

**Universidad de Las Palmas de Gran Canaria**  
**Departamento de Métodos Cuantitativos en Economía y Gestión**



**Eficiencia de las sucursales bancarias.**  
**El caso de un banco comercial español**

**Tesis Doctoral**

**Idaira Esther Cabrera Suárez**

**2022**

**Director:** Dr. Jorge V. Pérez Rodríguez  
Universidad de Las Palmas de Gran Canaria

**UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA  
ESCUELA DE DOCTORADO**

Programa de doctorado en **Turismo, Economía y Gestión**

Título de la Tesis **“Eficiencia de las Sucursales Bancarias. El  
Caso de un Banco Comercial Español”**

Tesis Doctoral presentada por **D<sup>a</sup> Idaira Esther Cabrera Suárez**

Dirigida por el Dr. **D. Jorge Vicente Pérez Rodríguez**

Codirigida por el Dr/a. D/D<sup>a</sup>. .....

Las Palmas de Gran Canaria, a **15** de **06** de 20**22**

El/la Director/a,

El/la Codirector/a

El/la Doctorando/a,

PEREZ  
RODRIGUEZ  
JORGE  
VICENTE -

(firma)

(firma)

(firma)



# Agradecimientos

El tema de esta tesis doctoral se cristalizó dentro del Departamento de Control de Gestión del Banco de Bilbao Vizcaya Argentaria (BBVA). Actualmente, desarrollo mi labor como Técnico Global dentro del Departamento de Transformación y Productividad de la Dirección Territorial desde 2021 y anteriormente, como especialista de Control de Gestión dentro del Área de Finanzas España. Agradezco a todo el personal del BBVA que me ha ayudado, pues su experiencia y asistencia ha sido de gran valor, así como agradecer el permitirme que utilizara recursos y bases de datos del BBVA. A todas ellas deseo manifestar mi más sincera gratitud.

Sin embargo, mi principal agradecimiento se lo dedico a la persona que me ha animado y motivado a culminar esta tesis que ha sido mi director de Tesis, Dr. Jorge V. Pérez-Rodríguez, su consejo, apoyo y dirección ha sido esencial para terminar este trabajo. Su constante estímulo y orientación me han permitido ir consiguiendo progresos en este proyecto. Su dedicación y sugerencias, junto a su rigor intelectual y su contagioso entusiasmo, han constituido la mejor fuente de estímulo para la culminación de esta Tesis. También ha sido de gran valor en relación con la revisión de la literatura, así como la metodología econométrica utilizada para estimar la eficiencia de las sucursales del Banco analizado.

A mi familia, mi marido Nicolás y mis hijas, Esther María y Laura Inés, por su amor, su paciencia y por todas las horas que no les he podido dedicar durante muchos días de muchos meses y años. Sé que esos momentos que no se repetirán, pero trataré de compensarlos una vez conseguida esta meta personal que me he propuesto. A mi marido por su especial y continuo apoyo tanto por el ánimo, la confianza incondicional depositada en mí para culminar este reto como por las horas brindadas a nuestras hijas

permitiéndome dedicar tiempo al estudio, sin él no hubiese sido posible terminar este proyecto.

A mis padres por sus sabios consejos, a veces era suficiente con solo recibir una llamada para entrar en paz con mi conciencia y así dar con la manera de encauzar todas las dificultades que se me han presentado. Me enseñaron a ser perseverante, confiar en que todo es posible para alcanzar lo que me proponía, y a no rendirme para conseguir los sueños. A mi hermana Beatriz, por su espíritu de mejora, capacidad de escuchar a los demás y sus palabras de motivación en los momentos difíciles, cuando se hacía arduo compaginar el mundo profesional con el ámbito familiar y las aspiraciones personales.

Tengo un pensamiento especial para mis abuelos y mi ahijada Sara María que me cuidan desde un lugar mejor y que se sentirían muy orgullosos de mí.

A mis amigos por su acompañamiento en este camino dando lugar a buenos momentos vividos permitiéndome liberar tensión y estrés.

“Nada te turbe, nada te espante todo se pasa, Dios no se muda, la paciencia todo lo alcanza, quien a Dios tiene nada le falta sólo Dios basta.”

*Santa Teresa*

# Resumen

El objetivo principal de esta Tesis Doctoral ha sido analizar la eficiencia en la producción y costes de las sucursales bancarias en España desde diferentes ángulos. En primer lugar, detectar las ramas totalmente eficientes e ineficientes, así como las supereficientes, y evaluar cuáles podrían ser los factores explicativos de las eficiencias estimadas. En segundo lugar, estudiar los cambios a lo largo del tiempo en la productividad de las sucursales. Y tercero, evaluar si las eficiencias de costes varían con el tiempo y cuáles pueden ser sus factores explicativos.

Para ello, hemos utilizado una muestra de sucursales bancarias de un gran banco comercial español durante el periodo 2011-2014. Dicho banco es uno de los de mayor volumen de recursos y está catalogado como uno de los más importantes y eficientes del mundo, a pesar de la crisis sistémica que lo afectó.

Las medidas de eficiencia estudiadas en esta Tesis han sido la eficiencia técnica, de escala y de costes, así como el cambio de productividad en el tiempo. Para ello, se han utilizado diversas metodologías como son el enfoque no paramétrico para fronteras completas usando DEA y FDH, pero también, los métodos no paramétricos de fronteras parciales (robustos) como *order- $\alpha$* ; los métodos de análisis de frontera estocástica paramétrica (SFA) para estimar la eficiencia de costes en un marco de datos de panel; y el método DEA-Malmquist para estimar el crecimiento de la productividad de las oficinas durante el periodo afectado por la crisis de 2012 y las reformas regulatorias del sistema bancario español.

Las principales características y resultados del análisis empírico realizado se pueden resumir de la siguiente manera:

### ***Evaluación de las sucursales bancarias mediante métodos no paramétricos***

Utilizamos un procedimiento de dos etapas para evaluar los determinantes de la eficiencia en las dimensiones operativa y de beneficios, como indicadores del desempeño de los gerentes de las sucursales, antes y después del rescate del sistema bancario español en 2012. Este procedimiento se basa tanto en el uso de métodos de frontera no paramétricos y no paramétricos robustos, como en la detección de valores atípicos (sucursales supereficientes) basados en los cálculos obtenidos mediante el uso de la frontera parcial siguiendo a Daraio y Simar. Este último método, que permite detectar posibles valores atípicos en los datos en una primera fase, facilita después un análisis de eficiencia no paramétrico clásico. En la primera etapa, las eficiencias se obtienen a partir de los datos anuales suministrados por las sucursales bancarias correspondientes al período 2011-2014. En general, se obtuvieron puntuaciones altas de las eficiencias operativas y de beneficios usando los estimadores DEA (eficiencias técnicas puras y de escala), FDH y estimadores *order- $\alpha$* . También, se obtuvieron las siguientes jerarquías en las puntuaciones:  $DEA < FDH < order-\alpha$ , lo que indica la preferibilidad de las estimaciones no paramétricas robustas al eliminar las observaciones atípicas. En la segunda etapa de nuestro análisis, se aplicó el método de Simar y Wilson usando un modelo econométrico censurado para identificar los factores explicativos de la eficiencia. Los resultados revelan la existencia de causalidad entre la eficiencia y algunas variables de interés, que podrían contribuir a una mejor comprensión del desempeño de las sucursales. Por ejemplo, algunas métricas propias de desempeño del banco, como la productividad basada en varlor, están relacionadas con la eficiencia técnica (pura).

### ***Análisis de la frontera estástica de costes de las sucursales bancarias***

Analizamos la relación entre los métodos de evaluación del desempeño utilizados por los bancos para evaluar sus sucursales y la correspondiente eficiencia de costes variable en el tiempo. Para hacer esto, empleamos un marco de datos de panel y consideramos la heterogeneidad latente de las sucursales bancarias, que podría surgir de problemas de gestión no sistemáticos no observados. Nuestro análisis utiliza datos mensuales obtenidos de las sucursales de un gran banco comercial que opera en España durante el período 2013–2014. Los resultados indican que existe heterogeneidad no observada, la eficiencia de costes de las sucursales bancarias es variable en el tiempo, y las puntuaciones de ineficiencia son bajas, entre 1 y 15 %. Además, encontramos que las eficiencias de costes variables en el tiempo se asocian positivamente con las medidas del volumen de negocios (la productividad basada en el valor alcanzado por la sucursal) y negativamente con la superficie de la sucursal y con el tiempo. Sin embargo, no se encontró relación entre el balance y el estado de pérdidas y ganancias (es decir, desviaciones negativas de las metas de ingresos brutos).

### ***Crecimiento de la productividad de las oficinas bancarias***

Analizamos el efecto de la reestructuración financiera y la crisis en la evolución del crecimiento de la productividad de las sucursales bancarias. Con ese fin, utilizamos el marco DEA-Malmquist en los enfoques de producción o beneficios. Utilizamos datos de los balances anuales de un importante banco comercial español para el período 2011-2014, que comprende el rescate del sistema bancario español y la segunda recesión de la economía española. Los principales resultados son tres. En primer lugar, a pesar de los duros ajustes y la desfavorable coyuntura económica, la productividad de las oficinas creció en el conjunto del período. En segundo lugar, el cambio de productividad se debió principalmente a un cambio tecnológico, más que a un cambio de eficiencia, lo



que destaca el papel significativo de la innovación implementada por el banco durante este período. Y tercero, hubo una contracción en la tasa de crecimiento de la productividad y sus componentes, y un aumento en el número de sucursales que experimentaron cambios negativos, especialmente para las sucursales no supereficientes.

**Palabras clave:** Industria bancaria, análisis de frontera determinista DEA y FDH, frontera parcial (*order- $\alpha$* ), frontera estocástica o SFA, eficiencia técnica y de costes, economías de escala, productividad total de los factores, cambio tecnológico, oficinas ineficientes, eficientes y supereficientes.

# Abstract

The main objective of this Doctoral Thesis has been to analyze the efficiency in the production and costs of bank branches in Spain from different angles. First, to detect the fully efficient and inefficient branches, as well as the super-efficient ones, and to evaluate what could be the explanatory factors of the estimated efficiencies. Second, to study changes over time in branch productivity. And third, to assess whether cost efficiencies vary over time and what their explanatory factors may be.

To do this, we have used a sample of bank branches of a large Spanish commercial bank during the period 2011-2014. It is one of the largest in volume of resources which is classified as one of the most important and most efficient in the world, despite the systemic crisis that has affected it.

The efficiency measures studied in this Thesis have been technical, scale and cost efficiency, as well as the productivity change using various econometric methodologies such as the non-parametric DEA, FDH and partial (robust) nonparametric methods (e.g., *order- $\alpha$* ) for technical efficiencies; parametric stochastic frontier analysis methods (SFA) for cost efficiency in a panel data framework; and the productivity growth of the branches using DEA-Malmquist during the period affected by the 2012 crisis and the regulatory reforms of the Spanish banking system.

The main characteristics and results of the empirical analysis carried out can be summarized as follows:

## ***Assessment of bank branches using non-parametric methods***

We use a two-stage procedure used to evaluate determinants of operational efficiency and profit dimensions, as an indication of the performance of branchmanagers in a large Spanish commercial bank, before and after the bailout of the

Spanish banking system in 2012. This procedure is based on both the non-parametric and robust non-parametric frontier methods and on partial frontier-based outlier detection (following Daraio and Simar) which enables us to detect potential outliers in the data, after which a classical non-parametric efficiency analysis is performed. In the first stage, efficiencies are obtained from annual data supplied by bank branches regarding the period 2011–2014. High scores were obtained in operational and profit efficiencies for DEA (pure technical and scale efficiencies), FDH and *order- $\alpha$*  estimators. In general, the following hierarchies of scores were obtained:  $DEA < FDH < order-\alpha$ . In the second stage of our analysis, a censored econometric model approach was applied (following Simar and Wilson) to identify the drivers of efficiency. Finally, the implications of this study for policymakers are discussed.

#### ***Stochastic cost frontier analysis for bank branches***

We analyse the relationship between performance assessment methods used by banks to evaluate their branches and the corresponding time-varying cost efficiency. To do this, we employ a panel data framework and consider bank branches' latent heterogeneity, which might arise from unobserved non-systematic management problems. Our analysis is based on monthly data obtained from the branches of a large commercial bank operating in Spain during the period 2013–2014. The results indicate that there is unobserved heterogeneity and time-varying cost efficiency in the bank branches, and that inefficiency scores are low, at 1–15%. Time-varying cost efficiencies are positively associated with measures of the volume of business (the productivity based on value achieved by the branch) and negatively associated with the surface area of the branch and with time. However, no relation was found regarding balance sheet and profit and loss statements (i.e., negative deviations from gross income targets).

### ***Productivity growth for bank branches***

We analyse the effect of financial restructuring and crisis in the evolution of bank branches productivity growth. To that ends, we use the DEA-Malmquist framework in either the production or the profit approaches. We use data for a major Spanish commercial bank's annual balance sheets for the period 2011-2014, which comprises the bailout of the Spanish banking system and the second downturn of the Spanish economy. The main results are threefold. First, despite the harsh adjustments and the unfavourable economic situation, the productivity of the branches grew during the overall period. Second, the productivity change was mainly due to a technological change, rather than efficiency change, which highlights the significant role of the innovation implemented by the bank during this period. And third, there had a contraction in the productivity growth rate and its components, and an increase in the number of branches that experienced negative changes, especially for non-superefficient branches.

**Key words:** Banking industry, DEA and FDH deterministic frontier analysis, partial frontier (*order- $\alpha$* ), stochastic frontier or SFA, technical and cost efficiency, economies of scale, total factor productivity, technological change, inefficient, efficient and super-efficient branches.



