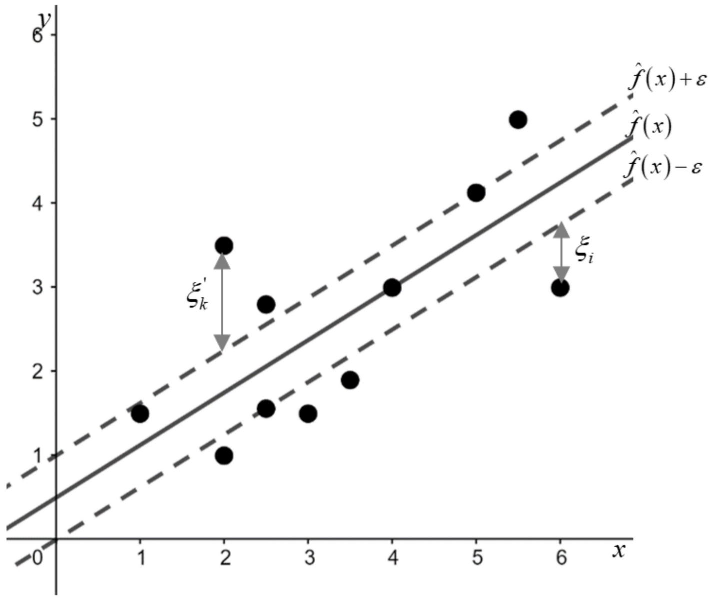

Figure 1. Support Vector Regression.

The next subsection explains how Data Envelopment Analysis (DEA) works.

2.2. Data Envelopment Analysis (DEA)

Let us consider the observation of n Decision Making Units (DMUs). DMU_i takes up $\boldsymbol{x}_i = \left(x_i^{(1)}, \ldots, x_i^{(m)}\right) \in R_+^m$ amounts of inputs to generate $\boldsymbol{y}_i = \left(y_i^{(1)}, \ldots, y_i^{(s)}\right) \in R_+^s$ amounts of outputs. The relative efficiency of each unit in the sample is evaluated by referring to the so-called production possibility set or technology, which is essentially the set of producible bundles of $(\boldsymbol{x}, \boldsymbol{y})$. It is generally defined as:

$$T = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in R_+^{m+s} : \boldsymbol{x} \text{ can produce } \boldsymbol{y}\right\} \tag{5}$$

Under Data Envelopment Analysis (DEA) (Charnes et al. [5] and Banker et al. [6] and more recently, Villa et al. [21], Sahoo et al. [22], and Amirteimoori [23]), T is usually assumed to satisfy free disposability with regard to inputs and outputs; that is, if $(\boldsymbol{x}, \boldsymbol{y}) \in T$, then $(\boldsymbol{x}', \boldsymbol{y}') \in T$ with $\boldsymbol{x}' \geq \boldsymbol{x}$ and $\boldsymbol{y}' \leq \boldsymbol{y}$. Convexity of T is also generally assumed (see, e.g., Färe and Primont [24]).

Insomuch as the measurement of technical efficiency is concerned, a certain subset of T is of interest. We allude to the weakly efficient set of T, defined as $\partial^W(T) := \{(\boldsymbol{x}, \boldsymbol{y}) \in T : \hat{\boldsymbol{x}} < \boldsymbol{x}, \hat{\boldsymbol{y}} > \boldsymbol{y} \Rightarrow (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \notin T\}$ (Let $\boldsymbol{z} = \left(z^{(1)}, \ldots, z^{(q)}\right)$ and $\boldsymbol{t} = \left(t^{(1)}, \ldots, t^{(q)}\right)$. Then, $\boldsymbol{z} < \boldsymbol{t}$ means $z^{(j)} < t^{(j)}$ for all $j = 1, \ldots, q$.). Some authors (see, for example, Briec and Lesourd [25]) define technical efficiency as the distance from a point in T to the weakly efficient set.

When $s = 1$, this context is confined to the central concept of production function f. Accordingly, m input variables are used to yield a univariate output, and hence, we can define the technology as:

$$T = \left\{(\boldsymbol{x}, y) \in R_+^{m+1} : y \leq f(\boldsymbol{x})\right\}.$$

According to the selected distance for measuring technical inefficiency, different DEA models emerge (Cooper et al. [26]). The directional distance function (DDF) is a relevant example of them. For m inputs and one output, resorting to the directional vector

$\boldsymbol{g} = (\boldsymbol{g}^-, g^+)$, where $\boldsymbol{g}^- = \mathbf{1}_m$ and $g^+ = 1$, the DDF problem has the following structure when the efficiency level of DMU_i is assessed, $i = 1, \ldots, n$:

$$\begin{array}{cl} \underset{\beta_i,\lambda_1,\ldots,\lambda_n}{Max} & \beta_i \\ s.t. & \\ & \sum\limits_{k=1}^{n} \lambda_k x_k^{(j)} \le x_i^{(j)} - \beta_i, \quad \forall j = 1, \ldots, m \\ & \sum\limits_{k=1}^{n} \lambda_k y_k \ge y_i + \beta_i, \\ & \sum\limits_{k=1}^{n} \lambda_k = 1, \\ & \lambda_k \ge 0 \qquad\qquad\qquad \forall k = 1, \ldots, n \end{array} \tag{6}$$

Given that (6) is a linear program, we can equivalently solve its corresponding dual formulation:

$$\begin{array}{cl} \underset{\boldsymbol{c}_i,p_i,\alpha_i}{Min} & -p_i y_i + \boldsymbol{c}_i \boldsymbol{x}_i + \alpha_i \\ s.t. & \\ & p_i y_k - \boldsymbol{c}_i \boldsymbol{x}_k - \alpha_i \le 0, \quad \forall k = 1, \ldots, n \\ & \|(\boldsymbol{c}_i, p_i)\|_1 = 1, \\ & p_i \ge 0, \\ & c_i^{(j)} \ge 0, \qquad\qquad \forall j = 1, \ldots, m \end{array} \tag{7}$$

DEA models must be solved for each DMU_i, $i = 1, \ldots, n$, in the sample.

Figure 2 shows an example of the DDF model with a distance vector $\boldsymbol{g} = (\boldsymbol{g}^-, g^+) = (\mathbf{1}_m, 1)$. Note that DEA generates a piece-wise linear technology (the region below the line), satisfying free disposability in inputs and outputs and convexity. Note also that the DEA estimate envelops all the observations from above. In this case, with $\boldsymbol{g} = (\mathbf{1}_m, 1)$, the DDF coincides with a particular distance between data and $\partial^W(T)$: the l_∞-distance (Briec [27] and Briec and Lesourd [25]).

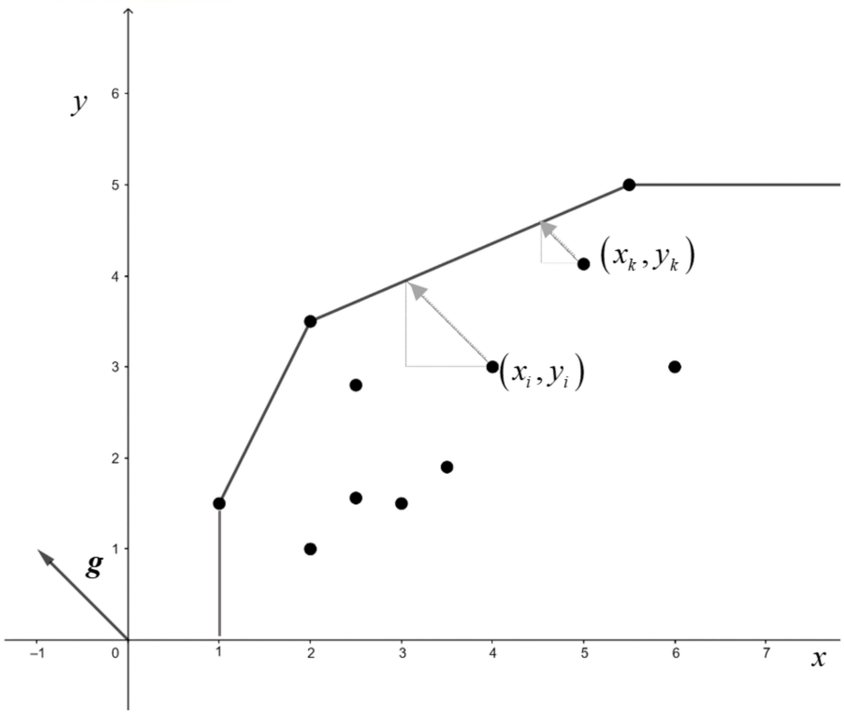

Figure 2. Illustration of the Directional Distance Function in Data Envelopment Analysis.

In this paper, our purpose is to construct a method that generates piece-wise linear frontiers as in Figure 2, by implementing the minimization of the generalization error of the model.

3. New PAC Learning with Piece-Wise Linear Hypothesis

This section revolves around two stages in the search for the generalization error bound: the first stage is based on the construction of the class of piecewise linear hypotheses whose elements are hyperplanes that are located as close as possible to the data sample through l_∞-*distance*, and the second stage is based on the construction of the bound of the fat-shattering dimension of the class of hypothesis constructed in the first stage. The minimization of the bound of the expected error using the bound of the fat-shattering dimension calculated gives rise to the Data Envelopment Analysis-based Machines (DEAM) model as a method for estimating piecewise linear production functions, which minimizes the generalization error as well as the empirical error.

To obtain this bound of the class of functions of our interest, we must derive the fat-shattering dimension bound for the hypothesis class with the piece-wise structure we desire. Then, minimizing the generalization error will be implemented through the minimization of the fat-shattering dimension bound. For this task, a previous step must be taken: a class of piece-wise linear hypothesis must be defined. A piece-wise linear hypothesis target is defined by a combination of n hyperplanes $\{H_p\}_{p=1,\dots,n}$ that are selected to evaluate the data depending on their input values. The hyperplanes will be defined for each input value $\boldsymbol{x} \in R^m_+$ as follows:

$$H_{p_x} = \left\{(\boldsymbol{x}, y) \in R^{m+1} : \boldsymbol{w}_{p_x}\boldsymbol{x} + \beta_{p_x} - \delta_{p_x} y = 0\right\}$$

Then, if we suppose $\delta_{p_x} > 0, \forall p_x \in \{1, \dots, n\}$, each output value estimation through the set of n hyperplanes $\{H_p\}_{p=1,\dots,n}$ can be written as a function of the input value vector $\boldsymbol{x}$:

$$h(\boldsymbol{x}) = \tfrac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \tfrac{\beta_{p_x}}{\delta_{p_x}}, \quad p_x \in \{1, \dots, n\} \ ,$$

with $\boldsymbol{w}_{p_x} \in R^m_+$ and $\beta_{p_x} \in R$. The value of $p_x \in \{1, \dots, n\}$, in our case, is chosen by considering two desired conditions that are inherited from production theory:

$$h(\boldsymbol{x}) = \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} \geq 0, \tag{8}$$

and

$$\tfrac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \tfrac{\beta_{p_x}}{\delta_{p_x}} \leq \tfrac{\boldsymbol{w}_{p}}{\delta_{p}}\boldsymbol{x} + \tfrac{\beta_{p}}{\delta_{p}}, \quad \forall p \in \{1, \dots, n\} \ . \tag{9}$$

Condition (8) ensures that the estimation of the output value associated with an input $\boldsymbol{x} \in R^m_+$ will be always non-negative. Additionally, condition (9) guarantees that the estimation $h(\boldsymbol{x})$ through the hyperplane H_{p_x} is less or equal than the estimation through any other hyperplane H_p. Condition (9) is the one that imposes concavity on the model. This type of condition was the key for stating concavity in the general multiple-regressor modeling in microeconomics (Afriat [28]; Kuosmanen et al. [13]). In particular, if the production function is concave, then the technology defined from this production function is convex.

The function class of piece-wise linear hypothesis can be constructed as follows:

$$\begin{aligned} F = \ & \Big\{x \mapsto \tfrac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \tfrac{\beta_{p_x}}{\delta_{p_x}} : \|\boldsymbol{x}\| \leq R, \tfrac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \tfrac{\beta_{p_x}}{\delta_{p_x}} \geq 0, \\ & \tfrac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \tfrac{\beta_{p_x}}{\delta_{p_x}} \leq \tfrac{\boldsymbol{w}_{p}}{\delta_{p}}\boldsymbol{x} + \tfrac{\beta_{p}}{\delta_{p}}, \forall p_x, p \in \{1, \dots, n\}\Big\}. \end{aligned} \tag{10}$$

Now, we can proceed with the second step: to establish a bound for the fat-shattering dimension of this function class to control the generalization error. Before proving the main theorem of this section, we need to state a necessary technical lemma. In the results, $r \in R_+$ is the radius of the ball centered in $\mathbf{0}_m$ that bounds the input data in the data sample.

Lemma 3. *If an input learning sample,* $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d\}$ *is* γ*-shattered through F defined in (10), then every subset* $S_0 \subseteq S$ *satisfies*

$$\left\|\sum S_0 - \sum(S - S_0)\right\| \geq |S| \left(\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{\boldsymbol{w}_p}{\delta_p}\right\|} - 2r \right), \tag{11}$$

denoting as $\sum S_0$ and $\sum(S - S_0)$ the sum of the elements in S_0 and $S - S_0$, respectively, and as $|S|$ the cardinal of the set S.

Proof. See Appendix A. □

Next, we prove the main theorem of this section. In particular, we state the bound for the fat-shattering dimension for piece-wise linear hypothesis classes.

Theorem 3. *Let X be the ball of radius R and center* $\mathbf{0}_m$ *in* R^m*, i.e.,* $X = \{\boldsymbol{x} \in R^m : \|\boldsymbol{x}\| \leq r\}$*, and let the hypothesis class be as follows*

$$F = \left\{ x \mapsto \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} : \|x\| \leq R, \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} \geq 0, \right. \\ \left. \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} \leq \frac{\boldsymbol{w}_p}{\delta_p}\boldsymbol{x} + \frac{\beta_p}{\delta_p}, \forall p_x, p \in \{1, \ldots, n\} \right\}, \tag{12}$$

then

$$fat_F(\gamma) \leq \left(\frac{r}{\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r} \right)^2. \tag{13}$$

Proof. See Appendix A. □

The next section involves the task of achieving a model that minimizes the established generalization error through the l_∞-distance.

4. Data Envelopment Analysis-Based Machines (DEAM)

Data Envelopment Analysis-based Machines (DEAM) can be defined from the idea of minimizing the expected error proposed in (1). If we do not consider the logarithmic factors, we can directly focus on minimizing d in this expression, for which a bound on the generalization error has been found in the case of the piece-wise linear hypothesis class F defined in (12):

$$d = fat_F(\gamma^-/16) + \left(\frac{16(D(S,f,\gamma)+\kappa)}{\gamma}\right)^2 \leq \underbrace{\left(\frac{r}{\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\frac{\gamma}{16} - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r} \right)^2}_{A} + \underbrace{\left(\frac{16(D(S,f,\gamma)+\kappa)}{\gamma}\right)^2}_{B} \tag{14}$$

Because of the complexity of implementing an optimization model in which the objective function has the aim of minimizing the above bound, we will break up the minimization of the whole bound into different objectives, which will be collected in an aggregation function that will conform the objective function of the final optimization program associated with DEAM, which will be shown later in this section.

Once the number of different hyperplanes in each hypothesis is set as the number of elements in the learning sample $|S| = n$, minimizing the bound of the fat-shattering dimension requires minimizing part A in (14). This is equivalent to maximizing $\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\frac{\gamma}{16}-\frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|}$. Regarding this last expression, we must maximize the numerator and minimize the denominator, as follows:

(i) The vector of coefficients (slopes) corresponding to the hyperplane H_p is $(\boldsymbol{w}_p, \delta_p)$. We can consider, without loss of generality, that $\|(\boldsymbol{w}_p, \delta_p)\|_1 = 1$. Then, minimizing $\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|$, is equivalent to minimizing $\underset{p\in\{1,\ldots,n\}}{Max}\left(\frac{1}{\frac{1}{\|w_p\|}-1}\right)$ since $\delta_p \geq 0$, $\forall p \in \{1,\ldots,n\}$. Focusing on that last equivalence, this objective can be directly translated into minimizing $\underset{p\in\{1,\ldots,n\}}{Max}\|\boldsymbol{w}_p\|$.

(ii) Maximizing $\underset{p\in\{1,\ldots,n\}}{Min}\left\{\frac{\gamma}{16}-\frac{\beta_p}{\delta_p}\right\}$ with a fixed value of the margin γ is equivalent to minimizing $\underset{p\in\{1,\ldots,n\}}{Max}\left\{\frac{\beta_p}{\delta_p}\right\}$. Because of $\|(\boldsymbol{w}_p, \delta_p)\|_1 = 1$, by minimizing $\underset{p\in\{1,\ldots,n\}}{Max}\|\boldsymbol{w}_p\|$ in (i), at the same time, the maximization of the elements $\{\delta_p\}_{p\in\{1,\ldots,n\}}$ is achieved. In this way, it is only necessary to minimize $\underset{p\in\{1,\ldots,n\}}{Max}\{\beta_p\}$ to maximize $\underset{p\in\{1,\ldots,n\}}{Max}\left\{\frac{\beta_p}{\delta_p}\right\}$.

Finally, a way of implementing (i) and (ii) is minimizing $u+v$, where $\|\boldsymbol{w}_p\| \leq u$, $\beta_p \leq v, \forall p \in \{1,\ldots,n\}$. Accordingly, minimizing the bound of the fat-shattering dimension, $A+B$, leads to minimizing $(u+v)+CD^2$, where $D^2 = D(S,f,\gamma)^2 = \|\xi\|_2^2$ and C is a parameter to be tuned by, for example, a cross-validation process. As a loss function we use the following: $\xi((\boldsymbol{x},y),f,\gamma) = \max\left\{0, D_{\|\cdot\|_\infty}((\boldsymbol{x},y),f)-(\theta-\gamma)\right\}$. Finally, the objective function has the following structure:

$$z(u,v,\xi_1,\ldots,\xi_n) = u+v+C\|\xi\|_2^2. \tag{15}$$

Accordingly, we introduce the optimization model that defines DEAM:

$$\begin{array}{lll}
\underset{w,\beta,\delta,\xi,\xi',u,v}{Min} & u+v+C\left(\sum\limits_{i=1}^{n}\xi_i^2+\sum\limits_{i=1}^{n}\xi_i'^2\right) & \\
s.t. & & \\
& \|\boldsymbol{w}_i\|_1 \leq u & \forall i=1,\ldots,n \quad (16.1)\\
& \beta_i \leq v & \forall i=1,\ldots,n \quad (16.2)\\
& \delta_p y_i \leq \boldsymbol{w}_p\boldsymbol{x}_i+\beta_p & \forall i,p=1,\ldots,n \quad (16.3)\\
& \boldsymbol{w}_i,\delta_i \geq 0 & \forall i=1,\ldots,n \quad (16.4)\\
& \boldsymbol{w}_i\boldsymbol{x}_i+\beta_i-\delta_i y_i \leq \varepsilon+\xi_i & \forall i=1,\ldots,n \quad (16.5)\\
& \delta_i y_i-\boldsymbol{w}_i\boldsymbol{x}_i-\beta_i \leq \varepsilon+\xi_i' & \forall i=1,\ldots,n \quad (16.6)\\
& \xi_i,\xi_i' \geq 0 & \forall i=1,\ldots,n \quad (16.7)\\
& \|(\boldsymbol{w}_i,\delta_i)\|_1 = 1 & \forall i=1,\ldots,n \quad (16.8)\\
& \boldsymbol{w}_i\boldsymbol{x}_i+\beta_i-\delta_i y_i \leq \boldsymbol{w}_p\boldsymbol{x}_i+\beta_p-\delta_p y_i & \forall i,p=1,\ldots,n \quad (16.9)
\end{array} \tag{16}$$

Model (16) determines a maximum of n different hyperplanes. The intersection of the half-spaces defined from these hyperplanes gives rise to the estimator of the underlying (convex) production technology. The number of hyperplanes to be considered in the implementation of the DEAM model can be seen as a key parameter of our approach since the results could be different depending on it. However, we suggest using n hyperplanes, which coincide with the number of DMUs. This is due to the experimental evidence found in the simulation study carried out in Section 5. We analyzed 2000 databases, and in all these cases, the number of hyperplanes at optimum were less than the number of DMUs in the corresponding data sample. This situation can be identified because some hyperplanes are repeated at the optimal solution of each problem.

Let us now explain each constraint of model (16) in detail. Constraints (16.1) and (16.2) come from $\|\mathbf{w}_p\| \leq u$, $\beta_p \leq v, \forall p = 1, \ldots, n$, respectively. The norm l_1 is used to be consistent with constraint (16.8). Additionally, this type of norm is associated with the definition of linear constraints, which are easier to be solved from a computational point of view. Constraint (16.3) is equivalent to $y_i \leq \frac{w_p}{\delta_p}\mathbf{x}_i + \frac{\beta_p}{\delta_p}$, $i, p = 1, \ldots, n$, i.e., it ensures that the hyperplanes envelop the data sample from above. Condition (16.4) forces that the n hyperplanes are monotonic non-decreasing and will be responsible for the satisfaction of the property of free disposability, as we will show later in the text (see Proposition 2 below). Constraints (16.5), (16.6), (16.7), and (16.8) allow for characterizing $\xi((\mathbf{x}, y), f, \gamma)$ as $\max\left\{0, D_{\|\cdot\|_\infty}((\mathbf{x}, y), f) - (\theta - \gamma)\right\}$. The parameter ε $(= \theta - \gamma \geq 0)$ will be chosen by cross validation. Let us now interpret specifically the value at optimum of the decision variable ξ_i. Let us pay attention to constraint (16.5). If $\mathbf{w}_i\mathbf{x}_i + \beta_i - \delta_i y_i - \varepsilon \geq 0$, then $\xi_i = \mathbf{w}_i\mathbf{x}_i + \beta_i - \delta_i y_i - \varepsilon$ since $\sum_{i=1}^{n} \xi_i^2$ is minimized in the objective function. In this way, considering (16.8), (16.3) and $\varepsilon \geq 0$, ξ_i can be interpreted as the l_∞-distance from the observation $(\mathbf{x}_i, y_i)$ to the hyperplane $H_{i\varepsilon}$:

$$\xi_i = \frac{|\mathbf{w}_i\mathbf{x}_i + \beta_i - \delta_i y_i - \varepsilon|}{\|(\mathbf{w}_i, \delta_i)\|_1} = D_{l_\infty}((\mathbf{x}_i, y_i), H_{i\varepsilon}),$$

where $H_{i\varepsilon} = \left\{(\mathbf{x}, y) \in R^{m+1} : \mathbf{w}_i\mathbf{x} + \beta_i - \delta_i y - \varepsilon = 0\right\}$ (Mangasarian [29]). If $\mathbf{w}_i\mathbf{x}_i + \beta_i - \delta_i y_i - \varepsilon < 0$, then $\xi_i = 0$ by (16.7) and the minimization of $\sum_{i=1}^{n} \xi_i^2$. Additionally, regarding the value of ξ_i', $i = 1, \ldots, n$, by constraints (16.3), (16.6), $\varepsilon \geq 0$ and the minimization of $\sum_{i=1}^{n} \xi'^2_i$, we obtain $\xi_i' = 0$ for all $i = 1, \ldots, n$ at optimum. This point has computational implications on the model since constraint (16.6) can be removed from it because (16.3) holds. Finally, constraint (16.9) guarantees that, for each $(\mathbf{x}_i, y_i)$ in the data sample, the hyperplane of the piece-wise linear production function associated with that point is the closest one to $(\mathbf{x}_i, y_i)$. Note that constraint (16.9), by (16.3) and (16.8), is equivalent to writing $D_{l_\infty}((\mathbf{x}_i, y_i), H_i) \leq D_{l_\infty}((\mathbf{x}_i, y_i), H_p)$ $\forall i, p = 1, \ldots, n$ (see Mangasarian [29]).

Figure 3 shows the shape of the function that will be generated by the model as an estimate of the underlying production function. Note that the estimate satisfies monotonicity and concavity, as happens with the DEA estimator. However, the DEAM estimator does not satisfy minimal extrapolation. Additionally, it implements a certain idea of robustness because of the margin notion inherited from SVR. Additionally, Figure 3 shows the possible interpretation of ξ_i as $D_{l_\infty}((\mathbf{x}_i, y_i), H_{i\varepsilon})$. In particular, ξ_i is the 'radius' of the squared ball in the figure.

As the technology generated by DEA, DEAM provides a piece-wise linear technology that can be defined as $T_{DEAM} := \left\{(\mathbf{x}, y) \in R_+^{m+1} : \mathbf{w}_p^*\mathbf{x} + \beta_p^* - \delta_p^* y \geq 0, \forall p \in \{1, \ldots, n\}\right\}$, given an optimal solution $\left(\left\{\mathbf{w}_p^*, \beta_p^*, \delta_p^*, \xi_p^*, \xi_p^{*\prime}\right\}_{p=1,\ldots,n}, u^*, v^*\right)$ of model (16).

The next propositions state that the derived technology from model (16) satisfies convexity and free disposability.

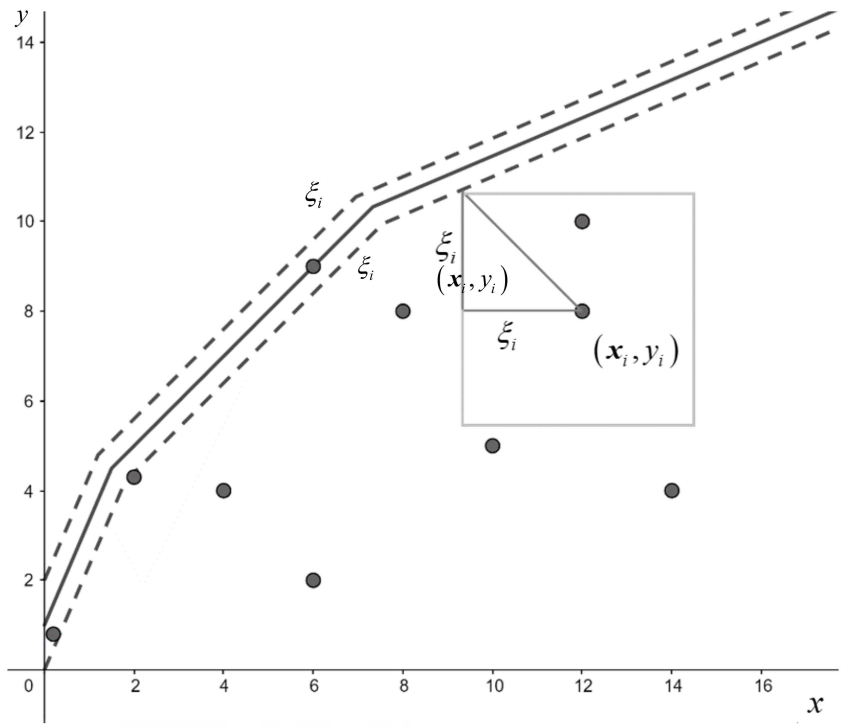

Figure 3. Illustration of the DEAM estimation of a production function.

Proposition 1. *T_{DEAM} is a convex set.*

Proof. The intersection of half-spaces is a convex set. □

Proposition 2. *T_{DEAM} satisfies free disposability in inputs and outputs.*

Proof. The result holds because $\boldsymbol{w}_p, \delta_p \geq 0$ for all $p \in \{1, \ldots, n\}$ (see Kuosmanen and Johnson [13]). □

Additionally, by constraint (16.3), we have that $\boldsymbol{w}_p^* \boldsymbol{x}_i + \beta_p^* - \delta_p^* y_i \geq 0, \forall i, p = 1, \ldots, n$. Therefore, for any observation $(\boldsymbol{x}_{i'}, y_{i'})$, we have that $\boldsymbol{w}_p^* \boldsymbol{x}_{i'} + \beta_p^* - \delta_p^* y_{i'} \geq 0, \forall p = 1, \ldots, n$, which implies that $(\boldsymbol{x}_{i'}, y_{i'}) \in T_{DEAM}$ since $T_{DEAM} = \Big\{ (\boldsymbol{x}, y) \in R_+^{m+1} : \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \delta_p^* y \geq 0,$ $\forall p \in \{1, \ldots, n\} \Big\}$. In this way, we can establish the following corollary.

Corollary 1. *The production possibility set generated by DEA is a subset of the production possibility set generated by DEAM.*

Proof. The result holds because the production possibility set generated by DEA and the production possibility set yielded by DEAM satisfy convexity, free disposability, and contain all observations, but only the technology related to DEA meets minimal extrapolation. □

In this way, we have that DEAM does not satisfy the minimal extrapolation principle, but its associated estimation of the technology always contains the observations.

As for the measurement of technical inefficiency of the observations, due to the nature of the technique used and based on the original ideas derived from Support Vector Regression, any $(\boldsymbol{x}_i, y_i)$ located within the margin will be identified as technically efficient (with $\xi_i^* = 0$). Otherwise, i.e., if $(\boldsymbol{x}_i, y_i)$ is located below the margin (see Figure 3), we have that ξ_i^* is the l_∞-distance from the observation to the (efficient) frontier of a 'robust' technology. This robust technology is defined by the translation of the original technology T_{DEAM} downward following the value of the margin ε. If we define this translated technology as $T^{\varepsilon}_{DEAM} = \left\{(\boldsymbol{x}, y) \in R^{m+1}_{+} : \boldsymbol{w}^*_p\boldsymbol{x} + \beta^*_p - \delta^*_p y - \varepsilon \geq 0, \forall p \in \{1, \ldots, n\}\right\}$, then $\xi^*_i = D_{l_\infty}((\boldsymbol{x}_i, y_i), \partial^W(T^{\varepsilon}_{DEAM}))$ (this result can be derived from Aparicio and Pastor [30]).

Now, we show the relationship between the Directional Distance Function (DDF) in DEA, model, and the DEAM model (16): The DDF model always yields a feasible solution of the model associated with Data Envelopment Analysis-based Machines.

Theorem 4. *Let* $\{(\boldsymbol{c}^*_i, \alpha^*_i, p^*_i)\}_{i=1,\ldots,n}$ *be a set of optimal solutions of model (7) for each* DMU_i, *i* = 1, ..., *n. Then,* $\left\{(\boldsymbol{c}^*_i, \alpha^*_i, p^*_i, \vartheta^*_i, \vartheta'^*_i)_{i=1,\ldots,n}, a^*, b^*\right\}$, *with* $\vartheta^*_i = -p^*_i y_i + \boldsymbol{c}^*_i\boldsymbol{x}_i + \alpha^*_i$, $\vartheta'^*_i = p^*_i y_i - \boldsymbol{c}^*_i\boldsymbol{x}_i - \alpha^*_i = 0$, $\forall i = 1, \ldots, n$, $a^* = \underset{i=1,\ldots,n}{max}\|\boldsymbol{c}^*_i\|$, $b^* = \underset{i=1,\ldots,n}{max}\{\alpha^*_i\}$ *is a feasible solution of model (16).*

Proof. Let $\{(\boldsymbol{c}^*_i, \alpha^*_i, p^*_i)\}_{i=1,\ldots,n}$ be a set of optimal solutions of model (7) for each DMU_i, *i* = 1, ..., *n*. By the characterization of a^* and b^* as $a^* = \max_{i=1,\ldots,n}\|c^*_i\|$ and $b^* = \max_{i=1,\ldots,n}\{\alpha^*_i\}$, the following inequalities are true:

$$\|\boldsymbol{c}^*_i\| \leq a^* \tag{17}$$

$$\alpha^*_i \leq b^* \tag{18}$$

Then, $\left\{(\boldsymbol{c}^*_i, \alpha^*_i, p^*_i, \vartheta^*_i, \vartheta'^*_i)_{i=1,\ldots,n}, a^*, b^*\right\}$ satisfies (16.1) and (16.2) in the DEAM model. Because of the fact that $\{(c^*_i, \alpha^*_i, p^*_i)\}_{i=1,\ldots,n}$ is a set of optimal solutions of model (7) for each DMU_i, *i* = 1, ..., *n*, the constraints of this model are satisfied for this solution:

$$\begin{aligned} & p^*_i y_k \leq \boldsymbol{c}^*_i\boldsymbol{x}_k + \alpha^*_i \quad \forall k = 1, \ldots, n; \forall i = 1, \ldots, n && (19.1) \\ & \|(\boldsymbol{c}^*_i, p^*_i)\|_1 = 1 \quad \forall i = 1, \ldots, n && (19.2) \\ & \boldsymbol{c}^*_i, p^*_i \geq 0 \quad \forall i = 1, \ldots, n && (19.3) \end{aligned} \tag{19}$$

Then, $\left\{(\boldsymbol{c}^*_i, \alpha^*_i, p^*_i, \vartheta^*_i, \vartheta'^*_i)_{i=1,\ldots,n}, a^*, b^*\right\}$ trivially satisfies (16.3), (16.4) and (16.8) in the DEAM model. Because of the definition of the variables ϑ^*_i and ϑ'^*_i as $\vartheta^*_i = -p^*_i y_i + \boldsymbol{c}^*_i\boldsymbol{x}_i + \alpha^*_i$, $\vartheta'^*_i = p^*_i y_i - \boldsymbol{c}^*_i\boldsymbol{x}_i - \alpha^*_i = 0$, $\forall i = 1, \ldots, n$, we have that:

$$-p_i y_i + \boldsymbol{c}_i\boldsymbol{x}_i + \alpha_i \leq \vartheta^*_i + \varepsilon \tag{20}$$

and

$$p^*_i y_i - \boldsymbol{c}^*_i\boldsymbol{x}_i - \alpha^*_i \leq \vartheta'^*_i + \varepsilon, \tag{21}$$

$\forall i = 1, \ldots, n$, and $\forall \varepsilon \geq 0$. Then, $\left\{(\boldsymbol{c}^*_i, \alpha^*_i, p^*_i, \vartheta^*_i, \vartheta'^*_i)_{i=1,\ldots,n}, a^*, b^*\right\}$ satisfies (16.5) and (16.6) in the DEAM model. Additionally, we have

$$0 \leq -p^*_i y_i + \boldsymbol{c}^*_i\boldsymbol{x}_i + \alpha^*_i = \vartheta^*_i \quad \forall i = 1, \ldots, n \tag{22}$$

and,

$$0 \leq p^*_i y_i - \boldsymbol{c}^*_i\boldsymbol{x}_i - \alpha^*_i = \vartheta'^*_i \quad \forall i = 1, \ldots, n\ . \tag{23}$$

Constraint (22) is satisfied by (19.1), and (23) is trivially satisfied. Then, $\left\{(c^*_i, \alpha^*_i, p^*_i, \vartheta^*_i, \vartheta'^*_i)_{i=1,\ldots,n}, a^*, b^*\right\}$ satisfies (16.7) in the DEAM model. Finally, the objective in (7) is to minimize $\vartheta_i = -p_i y_i + c_i x_i + \alpha_i$, $\forall i = 1, \ldots, n$, that implies

$$\vartheta_i^* = -p_i^* y_i + \boldsymbol{c}_i^* \boldsymbol{x}_i + \alpha_i^* \leq -p_k y_i + \boldsymbol{c}_k \boldsymbol{x}_i + \alpha_k \quad \forall k = 1, \ldots, n \ . \tag{24}$$

Then, $\left\{ \left(\boldsymbol{c}_i^*, \alpha_i^*, p_i^*, \vartheta_i^*, \vartheta_i'^* \right)_{i=1,\ldots,n}, a^*, b^* \right\}$ satisfies (16.9) in the DEAM model. Consequently, $\left\{ \left(\boldsymbol{c}_i^*, \alpha_i^*, p_i^*, \vartheta_i^*, \vartheta_i'^* \right)_{i=1,\ldots,n}, a^*, b^* \right\}$ is a feasible solution of (16). □

However, it can be shown that the DDF model (7) does not always yield an optimal solution of model (16).