# Índice de Contenidos

<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>1.1. MOTIVACIÓN.....</b>	<b>1</b>
<b>1.2. EL SISTEMA BANCARIO ESPAÑOL EN LA RECIENTE CRISIS FINANCIERA. UNA BREVE SINOPSIS .....</b>	<b>5</b>
1.2.1. SITUACIÓN ECONÓMICA ACTUAL EN ESPAÑA .....	5
1.2.2. SITUACIÓN ECONÓMICA ACTUAL DE LAS ENTIDADES FINANCIERAS .....	6
1.2.3. OBJETIVOS, METODOLOGÍA Y DATOS DE LA TESIS DOCTORAL.....	14
<b>1.3. ESTRUCTURA DE LA TESIS DOCTORAL .....</b>	<b>16</b>
<b>2. CONTEXTO ANALÍTICO - CONCEPTOS .....</b>	<b>19</b>
<b>2.1. INTRODUCCIÓN .....</b>	<b>19</b>
<b>2.2. TEORÍA DE LA PRODUCCIÓN.....</b>	<b>20</b>
2.2.1. AXIOMAS BÁSICOS DE LA PRODUCCIÓN .....	24
<b>2.3. EFICIENCIA.....</b>	<b>26</b>
2.3.1. LA EFICIENCIA TÉCNICA.....	26
2.3.2. MEDIDAS DE EFICIENCIA TÉCNICA DE DEBREU-FARRELL .....	27
2.3.3. FUNCIÓN DISTANCIA, EFICIENCIA TÉCNICA Y ORIENTACIONES .....	29
2.3.4. EFICIENCIA TÉCNICA Y FUNCIÓN DE PRODUCCIÓN .....	30
<b>2.4. EFICIENCIA DE COSTES (O ECONÓMICA) Y ASIGNATIVA .....</b>	<b>32</b>
2.4.1. EFICIENCIA DE COSTES .....	32
2.4.2. EFICIENCIA ASIGNATIVA .....	33
2.4.3. EFICIENCIA DE COSTES Y FUNCIÓN DE COSTES .....	36
2.4.4. ILUSTRACIÓN DE LA EFICIENCIA ECONÓMICA, ASIGNATIVA Y TÉCNICA.....	37
<b>2.5. LA FUNCIÓN DE COSTES.....</b>	<b>40</b>
2.5.1. FUNCIÓN DE COSTES TRANSLOG .....	40
2.5.2. DEMANDA DE FACTORES.....	43
<b>2.6. METAFRONTERAS DE PRODUCCIÓN.....</b>	<b>46</b>

2.6.1. METAFRONTAS.....	48
2.6.2. FRONTERAS ESPECÍFICAS DE GRUPO .....	49
2.6.3. EFICIENCIA TÉCNICA Y RATIO DE GAP TECNOLÓGICO.....	51
<b>2.7. MÉTODOS USADOS EN EL ANÁLISIS DE LA EFICIENCIA .....</b>	<b>53</b>
<b>3. MEDICIÓN DE LA EFICIENCIA – FRONTERAS DETERMINISTAS PARAMÉTRICAS Y NO PARAMÉTRICAS .....</b>	<b>57</b>
<b>3.1. INTRODUCCIÓN .....</b>	<b>57</b>
<b>3.2. MODELOS DE REGRESIÓN DE FRONTERA DETERMINISTAS .....</b>	<b>57</b>
<b>3.3. MÉTODOS NO PARAMÉTRICOS .....</b>	<b>63</b>
3.3.1. ANÁLISIS ENVOLVENTE DE DATOS (DEA) .....	63
3.3.2. EFICIENCIA TÉCNICA Y DE ESCALA .....	65
3.3.2.1. MODELOS DEA DE SUPEREFICIENCIA .....	70
3.3.2.2. MODELOS SBM Y SBM DE SUPEREFICIENCIA.....	72
3.3.2.3. TRATANDO OUTPUTS INDESEABLES EN DEA .....	74
3.3.2.4. DEA CON DATOS NEGATIVOS .....	77
3.3.3. EFICIENCIA DE COSTES Y ASIGNATIVA .....	79
3.3.3.1. UNA VARIACIÓN DEL MODELO DEA DE COSTES .....	82
3.3.4. CAMBIO DE PRODUCTIVIDAD EN EL TIEMPO .....	84
3.3.4.1. ÍNDICE DE PRODUCTIVIDAD DE MALMQUIST.....	85
3.3.4.2. ILUSTRACIÓN GRÁFICA .....	88
3.3.4.3. OBTENCIÓN DE LAS EFICIENCIAS .....	90
3.3.5. METAFRONTAS DE PRODUCCIÓN USANDO DEA .....	92
<b>3.4. RELAJANDO EL SUPUESTO DE CONVEXIDAD: <i>FREE DISPOSAL HULL</i> (FDH).....</b>	<b>93</b>
3.4.1. LA PRÁCTICA COMPUTACIONAL DE FDH.....	95
3.4.2. UNA ILUSTRACIÓN GRÁFICA .....	96
<b>3.5. MÉTODOS NO PARAMÉTRICOS ROBUSTOS.....</b>	<b>98</b>
3.5.1. ESTIMADORES <i>ORDER-M</i> .....	99
3.5.2. CAMBIO DE PRODUCTIVIDAD EN EL TIEMPO USANDO EL MÉTODO <i>ORDER-M</i> .....	102
3.5.3. ESTIMADORES <i>ORDER-ALPHA</i> .....	104
3.5.4. ILUSTRACIÓN GRÁFICA DE LOS ESTIMADORES.....	105
3.5.5. DETECCIÓN DE OUTLIERS BASADA EN LOS MODELOS DE FRONTERA PARCIAL ..	106

<b>3.6. DETERMINANTES DE LA EFICIENCIA. ¿CÓMO EXPLICAR LOS DIFERENCIALES DE EFICIENCIA?</b> .....	<b>109</b>
<b>4. MEDICIÓN DE LA EFICIENCIA – FRONTERAS ESTOCÁSTICAS.....</b>	<b>113</b>
<b>4.1. INTRODUCCIÓN .....</b>	<b>113</b>
<b>4.2. ANÁLISIS PARAMÉTRICO DE FRONTERAS ESTOCÁSTICAS (SFA).....</b>	<b>114</b>
4.2.1. ENFOQUE FRECUENTISTA .....	114
4.2.1.1. MODELOS DE FRONTERA ESTOCÁSTICA EN CORTE TRANSVERSAL.....	114
4.2.1.2. INDEPENDENCIA DE LOS ERRORES.....	122
4.2.1.3. DEPENDENCIA ENTRE LOS ERRORES.....	135
4.2.1.4. MODELOS DE FRONTERA ESTOCÁSTICA EN DATOS DE PANEL .....	143
4.2.1.5. MODELOS DE EFECTOS FIJOS Y ALEATORIOS .....	143
4.2.2. ENFOQUE BAYESIANO.....	160
4.2.2.1. MODELO BAYESIANO PARA CORTE TRANSVERSAL .....	161
4.2.2.2. MODELO BAYESIANO PARA PANEL DE DATOS.....	164
4.2.2.3. EFECTOS FIJOS .....	165
4.2.2.4. EFECTOS ALEATORIOS .....	166
<b>4.3. MODELOS DE FRONTERA ESTOCÁSTICA SEMIPARAMÉTRICOS Y NO PARAMÉTRICOS .</b> .....	<b>167</b>
4.3.1. ENFOQUE SEMIPARAMÉTRICO .....	167
4.3.2. ANÁLISIS ESTOCÁSTICO NO PARAMÉTRICO ENVOLVENTE DE DATOS (STONED) .....	170
<b>4.4. METAFRONTERAS .....</b>	<b>174</b>
4.4.1. FORMULACIÓN DEL MARCO TEÓRICO .....	175
4.4.2. ESTIMACIÓN .....	176
<b>4.5. UN MODELO DE FRONTERA DE PRODUCCIÓN ESTOCÁSTICA CON FUNCIÓN DE DISTANCIA.....</b>	<b>177</b>
<b>5. REVISIÓN DE LA LITERATURA SOBRE EFICIENCIA BANCARIA Y SUCURSALES .....</b>	<b>181</b>
<b>5.1. INTRODUCCIÓN .....</b>	<b>181</b>
<b>5.2. EFICIENCIA BANCARIA .....</b>	<b>182</b>
5.2.1. LA EFICIENCIA A TRAVÉS DE DISTINTAS TÉCNICAS.....	182



5.2.2. CAMBIOS DE PRODUCTIVIDAD .....	187
<b>5.3. DETERMINANTES DE LA EFICIENCIA .....</b>	<b>190</b>
5.3.1. DETERMINANTES DE LA VARIANZA.....	192
<b>5.4. ESTUDIOS SOBRE LA BANCA ESPAÑOLA Y COMPARACIONES INTERNACIONALES ..</b>	<b>194</b>
5.4.1. BANCA ESPAÑOLA.....	194
5.4.2. COMPARACIÓN INTERNACIONAL.....	195
5.4.2.1. ESTIMACIÓN DE UNA FRONTERA NACIONAL PARA CADA PAÍS .....	196
5.4.2.2. ESTIMACIÓN DE UNA FRONTERA COMÚN PARA VARIOS PAÍSES .....	198
<b>5.5. EFICIENCIA DE LAS SUCURSALES BANCARIAS.....</b>	<b>204</b>
<b>5.6. ESTUDIOS SOBRE SUCURSALES BANCARIAS USANDO MÉTODOS NO PARAMÉTRICOS ..</b>	<b>205</b>
.....	
<b>5.7. CAMBIO DE PRODUCTIVIDAD DE LAS SUCURSALES BANCARIAS .....</b>	<b>210</b>
<b>5.8. ESTUDIOS USANDO FRONTERAS ESTOCÁSTICAS .....</b>	<b>210</b>
<b>6. MODELO DE DIRECCIÓN, DATOS Y DIMENSIONES .....</b>	<b>215</b>
<b>6.1. INTRODUCCIÓN .....</b>	<b>215</b>
<b>6.2. MODELO DE PRODUCCIÓN/DIRECCIÓN DEL BANCO ANALIZADO.....</b>	<b>217</b>
6.2.1. TIPO DE NEGOCIO Y OBJETIVOS .....	217
6.2.2. LAS OFICINAS .....	218
6.2.3. PERSONAL .....	219
6.2.4. INFRAESTRUCTURAS .....	222
6.2.5. TRANSACCIONES Y SALDOS DE CLIENTES .....	225
6.2.6. INGRESOS .....	226
6.2.7. EVALUACIÓN DE LAS OFICINAS .....	228
<b>6.3. DATOS .....</b>	<b>230</b>
<b>6.4. DIMENSIONES DEL ANÁLISIS DE LA EFICIENCIA TÉCNICA EN LA RED DE OFICINAS</b>	<b>234</b>
6.4.1. ENFOQUE DE PRODUCCIÓN.....	236
6.4.2. ENFOQUE DE INTERMEDIACIÓN.....	237
6.4.3. ENFOQUE DE RENTABILIDAD.....	239

<b>6.5. DIMENSIONES EN EL ESTUDIO DE LA EFICIENCIA DE COSTES .....</b>	<b>239</b>
<b>7. ESTIMACIÓN NO PARAMÉTRICA DE LA EFICIENCIA TÉCNICA DE LAS SUCURSALES BANCARIAS.....</b>	<b>241</b>
<b>7.1. INTRODUCCIÓN .....</b>	<b>241</b>
<b>7.2. REVISIÓN DE LA LITERATURA .....</b>	<b>246</b>
<b>7.3. ESTIMADORES CLÁSICOS Y NO-PARAMÉTRICOS ROBUSTOS .....</b>	<b>251</b>
7.3.1. FUNCIONES DE FRONTERA COMPLETA ESTIMADAS POR MÉTODOS NO PARAMÉTRICOS .....	252
7.3.2. FUNCIONES DE FRONTERA PARCIAL ESTIMADAS POR MÉTODOS NO- PARAMÉTRICOS ROBUSTOS.....	254
7.3.3. DETECCIÓN DE VALORES ATÍPICOS BASADOS EN FRONTERAS PARCIALES .....	256
7.3.4. PROCEDIMIENTO DE SIMAR Y WILSON .....	257
<b>7.4. ANÁLISIS EMPÍRICO .....</b>	<b>258</b>
7.4.1. DATOS .....	258
7.4.2. MEDICIÓN DEL DESEMPEÑO DE LAS SUCURSALES: DIMENSIONES, INPUTS Y OUTPUTS .....	259
7.4.3. FACTORES AMBIENTALES.....	261
7.4.4. EFICIENCIAS ESTIMADAS.....	266
7.4.4.1. ESTIMACIONES DE VALORES ATÍPICOS BASADOS EN DEA-VRS, FDH, <i>ORDER-                     ALPHA</i> Y FRONTERA PARCIAL .....	266
7.4.4.2. ESTIMACIONES DE EFICIENCIA DE ESCALA .....	271
7.4.5. RELACIÓN ENTRE ESTIMACIONES DE EFICIENCIA Y FACTORES DE LA SUCURSAL BANCARIA.....	274
<b>8. EL GRADO DE CUMPLIMIENTO DE LAS OFICINAS Y LA EFICIENCIA DE COSTES. UNA APROXIMACIÓN DE FRONTERAS ESTOCÁSTICAS.....</b>	<b>281</b>
<b>8.1. INTRODUCCIÓN .....</b>	<b>281</b>
<b>8.2. REVISIÓN DE LA LITERATURA .....</b>	<b>284</b>
8.2.1. DESEMPEÑO Y EFICIENCIA DE SUCURSALES BANCARIAS .....	284
8.2.2. EFICIENCIAS VARIABLES EN EL TIEMPO DE LAS SUCURSALES BANCARIAS.....	285
8.2.2.1. EFICIENCIA VARIABLE EN EL TIEMPO EN DATOS DE PANEL.....	285