5. Computational Experience

This section compares the performance of DEA and DEAM for estimating production functions. For this task, we designed five typical production scenarios in Table 1.

Table 1. Simulated scenarios.

Scenario	Inputs	Production Function
I	x_1	$y = {x_1}^{0.5}$
II	x_1, x_2	$y = x_1^{0.35} \cdot x_2^{0.15}$
III	x_1, x_2, x_3	$y = x_1^{0.30} \cdot x_2^{0.15} \cdot x_3^{0.05}$
IV	x_1, x_2, x_3, x_4	$y = x_1^{0.25} \cdot x_2^{0.15} \cdot x_3^{0.05} \cdot x_4^{0.05}$
V	x_1, x_2, x_3, x_4, x_5	$y = x_1^{0.25} \cdot x_2^{0.10} \cdot x_3^{0.05} \cdot x_4^{0.05} \cdot x_5^{0.05}$

The simulations implement Cobb–Douglas production functions, which are frequently used in econometrics for establishing the relation between the maximum amount of outputs that can be produced from a set of inputs. Thereby, scenario I implements a mono-input mono-output case, while the other scenarios represent multi-input mono-output cases. For each scenario, we ran 100 trials $(t = 1, \ldots, 100)$ with sample sizes: $n \in \{25, 50, 75, 100\}$. The inputs were calculated randomly from $Uni[1, 10]$. For simulating inefficiencies, we selected a random distribution $\exp(1/3)$ for u. Mean squared error (MSE) and bias were the two measures employed to assess the performance of each method.

The DEAM model (16), as other machine learning techniques, needs to find the best model through a cross-validation process. For this task and exclusively for the DEAM model, we implemented a five-fold cross validation using a certain grid of hyperparameters. This grid was arbitrarily set as: $C \in \{1,\ 10,\ 50,\ 100,\ 10^6\}$ and $\varepsilon \in \{0,\ 0.001,\ 0.01,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 1\}$. Note that DEA does not need to apply a cross-validation process. Instead, DEA uses the whole dataset to evaluate efficiency scores.

Table 2 sums up the results obtained for each scenario when DEA (without cross validation) and DEAM (with cross validation) are applied. The first two columns present the type of scenario and the sample size. The following columns show the mean and standard deviation (in brackets) of MSE obtained by DEA and DEAM. Fraction of trial reports the proportion of trials in which DEAM either improves upon or equals the MSE given by the DEA method, while the next column illustrates the percentage of improvement of DEAM with respect to DEA. The four subsequent columns are similar to the previous ones, but with regard to bias.

Regarding the results, the DEAM method performed better than DEA, with improvements ranging from 5% to 45% on average in MSE and 2% to 28% in bias. This fact increased when the number of inputs were higher. In addition, the results illustrate how the model worked better when the number of DMUs was around 50–75. Scenario I, i.e., the single input single output framework, shows small differences between the two methods. Nevertheless, in the trials, DEAM outperformed DEA in more than 95% of the cases. In contrast, the best analyzed situation was scenario V (one output and five inputs) with $n = 25$, showing a 45% reduction in MSE and 28% in bias, on average. This last result could be interpreted in favor of the DEAM approach as an indication that DEAM also seemed to outperform DEA with respect to the curse of dimensionality (Charles et al. [31]).

Table 2. Performance of DEA and DEAM.

		MSE				BIAS			
				Fraction of Trials	Improvement (%)			Fraction of Trials	Improvement (%)
Scenario	**Number of Obs.**	**DEA**	**DEAM**	**DEAM<= DEA**	**DEAM vs. DEA**	**DEA**	**DEAM**	**DEAM<= DEA**	**DEAM vs. DEA**
I	**25**	0.027(0.020)	0.024(0.019)	1.000	11.609%	0.125(0.046)	0.119(0.046)	1.000	4.873%
I	**50**	0.011(0.007)	0.010(0.007)	0.990	8.005%	0.076(0.026)	0.075(0.026)	0.990	2.822%
I	**75**	0.007(0.005)	0.007(0.005)	0.990	7.622%	0.060(0.019)	0.059(0.019)	0.980	2.194%
I	**100**	0.005(0.004)	0.005(0.004)	0.990	5.231%	0.051(0.019)	0.050(0.019)	0.950	1.936%
II	**25**	0.151(0.084)	0.108(0.067)	1.000	27.109%	0.276(0.071)	0.240(0.072)	1.000	13.460%
II	**50**	0.091(0.043)	0.067(0.037)	0.980	24.012%	0.206(0.045)	0.184(0.045)	0.990	10.587%
II	**75**	0.060(0.029)	0.040(0.024)	1.000	32.846%	0.160(0.032)	0.138(0.035)	1.000	14.252%
II	**100**	0.049(0.022)	0.033(0.019)	1.000	32.636%	0.140(0.031)	0.122(0.030)	1.000	13.285%
III	**25**	0.451(0.236)	0.287(0.199)	0.960	35.967%	0.470(0.126)	0.380(0.125)	0.960	19.215%
III	**50**	0.270(0.121)	0.165(0.090)	0.990	36.812%	0.347(0.077)	0.280(0.072)	0.980	19.075%
III	**75**	0.211(0.091)	0.119(0.050)	0.990	39.786%	0.291(0.056)	0.229(0.050)	0.980	20.996%
III	**100**	0.171(0.076)	0.112(0.053)	1.000	32.405%	0.257(0.047)	0.213(0.043)	1.000	16.971%
IV	**25**	1.046(0.457)	0.804(1.070)	0.880	14.949%	0.727(0.177)	0.623(0.264)	0.860	12.086%
IV	**50**	0.728(0.246)	0.471(0.265)	0.960	35.859%	0.571(0.113)	0.469(0.146)	0.880	18.295%
IV	**75**	0.605(0.191)	0.384(0.154)	0.990	35.084%	0.497(0.079)	0.403(0.079)	0.960	18.539%
IV	**100**	0.462(0.162)	0.308(0.114)	1.000	30.776%	0.418(0.068)	0.342(0.064)	0.990	17.912%
V	**25**	1.896(0.766)	1.009(0.563)	0.980	44.803%	0.984(0.224)	0.703(0.211)	0.980	28.043%
V	**50**	1.396(0.478)	0.922(0.566)	0.900	32.353%	0.801(0.140)	0.648(0.218)	0.870	18.492%
V	**75**	1.057(0.303)	0.750(0.315)	0.950	28.473%	0.673(0.107)	0.567(0.152)	0.880	15.677%
V	**100**	0.914(0.261)	0.624(0.211)	0.980	29.296%	0.613(0.090)	0.502(0.087)	0.970	17.543%

6. Discussion

In this section, we briefly discuss the main results of this paper and how they can be interpreted from the perspective of previous studies, mainly those based on Data Envelopment Analysis. Our findings and their implications are also discussed. Some limitations of our approach are highlighted.

In this paper, we have introduced a new way of estimating production frontiers in engineering and microeconomics, which is based upon the same fundamentals of Support Vector Machines (SVM), which is a well-known machine learning technique. Our numerical results have demonstrated that the frontier estimator derived from the new methodology (DEAM) is better than that associated with Data Envelopment Analysis (DEA), which represents the standard non-parametric technique for determining technical efficiency in the literature. The bias and mean squared error obtained for DEAM are smaller in all the scenarios analyzed, regardless of the number of variables and DMUs.

In comparison with the standard literature, the new methodology is more flexible. It generates production possibility sets that satisfy convexity, free disposability in inputs and outputs, and contain all the observations, but they do not meet the postulate of minimal extrapolation. In contrast, DEA satisfies all the above properties. In particular, minimal extrapolation is the reason why DEA can be seen as an overfitted model to estimate the underlying Data Generating Process (DGP) that is behind the generation of the data sample. DEAM does not suffer from this overfitting problem. However, it is not evident where the production possibility set, estimated by a non-overfitted model, should be located in the input–output space to correctly approximate the underlying technology, which, by definition, is unknown to us. In this regard, in this paper, we have implemented for the first time a strategy based on the idea of Structural Risk Minimization (Vapnik [1]) and cross validation, introducing a new PAC (Probably Approximately Correct) bound in production theory with the aim of solving the overfitting problem linked to DEA.

Some other authors have tried to modify the standard DEA technique such that the new approaches work as inferential methods (with the focus on the DGP) rather than as mere descriptive tools. For example, Banker and Maindiratta [8] and Banker [9] associated

DEA with maximum likelihood. Simar and Wilson [10–12] adapted bootstrapping to DEA. Kuosmanen and Johnson [13,14] introduced the Corrected Concave Nonparametric Least Squares. Unfortunately, despite the importance of machine learning techniques in the current literature, there have been few attempts to adapt DEA to the field of machine learning (see, for example, Esteve et al. [7], or Olesen and Ruggiero [15]). In this sense, DEAM has allowed us to build a new bridge between these two worlds: machine learning and efficiency measurement.

Finally, we would like to highlight a clear limitation associated with the new approach. DEAM is linked to an intensive computational procedure based on cross validation. This feature contrasts sharply with the simplicity of Data Envelopment Analysis.

7. Conclusions and Future Work

In this paper, for the first time, a bound on the generalization error for a piece-wise linear hypothesis has been established in the context of Support Vector Regression (SVR), by also considering typical axioms from production theory: convexity and free disposability. It shapes a new nexus between non-parametric frontier analysis and machine learning in the line recently followed by Esteve et al. [7], Valero-Carreras et al. [32], and Olesen and Ruggiero [15]. The new formulation on the bound of the generalization error of this kind of hypothesis gives rise to a new way of bounding the whole expected error when we approximate a target function through a piece-wise linear function, also controlling the empirical error. Minimizing this bound led to the definition of a new model, called Data Envelopment Analysis-based Machines (DEAM), which generates production function estimations that seek a balance between the empirical error and the generalization error.

Classical non-parametric techniques, such as DEA, suffer from the overfitting problem because they assume the axiom of minimal extrapolation (Banker et al. [6], Afriat [28], and Farrell [33]). The DEAM model, however, is more flexible when it comes to estimating production frontiers through a cross-validation process, disregarding the minimal extrapolation axiom, as was shown by a computational experience in this paper.

Finally, we finish by mentioning several lines that pose interesting avenues for further research. The first one is the possibility of extending the method to model multi-output situations. This could be interesting for dealing with more realistic production situations, considering information on the correlation among several outputs. Second, we could use other transformation functions (kernel methods) for the input space, in the same way as standard Support Vector Regression.

Author Contributions: Conceptualization, N.M.G. and J.A.; methodology, N.M.G. and J.A.; software, D.V.-C.; validation, N.M.G. and D.V.-C.; formal analysis, N.M.G., J.A. and D.V.-C.; investigation, N.M.G., J.A. and D.V.-C.; resources, N.M.G., J.A. and D.V.-C.; data curation, N.M.G. and D.V.-C.; writing—original draft preparation, N.M.G., J.A. and D.V.-C.; writing—review and editing, N.M.G., J.A. and D.V.-C.; visualization, N.M.G. and D.V.-C.; supervision, J.A.; project administration, J.A.; funding acquisition, J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033 grant number PID2019-105952GB-I00, by Generalitat Valenciana grant number ACIF/2020/155, and by Miguel Hernández University of Elche grant number 01623/2020.

Acknowledgments: The authors are grateful to the two anonymous reviewers for providing constructive comments and helping in improving the contents and presentation of this paper. Additionally, the authors are thankful for grant PID2019-105952GB-I00 funded by Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación/10.13039/501100011033. D. Valero-Carreras is thankful for the financial support from the Generalitat Valenciana under grant ACIF/2020/155. Finally, N. Guerrero is thankful for the financial support from the Miguel Hernández University of Elche under grant 01623/2020.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Lemma 3. Let $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d\}$ be γ-shattered by

$$F = \left\{ x \mapsto \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} : \|\boldsymbol{x}\| \leq R, \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} \geq 0, \frac{\boldsymbol{w}_{p_x}}{\delta_{p_x}}\boldsymbol{x} + \frac{\beta_{p_x}}{\delta_{p_x}} \leq \frac{\boldsymbol{w}_p}{\delta_p}\boldsymbol{x} + \frac{\beta_p}{\delta_p}, \forall p_x, p \in \{1, \ldots, n\} \right\}$$

witnessed by $r_1, \ldots, r_d \in R$. Then, for all $b = (b_1, \ldots, b_d) \in \{-1, 1\}^d$, there are $\left\{ (\boldsymbol{w}_p)_b \right\}_{p \in \{1, \ldots, n\}}$, $\left\{ (\delta_p)_b \right\}_{p \in \{1, \ldots, n\}}$ and $\left\{ (\beta_p)_b \right\}_{p \in \{1, \ldots, n\}}$ satisfying for all $i \in \{1, \ldots, d\}$ the following inequality:

$$b_i \left[\left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b x_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right) - r_i \right] \geq \gamma$$

Let set $S_0 \subset S$, and consider two cases:

- Case 1: If $\sum\{r_i : \boldsymbol{x}_i \in S_0\} \geq \sum\{r_i : \boldsymbol{x}_i \in S - S_0\}$, then $b_i = 1$ if and only if $\boldsymbol{x}_i \in S_0$
- Case 2: If $\sum\{r_i : \boldsymbol{x}_i \in S_0\} < \sum\{r_i : \boldsymbol{x}_i \in S - S_0\}$, then $b_i = 1$ if and only if $\boldsymbol{x}_i \in S - S_0$

Let us suppose that $\sum\{r_i : \boldsymbol{x}_i \in S_0\} \geq \sum\{r_i : \boldsymbol{x}_i \in S - S_0\}$, with $b_i = 1$ if and only if $\boldsymbol{x}_i \in S_0$ (*CASE 1*). For all $\boldsymbol{x}_i \in S_0$, we have

$$\begin{aligned} b_i \left[\left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right) - r_i \right] &= 1 \cdot \left[\left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right) - r_i \right] \\ &= \left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right) - r_i \underset{x_i \in S_0 \subset S}{\geq} \gamma, \end{aligned}$$

that is

$$\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \geq r_i + \gamma.$$

Then, taking the sum over the elements in the set S_0, we obtain the expression $\sum_{i / x_i \in S_0} \left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right) \geq \sum_{i / x_i \in S_0} (r_i + \gamma)$; which yields the following inequality:

$$\sum_{i / x_i \in S_0} \left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right) \geq \sum \{r_i : x_i \in S_0\} + |S_0|\gamma.$$

From F, we have $\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \leq \left(\frac{\boldsymbol{w}_p}{\delta_p} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_p}{\delta_p} \right)_b$, for all $p \in \{1, \ldots, P\}$. Thereby, the inequality

$$\left(\frac{\boldsymbol{w}_p}{\delta_p} \right)_b \underbrace{\sum_{i / x_i \in S_0} \boldsymbol{x}_i}_{\sum S_0} + |S_0| \left(\frac{\beta_p}{\delta_p} \right)_b = \sum_{i / x_i \in S_0} \left(\left(\frac{\boldsymbol{w}_p}{\delta_p} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_p}{\delta_p} \right)_b \right) \geq \sum_{i / x_i \in S_0} \left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}} \right)_b \right)$$

is satisfied $\forall p \in \{1, \ldots, n\}$. Finally,

$$\left(\frac{\boldsymbol{w}_p}{\delta_p} \right)_b \sum S_0 + |S_0| \left(\frac{\beta_p}{\delta_p} \right)_b \geq \sum \{r_i : \boldsymbol{x}_i \in S_0\} + |S_0|\gamma. \tag{A1}$$

Now, let $x_i \in S - S_0$, then

$$b_i\left[\left(\left(\frac{w_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right) - r_i\right] = (-1)\cdot\left[\left(\left(\frac{w_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right) - r_i\right] =$$

$$-\left(\left(\frac{w_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b x_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right) + r_i \underset{\boldsymbol{x}_i\in S-S_0\subset S}{\geq} \gamma$$

Then,

$$\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \leq r_i - \gamma.$$

Following the idea of applying the summary of elements, but now considering $x_i \in S - S_0$, the inequality $\sum\limits_{i/x_i\in S-S_0}\left(\left(\frac{w_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right) \leq \sum\limits_{i/\boldsymbol{x}_i\in S-S_0}(r_i-\gamma)$ holds. It can be rewritten as

$$\sum_{i/x_i\in S-S_0}\left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right) \leq \sum\{r_i : \boldsymbol{x}_i \in S - S_0\} - |S-S_0|\gamma. \quad \text{(A2)}$$

Now, there is $x_{i'} \in S - S_0$ such that $\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b x_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \leq \left(\frac{w_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b$ for all $x_i \in S - S_0$. Consequently, we have that $|S-S_0|\left(\frac{\boldsymbol{w}_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \boldsymbol{x}_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \leq \sum\limits_{i/x_i\in S-S_0}\left(\left(\frac{w_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right)$. $|S-S_0| \geq 0$ because $|S-S_0|$ is the cardinal of the set $S-S_0$. Conversely, $\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \boldsymbol{x}_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \geq 0$ by definition of F. Then, $-|S-S_0|\left(\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \boldsymbol{x}_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b\right) \leq |S-S_0|\left(\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b x_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b\right)$. Now, we can guarantee that

$$-|S-S_0|\left(\left(\frac{\boldsymbol{w}_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b x_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b\right) \leq \sum_{i/x_i\in S-S_0}\left(\left(\frac{\boldsymbol{w}_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b \boldsymbol{x}_i + \left(\frac{\beta_{p_{x_i}}}{\delta_{p_{x_i}}}\right)_b\right), \quad \text{(A3)}$$

and by (A2), the following inequality holds:

$$-|S-S_0|\left(\left(\frac{\boldsymbol{w}_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \boldsymbol{x}_{i'} + \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b\right) \leq \sum\{r_i : \boldsymbol{x}_i \in S - S_0\} - |S-S_0|\gamma. \quad \text{(A4)}$$

Considering inequalities (25) and (28), for $p_{x_{i'}} \in \{1,\ldots,n\}$, we have that ($\left.\begin{matrix} A \geq B \\ C \leq D \end{matrix}\right\} \Rightarrow A - C \geq B - D$.)

$$\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \sum S_0 + |S_0|\left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b + |S-S_0|\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \boldsymbol{x}_{i'} + |S-S_0|\left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \geq$$
$$\sum\{r_i : \boldsymbol{x}_i \in S_0\} + |S_0|\gamma - \sum\{r_i : \boldsymbol{x}_i \in S - S_0\} + |S-S_0|\gamma$$

and then,

$$\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b [\sum S_0 + |S - S_0| x_{i'}] + |S| \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \geq$$

$$\sum\{r_i : x_i \in S_0\} - \sum\{r_i : x_i \in S - S_0\} + |S|\gamma.$$

Under the supposition in the case 1 that $\sum\{r_i : x_i \in S_0\} \geq \sum\{r_i : x_i \in S - S_0\}$, we have that

$$\sum\{r_i : x_i \in S_0\} - \sum\{r_i : x_i \in S - S_0\} \geq 0,$$

which implies that

$$\left(\sum\{r_i : x_i \in S_0\} - \sum\{r_i : x_i \in S - S_0\}\right) + |S|\gamma \geq |S|\gamma.$$

Therefore, for $p_{x_{i'}} \in \{1, \ldots, n\}$, we have

$$\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b [\sum S_0 + |S - S_0| x_{i'}] + |S| \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \geq |S|\gamma,$$

that is,

$$\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b [\sum S_0 + |S - S_0| x_{i'}] \geq |S| \left(\gamma - \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b\right). \quad \text{(A5)}$$

Under the Cauchy–Schwarz inequality, we have

$$\left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b [\sum S_0 + |S - S_0| x_{i'}] \leq \left\| \left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \right\| \|\sum S_0 + |S - S_0| x_{i'}\|,$$

and then, we have

$$\|\sum S_0 + |S - S_0| x_{i'}\| = \|\sum S_0 - \sum(S - S_0) + \sum(S - S_0) + |S - S_0| x_{i'}\| \underset{Triangular}{\leq} \|\sum S_0 - \sum(S - S_0)\| +$$
$$+ \|\sum(S - S_0)\| + |S - S_0| \|x_{i'}\| \underset{Triangular}{\leq} \|\sum S_0 - \sum(S - S_0)\| + \sum_{i/x_i \in S - S_0} \|x_i\| + |S - S_0| \|x_{i'}\| \underset{\|x_i\| \leq r, \forall i \in \{1,\ldots,n\}}{\leq}$$
$$\|\sum S_0 - \sum(S - S_0)\| + |S - S_0| r + |S - S_0| r = \|\sum S_0 - \sum(S - S_0)\| + 2|S - S_0| r \underset{|S - S_0| \leq |S|}{\leq}$$
$$\|\sum S_0 - \sum(S - S_0)\| + 2|S| r.$$

In this way, it is possible to obtain

$$\left\| \left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \right\| (\|\sum S_0 - \sum(S - S_0)\| + 2|S| r) \geq \left\| \left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \right\| \|\sum S_0 + |S - S_0| x_{i'}\| \geq |S| \left(\gamma - \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b\right).$$

Because $\left\| \left(\frac{w_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \right\| \leq \underset{p \in \{1,\ldots,n\}}{Max} \left\| \left(\frac{w_p}{\delta_p}\right)_b \right\|$ and $\gamma - \left(\frac{\beta_{p_{x_{i'}}}}{\delta_{p_{x_{i'}}}}\right)_b \geq \underset{p \in \{1,\ldots,n\}}{Min} \left(\gamma - \left(\frac{\beta_p}{\delta_p}\right)_b\right)$, then

$$\underset{p \in \{1,\ldots,n\}}{Max} \left\| \left(\frac{w_p}{\delta_p}\right)_b \right\| (\|\sum S_0 - \sum(S - S_0)\| + 2|S| r) \geq |S| \underset{p \in \{1,\ldots,n\}}{Min} \left(\gamma - \left(\frac{\beta_p}{\delta_p}\right)_b\right).$$

Finally,

$$\left\|\sum S_0 - \sum(S - S_0)\right\| \geq |S| \left(\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left(\gamma - \left(\frac{\beta_p}{\delta_p}\right)_b\right)}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\left(\frac{w_p}{\delta_p}\right)_b\right\|} - 2r \right), \forall S_0 \subset S.$$

The proof for case 2 is analogous. □

Proof of Theorem 3. By Lemma 3, we have

$$\left\|\sum S_0 - \sum(S - S_0)\right\| \geq |S| \left(\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r \right),$$

for every subset $S_0 \subseteq S$, with $S = \{x_1, \ldots, x_d\}$ being an input learning sample γ-shattered through F defined in (10). Additionally, by Lemma 2, for all $S \subseteq R_+^m$ with $\|x\| \leq r$ for $x \in S$, some $S_0 \subseteq S$ satisfies the following condition:

$$\left\|\sum S_0 - \sum(S - S_0)\right\| \leq \sqrt{|S|}r.$$

Then, for certain $S_0 \subseteq S$, we have

$$\left\|\sum S_0 - \sum(S - S_0)\right\| \geq |S| \left(\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r \right) \text{ and } \left\|\sum S_0 - \sum(S - S_0)\right\| \leq \sqrt{|S|}r.$$

Therefore,

$$|S| \left(\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r \right) \leq \sqrt{|S|}r.$$

Finally,

$$|S| \leq \left(\frac{r}{\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r} \right)^2.$$

Because this is true for all S γ-shattered by F, it will be also true for the largest set γ-shattered by F, which means that $fat_F(\gamma)$ will be bound in that way:

$$fat_F(\gamma) \leq \left(\frac{r}{\frac{\underset{p\in\{1,\ldots,n\}}{Min}\left\{\gamma - \frac{\beta_p}{\delta_p}\right\}}{\underset{p\in\{1,\ldots,n\}}{Max}\left\|\frac{w_p}{\delta_p}\right\|} - 2r} \right)^2.$$

□

References

1. Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.
2. Vapnik, V. Principles of risk minimization for learning theory. In *Advances in Neural Information Processing Systems*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1992; pp. 831–838.
3. Blanco, V.; Puerto, J.; Salmerón, R. Locating hyperplanes to fitting set of points: A general framework. *Comput. Oper. Res.* **2018**, *95*, 172–193.
4. Blanco, V.; Puerto, J.; Rodriguez-Chia, A.M. On lp-Support Vector Machines and Multidimensional Kernels. *J. Mach. Learn. Res.* **2020**, *21*, 14.
5. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
6. Banker, R.D.; Charnes, A.; Cooper, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag. Sci.* **1984**, *30*, 1078–1092. [CrossRef]
7. Esteve, M.; Aparicio, J.; Rabasa, A.; Rodriguez-Sala, J.J. Efficiency analysis trees: A new methodology for estimating production frontiers through decision trees. *Expert Syst. Appl.* **2020**, *162*, 113783. [CrossRef]
8. Banker, R.D.; Maindiratta, A. Maximum likelihood estimation of monotone and concave production frontiers. *J. Product. Anal.* **1992**, *3*, 401–415. [CrossRef]
9. Banker, R.D. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Manag. Sci.* **1993**, *39*, 1265–1273. [CrossRef]
10. Simar, L.; Wilson, P.W. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Manag. Sci.* **1998**, *44*, 49–61.
11. Simar, L.; Wilson, P.W. A general methodology for bootstrapping in non-parametric frontier models. *J. Appl. Stat.* **2000**, *27*, 779–802.
12. Simar, L.; Wilson, P.W. Statistical inference in nonparametric frontier models: The state of the art. *J. Product. Anal.* **2000**, *13*, 49–78. [CrossRef]
13. Kuosmanen, T.; Johnson, A.L. Data envelopment analysis as nonparametric least-squares regression. *Oper. Res.* **2010**, *58*, 149–160. [CrossRef]
14. Kuosmanen, T.; Johnson, A. Modeling joint production of multiple outputs in StoNED: Directional distance function approach. *Eur. J. Oper. Res.* **2017**, *262*, 792–801.
15. Olesen, O.B.; Ruggiero, J. The hinging hyperplanes: An alternative nonparametric representation of a production function. *Eur. J. Oper. Res.* **2022**, *296*, 254–266.
16. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
17. Valiant, L.G. A theory of the learnable. *Commun. ACM* **1984**, *27*, 1134–1142.
18. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, MA, USA, 2000.
19. Bartlett, P.; Shawe-Taylor, J. *Generalization Performance of Support Vector Machines and Other Pattern Classifiers. Adv. Kernel Methods Support Vector Learn*; MIT Press: Cambridge, MA, USA, 1999; pp. 43–54.
20. Vazquez, E.; Walter, E. Multi-output suppport vector regression. *IFAC Proc. Vol.* **2003**, *36*, 1783–1788. [CrossRef]
21. Villa, G.; Lozano, S.; Redondo, S. Data envelopment analysis approach to energy-saving projects selection in an energy service company. *Mathematics* **2021**, *9*, 200. [CrossRef]
22. Sahoo, B.K.; Saleh, H.; Shafiee, M.; Tone, K.; Zhu, J. An Alternative Approach to Dealing with the Composition Approach for Series Network Production Processes. *Asia-Pac. J. Oper. Res. (APJOR)* **2021**, *38*, 2150004.
23. Amirteimoori, A.; Sahoo, B.K.; Charles, V.; Mehdizadeh, S. Stochastic Network Data Envelopment Analysis. In *Stochastic Benchmarking*; Springer: Cham, Switzerland, 2022; pp. 77–117.
24. Färe, R.; Primont, D. Distance functions. In *Multi-Output Production and Duality: Theory and Applications*; Springer: Dordrecht, The Netherlands, 1995; pp. 7–41.
25. Briec, W.; Lesourd, J.B. Metric distance function and profit: Some duality results. *J. Optim. Theory Appl.* **1999**, *101*, 15–33.
26. Cooper, W.W.; Seiford, L.M.; Tone, K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*; Springer: New York, NY, USA, 2007; Volume 2.
27. Briec, W. Hölder distance function and measurement of technical efficiency. *J. Product. Anal.* **1999**, *11*, 111–131.
28. Afriat, S.N. Efficiency estimation of production functions. *Int. Econ. Rev.* **1972**, *13*, 568–598.
29. Mangasarian, O.L. Arbitrary-norm separating plane. *Oper. Res. Lett.* **1999**, *24*, 15–23.
30. Aparicio, J.; Pastor, J.T. A well-defined efficiency measure for dealing with closest targets in DEA. *Appl. Math. Comput.* **2013**, *219*, 9142–9154.
31. Charles, V.; Aparicio, J.; Zhu, J. The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *Eur. J. Oper. Res.* **2019**, *279*, 929–940.
32. Valero-Carreras, D.; Aparicio, J.; Guerrero, N.M. Support vector frontiers: A new approach for estimating production functions through support vector machines. *Omega* **2021**, *104*, 102490.
33. Farrell, M.J. The measurement of productive efficiency. *J. R. Stat. Soc. Ser. A* **1957**, *120*, 253–281.

Guerrero, N.M., Aparicio, J., Moragues, R., Valero-Carreras, D. (2024). Merging Data Envelopment Analysis and Structural Risk Minimization: Some Examples of Use of Multi-output Machine Learning Techniques on Real-World Data. In: Edalatpanah, S., Hosseinzadeh Lotfi, F., Kerstens, K., Wanke, P. (eds) Analytical Decision Making and Data Envelopment Analysis. Infosys Science Foundation Series(). Springer, Singapore. https://doi.org/10.1007/978-981-97-6972-8_1

Merging Data Envelopment Analysis and Structural Risk Minimization: Some Examples of Use of Multi-output Machine Learning Techniques on Real-World Data

Nadia M. Guerrero, Juan Aparicio, Raul Moragues, and Daniel Valero-Carreras

Abstract Data Envelopment Analysis (DEA) is nowadays a very famous nonparametric technique for the measurement of technical efficiency. It does so by building a production possibility set that satisfies certain microeconomic and mathematical axioms, such as free disposability in inputs and outputs and convexity, and determines the most conservative estimate of technical inefficiency of each assessed unit via the minimal extrapolation principle (i.e., the Occam's Razor view). Given a data sample and from a statistical point of view, this last axiom implies an estimation of technical inefficiency by exclusively minimizing the empirical error, resulting in overfitting to the data as a by-product. To overcome this statistical problem when the objective is measuring technical inefficiency beyond the data sample, a methodology has been recently introduced which follows the Structural Risk Minimization (SRM) principle. It controls both the empirical and the generalization (prediction) error of the model. This methodology, called Data Envelopment Analysis-based Machines (DEAM) was introduced for the single-output setting [(Guerrero, Aparicio, & Valero-Carreras 2022). Combining Data Envelopment Analysis and Machine Learning. Mathematics 2022, 10, 909.]. In this chapter, we extend the DEAM approach for the estimation of production functions to the multi-output production framework,

N. M. Guerrero · J. Aparicio (✉) · R. Moragues · D. Valero-Carreras
Center of Operations Research (CIO), Miguel Hernandez University, Elche, Spain
e-mail: j.aparicio@umh.es

N. M. Guerrero
e-mail: nadia.guerrero@goumh.umh.es

R. Moragues
e-mail: rmoragues@umh.es

D. Valero-Carreras
e-mail: dvalero@umh.es

J. Aparicio
ValgrAI—Valencian Graduate School and Research Network of Artificial Intelligence, Valencia, Spain

[redacted]

evaluate technical efficiency with respect to a variety of measures, and illustrate its performance via some empirical applications. In this way, we provide some examples of use of multi-output machine learning techniques for measuring technical efficiency on real-world data.