8.2.2.2. EFICIENCIA DE LAS SUCURSALES EN UN MARCO DE FRONTERA ESTOCÁSTICA.....	287
<b>8.3. LA HETEROGENEIDAD LATENTE Y EFICIENCIA DE COSTES VARIABLE EN EL TIEMPO DE LAS SUCURSALES BANCARIAS.....</b>	<b>290</b>
<b>8.4. RESULTADOS EMPÍRICOS .....</b>	<b>294</b>
8.4.1. DATOS .....	294
8.4.1.1. DOS MEDIDAS DE DESEMPEÑO PROPIAS PARA EVALUAR LAS SUCURSALES BANCARIAS .....	295
8.4.1.2. ESTADÍSTICOS DESCRIPTIVOS .....	296
8.4.2. ENFOQUE DE PRODUCCIÓN, RESTRICCIONES DE HOMOGENEIDAD Y EFICIENCIA DE COSTES .....	300
8.4.3. ESTIMACIONES DEL MODELO .....	302
<b>9. REESTRUCTURACIÓN FINANCIERA, CRISIS Y CRECIMIENTO DE LA PRODUCTIVIDAD EN LAS OFICINAS BANCARIAS .....</b>	<b>311</b>
<b>9.1. INTRODUCCIÓN .....</b>	<b>311</b>
<b>9.2. CAMBIOS DE PRODUCTIVIDAD EN EL TIEMPO: DEA-MALMQUIST.....</b>	<b>316</b>
<b>9.3. ANÁLISIS EMPÍRICO.....</b>	<b>321</b>
9.3.1. DATOS .....	321
9.3.2. DIMENSIONES DEL DESEMPEÑO DE LA SUCURSAL BANCARIA .....	323
9.3.3. PRODUCTIVIDAD TOTAL DE LOS FACTORES PARA TODAS LAS SUCURSALES EN EL PERÍODO 2011-2014 .....	324
9.3.4. EFECTOS DE LA CRISIS Y LA REESTRUCTURACIÓN FINANCIERA EN LAS SUCURSALES BANCARIAS .....	328
9.3.4.1. TODAS LAS SUCURSALES .....	328
9.3.4.2. SUCURSALES COMPLETAMENTE EFICIENTES E INEFICIENTES.....	333
<b>10. CONCLUSIONES .....</b>	<b>341</b>
<b>10.1. VALORACIÓN DE LA EFICIENCIA DE LAS SUCURSALES USANDO MÉTODOS NO PARAMÉTRICOS ROBUSTOS .....</b>	<b>343</b>
<b>10.2. EFICIENCIA DE COSTES USANDO MODELOS DE FRONTERAS ESTOCÁSTICAS.....</b>	<b>347</b>
<b>10.3. CAMBIOS DE PRODUCTIVIDAD DE LAS OFICINAS.....</b>	<b>350</b>

<b>10.4. FUTURAS LÍNEAS DE INVESTIGACIÓN .....</b>	<b>353</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>357</b>
<b>APÉNDICE I. ALGUNAS CÓPULAS ELÍPTICAS Y ARCHIMEDEANAS</b>	<b>395</b>

# Índice de Figuras

<i>FIGURA 2.1. CONJUNTOS DE TECNOLOGÍA, INPUTS Y OUTPUTS.</i> .....	22
<i>FIGURA 2.2. EFICIENCIA TÉCNICA CON FUNCIÓN DE PRODUCCIÓN DETERMINISTA.</i> .....	31
<i>FIGURA 2.3. MEDIDAS DE EFICIENCIA SEGÚN LA ORIENTACIÓN.</i> .....	38
<i>FIGURA 2.4. METAFRONTERAS.</i> .....	50
<i>FIGURA 3.1. EFICIENCIA TÉCNICA CON UNA FUNCIÓN DE PRODUCCIÓN DETERMINISTA.</i> .....	58
<i>FIGURA 3.2. FRONTERAS DE PRODUCCIÓN OBTENIDAS POR OLS, COLS Y MOLS.</i> .....	62
<i>FIGURA 3.3. FRONTERAS CRS Y VRS DEA INPUT ORIENTADAS (COELLI, 1996).</i> .....	68
<i>FIGURA 3.4. FRONTERAS DEA-CRS, VRS Y NIRS (COELLI, 1996).</i> .....	70
<i>FIGURA 3.5. EFICIENCIA DE COSTES EN EL MODELO DEA-CRS USANDO LA ISOCUANTA CONVEXA LINEAL A TROZOS Y LA LÍNEA DE ISOCOSTES (COELLI, 1996).</i> .....	82
<i>FIGURA 3.6. CAMBIO EN EFICIENCIA Y CAMBIO TÉCNICO SEGÚN ORIENTACIÓN-OUTPUT E INPUT.</i> ..	89
<i>FIGURA 3.7. CONJUNTO DE POSIBILIDADES DE PRODUCCIÓN EN CRS, VRS Y FDH (JI AND LEE, 2010).</i> .....	97
<i>FIGURA 3.8. MEDIDAS DE EFICIENCIA FDH EN LA ORIENTACIÓN-INPUT (JI AND LEE, 2010).</i> .....	97
<i>FIGURA 3.9. FRONTERAS DE POSIBILIDADES DE PRODUCCIÓN (ISOCUANTAS) ESTIMADAS NO PARAMÉTRICAMENTE (TAUCHMANN, 2012).</i> .....	105
<i>FIGURA 3.10. PRODUCCIÓN DE SALUD A NIVEL DE DISTRITO EN ALEMANIA EN 2004. ILUSTRACIÓN DEA, FDH, ORDER-M Y ORDER-A PARA EL CASO DE UN SOLO INPUT Y OUTPUT (TAUCHMANN, 2011).</i> .....	106
<i>FIGURA 3.11. PUNTOS DE DISCONTINUIDAD EN EL MODELO DE PRODUCCIÓN REGIONAL DE SALUD EN BAVARIA EN 2004 (TAUCHMANN, 2011).</i> .....	108
<i>FIGURA 4.1. DISTRIBUCIONES MARGINALES <math>f(\varepsilon)</math> Y <math>f(\xi)</math>, SEGÚN SEAN DE PRODUCCIÓN O COSTES, PARA DISTINTOS VALORES DE LOS PARÁMETROS.</i> .....	125

<i>FIGURA 4.2. FUNCIONES DE DENSIDAD CONJUNTAS PARA DIFERENTES VALORES DE LOS PARÁMETROS.....</i>	<i>140</i>
<i>FIGURA 6.1. UN MODELO DE EFICIENCIA DE LA RED DE OFICINAS BANCARIAS (YANG Y LIU, 2012). .....</i>	<i>216</i>
<i>FIGURA 8.1. TRAYECTORIA TEMPORAL DE LA MEDIANA SPLINE (CÚB.) DE LOS COSTES OBSERVADOS POR MES.....</i>	<i>299</i>
<i>FIGURA 8.2. TRAYECTORIA TEMPORAL DE LA MEDIANA SPLINE (CÚBICA) DE LA PBV Y LA DESVIACIÓN DEL MARGEN BRUTO POR MES.....</i>	<i>300</i>
<i>FIGURA 8.3. RENTABILIDAD EN EL MODELO TRE.....</i>	<i>306</i>
<i>FIGURA 9.1. GRÁFICOS DE DENSIDAD DEL ÍNDICE DE PRODUCTIVIDAD ESTIMADO DE MALMQUIST Y SUS COMPONENTES POR SUBPERÍODOS. ENFOQUE DE PRODUCCIÓN.....</i>	<i>331</i>
<i>FIGURA 9.2. GRÁFICOS DE DENSIDAD DEL ÍNDICE DE PRODUCTIVIDAD ESTIMADO DE MALMQUIST Y SUS COMPONENTES POR SUBPERÍODOS. ENFOQUE DE RENTABILIDAD.....</i>	<i>332</i>
<i>FIGURA 9.3. GRÁFICAS DE DENSIDAD DEL ÍNDICE DE PRODUCTIVIDAD ESTIMADO DE MALMQUIST Y SUS COMPONENTES POR SUBPERÍODOS. ENFOQUE DE PRODUCCIÓN.....</i>	<i>337</i>
<i>FIGURA 9.4. GRÁFICOS DE DENSIDAD DEL ÍNDICE DE PRODUCTIVIDAD ESTIMADO DE MALMQUIST Y SUS COMPONENTES POR SUBPERÍODOS. ENFOQUE DE RENTABILIDAD.....</i>	<i>338</i>

# Índice de Tablas

<i>TABLA 2.1. FUNCIONES DE PRODUCCIÓN.....</i>	<i>23</i>
<i>TABLA 2.2 MÉTODOS USADOS EN LOS MODELOS DE FRONTERAS EFICIENTES. ....</i>	<i>54</i>
<i>TABLA 6.1. DIMENSIÓN DE PRODUCCIÓN.....</i>	<i>237</i>
<i>TABLA 6.2. DIMENSIÓN DE INTERMEDIACIÓN.....</i>	<i>238</i>
<i>TABLA 6.3. DIMENSIÓN DE RENTABILIDAD.....</i>	<i>239</i>
<i>TABLA 7.1. ESTADÍSTICOS DESCRIPTIVOS PARA TODAS LAS SUCURSALES DURANTE EL PERÍODO 2011-2014.....</i>	<i>265</i>
<i>TABLA 7.2. EFICIENCIA POR MÉTODOS NO-PARAMÉTRICOS AND NO-PARAMÉTRICOS ROBUSTOS CONSIDERANDO LAS DIMENSIONES DE PRODUCCIÓN Y RENTABILIDAD.....</i>	<i>270</i>
<i>TABLA 7.3. ESTIMACIONES DE EFICIENCIA DE ESCALA.....</i>	<i>272</i>
<i>TABLA 7.4. FACTORES AMBIENTALES POR DIMENSIÓN PARA EFICIENCIAS VRS Y FDH. PROCEDIMIENTO DE SIMAR Y WILSON.....</i>	<i>277</i>
<i>TABLA 7.5. EFICIENCIAS DE ESCALA (ESCALA-DEA) Y FACTORES AMBIENTALES POR DIMENSIÓN. PROCEDIMIENTO DE SIMAR Y WILSON.....</i>	<i>278</i>
<i>TABLA 8.1. ESTADÍSTICOS DESCRIPTIVOS DE LAS SUCURSALES BANCARIAS ANALIZADAS. PERÍODO ENERO 2013 – DICIEMBRE 2014.....</i>	<i>298</i>
<i>TABLA 8.2. ESTIMACIONES DE DATOS DE PANEL DEL MODELO DE FRONTERA ESTOCÁSTICA DE COSTES TRANSLOGARÍTMICAS (LA INEFICIENCIA SIGUE UNA DISTRIBUCIÓN NORMAL TRUNCADA).....</i>	<i>305</i>
<i>TABLA 9.1. DESCRIPTIVOS DE TODAS LAS SUCURSALES ANALIZADAS EN EL PERIODO 2011- 2014.....</i>	<i>322</i>

*TABLA 9.2. ÍNDICE DE PRODUCTIVIDAD DEA-MALMQUIST ELIMINANDO LOS VALORES ATÍPICOS Y SUS COMPONENTES POR SUBPERÍODOS Y PERÍODO GENERAL. RESUMEN DE MEDIAS ANUALES. .... 326*

*TABLA 9.3. ÍNDICE DE PRODUCTIVIDAD DEA-MALMQUIST ELIMINANDO LOS VALORES ATÍPICOS Y SUS COMPONENTES POR SUBPERÍODOS Y PERÍODO GENERAL. RESUMEN DE MEDIAS ANUALES. .... 336*





# CAPÍTULO 1

## INTRODUCCIÓN

---

### 1.1. MOTIVACIÓN

El análisis de la eficiencia ha sido un tema importante en la investigación económica desde hace más de 50 años. De hecho, el campo de análisis de la eficiencia es diverso, abarcando muchas disciplinas académicas como, por ejemplo, la agricultura, la organización industrial, la ingeniería y la economía. Pero, es en nuestra disciplina, la economía, donde la utilización de técnicas de análisis de eficiencia ha cubierto prácticamente a todos los sectores: la salud, el turismo, transporte (aeropuertos, puertos, ferrocarril), energético (abastecimiento de agua, electricidad), pero también bancos y sucursales bancarias, entre muchos otros.

La banca es una de las industrias más complejas del mundo y contribuye en gran medida a la riqueza de un país (por ejemplo, en el Reino Unido aproximadamente el 25% del PIB es producido por su sector de servicios financieros). El sector bancario juega un papel importante para la transmisión de capital de los agentes económicos con excedentes de ahorro a los agentes que necesitan financiación. En una perspectiva macroeconómica, la eficiencia del sector bancario en su provisión de servicios de intermediación financiera es importante para la asignación global de recursos. Además, la eficiencia del sector bancario desempeña un papel para el mecanismo de transmisión monetaria y para la estabilidad del sistema financiero, de tal manera que la eficiencia financiera afecta al crecimiento bancario, y las insolvencias bancarias pueden dar lugar a crisis sistémicas con consecuencias adversas para el conjunto de la economía.

Los bancos ofrecen una amplia gama de productos y servicios que van desde simples cuentas corrientes, cuentas a plazo, planes de jubilación, fondos mutuales, hipotecas, préstamos al consumo y muchos otros. El mecanismo a través del cual los bancos manejan estas transacciones es la red de sucursales que sirve como contacto principal con los clientes actuales y potenciales.

A pesar de la rápida expansión del uso de Internet en la banca y de muchos otros canales de transacción disponibles, es a través de una sucursal que los clientes realizan un gran porcentaje de sus actividades bancarias, incluyendo hipotecas, préstamos, cuentas de inversión y corretaje de valores, por nombrar sólo unos pocos. Las sucursales tienen una variedad de tamaños, y ofrecen diversos servicios a diferentes clientes mientras operan en diferentes regiones e incluso países. Por todo ello, la medición del rendimiento de las sucursales bancarias es una tarea muy difícil.

El objetivo principal de esta Tesis Doctoral se centra en el estudio de la eficiencia de las sucursales bancarias (oficinas) de la banca española.

La razón por la que analizaremos la eficiencia de la red de oficinas bancarias españolas es triple.