Keywords Data envelopment analysis · Machine learning · Support vector regression · Structural risk minimization · Multi-output production frontiers

1 Introduction

Efficiency measurement is an area of research which has grown immensely since the original definition proposed by (Koopmans, 1951) and the first approaches to their measurement by Debreu and Farrell (Debreu, 1951; Farrell, 1957). Proposed techniques are often classified into parametric and nonparametric approaches, from which some of the most extended methods are Stochastic Frontier Analysis in the parametric family (Aigner et al., 1977; Meeusen & van Den Broeck, 1977), while in the nonparametric family, Data Envelopment Analysis (DEA) has become one of the most commonly used methodologies. The parametric approach usually assumes some functional form for the production function and aims to estimate the coefficients or parameters associated with this form, using a method such as the minimization of an error function. Upon this basis, it can apply statistical inference tools to the parameters, but it requires some assumptions on the distributions of the error terms and technical inefficiency. On the other hand, the nonparametric approach does not require an a priori assumption about the shape of the production function, nor does it require assumptions about the probability distributions associated with the error terms or the data generating process. Instead, nonparametric methodologies usually base their estimation on the satisfaction of certain properties of the underlying production process. Additionally, the nonparametric approaches have a more natural multi-output extension in comparison with the parametric approaches, which necessarily represent this more complex framework through an indirect notion: the distance function (Orea & Zofío, 2019).

When measuring efficiency, it is often measured with respect to some production possibility set or technology, which consists of those combinations of resources (inputs) and products (outputs) which are feasible to be produced by the production process under study. In particular, DEA is a nonparametric technique originally introduced in (Banker et al., 1984; Charnes et al., 1978) which is based on the microeconomic assumptions on the technology of convexity, envelopment of the data, free disposability of inputs and outputs, and the principle of minimal extrapolation. The family of DEA techniques also includes different estimations of the production possibility set, or technology, according to a variety of assumptions and choices, such as returns to scale.

Regardless of the particular form of the production technology, the measurement of efficiency pays particular interest to the efficient frontier of the technology, which consists of those points which cannot be improved while remaining feasible. In this context, the technical efficiency of a Decision Making Unit (DMU) is defined as the potential improvement of said unit while remaining in the technology (reaching the efficient frontier). However, there are multiple and varied paths to improvement which can be followed to reach efficiency, with a variety of properties being satisfied by various approaches. For example, some of the first approaches introduced were the Farrell or radial measures (Farrell, 1957). The input-oriented radial measure allowed for the equiproportional reduction of the inputs while leaving the outputs constant, while the output-oriented radial measure allows for the increase in outputs while leaving the inputs constant (Banker et al., 1984; Charnes et al., 1978). In particular, the Farrell measures keep the relative proportions of the inputs (outputs), or mix, constant. Another approach to measure efficiency, the Directional Distance Function (DDF), considers an a priori specified directional vector and projects DMUs along it (Chambers et al., 1998; Luenberger, 1992). These measures do retain certain properties of the original DMU, but they do not guarantee that the projection of the DMU in question is fully efficient, since there may be potential improvements to be made along some variables but not others, and they measure so-called weak efficiency. A stronger concept is that of strong or Pareto efficiency, where the units that the projections are made into do not allow for further improvement along any of their components without the worsening of another component. In order to overcome this limitation, measures such as the Russell measures were proposed which generalise the Farrell measures. The Russell measures are still oriented, but not radial. They allow the rescaling of different input (output) variables by different amounts (Färe & Lovell, 1978). Along these lines, an alternative formulation of the DEA models using slacks along individual variables (inputs and outputs) was introduced via the additive model (Charnes et al., 1985), which was later extended by the addition to weights to the slacks, and the introduction of a variety of Weighted Additive measures of efficiency such as (Cooper et al., 1999, 2011; Tone, 2001). All these measures rely on estimations of the underlying production technology and modify the permitted improvement paths of the DMUs.

Despite all its nice properties, DEA has been criticised for its nature as a descriptive tool. In particular, the use of the principle of minimal extrapolation is an approach which aims to minimize the distance from the frontier to the dataset, which is related to exclusively minimizing the empirical error (Esteve et al., 2020). This results in overfitting to the dataset being used and a potential lack of generalization to unseen data when the sample size is small. This has led authors to turn to areas such as statistical learning in order to attempt to perform statistical inference tasks on these datasets in order to estimate more robust efficiencies. One avenue of research in this direction is the study of the asymptotic properties of the DEA estimators such as consistency, rates of convergence, and bias (Kneip et al., 1998, 2008, 2011, 2015). Building upon this, bias correction methods have been developed using bootstrapping methodologies.

Another approach towards the attempts to overcome this limitation can be found in generalization theory, and the approaches to attempt to bind not only the empirical error but also the prediction or generalization error. In this area, a fruitful and interesting direction is the Structural Risk Minimization (SRM) principle, which attempts to obtain and minimize an upper bound on the expected risk over a hypothesis class (Vapnik, 1991). These types of bounds are called Probably Approximately Correct, since they are bounds which have a small probability of the bound failing (Probably) when the bound is achieved via a classifier which has a low error rate (Approximately Correct). Then, using these bounds, techniques are developed to minimize them. A family of estimators which are related to the SRM principle is that of Support Vector Machines (SVM) (Vapnik, 1998), which aim to minimize two factors: the value of the empirical risk and the generalization error. These factors control the bound on the expected risk. In particular, Support Vector Regression adapts the SVM family to regression problems.

Based on these concepts, a recent contribution (Guerrero et al., 2022) proposed an estimator called Data Envelopment Analysis-based Machines (DEAM) to estimate polyhedral technologies satisfying microeconomic axioms as in DEA except for the minimal extrapolation principle. In that article, the authors defined a piece-wise linear hypothesis class and provided a PAC bound for the fat-shattering dimension of this class. Using these bounds, they introduced a methodology inspired by Support Vector Regression, enabling it to control the generalization error of the model, thus allowing it to generalize to unseen data. Guerrero et al., however, only introduced the technique in a single-output scenario, which is a very limited production situation in practice.

In this chapter, we will extend the DEAM technique to estimate technical efficiency in multi-input multi-output contexts. This extension satisfies the microeconomic axioms of DEA except for the minimal extrapolation principle, and estimates, as in the original DEAM technique, production technologies which do not overfit to the data. Furthermore, we will tailor the most usual list of technical efficiency measures to be used in the case of resorting to DEAM for estimating the production technology. In this sense, we will provide the mathematical programming models that should be implemented. Finally, we will illustrate the performance of the new models, in comparison with DEA, through two real-world datasets.

The area of statistical learning is also related to machine learning. The worlds of efficiency estimation and machine learning have some similarities which have recently begun to be exploited in force, with a variety of authors attempting to use techniques to avoid overfitting in efficiency estimation via the use of machine learning techniques. Some early contributions toward this area can be seen in work of (Kuosmanen & Johnson, 2010), who introduced a piecewise linear estimator called Corrected Concave Nonparametric Least Squares. Nonparametric estimators using kernels were introduced in (Du et al., 2013), and a methodology involving quadratic and cubic splines was proposed in (Daouia et al., 2016). Methods involving decision trees were tailored to measure technical efficiency in (Aparicio et al., 2021; Esteve et al., 2020) and (Tsionas, 2022), and were enriched by bagging (bootstrap aggregating) techniques such as random forest (Esteve et al., 2023) or boosting methodologies such as gradient boosting in (Guillen et al., 2023). Support Vector

Regression-inspired methods were also introduced in (Valero-Carreras et al., 2021, 2022).

The rest of this chapter is structured as follows. Section 2 briefly reviews background concepts about Data Envelopment Analysis, measures of efficiency, and the DEAM technique. Section 3 extends DEAM to the context of multi-output production processes and enriches DEAM with the measurement of efficiency with respect to multiple measures of efficiency. Section 4 illustrates the efficiencies obtained by multi-output DEAM using two empirical examples with multiple outputs. Finally, Sect. 5 presents some conclusions obtained, and outlines some potential avenues for further research.

2 Background

We begin by setting up some notation. We will denote vectors using bold characters while scalars will be in italics. We will denote the scalar product of two vectors $\boldsymbol{a}$, $\boldsymbol{b}$ by $\boldsymbol{ab}$, while their component-wise product will be indicated by $\boldsymbol{a} \odot \boldsymbol{b}$. We now briefly introduce the basic concepts of Data Envelopment Analysis, including a variety of measures of efficiency, and we describe the Data Envelopment Analysis-based Machines (DEAM) methodology.

Data Envelopment Analysis is a nonparametric methodology to estimate the technical efficiency of a Decision Making Unit (DMU) from the information obtained about the inputs and outputs of some observations. We consider a set of n DMUs, where unit k takes inputs $\boldsymbol{x}_k = \left(x_k^{(1)}, \ldots, x_k^{(m)}\right) \in R_+^m$ in order to produce $\boldsymbol{y}_k = \left(y_k^{(1)}, \ldots, y_k^{(s)}\right) \in R_+^s$ outputs. The data will be arranged into a matrix $X = (\boldsymbol{x}_k)$ of inputs and a matrix $Y = \left(\boldsymbol{y}_k\right)$ of outputs, where each DMU is a column of these matrices. The production possibility set, or technology, is defined in general as:

$$T = \{(\boldsymbol{x}, \boldsymbol{y}) \in R_+^{m+s} : \boldsymbol{x} \text{ can produce } \boldsymbol{y}\}. \tag{1}$$

The technical efficiency of a DMU can be defined as the improvement in inputs and/or outputs which can be performed while remaining within the production technology. Of relevance to its measurement is a particular region of the technology called the weak efficient frontier of T:

$$\delta^W(T) = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in T : \hat{\boldsymbol{x}} < \boldsymbol{x}, \hat{\boldsymbol{y}} > \boldsymbol{y} \Rightarrow \left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\right) \notin T\right\}. \tag{2}$$

The weak efficient frontier of a technology consists of the points that are not strictly Pareto dominated by any other points in the technology. Other forms of the efficient frontier result in slightly different definitions, such as the strong efficient frontier of T, which consists of those points for which any improvement along any variable results in leaving the technology:

$$\delta^S(T) = \{(\boldsymbol{x}, \boldsymbol{y}) \in T : \hat{\boldsymbol{x}} \leq \boldsymbol{x}, \hat{\boldsymbol{y}} \geq \boldsymbol{y}, (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \neq (\boldsymbol{x}, \boldsymbol{y}) \Rightarrow (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) \notin T\}. \quad (3)$$

Since the theoretical technology is unknown in practice, we will work with estimates of this technology. Under Variable Returns to Scale, the DEA estimate of the technology is as follows:

$$\widehat{T}_{DEA} = \{ (\boldsymbol{x}, \boldsymbol{y}) \in R_+^{m+s} : x \geq X\lambda, y \leq Y\lambda, \lambda \geq 0, \lambda 1 = 1 \} \quad (4)$$

In particular, $\widehat{T}_{DEA}$ is the smallest set containing the observations, which satisfies convexity and free disposability in inputs and outputs (Banker et al., 1984). Convexity of the technology means that, for any two DMUs $(\boldsymbol{x}_k, \boldsymbol{y}_k), (\boldsymbol{x}_l, \boldsymbol{y}_l) \in \widehat{T}_{DEA}$, then any convex linear combination of them is also feasible, that is, $\lambda_k(\boldsymbol{x}_k, \boldsymbol{y}_k) + (1 - \lambda_k)(\boldsymbol{x}_l, \boldsymbol{y}_l) \in \widehat{T}_{DEA}, 0 \leq \lambda_k \leq 1$. Free disposability in inputs and outputs assumes that, if $(\boldsymbol{x}, \boldsymbol{y}) \in \widehat{T}_{DEA}$ then, whenever $\boldsymbol{x}' \geq \boldsymbol{x}$ and $\boldsymbol{y}' \leq \boldsymbol{y}$, we have $(\boldsymbol{x}', \boldsymbol{y}') \in \widehat{T}_{DEA}$. Due to the principle of minimal extrapolation, which is related to the selection of the smallest set satisfying the previous conditions, $\widehat{T}_{DEA}$ is a conservative estimate, which will fit the data perfectly at a frontier level, but will overestimate the technical efficiency of the DMUs, since the true efficient frontier will be slightly higher and unobserved. This is due to the property of minimal extrapolation, and it is a characteristic often denoted overfitting in the data-based modelling and machine learning literature. It usually means that an estimator fits the data so well that it may not generalise to new data, which was not used to train the estimator. This is especially a problem if the objective of the study is to state something about the underlying production function of the sector instead of evaluating and ranking exclusively the set of DMUs under assessment with respect to their performance. Under the statistical approach to production theory (Daraio & Simar, 2007), there is a Data Generating Process (DGP), which is unknown for the observers, which relates inputs to outputs. We only observe a sample of n possible combinations of these inputs and outputs. At this point, researchers could be interested in determining two alternative notions of technical efficiency: relative vs absolute efficiency (Aparicio & Esteve, 2022). Data Envelopment Analysis identifies the degree of relative technical efficiency of each DMU, that is, the degree of efficiency measured in comparative terms regarding the performance of exactly the observations in the sample. On the other hand, absolute technical efficiency is the performance determined with respect to the underlying DGP. It stands to reason that, when the sample size increases significantly, the two notions (approximately) coincide by invoking some property of consistency. But this is not true in the case of relatively small sample sizes. In certain real frameworks, calculating the relative technical efficiency of each DMU may not be the only goal of the study. For example, in the well-known PISA (Programme for International Student Assessment) survey by the OECD (Organisation for Economic Co-operation and Development), researchers play with information about a (random) sample of schools of each participant (country). In fact, data are provided by OECD in such a way that it is almost impossible to recognize the schools under study. Thus, the schools are anonymous for any external observer. Consequently, estimating the efficiency of these n anonymous units (schools) in the observed sample is not the only objective. Instead, a more useful task is to go beyond it and attempt to estimate the underlying (education) production function behind the data. The information obtained about the production technology will allow a decision maker to obtain guidance in practice about how to improve performance in the actual units (schools), instead of on this sample of anonymous observations, where it may not be very useful. In this context, determining absolute technical efficiency could make more sense than identifying relative technical efficiency. In particular, one of the real-world datasets that we will use to illustrate the performance of our model corresponds to information provided in the PISA report. Additionally, even in the case of having information about all the units that make up a given population (for example, all banks in Europe, all tax offices in a country, etc.), we could suppose, as it is usual to assume in Statistics, that these observations are a random sample of all the different combinations that we could have observed. Behind these observations is the DGP and we could be interested in assessing each DMU with respect to it rather than exclusively the observed performance of the remaining units in the population under evaluation. Although we recognize that, in this second scenario, researchers might be tempted to exclusively measure relative efficiency due to having data about all the elements of the population. In this sense, relative efficiency would work as a descriptive tool of the performance in the sample and the calculation of absolute efficiency could be seen as complementary information (determined from an inferential perspective). We will also illustrate our approach using a framework like the last one mentioned here.

With respect to $\widehat{T}_{DEA}$, or any production technology satisfying similar microeconomic axioms, a variety of measures of efficiency have been proposed, which relate to the distance of each DMU to the frontier (weak or strong). These are various ways of considering how to improve the efficiency of a DMU, which may be accomplished by increasing the produced outputs without increasing the required inputs, by producing the same outputs with fewer inputs, or a combination of these two approaches. We next present this variety of measures.

2.1 Measures of Technical Efficiency in Efficiency Analysis

In general terms, the technical efficiency level of a DMU k is defined as the distance between the input–output bundle $\left(\boldsymbol{x}_k, \boldsymbol{y}_k\right)$ and the so-called efficient frontier of a technology T, defined as those points whose efficiency cannot be increased without leaving the technology. The radial measures of efficiency are some of the first proposed methods for measuring technical efficiency (Farrell, 1957). The output-oriented radial measure (RO) maintains the inputs of each DMU constant while increasing every output by the same scalar value ϕ:

$$\phi(\boldsymbol{x}, \boldsymbol{y}) = max\{\phi : (\boldsymbol{x}, \phi\boldsymbol{y}) \in T\}. \tag{5}$$

Similarly, the input-oriented radial measure (RI) keeps the output values constant while scaling all inputs equiproportionally by a scalar θ (Banker et al., 1984). It is defined by:

$$\theta(\boldsymbol{x}, \boldsymbol{y}) = min\{\theta : (\theta \boldsymbol{x}, \boldsymbol{y}) \in T\}. \tag{6}$$

These measures have a radial orientation, as they scale all inputs or outputs by the same amount. They keep the mix of outputs (inputs) constant. Another approach, the directional distance function (DDF) (Chambers et al., 1998; Luenberger, 1992), considers the potential improvement along a direction specified by a nonzero vector $\boldsymbol{g} = \left(\boldsymbol{g}_0^x, \boldsymbol{g}_0^y\right)$, where $\boldsymbol{g}_0^x \in R_+^m$ and $\boldsymbol{g}_0^y \in R_+^s$, as a measure of inefficiency β of an input–output bundle $(\boldsymbol{x}, \boldsymbol{y})$ given by the following program:

$$\beta(\boldsymbol{x}, \boldsymbol{y}) = max\left\{\beta : \left(\boldsymbol{x} - \beta \boldsymbol{g}_0^x, \boldsymbol{y} + \beta \boldsymbol{g}_0^y\right) \in T\right\}. \tag{7}$$

Regarding the directional vector $\boldsymbol{g}$, in this chapter, we particularly use the values of the DMU itself, that is, $\boldsymbol{g} = (\boldsymbol{x}, \boldsymbol{y})$. The measures described so far, RI, RO and DDF, project DMUs to the weak efficient frontier of the technology in either a radial direction or a direction previously specified. At this stage, some extra improvement may still be possible, but not simultaneously in all inputs (outputs), by focusing on the strong efficient frontier.

Other proposals directly project DMUs to the strong efficient frontier. For example, an alternative model, which measures slacks independently along each variable, is the weighted additive (WA) model (Knox Lovell & Pastor, 1995). Using the weighted additive model, the Range Adjusted Measure (Cooper et al., 1999) is defined by choosing weights as follows. Let $R_j^- = \overline{x}_j - \underline{x}_j$ and $R_r^+ = \overline{y}_r - \underline{y}_r$ be the range of each variable. Then, the WA uses weights $\left(\boldsymbol{\alpha}_o^x, \boldsymbol{\mu}_o^y\right) \in R_{++}^{m+s}$, where $\boldsymbol{\alpha}_0^{x(j)} = \frac{1}{(m+s)R_j^-}$ and $\boldsymbol{\mu}_0^{y(r)} = \frac{1}{(m+s)R_r^+}$:

$$WA(\boldsymbol{x}, \boldsymbol{y}) = max\left\{\boldsymbol{\alpha}_0^x \boldsymbol{s}_0^x + \boldsymbol{\mu}_0^y \boldsymbol{s}_0^y \in R : \left(\boldsymbol{x} - \boldsymbol{s}_0^x, \boldsymbol{y} + \boldsymbol{s}_0^y\right) \in T, \left(\boldsymbol{s}_0^x, \boldsymbol{s}_0^y\right) \in R_+^{m+s}\right\}. \tag{8}$$

A generalization of the radial measures which we also consider are the Russell measures, which scale each input (output) by a different constant (Färe & Lovell, 1978). This allows for the identification of inefficiencies in some variables. The input-oriented Russell model (RUI) scales each input j by a scalar $0 \leq \theta^{(j)} \leq 1$, and is given by:

$$RUI(x, y) = min\left\{\frac{1}{m}\sum\nolimits_{(j=1)}^{m} \theta^{(j)} : (\theta \ \Theta \ x, y) \in T, \ 0 \ \leq \ \theta \ \leq \ 1\right\}. \tag{9}$$

Similarly, the output-oriented Russell model (RUO), which allows for the increase of each output by a different factor, is given by:

$$RUO(x, y) = max\left\{\frac{1}{s}\sum\nolimits_{r=1}^{s}\phi^{(r)} : (x, \phi \odot y) \in T, \phi \geq 1\right\}. \quad (10)$$

The final measure that we consider is the Enhanced Rusell Graph (ERG) measure, introduced in (Pastor et al., 1999) and (Tone, 2001), which uses the ratio between the terms in the Russell Input and Output measures. This model is not linear, but it can be linearized using a slacks-based model via a standard transformation in the literature:

$$ERG(x, y) = min\left\{\left(\frac{1}{m}\sum\nolimits_{j=1}^{m}\theta^{(j)}\right)/\left(\frac{1}{s}\sum\nolimits_{r=1}^{s}\phi^{(r)}\right) : (\theta \odot x, \phi \odot y) \in T, \phi \geq 1, 0 \leq \theta \leq 1\right\}. \quad (11)$$

In this chapter, we adapt each of the measures presented to the proposed multi-output generalization of DEAM in (Guerrero et al., 2022). Also, we will compare the efficiency scores obtained by DEA and multi-output DEAM with respect to each measure in empirical examples. We remark that the comparisons between efficiency scores are only appropriate when comparing efficiencies calculated with respect to the same measure with respect to each model, since different measures have different properties and cannot be compared.

We remark here the different nature of some of the measures. The Radial Input and Russell input measures are input-oriented, such that values of 1 indicate technical efficiency, and the obtained efficiency values lie between 0 and 1. On the other hand, the Radial Output and Russell Output measures have the property that the efficient units attain the value 1, and larger values indicate lower efficiency, since their outputs can be increased by larger amounts while remaining feasible. The DDF and the WA measure can be considered measures of inefficiency, since efficient units achieve a score of 0, and larger values indicate less efficiency. Finally, the ERG provides values between 0 and 1, with this last value signaling efficiency.

2.2 *Data Envelopment Analysis-Based Machines*

We now present the basics of Data Envelopment Analysis-based Machines as introduced in (Guerrero et al., 2022), where the single-output multi-input model was stated. This technique is useful to estimate production technologies, without suffering from overfitting, via the implementation of bounds on the generalization error of the model following the Structural Risk Minimization (SRM) principle. The DEAM model is:

$$\begin{array}{lll}
\underset{w,\beta,\delta,\xi,\xi\prime,u,v}{Min} & u+v+C\left(\sum_{i=1}^{n}\xi_i^2+\sum_{i=1}^{n}\xi_i\prime^2\right) & \\
s.t. & & \\
& \|\boldsymbol{w}_i\|_1 \le u & \forall i=1,\ldots,n \quad (12.1)\\
& \beta_i \le v & \forall i=1,\ldots,n \quad (12.2)\\
& \delta_p y_i \le \boldsymbol{w}_p\boldsymbol{x}_i+\beta_p & \forall i,p=1,\ldots,n \quad (12.3)\\
& \boldsymbol{w}_i,\delta_i \ge 0 & \forall i=1,\ldots,n \quad (12.4)\\
& \boldsymbol{w}_i\boldsymbol{x}_i+\beta_i-\delta_i y_i \le \varepsilon+\xi_i & \forall i=1,\ldots,n \quad (12.5)\\
& \delta_i y_i-\boldsymbol{w}_i\boldsymbol{x}_i-\beta_i \le \varepsilon+\xi_i\prime & \forall i=1,\ldots,n \quad (12.6)\\
& \xi_i,\xi_i\prime \ge 0 & \forall i=1,\ldots,n \quad (12.7)\\
& \|(\boldsymbol{w}_i,\delta_i)\|_1 = 1 & \forall i=1,\ldots,n \quad (12.8)\\
& \boldsymbol{w}_i\boldsymbol{x}_i+\beta_i-\delta_i y_i \le \boldsymbol{w}_p\boldsymbol{x}_i+\beta_p-\delta_p y_i & \forall i,p=1,\ldots,n \quad (12.9)
\end{array} \tag{12}$$

This model is a Support Vector Regression-inspired model where the terms u and v are bounds to each of the components in the fat-shattering bound obtained for a class of piece-wise linear hypothesis classes in (Theorem 3, Guerrero et al., 2022). This model involves a number n of half-spaces (and corresponding hyperplanes), which estimate a polyhedral production technology via their intersection. The first two conditions (12.1) and (12.2) ensure that the bounds on the generalization error hold, using the 1-norm, which is linear, for consistency with (12.8). Constraint (12.3) guarantees that the estimated production technology envelops the data from above, by forcing each hyperplane to be above the data. Moreover, constraint (12.4) ensures free disposability (see (Kuosmanen & Johnson, 2010)). The constraints (12.5), (12.6) and (12.7) relate to the variables $\xi\prime s$, which, at optimum, can be interpreted as the l_∞ distance from the observation to a 'robust' definition of efficient frontier, which uses the notion of margin (here represented by ε). Given that this frontier is made up of hyperplanes, the distance is achieved in one of them. Finally, this hyperplane is guaranteed to be the one closest to the observation by restriction (12.9).

The DEAM model also involves two hyperparameters, C and ε, which will be chosen by cross-validation. The hyperparameter C weighs the terms in the objective function, which combines the empirical error term involving ξ and the bound term $u+v$. The other hyperparameter, ε, is a margin hyperparameter, which adds robustness to SVR and DEAM.

The technology estimated by the single-output DEAM is a piece-wise linear technology that can be defined as

$$\hat{T}_{DEAM} := \left\{(x,\ y) \in R_+^{m+1} : \mathbf{w}_p^* x + \beta_p^* - \delta_p^* y \ge 0,\ \forall p \in \{1,\ \ldots,\ n\}\right\},$$

given an optimal solution $(\{\mathbf{w}_p^*,\ \beta_p^*,\ \delta_p^*,\ \xi_p^*,\ \xi_p^{*'}\}_{p=1,\ldots,n},\ u^*,\ v^*)$ of model (12). This technology is the intersection of the half-spaces defined by the set of n hyperplanes identified by the model at optimum. It satisfies convexity and free disposability

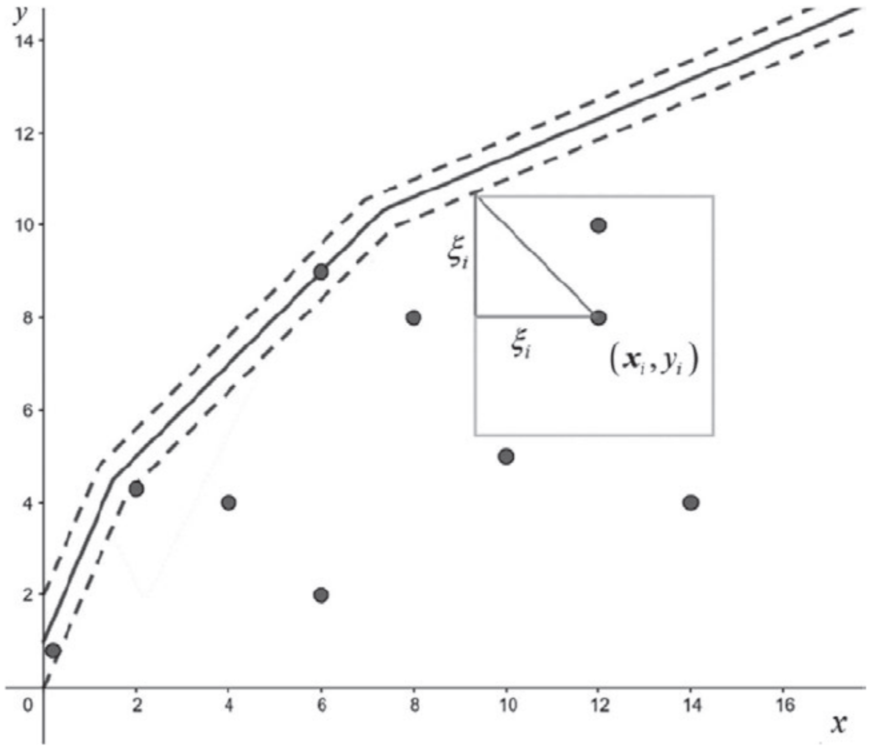

Fig. 1 Illustration of the DEAM estimation of a production function in a single-input single-output example

(Propositions 1 and 2, Guerrero et al., 2022) and includes all the observations. Consequently, it contains the DEA-estimated technology as a subset (Corollary 1, Guerrero et al., 2022).

Figure 1 illustrates the technology that can be estimated by DEAM in a single-input single-output example. The solid line represents the efficient frontier, whereas the dashed lines indicate its corresponding robust estimate (plus/minus the margin ε). The DEAM technology, i.e., the intersection between the non-negative orthant and the half-spaces defined from the hyperplanes, is a polyhedron. The technology contains all the DMUs, is a convex set and satisfies free disposability. However, it does not meet the minimal extrapolation principle. Additionally, the value of the variable ξ_i at optimum represents the least distance from DMU_i to the margin (using the l_∞-norm).

3 Methodology

In this section, we extend the DEAM model to the multi-output context. To do that, we will provide a natural extension of the single-output model by transforming the product of scalars δy into the product of corresponding vectors $\boldsymbol{\delta y}$. Accordingly, the multi-output DEAM model is as follows:

$$\begin{array}{lll} \underset{w,\beta,\delta,\xi,\xi\prime,u,v}{Min} & u+v+C\left(\sum_{i=1}^{n}\xi_i^2+\sum_{i=1}^{n}\xi_i\prime^2\right) & \\ s.t. & & \\ & \|\boldsymbol{w}_i\|_1 \le u & \forall i=1,\ldots,n \quad (13.1) \\ & \beta_i \le v & \forall i=1,\ldots,n \quad (13.2) \\ & \boldsymbol{\delta}_p\boldsymbol{y}_i \le \boldsymbol{w}_p\boldsymbol{x}_i+\beta_p & \forall i,p=1,\ldots,n \ (13.3) \\ & \boldsymbol{w}_i,\boldsymbol{\delta}_i \ge 0 & \forall i=1,\ldots,n \quad (13.4) \\ & \boldsymbol{w}_i\boldsymbol{x}_i+\beta_i-\boldsymbol{\delta}_i\boldsymbol{y}_i \le \varepsilon+\xi_i & \forall i=1,\ldots,n \quad (13.5) \\ & \boldsymbol{\delta}_i\boldsymbol{y}_i-\boldsymbol{w}_i\boldsymbol{x}_i-\beta_i \le \varepsilon+\xi_i\prime & \forall i=1,\ldots,n \quad (13.6) \\ & \xi_i,\xi_i\prime \ge 0 & \forall i=1,\ldots,n \quad (13.7) \\ & \|(\boldsymbol{w}_i,\boldsymbol{\delta}_i)\|_1=1 & \forall i=1,\ldots,n \quad (13.8) \\ & \boldsymbol{w}_i\boldsymbol{x}_i+\beta_i-\boldsymbol{\delta}_i\boldsymbol{y}_i \le \boldsymbol{w}_p\boldsymbol{x}_i+\beta_p-\boldsymbol{\delta}_p\boldsymbol{y}_i & \forall i,p=1,\ldots,n \ (13.9) \end{array} \quad (13)$$

This model extends DEAM to the multi-output context by considering a number s of outputs, so that the hyperplanes envelop the data in every output simultaneously, where now, correspondingly, each $\boldsymbol{\delta}_p$ is a vector with s components. Its production technology is defined as the intersection of a number of at most n different hyperplanes estimated by the model, so that multi-output DEAM provides a convex, piece-wise linear technology which, given an optimal solution $(\{w_p^*,\ \beta_p^*,\ \delta_p^*,\ \xi_p^*,\ \xi_p^{*'}\}_{p=1,\ldots,n},\ u^*,\ v^*)$ of model (13), is defined by:

$$\widehat{T}_{DEAM} := \left\{(\boldsymbol{x},\boldsymbol{y}) \in R_+^{m+s} : \boldsymbol{w}_p^*\boldsymbol{x}+\beta_p^*-\boldsymbol{\delta}_p^*\boldsymbol{y} \ge 0, \forall p \in \{1,\ldots,n\}\right\}.$$

This estimator satisfies the properties of envelopment of the data, convexity and free disposability of inputs and outputs, inherited from the single-output model.

Proposition 1. $\widehat{T}_{DEAM}$ is a convex set.

Proof. $\widehat{T}_{DEAM}$ is defined as the intersection of a set of half-spaces, which is a convex set. ■

Proposition 2. $\widehat{T}_{DEAM}$ satisfies free disposability of inputs and outputs.

Proof. Free disposability holds as $\boldsymbol{w}_i, \boldsymbol{\delta}_i \ge 0$ for all $i = 1, \ldots, n$, which is guaranteed by constraint (13.4). See (Kuosmanen & Johnson, 2010). ■

Proposition 3. $\widehat{T}_{DEAM}$ envelops the data. That is, $(\boldsymbol{x}_k, \boldsymbol{y}_k) \in \widehat{T}_{DEAM}$ for all $k \in \{1, \ldots, n\}$.

Proof. Let $(\boldsymbol{x}_k, \boldsymbol{y}_k)$ be an observation. Then, by constraint (13.3), $\boldsymbol{w}_p\boldsymbol{x}_k+\beta_p-\boldsymbol{\delta}_p\boldsymbol{y}_k \ge 0$ for all $p = 1, \ldots, n$. This guarantees that every hyperplane envelops the data from above. Thus, so does their intersection, $\widehat{T}_{DEAM}$. Therefore, $(\boldsymbol{x}_k, \boldsymbol{y}_k) \in \widehat{T}_{DEAM}$ ■.

By collecting these results together, we also satisfy the property that $\widehat{T}_{DEA} \subseteq \widehat{T}_{DEAM}$, in other words, the technology estimated by the multi-output DEAM estimator contains the DEA estimate of the technology as a subset. The hyperparameters

C and ε involved in the DEAM model will be tuned using a train-test split with a set of possible values for each hyperparameter.