En primer lugar, el análisis de la eficiencia de las oficinas es importante porque permite:

1. Identificar aquellas oficinas problemáticas a los directores de sucursales, destacando las fortalezas y debilidades de sus operaciones y minimizando las

ineficiencias en las operaciones bancarias, enfatizando las desviaciones a la unidad de "buenas prácticas" para cualquier escala o combinación de productos.<sup>1</sup>

2. Realizar evaluaciones y comparaciones con los objetivos del banco. Las puntuaciones de eficiencia obtenidas por cada oficina pueden compararse con las medidas que utiliza el propio banco para cada oficina a nivel de objetivos o desviaciones presupuestarias.
3. Realizar recomendaciones de política y gestión sobre las oficinas: tamaño óptimo de la sucursal, tamaño del mercado, relación con los ingresos / margen bruto, cobertura de vacantes en relación al ahorro de costes, entre otros.

En segundo lugar, porque existe escasa o nula investigación sobre la eficiencia de las sucursales en la industria bancaria española, ya sea mediante modelado no paramétrico del Análisis Envolvente de Datos (DEA) o *Free Disposal Hull* (FDH) y sus extensiones, o utilizando la metodología de Análisis de Fronteras Estocásticas (SFA) paramétrica.<sup>2</sup> Por ejemplo, resulta curioso comprobar cómo, a pesar de que existe un extenso cuerpo teórico y empírico de literatura internacional sobre el estudio de la eficiencia de las sucursales bancarias (véase Fethi y Pasiouras, 2010; Paradi y Zhu, 2013; para una visión general), para nuestro conocimiento, Lovell y Pastor (1997) es el único estudio que analiza la eficiencia de las sucursales bancarias en el marco de los

---

<sup>1</sup> Berger et al. (1997) indicaron que la información sobre el desempeño de las sucursales puede mejorar nuestra comprensión de los fundamentos de la eficiencia a nivel de los bancos y ayudar a resolver algunos problemas de medición en el análisis estándar de los bancos. Paradi y Zhu (2013) se refirieron a otros aspectos, como el hecho de que la red de sucursales de un banco suele representar su mayor fuente de gastos operacionales. Además, es a través de las sucursales donde los clientes llevan a cabo una gran proporción de sus tareas bancarias.

<sup>2</sup> Sin embargo, a nivel de bancos existen muchos estudios sobre la economía española que han analizado la eficiencia de su sistema bancario globalmente. Por ejemplo, usando DEA destacan Pastor (1995), Grifell y Lovell, (1997), Pastor et al. (2002), Maudos y Pastor (2003), Guzmán y Reverte (2008), Färe et al. (1997), Orea (2000), Maudos y otros (2002), Cuesta y Orea (2002) (2010) y San-José et al. (2014), mientras que en SFA destacan Maudos (1996), Maudos (2001), Maudos y Pastor (2001a, 2001b), y Orea y Kumbhakar (2004).

modelos DEA, utilizando datos de una gran institución financiera en España. Estos autores examinaron el desempeño del procedimiento de fijación de objetivos para las sucursales empleado para evaluar su práctica operativa.<sup>3</sup>

Y, en tercer lugar, la oficina bancaria muestra directamente el efecto de la crisis reciente de 2012. Debemos reconocer que el sistema bancario español ha sido acreditado como uno de los más sólidos y mejor equipados entre todas las economías occidentales para hacer frente a la crisis de liquidez mundial, gracias a las normas y prácticas bancarias conservadoras del país (por ejemplo, se requiere que los bancos tengan elevadas provisiones de capital y exijan condiciones a los prestatarios). Sin embargo, esta práctica se relajó durante la burbuja inmobiliaria, ya que el regulador (Banco de España) no supo anticipar una fase tan recesiva, tan intensa, provocada por el crecimiento del crédito a la vivienda y a las empresas vinculadas al sector inmobiliario. Así, pues, en el entorno de transformación financiera, acentuado por la reciente crisis de 2008 y el rescate del sistema bancario español de 2012 que ha supuesto un importante proceso de reestructuración del sistema y provocado que las entidades reduzcan su capacidad instalada, ya situada en un 30%. Según el Banco de España hemos pasado de las 44.362 oficinas de 2009 hasta las 31.817 a finales del 2014 y, se ha seguido reduciendo llegando a 24.855 a finales del 2019. Este reajuste se aprecia no solo en el número de entidades, en las redes de distribución que se han reducido de manera muy relevante para poder mejorar la eficiencia sino en el número de empleados que ha

---

<sup>3</sup> Hasta la fecha no se ha publicado ninguna investigación sobre medidas de eficiencia técnica, de escala, de coste y de asignación, de cambios de productividad total de factores y de evaluación de relaciones (descriptivas y estadísticamente) entre los factores relevantes para las oficinas bancarias en España.

supuesto una caída de un tercio de su pantalla, según datos de la Fundación de las Cajas de Ahorro (Funcas).

A continuación, comentaremos brevemente la situación del sistema bancario español, para luego comentar los objetivos principales y específicos que pretendemos abordar con este trabajo y, finalmente, describir la estructura de la Tesis Doctoral.

## **1.2. EL SISTEMA BANCARIO ESPAÑOL EN LA RECIENTE CRISIS**

### **FINANCIERA. UNA BREVE SINOPSIS**

#### **1.2.1. SITUACIÓN ECONÓMICA ACTUAL EN ESPAÑA**

Hasta el año 2008 España vivió uno de los periodos más prósperos de su historia desde 1975. Las razones que impulsaron este periodo de auge económico fueron la incorporación de España a la Unión Monetaria y la llegada masiva de inmigrantes atraídos por la creciente demanda de empleo en el sector de la construcción y de algunos servicios básicos como la hostelería y el servicio doméstico. Sin embargo, tras la caída de Lehman Brothers en 2007, la situación económica fue empeorando paulatinamente hasta la coyuntura del 2016. Esta crisis se enmarca dentro de la denominada Gran Recesión que afectó a la mayor parte de países del mundo, en especial a los países desarrollados. A partir de 2008, la crisis financiera se recrudeció y se extendieron los graves problemas de solvencia de las entidades financieras internacionales, que parecían haberse concentrado en la banca de inversión estadounidense y en bancos comerciales excesivamente orientados a fórmulas de titulización muy complejas y a través de vehículos especiales. La quiebra de Lehman Brothers y el rescate de la aseguradora AIG (American International Group) generaron un grave deterioro en la situación de incertidumbre y de crisis de confianza.

A partir de septiembre de 2008, la crisis financiera internacional se recrudeció y se extendieron los graves problemas de solvencia de las entidades financieras internacionales. Las primas de riesgo de los bonos corporativos experimentaron también fuertes crecimientos, y las condiciones crediticias se endurecieron adicionalmente. En este contexto de turbulencias en los mercados de capitales, se produjo un incremento generalizado de los costes de la intermediación financiera que tiene por sí mismo un efecto negativo sobre la actividad, similar a los que se derivarían de un descenso de la productividad, ya que, al aumentar el coste de uso del capital, se reduce la inversión productiva.

Las consecuencias de una crisis económica mundial que se vió aumentada por sus efectos sobre la economía española por el estallido de la burbuja inmobiliaria y el endeudamiento excesivo del sector privado.

### **1.2.2. SITUACIÓN ECONÓMICA ACTUAL DE LAS ENTIDADES FINANCIERAS**

En España, además de una crisis económica se había producido una crisis financiera, una crisis de la deuda soberana, una crisis inmobiliaria, una crisis bancaria y una crisis crediticia. Esta crisis económica ha comprometido la estabilidad de las entidades financieras, lo cual ha provocado fusiones, ha obligado a bloquear el reintegro de las inversiones de determinados fondos inmobiliarios y ha convertido a los bancos y cajas de ahorro en improvisadas agencias inmobiliarias.

En 2010, España disponía de una red bancaria de 45.000 sucursales siendo la cuarta más grande del mundo después de China (120.420), USA (100.130) e India (66.663). Si comparamos el número de oficinas bancarias con la población, el sector bancario

español era uno de los más bancarizados del mundo, con cerca de 1.000 oficinas por cada millón de habitantes, ocupando los primeros lugares en el mundo con 948 oficinas. Duplicaba la densidad de Alemania y triplicaba la de USA. Esto se debía entre otras razones a que ha primado el contacto directo con el cliente a través de las oficinas bancarias y a que el modelo de negocio bancario en España es más universal que el modelo anglosajón. Sin embargo, la diferencia de la ratio de oficinas bancarias por habitante en relación con otros países es ahora sustancialmente más baja que antes de la crisis. La crisis ha sido un factor que ha acelerado la evolución natural que, como ocurre en los demás países europeos, implica menos oficinas por número de habitantes. El cierre de oficinas bancarias en España ha supuesto, en definitiva, un ajuste de capacidad para corregir desequilibrios y mejorar su eficiencia.

Entre 2008 y 2017, según los datos publicados por el Banco de España, el número de entidades bancarias en España ha disminuido un 28% y las oficinas bancarias han reducido su número en un 39%. En el año 2017 el número de oficinas bancarias operativas se redujo en 937, incluyendo sucursales de bancos, cajas de ahorros y cooperativas de crédito. La variación más significativa se produjo en la primera parte del año, contabilizando 833 oficinas menos. El número ha descendido de forma constante desde 2008, aunque no todo han sido cierres, sino que se han registrado algunas aperturas para reforzar la red en determinadas zonas geográficas. De esta forma, el número de oficinas operativas de entidades de depósito a finales de 2017 se sitúa en 27.706. Los bancos tradicionales y las cajas de ahorros son los que más han ajustado su red desde 2008, mientras que las cooperativas de crédito apenas han alterado su número en todo el tiempo que duró la crisis. El cierre de oficinas no ha afectado de forma lineal a la población española, ya que ha sido más acusado en municipios de más de 10.000 habitantes, donde la concentración de sucursales y cajeros era inicialmente mayor.



Además, este proceso de transformación que ha motivado una reducción significativa del número de empleados y de oficinas, ha venido acompañado de un aumento del número de empleados por cada oficina generando sucursales de mayor tamaño con más gestores especializados.

En el contexto de la profunda crisis económica que hemos pasado, el sistema financiero ha iniciado una estrategia de reducción de costes que incluye el cierre de oficinas. No se sospechaba hace unos años que los tipos bajarían al 0%, tras desaparecer más de 50 entidades bancarias en España, casi todas cajas de ahorros, y nacionalizar otras, parecía que en 2015 la banca volvería a recuperar lustre, sin embargo, la recuperación de la cuenta resultados ha sido menor de la esperada. El principal problema al que se enfrentan los bancos es la presión sobre su margen de interés, núcleo de la actividad tradicional bancaria y de donde procede el grueso de sus ingresos. Por lo que, las recomendaciones del BCE y del Banco de España al sector financiero son la de continuar con la reducción de costes, para buscar mejorar la rentabilidad y los beneficios como consecuencia de esa caída de márgenes.

A diciembre de 2019, la banca española cuenta con un total de 23.851 oficinas, lo que supone un 47,8% menos que las sucursales operativas que había en 2008, cuando las entidades financieras sumaban 45.707 locales, el número más alto de la historia del sector (desde que el Banco de España elabora estas estadísticas). En 2019 la banca española sigue aligerando la red de sucursales recortando 2.160 oficinas, superando la cifra del año anterior que supuso 1.309 cierres. Y es que, ante la dificultad que supone para el sector bancario generar ingresos en un entorno de tipos negativos que se mantienen en mínimos históricos, una de las fórmulas que han venido aplicando las entidades ha sido recortar en gastos operativos a través del cierre de oficinas. Todo ello,

unido a la transformación digital del negocio, ha provocado que la estrategia de muchos bancos se haya basado en clausurar dos sucursales que se encontraban en localizaciones cercanas para abrir macrooficinas con mayores prestaciones en materia de asesoramiento y atención al cliente y con horarios más amplios que las oficinas tradicionales.

La banca superviviente de la crisis se podía decir que afronta seis retos principales e interrelacionados:

### **1. El desafío a los modelos de negocio tradicionales**

Actualmente, la banca está experimentando un cambio significativo y los modelos de negocio bancario tradicionales están en transformación. Debido a los constantes retos, por aumentar la rentabilidad, garantizar su participación en el mercado y mitigar riesgos, los bancos deben evaluar sus propias capacidades y evaluar sus fortalezas y debilidades. La digitalización y la disrupción de nuevos participantes del mercado y competidores, como las empresas Fintech con un coste de capital inferior al de los Bancos, y con exigencias regulatorias y de capitales menores, están reduciendo la cuota de participación en el mercado a las Instituciones financieras.

- Erosión de márgenes

El debilitado crecimiento macroeconómico de las economías maduras ha instaurado un periodo de tipos de interés muy bajos, cuya tendencia es que se prolongue en el tiempo. Ello ha erosionado el margen de intermediación de todo el sector bancario, primer y más importante escalón en su cuenta de resultados.

Adicionalmente, el efecto de la crisis financiera supuso un impacto en la economía productiva elevando significativamente la ratio de morosidad y las provisiones.

- Desarrollo del *Shadow Banking*. Competidores no tradicionales

La crisis financiera ha debilitado la confianza de los clientes en sus bancos. Aprovechando esa brecha donde los consumidores no vieron cubiertas sus necesidades, se han consolidado unos gigantes tecnológicos como PayPal, Amazon, Facebook, Google, entre otros y, han florecido una gran cantidad de compañías tecnológicas de menor tamaño, Fintech.

Estas startups utilizan la tecnología financiera como medio para abaratar costes y simplificar procesos, dando como resultado un servicio eficiente en internet para el usuario, a la vez que generan menores comisiones en relación con los sistemas tradicionales, ganando cuota en las líneas de negocios más rentables de la banca manual. Basan su fortaleza en la tecnología aplicada a ofrecer respuestas a los usuarios de productos financieros y se mantienen alejados de la costosa y compleja regulación bancaria, lo que les permite tener bajos costes y ser más ágiles y eficientes.

## **2. El aumento de la presión regulatoria**

Como consecuencia de pérdidas derivadas de esta crisis, y los millones que tuvieron que absorber los contribuyentes, se estableció una ola de regulación, implementadas posteriormente a la crisis, para devolver la confianza y seguridad de los depositantes y el mercado.

Los nuevos requerimientos regulatorios han tenido un impacto directo en las actividades bancarias y sus modelos de negocio. Antes de la crisis, los niveles de rentabilidad de los bancos se incrementaron por el alto grado de apalancamiento y por mantener un fondeo mayorista relativamente de bajo coste, y en algunos casos, realizar

actividades de alto riesgo, como ser los préstamos inmobiliarios y sus titulaciones, para poder aumentar sus ganancias. Los cambios en el comportamiento bancario y en el marco regulatorio han hecho que algunas de las estrategias bancarias más rentables sean menos viables.

Las mayores exigencias, al requerir capital de alta calidad, activos líquidos, fuentes de financiamientos más estables y mayor solvencia pronostican que el negocio bancario futuro será distinto al conocido hasta hoy.

La regulación impacta doblemente en la rentabilidad:

- Exige un desapalancamiento, lo que mina la rentabilidad sobre recursos y limita las posibilidades de crecimiento.

- Exige un incremento de costes operativos, principalmente en tecnología y recursos humanos. Así, como ciertas líneas de negocio, por ejemplo, las actividades de negociación haciendo que algunos bancos reduzcan su participación en este tipo de actividades.

### **3. La pérdida de confianza por la sociedad**

La sociedad tiene la percepción de que la banca tradicional cometió errores y no ejecutó con diligencia su disposición social en los años anteriores a la crisis, que se fundamenta básicamente, en promover el crecimiento económico y social, proteger el patrimonio de ahorradores e inversores y fomentar la estabilidad financiera y la solvencia del sistema financiero.

La banca puede ampliar su papel en la sociedad, y una buena oportunidad son los clientes no bancarizados. Una ventaja, es la experiencia en el sector financiero, que no

mantienen los competidores no tradicionales, y se puede lograr a través de la recuperación de la confianza del sector financiero.

#### **4. Innovaciones disruptivas**

Mientras la banca se esfuerza en una carrera para evolucionar su pesada infraestructura tecnológica, siguen apareciendo, a un ritmo sin precedentes, avances disruptivos como la Realidad Aumentada, la Computación Cognitiva, la Computación Cuántica, la Inteligencia Artificial (que puede llevar el asesoramiento financiero a una nueva dimensión) y la Biometría.

La banca a futuro tiene diferentes oportunidades que pueden ser utilizadas para crear ese valor que se necesita para atraer a las nuevas generaciones. A continuación, se muestran diferentes herramientas tecnológicas que se pueden aplicar para lograr dicho objetivo:

- a) **Data sin estructura:** esto se refiere a la información que no tiene un modelo de datos predefinido, sin embargo, los bancos poseen abundantes reservas de este tipo de data. Este tipo de data se usa en la vida diaria de los clientes por ejemplo en mensajes de texto, búsquedas en la web, imágenes, documentos que contienen información legal, tales como los certificados de nacimiento y de matrimonio, y los contratos por ejemplo escrituras de vivienda. Algunos bancos han logrado aprovecharlo a través de minería de datos y analítica avanzada y han obtenido una visión de mucho valor sobre el comportamiento de sus clientes, preferencias y situaciones que han impactado su vida y por ende su comportamiento crediticio.

- b) **Automatización:** actividades repetitivas, serán realizadas por robots y por algoritmos automatizados. Esto ha aumentado su valor ya que se ha probado que la inclusión en algunos bancos aumenta el índice de eficiencia debido a una reducción de mano de obra que impacta en los costes de la institución y ofrece agilidad al cliente al reducir el tiempo de respuesta.
- c) **Nueva Infraestructura de Pagos:** los pagos digitales a través del móvil u online dan a los bancos la habilidad para poder realizar varias actividades para incrementar rentabilidad y eficiencia.
- d) **Tecnología *Blockchain*:** estructura de datos en la que la información contenida se agrupa en conjuntos (bloques).
- e) **Internet de las cosas (IoT):** es un concepto que se refiere a una interconexión digital de objetos cotidianos con internet. Constituye un cambio radical en la calidad de vida de las personas en la sociedad, ofrece una gran cantidad de nuevas oportunidades de acceso a datos, servicios específicos en la educación, seguridad, asistencia sanitaria y en el transporte, entre otros campos.

## **5. La construcción de un nuevo modelo de relación con el cliente**

El modelo de relación con el Cliente basado en la oficina se está transformando porque ha cambiado el comportamiento de éstos. Una mejor experiencia de usuario, un nuevo modelo de relación, atención digital y aumentar los niveles de seguridad son las tendencias en el sector.

## **6. El cambio en la estructura del mercado**

La crisis ha potenciado la concentración bancaria. A pesar de que la presencia de entidades sistémicas representa un problema potencial, la crisis ha alentado la consolidación y concentración de muchos sectores financieros, alimentando el debate sobre los efectos de una mayor concentración sobre el nivel de competencia y la estabilidad del sistema financiero.

La concentración de entidades persigue la obtención de sinergias y economías de alcance que mejoren la eficacia de la intermediación financiera. Un mayor tamaño de los grupos resultantes debería hacer más fácil la consecución de aumentos adicionales en el grado de eficiencia, y facilitar el acceso a los mercados de capitales.

Sin embargo, la ola de fusiones y operaciones corporativas que hemos vivido durante los últimos años supone una mayor concentración del mercado sobre el nivel de competencia en la industria bancaria, especialmente en mercados minoristas locales, a la vez que el mayor tamaño de determinados bancos domésticos generan preocupación sobre la futura estabilidad del sistema financiero.

Con todos estos retos, es de gran importancia el análisis de la eficiencia de las oficinas que conforman los bancos, así como determinar los principales factores.

### **1.2.3. OBJETIVOS, METODOLOGÍA Y DATOS DE LA TESIS DOCTORAL**

Esta tesis Doctoral tiene como principal objetivo analizar la eficiencia de las sucursales bancarias de un gran banco comercial español.

Los objetivos específicos de esta Tesis son tres. Primero, detectar las sucursales totalmente eficientes e ineficientes, así como las supereficientes y evaluar cuáles pueden

ser los factores explicativos de las eficiencias estimadas. Segundo, estudiar los cambios en el tiempo de la productividad de las sucursales. Y, tercero, evaluar si las eficiencias de coste varían en el tiempo y cuáles pueden ser sus factores explicativos. Estos análisis aún no han sido explorados para el caso de España en el nivel de las sucursales de un Banco.

La metodología utilizada para analizar la eficiencia de las sucursales atendiendo a los objetivos específicos comentados es triple.

Por un lado, para evaluar las ineficiencias de las sucursales usamos un procedimiento en dos etapas. Primero, se estiman las eficiencias operativas, así como las dimensiones de beneficios como una indicación del desempeño de los gerentes o directores de las sucursales y, segundo, se evalúan los determinantes de dichas eficiencias. El enfoque utilizado para estimar las eficiencias es no paramétrico de tipo DEA - que fue desarrollado originalmente por Charnes et al. (1978) y Banker et al. (1984) -, así como FDH. Usando DEA podemos estimar no solo las eficiencias técnicas, sino que las de escala. Pero, también, usamos métodos de fronteras parciales no paramétricos robustos (order- $\alpha$ ) que permiten distinguir entre oficinas supereficientes, totalmente eficientes e ineficientes. Este método también se utiliza para la detección de valores atípicos o comportamientos anómalos de las sucursales siguiendo el método propuesto por Daraio y Simar, después de lo cual se realiza un análisis clásico de eficiencia no paramétrica tipo DEA y FDH, respectivamente.

Por otro lado, también construimos el índice de Malmquist que se usa para determinar el cambio en la productividad total de los factores (TFP) en un marco DEA, el cual se divide en dos componentes: el cambio en la eficiencia técnica y el progreso tecnológico, respectivamente.



Finalmente, empleamos el enfoque paramétrico basado en la metodología de Análisis de Frontera Estocástica (SFA), cuyos modelos seminales fueron desarrollados por Aigner et al. (1977) y Meeusen y van de Broeck (1977), pero en el entorno de los modelos de panel de datos. Los modelos que emplearemos fueron propuestos por Greene (2005). Estos modelos permiten estimar las eficiencias de costes variables en el tiempo y son recientes en la literatura econométrica. En particular, estos modelos permiten evaluar la heterogeneidad no observada y la ineficiencia de las sucursales bancarias.

Los datos del estudio se obtuvieron de los balances anuales del banco para el período 2011-2014. Este periodo se caracteriza porque incluye el segundo periodo de recesión de la economía española (2011-2013) y tiene incluido el periodo donde se produce el rescate del sistema bancario español en junio de 2012.

### **1.3. ESTRUCTURA DE LA TESIS DOCTORAL**

La Tesis Doctoral está compuesta por diez capítulos divididos en dos partes. El Capítulo 1 es la introducción de esta Tesis Doctoral. A continuación, aparece la Parte 1, compuesta por cuatro capítulos, donde se describen el marco metodológico empleado para estimar las fronteras eficientes de producción y coste, así como una revisión de la literatura empírica sobre la eficiencia de bancos y sucursales bancarias. El Capítulo 2 explica los fundamentos teóricos y los conceptos clave en el análisis de la eficiencia, haciendo hincapié en la teoría de la producción. Los Capítulos 3 y 4 tratan de los métodos de estimación de fronteras deterministas y estocásticas, respectivamente, más ampliamente utilizados hasta el momento. Finalmente, el Capítulo 5 hace una revisión

de la literatura sobre eficiencia bancaria y de las sucursales, comentando algunos estudios que comparan internacionalmente los sistemas bancarios nacionales.

La segunda parte de la Tesis se ocupa íntegramente del análisis empírico, y consta de tres capítulos. El Capítulo 6 describe los datos utilizados, el modelo de producción/dirección del banco, y la metodología para evaluar el grado de cumplimiento de los objetivos de las oficinas usando medidas diseñadas por el Banco. Los Capítulos 7, 8 y 9 presentan los resultados de la estimación de las eficiencias usando distintos métodos. El Capítulo 7 compara las eficiencias obtenidas por los modelos DEA, FDH y de fronteras no paramétricas parciales usando los métodos no paramétricos robustos como order- $\alpha$ . Además, en este capítulo se analizan los determinantes de estas usando la metodología de Simar y Wilson. Cabe destacar que este capítulo se construye sobre el artículo publicado y titulado: “*Assessing branch efficiency and managerial behaviour in a large Spanish commercial bank*” (I. Cabrera-Suárez, J.V. Pérez-Rodríguez): *Revista Española de Financiación y Contabilidad – Spanish Journal of Finance and Accounting*, 2020, 49 (1), pp. 48-73, que pertenece al 4º cuartil (Q4) en el catálogo del *Journal Citation Report (JCR)* en su edición de *Social Science* y Q3 en *Scimago Journal Rank (SJR)*.

El Capítulo 8 versa sobre la modelización de frontera estocástica de costes analizando la relación entre algunos métodos de evaluación del desempeño de las sucursales utilizados por el Banco y la eficiencia de costes estimada que varía con el tiempo. Para hacer esto, se emplea un marco de datos de panel y consideramos la heterogeneidad latente de las sucursales bancarias, que podría surgir de problemas de administración no sistemáticos y también no observados. Este análisis utiliza datos mensuales obtenidos de las sucursales durante el período 2013-2014. El artículo derivado de este capítulo fue titulado: “*Bank branch performance and cost efficiency: A*

*stochastic frontier panel data approach*” (I. Cabrera-Suárez, J.V. Pérez-Rodríguez) y publicado en la revista *International Journal of Finance and Economics*, 2021, 26 (4), 5850-5863, que pertenece al Q2 en JCR-*Social Science* y en SJR en el año 2020.

El Capítulo 9 trata del análisis de los cambios de productividad en el tiempo, así como los componentes que la caracterizan. El artículo derivado de este capítulo está siendo evaluado en *Applied Economics*, que es Q2 en el JCR-*Social Science*. El artículo se titula: “*Productivity growth in bank branches during periods of regulatory change*”. Tanto el Capítulo 7 como 9 utilizan datos anuales en el periodo 2011-2014.

Finalmente, el Capítulo 10 presenta las conclusiones más importantes obtenidas de los tres objetivos específicos analizados y a partir de los resultados obtenidos de los modelos estimados.

# CAPÍTULO 2

## CONTEXTO ANALÍTICO - CONCEPTOS

---

### 2.1. INTRODUCCIÓN

La motivación teórica y los fundamentos conceptuales para la medición de la ineficiencia surgen con las aportaciones de Debreu (1951) y Farrell (1957).

Farrell (1957) sentó las bases para medir la eficiencia y productividad en el nivel microeconómico. Sus principales contribuciones consistieron en dos cuestiones: la definición de la eficiencia y la productividad, y el cálculo de la tecnología de referencia y las medidas de eficiencia. El supuesto fundamental es el de una asignación de input-output perfecta que permite operaciones ineficientes.

La ineficiencia se define como la distancia entre la producción real (observada) de una empresa individual (*Decision Making Unit* – DMU, o Unidad de Toma de Decisión, en español) a la función de producción frontera aceptada como punto de referencia (objetivo y no observado). La base de esta medida es la contracción o expansión radial de una observación a la frontera, es decir, la contracción radial de factores productivos (inputs) o expansión radial de productos (outputs) conectando puntos ineficientes observados con puntos de referencia en la frontera de producción. Si el punto de producción real de una empresa se encuentra en la frontera, ésta es totalmente eficiente. Si se encuentra por debajo de la frontera, entonces ésta es ineficiente. La relación entre la producción real y potencial define el nivel de eficiencia de la empresa.