Next, we adapt the list of most usual efficiency measures in DEA to be calculated with respect to the technology $\widehat{T}_{DEAM}$. The adaptation is performed by substituting $\widehat{T}_{DEAM}$ instead of the generic technology T in the corresponding definitions from Sect. 2. In particular, we show the models which calculate the efficiency of a generic input–output bundle $(\boldsymbol{x}_0, \boldsymbol{y}_0)$. We begin with the input-oriented radial measure (RI), which is calculated via the linear program:

$$\begin{array}{ll} min_{\theta,\boldsymbol{x},\boldsymbol{y}} & \theta \\ s.t. & \boldsymbol{x} \le \theta \boldsymbol{x}_o, \\ & \boldsymbol{y} \ge \boldsymbol{y}_o, \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \ge 0,\ p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y} \ge 0 \end{array} \tag{14}$$

Similarly, the output-oriented radial measure (RO) is calculated using:

$$\begin{array}{ll} max_{\phi,\boldsymbol{x},\boldsymbol{y}} & \phi \\ s.t. & \boldsymbol{x} \le \boldsymbol{x}_o, \\ & \boldsymbol{y} \ge \phi \boldsymbol{y}_o, \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \ge 0,\ p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y} \ge 0 \end{array} \tag{15}$$

We now adapt the DDF with directional vector $\boldsymbol{g} = \left(\boldsymbol{g}_o^x, \boldsymbol{g}_o^y\right) \in R_+^{m+s}$. Later, in the section devoted to the empirical applications, we will use the values of the DMU itself as components of the directional vector, that is, $\boldsymbol{g} = \left(\boldsymbol{x}_0, \boldsymbol{y}_0\right)$:

$$\begin{array}{ll} max_{\beta,\boldsymbol{x},\boldsymbol{y}} & \beta \\ s.t. & \boldsymbol{x} \le \boldsymbol{x}_o - \beta \boldsymbol{g}_o^x, \\ & \boldsymbol{y} \ge \boldsymbol{y}_o + \beta \boldsymbol{g}_o^y, \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \ge 0,\ p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y} \ge 0 \end{array} \tag{16}$$

The next measure which we present is the Weighted Additive (WA) measure with weights $\left(\boldsymbol{\alpha}_o^x, \boldsymbol{\mu}_o^y\right) \in R_{++}^{m+s}$. In the applications shown in this chapter, we choose the weights corresponding to the Range Adjusted Measure, that is, we consider weights $\boldsymbol{\alpha}_0^{x(j)} = \frac{1}{(m+s)R_j^-}$ and $\boldsymbol{\mu}_0^{y(r)} = \frac{1}{(m+s)R_r^+}$, where $R_j^- = \overline{x}_j - \underline{x}_j$ and $R_r^+ = \overline{y}_r - \underline{y}_r$ are the ranges of values of each variable. The WA is calculated by solving the slacks-based linear model as follows:

$$\begin{aligned} max_{s_0^x, s_0^y, \boldsymbol{x}, \boldsymbol{y}} \quad & \boldsymbol{\alpha}_o^x \boldsymbol{s}_o^x + \boldsymbol{\mu}_o^y \boldsymbol{s}_o^y \\ s.t. \quad & \boldsymbol{x} \leq \boldsymbol{x}_o - \boldsymbol{s}_o^x, \\ & \boldsymbol{y} \geq \boldsymbol{y}_o + \boldsymbol{s}_o^y, \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \geq 0, \; p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}_o^x, \boldsymbol{s}_o^y \geq 0 \end{aligned} \tag{17}$$

The Russell Input measure (RUI) is calculated by solving:

$$\begin{aligned} min_{\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}} \quad & \frac{1}{m} \sum_{j=1}^{m} \theta^{(j)} \\ s.t. \quad & x^{(j)} \leq \theta^{(j)} x_o^{(j)}, \qquad j = 1, \ldots, m \\ & \boldsymbol{y} \geq \boldsymbol{y}_o, \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \geq 0, \; p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y} \geq 0, \boldsymbol{\theta} \leq 1 \end{aligned} \tag{18}$$

Analogously, the Russell Output-oriented measure (RUO) has the following model:

$$\begin{aligned} max_{\boldsymbol{\phi}, \boldsymbol{x}, \boldsymbol{y}} \quad & \frac{1}{s} \sum_{r=1}^{s} \phi^{(r)} \\ s.t. \quad & \boldsymbol{x} \leq \boldsymbol{x}_o, \\ & y^{(r)} \geq \phi^{(r)} y_o^{(r)}, \qquad r = 1, \ldots, s \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \geq 0, \; p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y} \geq 0, \boldsymbol{\phi} \geq 1 \end{aligned} \tag{19}$$

Finally, the Enhanced Russell Graph (ERG) measure requires a bit more work. It is a solution of the following nonlinear problem:

$$\begin{aligned} min_{\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{x}, \boldsymbol{y}} \quad & \frac{\frac{1}{m} \sum_{j=1}^{m} \theta^{(j)}}{\frac{1}{s} \sum_{r=1}^{s} \phi^{(r)}} \\ s.t. \quad & x^{(j)} \leq \theta^{(j)} x_o^{(j)}, \qquad j = 1, \ldots, m \\ & y^{(r)} \geq \phi^{(r)} y_o^{(r)}, \qquad r = 1, \ldots, s \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \geq 0, \; p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y} \geq 0, \boldsymbol{\phi} \geq 1, \boldsymbol{\theta} \leq 1 \end{aligned} \tag{20}$$

This problem is not linear, but it can be converted into a linear problem by considering as a basis its associated slacks-based formulation (Tone, 2001):

$$\begin{aligned} min_{s_0^x, s_0^y, \boldsymbol{x}, \boldsymbol{y}} \quad & \frac{1 - \frac{1}{m} \sum_{j=1}^{m} \frac{s_o^{x(j)}}{x_o^{(j)}}}{1 + \frac{1}{s} \sum_{r=1}^{s} \frac{s_o^{y(r)}}{y_o^{(r)}}} \\ s.t. \quad & \boldsymbol{x} \leq \boldsymbol{x}_o - \boldsymbol{s}_o^x, \\ & \boldsymbol{y} \geq \boldsymbol{y}_o + \boldsymbol{s}_o^y, \\ & \boldsymbol{w}_p^* \boldsymbol{x} + \beta_p^* - \boldsymbol{\delta}_p^* \boldsymbol{y} \geq 0, \; p = 1, \ldots, n \\ & \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{s}_o^x, \boldsymbol{s}_o^y \geq 0 \end{aligned} \tag{21}$$

This model can be linearized using the method of (Charnes & Cooper, 1962), where they normalise the denominator to one via an auxiliary variable $\chi = \left(1 + \frac{1}{s}\sum_{r=1}^{s} \frac{t_o^{y(r)}}{y_o^{(r)}}\right)^{-1} > 0$. We then define transformed slack variables by $\boldsymbol{t}_0^x = \chi \boldsymbol{s}_0^x$ and $\boldsymbol{t}_0^y = \chi \boldsymbol{s}_0^y$, and substitute the original input and output vectors via $\widetilde{\boldsymbol{x}} = \chi \boldsymbol{x}$ and $\widetilde{\boldsymbol{y}} = \chi \boldsymbol{y}$. This results in the following linear problem in $\chi, \boldsymbol{t}_0^x, \boldsymbol{t}_0^y, \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}$, which we solve:

$$\begin{aligned} min_{\chi, \boldsymbol{t}_o^x, \boldsymbol{t}_o^y, \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}} \quad & \chi - \frac{1}{m}\sum_{j=1}^{m} \frac{t_o^{x(j)}}{x_o^{(j)}} \\ s.t. \quad & \chi + \frac{1}{s}\sum_{r=1}^{s} \frac{t_o^{y(r)}}{y_o^{(r)}} = 1, \\ & \widetilde{\boldsymbol{x}} \leq \chi \boldsymbol{x}_o - \boldsymbol{t}_o^x, \\ & \widetilde{\boldsymbol{y}} \geq \chi \boldsymbol{y}_o + \boldsymbol{t}_o^y, \\ & \boldsymbol{w}_p^* \widetilde{\boldsymbol{x}} + \chi \beta_p^* - \boldsymbol{\delta}_p^* \widetilde{\boldsymbol{y}} \geq 0, \; p = 1, \ldots, n \\ & \widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}}, \boldsymbol{t}_o^x, \boldsymbol{t}_o^y \geq 0, \\ & \chi \geq 0 \end{aligned} \tag{22}$$

Thus, we can calculate all six measures by solving linear programming models.

4 Computational Experiences

In this section, we illustrate the use of the multi-output DEAM model using two empirical datasets. We use these two examples to illustrate the efficiencies obtained with respect to the various measures introduced using both models, and the differences between DEAM and DEA.

The first dataset contains 51 schools located in the region of Madrid (Spain). The dataset comes from the Organisation for Economic Co-operation and Development (OECD), from its report on the quality of teaching in countries through the Programme for International Student Assessment (PISA) (OECD, 2014). The second dataset, used in (Aparicio et al., 2020), contains 44 DMUs. In this case, the DMUs are regional tax offices in Spain that oversee the tax collection in almost all Spanish provinces.

We calculate the efficiency of each DMU according to the six measures of efficiency presented in the previous section. These are the Radial measures (both input and output-oriented; RI and RO), the Directional Distance Function (DDF), the Weighted Additive (WA) measure, the Russell measures (RUI, RUO), and the Enhanced Russell Graph (ERG).

We recall that the input-oriented measures of efficiency (RI and RUI) have efficient units with a score of 1, and their range of possible values is (0,1]. Meanwhile, the output-oriented measures (RO and RUO) also have values of 1 for efficient units, but in these case inefficient DMUs have values strictly greater than 1. The DDF and WA are measures of inefficiency and thus efficient units achieve values of 0, while inefficient units attain strictly positive values. Finally, the ERG also has values in (0,1], with efficient units achieving the value 1.

The DEAM estimates of the efficiencies were obtained by first selecting the best hyperparameters from a grid of possible values. In particular, $C \in \{1,10,50,100,10^6\}$, while $\varepsilon \in \{0,0.001,0.01,0.1,0.2,0.3,0.4,0.5,1\}$. The dataset was split, at random, into a training set consisting of two thirds of the DMUs while the remaining third was reserved for the test set. Once the best hyperparameters were selected with this train-test split, the DEAM model was solved on the whole dataset and the efficiency of each DMU was calculated for each measure of efficiency.

The models were solved using a PC with an Intel® Core™ i7-8700 CPU 3.2 GHz with 6 cores, 32 Gigabyte of RAM and operating system Windows 10 Home 64 bit. CPLEX v12.10 was used to solve the optimization problems. DEAM spent 143 s on the first dataset, and 74 s for the second one.

4.1 Schools in Madrid

The first dataset that we consider comes from the education context. It is a dataset obtained from the PISA report by the OECD in 2012, where the DMUs are schools in the region of Madrid, in Spain. The inputs used are the quality of the educational resources of the school (SCMATEDU, x_1), the economic, social and cultural status index of the students (ESCS, x_2), and the number of teachers per (hundred) students (TEACHERS, x_3). This means that one indicator represents the physical resources available, one represents the socioeconomic status of the students, and one represents the human capital in each school. The variables used as outputs of the process are the scores in standardized tests in mathematics (PVMATH, y_1) and reading (PVREAD, y_2). The dataset is made up of 51 schools that were selected by the OECD at random as a representative sample of the schools in the region of Madrid. Note how, in this situation, it seems more interesting to determine absolute technical efficiency rather than relative technical efficiency, since we do not have information about all the schools in this region of Spain. Additionally, DEAM is able to provide certain insights on the education production function of the sector beyond the sample observed.

The efficiencies calculated by DEA and DEAM with respect to each measure are reported in Table 1. Each row indicates one DMU, with the final rows indicating the average efficiency obtained by each method and the number of DMUs considered efficient in each case. We observe a higher value of average inefficiency estimated by DEAM than DEA in all measures. This is because the DEA frontier is closer to the data than the DEAM frontier, as a consequence of the minimal extrapolation principle. In contrast, DEAM tries to estimate the actual education production function that is behind the generation of this particular sample of schools. Additionally, the differences in the average efficiency values range between 2 and 25%, with the latter attained at the DDF. Furthermore, Fig. 2 shows the distributions of the efficiencies measured by both methods.

Table 1 Results on the PISA dataset involving schools in Madrid

						DEA							DEAM						
DMU	x_1	x_2	x_3	y_1	y_2	RI	RO	DDF	WA	RUI	RUO	ERG	RI	RO	DDF	WA	RUI	RUO	ERG
1	2.547	5.28	7.669	538.1	540.12	0.943	1.010	0.009	0.065	0.912	1.011	0.911	0.896	1.054	0.035	0.110	0.888	1.071	0.880
2	3.886	5.61	8.355	526.55	541.01	0.774	1.056	0.051	0.226	0.725	1.066	0.724	0.770	1.140	0.088	0.240	0.715	1.158	0.705
3	3.43	3.91	0.719	462.92	435.57	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.002	0.996	1.000	0.989
4	3.079	4.52	5.569	482.7	504.57	0.907	1.064	0.043	0.121	0.857	1.083	0.843	0.900	1.078	0.049	0.127	0.834	1.090	0.829
5	5.576	4.27	9.219	475.03	490.62	0.879	1.124	0.068	0.346	0.622	1.124	0.611	0.840	1.159	0.089	0.357	0.609	1.174	0.601
6	3.618	5.34	7.858	515.33	497.42	0.790	1.102	0.079	0.256	0.733	1.115	0.707	0.790	1.134	0.085	0.259	0.717	1.188	0.698
7	2.73	4.83	7.426	522	501.03	0.893	1.038	0.029	0.137	0.864	1.066	0.831	0.893	1.054	0.036	0.145	0.842	1.111	0.822
8	3.43	3.56	8.293	466.25	462.29	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.009	0.988	1.000	0.961
9	3.43	4.55	5.504	521.99	528.37	0.962	1.023	0.015	0.053	0.925	1.028	0.924	0.953	1.027	0.017	0.065	0.902	1.031	0.891
10	3.618	4.31	7.843	439.71	477.43	0.876	1.176	0.079	0.278	0.712	1.187	0.677	0.872	1.195	0.091	0.283	0.680	1.203	0.664
11	5.576	5.89	8.247	575.22	596.57	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.859	1.087	0.054	0.233	0.773	1.092	0.773
12	3.824	4.24	12.55	476.09	485.35	0.873	1.117	0.065	0.296	0.660	1.127	0.639	0.867	1.119	0.073	0.311	0.643	1.169	0.633
13	4.352	5.64	8.055	502.06	491.3	0.721	1.145	0.126	0.350	0.639	1.157	0.627	0.721	1.208	0.127	0.354	0.629	1.249	0.618
14	4.352	5.6	6.407	563.5	559.93	0.912	1.020	0.019	0.104	0.898	1.021	0.898	0.888	1.064	0.041	0.131	0.857	1.072	0.851
15	4.352	3.76	10.241	399.56	462.14	0.947	1.070	0.030	0.235	0.765	1.141	0.703	0.930	1.143	0.051	0.266	0.707	1.188	0.664
16	3.618	4.89	7.341	488.98	475.76	0.808	1.148	0.096	0.265	0.695	1.166	0.689	0.806	1.150	0.096	0.269	0.684	1.196	0.677
17	4.352	4.48	6.818	527.74	518.22	0.944	1.041	0.025	0.155	0.826	1.049	0.812	0.943	1.042	0.025	0.177	0.808	1.071	0.794
18	2.73	4.03	7.979	506.97	517.54	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.021	0.920	1.000	0.973
19	2.353	4.98	8.432	487.05	489.73	0.910	1.101	0.072	0.202	0.837	1.103	0.783	0.875	1.132	0.087	0.206	0.762	1.144	0.763
20	3.252	4.53	8.139	485.73	488.29	0.868	1.114	0.073	0.214	0.736	1.134	0.732	0.866	1.114	0.074	0.218	0.717	1.154	0.717
21	3.824	4.79	6.677	557.72	558.29	0.995	1.002	0.001	0.016	0.962	1.005	0.962	0.993	1.003	0.002	0.050	0.948	1.017	0.933

(continued)

Table 1 (continued)

						DEA							DEAM						
DMU	x_1	x_2	x_3	y_1	y_2	RI	RO	DDF	WA	RUI	RUO	ERG	RI	RO	DDF	WA	RUI	RUO	ERG
22	3.618	3.89	9.254	481.47	484.08	0.955	1.039	0.021	0.102	0.826	1.046	0.825	0.943	1.056	0.030	0.129	0.783	1.079	0.802
23	4.061	4.91	11.801	518.66	520.18	0.835	1.089	0.062	0.279	0.652	1.090	0.639	0.830	1.095	0.062	0.283	0.646	1.152	0.635
24	3.252	4.06	8.379	455.6	490.63	0.936	1.106	0.042	0.148	0.797	1.112	0.769	0.935	1.133	0.048	0.179	0.783	1.137	0.757
25	3.824	5.16	4.889	535.09	515.14	0.932	1.021	0.016	0.084	0.927	1.030	0.920	0.906	1.049	0.032	0.100	0.880	1.063	0.874
26	3.252	3.86	10.131	450.2	465.29	0.966	1.095	0.033	0.143	0.810	1.095	0.788	0.966	1.115	0.034	0.171	0.780	1.116	0.766
27	5.576	4.17	7.986	544.77	531.85	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.012	0.999	1.000	0.970
28	3.252	5.37	7.253	559.58	530.52	0.977	1.005	0.005	0.063	0.943	1.026	0.925	0.924	1.040	0.026	0.134	0.909	1.091	0.875
29	3.252	5.35	8.157	514.38	507.43	0.789	1.093	0.071	0.229	0.744	1.094	0.726	0.789	1.130	0.083	0.233	0.726	1.178	0.714
30	5.576	4.39	7.39	492.08	515.49	0.911	1.083	0.048	0.284	0.697	1.085	0.688	0.867	1.105	0.067	0.288	0.682	1.122	0.679
31	3.43	4.46	7.458	492.5	480.55	0.882	1.095	0.059	0.209	0.744	1.129	0.736	0.880	1.095	0.059	0.213	0.731	1.142	0.725
32	4.762	4.04	5.155	480.81	467.74	0.953	1.054	0.025	0.160	0.708	1.054	0.708	0.933	1.062	0.033	0.176	0.697	1.094	0.701
33	4.061	5.08	6.193	574.48	569.54	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.994	1.003	0.002	0.014	0.991	1.011	0.985
34	3.683	5.13	8.432	489.58	476.31	0.770	1.160	0.109	0.307	0.659	1.171	0.652	0.768	1.174	0.109	0.310	0.644	1.232	0.639
35	4.061	4.7	5.61	557.63	551.83	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.002	1.000	1.000	0.995
36	3.43	4.81	7.838	477.37	516.09	0.840	1.100	0.082	0.219	0.748	1.125	0.727	0.833	1.150	0.093	0.223	0.728	1.157	0.713
37	2.73	4.17	8.429	518.05	560.68	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	0.957
38	4.061	4.53	7.326	520.44	517.37	0.918	1.055	0.036	0.170	0.795	1.063	0.776	0.918	1.055	0.037	0.176	0.780	1.088	0.768
39	3.252	5.69	6.572	551.23	559.5	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.900	1.053	0.034	0.108	0.896	1.054	0.887
40	2.906	4.58	8.432	478.26	487.24	0.886	1.128	0.083	0.213	0.760	1.132	0.744	0.886	1.129	0.084	0.217	0.730	1.164	0.726
41	2.353	4.25	8.705	503.51	500.29	0.989	1.028	0.011	0.098	0.879	1.059	0.853	0.989	1.029	0.011	0.113	0.846	1.070	0.846

(continued)

Table 1 (continued)

						DEA							DEAM						
DMU	x_1	x_2	x_3	y_1	y_2	RI	RO	DDF	WA	RUI	RUO	ERG	RI	RO	DDF	WA	RUI	RUO	ERG
42	3.252	4.22	7.727	472.39	472.68	0.914	1.112	0.059	0.195	0.744	1.137	0.736	0.911	1.113	0.060	0.213	0.722	1.152	0.710
43	3.43	4.11	9.615	476.23	475.28	0.910	1.092	0.049	0.193	0.716	1.118	0.709	0.910	1.096	0.057	0.218	0.700	1.152	0.693
44	2.142	4.31	6.509	518.9	541.13	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.007	0.990	1.000	0.993
45	4.352	4.78	6.03	531.72	566.63	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.981	1.010	0.007	0.070	0.941	1.029	0.911
46	2.547	5.13	8.432	500.74	555.01	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.867	1.055	0.039	0.134	0.845	1.089	0.833
47	3.43	4.11	7.759	478.77	489.54	0.927	1.082	0.040	0.139	0.763	1.089	0.753	0.922	1.084	0.044	0.166	0.758	1.113	0.739
48	2.547	5.51	6.587	552	542.14	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.987	1.006	0.004	0.050	0.976	1.017	0.962
49	3.43	4.63	7.24	531.27	546.76	0.942	1.030	0.020	0.077	0.868	1.031	0.868	0.939	1.031	0.021	0.098	0.853	1.044	0.846
50	3.824	4.87	5.82	499.34	529.09	0.899	1.051	0.036	0.154	0.828	1.086	0.817	0.875	1.074	0.047	0.160	0.812	1.094	0.807
51	3.079	4.35	7.865	483.88	517.75	0.922	1.086	0.051	0.150	0.808	1.089	0.789	0.916	1.096	0.057	0.153	0.788	1.100	0.776
MEAN						0.923	1.058	0.036	0.136	0.839	1.068	0.828	0.906	1.077	0.045	0.160	0.809	1.100	0.798
# Efficient DMUs						13	13	13	13	13	13	13	7	7	7	1	2	7	0

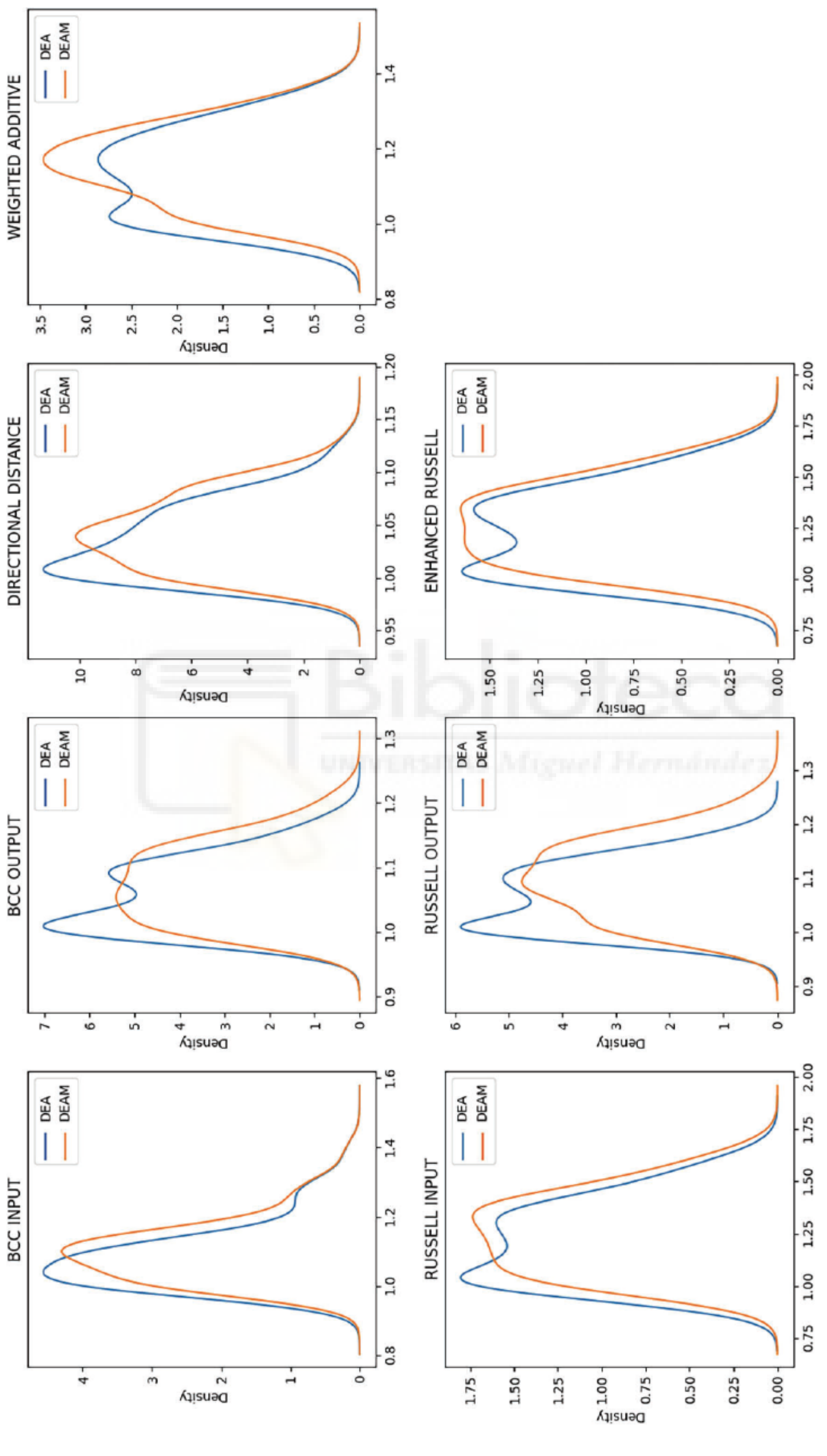

Fig. 2 Efficiency distributions measured by DEA and DEAM for the dataset of schools in Madrid

Regarding the number of DMUs classified as efficient by each method, we observe that DEAM has higher discriminating capability that DEA. DEAM classifies 7 or fewer DMUs as efficient while DEA classifies 13 DMUs as efficient, so that the number of efficient DMUs is reduced by almost a half. Every DMU considered efficient by DEAM was already considered efficient by DEA. We also observe that, with respect to certain measures, the number of DMUs considered efficient is even lower. These measures are the Weighted Additive (1 efficient DMU), Russell Input (2 efficient DMUs) and Enhanced Russell Graph (zero efficient units). Thus, we observe that DEAM can detect inefficiencies in certain components better than DEA in this application.

To illustrate the differences measured, we examine the estimated efficiencies of two schools in the dataset. The first school that we compare is school 46, which is considered efficient by DEA, but inefficient by DEAM. This school has above-average values of ESCS (5.13) and teachers per 100 students (8.432), but its educational resources are close to the minimum of the dataset (2.547). Its students attain scores slightly lower than the average in mathematics (500.74), while their scores in reading are high (555.01). With these values, this unit is considered efficient by DEA, while DEAM estimates its RI at 0.867. Its DEAM efficiency with respect to the RUI and ERG are below 0.85. Its WA is 0.134, which is slightly below the average of 0.160. The RO, DDF, RUO efficiency values of this DMU are closer to efficient. These values indicate that DEAM proposes that its inputs can be decreased by around 15% while still achieving similar outputs (test scores), but that, with the resources available, little further improvement on the test scores is possible. The second school that we observe in more detail is school 5, which has very high educational resources (5.576) and teachers (9.219), but its ESCS is below average (4.27). This school, despite the resources available, only achieves low scores in the standardized tests. In fact, this school obtains results below 490 in both subjects. These scores fall short of the average scores of the PISA overall dataset, which are 494 in reading and 496 in mathematics in 2012. This school is considered inefficient by both methods, but while DEA estimates at 0.879 its Radial input efficiency, DEAM evaluates it at 0.84. In this case, DEAM is capable of detecting further inefficiencies than DEA in all measures, but particularly with respect to the radial (RI and RO), DDF and Russell Output measures.

Due to the nature of the data as a sample of a larger dataset, another viewpoint in which we wish to focus is the possible perspective of a decision maker when faced with data not in the dataset. In this case, the estimation of absolute technical efficiency and information about the production function beyond the sample is useful. A decision maker, with data about a school potentially not in the observed sample, may be interested in evaluating it with respect to the estimated (education) production technology. In order to work with the DEAM estimated technology, one needs only the hyperplane parameters estimated by DEAM. Given these parameters, one can compare any point in the input–output space with this convex set by solving a system of linear equations and determining whether it satisfies each inequality. One possible case of interest is the allocation of educational resources and teachers to enable schools to obtain at least the overall average PISA scores in 2012, which are 494 in

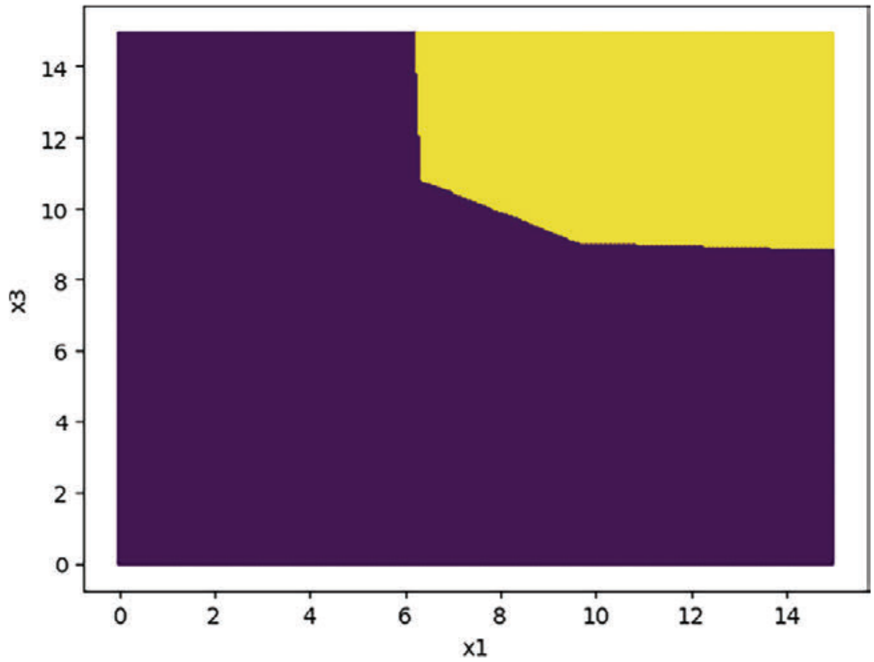

Fig. 3 DEAM level curve of feasible average PISA results given ESCS = 3.60

mathematics and 496 in reading. For this task, the ESCS of the students, which is determined by the area in which the school is located, is not easily modified, and can be considered fixed. For example, given a school with an ESCS close to the lowest observed value of 3.56, to allocate resources so that these scores were feasible with respect to the estimated production function. A hypothetical school with ESCS of 3.60 and average values in the remaining input cannot achieve these scores. Figure 3 represents the combinations of educational resources and teachers which yield these scores feasible. It can be observed that large values would be necessary, which are not present in any observed unit. Even a hypothetical allocation equal to the largest values in the observed sample, with $x_1 = 5.576$ and $x_3 = 12.55$, would not yield a feasible unit, given $x_2 = 3.60$. Figure 4 represents the feasible region with an ESCS value of 4.00, where the feasible region includes much lower values of allocated resources. Thus, an increase of around 10% in the ESCS of this hypothetical school would make it much easier to allocate resources towards reaching the proposed objective. Therefore, something else other than investment in teachers or resources is necessary, such as mid- or long-term changes in the socioeconomic fabric of the neighborhood, or a change in the production technology (new disruptive educational strategies), among other possibilities.

4.2 Spanish Regional Tax Offices

We now analyse the efficiency of tax offices from the regions (provinces) into which Spain is subdivided. This dataset can be found in (Aparicio et al., 2020). The data

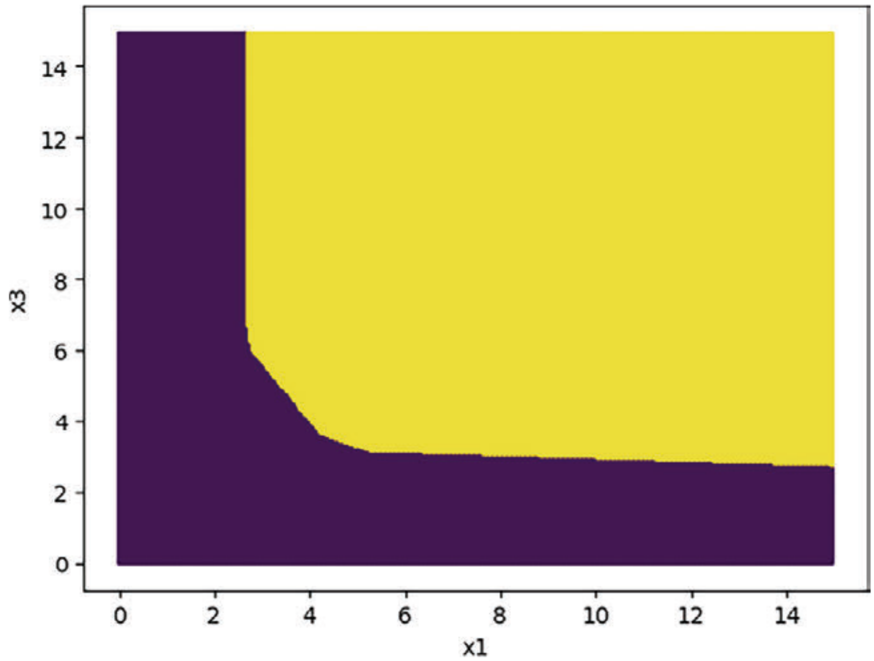

Fig. 4 DEAM level curve of feasible average PISA results given ESCS = 4.00

is from the year 2008. In this case, we consider three variables as inputs and two as outputs. There are two inputs related to labor: the number of tax inspectors and specialists (x_1) and the number of workers in the rest of the workforce (x_2). The third variable considered as an input (x_3) indicates the number of complaints against tax authority that have been positively resolved in favor of the taxpayers, showing mistakes by the tax office personnel. This is a bad output, that is, an output variable which is undesirable and thus it is a variable which a manager should aim to minimize, so this variable is considered as an input in our model, as in (Hailu & Veeman, 2000; Mahlberg & Sahoo, 2011). The outputs considered are the two main types of taxes which are collected by the offices under study: the inheritance and gift tax (y_1), and the real estate transfer tax settlements processed (y_2).

We report the results obtained in Table 2, where we indicate the estimated efficiencies for each DMU, as well as the average efficiency according to each measure, and the number of DMUs considered efficient. The average efficiency is deemed slightly lower by DEAM than by DEA with respect to all measures. Furthermore, we can observe that DEAM classifies 10 DMUs as efficient, while DEA considers 15 DMUs efficient, showing the higher discrimination capability of DEAM. These observations indicate that DEAM obtains estimates of efficiency that fit the data less closely, thus obtaining estimates of the production frontier that do not suffer from the overfitting problem of DEA.