Farrell propone dos componentes para definir la eficiencia: la eficiencia técnica y la eficiencia distributiva (o de asignación). La primera refleja la capacidad de una DMU para reducir al mínimo el uso de inputs con el fin de producir una cantidad dada de producto. La segunda, refleja la capacidad de una DMU para utilizar inputs en proporciones óptimas, dados sus precios respectivos y la tecnología de producción. Ambas medidas representan una medida total de eficiencia (Coelli et al., 1997).

La interpretación de los ratios de eficiencia puede hacerse como sigue. Los ratios de eficiencia toman un valor entre cero y uno, donde uno indica que la DMU es totalmente eficiente. Por ejemplo, una puntuación de eficiencia medida contra una frontera de costes del 90% significa que la DMU podría haber reducido los costes en un 10% sin alterar su vector de producción/producto.

A continuación, comentaremos brevemente algunos aspectos analíticos y conceptuales clave de la teoría de la producción y costes que serán empleados a lo largo de esta Tesis.

Conviene destacar que este capítulo se construye basándonos en definiciones e ideas que pueden encontrarse en algunos manuales de texto que explican el análisis de la eficiencia como, por ejemplo, Kumbhakar y Lovell (2000), Coelli, Prasada Rao, O'Donnell y Battese (2005) y Fried, Lovell y Schmidt (2008).

## **2.2. TEORÍA DE LA PRODUCCIÓN**

En el contexto del análisis de la eficiencia, la producción se considera como un proceso en el que los productores utilizan inputs (denotados por el vector de inputs,  $x$ ) para producir algún output (denotado por  $y$ ). Los productores usan los inputs para

obtener outputs usando alguna tecnología de producción, la cual puede describirse bien con conceptos de teoría de conjuntos o con el enfoque familiar de la función de producción.

### En relación con la teoría de conjuntos

Comenzaremos introduciendo el conjunto de outputs y de inputs junto al conjunto de la tecnología para alguna tecnología de producción.

En primer lugar, el conjunto de la tecnología  $T$  se define como el conjunto de producción factible  $(y, x)$  que puede producirse con cierta tecnología de producción específica de la unidad productiva bajo observación. Formalmente, podemos escribir:

$$T = \{(y, x) : x \text{ puede producir } y\}$$

Intuitivamente, el límite de este conjunto es la frontera de producción, que es la producción máxima producible para cualquier vector de inputs dado.

En segundo lugar, los conjuntos de inputs de la misma tecnología de producción se definen entonces como los conjuntos de vectores de inputs que son viables para cada elemento del vector de output  $y$ .

$$L(y) = \{x : (y, x) \in T\}$$

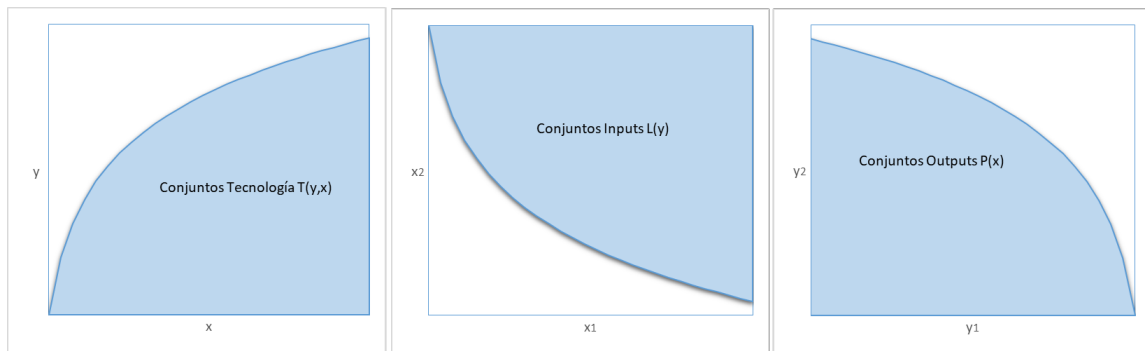
Del mismo modo, el límite de este conjunto forma las isocuantas de inputs para la tecnología de producción.

Finalmente, el conjunto de output se define como el conjunto de outputs factibles, para cada posible vector de input  $x$ .

$$P(x) = \{y : (y, x) \in T\}$$

La Figura 2.1 muestra los tres conjuntos definidos anteriormente. El conjunto de la tecnología y output es compacto mientras que el conjunto de inputs es cerrado, limitado por debajo por la isocuanta de inputs. Las propiedades generales de estos conjuntos son dadas en Coelli et al. (2005), y son omitidas aquí por cuestión de brevedad, dado que nuestro interés se centra en la ilustración posterior de las definiciones y propiedades de las eficiencias técnica, de costes y asignativa.

**Figura 2.1.** Conjuntos de tecnología, inputs y outputs.



### En relación con la función de producción

En contraste con la representación de conjuntos anterior, la tecnología también puede caracterizarse por una función de producción, la cual puede verse como representación paramétrica del proceso de producción para un productor dado. Esta representación, sin embargo, requiere que el proceso de producción sea de output único, o alternativamente que el vector de output pueda agregarse a un vector de output compuesto usando algunos pesos óptimos (Kumbhakar y Lovell, 2000). La función misma proporciona la relación entre inputs y outputs, con propiedades específicas que dependen de la forma funcional elegida.

Sea la función de producción general del tipo:

$$y_i = f(x),$$

donde  $y_i$  es el nivel de producción de la empresa  $i$ -ésima ( $i=1, \dots, n$ ),  $f$  es la función de producción,  $x = (x_1, \dots, x_k)$  es un vector de  $k$  inputs, siendo  $x_j$  el  $j$ -ésimo input,  $j=1, \dots, k$ . Por simplicidad, se omite el subíndice  $i$  representativo de la empresa.

Las formas funcionales para las funciones de producción,  $f$ , van desde las muy simples hasta las más complejas. La Tabla 2.1 recoge algunas expresiones de las funciones de producción en el análisis de la eficiencia (Coelli et al., 2005).

**Tabla 2.1.** Funciones de producción.

Función	Especificación ( $y_i = f(x)$ )
Lineal	$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}$
Cobb-Douglas	$y_i = \beta_0 + \prod_{j=1}^k x_{ij}^{\beta_j}$
Cuadrático	$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \frac{1}{2} \sum_{j=1}^k \sum_{l=1}^k \beta_{ijl} x_{ij} x_{il}$
Cuadrática normalizada	$y_i = \beta_0 + \sum_{j=1}^{k-1} \beta_{ij} \left( \frac{x_j}{x_k} \right) + \frac{1}{2} \sum_{j=1}^{k-1} \sum_{l=1}^{k-1} \beta_{ijl} \left( \frac{x_{ij}}{x_{ik}} \right) \left( \frac{x_{il}}{x_{ik}} \right)$
Translog	$y_i = \exp \left( \beta_0 + \sum_{j=1}^k \beta_j \log x_{ij} + \frac{1}{2} \sum_{j=1}^k \sum_{l=1}^k \beta_{ijl} \log x_{ij} \log x_{il} \right)$
Generalizada de Leontieff	$y_i = \sum_{j=1}^k \sum_{l=1}^k \beta_{ijl} (x_{ij} x_{il})^{0.5}$
Elasticidad constante de sustitución (CES)	$y_i = \beta_0 \left( \sum_{j=1}^k \beta_j x_{ij}^\gamma \right)^{1/\gamma}$

*Notas:* log indica el logaritmo natural.  $\beta_0, \beta_j$  ( $j=1, \dots, k$ ) y  $\gamma$  son parámetros de la tecnología. En concreto,  $\gamma$  es la elasticidad constante de sustitución.



Por ejemplo, si asumimos la función de producción log-lineal de Cobb-Douglas (véase la Tabla 2.1), originalmente desarrollada por Cobb y Douglas (1928), entonces el modelo de frontera de producción se escribe como:

$$\log y_i = \beta_0 + \sum_{j=1}^k \beta_j \log x_{ij}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, k,$$

donde  $\log y_i$  es el logaritmo natural de la producción de la  $i$ -ésima empresa;  $\log x_i$  es un vector  $k \times 1$  de transformaciones logarítmicas de las cantidades de input de la  $i$ -ésima empresa. La función de producción de Cobb-Douglas presenta algunas propiedades útiles para los fines de la investigación, como la interpretación práctica de los parámetros estimados y la inclusión sencilla del progreso técnico neutral de Hicks en las fronteras de producción estimadas.

### 2.2.1. AXIOMAS BÁSICOS DE LA PRODUCCIÓN

Los axiomas básicos de producción, siguiendo a Kumbhakar y Lovell (2000), pueden sintetizarse en:

- No negatividad. Este axioma indica que la función de producción transforma los inputs a outputs en el sentido que el output resultante es un número real finito y no negativo. Formalmente,  $f : \mathfrak{R}^k \rightarrow \mathfrak{R}^+$ .
- Esenciabilidad débil. Este axioma indica que para que haya producción estrictamente positiva se necesita al menos un input. Es decir,  $f(x) > 0$ , entonces el vector de inputs debe tener al menos un elemento estrictamente

positivo. Esencialidad fuerte exige que se utilicen todos los factores productivos para que la producción del output sea positiva.

- Monotonicidad. Este axioma indica que la producción es no decreciente en los inputs. Formalmente, si  $x_1 > x_2$  entonces  $f(x_1) > f(x_2)$ . Este axioma también implica que, para las funciones de producción continuamente diferenciables, los productos marginales son no negativos para todos los inputs. La monotonicidad puede ser fuerte o débil. La monotonicidad fuerte exige que los factores de producción que se incorporan a la producción posean una productividad marginal positiva. La versión débil permite productividades marginales positivas, pero, además, que una productividad marginal sea nula para alguno de los factores productivos.
- Concavidad. Este axioma indica que la función de producción debe ser cóncava en inputs. Es decir, para dos vectores de inputs  $x_i$  y  $x_j$ , y la ponderación  $\omega$ , sucede que:  $f(\omega x_i + (1-\omega)x_j) \geq \omega f(x_i) + (1-\omega)f(x_j)$ . Esta propiedad implica la ley de productividad marginal decreciente.

En general, estos axiomas garantizan que la producción se realiza de una manera económicamente factible (por ejemplo, "no se puede conseguir algo por nada" (esencialidad débil)). Así, puede decirse que las funciones de producción que cumplen estos axiomas se comportan bien en el contexto del análisis de eficiencia, es decir, si cumplen dichos axiomas, entonces se dice que la frontera estimada es axiomática.

Aún así, debemos tener en cuenta que incluso estos axiomas básicos pueden ser vulnerados en situaciones donde se consideran procesos de producción más complejos. Por ejemplo, la propiedad de monotonicidad puede ser fácilmente incumplida si hay congestión de inputs (*input congestion*), lo cual puede ocurrir en situaciones de la vida

real. Tales vulneraciones también pueden ocurrir si la definición del vector de inputs se amplía para incluir los llamados inputs malos (*bad inputs*) tales como emisiones de CO<sub>2</sub> u otros contaminantes.

## 2.3. EFICIENCIA

### 2.3.1. LA EFICIENCIA TÉCNICA

Una definición formal de la eficiencia técnica fue dada por Koopmans (1951):

- Un vector output-input  $(y, x) \in T$ , es *técnicamente eficiente* sí, y solo sí,  $(y^*, x^*) \notin T$  para  $(y^*, -x^*) \geq (y, -x)$ .

Dos casos especiales de la definición de Koopmans están relacionados con que la eficiencia puede analizarse desde dos tipos de orientaciones. Primero, la orientación-output que permite responder a la pregunta ¿cuántas cantidades de input podemos reducir proporcionalmente sin cambiar la cantidad producida de output? Es decir,

- Un vector de inputs,  $x \in L(y)$ , es *técnicamente eficiente* sí, y solo sí,  $x^* \notin L(y)$  para  $x^* \leq x$ . Por lo tanto, para un output dado, se define a un productor técnicamente eficiente como aquel cuya combinación de inputs coincide con la isocuanta, sin posibilidad de reducción de los inputs con el mismo nivel de output.

Y, segundo, la orientación-input que permite responder a la pregunta, ¿cuánta cantidad de output podemos incrementar proporcionalmente sin altear las cantidades de inputs utilizados? Es decir,

- Un vector de outputs,  $y \in P(x)$ , es *técnicamente eficiente* si, y solo si,  $y^* \notin P(x)$  para  $y^* \geq y$ . Por lo tanto, para un input dado, se define a un productor técnicamente eficiente como aquel cuya producción coincide con la frontera de producción, sin posibilidad de expansión del producto con el mismo nivel de inputs.

### **2.3.2. MEDIDAS DE EFICIENCIA TÉCNICA DE DEBREU-FARRELL**

Una definición menos restrictiva de la eficiencia técnica dada por Koopmans (1951), que se deriva de los trabajos de Debreu (1951) y Farrell (1957), es que una unidad de producción es técnicamente eficiente si para un determinado vector input-output no existe una reducción (contracción) de las cantidades de inputs o no se puede ampliar la cantidad de output que es factible manteniendo un nivel dado de producción e inputs, respectivamente. En otras palabras, la unidad de producción está asignando sus inputs de manera que su producción es máxima, es decir, la producción coincide con la verdadera frontera de producción.

Siguiendo a Kumbhakar y Lovell (2000, páginas 44-45), las medidas de Debreu-Farrell de eficiencia técnica (TE) (también denominadas de eficiencia radial) y sus propiedades son las siguientes:

### *Orientación-input*

La definición de la eficiencia técnica orientada hacia los inputs (input-orientada) es una función  $TE_I(y, x) = \min \{ \theta : \theta x \in L(y) \}$ , representando la contracción igualmente proporcional de todos los inputs, que satisface las siguientes propiedades:

- $TE_I(y, x) \leq 1$ , siendo  $TE_I(y, x) = 1$  si  $x \in L(y)$ .
- $TE_I(y, x)$  es no decreciente en  $x$ .
- $TE_I(y, x)$  es homogénea de grado -1 en  $x$ . Esto es, cuando se dobla la cantidad de inputs, la eficiencia técnica solo lo hace en la mitad.
- $TE_I(y, x)$  es invariante a las unidades de medida de  $(y, x)$ .

Otra definición de la eficiencia técnica con orientación input es que si solo se produce un solo output, entonces  $TE_I(y, x) = \min \{ \theta : y \in f(\theta x) \}$ . Así, la contracción del vector de inputs estará en la frontera.

### *Orientación-output*

La definición de la eficiencia técnica orientada hacia el output (output-orientada) es una función  $TE_O(x, y) = \max \{ \phi : \phi y \in P(x) \}$ , representando la expansión igualmente proporcional de todos los outputs, que satisface las siguientes propiedades:

- $TE_O(x, y) \leq 1$ , siendo  $TE_O(x, y) = 1$  si  $y \in P(x)$ .

- $TE_o(x, y)$  es no decreciente en  $y$ .
- $TE_o(x, y)$  es homogénea de grado +1 en  $y$ .
- $TE_o(x, y)$  es invariante a las unidades de medida de  $(x, y)$ .

Otra definición de la eficiencia técnica con orientación output es que si solo se produce un solo output, entonces  $TE_o(x, y) = \left[ \max \{ \phi : \phi y \in f(x) \} \right]^{-1}$ . Así, la ampliación del output estará en la frontera.

### 2.3.3. FUNCIÓN DISTANCIA, EFICIENCIA TÉCNICA Y ORIENTACIONES

Cuando una tecnología de producción usa múltiples inputs para producir múltiples outputs se utilizan funciones distancia, que denominaremos por  $D$ .

Las funciones distancia a las que aludiremos, y que por ejemplo serán empleadas en los índices de Malmquist dentro de la metodología DEA (Capítulo 3), pueden definirse atendiendo al tipo de orientación como:

- Orientación-input:  $D_I(y, x) = \max \{ \lambda : x/\lambda \in L(y) \}$ , siendo  $L(y)$  el conjunto de inputs de la tecnología de producción.  $y$  puede producirse con menor cantidad de inputs,  $x/\lambda$ ,  $\lambda > 1$ .
- Orientación-output:  $D_O(x, y) = \min \{ \mu : y/\mu \in P(x) \}$ . Definido en términos del conjunto de output,  $P(x)$ .  $y$  puede producirse con el vector de inputs,  $x$ , pero puede ser expandido hacia  $y/\mu$ ,  $\mu < 1$ .

Para no extendernos en este apartado, sus propiedades pueden consultarse en Kumbhakar y Lovell (2000, páginas 30-32).

La eficiencia técnica puede definirse, según sea la orientación al input o output, como:

- Orientación-input:  $TE_I(y, x) = \min \{ \theta : D_I(y, \theta x) \geq 1 \}$ .
- Orientación-output:  $TE_O(x, y) = \left[ \max \{ \phi : D_O(x, \phi y) \leq 1 \} \right]^{-1}$ . Es el recíproco de la eficiencia de Farrell (1957).

La relación entre la función distancia y la eficiencia radial es:

- $TE_I(y, x) = \left[ D_I(y, x) \right]^{-1}$ , es decir, la eficiencia técnica con orientación-input es igual al recíproco de la distancia con orientación-input.
- $TE_O(x, y) = D_O(x, y)$ , es decir, la eficiencia radial con orientación-output coincide con la distancia de orientación-output.

#### **2.3.4. EFICIENCIA TÉCNICA Y FUNCIÓN DE PRODUCCIÓN**

La eficiencia técnica puede construirse a partir de la ratio entre la producción observada y la función de producción frontera.

Sea la función de producción  $f(x_i)$ , de las definidas en la Tabla 2.1. Entonces, la eficiencia técnica ( $TE_i$ ) para el productor  $i$ -ésimo ( $i=1,2,\dots, n$ ) usando el vector de

inputs,  $x_i = (x_{i1}, \dots, x_{ik})'$  para producir  $y_i$ , es la relación entre los valores observados de la producción y la función de producción frontera ( $f(x_i)$ ), tal que:

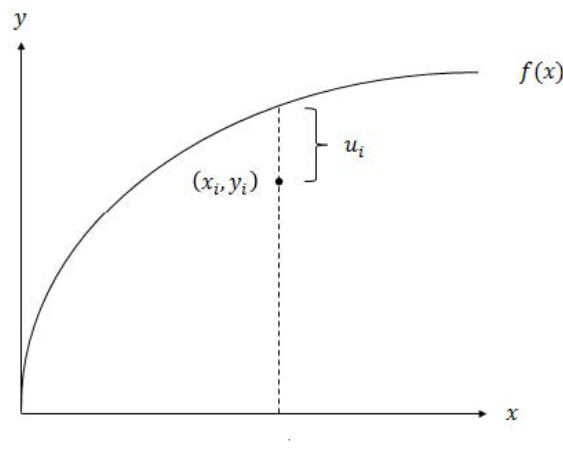
$$TE_i = \frac{y_i}{f(x_i)}$$

Esta ratio se encuentra entre 0 (totalmente ineficiente) y 1 (totalmente eficiente). A partir de dicha relación es fácil ver que:

$$y_i = f(x_i) \times TE_i$$

La Figura 2.2 ilustra el concepto de eficiencia técnica en un proceso de producción determinista con un sólo input ( $x$ ) y un sólo output ( $y$ ). La ineficiencia es la distancia que representa  $u_i$ .

**Figura 2.2.** Eficiencia técnica con función de producción determinista.



Conviene destacar que para estimar  $TE_i$ , la función  $f(x_i)$  puede aproximarse de forma paramétrica, es decir, incluyendo un vector de parámetros de la tecnología de producción, tal que:  $f(x_i; \beta)$ , siendo  $\beta$  dicho vector de parámetros, y estableciendo



supuestos sobre los errores del modelo a estimar; o de forma no paramétrica. Ambos métodos de estimación serán descritos en el Capítulo 4.

## 2.4. EFICIENCIA DE COSTES (O ECONÓMICA) Y ASIGNATIVA

La eficiencia económica global (o total) de una empresa posee dos componentes. La eficiencia técnica – es decir, aquella que refleja la habilidad de la empresa para obtener el máximo output dado el nivel de inputs empleado –, y la eficiencia asignativa – es decir, aquella que refleja la habilidad de la empresa para utilizar la combinación correcta de inputs dados los precios relativos de los mismos (es decir, usar los inputs en óptimas proporciones) –.

En esta sección, siguiendo a Kumhakar y Lovell (2000), definiremos algunos conceptos (definiciones y propiedades) relacionadas con la eficiencia de costes (CE) y la eficiencia asignativa (AE), respectivamente.

### 2.4.1. EFICIENCIA DE COSTES

Sea  $c(y, w)$  una función de costes que depende del nivel de output (o vector de outputs),  $y$ , pero también de los precios de los inputs,  $w$ .

Las propiedades de esta función son:

- No negatividad:  $c(y, w) \geq 0, y \geq 0$ .
- Homogénea de grado 1 en los precios de los inputs. Esto es,  $c(y, \lambda w) = \lambda c(y, w), \lambda > 0$ . Es decir, si los precios se doblan, los costes también.

- Concavidad y continuidad en  $w$ .
- Monotonicidad. Es decir,  $c(\lambda y, w) \leq c(y, w)$ ,  $0 \leq \lambda \leq 1$ .
- Convexidad en  $y$ .

La eficiencia de costes puede interpretarse como una medida que evalúa la capacidad de una empresa para producir outputs a un coste mínimo. Así, una medida de la eficiencia de costes es una función  $CE(y, x, w) = \frac{c(y, w)}{w'x}$  o ratio mínimo de costes respecto del observado, siendo  $w'$  un vector de precios de los inputs.

Las propiedades de la eficiencia de costes para  $x \in L(y)$ , son las siguientes:

- $0 < CE(y, x, w) \leq 1$ , siendo  $CE(y, x, w) = 1$  si  $c(y, w) = w'x$ .
- Homogénea de grado -1 en inputs, es decir,  $CE(y, \lambda x, w) = \lambda^{-1} CE(y, x, w)$ ,  $\lambda > 0$ . O sea, doblando los inputs, la eficiencia de costes crece en la mitad.
- No decreciente en output. Es decir,  $CE(\lambda y, x, w) \geq CE(y, x, w)$ ,  $\lambda \geq 1$ .
- Homogénea de grado cero en el precio de los inputs. Es decir,  $CE(y, x, \lambda w) = CE(y, x, w)$ ,  $\lambda > 0$ . O sea, doblar todos los precios no tiene efecto sobre la eficiencia de costes.

#### 2.4.2. EFICIENCIA ASIGNATIVA

La eficiencia asignativa, que también servirá para construir la eficiencia de costes conocida la eficiencia técnica, refleja la capacidad de la empresa de utilizar los inputs en proporciones óptimas, dados los precios de sus inputs.

La eficiencia asignativa puede construirse según sea la orientación. Así, podemos distinguir la eficiencia input-orientada y output-orientada.

### *Orientación-input*

La definición de la eficiencia asignativa en su orientación-input puede escribirse como:

$$AE_I(y, x, w) = CE(y, x, w) / TE_I(y, x)$$

Las propiedades de esta medida de eficiencia son:

- $0 < AE_I(y, x, w) \leq 1$ . Si  $AE_I(y, x, w) = 1$ , entonces  $CE(y, x, w) = TE_I(y, x)$ .
- Homogénea de grado cero en inputs, es decir,  
 $AE_I(y, \lambda x, w) = AE_I(y, x, w)$ ,  $\lambda > 0$ .
- Homogénea de grado cero en precio de los inputs. Es decir,  
 $AE_I(y, x, \lambda w) = AE_I(y, x, w)$ ,  $\lambda > 0$ .

Teniendo en cuenta la relación:  $AE_I(y, x, w) = CE(y, x, w) / TE_I(y, x)$ , la eficiencia de costes puede descomponerse fácilmente entre eficiencia técnica y eficiencia asignativa. Así,

$$CE(y, x, w) = TE_I(y, x) \times AE_I(y, x, w)$$

De esta forma, una condición necesaria y suficiente para que  $CE(y, x, w) = 1 \Leftrightarrow TE_I(y, x, w) = AE_I(y, x, w) = 1$ . Es decir, un productor alcanza la

eficiencia de costes máxima cuando éste es eficiente tanto técnicamente como asignativamente.

Conviene destacar que la eficiencia asignativa también puede definirse un su orientación-output, para lo cual es necesario definir la eficiencia de los ingresos (RE). Para ello, se asumirá que los productores determinan el precio del output  $p$  y buscan maximizar sus ingresos:  $p'y$ . La eficiencia de los ingresos es igual a:

$$RE(x, y, p) = \frac{p'y}{r(x, p)}$$

siendo  $r(x, p)$  la frontera de los ingresos (no negativa, convexa, monótona, homogénea de grado +1 en precios). Para un  $y \in P(x)$ , la eficiencia de los ingresos es una medida que está comprendida entre 0 y 1 ( $0 < RE(x, y, p) \leq 1$ ), es homogénea de grado +1 en output ( $RE(x, \lambda y, p) = \lambda RE(x, y, p)$ ,  $\lambda > 0$ ), es homogénea de grado cero en precios ( $RE(x, y, \lambda p) = RE(x, y, p)$ ,  $\lambda > 0$ ), y no es decreciente en inputs ( $RE(\lambda x, y, p) \leq RE(x, y, p)$ ,  $\lambda \geq 1$ ).

### *Orientación-output*

Una medida de eficiencia asignativa output-orientada es igual a:

$$AE_o(x, y, p) = RE(x, y, p) / TE_o(x, y)$$

cuyas propiedades son que:

$$0 < AE_o(x, y, p) \leq 1$$

- Homogénea de grado cero en outputs, es decir,

$$AE_o(x, \lambda y, p) = AE_o(x, y, p), \lambda > 0.$$

- Homogénea de grado cero en precio de los outputs. Es decir,

$$AE_o(x, y, \lambda p) = AE_o(x, y, p), \lambda > 0.$$

También, la medida de eficiencia asignativa output-orientada puede descomponerse como:

$$RE(x, y, p) = TE_o(x, y) \times AE_o(x, y, p)$$

Luego, también ocurrirá que  $RE(x, y, p) = 1$  cuando  $TE_o(x, y) = AE_o(x, y, p) = 1$ .

Es decir, un productor alcanza la eficiencia de ingresos máxima cuando es técnicamente eficiente y asignativamente eficiente produciendo la correcta combinación de outputs.

### 2.4.3. EFICIENCIA DE COSTES Y FUNCIÓN DE COSTES

Al igual que hicimos con la eficiencia técnica, podemos obtener la expresión de la eficiencia de costes ( $CE(y, x, w)$ ) y asignativa ( $AE(y, x, w)$ ) para cada productor comparando la función de costes frontera con los costes observados.

Por ejemplo, la expresión para obtener la eficiencia de costes para el  $i$ -ésimo productor,  $i = 1, 2, \dots, n$ , puede calcularse a través de la expresión,  $CE_i = CE(y_i, x_i, w_i)$ ,  $i = 1, 2, \dots, n$ , tal que:

$$CE_i = \frac{c(y_i, w_i)}{c_i}, i = 1, 2, \dots, n$$

donde  $c(y_i, w_i)$  es el coste óptimo (mínimo) o frontera de costes común a todos los productores y  $c_i = \sum_{j=1}^k w_{ij}x_{ij} = w'_i x_i$  es el coste observado del  $i$ -ésimo productor, siendo  $w'_i$  un vector fila  $1 \times k$  de precios de los factores.  $c_i \geq c(y_i, w_i)$ ,  $i = 1, 2, \dots, n$  y  $CE_i \leq 1$ .

En general, la expresión para obtener la eficiencia asignativa ( $AE_i$ ) para el  $i$ -ésimo productor,  $i = 1, 2, \dots, n$ , es:

$$AE_i = CE_i / TE_i, i = 1, 2, \dots, n,$$

es decir, el cociente entre la eficiencia de costes y la eficiencia técnica.