Figure 5 then shows the density of the efficiencies measured, where we can see the higher average efficiencies obtained by DEAM, particularly in the cases of the Russell Input and Enhanced Russell Graph measures, which could already be observed in the PISA dataset.

Table 2 Results on the Spanish tax offices dataset

	DEA							DEAM						
DMU	RI	RO	DDF	WA	RUI	RUO	ERG	RI	RO	DDF	WA	RUI	RUO	ERG
ALMERIA	0.554	2.150	0.331	0.107	0.414	2.236	0.317	0.527	2.150	0.335	0.110	0.361	2.254	0.315
CADIZ	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	0.996
CORDOBA	0.523	1.994	0.323	0.069	0.412	2.132	0.378	0.523	1.994	0.323	0.070	0.369	2.157	0.354
GRANADA	0.653	1.609	0.224	0.103	0.456	1.671	0.448	0.653	1.609	0.224	0.104	0.451	1.671	0.446
HUELVA	0.521	2.473	0.377	0.064	0.393	2.565	0.290	0.506	2.474	0.379	0.069	0.317	2.569	0.287
JAEN	0.595	1.743	0.265	0.066	0.470	2.161	0.394	0.589	1.752	0.268	0.067	0.428	2.193	0.379
MALAGA	0.577	1.742	0.270	0.099	0.513	1.948	0.428	0.554	1.821	0.289	0.100	0.426	1.948	0.388
SEVILLA	0.739	1.305	0.156	0.138	0.537	1.504	0.518	0.739	1.385	0.156	0.139	0.532	1.504	0.514
HUESCA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
TERUEL	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.932	1.180	0.049	0.004	0.589	1.261	0.561
ZARAGOZA	0.760	1.325	0.138	0.041	0.672	1.347	0.601	0.760	1.325	0.138	0.042	0.546	1.424	0.541
OVIEDO	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
BALEARES	0.977	1.024	0.012	0.022	0.821	1.034	0.821	0.930	1.075	0.036	0.024	0.771	1.124	0.768
CANTABRIA	0.862	1.156	0.073	0.024	0.704	1.493	0.531	0.862	1.156	0.073	0.026	0.603	1.493	0.512
ALBACETE	0.969	1.041	0.018	0.007	0.873	1.056	0.873	0.945	1.073	0.031	0.008	0.860	1.097	0.854
CIUDAD REAL	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
CUENCA	0.990	1.017	0.006	0.004	0.923	1.018	0.923	0.962	1.068	0.024	0.005	0.840	1.090	0.831
GUADALAJARA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.995	1.007	0.003	0.004	0.943	1.017	0.935
TOLEDO	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.995	1.006	0.003	0.002	0.972	1.008	0.972
AVILA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.978	1.061	0.016	0.004	0.799	1.074	0.748
BURGOS	0.879	1.167	0.070	0.014	0.757	1.373	0.649	0.870	1.184	0.077	0.015	0.683	1.412	0.647

(continued)

Table 2 (continued)

	DEA							DEAM						
DMU	RI	RO	DDF	WA	RUI	RUO	ERG	RI	RO	DDF	WA	RUI	RUO	ERG
LEÓN	0.975	1.026	0.013	0.010	0.943	1.292	0.774	0.969	1.033	0.016	0.011	0.922	1.308	0.743
PALENCIA	0.850	1.448	0.126	0.009	0.659	1.674	0.534	0.803	1.451	0.143	0.013	0.490	1.878	0.403
SALAMANCA	0.953	1.060	0.026	0.005	0.879	1.156	0.834	0.941	1.078	0.034	0.006	0.874	1.179	0.803
SEGOVIA	0.828	1.457	0.131	0.009	0.620	1.608	0.559	0.804	1.457	0.143	0.013	0.496	1.702	0.463
SORIA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.001	1.000	1.000	0.905
VALLADOLID	0.745	1.462	0.164	0.033	0.551	1.574	0.471	0.745	1.462	0.164	0.034	0.516	1.574	0.466
ZAMORA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
BARCELONA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
GIRONA	0.825	1.244	0.105	0.013	0.724	1.245	0.707	0.818	1.246	0.105	0.014	0.565	1.256	0.562
LLEIDA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	0.992	1.009	0.004	0.002	0.955	1.013	0.948
TARRAGONA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
BADAJOZ	0.650	1.687	0.232	0.081	0.407	1.715	0.388	0.650	1.687	0.232	0.083	0.407	1.715	0.380
CACERES	0.955	1.077	0.029	0.006	0.795	1.077	0.795	0.928	1.124	0.046	0.009	0.775	1.126	0.775
A CORUÑA	0.722	1.410	0.166	0.055	0.588	1.527	0.528	0.722	1.410	0.166	0.057	0.543	1.539	0.518
LUGO	0.715	1.564	0.190	0.033	0.526	2.098	0.398	0.715	1.564	0.190	0.034	0.495	2.106	0.380
OURENSE	0.458	3.122	0.453	0.051	0.331	3.231	0.227	0.441	3.152	0.464	0.055	0.260	3.303	0.202
PONTEVEDRA	0.545	1.995	0.313	0.100	0.416	2.066	0.363	0.545	2.004	0.313	0.101	0.400	2.066	0.358
LA RIOJA	0.859	1.330	0.099	0.030	0.545	1.331	0.521	0.859	1.330	0.099	0.033	0.501	1.331	0.477
MADRID	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
MURCIA	0.470	2.029	0.366	0.142	0.397	2.194	0.359	0.470	2.169	0.366	0.146	0.380	2.196	0.352

(continued)

Table 2 (continued)

	DEA							DEAM						
DMU	RI	RO	DDF	WA	RUI	RUO	ERG	RI	RO	DDF	WA	RUI	RUO	ERG
ALICANTE	0.994	1.006	0.003	0.140	0.741	1.007	0.741	0.936	1.068	0.033	0.210	0.676	1.089	0.620
CASTELLÓN	0.905	1.108	0.051	0.089	0.655	1.299	0.544	0.893	1.122	0.057	0.090	0.654	1.304	0.536
VALENCIA	1.000	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	1.000	1.000	1.000
MEAN	0.842	1.336	0.108	0.036	0.744	1.423	0.703	0.831	1.356	0.114	0.039	0.691	1.454	0.658
# Efficient DMUs	15	15	15	15	15	15	15	10	10	10	9	10	10	8

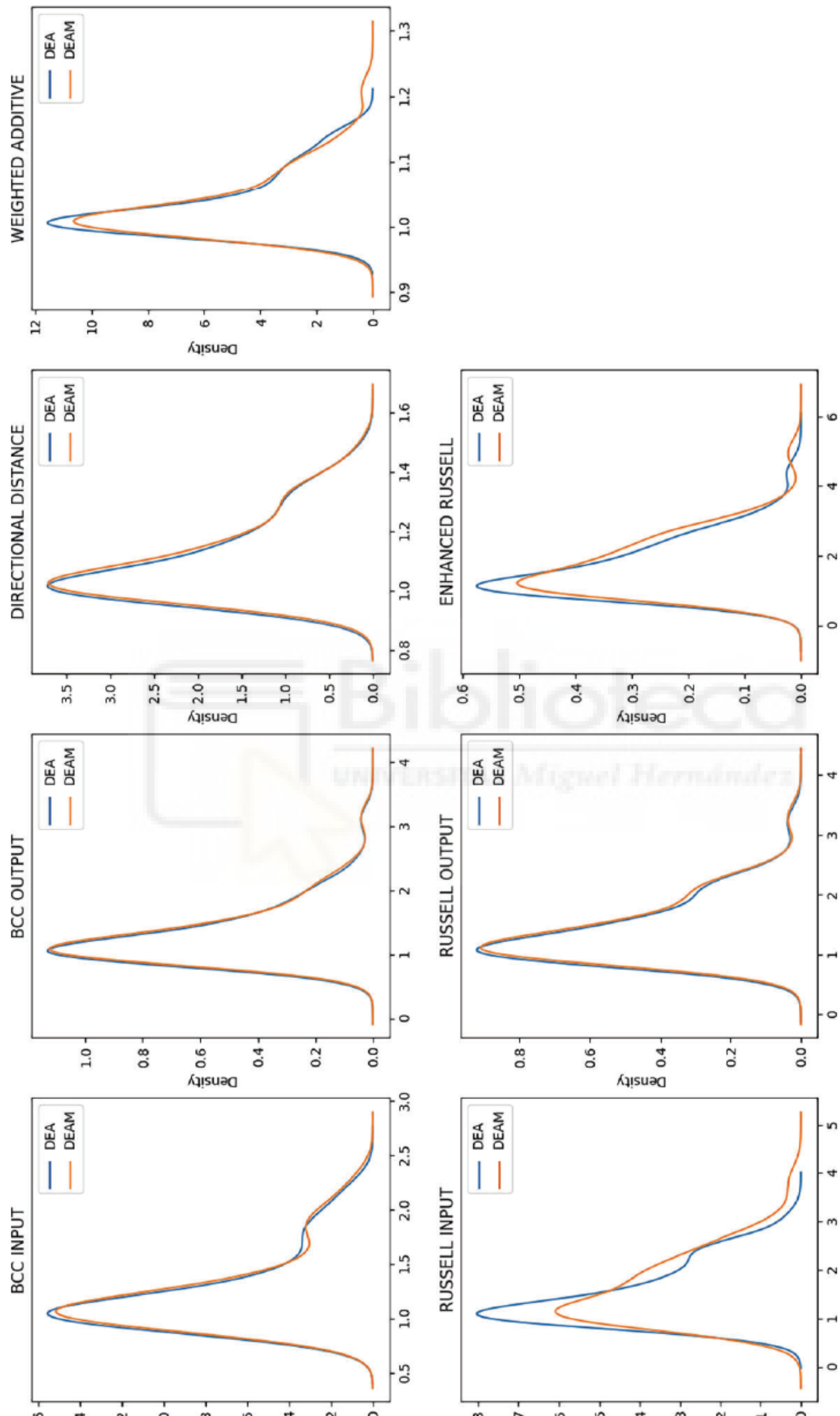

Fig. 5 Efficiency distributions measured by DEA and DEAM for the dataset of tax offices in Spain

As with the previous dataset, the units considered efficient by DEAM were already considered efficient by DEA, but the converse does not hold. The provinces which are considered efficient by DEA but not by DEAM include some of the provinces with the smallest populations, such as Teruel (0.932 RI) and Avila (0.978 RI), where the changes in efficiency are slightly higher, but also some larger provinces such as Guadalajara, Toledo, and Lleida, which are considered almost fully efficient by DEAM. For example, the differences between DEA and DEAM in the case of Lleida are smaller than 0.01 with respect to the RI, RO, DDF and RUO measures, with larger changes of up to 0.045 in the RUI measure and 0.052 in the ERG.

The largest changes in efficiency can be observed in Teruel, Alicante, Baleares, and Palencia. In particular, Alicante has a large variation between the measured efficiencies depending on the measure. DEAM considers this unit almost efficient in the output-oriented measures (RO and RUO) and the DDF, but it is the least efficient according to the WA and achieves efficiencies lower than 0.70 with respect to the RUI and ERG. These measures are capable of detecting inefficiencies in a single input, and it can be noted that this unit attains the maximum of the dataset in input x_3, with above average values in the remaining inputs and outputs.

5 Conclusions

In the nonparametric literature of efficiency estimation, one of the advantages of DEA and related estimators is the natural multi-output extension and interpretation of the models. However, methods such as DEA are descriptive in nature, and do not naturally work to perform inference. There have been recent contributions of models which combine DEA with machine learning principles (see, for example, (Kuosmanen & Johnson, 2010), (Du et al., 2013), (Esteve et al., 2020), (Valero-Carreras et al., 2021), (Olesen & Ruggiero, 2022)). A recent contribution, (Guerrero et al., 2022) proposed a technique, called Data Envelopment Analysis-based Machines (DEAM), which follows the Structural Risk Minimization principle in order to overcome the descriptive nature of DEA. This is achieved via the statement of theoretical bounds on both the empirical and generalization error of the model. However, the DEAM technique was only introduced in the single-output production context. In this chapter, we have extended DEAM to estimate efficiency in production processes with multiple outputs and multiple inputs while satisfying the properties of convexity, free disposability, and envelopment of the observed data, but not that of minimal extrapolation.

Moreover, we have described how some well-known measures of efficiency can be adapted to calculate efficiency with the multi-output DEAM technique. These include both oriented and non-oriented measures which calculate weak or Pareto efficiency, in a variety of possible orientations. We have illustrated this approach with applications to two real datasets in the education and tax collection contexts. In both cases, DEAM shows a higher discriminating power than DEA, classifying fewer DMUs as efficient than DEA, and obtaining efficiency scores which indicate lower efficiencies. This indicates that we obtain results which overcome the overfitting present in DEA, estimating production frontiers which are more robust. Furthermore, this can be understood as identifying more sources of inefficiency.

These additional sources of inefficiency can be observed in all the adapted measures, but they are highlighted even more with respect to some measures, such as the Weighted Additive, Russell Input and Enhanced Russell Graph measures, where DEAM classifies an even smaller number of units as fully efficient, thus being able to identify more inefficiencies with respect to these measures. These are slacks-based measures which measure Pareto efficiency, and can identify inefficiencies along individual variables, in the inputs (Russell Input) or in both inputs and outputs simultaneously (WA, ERG). More examples should be considered to evaluate whether this applies in general.

A limitation of the DEAM estimator is the intensive computational burden placed by cross-validation of the hyperparameters. This could be further explored and improved upon by the consideration of other methods for estimating the hyperparameters or by the simplification of models.

Finally, we mention some possible avenues for further research. Given the Structural Risk Minimization bounds on DEAM, a direction worth studying are the properties of asymptotic convergence and consistency results for DEAM (Christmann & Steinwart, 2007; Steinwart, 2002, 2005). Another possible direction worth considering is the extension of DEAM to work with panel data, evaluating the differences in efficiency arising from the changes in efficiency and productivity over time (efficiency change, scale efficiency change, technical change) (Tsionas, 2022). Another interesting generalization could consider the measurement and decomposition of the economic inefficiencies into their components (allocative, technical, etc.) (Aparicio & Zofío, 2023). Another potential avenue comes from the Support Vector Regression literature, where transformation functions are often considered which transform the input space into a higher-dimensional space allowing for better estimations and yielding more flexibility to the technique (Vapnik, 1998). Finally, with single-output estimation methods, there are often multiple proposed approaches to generalize them to the multi-output prediction context, and other such techniques could also be considered (Vazquez & Walter, 2003).

Acknowledgements The authors thank the grant PID2022-136383NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU. This work was also supported by the Generalitat Valenciana under Grant ACIF/2020/155. Additionally, J. Aparicio thanks the grant PROMETEO/2021/063 funded by the Valencian Community (Spain).

References

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*(1), 21–37. https://doi.org/10.1016/0304-4076(77)90052-5

Aparicio, J., Cordero, J. M., & Díaz-Caro, C. (2020). Efficiency and productivity change of regional tax offices in Spain: An empirical study using Malmquist-Luenberger and Luenberger indices. *Empirical Economics, 59*(3), 1403–1434. https://doi.org/10.1007/s00181-019-01667-8

Aparicio, J., & Esteve, M. (2022). How to peel a data envelopment analysis frontier: A cross-validation-based approach. *Journal of the Operational Research Society, 1–15*. https://doi.org/10.1080/01605682.2022.2157765

Aparicio, J., Esteve, M., Rodriguez-Sala, J. J., & Zofio, J. L. (2021). The Estimation of Productive Efficiency Through Machine Learning Techniques: Efficiency Analysis Trees. In *International Series in Operations Research and Management Science* (Vol. 312, pp. 51–92). Springer. https://doi.org/10.1007/978-3-030-75162-3_3.

Aparicio, J., & Zofío, J. L. (2023). Decomposing profit change: Konüs, Bennet and Luenberger indicators. *Socio-Economic Planning Sciences, 101573*. https://doi.org/10.1016/J.SEPS.2023.101573

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications, 98*(2), 351–364. https://doi.org/10.1023/A:1022637501082

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly, 9*(3–4), 181–186. https://doi.org/10.1002/nav.3800090303

Charnes, A., Cooper, W. W., Golany, B., Seiford, L., & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics, 30*(1–2), 91–107. https://doi.org/10.1016/0304-4076(85)90133-2

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

Christmann, A., & Steinwart, I. (2007). Consistency and robustness of kernel-based regression in convex risk minimization. *Bernoulli, 13*(3), 799–819. https://doi.org/10.3150/07-BEJ5102

Cooper, W. W., Park, K. S., & 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*(1), 5–42. https://doi.org/10.1023/A:1007701304281.

Cooper, W. W., Pastor, J. T., Borras, F., Aparicio, J., & Pastor, D. (2011). BAM: A bounded adjusted measure of efficiency for use with bounded additive models. *Journal of Productivity Analysis, 35*(2), 85–94. https://doi.org/10.1007/s11123-010-0190-2

Daouia, A., Noh, H., & 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. https://doi.org/10.1111/RSSB.12098.

Daraio, C., & Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis* (Vol. 4). Springer US. https://doi.org/10.1007/978-0-387-35231-2.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica, 19*(3), 273. https://doi.org/10.2307/1906814

Du, P., Parmeter, C. F., & Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. *Statistica Sinica*, 1347–1371.

Esteve, M., Aparicio, J., Rabasa, A., & Rodriguez-Sala, J. J. (2020). Efficiency analysis trees: A new methodology for estimating production frontiers through decision trees. *Expert Systems with Applications, 162*, 113783. https://doi.org/10.1016/j.eswa.2020.113783

Esteve, M., Aparicio, J., Rodriguez-Sala, J. J., & 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. https://doi.org/10.1016/J.EJOR.2022.04.024

Färe, R., & Lovell, C. A. 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 (General), 120*(3), 253. https://doi.org/10.2307/2343100.

Guerrero, N. M., Aparicio, J., & Valero-Carreras, D. (2022). Combining data envelopment analysis and machine learning. *Mathematics, 10*(6), 909. https://doi.org/10.3390/MATH10060909

Guillen, M. D., Aparicio, J., & Esteve, M. (2023). Gradient tree boosting and the estimation of production frontiers. *Expert Systems with Applications, 214*, 119134. https://doi.org/10.1016/J.ESWA.2022.119134

Hailu, A., & Veeman, T. S. (2000). Environmentally sensitive productivity analysis of the Canadian pulp and paper industry, 1959–1994: an input distance function approach. *Journal of Environmental Economics and Management, 40*(3), 251–274. https://doi.org/10.1006/JEEM.2000.1124

Kneip, A., Park, B. U., & Simar, L. (1998). A note on the convergence of nonparametric Dea estimators for production efficiency scores. *Econometric Theory, 14*(6), 783–793. https://doi.org/10.1017/S0266466698146042

Kneip, A., Simar, L., & Wilson, P. W. (2008). Asymptotics and consistent bootstraps for Dea estimators in nonparametric frontier models. *Econometric Theory, 24*(6), 1663–1697. https://doi.org/10.1017/S0266466608080651

Kneip, A., Simar, L., & Wilson, P. W. (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. *Computational Economics, 38*(4), 483–515. https://doi.org/10.1007/S10614-010-9217-Z

Kneip, A., Simar, L., & Wilson, P. W. (2015). When Bias kills the variance: central limit theorems for Dea and FDH efficiency scores. *Econometric Theory, 31*(2), 394–422. https://doi.org/10.1017/S0266466614000413

Koopmans, T. C. (1951). Efficient allocation of resources. *Econometrica, 19*(4), 455. https://doi.org/10.2307/1907467

Kuosmanen, T., & Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research, 58*(1), 149–160. https://doi.org/10.1287/opre.1090.0722

Knox Lovell, C. A., & Pastor, J. T. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters, 18*(3), 147–151. https://doi.org/10.1016/0167-6377(95)00044-5

Luenberger, D. G. (1992). Benefit functions and duality. *J. Math. Econom., 21*(5), 461–481. https://doi.org/10.1016/0304-4068(92)90035-6

Mahlberg, B., & Sahoo, B. K. (2011). Radial and non-radial decompositions of Luenberger productivity indicator with an illustrative application. *International Journal of Production Economics, 131*(2), 721–726. https://doi.org/10.1016/J.IJPE.2011.02.021

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review, 18*(2), 435. https://doi.org/10.2307/2525757

OECD. (2014). *PISA 2012 Technical Report*.

Olesen, O. B., & Ruggiero, J. (2022). The hinging hyperplanes: An alternative nonparametric representation of a production function. *European Journal of Operational Research, 296*(1), 254–266. https://doi.org/10.1016/j.ejor.2021.03.054

Orea, L., & Zofío, J. L. (2019). Common methodological choices in nonparametric and parametric analyses of firms' performance. In *The Palgrave Handbook of Economic Performance Analysis* (pp. 419–484). Springer International Publishing. https://doi.org/10.1007/978-3-030-23727-1_12.

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (1999). An enhanced DEA Russell graph efficiency measure. *European Journal of Operational Research, 115*(3), 596–607. https://doi.org/10.1016/S0377-2217(98)00098-8

Steinwart, I. (2002). Support vector machines are universally consistent. *Journal of Complexity, 18*(3), 768–791. https://doi.org/10.1006/JCOM.2002.0642
Steinwart, I. (2005). Consistency of support vector machines and other regularized kernel classifiers. *IEEE Transactions on Information Theory, 51*(1), 128–142. https://doi.org/10.1109/TIT.2004.839514
Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research, 130*(3), 498–509. https://doi.org/10.1016/S0377-2217(99)00407-5
Tsionas, M. (2022). Efficiency estimation using probabilistic regression trees with an application to Chilean manufacturing industries. *International Journal of Production Economics, 249*, 108492. https://doi.org/10.1016/J.IJPE.2022.108492
Valero-Carreras, D., Aparicio, J., & Guerrero, N. M. (2021). Support vector frontiers: A new approach for estimating production functions through support vector machines. *Omega*, *104*, 102490. https://doi.org/10.1016/j.omega.2021.102490.
Valero-Carreras, D., Aparicio, J., & Guerrero, N. M. (2022). Multi-output Support vector frontiers. *Computers & Operations Research, 143*, 105765. https://doi.org/10.1016/J.COR.2022.105765
Vapnik, V. (1991). Principles of risk minimization for learning theory. *Advances in Neural Information Processing Systems, 4*, 831–838.
Vapnik, V. (1998). *Statistical learning theory*. Wiley.
Vazquez, E., & Walter, E. (2003). Multi-output support vector regression. *IFAC Proceedings Volumes, 36*(16), 1783–1788. https://doi.org/10.1016/S1474-6670(17)35018-8

Guerrero, N. M., Moragues, R., Aparicio, J., & Valero-Carreras, D. (2024). Support Vector Frontiers with kernel splines. Omega, 103130.

https://doi.org/10.1016/j.omega.2024.103130

Omega 128 (2024) 103130

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Support Vector Frontiers with kernel splines

Nadia M. Guerrero [a], Raul Moragues [a], Juan Aparicio [a,b,*], Daniel Valero-Carreras [a]

[a] Center of Operations Research (CIO), Miguel Hernandez University of Elche (UMH), Av. de la Universidad s/n, Elche 03202 Spain
[b] Joint Research Unit, Valencian Graduate School and Research Network of Artificial Intelligence (valgrAI), Camino de Vera s/n, Valencia 46022 Spain

ARTICLE INFO

Keywords:
Data envelopment analysis
Support vector regression
Linear splines
Convexity

ABSTRACT

Among recent methodological proposals for efficiency measurement, machine learning methods are playing an important role, particularly in the reduction of overfitting in classical statistical methods. In particular, Support Vector Frontiers (SVF) is a method which adapts Support Vector Regression (SVR) to the estimation of production technologies through stepwise frontiers. The SVF estimator is convexified in a second stage to deal with convex technologies. In this paper, we propose SVF-Splines, an extension of SVF for the estimation of efficiency in multi-input multi-output production processes which uses a transformation function generating linear splines to directly estimate convex production technologies. The proposed methodology reduces the computational complexity of the original SVF and does not require a two-step estimation process to obtain convex production technologies. A simulated experiment comparing SVF-Splines with standard DEA and (convexified) SVF indicates better performance of the proposed methodology, with improvements of up to 95 % in mean squared error when compared with DEA. The computational advantages of SVF-Splines are also observed, with runtime over 70 times faster than SVF in certain scenarios, with better scaling as the size of the problem increases. Finally, an empirical illustration is provided where SVF-Splines is calculated with respect to various typical technical efficiency measures of the literature.

1. Introduction

When faced with a group of companies or other entities which an analyst wants to evaluate and compare from a benchmarking point of view, an important line of research is the determination of the underlying production process that is behind the observed data. Many of the existing approaches in the literature can be split into two families, parametric and non-parametric methods. Among the most widely used parametric approaches, we encounter Stochastic Frontier Analysis (SFA) [1,26] while among the non-parametric perspectives, Data Envelopment Analysis (DEA) [5,8] has received enough attention to develop into its own research topic.

Among the advantages of non-parametric approaches, their flexibility, the mild conditions required for their use, and the natural way in which they deal with multi-input multi-output production processes have been pointed out [11]. In particular, DEA is characterised by its estimation of the production technology as the smallest set which satisfies envelopment of the data from above, free disposability of inputs and outputs, and convexity. The smallest set is achieved via the principle of minimal extrapolation. Within this context, various types of assumptions are possible, yielding different estimators. For example, a related estimator is Free Disposal Hull (FDH) [12], which removes the convexity postulate. This results in stepwise frontiers in FDH as opposed to the piecewise linear frontiers estimated by DEA.

As one of the most well-known non-parametric models, many properties of DEA have been considered. In particular, the postulate of minimal extrapolation has led to criticisms being leveraged that it is a conservative estimator, sometimes even labelling it as a pure descriptive approach [14]. This results in a set which fits too closely to the observed data and may not correctly estimate the underlying production process. Various authors have attempted to overcome this issue and endow DEA with inferential capabilities from the statistical point of view, such as [11], who propose a characterization of the Data Generating Process (DGP) that is behind the observations. They assume that the observations are a sample of identically and independently distributed random variables with an unknown joint distribution. The task of the estimation of the production technology can then be identified with estimating the support of the underlying DGP. They used this setting to perform

Area: Data-Driven Analytics This manuscript was processed by Associate Editor Stanko Dimitrov.
* Corresponding author at: Center of Operations Research (CIO). Miguel Hernandez University of Elche (UMH), Av. de la Universidad s/n, Elche 03202 Spain.
E-mail address: j.aparicio@umh.es (J. Aparicio).

https://doi.org/10.1016/j.omega.2024.103130
Received 26 August 2023; Accepted 4 June 2024
Available online 5 June 2024

inference tasks such as proving consistency and performing bootstrapping to estimate confidence intervals on DEA models.

Recently, another approach which has become more important makes use of some of the similarities between nonparametric methods and the machine learning literature. Among machine learning models, Support Vector Machines (SVM) [38,39] are an interesting family of machine learning algorithms, since the approach is based on solid statistical learning theory. Support Vector algorithms use the principle of Structural Risk Minimization to aim to obtain models with good generalization capabilities via bounds on estimates of the out-of-sample generalization error (prediction error) of models. Some recent contributions in this line of research ([36], 2022) proposed an adaptation of Support Vector Regression (SVR), called Support Vector Frontiers (SVF) to the estimation of stepwise production frontiers, i.e., comparable with FDH, and a Convexified SVF (CSVF), which is comparable to DEA.

Other methodology works which use machine learning principles for measuring efficiency can be seen in the Corrected Convex Nonparametric Least Squares (CCNLS) proposal by [21], while [31] proposed a smooth nonparametric kernel frontier estimator. [10] introduced an estimator based on quadratic and cubic splines with shape constraints. Decision trees have been adapted in various ways, such as [14,35]. The Structural Risk Minimization was used to construct a technology estimator by [16]. A representation of production frontiers using hinging hyperplanes was introduced by [29]. Boosting methods have been adapted by [17,18]. Additive models based on splines have been proposed by [13]. In addition to the regression-based approaches, based on supervised learning methods, a recent contribution has proposed an unsupervised learning-based generalization of DEA [27,28], among others.

One of the tools which allow SVMs to be very flexible is the use of transformation functions with associated kernels. These map the original space of predictors into a higher-dimensional space, where the classification/regression task is performed via a hyperplane which, when transformed back to the original space, can have different and flexible shapes. Usual kernels can be linear, polynomial, splines, gaussian, RBF kernels, among others. Within the SVM family, we encounter the Support Vector Regression (SVR) algorithm, which applies the SVM approach to regression problems. The flexibility of SVM using kernels allows SVR to estimate functions satisfying a variety of properties. An important family of kernels is given by the splines generating kernels, which allow the flexibility of splines interpolation to be used in conjunction with Support Vector Machines [39].

In the context of efficiency measurement, SVF [36,37] resorts to SVR that partitions the input space into a grid of cells, and associates to each cell values of 0 or 1 according to the location of data points on the grid. The use of constant values results in the use of step functions, which yields a stepwise estimation of the production frontier, in line with FDH, which is at a second stage convexified to obtain a production technology along the lines of DEA. However, this choice of transformation function causes the method to have a large computational expense. We remark that the Boolean grid of values used by SVF can be seen as a kernel generating splines of order 0, ([39], p. 464).

In this paper, we propose SVF-Splines, an extension of SVF which uses a transformation function involving splines of order 1. This results in piecewise linear estimators which can be directly compared with DEA while, at the same time, reducing the computational complexity of SVF, as we will show. We consider the restrictions which ensure that the estimator satisfies the microeconomic postulates of convexity, free disposability in inputs and outputs and data envelopment. We then compare this estimator in a computational experiment with traditional DEA and with CSVF, that is, the convexified version of SVF. We observe better results with lower computation times, particularly as the number of observations and dimensions increase. We also adapt a variety of classical measures of efficiency to the SVF-Splines estimator and illustrate with an empirical example the efficiencies obtained by DEA and the new approach.

The rest of the paper is structured as follows. Section 2 describes background concepts about Data Envelopment Analysis, Support Vector Regression and (Convex) Support Vector Frontiers. Section 3 adapts the linear splines kernel to SVF to introduce the SVF-Splines algorithm, proves that it satisfies the microeconomic postulates, characterizes the estimated technology as a DEA-type technology, and shows how to calculate a variety of measures of efficiency with the SVF-Splines estimator. Section 4 results from a computational experiment comparing the proposed SVF-Splines algorithm to DEA and Convex Support Vector Frontiers, as well as a discussion about their computational characteristics. Section 5 then illustrates the results obtained by SVF-Splines in an empirical example. Finally, Section 6 presents the conclusions obtained and outlines further possible lines of research.

2. Background

2.1. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is one of the most well-known techniques for measuring the efficiency of a set of units which use a variety of inputs to produce a variety of outputs. It is a nonparametric technique which estimates technical efficiency as the “distance” (along some permissible direction) to the efficient frontier of a production technology. The DEA production technology consists of the unique smallest set which envelops the observations, while satisfying convexity and free disposability of inputs and output [5]. In a production process with n DMUs (Decision Making Units) which use m inputs and produce s outputs, we denote inputs as $\mathbf{x} \in \mathbb{R}^m_+$ and outputs as $\mathbf{y} \in \mathbb{R}^s_+$. Let $X \in \mathbb{R}^{m\times n}_+$ ($Y \in \mathbb{R}^{s\times n}_+$) be the matrix containing all the inputs (outputs) of the DMUs in the dataset, with each DMU as a column. The DEA estimate of the technology under Variable Returns to Scale (VRS) is [5]:

$$\widehat{T}_{DEA} = \left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{m+s}_+ : \boldsymbol{x} \geq X\boldsymbol{\lambda}, \boldsymbol{y} \leq Y\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda}\mathbf{1} = 1\right\} \quad (1)$$

DEA assumes convexity of its production technology, which is a polyhedral set. When the convexity assumption is relaxed, we obtain the Free Disposal Hull (FDH) estimator, which envelops the data and satisfies free disposability of inputs and outputs and minimal extrapolation but does not satisfy convexity. The production technology estimated by FDH is stepwise, and the convexification of this technology is the technology estimated by DEA on the same data.

A region of the production technology of particular importance for the measurement of the efficiency of DMUs is the efficient frontier. There are various possible characterizations of this subset, such as the weakly efficient frontier $\delta^W(T)$ and strongly efficient frontier $\delta^S(T)$:

$$\delta^W(T) = \{(\boldsymbol{x},\boldsymbol{y}) \in T : \widehat{\boldsymbol{x}} < \boldsymbol{x}, \widehat{\boldsymbol{y}} > \boldsymbol{y} \Rightarrow (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \in T\} \quad (2)$$

$$\delta^S(T) = \{(\boldsymbol{x},\boldsymbol{y}) \in T : \widehat{\boldsymbol{x}} \leq \boldsymbol{x}, \widehat{\boldsymbol{y}} \geq \boldsymbol{y}, (\boldsymbol{x},\boldsymbol{y}) \neq (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \Rightarrow (\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \notin T\} \quad (3)$$

Elements of the strongly efficient frontier do not admit any improvement along any variable (input or output) without worsening along some other component (input or output) while remaining feasible. The weakly efficient frontier, however, consists of those elements that are not strictly Pareto dominated by any other feasible bundle, i.e., it also contains those elements which allow for improvement along one dimension while keeping the remaining variables constant. These elements do not belong to the strongly efficient frontier. Hence, the strongly efficient frontier is a subset of the weakly efficient frontier, though they do not necessarily coincide. In the case of DEA, the efficient frontiers are both piecewise linear sets. Various measures of efficiency project to either of the two efficient frontiers.

In this paper, as technical efficiency measures, we consider the radial measures, both input and output oriented, which were the first introduced in [15,5]. We also consider the Directional Distance Function [7], and the Weighted Additive Measure [9].

With respect to an arbitrary production technology T, the output-

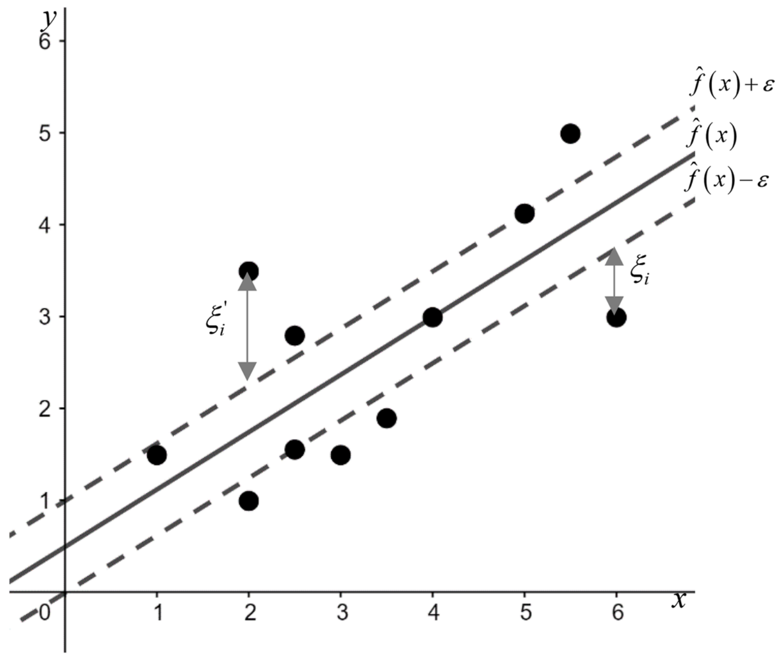

Fig. 1. Linear Support Vector Regression estimation.

oriented radial measure measures how much the outputs can be increased by the same proportion while remaining feasible. It can be calculated using the following model:

$$\psi(\boldsymbol{x},\boldsymbol{y}) = \max\{\psi \in \mathbb{R} : (\boldsymbol{x},\psi\boldsymbol{y}) \in T\} \tag{4}$$

Similarly, the input-oriented radial measure describes how much every input can be reduced by the same amount while the DMU does not become infeasible, and is calculated by solving the following model:

$$\theta(\boldsymbol{x},\boldsymbol{y}) = \min\{\theta \in \mathbb{R} : (\theta\boldsymbol{x},\boldsymbol{y}) \in T\} \tag{5}$$

The directional distance function (DDF) projects the given bundle of inputs and outputs along a pre-specified direction, given by a nonzero directional vector $\boldsymbol{g} = (\boldsymbol{g}^-,\boldsymbol{g}^+)$. It is a graph type measure, since it seeks to improve both inputs and outputs simultaneously. It was introduced by [7]:

$$\beta(\boldsymbol{x},\boldsymbol{y}) = \max\{\beta \in \mathbb{R} : (\boldsymbol{x}-\beta\boldsymbol{g}^-,\boldsymbol{y}+\beta\boldsymbol{g}^+) \in T\} \tag{6}$$

In this paper, we choose the directional vector $\boldsymbol{g} = (\boldsymbol{x},\boldsymbol{y})$ given by the values of the input-output bundle itself. This choice results in a units-invariant measure. The DDF and both radial measures project DMUs to the weakly efficient frontier, so there may be additional potential improvements (slacks) along some directions.

The Weighted Additive (WA) measure which we consider ensures that DMUs are projected to the strongly efficient frontier, as it detects slacks along any input or output. It takes as its basis a slightly different DEA model, introduced in [23]. Given input-output weights $(\boldsymbol{\rho}^-, \boldsymbol{\rho}^+) \in \mathbb{R}^{m+s}_{++}$, the Weighted Additive Model is calculated as follows:

$$WA(\boldsymbol{x},\boldsymbol{y}) = \max\left\{\boldsymbol{\rho}^-\boldsymbol{s}^- + \boldsymbol{\rho}^+\boldsymbol{s}^+ : (\boldsymbol{x}-\boldsymbol{s}^-,\boldsymbol{y}+\boldsymbol{s}^+) \in T, (\boldsymbol{s}^-,\boldsymbol{s}^+) \in \mathbb{R}^{m+s}_{+}\right\} \tag{7}$$

In particular, we use weights corresponding to the Range Adjusted Measure [9]. These weights are given by $\rho^{-(j)} = \frac{1}{(m+s)R_j^-}$ and $\rho^{+(r)} = \frac{1}{(m+s)R_r^+}$, where R_j^- is the range of input j and R_r^+ is the range of values of output $\boldsymbol{r}$. This choice results in a graph measure which is invariant to units of measurement.

The radial measures both determine as efficient those DMUs with efficiency 1. However, in the case of the output orientation, every DMU attains values larger than unity, while in the input oriented measure the efficiencies attain values between 0 and 1. Meanwhile, the DDF and WA can be considered measures of inefficiency, as efficient DMUs attain values of 0, and larger values indicate less efficient units. With the choices of weights and directional vector above, they are bounded above by 1.

2.2. Support Vector Regression and splines kernel

We now describe the Support Vector Regression (SVR) algorithm and the splines kernel that we will use. SVR is a regression algorithm that adapts the Structural Risk Minimization problem to estimate a regression on a variable while not overfitting too much to the data. Originally introduced with the Euclidean or l_2 norm, it has been extended to deal with other norms, such as the l_1 norm, which results in a linear objective function, other l_p or the l_∞ norms (see e.g. [6,32], and [41]). The base SVR model with respect to such a norm (with margin ε) is given by:

$$\begin{aligned} \underset{\boldsymbol{w},b,\xi'_i,\xi_i}{Min} \quad & \| \boldsymbol{w} \| + C\sum_{i=1}^{n} \left(\xi_i'^2 + \xi_i^2\right) \\ & y_i - (\boldsymbol{w}\cdot\boldsymbol{x}_i + b + \varepsilon) \le \xi'_i, \quad i = 1,\dots,n \\ & (\boldsymbol{w}\cdot\boldsymbol{x}_i + b - \varepsilon) - y_i \le \xi_i, \quad i = 1,\dots,n \\ & \xi'_i, \xi_i \ge 0, \qquad i = 1,\dots,n \end{aligned} \tag{8}$$

The objective function of this model consists of two parts, a regularization term $\| \boldsymbol{w} \|$ and an empirical error term $\sum_{i=1}^{n}\left(\xi_i'^2 + \xi_i^2\right)$, which are combined via a weight C. This C is a hyperparameter which, together with the margin hyperparameter ε, is obtained via a cross-validation process. The SVR model estimates a decision function given by $\widehat{f}(\boldsymbol{x}) = \boldsymbol{w}^*\cdot\boldsymbol{x} + b^*$ with errors of ξ,ξ' for the observed data outside of an ε-insensitive region. Thus, the observations within an ε margin of the estimated function $\widehat{f}(\boldsymbol{x})$ attain an error of 0, and the empirical errors are measured to this ε margin of the decision function. A graphical illustration of a function estimated by the model can be found in Fig. 1.

The SVR method can be adapted to estimate nonlinear functions via the use of a transformation function ϕ which maps the space of predictor variables into a higher-dimensional space in such a way that the obtained estimation function is linear in the transformed space but not in the original space. This is sometimes called the "kernel trick" in the

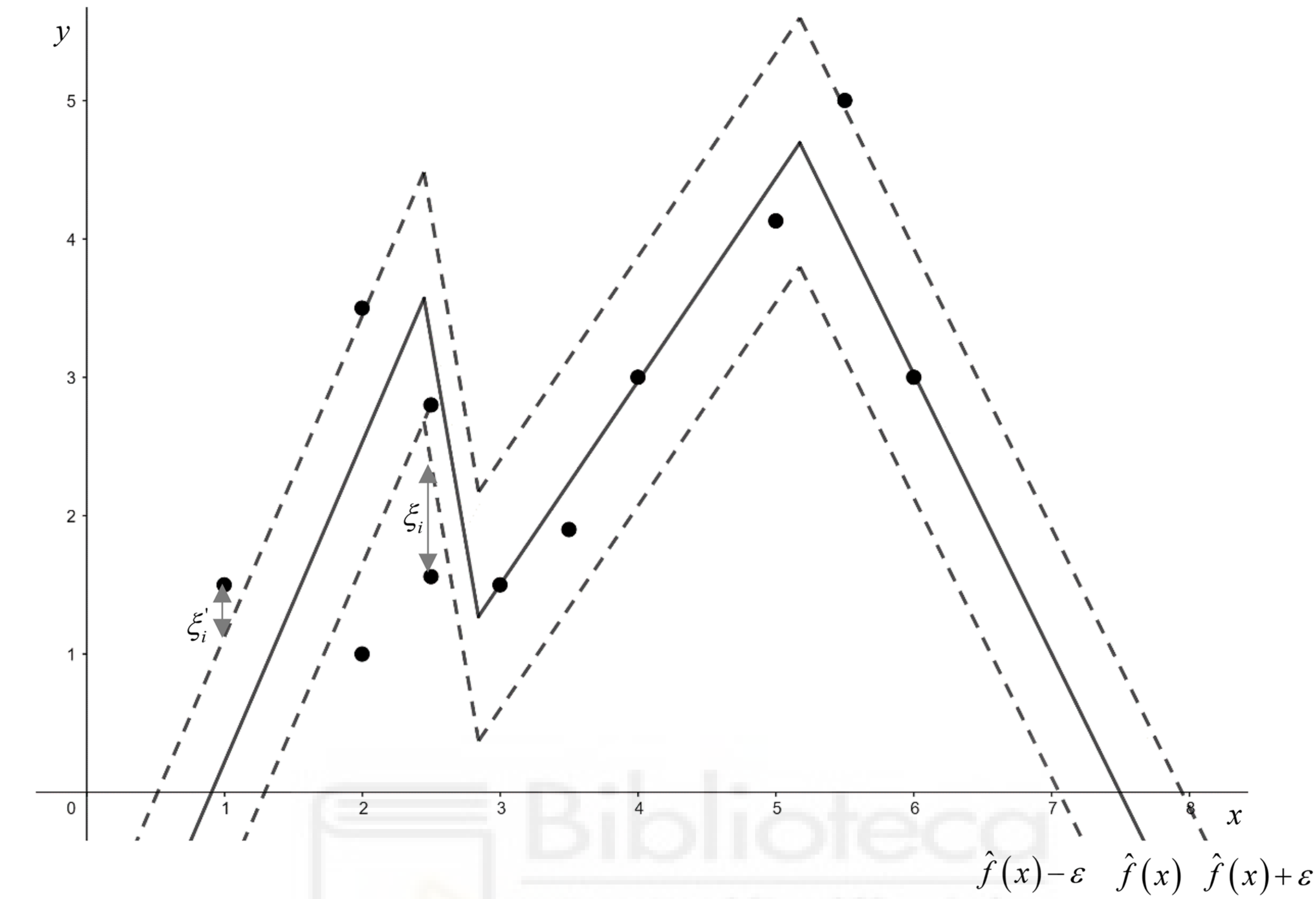

Fig. 2. Support Vector Regression with linear splines.

literature. The kernel SVR model with transformation function ϕ is the following:

$$\begin{aligned} \underset{\boldsymbol{w},b,\xi'_i,\xi_i}{Min} \quad & \| \boldsymbol{w} \| + C\sum_{i=1}^{n} \left(\xi_i'^2 + \xi_i^2\right) \\ & y_i - (\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i) + b + \varepsilon) \leq \xi'_i, \quad i = 1, \dots, n \\ & (\boldsymbol{w}\cdot\phi(\boldsymbol{x}_i) + b - \varepsilon) - y_i \leq \xi_i, \quad i = 1, \dots, n \\ & \xi'_i, \xi_i \geq 0, \quad i = 1, \dots, n \end{aligned} \tag{9}$$

When a transformation function ϕ is used, the function estimated by the model is $\widehat{f}(\boldsymbol{x}) = \boldsymbol{w}^*\cdot\phi(\boldsymbol{x}) + b^*$, and its shape depends on the characteristics of the transformation function ϕ used. There is a wide variety of possible kernels to use in this context, such as linear, polynomial, Gaussian, grid-like and splines kernels.

In this paper, we focus our attention on the splines transformation function proposed by Vapnik in ([39], p. 464). This transformation uses a finite number of knots to construct splines of order q by splitting each input dimension into a finite number of knots. Splines are flexible functions defined piecewise by polynomials of degree q, whose formulation is given by:

$$\begin{aligned} \phi(\boldsymbol{x}_i) = \Big(& 1, x_i^{(1)}, \left(x_i^{(1)}\right)^2, \dots, \left(x_i^{(1)}\right)^q, \left(x_i^{(1)} - t_1^{(1)}\right)_+^q, \dots, \left(x_i^{(1)} - t_{k_1}^{(1)}\right)_+^q, \\ & 1, x_i^{(2)}, \left(x_i^{(2)}\right)^2, \dots, \left(x_i^{(2)}\right)^q, \left(x_i^{(2)} - t_1^{(2)}\right)_+^q, \dots, \left(x_i^{(2)} - t_{k_2}^{(2)}\right)_+^q, \\ & \vdots \\ & 1, x_i^{(m)}, \left(x_i^{(m)}\right)^2, \dots, \left(x_i^{(m)}\right)^q, \left(x_i^{(m)} - t_1^{(m)}\right)_+^q, \dots, \left(x_i^{(m)} - t_{k_m}^{(m)}\right)_+^q \Big) \end{aligned} \tag{10}$$

In this transformation function, the jth component of $\boldsymbol{x}$ is transformed into a $(1 + q + k_j)$-dimensional vector, where k_j is the number of knots along dimension j. The first component of such vector is a constant value of 1, the next q components are powers of the original component of $\boldsymbol{x}$, and the final k_j elements are defined by:

$$\left(x_i^{(j)} - t_{l_j}^{(j)}\right)_+^q = \begin{cases} \left(x_i^{(j)} - t_{l_j}^{(j)}\right)^q & if \quad x_i^{(j)} > t_{l_j}^{(j)} \\ 0 & if \quad x_i^{(j)} \leq t_{l_j}^{(j)} \end{cases}, l_j = 1, \dots, k_j, j = 1, \dots, m \tag{11}$$

In particular, splines of order $q = 0$ yield step functions as estimators, while splines of order $q = 1$, also called linear splines, produce piecewise linear estimators, see Fig. 2 for an example. Higher values of q provide piecewise approximations using polynomials of degree q. In this paper, we focus on splines of order $q = 1$, inspired by the nature of DEA, which estimates a piecewise linear production function.

2.3. Support Vector Frontiers

A recent contribution [36] shows how to adapt Support Vector Regression to the estimation of production functions in the single-output case. The authors adapt the SVR model to satisfy the properties of production functions of envelopment from above and monotonicity and develop Support Vector Frontiers (SVF). They propose a transformation function $\phi^G_{SVF}(\boldsymbol{x})$ of the space of inputs which consists of binary values on a grid with its components taking values of 0 or 1 whenever they are dominated by $\boldsymbol{x}$. As a consequence, SVF estimates step production functions, in other words, comparable with the Free Disposal Hull estimator. The authors also show how to convexify this technology in order to estimate DEA-like production functions, via Convexified Support Vector Frontiers (CSVF). Additionally, SVF has later been extended to the multi-output context in [37]. The multi-output SVF model is given by:

$$
\begin{aligned}
\min_{w_r,\xi_{ri}} \quad & \sum_{r=1}^{s}\left\|\boldsymbol{w}^{(r)}\right\| + C\sum_{r=1}^{s}\sum_{i=1}^{n}\xi_i^{(r)} \\
\text{s.t.} \quad & \boldsymbol{w}^{(r)}\cdot\phi^G_{SVF}(\boldsymbol{x}_i) - y_i^{(r)} \geq 0, && i=1,\dots,n && r=1,\dots,s \\
& \boldsymbol{w}^{(r)}\cdot\phi^G_{SVF}(\boldsymbol{x}_i) - y_i^{(r)} \leq \varepsilon + \xi_i^{(r)}, && i=1,\dots,n && r=1,\dots,s \\
& W^{(r)}_{l_1\dots s_j\dots l_m} \leq W^{(r)}_{l_1\dots l_m}, && l_1=1,\dots,k_1,\dots,l_m=1,\dots,k_m && s_j=l_j-1 \\
& && l_j=1,\dots,k_j,\ j=1,\dots,m && r=1,\dots,s \\
& \xi_i^{(r)} \geq 0, && i=1,\dots,n
\end{aligned}
\tag{12}
$$

We remark that this model, when $s=1$, is equivalent to the single-output model introduced in [36]. It involves a transformation function $\phi^G_{SVF}(\boldsymbol{x})$ associated to a grid G common to all outputs, which partitions the input space into a grid of binary values, as well as a different set of weights $\boldsymbol{w}^{(r)}$for each output dimension. Each vector of weights has one component for each cell of the grid G. In (12), $\phi^G_{SVF}(\boldsymbol{x})$ can be considered a splines transformation of order 0. It results in an estimation for each output $\widehat{y}^{(r)}_{SVF}(\boldsymbol{x}_i) := \boldsymbol{w}^{(r)}\cdot\phi^G_{SVF}(\boldsymbol{x}_i)$, $r=1,\dots,s$, which is a stepwise production frontier in line with Free Disposal Hull.

The SVF model has an empirical error $\xi_i^{(r)}$ associated with each $y_i^{(r)}$, the output dimension $r$of DMU i, and model (12) minimizes the weighted sum $\xi_i^{(r)}$ of the regularization term $\sum_{r=1}^{s}\|\boldsymbol{w}^{(r)}\|$ and the empirical error $\sum_{r=1}^{s}\sum_{i=1}^{n}\xi_i^{(r)}$, weighted by a parameter C. The first constraint in (12) ensures that the estimator envelops the observed data from above, while the second constraint will penalize the error committed beyond a margin parameter ε via the empirical error $\xi_i^{(r)}$, which will be nonnegative by the final restriction. Finally, the third restriction, which involves variables $W^{(r)}_{l_1\dots l_m}$ defined by $W^{(r)}_{l_1\dots l_m} := \sum_{\substack{s_1=1,\dots,l_1\\ \vdots \\ s_m=1,\dots,l_m}} w^{(r)}_{s_1\dots s_m}$, for each output dimension $r=1,\dots,s$, will ensure that the estimated output $\widehat{y}^{(r)}_{SVF}(\boldsymbol{x})$ is a monotonic non-decreasing function. Once model (12) is solved, optimal values $\boldsymbol{w}^{*(r)}$ are obtained, and the estimated output is given by $\widehat{y}^{(r)}_{SVF}(\boldsymbol{x}) = \boldsymbol{w}^{*(r)}\cdot\phi^G_{SVF}(\boldsymbol{x})$ for each $r=1,\dots,s$. The associated production technology is given by:

$$\widehat{T}_{SVF} := \left\{(\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}^{m+s}_{+} : \boldsymbol{y}\leq\widehat{\boldsymbol{y}}_{SVF}(\boldsymbol{x})\right\} \tag{13}$$

By ([37], Section 3.1), this technology satisfies envelopment of the data, free disposability, and it is an FDH-style production technology. In particular, it can be characterised as an FDH-type estimator on a set of "virtual points" defined by the grid involved in the transformation function $\phi^G_{SVF}(\boldsymbol{x})$. These points are virtual in the sense that they are not necessarily in the original data. Instead, they are defined using the predictions of the model on extreme points of the grid cells. In a second stage, this production technology is convexified to obtain a DEA-style technology with respect to these virtual points, yielding the Convexified SVF (CSVF) estimator. This results in a two-stage estimation of convex production technologies. In particular, to simplify computation, each split input dimension is split in the same number of nodes, which leads to a single hyperparameter d to be tuned together with $C$and ε via a five-fold cross-validation procedure.

The margin hyperparameter ε can be used to define a more robust notion of technical efficiency, which the authors name ε-insensitive technical efficiency. This notion considers any DMU within a margin ε of the technical efficient frontier to be ε-insensitive technically efficient, and allows for the identification of those units which are not considered ε-insensitive technically efficient as units which are far from being technically efficient.

As recognized by its authors, the SVF estimator has some limitations, particularly of computational complexity. The proposed model involves a large amount of variables and restrictions, given that there is one w associated to each cell of the grid, which results in an exponential ($\approx md^m$) number of restrictions involving the W.

In this paper, we propose a way to overcome these limitations by using a model based on linear splines in order to reduce the computational complexity and directly estimate convex production technologies in a single stage.

3. Support Vector Frontiers with kernel splines

3.1. The single-output case

In this section, we show how to adapt a linear splines kernel (with $q=1$) to measure efficiency via the modification of the Support Vector Frontiers estimation with this transformation function. This adaptation reduces the exponential complexity of the restrictions of model (12) to a set of a linear number of restrictions via the choice of nodes along each of the input components. We begin by presenting the single-output model and proving its properties, before generalizing it to the multi-output case:

$$
\begin{aligned}
&\underset{\boldsymbol{w},\xi_i}{Min} \quad \|\boldsymbol{w}\| + C\sum_{i=1}^{n}\xi_i \\
&s.t. \\
&\boldsymbol{w}\cdot\phi^G_{SVF-SP}(\boldsymbol{x}_i) - y_i \le \varepsilon + \xi_i \quad i=1,\dots,n \qquad (14.1)\\
&y_i - \boldsymbol{w}\cdot\phi^G_{SVF-SP}(\boldsymbol{x}_i) \le 0 \quad i=1,\dots,n \qquad (14.2)\\
&W^{(j)}_{l_j} \ge 0 \quad l_j=0,\dots,k_j \quad j=1,\dots,m \qquad (14.3)\\
&w^{(j)}_0, w^{(j)}_{-1} \ge 0 \quad j=1,\dots,m \qquad (14.4)\\
&w^{(j)}_k \le 0 \quad k=1,\dots,k_j \quad j=1,\dots,m \qquad (14.5)\\
&\xi_i \ge 0 \quad i=1,\dots,n \qquad (14.6)
\end{aligned} \tag{14}
$$

where

$$\phi^G_{SVF-SP}(\boldsymbol{x}_i) = \left(1, x_i^{(1)}, \left(x_i^{(1)} - t_1^{(1)}\right)_+, \dots, \left(x_i^{(1)} - t_{k_1}^{(1)}\right)_+, \quad \dots, \quad 1, x_i^{(m)}, \left(x_i^{(m)} - t_1^{(m)}\right)_+, \dots, \left(x_i^{(m)} - t_{k_m}^{(m)}\right)_+\right) \tag{15}$$

with

$$\left(x_i^{(j)} - t_{l_j}^{(j)}\right)_+ = \begin{cases} x_i^{(j)} - t_{l_j}^{(j)} & if \quad x_i^{(j)} > t_{l_j}^{(j)} \\ 0 & if \quad x_i^{(j)} \le t_{l_j}^{(j)} \end{cases}, l_j = 1,\dots,k_j, j=1,\dots,m \tag{16}$$

Thus, $\phi^G_{SVF-SP}(\boldsymbol{x}_i)$ is a transformation from $\mathbb{R}^m \rightarrow \mathbb{R}^{2m+\sum_{j=1}^m k_j}$ (i.e., a $\sum_{j=1}^m (k_j + 2)$-dimensional space). We denote this dimension by h. The corresponding weights vector is:

$$\boldsymbol{w} = \left(w_{-1}^{(1)}, w_0^{(1)}, w_1^{(1)}, \dots, w_{k_1}^{(1)}, \quad \dots \quad, w_{-1}^{(m)}, w_0^{(m)}, w_1^{(m)}, \dots, w_{k_m}^{(m)}\right) \tag{17}$$

We remark that weights $w_1^{(j)}, \dots, w_{k_j}^{(j)}$ correspond to the nodes $t_1^{(j)}, \dots, t_{k_j}^{(j)}$ in (15). Hence, we denote the first two components along each input dimension j, which do not correspond to any such nodes, by $w_{-1}^{(j)}$ and $w_0^{(j)}$. For ease of notation, we combine those weights which get used in each interval between two consecutive nodes of the splines transformation:

$$W_{l_j}^{(j)} = \sum_{k=0}^{l_j} w_k^{(j)} \quad l_j = 0,\dots,k_j, \quad j=1,\dots,m \tag{18}$$

This transformation function and associated kernel involves various parameters, which must be estimated or chosen appropriately. Each input dimension j is split into a number k_j of nodes, and defines $k_j + 2$ components of both $\boldsymbol{w}$ and ϕ. We choose to divide each input dimension into a number d of nodes between the minimum and maximum values observed along each dimension of the same width $(\max - \min)/d$. This defines $d+1$ nodes for each input, so that the dimension of the transformation function, as well as that of $\boldsymbol{w}$, is $h = m(d+3)$. This value d is a hyperparameter of the algorithm which will be estimated via a five-fold cross-validation process together with C and ε. The input space is thus divided into $(d+2)^m$ grid cells.

Solving problem (14) yields optimal values $\boldsymbol{w}^*$ and $\boldsymbol{\xi}^*$ and the corresponding piecewise linear production function is given by $\hat{f}(\boldsymbol{x}) = \hat{y}_{SVF-SP}(\boldsymbol{x}) = \boldsymbol{w}^*\cdot\phi^G_{SVF-SP}(\boldsymbol{x})$. We will denote the corresponding $W_{l_j}^{(j)}$ at optimum by ${W_{l_j}^{(j)}}^*$ as well. We now prove that $\hat{f}(\boldsymbol{x})$ satisfies data envelopment, monotonicity and concavity. Note that the properties of free disposability and convexity for the multi-input multi-output framework are translated into (non-decreasing) monotonicity and concavity of the corresponding production function for the multi-input single-output case, respectively.

Proposition 3.1. *For each $i = 1,\dots,n$, $\hat{f}(\boldsymbol{x})$ satisfies envelopment of the data. That is, for each $i = 1,\dots,n$, we have $y_i \le \hat{f}(\boldsymbol{x}_i)$.*

Proof. Holds by constraint (14.2) and definition of $\hat{f}(\boldsymbol{x})$, which forces $y_i - \boldsymbol{w}^*\cdot\phi^G_{SVF-SP}(\boldsymbol{x}_i) = y_i - \hat{f}(\boldsymbol{x}_i) \le 0$ for each i.■

We now prove that $\hat{f}(\boldsymbol{x})$ is monotonic non-decreasing.

Proposition 3.2. *If $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{z} \ge \mathbf{x}$, then $\hat{f}(\mathbf{z}) \ge \hat{f}(\mathbf{x})$.*

Proof. Assume that $\mathbf{z} \ge \mathbf{x}$. We construct a series of inequalities where at each step only one component changes. That is,

$$\mathbf{x} = \boldsymbol{\alpha}_0 = (x(1),\dots,x(m)) \le \boldsymbol{\alpha}_1 = (z(1),\dots,x(m)) \le \dots \le \boldsymbol{\alpha}_m = (z(1),\dots,z(m)) = \mathbf{z}.$$

We will prove the inequality between $\boldsymbol{\alpha}_{j-1}$ and $\boldsymbol{\alpha}_j$ for each $j = 1,\dots,m$. Hence, we consider $\hat{f}(\boldsymbol{\alpha}_j) - \hat{f}(\boldsymbol{\alpha}_{j-1}) = \boldsymbol{w}^*\cdot\phi^G_{SVF-SP}(\boldsymbol{\alpha}_j) - \boldsymbol{w}^*\cdot\phi^G_{SVF-SP}(\boldsymbol{\alpha}_{j-1}) = \boldsymbol{w}^*\cdot(\phi^G_{SVF-SP}(\boldsymbol{\alpha}_j) - \phi^G_{SVF-SP}(\boldsymbol{\alpha}_{j-1}))$. Since the only component that changes between $\boldsymbol{\alpha}_{j-1}$ and $\boldsymbol{\alpha}_j$ is the jth component and $\boldsymbol{w}^*$ is fixed, the only components of ϕ^G_{SVF-SP} that change are those involving the jth component. Furthermore, $\beta^{(j)} = \alpha_j^{(j)} - \alpha_{j-1}^{(j)} \ge 0$. The other terms cancel out, and we have:

$$\hat{f}(\boldsymbol{\alpha}_j) - \hat{f}(\boldsymbol{\alpha}_{j-1}) = w_{-1}^{(j)*}(1-1) + w_0^{(j)*}\beta^{(j)} + \sum_{k=1}^{k_j} w_k^{(j)*}\left(\left(\alpha_j^{(j)} - t_k^{(j)}\right)_+ - \left(\alpha_{j-1}^{(j)} - t_k^{(j)}\right)_+\right).$$

For each k, the terms involving $t_k^{(j)}$ within the summatory can either both be active, i.e., $\alpha_j^{(j)} \ge \alpha_{j-1}^{(j)} \ge t_k^{(j)}$, both inactive, that is, $t_k^{(j)} \ge \alpha_j^{(j)} \ge \alpha_{j-1}^{(j)}$, or only the $\alpha_j^{(j)}$ active, which happens when $\alpha_j^{(j)} \ge t_k^{(j)} \ge \alpha_{j-1}^{(j)}$. The term $\left(\alpha_j^{(j)} - t_k^{(j)}\right)_+ - \left(\alpha_{j-1}^{(j)} - t_k^{(j)}\right)_+$ is then $\left(\alpha_j^{(j)} - t_k^{(j)}\right)_+ - \left(\alpha_{j-1}^{(j)} - t_k^{(j)}\right)_+ = \alpha_j^{(j)} - \alpha_{j-1}^{(j)} = \beta^{(j)}$ (both active), $0 - 0 = 0 \le \beta^{(j)}$ if both are inactive, or $\left(\alpha_j^{(j)} - t_k^{(j)}\right) - 0 = \alpha_j^{(j)} - t_k^{(j)} \le \alpha_j^{(j)} - \alpha_{j-1}^{(j)} = \beta^{(j)}$. Thus, in any case, these terms are bounded above by $\beta^{(j)}$. Then, each such term is multiplied by $w_k^{(j)*}$, which is non-positive by restriction (14.5). Thus, we have $\hat{f}(\boldsymbol{\alpha}_j) - \hat{f}(\boldsymbol{\alpha}_{j-1}) \ge w_0^{(j)*}\beta^{(j)} + \sum_{k=1}^{k_j} w_k^{(j)*}\beta^{(j)} = W_{k_j}^{(j)*}\beta^{(j)} \ge 0$.

We remark that this final inequality holds as $W_{k_j}^{(j)*} \ge 0$, by constraint (14.3). Therefore, we have $\hat{f}(\boldsymbol{\alpha}_j) \ge \hat{f}(\boldsymbol{\alpha}_{j-1})$ for each j. By applying this argument repeatedly, we obtain that $\hat{f}(\mathbf{z}) - \hat{f}(\mathbf{x})$, as claimed. Thus, $\hat{f}(\mathbf{x})$ is monotonic non-decreasing. ■

Next, to prove concavity of the estimated production function, we consider the nature of the components of the transformations, and the signs of the components of $\mathbf{w}^*$. Intuitively, as each of the individual w (except the first ones) are negative, so the slope of the estimated frontier only decreases as more terms activate, resulting in a concave production function.

Proposition 3.3. *The estimated production function $\hat{f}(\mathbf{x})$ is concave.*

Proof. We consider each component of $\phi^G_{SVF-SP}(\boldsymbol{x})$. The first two terms along each input dimension are linear, and their corresponding weights in $\mathbf{w}^*$, that is, w_{-1}^* and w_0^*, are non-negative by constraint (14.4). Thus, the corresponding summands in the expression for $\hat{f}(\mathbf{x})$ are linear, hence both convex and concave. The remaining terms are defined as the maximum between two linear functions, the constant 0 and $x^{(j)} - t^{(j)}_{l_j}$. Since both terms are linear, they are convex, i.e., the area above their respective curves is convex. The area of the maximum between two functions corresponds to the intersection of both regions, so each term in $\phi^G_{SVF-SP}(\boldsymbol{x})$ involving a maximum is convex. These terms are then multiplied by the corresponding weights $w^{(j)*}_{l_j}$, which are non-positive by restriction (14.5), so the corresponding products are concave functions. Thus, the expression for $\hat{f}(\mathbf{x})$ is a sum of concave functions, and is thus concave. ■

Having established these properties of the estimated production function, we now move on to the multi-output case, and consider the corresponding production technology.

3.2. The multi-output case

We now present the multioutput SVF-Splines model, which extends the above single-output model to the multi-output case following the approach by [40]. The idea is that a multi-output model can be obtained by using a transformation of the outputs and estimating weights for each of the outputs separately, so that with the same grid in the input space (which a common structure of knots for all the outputs), different estimations are obtained for each of the multiple outputs. The model is as follows:

$$
\begin{aligned}
&\underset{\boldsymbol{w}_r,\xi_{ri}}{Min} \quad \sum_{r=1}^{s} \| \boldsymbol{w}^{(r)} \| + C\sum_{r=1}^{s}\sum_{i=1}^{n}\xi_i^{(r)} \\
&s.t. \\
&\boldsymbol{w}^{(r)}\cdot\phi^G_{SVF-SP}(\boldsymbol{x}_i) - y_i^{(r)} \le \varepsilon + \xi_i^{(r)} \quad i=1,\ldots,n \quad r=1,\ldots,s \quad (19.1)\\
&y_i^{(r)} - \boldsymbol{w}^{(r)}\cdot\phi^G_{SVF-SP}(\boldsymbol{x}_i) \le 0 \quad i=1,\ldots,n \quad r=1,\ldots,s \quad (19.2)\\
&W^{(j)^{(r)}}_{l_j} \ge 0 \quad l_j=0,\ldots,k_j \quad j=1,\ldots,m \quad r=1,\ldots,s \quad (19.3)\\
&w^{(j)^{(r)}}_0, w^{(j)^{(r)}}_{-1} \ge 0 \quad j=1,\ldots,m \quad r=1,\ldots,s \quad (19.4)\\
&w^{(j)^{(r)}}_k \le 0 \quad k=1,\ldots,k_j \quad j=1,\ldots,m \quad r=1,\ldots,s \quad (19.5)\\
&\xi_i^{(r)} \ge 0 \quad i=1,\ldots,n \quad r=1,\ldots,s \quad (19.6)
\end{aligned}
\tag{19}
$$

Here, $\phi^G_{SVF-SP}(\boldsymbol{x})$ is defined as in the single-output model by (15), whereas there is a $\mathbf{w}^{(r)}$ associated with each output. Correspondingly, the $W$are defined by:

$$W^{(j)^{(r)}}_{l_j} = \sum_{k=0}^{l_j} w^{(j)(r)}_k, \quad l_j = 0,\ldots,k_j, \quad j=1,\ldots,m, \quad r=1,\ldots,s. \tag{20}$$

In this setting, once model (19) is solved yielding optimal values $\mathbf{w}^*$ and $\boldsymbol{\xi}^*$, we define the rth component of the output vector corresponding to input profile$\boldsymbol{x}$ by:

$$\hat{f}^{(r)}(\boldsymbol{x}) = \hat{y}^{(r)}{}_{SVF-SP}(\boldsymbol{x}) = \boldsymbol{w}^{(r)*}\cdot\phi^G_{SVF-SP}(\boldsymbol{x}). \tag{21}$$

This defines the multi-output production frontier $\hat{\mathbf{f}}(\mathbf{x})$. The corresponding production possibility set or technology is defined as the set of those collections of inputs and outputs whose outputs lie below the production frontier:

$$\hat{T}_{SVF-SP} := \left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{m+s}_+ : \boldsymbol{y} \le \hat{\boldsymbol{f}}(\boldsymbol{x})\right\} \tag{22}$$

We now prove that $\hat{T}_{SVF-SP}$ satisfies data envelopment, free disposability in inputs and outputs and convexity. With these definitions, we can extend the results above to the multi-output estimator. The proofs are very similar to the ones in [37] Lemma 1, Propositions 1, 2, 3, 4. As in the single-output model, the hyperparameters C, ε and d are estimated by five-fold cross-validation.

Proposition 3.4. *For all $i = 1,\ldots,n$, we have $(\mathbf{x}_i, \mathbf{y}_i) \in \hat{T}_{SVF-SP}$.*

Proof. Follows by constraint (19.2) and the definition of the technology. ■

Proposition 3.5. *$\hat{T}_{SVF-SP}$ satisfies free disposability in inputs and outputs.*

Proof. Let $(\mathbf{x}, \mathbf{y}_x) \in \hat{T}_{SVF-SP}$ and $(\mathbf{z}, \mathbf{y}_z)$ satisfy $\mathbf{z} \ge \mathbf{x}$ and $\mathbf{y}_z \le \mathbf{y}_x$. Then, by Proposition 3.2 we have $\hat{f}^{(r)}(\mathbf{z}) \ge \hat{f}^{(r)}(\mathbf{x})$ for each component of $\hat{\mathbf{f}}(\mathbf{x})$. Since $\mathbf{y}_z \le \mathbf{y}_x$, we have that $\mathbf{y}_z \le \mathbf{y}_x \le \hat{\mathbf{f}}(\mathbf{x}) \le \hat{\mathbf{f}}(\mathbf{z})$, so that $(\mathbf{z}, \mathbf{y}_z) \in \hat{T}_{SVF-SP}$ and $\hat{T}_{SVF-SP}$ satisfies free disposability. ■

Proposition 3.6. *$\hat{T}_{SVF-SP}$ is convex.*

Proof. By Proposition 3.3, each component of $\hat{\mathbf{f}}(\mathbf{x})$ is concave. Thus, so is $\hat{\mathbf{f}}(\mathbf{x})$. Now, ([24], p. 81) shows that a function $\hat{\mathbf{f}}(\mathbf{x})$ is concave if and only if its hypograph is a convex set, where the hypograph is $HG_{\hat{\mathbf{f}}} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{m+s} : \mathbf{y} \le \hat{\mathbf{f}}(\mathbf{x})\}$. In this context, the production technology $\hat{T}_{SVF-SP} = \left\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{m+s}_+ : \boldsymbol{y} \le \hat{\boldsymbol{f}}(\boldsymbol{x})\right\} = HG_{\hat{\boldsymbol{f}}} \cap \mathbb{R}^{m+s}_+$ is the intersection of the hypograph and the non-negative quadrant of $\mathbb{R}^{m+s}$, and both these sets are convex. Therefore, since the intersection of convex sets is a convex set, the production technology $\hat{T}_{SVF-SP}$ is a convex set.

Proposition 3.7. *If $s = 1$, then the multi-output model coincides with the single-output model.*

Proof. Clear from the formulation of the models. ■

Corollary. $\hat{T}_{DEA} \subseteq \hat{T}_{SVF-SP}$.

Proof. The set $\hat{T}_{DEA}$ is, by the principle of minimal extrapolation, the intersection of all sets satisfying envelopment, free disposability and convexity. By Propositions 3.4, 3.5 and 3.6, $\hat{T}_{SVF-SP}$ satisfies envelopment, free disposability and convexity. Thus, $\hat{T}_{DEA}$ is a subset of $\hat{T}_{SVF-SP}$. ■

3.3. Characterization of the estimated technology as a DEA-type technology

We now proceed to characterise the estimated technology as a DEA-type technology with respect to a set of “virtual points”, which are not observed in the data but rather constructed from the SVF-Splines estimations and the extreme points of the grid cells involved in the splines transformation function. In other words, we prove that the estimated technology consists of the smallest convex set enveloping these virtual points which satisfies free disposability, i.e., a DEA-type technology with these virtual points as observations.

A grid consists of cells $C_{l_1 l_2 \ldots l_m}$, where $l_j = 0, \ldots, k_j$ for each input $j = 1, \ldots, m$, with corresponding lower extreme knot-point $\boldsymbol{a}_{l_1 l_2 \ldots l_m} = \left(t_{l_1}^{(1)}, t_{l_2}^{(2)}, \ldots, t_{l_m}^{(m)}\right)$. Each grid cell has 2^m extreme points and is the convex closure of its extreme points. The output values estimated by SVF-Splines at each extreme point of a grid cell is given by $\widehat{\boldsymbol{f}}\left(\boldsymbol{a}_{l_1 l_2 \ldots l_m}\right) = \widehat{\boldsymbol{y}}_{SVF-SP}\left(\boldsymbol{a}_{l_1 l_2 \ldots l_m}\right)$, and the set of all pairs $\left(\boldsymbol{a}_{l_1 l_2 \ldots l_m}, \widehat{\boldsymbol{f}}\left(\boldsymbol{a}_{l_1 l_2 \ldots l_m}\right)\right)$ forms the virtual data of $\widehat{T}_{SVF-SP}$. In other words, if we let A be the matrix containing the inputs $\boldsymbol{a}_{l_1 l_2 \ldots l_m}$ as columns and $\widehat{\boldsymbol{f}}(A)$ be the matrix of corresponding estimated outputs, we have:

$$\widehat{T}_{SVF-SP-DEA} = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} : \boldsymbol{x} \geq A\lambda, \boldsymbol{y} \leq \widehat{\boldsymbol{f}}(A)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1\right\} \tag{23}$$

By the properties of DEA, this set is the smallest set satisfying envelopment of the "virtual points" $(\boldsymbol{a}_{l_1 l_2 \ldots l_m}, \widehat{\boldsymbol{f}}(\boldsymbol{a}_{l_1 l_2 \ldots l_m}))$ determined by SVF-Splines, free disposability in inputs and outputs, and convexity. We now prove the following equality:

Proposition 4. $\widehat{T}_{SVF-SP-DEA} = \widehat{T}_{SVF-SP}$. *In other words,* $\widehat{T}_{SVF-SP}$ *is a DEA-type production technology with respect to the virtual points* $\left(\boldsymbol{a}_{l_1 l_2 \ldots l_m}, \widehat{\boldsymbol{f}}\left(\boldsymbol{a}_{l_1 l_2 \ldots l_m}\right)\right)$ *for each* $l_1 = 0, \ldots, k_1; \ldots; l_m = 0, \ldots, k_m$.

Proof. Recall the definition of the technology estimated by SVF-Splines (22):

$$\widehat{T}_{SVF-SP} := \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{m+s} : \boldsymbol{y} \leq \widehat{\boldsymbol{f}}(\boldsymbol{x})\right\}.$$

We now prove that these two characterizations (22) and (23) of the production technology coincide. By Propositions 3.5 and 3.6, we have that $\widehat{T}_{SVF-SP}$ is a production possibility set satisfying free disposability of inputs and outputs and convexity. Regarding envelopment of the virtual points, we have proved envelopment of the original data in Proposition 3.4, but now the data defining the technology are not the original data but instead the virtual data given by $\widehat{\boldsymbol{f}}(\boldsymbol{x})$. These points are also contained in $\widehat{T}_{SVF-SP}$ by the definition of the technology, thus $\widehat{T}_{SVF-SP}$ also envelops the virtual data $(\boldsymbol{a}_{l_1 l_2 \ldots l_m}, \widehat{\boldsymbol{f}}(\boldsymbol{a}_{l_1 l_2 \ldots l_m}))$ for each extreme of the grid. Therefore, by the principle of minimal extrapolation, we have $\widehat{T}_{SVF-SP-DEA} \subseteq \widehat{T}_{SVF-SP}$.

For the reverse inclusion, we consider an arbitrary $(\boldsymbol{x}, \boldsymbol{y}) \in \widehat{T}_{SVF-SP}$, and we will prove that $(\boldsymbol{x}, \boldsymbol{y}) \in \widehat{T}_{SVF-SP-DEA}$. The input profile $\boldsymbol{x}$ belongs to a cell of the grid, which we denote by C, and is defined by a set of extreme points $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{2^m}$. In particular, C is the convex closure of these points. As $\boldsymbol{x} \in C$, it can be written as a convex linear combination of its corner points, that is, there exists $\lambda \geq \boldsymbol{0}$ with $\sum_{v=1}^{2^m} \lambda_v = 1$ such that $\boldsymbol{x} = \sum_{v=1}^{2^m} \boldsymbol{\alpha}_v \lambda_v$ We now show that $\boldsymbol{y} \leq \widehat{\boldsymbol{f}}(A)\lambda$. Now, $\boldsymbol{f}(\boldsymbol{x})$ is a piecewise linear function, which is linear within the confines of each cell of the grid. Thus, since $\boldsymbol{x} \in C$, we have $\widehat{\boldsymbol{f}}(\boldsymbol{x}) = \widehat{\boldsymbol{f}}\left(\sum_{v=1}^{2^m} \boldsymbol{\alpha}_v \lambda_v\right) = \sum_{v=1}^{2^m} \widehat{\boldsymbol{f}}(\boldsymbol{\alpha}_v)\lambda_v = \widehat{\boldsymbol{f}}(A)\lambda$. As, by assumption, $(\boldsymbol{x}, \boldsymbol{y}) \in \widehat{T}_{SVF-SP}$, we have that $\boldsymbol{y} \leq \widehat{\boldsymbol{y}}_{SVF-SP}(\boldsymbol{x}) = \widehat{\boldsymbol{f}}(\boldsymbol{x}) = \widehat{\boldsymbol{f}}(A)\lambda$. Thus, $\widehat{T}_{SVF-SP} \subseteq \widehat{T}_{SVF-SP-DEA}$, and equality follows. ■

3.4. Measures of efficiency using SVF-Splines

We will use this characterization of the estimated technology $\widehat{T}_{SVF-SP}$ as a DEA-like technology with respect to a set of virtual points to adapt some of the measures of efficiency available in the literature to this context. We adapt the radial measures, both input and output oriented, as well as the models defining the Directional Distance Function and the Weighted Additive Measure. With respect to the virtual points $(\boldsymbol{a}_{l_1 l_2 \ldots l_m}, \widehat{\boldsymbol{f}}(\boldsymbol{a}_{l_1 l_2 \ldots l_m}))$ defined above, the output-oriented radial measure can be calculated using the following model [5]:

$$\psi(\boldsymbol{x}, \boldsymbol{y}) = \max\{\psi \in \mathbb{R} : (\boldsymbol{x}, \psi\boldsymbol{y}) \in \widehat{T}_{SVF-SP}\} = \max\{\psi : \boldsymbol{x} \geq A\lambda, \psi\boldsymbol{y} \leq \widehat{\boldsymbol{f}}(A)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1\} \tag{24}$$

Similarly, the input-oriented radial measure is calculated by solving model [5]:

$$\theta(\boldsymbol{x}, \boldsymbol{y}) = \min\{\theta \in \mathbb{R} : (\theta\boldsymbol{x}, \boldsymbol{y}) \in \widehat{T}_{SVF-SP}\} = \min\{\theta : \theta\boldsymbol{x} \geq A\lambda, \boldsymbol{y} \leq \widehat{\boldsymbol{f}}(A)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1\} \tag{25}$$

The directional distance function (DDF), with directional vector $\mathbf{g} = (\mathbf{g}^-, \mathbf{g}^+)$ [7] is the solution to the following model:

$$\begin{aligned} \beta(\boldsymbol{x}, \boldsymbol{y}) &= \max\{\beta \in \mathbb{R} : (\boldsymbol{x} - \beta\mathbf{g}^-, \boldsymbol{y} + \beta\mathbf{g}^+) \in \widehat{T}_{SVF-SP}\} \\ &= \max\{\beta : \boldsymbol{x} - \beta\mathbf{g}^- \geq A\lambda, \boldsymbol{y} + \beta\mathbf{g}^+ \leq \widehat{\boldsymbol{f}}(A)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1\} \end{aligned} \tag{26}$$

The weighted additive (WA) model [23] is calculated in the case of resorting to the new approach as:

$$\begin{aligned} WA(\boldsymbol{x}, \boldsymbol{y}) &= \max\{\boldsymbol{\rho}^-\boldsymbol{s}^- + \boldsymbol{\rho}^+\boldsymbol{s}^+ \in \mathbb{R} : (\boldsymbol{x} - \boldsymbol{s}^-, \boldsymbol{y} + \boldsymbol{s}^+) \in \widehat{T}_{SVF-SP}, (\boldsymbol{s}^-, \boldsymbol{s}^+) \in \mathbb{R}_+^{m+s}\} \\ &= \max\{\boldsymbol{\rho}^-\boldsymbol{s}^- + \boldsymbol{\rho}^+\boldsymbol{s}^+ : \boldsymbol{x} - \boldsymbol{s}^- \geq A\lambda, \boldsymbol{y} + \boldsymbol{s}^+ \leq \widehat{\boldsymbol{f}}(A)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq \boldsymbol{0}\} \end{aligned} \tag{27}$$

These linear models allow for the estimation of the efficiency of an input-output bundle with respect to the SVF-Splines estimated production technology. They can further be extended to calculate ε-insensitive technical efficiency, as in [37], by substituting the terms $\widehat{\boldsymbol{f}}(A)$ by $\widehat{\boldsymbol{f}}(A) - \varepsilon$ in the second restriction of each of the models above. For example, the ε-insensitive radial output efficiency would be calculated using:

$$\psi(\boldsymbol{x}, \boldsymbol{y}) = \max\{\psi : \boldsymbol{x} \geq A\lambda, \psi\boldsymbol{y} \leq (\widehat{\boldsymbol{f}}(A) - \varepsilon)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1\} \tag{28}$$

The rest of the models are analogous.

These models, however, involve many virtual points ($(d+2)^m$). In practice, we often substitute $\boldsymbol{a}_{l_1 \ldots l_m}$ by the original data to obtain a simpler model, based only on the set of points $(\boldsymbol{x}_i, \widehat{\boldsymbol{f}}(\boldsymbol{x}_i))$, $i = 1, \ldots, n$, which is computationally less expensive. The radial output-oriented model would then be calculated using the following model, with the remaining measures being analogous:

$$\psi(\boldsymbol{x}, \boldsymbol{y}) = \max\{\psi : \boldsymbol{x} \geq X\lambda, \psi\boldsymbol{y} \leq f(X)\lambda, \lambda \geq \boldsymbol{0}, \lambda\boldsymbol{1} = 1\} \tag{29}$$

4. Computational experience

In this section, we assess the quality of the SVF-Splines method by comparing it with other techniques. In particular, we perform a simulated experiment where we compare the performance of SVF-Splines with first DEA and then CSVF using different production technologies with multiple inputs and a single output, within a simulated experience extracted from [27].

The details of the simulated production functions appear in Table 1. The simulated production functions are Cobb-Douglas functions, which are widely known in the economic literature. Each scenario uses a different number of inputs, and the exponent associated with each input indicates the level of theoretical marginal contribution of each variable to the output. The exponents of the inputs add up to 0.5 in every case, which is related to non-increasing returns to scale. The input values were obtained independently and identically distributed from $Uni[1,10]$. The observed output value is then calculated by multiplying by an extra inefficiency term e^{-u}, with $u \sim \exp(1/3)$.

Each scenario was simulated with sample sizes $n = 30, 50, 70, 100$. Each combination of scenario and sample size was then replicated 50 times. The performance of each algorithm was measured using the standard Mean Squared Error (MSE) between the real, unobserved production frontier and the estimated production frontier over the 50 trials, which is calculated as $\sum_{t=1}^{50} \sum_{i=1}^{n} \left(f(\boldsymbol{x}_i^t) - \widehat{f}(\boldsymbol{x}_i^t)\right)^2 / 50n$, as well as the corresponding Bias, with formulation $\sum_{t=1}^{50} \sum_{i=1}^{n} |f(\boldsymbol{x}_i^t) - \widehat{f}(\boldsymbol{x}_i^t)| / 50n$, where the superscript t corresponds to the trial considered.

The experiments were performed in the Scientific Computation Cluster at the Miguel Hernandez University of Elche, which has a Supermicro SYS-1029GQ-TRT node, with two Intel(R) Xeon(R) Gold 6242R CPU @ 3.10 GHz processors, 80 cores and 768 GB of RAM. The algorithm was programmed in Python and CPLEX v20.1.0 was used for solving the optimization problems.

4.1. Comparison between DEA and SVF-Splines

We first compare the SVF-Splines methodology with standard DEA via the MSE and Bias scores obtained in the simulated scenarios. Table 2 reports the comparison between these two methods with respect to both MSE and Bias for each combination of scenario and sample size. The first two columns indicate the scenario (which corresponds to the number of inputs) and the sample size. The next three columns refer to the MSE associated with the DEA and SVF-Splines estimators, and the relative difference between the two methods. The final three columns indicate the Bias of the techniques and their relative difference.

Regarding the hyperparameters C, ε and d involved in SVF-Splines, we tune them using a five-fold cross validation procedure, which evaluates which combination of values yields a better estimator, as measured by the out-of-sample MSE at each fold. The sets of possible values which we consider for the hyperparameters are $C \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1, 2, 5, 10, 100\}$, $\varepsilon \in \{0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1\}$, whereas the number of partitions d along each input dimension depends on the number n of DMUs via $d = 0.1 \cdot h \cdot n$, rounded to the nearest integer, with $h = 1, \ldots, 10$.

We observe that SVF-Splines outperforms DEA in every scenario, with improvements in MSE of up to 95 % and up to 77 % in the case of Bias. The relative improvements increase as the sample size increases (except in the scenario with a single input). As the dimensionality of the problem increases, the MSE and Bias of SVF-Splines grows slower than those of DEA. These results indicate that SVF-Splines is capable of estimating production functions closer to the unobserved theoretical production function than DEA, so that we can conclude that SVF-Splines seems not overfit to the data as much as DEA. In the case of DEA, this overfitting could be attributed to the principle of minimal extrapolation (see, for example, [14]).

Finally, we point out that the new approach (SVF-splines) has demonstrated superior performance over the standard DEA approach in terms of MSE and bias within a finite-sample analysis (with $n \leq 200$). However, claiming complete superiority would be premature. It is important to consider various properties, such as consistency, from a statistical perspective. While consistency has been extensively studied for the DEA estimator (see, for example, [20]), a similar analysis is lacking for SVF-splines. Consequently, our comparison with the standard DEA model is confined to finite-sample analysis and our findings may be particularly relevant in scenarios with limited data sample sizes. Notably, a detailed analysis involving a greater number of DMUs and variables exceeds the scope of this paper but presents a potential avenue for future research.

4.2. Comparison between SVF-Splines and CSVF

In addition to standard DEA, we compare SVF-Splines with Convexified Support Vector Frontiers (CSVF). Table 3 shows those comparisons which could be performed, which were not as extensive as in the previous comparison due to the high computational cost with CSVF, as described in Table 4. The structure of Table 3 is analogous to that of Table 2. The same set of potential hyperparameter values described in the previous section was used in the Cross-Validation process of both CSVF and SVF-Splines.

We can observe in Table 3 that, while in the single-input case, CSVF performs slightly better than SVF-Splines, as soon as there are at least two inputs, SVF-Splines commits a smaller MSE and Bias than CSVF. This could be due to the fact that SVF-Splines directly estimates a convex production function while CSVF first estimates a stepwise function which is later convexified.

We now discuss the computational workload associated with CSVF and SVF-Splines. Table 4 reports the average time spent by both SVF-Splines and CSVF in the overall Cross-Validation process (CV), as well as the time used to create the Best Model (BM) and the time spent to Solve the Best Model (SBM) on average in each configuration. We can observe how CSVF is a very computationally expensive method. In particular, CSVF requires a long time to execute as soon as there are at least 2 inputs, and it scales much slower as the number of inputs and/or DMUs increases.

In particular, we can see how, already with only two inputs and 70 DMUs, the computational time needed by CSVF is around 6.5 times that required by the new approach, and this ratio only increases as the number of DMUs and inputs increase. With the largest setting solved by CSVF, SVF-Splines can solve the problem over 70 times faster than CSVF.

Table 1
Data generating processes used.

Scenario/ Num. Inputs	Data Generating Process ($f(\boldsymbol{x}) \cdot e^{-u}$)
1	$x_1^{0.5} \cdot e^{-u}$
2	$x_1^{0.4} \cdot x_2^{0.1} \cdot e^{-u}$
3	$x_1^{0.3} \cdot x_2^{0.1} \cdot x_3^{0.1} \cdot e^{-u}$
4	$x_1^{0.3} \cdot x_2^{0.1} \cdot x_3^{0.08} \cdot x_4^{0.02} \cdot e^{-u}$
5	$x_1^{0.3} \cdot x_2^{0.1} \cdot x_3^{0.08} \cdot x_4^{0.01} \cdot x_5^{0.01} \cdot e^{-u}$
6	$x_1^{0.3} \cdot x_2^{0.1} \cdot x_3^{0.08} \cdot x_4^{0.01} \cdot x_5^{0.006} \cdot x_6^{0.004} \cdot e^{-u}$
9	$x_1^{0.3} \cdot x_2^{0.1} \cdot x_3^{0.08} \cdot x_4^{0.005} \cdot x_5^{0.004} \cdot x_6^{0.001} \cdot x_7^{0.005} \cdot x_8^{0.004} \cdot x_9^{0.001} \cdot e^{-u}$
12	$x_1^{0.2} \cdot x_2^{0.075} \cdot x_3^{0.025} \cdot x_4^{0.05} \cdot x_5^{0.05} \cdot x_6^{0.08} \cdot x_7^{0.005} \cdot x_8^{0.004} \cdot x_9^{0.001} \cdot x_{10}^{0.005} \cdot x_{11}^{0.004} \cdot x_{12}^{0.001} \cdot e^{-u}$
15	$x_1^{0.15} \cdot x_2^{0.025} \cdot x_3^{0.025} \cdot x_4^{0.05} \cdot x_5^{0.025} \cdot x_6^{0.025} \cdot x_7^{0.05} \cdot x_8^{0.05} \cdot x_9^{0.08} \cdot x_{10}^{0.005} \cdot x_{11}^{0.004} \cdot x_{12}^{0.001} \cdot x_{13}^{0.005} \cdot x_{14}^{0.004} \cdot x_{15}^{0.001} \cdot e^{-u}$

Table 2
Comparison between DEA and SVF-Splines according to MSE and Bias.

Scenario	Sample Size	Mean Squared Error			Bias		
		DEA	SVF-SP	Improvement	DEA	SVF-SP	Improvement
1	**30**	0.02	0.016	17 %	0.104	0.097	7 %
	50	0.011	0.01	13 %	0.076	0.072	5 %
	70	0.007	0.006	11 %	0.059	0.057	4 %
	100	0.009	0.008	12 %	0.069	0.067	4 %
	200	0.000	0.000	3 %	0.01	0.009	2 %
2	**30**	0.074	0.035	52 %	0.198	0.135	32 %
	50	0.048	0.019	60 %	0.16	0.102	37 %
	70	0.031	0.012	60 %	0.127	0.08	38 %
	100	0.024	0.01	61 %	0.109	0.069	38 %
	200	0.001	0.001	34 %	0.026	0.018	29 %
3	**30**	0.132	0.05	62 %	0.285	0.174	40 %
	50	0.098	0.034	65 %	0.242	0.139	43 %
	70	0.07	0.022	70 %	0.203	0.105	49 %
	100	0.056	0.013	76 %	0.177	0.083	54 %
	200	0.003	0.001	63 %	0.045	0.023	49 %
4	**30**	0.2	0.064	67 %	0.342	0.191	44 %
	50	0.156	0.042	74 %	0.3	0.149	51 %
	70	0.114	0.025	78 %	0.252	0.119	53 %
	100	0.095	0.018	81 %	0.231	0.096	59 %
	200	0.006	0.001	79 %	0.060	0.024	60 %
5	**30**	0.312	0.121	61 %	0.468	0.292	38 %
	50	0.239	0.063	74 %	0.397	0.209	48 %
	70	0.225	0.058	74 %	0.387	0.2	49 %
	100	0.176	0.037	78 %	0.343	0.163	52 %
	200	0.009	0.000	82 %	0.059	0.022	62 %
6	**30**	0.313	0.098	68 %	0.465	0.256	45 %
	50	0.272	0.073	73 %	0.424	0.211	51 %
	70	0.227	0.05	78 %	0.385	0.17	56 %
	100	0.2	0.037	82 %	0.359	0.143	60 %
	200	0.012	0.001	90 %	0.084	0.026	69 %
9	**30**	0.392	0.129	68 %	0.52	0.286	45 %
	50	0.368	0.095	74 %	0.501	0.24	52 %
	70	0.337	0.067	80 %	0.475	0.203	57 %
	100	0.318	0.052	84 %	0.46	0.173	62 %
	200	0.023	0.001	94 %	0.114	0.028	76 %
12	**30**	0.413	0.135	68 %	0.534	0.296	45 %
	50	0.416	0.107	74 %	0.539	0.257	52 %
	70	0.396	0.084	79 %	0.526	0.229	57 %
	100	0.378	0.064	83 %	0.515	0.2	61 %
	200	0.030	0.002	94 %	0.135	0.031	77 %
15	**30**	0.439	0.165	63 %	0.566	0.34	40 %
	50	0.421	0.131	69 %	0.543	0.295	46 %
	70	0.427	0.093	78 %	0.55	0.245	56 %
	100	0.408	0.078	81 %	0.537	0.221	59 %
	200	0.034	0.002	95 %	0.147	0.033	77 %

Table 3
Comparison between SVF-Splines and CSVF according to MSE and Bias.

Scenario	Sample Size	Mean Squared Error			Bias		
		SVF-SP	CSVF	Improvement	SVF-SP	CSVF	Improvement
1	**30**	0.016	0.016	0 %	0.097	0.09	−8 %
	50	0.01	0.008	−25 %	0.072	0.066	−9 %
	70	0.006	0.005	−20 %	0.057	0.053	−8 %
	100	0.008	0.008	0 %	0.067	0.064	−5 %
2	**30**	0.035	0.05	30 %	0.135	0.164	18 %
	50	0.019	0.03	37 %	0.102	0.125	18 %
	70	0.012	0.021	43 %	0.08	0.105	24 %
	100	0.01	0.017	41 %	0.069	0.087	21 %
3	**30**	0.005	0.009	44 %	0.089	0.077	13 %
	50	0.034	0.058	41 %	0.139	0.192	28 %

Furthermore, the estimated MSE and Bias obtained are lower for the SVF-Splines than for the CSVF algorithm.

Computational times used by SVF-Splines with more inputs and DMUs are reported in Fig. 3. It can be seen that, even in the most expensive case of 100 DMUs with 15 inputs, the computational load is lower than that of CSVF with 3 inputs and 30 DMUs.

The computational burden can be observed to be more in that the creation of models, and in particular the Best Model (BM), which is much more computationally expensive for CSVF than for SVF-Splines. The Solution of the Best Model (SBM), despite the size of the models and how expensive they are to construct, is very fast in both cases. We can attribute this improvement of the new approach to the grid associated to the transformation function, which for CSVF requires a number of parameters which is exponential in the number of inputs m ($\approx md^m$),

Table 4
Execution time of the SVF and SVF-Splines algorithms.

Scenario	Sample Size	SVF-SP			CSVF		
		CV	BM	SBM	CV	BM	SBM
1	30	161.02	0.02	0.0	137.68	0.04	0.0
	50	196.22	0.0	0.02	157.02	0.06	0.0
	70	242.26	0.12	0.04	195.46	0.2	0.0
	100	325.68	0.06	0.02	262.3	0.24	0.0
2	30	188.68	0.08	0.0	324.18	0.4	0.0
	50	251.7	0.06	0.02	925.4	2.12	0.0
	70	335.52	0.1	0.06	2166.32	5.82	0.04
	100	489.0	0.14	0.02	5401.16	13.74	0.06
3	30	208.66	0.04	0.0	4697.6	7.64	0.06
	50	299.28	0.06	0.0	21,764.54	46.02	0.18

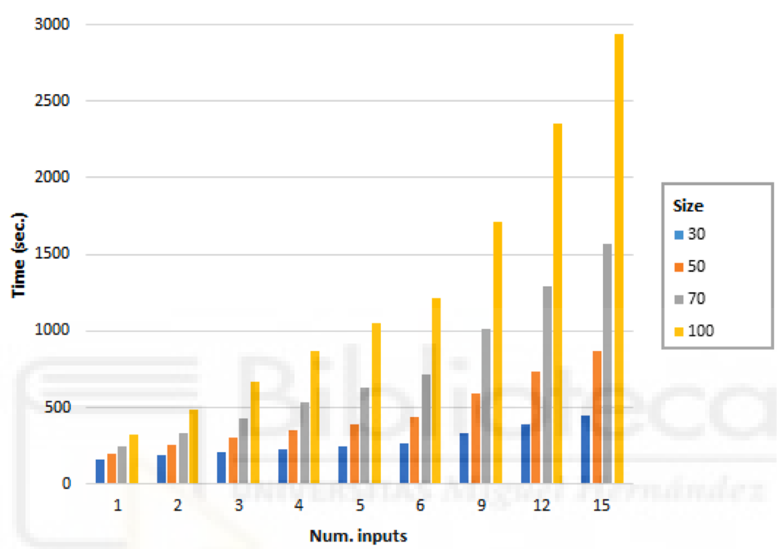

Fig. 3. SVF-Splines runtime according to sample size and number of inputs.

as well as corresponding restrictions. On the other hand, the transformation function related to SVF-Splines has a linear number of parameters in the inputs ($\approx m(d+3)$).

From this computational experiment, we can conclude that SVF-Splines is a method which reduces the overfitting present in DEA, obtaining estimations closer to the unobserved efficient frontier. It furthermore improves upon the results obtained by CSVF, with much lower computational time required and lower MSE and Bias as soon as the number of inputs is at least 2.

5. Empirical illustration

In this section, we present an illustration of the results that can be obtained by SVF-Splines in an empirical database from the literature.

Table 5
Descriptive statistics of the Spanish tax offices dataset.

	x1	x2	x3	y1	y2
Average	32	91	57.6	10.45	8.91
Median	23	70	22.8	7.39	6.59
Std. Dev.	30	74	88.3	9.49	9.60
Max.	191	368	420.2	49.27	60.02
90 %	60	194	176.7	20.65	18.97
75 %	39	109	63.7	10.71	10.18
25 %	17	43	8.9	5.30	3.62
10 %	11	32	5.0	3.88	2.14
Min.	7	21	0.8	2.51	1.48

This database consists of 44 regional tax offices in Spain, which oversee tax collection in almost all Spanish provinces. The data is from 2011, and the database was used in [3]. We calculate the efficiency scores using the efficiency measures introduced above (Section 3.4), that is, the output and input-oriented radial models (24), (25), the directional distance function (26) with directional vector $\boldsymbol{g} = (\boldsymbol{x},\boldsymbol{y})$, and the Weighted Additive (27) with the weights corresponding to the Range Adjusted Measure, that is, $\rho^{-(j)} = \frac{1}{(m+s)R_j^-}$ and $\rho^{+(r)} = \frac{1}{(m+s)R_r^+}$, where R_j^- is the range of input j and R_r^+ is the range of values of output r.

The dataset used contains three variables considered as inputs, and two outputs. There are two inputs related to labour: the number of tax inspectors and specialists (x_1) and the number of workers in the rest of the workforce (x_2). We also consider as an input the number of successful complaints by taxpayers against the tax authority (x_3). This variable indicates an output that a manager should desire to minimize, sometimes called a bad output in the literature. Thus, it is considered as an input in the model, as in [19,25]. The two outputs considered correspond to the two main taxes collected by the tax offices: the inheritance and gift tax (y_1) and the real estate transfer tax settlements processed (y_2).

The data was normalized so that the values of the variables lie between 0 and 1 by dividing the values of each variable by the maximum attained value. This does not change the values of the measures of efficiency, which are all units-invariant. Table 5 shows descriptive statistics of the dataset before normalization.

The sets of potential hyperparameter values considered in the five-

Table 6
Estimated efficiencies in the Spanish tax offices dataset.

DMU	Radial Output			Radial Input			Weighted Additive (RAM)			Directional Distance Function		
	DEA	SVF-SP	ε-insensitive efficient	DEA	SVF-SP	ε-insensitive efficient	DEA	SVF-SP	ε-insensitive efficient	DEA	SVF-SP	ε-insensitive efficient
ALMERIA	1.4075	1.5825	No	0.6966	0.5892	No	0.1046	0.1118	No	0.4075	0.4512	No
CADIZ	1.1414	1.2755	No	0.8866	0.7504	No	0.0649	0.0824	No	0.1297	0.2305	No
CORDOBA	1.5738	1.9151	No	0.4969	0.4468	No	0.0692	0.0745	No	0.3691	0.5684	No
GRANADA	1.2926	1.4025	No	0.7888	0.6659	No	0.0619	0.0716	No	0.2899	0.3494	No
HUELVA	1.4185	1.5963	No	0.7937	0.5286	No	0.0355	0.0416	No	0.4185	0.5416	No
JAEN	1.3502	1.6406	No	0.7441	0.5945	No	0.0480	0.0540	No	0.3217	0.4803	No
MALAGA	1.0000	1.0935	Yes	1.0000	0.9022	Yes	0.0000	0.0625	Yes	0.0000	0.0892	Yes
SEVILLA	1.1930	1.2155	No	0.8076	0.8012	No	0.0523	0.0808	No	0.1402	0.1947	No
HUESCA	1.0000	1.6517	Yes	1.0000	0.9511	Yes	0.0000	0.0202	Yes	0.0000	0.4394	Yes
TERUEL	1.0000	1.2668	Yes	1.0000	1.0000	Yes	0.0000	0.0141	Yes	0.0000	0.0000	Yes
ZARAGOZA	1.0000	1.0666	Yes	1.0000	0.9341	Yes	0.0000	0.0233	Yes	0.0000	0.0401	Yes
OVIEDO	1.0000	1.0000	Yes	1.0000	1.0000	Yes	0.0000	0.0247	Yes	0.0000	0.0000	Yes
BALEARES	1.2670	1.8781	No	0.8114	0.5058	No	0.0581	0.0625	No	0.2285	0.3861	No
CANTABRIA	1.2319	1.3858	Yes	0.8058	0.6540	Yes	0.0243	0.0293	Yes	0.2049	0.3531	Yes
ALBACETE	1.2181	1.3919	Yes	0.8598	0.6805	Yes	0.0135	0.0218	Yes	0.1867	0.3651	Yes
CIUDAD REAL	1.0000	1.0000	Yes	1.0000	0.9996	Yes	0.0000	0.0032	Yes	0.0000	0.0000	Yes
CUENCA	1.2651	1.9234	No	0.8621	0.6563	No	0.0178	0.0307	No	0.2398	0.6550	No
GUADALAJARA	1.0364	1.7132	Yes	0.9778	0.6335	Yes	0.0083	0.0233	Yes	0.0345	0.5594	Yes
TOLEDO	1.1788	1.2682	Yes	0.7632	0.7387	Yes	0.0211	0.0260	Yes	0.1512	0.2588	Yes
AVILA	1.0000	1.2552	Yes	1.0000	1.0000	Yes	0.0000	0.0152	Yes	0.0000	0.0000	Yes
BURGOS	1.1882	1.2954	Yes	0.8444	0.7433	Yes	0.0172	0.0225	Yes	0.1672	0.2772	Yes
LEÓN	1.0000	1.0067	Yes	1.0000	0.9897	Yes	0.0000	0.0151	Yes	0.0000	0.0058	Yes
PALENCIA	1.1641	1.3862	Yes	0.8632	0.8000	Yes	0.0084	0.0205	Yes	0.1544	0.3670	Yes
SALAMANCA	1.0139	1.0910	Yes	0.9875	0.9035	Yes	0.0014	0.0142	Yes	0.0135	0.0841	Yes
SEGOVIA	1.2139	1.4089	Yes	0.8711	0.7071	Yes	0.0093	0.0228	Yes	0.1682	0.3802	Yes
SORIA	1.0000	1.8165	Yes	1.0000	1.0000	Yes	0.0000	0.0203	Yes	0.0000	0.0000	Yes
VALLADOLID	1.3574	1.5670	No	0.7649	0.6038	No	0.0305	0.0360	No	0.3362	0.4688	No
ZAMORA	1.0000	1.0805	Yes	1.0000	0.9086	Yes	0.0000	0.0162	Yes	0.0000	0.0746	Yes
BARCELONA	1.0000	1.0000	Yes	1.0000	1.0000	Yes	0.0000	0.1021	Yes	0.0000	0.0000	Yes
GIRONA	1.0000	1.1410	Yes	1.0000	0.8714	Yes	0.0000	0.0286	Yes	0.0000	0.0802	Yes
LLEIDA	1.1634	1.2723	Yes	0.8622	0.7531	Yes	0.0160	0.0240	Yes	0.1457	0.2472	Yes
TARRAGONA	1.0000	1.1972	Yes	1.0000	0.7872	Yes	0.0000	0.0238	Yes	0.0000	0.1921	Yes
BADAJOZ	1.6460	1.7975	No	0.6317	0.4785	No	0.0699	0.0750	No	0.5002	0.7027	No
CACERES	1.4216	1.6921	No	0.7855	0.5000	No	0.0245	0.0321	No	0.3863	0.5989	No
A CORUÑA	1.0000	1.0000	Yes	1.0000	1.0000	Yes	0.0000	0.0046	Yes	0.0000	0.0000	Yes
LUGO	1.0000	1.0236	Yes	1.0000	0.9710	Yes	0.0000	0.0147	Yes	0.0000	0.0206	Yes
OURENSE	2.0479	2.1611	No	0.5134	0.4196	No	0.0469	0.0514	No	0.6079	0.6995	No
PONTEVEDRA	1.2544	1.3647	No	0.7844	0.7178	No	0.0545	0.0672	No	0.2519	0.2821	No
LA RIOJA	1.0000	1.3953	Yes	1.0000	0.7593	Yes	0.0000	0.0269	Yes	0.0000	0.3182	Yes
MADRID	1.0000	1.0000	Yes	1.0000	1.0000	Yes	0.0000	0.0105	Yes	0.0000	0.0000	Yes
MURCIA	1.2786	1.3135	No	0.7771	0.7422	No	0.1426	0.1808	No	0.2496	0.2610	No
ALICANTE	1.0000	1.2054	No	1.0000	0.8316	No	0.0000	0.1235	No	0.0000	0.1379	No
CASTELLÓN	1.6437	1.8878	No	0.6331	0.5140	No	0.0584	0.0637	No	0.5191	0.6429	No
VALENCIA	1.0000	1.0000	Yes	1.0000	0.9993	Yes	0.0000	0.0285	Yes	0.0000	0.0000	Yes
Average	1.1811	1.3779		0.8776	0.7735		0.0241	0.0432		0.1505	0.2683	
Std. dev.	0.2304	0.3143		0.1398	0.1829		0.0326	0.0362		0.1735	0.2279	
# eff. Units	19	6	27	19	7	27	19	0	27	19	9	27

fold cross-validation process were: $C \in \{0.001, 0.01, 0.05, 0.1, 0.5, 1, 2, 5, 10\}$, $\varepsilon \in \{0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2\}$. We remark that, since the values of each variable were normalized to lie between 0 and 1, a margin of 0.1 corresponds 10 % of the maximum value of each output variable, i.e., already quite a large margin.[1] The value of d was chosen within $d = 0.1 \cdot h \cdot n$ for $h = 1, \ldots, 20$, rounded to the closest integer. In other words, $d = \{4, 9, 13, 18, \ldots, 70, 75, 79, 84, 88\}$. The best hyperparameter combination was chosen as $C = 1.0$, $\varepsilon = 0.05$ and $d = 4$. This indicates an ε-insensitive margin of 5 % of the maximum along each output, with the input space being divided into $(d+2)^m = 6^3 = 216$ cells via $m(d+3) = 3 \cdot 7 = 21$ weight hyperparameters.

Table 6 then presents the efficiencies estimated by DEA and SVF-Splines with respect to each of the measures of efficiency, as well as whether each DMU is considered ε-insensitive efficient with respect to the estimated production technology. The final three rows of Table 6 indicate the average efficiency measured by each method, the corresponding standard deviation, and the number of DMUs considered efficient with respect to each algorithm and measure. We observe that SVF-Splines consistently estimates higher inefficiencies, indicating a frontier which fits the data less closely than that of DEA, which suffers from overfitting.[2] We observe that, while DEA always considers 19 units as efficient regardless of the measure, SVF-Splines considers 7 DMUs as efficient in the case of the Radial Input measure and 9 units when the Directional Distance Function is applied. Moreover, under the SVF-Splines model, 6 DMUs are considered efficient with respect to the Radial Output measure, and no DMUs are considered fully efficient with

[1] Support Vector Regression (SVR), foundational to the SVF-splines approach, depends on hyperparameters, including the margin. This margin defines the error-free region around the regression line and is crucial in SVR models. Tuning it involves considering various values, a non-trivial task. Normalizing data allows reinterpretation of margin values, with a margin of z corresponding to 100·z% of the maximum output value. For the empirical case, we chose a margin of 0.2, which seems large enough.

[2] In all cases, a higher value of the measure indicates a greater level of inefficiency, with the exception of the Radial Input measure, where the interpretation is reversed: higher values denote better technical efficiency.

respect to the Range Adjusted Measure (WA), which is capable of detecting additional sources of inefficiencies along any variables, and projects to the strongly efficient frontier.

Regarding ε-insensitive technical efficiency, 27 DMUs are considered efficient with this margin of $\varepsilon = 0.05$. This can be seen as an indication that the remaining 17 DMUs can be considered very far from efficient, whereas those 27 which are within the margin can be considered to be close to efficient with respect to this more robust notion of efficiency, even if they are not completely efficient.

Additionally, we use the Li test adapted to the production context [33], based on the Li test [22], as a tool to compare the vectors of efficiency scores estimated by each method with respect to the same measures of efficiency. This is a nonparametric statistical test for the similarity of two distributions of technical efficiency scores. We remark that the efficiencies cannot be compared between different measures, since each of them has different properties, whether in orientation, range of potential values, and other properties. Hence, we compare the efficiencies obtained, always with respect to the same measure, by both DEA and SVF-Splines.

The version of the Li test that we use requires that the efficiencies are in the output orientation. That is, that efficient units attain the value 1, and larger values indicate higher inefficiencies, without an upper bound, which is the same orientation as the output-oriented radial measure. Therefore, we transform the other measures appropriately. For the input-oriented measure, we transform it to its reciprocal $1\ /x$. For the DDF and RAM, which are bounded measures of inefficiency, the corresponding transformation is $1/(1-x)$. This is because, with the choices of weights used, they are inefficiency measures bounded between 0 and 1, with efficient units attaining 0. The corresponding measures of inefficiency are $1-x$, but these are oriented as an input measure. In order to orient them in the output sense as required for the Li Test, we therefore calculate their reciprocal.

We present in Fig. 4 the kernel density distributions comparing the efficiency scores estimated by DEA and SVF-Splines. We again observe that SVF-Splines estimates higher average inefficiencies. The Li test yields evidence that the distributions are indeed statistically significantly different according to this measure in the radial measures and RAM. Regarding the DDF, the p-value in question was 0.079, which does not allow for this conclusion at a significance level of 0.05, but would do so at the significance value of 0.10 also recommended by Simar and Zelenyuk.

Therefore, we can observe in this empirical illustration that SVF-Splines shows higher discriminatory power than DEA by considering fewer units as fully efficient, estimating higher average inefficiencies which show statistically significant differences compared to DEA. Furthermore, SVF-Splines adds an additional layer of classification in the difference between the estimated efficiencies and the ε-insensitive efficient units. This notion of ε-insensitivity can be seen as a more robust region of efficiency, which allows to label some DMUs as being highly inefficient.

6. Conclusions and future work

Traditional non-parametric methods for the measurement of

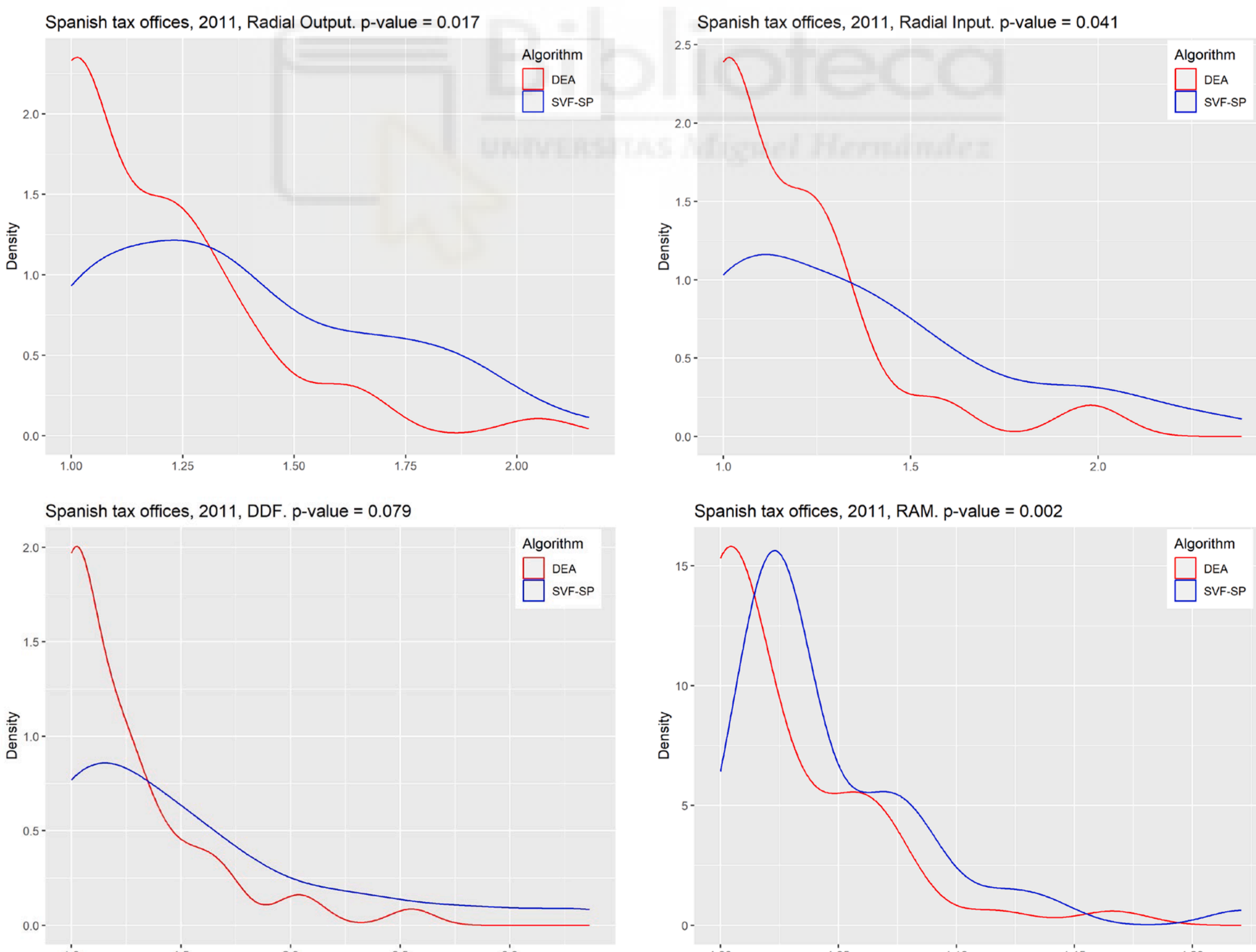

Fig. 4. Kernel density plots for different measures comparing efficiencies estimated by DEA and SVF-Splines.

efficiency such as FDH and DEA have many interesting properties, but have received criticism regarding their performance, such as that they suffer from a problem of overfitting to the observed data (see, for example, [14] or [35]), which can be attributed to the principle of minimal extrapolation and may result in lack of generalization capability (low inferential power). An area of recent interest in the measurement of efficiency is the use of machine-learning methodologies to improve the estimators by overcoming these overfitting issues and provide them with generalization capability and robustness.

One approach of particular interest is Support Vector Frontiers (SVF) ([36], 2022), which adapts Support Vector Regression to estimate stepwise production technologies satisfying the axioms of Free Disposal Hull except for minimal extrapolation. These production technologies can then be convexified to obtain convex production technologies, along the lines of DEA. However, the SVF approach presents some limitations regarding its computational time, as well as its two-stage estimation, where first a stepwise production frontier is calculated, which is later convexified.

In this paper, we have proposed an extension of Support Vector Frontiers by using a transformation function involving linear splines, that is, with a transformation function involving kernels which generate splines of order 1. The transformation function in SVF could be considered as a splines transformation of order 0. This extension, which we denote by SVF-Splines, allows for the direct estimation of convex, piecewise linear, DEA-type, production technologies, in both a single and multi-output context, with lower computational costs. This paper defines the SVF-Splines model and characterizes its corresponding production technology. The properties satisfied by DEA, except for minimal extrapolation, are established for the production possibility set estimated by SVF-Splines in a multi-input multi-output context. These properties are envelopment of the observed data, free disposability of inputs and outputs and convexity. Furthermore, the estimated set is identified with a DEA estimator defined on a set of virtual points. Subsequently, we have shown how to estimate technical inefficiency in the SVF-Splines context for a variety of standard measures of efficiency from the literature.

This contribution therefore proposes a significant improvement to the SVF estimator by using a different transformation function under the assumption of convexity. The change of kernel results in a different set of restrictions being used to guarantee the satisfaction of the microeconomic postulates.

The validity of the proposed approach is evaluated by a simulated experiment where it was compared with DEA and CSVF (the convexified version of the original SVF by Valero-Carreras, 2021, 2022). The results from the simulation show that SVF-Splines can estimate production technologies closer to the unobserved theoretical production frontier than DEA and SVF-Splines, as measured using MSE and Bias statistics. Furthermore, the computational time required by SVF-Splines is much lower than that of CSVF, which must create and solve models with a number of variables and restrictions which is exponential in the number of inputs, whereas SVF-Splines only has linearly many such variables and restrictions. Regarding the results, SVF-Splines obtained a similar performance to DEA and CSVF in the single-input single-output scenario. However, when the number of inputs is at least 2, SVF-Splines obtained improvements between 52 % and 95 % in MSE and of 32 % and 77 % in Bias with respect to DEA, and between 30 % and 44 % in MSE and between 18 % and 28 % in Bias for SVF-Splines when compared to CSVF. Regarding the comparison with CSVF, only those scenarios with at most 3 inputs and 50 DMUs were solved, since in larger scenarios the computation time required by CVSF becomes impracticable.

Furthermore, an empirical example is provided to illustrate the results that can be obtained using SVF-Splines. The database under study consists of 44 regional tax offices in Spain. From this example, it can be observed how SVF-Splines classifies a fewer DMUs as efficient than DEA and how it estimates overall smaller average efficiencies, thus identifying more sources of inefficiency. Efficiency scores are calculated with respect to the radial output and input measures, the Directional Distance Function, and the Weighted Additive model, to illustrate the variety of measures available in the literature of nonparametric efficiency estimation. In addition, comparisons between the estimated vectors of efficiencies with respect to the same measure show significant differences with respect to efficiency measures between DEA and SVF-Splines. Additionally, when the more robust concept of epsilon-insensitive efficiency is considered, the method classifies a larger number of DMUs as efficient, which can be interpreted as evidence pointing towards that those that are still labelled as inefficient through the new approach can be considered to be very far from technical efficiency.

Regarding the possible utilization by a decision-maker, while he may still be interested in a quick estimation and does not care about overfitting to the available data, DEA is a valid option. However, if the goal is to obtain more robust results and to perform inference, SVF-Splines is an option to consider. Regarding the limitations of the method, SVF-Splines requires a cross-validation procedure which involves large grid of values to establish three hyperparameters, which results in some computational load. It is still slower than DEA, which involves only linear programming. This could be further improved by narrowing down the potential ranges of values. The determination of knots may also be problematic and could be performed in other ways (for example, a non equi-distance split of each input dimension).

We finally mention various potential lines of future work, such as the use of estimators involving higher order splines (quadratic, cubic, etc.). Other types of transformation functions such as Radial Basis Functions or Gaussian kernels could also be used. We observe that the restrictions on the parameters which ensure that the production axioms hold are not the same as in SVF, and more information about which properties of SVR allow the satisfaction of the properties may be interesting, regarding other potential kernels and their associated constraints. In this paper we have focused our attention on measuring the efficiency of units in a single dataset, and the proposed algorithm could be adapted to work with panel data, and the study of whether the differences in efficiency arise from its various sources (efficiency change, scale efficiency change, technical change). Further validation using more real databases in different contexts would always be useful. Other robustness increasing methods such as relaxed support vector regression [30] or based on the directional distance function [4] could also be considered. Additional future research could entail comparative analyses among DEA-related methods, such as Supper-Efficiency DEA [2] and SBM (Slack-Based Measure) (see [34]), and the new approach (SVF-Splines) to elucidate their relative advantages and limitations in different contexts. Additionally, further research is warranted to delve into the 'black-box' mechanism underlying support vector-based approaches for constructing production frontiers. By addressing these areas, we aim to enhance our understanding of efficiency measurement techniques and contribute to the advancement of knowledge in this field.

CRediT authorship contribution statement

Nadia M. Guerrero: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Raul Moragues:** Writing – original draft, Visualization, Formal analysis. **Juan Aparicio:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition. **Daniel Valero-Carreras:** Validation, Software, Data curation.

Declaration of competing interest

The authors notify that there's no financial/personal interest or belief that could affect their objectivity.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank the grant PID2022–136383NB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU. D. Valero thanks the grant ACIF/2020/155 (*Generalitat Valenciana*). Additionally, J. Aparicio thanks the grant PROMETEO/2021/063 funded by the Valencian Community (Spain).

References

[1] Aigner D, Lovell CAK, Schmidt P. Formulation and estimation of stochastic frontier production function models. J Econom 1977;6(1):21–37. https://doi.org/10.1016/0304-4076(77)90052-5.
[2] Andersen P, Petersen NC. A procedure for ranking efficient units in data envelopment analysis. Manage Sci 1993;39(10):1261–4. https://doi.org/10.1287/mnsc.39.10.1261.
[3] Aparicio J, Cordero JM, Díaz-Caro C. Efficiency and productivity change of regional tax offices in Spain: an empirical study using Malmquist–Luenberger and Luenberger indices. Empir Econ 2020;59(3):1403–34. https://doi.org/10.1007/s00181-019-01667-8.
[4] Arabmaldar A, Sahoo BK, Ghiyasi M. A generalized robust data envelopment analysis model based on directional distance function. Eur J Oper Res 2023;311(2):617–32. https://doi.org/10.1016/j.ejor.2023.05.005.
[5] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Manage Sci 1984;30(9):1078–92. https://doi.org/10.1287/mnsc.30.9.1078.
[6] Blanco V, Puerto J, Rodriguez-Chia AM. On lp-support vector machines and multidimensional kernels. J Mach Learn Res 2020;21(14):1–29. http://jmlr.org/papers/v21/18-601.html.
[7] Chambers RG, Chung Y, Färe R. Profit, directional distance functions, and nerlovian efficiency. J Optim Theory Appl 1998;98(2):351–64. https://doi.org/10.1023/A:1022637501082.
[8] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. Eur J Oper Res 1978;2(6):429–44. https://doi.org/10.1016/0377-2217(78)90138-8.
[9] Cooper WW, Park KS, Pastor JT. RAM: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. J Productivity Anal 1999;11(1):5–42. https://doi.org/10.1023/A:1007701304281.
[10] Daouia A, Noh H, Park BU. Data envelope fitting with constrained polynomial splines. J R Stat Soc 2016;78(1):3–30. https://doi.org/10.1111/RSSB.12098. *Series B: Statistical Methodology*.
[11] Daraio C, Simar L. Advanced robust and nonparametric methods in efficiency analysis, Vol. 4. Springer US; 2007. https://doi.org/10.1007/978-0-387-35231-2.
[12] Deprins D, Simar L, Tulkens H. Measuring labor-efficiency in post offices. Université catholique de Louvain, Center for Operations Research and Econometrics (CORE); 1984. p. 243–67. https://doi.org/10.1007/978-0-387-25534-7_16. M. Marchand, P. Pestieau, & H. Tulkens, Eds.LIDAM Reprints CORE 571.
[13] España VJ, Aparicio J, Barber X, Esteve M. Estimating production functions through additive models based on regression splines. Eur J Oper Res 2023. https://doi.org/10.1016/j.ejor.2023.06.035.
[14] Esteve M, Aparicio J, Rabasa A, Rodriguez-Sala JJ. Efficiency analysis trees: a new methodology for estimating production frontiers through decision trees. Expert Syst Appl 2020;162:113783. https://doi.org/10.1016/j.eswa.2020.113783.
[15] Farrell MJ. The measurement of productive efficiency. J R Stat Soc Ser A 1957;120(3):253. https://doi.org/10.2307/2343100.
[16] Guerrero NM, Aparicio J, Valero-Carreras D. Combining data envelopment analysis and machine learning. Mathematics 2022;10(6):909. https://doi.org/10.3390/MATH10060909.
[17] Guillen MD, Aparicio J, Esteve M. Gradient tree boosting and the estimation of production frontiers. Expert Syst Appl 2023;214:119134. https://doi.org/10.1016/J.ESWA.2022.119134.
[18] Guillen MD, Aparicio J, Esteve M. Performance evaluation of decision-making units through boosting methods in the context of free disposal hull: some exact and heuristic algorithms. Int J Inf Technol Decis Mak 2023. https://doi.org/10.1142/S0219622023500050. published online.
[19] Hailu A, Veeman TS. Environmentally sensitive productivity analysis of the Canadian pulp and paper industry, 1959-1994: an input distance function approach. J Environ Econ Manage 2000;40(3):251–74. https://doi.org/10.1006/JEEM.2000.1124.
[20] Kneip A, Park BU, Simar L. A note on the convergence of nonparametric DEA estimators for production efficiency scores. Econ Theory 1998;14(6):783–93. https://doi.org/10.1017/S0266466698146042.
[21] Kuosmanen T, Johnson AL. Data envelopment analysis as nonparametric least-squares regression. Oper Res 2010;58(1):149–60. https://doi.org/10.1287/opre.1090.0722.
[22] Li Q. Nonparametric testing of closeness between two unknown distribution functions. Econom Rev 1996;15(3):261–74. https://doi.org/10.1080/07474939608800355.
[23] Lovell CAK, Pastor JT. Units invariant and translation invariant DEA models. Operations Research Letters 1995;18(3):147–51.
[24] Madden P. Concavity and optimization in microeconomics. B. Blackwell; 1986.
[25] Mahlberg B, Sahoo BK. Radial and non-radial decompositions of Luenberger productivity indicator with an illustrative application. Int J Prod Econ 2011;131(2):721–6. https://doi.org/10.1016/J.IJPE.2011.02.021.
[26] Meeusen W, van Den Broeck J. Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev (Philadelphia) 1977;18(2):435. https://doi.org/10.2307/2525757.
[27] Moragues R, Aparicio J, Esteve M. An unsupervised learning-based generalization of data envelopment analysis. Oper Res Perspect 2023;11:100284. https://doi.org/10.1016/j.orp.2023.100284.
[28] Moragues R, Aparicio J, Esteve M. Measuring technical efficiency for multi-input multi-output production processes through OneClass support vector machines: a finite-sample study. Oper Res 2023;23(3):47. https://doi.org/10.1007/s12351-023-00788-4.
[29] Olesen OB, Ruggiero J. The hinging hyperplanes: an alternative nonparametric representation of a production function. Eur J Oper Res 2022;296(1):254–66. https://doi.org/10.1016/j.ejor.2021.03.054.
[30] Panagopoulos OP, Xanthopoulos P, Razzaghi T, Şeref O. Relaxed support vector regression. Ann Oper Res 2019;276(1–2):191–210. https://doi.org/10.1007/S10479-018-2847-6/TABLES/6.
[31] Parmeter, C.F., & 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*, 463–88. 10.1007/978-1-4614-1653-1_18/FIGURES/6.
[32] Pedroso JP, Murata N. Support vector machines with different norms: motivation, formulations and results. Pattern Recognit Lett 2001;22(12):1263–72. https://doi.org/10.1016/S0167-8655(01)00071-X.
[33] Simar L, Zelenyuk V. On testing equality of distributions of technical efficiency scores. Econom Rev 2006;25(4):497–522. https://doi.org/10.1080/07474930600972582.
[34] Tone K. A slacks-based measure of efficiency in data envelopment analysis. Eur J Oper Res 2001;130(3):498–509. https://doi.org/10.1016/S0377-2217(99)00407-5.
[35] Tsionas M. Efficiency estimation using probabilistic regression trees with an application to Chilean manufacturing industries. Int J Prod Econ 2022;249:108492. https://doi.org/10.1016/J.IJPE.2022.108492.
[36] Valero-Carreras D, Aparicio J, Guerrero NM. Support vector frontiers: a new approach for estimating production functions through support vector machines. Omega (Westport) 2021;104:102490. https://doi.org/10.1016/j.omega.2021.102490.
[37] Valero-Carreras D, Aparicio J, Guerrero NM. Multi-output support vector frontiers. Comput Oper Res 2022;143:105765. https://doi.org/10.1016/J.COR.2022.105765.
[38] Vapnik V. Principles of risk minimization for learning theory. Adv Neural Inf Process Syst 1991;4:831–8.
[39] Vapnik V. Statistical learning theory. Wiley; 1998.
[40] Vazquez E, Walter E. Multi-output support vector regression. IFAC Proc Volumes 2003;36(16):1783–8. https://doi.org/10.1016/S1474-6670(17)35018-8.
[41] Zhu J, Rosset S, Tibshirani R, Hastie T. 1-norm support vector machines. Adv Neural Inf Process Syst 2003;16:1–8.

10. AGRADECIMIENTOS

Como autora de esta tesis, Dña. *Nadia María Guerrero Martínez*, traslado los siguientes agradecimientos:

A la Universidad Miguel Hernández, por ser un organismo a la par de importante cercano hacia el estudiante. Por ser una universidad que aboga por la innovación, por la originalidad y por las oportunidades, sin perder la contundencia de las disciplinas más rigurosas.

Al Centro de Investigación Operativa de la Universidad Miguel Hernández, por su gran profesionalidad, y apoyo hacia el estudiante y la investigación en general, hecho que me ha permitido sentirme en este tiempo dentro de una gran familia en la que confiar, en particular a la hora de enfrentarme a la docencia asociada al doctorado.

Al programa de doctorado en Estadística, Optimización y Matemática Aplicada, así como a su coordinador el Dr. D. Domingo Morales González, por ofrecer diferentes actividades que me permitieron acercarme a toda la materia objeto de estudio de manera natural, con dinamismo y motivación, nutriéndome de las distintas disciplinas.

Al Dr. D. Juan Aparicio Baeza, director de esta tesis, por ofrecerme la oportunidad de unirme a su innovador proyecto y por dirigir con gran rigor cada paso de la investigación, perfeccionando cada etapa a través de sus explicaciones en la materia y de la compartición de sus conocimientos. Así mismo, agradecerle todo el anhelo hacia la materia de estudio, pues es esto lo que me ha llevado a afianzar gran confianza en que los resultados obtenidos iban en la línea adecuada. Por último, por toda su atención, por todo el apoyo y por toda su perseverancia en el proceso de dar luz a esta tesis y los artículos que la componen.

A Daniel Valero y a Raúl Moragues por sacar adelante conmigo los artículos que dan lugar a esta tesis y por trabajar en ellos con todo el respeto y perfección, aportándoles todos los conocimientos traídos de sus disciplinas.

A la “sala P.I.F. (Personal Investigador en Formación)”, por hacerme sentir durante varios años todo el apoyo y la alegría de personas que comienzan su andadura en el mundo de la investigación, lo que tornó los momentos más tediosos en mero tránsito hacia la luz del conocimiento.

A todos los profesores que han pasado por mi vida, por haber compartido conmigo sus conocimientos y sus experiencias, que me han ayudado a crecer como persona y a recorrer mi vida académica con pasión.

A todos mis tíos y en especial a mis padrinos, por toda su motivación y por compartir conmigo todos sus conocimientos sobre el mundo académico.

A mi hermano, Rodrigo, por ayudarme a afianzar el valor de la responsabilidad al convertirme en hermana mayor, lo que me ha acompañado en todo este camino.

A mi padre, Francisco Rodrigo Guerrero González, por las tardes de matemáticas y dibujo técnico desde mis seis años de vida, y por trasladarme su cariño por los libros y por aprender, hecho que me ha permitido vivir la vida académica con desenfado y naturalidad.

A mi madre, Josefa María Martínez Gómez, por enseñarme el valor de la positividad y de ver el lado bueno de todas las cosas, por enseñarme la sencillez a la hora de resolver problemas y el valor del trabajo, valores que han impregnado toda mi carrera académica.

A mi marido, Samuel Mateo Riquelme, por ayudarme a sacar a la luz mi mejor versión en todos los ámbitos de mi vida, tanto en el terreno empresarial, académico como personal. Por permitirme conocer el mundo con él y poder aprender valores y

conocimientos que han influido de forma directa e indirecta en el desarrollo de esta tesis.

A mis hijos, por enseñarme el valor de la dimensión del tiempo, por motivarme a finalizar esta tesis solo con el simple hecho de ser, por impulsarme a conocer mejor nuestro mundo para mostrárselo a ellos y a ser mejor cada día, por ayudarme también a entrenar mi parte más ancestral complementando todo el entrenamiento realizado en estos años sobre mi parte más racional.

A las Matemáticas, por ser para mí más que una profesión, por ser mi modo de vida.