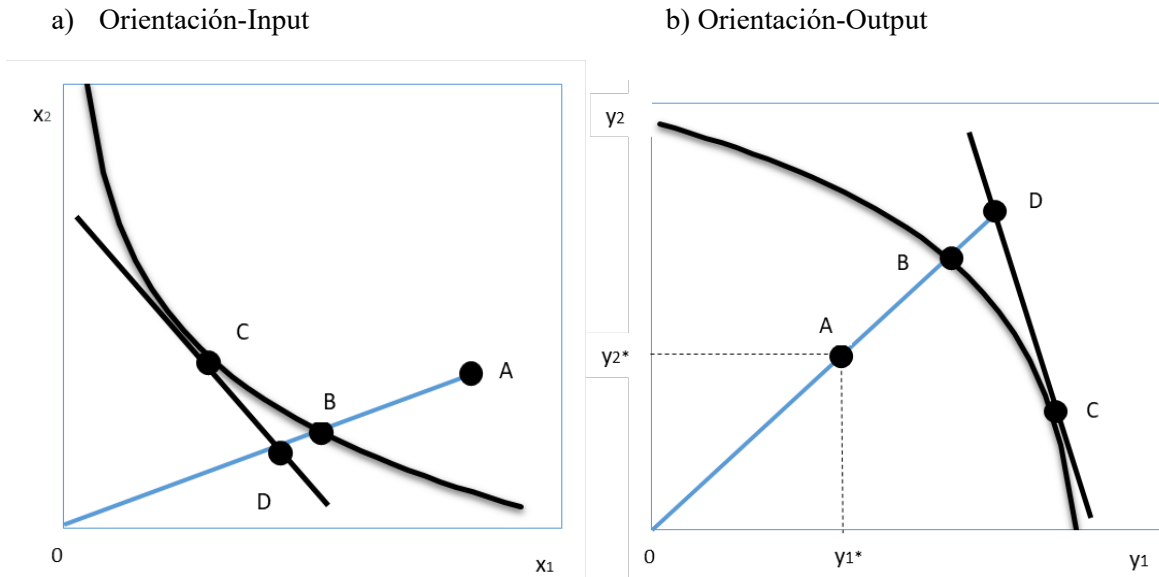
Los métodos de estimación de las expresiones de eficiencia vistas en esta Sección serán descritos en los Capítulos 3 y 4.

#### 2.4.4. ILUSTRACIÓN DE LA EFICIENCIA ECONÓMICA, ASIGNATIVA Y TÉCNICA

Los conceptos de eficiencia técnica, de costes y asignativa pueden ilustrarse gráficamente en la Figura 2.3 usando, como ejemplo, un proceso de producción de dos outputs ( $y_1, y_2$ ) que utiliza dos inputs ( $x_1, x_2$ ).

Como sabemos, la eficiencia puede analizarse en términos de la orientación-input (combinación óptima de inputs para lograr un nivel dado de output), o la orientación-output (el output óptimo que se podría producir dado un conjunto de inputs). En la Figura 2.3 se distinguen ambas situaciones.

**Figura 2.3.** Medidas de eficiencia según la orientación.



encuentra en la isoquanta asociada con el nivel mínimo de inputs requerido para producir  $(y_1^*, y_2^*)$  (es decir,  $\text{Iso}(y_1^*, y_2^*)$ ). La eficiencia técnica input-orientada es igual a:

$$TE_I(y, x) = OB/OA$$

Sin embargo, la combinación menos costosa de inputs que produce  $(y_1, y_2)$  viene dada por el punto C (es decir, el punto donde la tasa marginal de sustitución técnica es igual a la relación de precios de inputs  $w_2/w_1$ ). Para lograr el mismo nivel de coste (es decir, el gasto en inputs), los inputs tendrían que ser reducidos al punto D. Por lo tanto, la eficiencia de costes es igual a:

$$CE(y, x, w) = OD/OA$$

De esta forma, una vez ha sido calculada la eficiencia de costes y la eficiencia técnica, la eficiencia asignativa input-orientada es igual a:

$$AE_I(y, x, w) = CE(y, x, w)/TE_I(y, x) = OD/OB$$

En cuanto a la orientación-output, la Figura 2.3(b) muestra la frontera de posibilidades de producción para un conjunto dado de input. Como se muestra en esta figura, la producción de la empresa, que produce en el punto A, se puede expandir radialmente al punto B si los inputs empleados por la empresa son utilizados eficientemente. Por lo tanto, la medida de eficiencia técnica con orientación-output es igual a:

$$TE_o(x, y) = OA/OB$$

Mientras que el punto B es técnicamente eficiente, en el sentido de que se sitúa en la frontera de posibilidades de producción, podría obtenerse un mayor ingreso produciendo en el punto C (el punto donde la tasa marginal de transformación es igual a la relación de precios  $p_2/p_1$ ). En este caso, se debe producir más de  $y_1$  y menos de  $y_2$  con el fin de maximizar los ingresos. Para obtener el mismo nivel de ingresos que en el punto C, manteniendo la misma combinación de inputs y productos, la producción de la empresa tendría que ampliarse hasta el punto D. Por lo tanto, la eficiencia de los ingresos es igual a:

$$RE(x, y, p) = OA/OD$$

Así pues, una vez ha sido calculada la eficiencia técnica y la eficiencia de ingresos, la eficiencia asignativa con orientación-output es igual a:

$$AE_o(x, y, p) = RE(x, y, p)/TE_o(x, y) = OB/OD$$



## 2.5. LA FUNCIÓN DE COSTES

En esta sección centramos la exposición en una función de costes paramétrica y determinista que, en general, tiene la forma  $c(y_i, w_i; \beta)$ ,  $i = 1, 2, \dots, n$ , siendo  $\beta$  el vector de parámetros desconocidos que representan la tecnología de costes.<sup>4</sup>

### 2.5.1. FUNCIÓN DE COSTES TRANSLOG

Sea  $w$  un vector de precios de los  $k$  inputs utilizados en la producción de  $y$ . Sea  $c$  el coste total, cuya función puede escribirse como  $c(y, w)$ . Una representación muy usada para  $c(y, w)$  es la función de costes logarítmica trascendental no homotética (o translog), que anida la función de Cobb-Douglas. La función translog puede ser vista como una aproximación de Taylor en los logaritmos de una función de costes arbitraria.<sup>5</sup>

Siguiendo a Berndt (1991, págs. 469 y ss.), la función no homotética translog, siendo  $w_j$  el precio del  $j$ -ésimo input, puede escribirse como (omitimos el subíndice  $i=1, \dots, n$ ):

---

<sup>4</sup> Cabe destacar que en la Sección 2 de este Capítulo se han definido algunas funciones de producción determinista y paramétricas.

<sup>5</sup> Las funciones no homotéticas son muy generales, ya que los cocientes (ratios) de coste/demanda de inputs que son mínimos pueden depender de un nivel general de output, en contraposición a lo que suponen las funciones homotéticas, ya que, en este caso, las demandas de inputs son independientes del nivel general de output.

$$\log c = \log \alpha_0 + \sum_{j=1}^k \alpha_j \log w_j + \frac{1}{2} \sum_{j=1}^k \sum_{l=1}^k \gamma_{jl} \log w_j \log w_l + \alpha_y \log y + \frac{1}{2} \gamma_{yy} (\log y)^2 + \sum_{j=1}^k \gamma_{jy} \log w_j \log y$$

En la anterior ecuación pueden imponerse una serie de restricciones generales de interés. Estas restricciones pueden ser:

1) Restricción de simetría:  $\gamma_{jl} = \gamma_{lj}$ . Por ejemplo, si existiesen dos factores de producción: capital (K) y trabajo (L), la expresión del doble sumatorio sería equivalente a:

2)

$$\begin{aligned} \frac{1}{2} \sum_{j=1}^2 \sum_{l=1}^2 \gamma_{jl} \log w_j \log w_l &= \frac{1}{2} \left( \gamma_{11} (\log w_1)^2 + \gamma_{22} (\log w_2)^2 + \gamma_{12} \log w_1 \log w_2 + \gamma_{21} \log w_2 \log w_1 \right) \\ &= \frac{1}{2} \left( \gamma_{11} (\log w_1)^2 + \gamma_{22} (\log w_2)^2 + 2\gamma_{12} \log w_1 \log w_2 \right) \\ &= \frac{1}{2} \gamma_{11} (\log w_1)^2 + \frac{1}{2} \gamma_{22} (\log w_2)^2 + \gamma_{12} \log w_1 \log w_2 \end{aligned}$$

3) Restricción de homogeneidad de grado uno en precios. Para que una función de costes se comporte adecuadamente, entre otras cosas debe cumplir el ser homogénea de grado uno en los precios, dado  $y$ . Esto implica las siguientes restricciones sobre la ecuación anterior:

$$\sum_{j=1}^k \alpha_j = 1, \quad \sum_{j=1}^k \gamma_{jl} = \sum_{j=1}^k \gamma_{lj} = \sum_{j=1}^k \gamma_{jy} = 0$$

Otras restricciones adicionales que pueden introducirse en el modelo son:

a) Función homotética. Es necesario y suficiente que  $\gamma_{jy} = 0, \forall j = 1, \dots, k$ . Esta condición implica que las demandas de factores son independientes del output.

- b) Función homogénea de grado constante en el output. Es necesario que la función sea homotética y que además,  $\gamma_{yy} = 0$ . El grado de homogeneidad es  $1/\alpha_y$ .
- c) Rendimientos a escala constantes de la función de producción dual. Esto ocurre cuando la función de producción es homotética, homogénea de grado uno en precios  $y$ , además,  $\alpha_y = 1$ . Esto significa que los costes medios no dependen del output.
- d) Función de Cobb-Douglas de rendimientos constantes a escala. Esto ocurre cuando añadimos la restricción  $\gamma_{jl} = 0 \forall j, l = 1, \dots, k$  junto a todas las anteriores restricciones (homotecidad, homogeneidad, rendimientos constantes a escala). Esta función puede escribirse como:

$$\log c = \log \alpha_0 + \sum_{j=1}^k \alpha_j \log w_j + \alpha_y \log y$$

Para ilustrar con un sencillo ejemplo algunas de las anteriores expresiones, consideremos que  $k=4$ , y que estos cuatro factores son Capital ( $K$ ), Trabajo ( $L$ ), Energía ( $E$ ) e inputs intermedios no energéticos ( $M$ ). Con estas definiciones, la ecuación translog con restricciones de simetría puede escribirse como:

$$\begin{aligned} \log c = & \log \alpha_0 + \alpha_K \log w_K + \alpha_L \log w_L + \alpha_E \log w_E + \alpha_M \log w_M \\ & + \gamma_{KL} \log w_K \log w_L + \gamma_{KE} \log w_K \log w_E + \gamma_{KM} \log w_K \log w_M \\ & + \gamma_{LE} \log w_L \log w_E + \gamma_{LM} \log w_L \log w_M + \gamma_{EM} \log w_E \log w_M \\ & + \frac{1}{2} \gamma_{KK} (\log w_K)^2 + \frac{1}{2} \gamma_{LL} (\log w_L)^2 + \frac{1}{2} \gamma_{EE} (\log w_E)^2 + \frac{1}{2} \gamma_{MM} (\log w_M)^2 \\ & + \alpha_y \log y + \frac{1}{2} \gamma_{yy} (\log y)^2 \\ & + \gamma_{Ky} \log w_K \log y + \gamma_{Ly} \log w_L \log y + \gamma_{Ey} \log w_E \log y + \gamma_{My} \log w_M \log y \end{aligned}$$

Por último, además de las restricciones de interés y las adicionales que hemos comentado, podríamos definir las economías de escala presentes en la tecnología de producción utilizada.

Las economías de escala se definen en términos del incremento relativo en el output que resulta de un incremento proporcional en todos los inputs. Christensen y Greene (1976) apuntan, citando a Hanoch (1975), que es más apropiado representar las economías de escala mediante la relación entre el coste total y el output (donde los precios de los inputs son constantes y los costes se minimizan en cada nivel de output). De esta forma, Christensen y Greene (1976) comentan que una forma natural de expresar el nivel de las economías de escala es como el incremento proporcional de los costes resultante de un pequeño incremento proporcional en el nivel de output. Esta es la elasticidad del coste respecto al output. Estos autores definen la economía de escala ( $EE$ ) como:

$$EE = 1 - \frac{\partial \log c}{\partial \log y}$$

donde  $\frac{\partial \log c}{\partial \log y}$  es la elasticidad del coste respecto al output. Si  $EE$  es positivo, entonces existen economías de escala. Si  $EE$  es negativo, entonces existen diseconomías de escala. Como indican los citados autores,  $EE$  tiene una interpretación en términos de porcentaje.

## **2.5.2. DEMANDA DE FACTORES**

A partir de la función de costes pueden derivarse un sistema de ecuaciones de demanda de factores de producción.

Diferenciando la función de costes en logaritmos ( $\log c$ ) con respecto al logaritmo de los precios de los factores ( $\log w_j$ ), y utilizando el lema de Shephard, podemos obtener un sistema de ecuaciones de demandas de factores que no dependen del output,

las cuales son las condiciones de primer orden del problema de minimización del coste.

Así, por ejemplo,

$$\begin{aligned}
 S_j &= \frac{\partial \log c}{\partial \log w_j} \\
 &= \frac{\partial c}{\partial w_j} \frac{w_j}{c} \\
 &= \frac{w_j Q_j}{c} \\
 &= \alpha_j + \sum_{j=1}^k \gamma_j \log w_j + \gamma_{jy} \log y
 \end{aligned}$$

siendo  $Q_j$  la cantidad de input  $j$ -ésimo, y teniendo en cuenta que

$$c = \sum_{j=1}^k w_j Q_j, \quad \frac{\partial c}{\partial w_j} = Q_j \text{ y } \sum_{j=1}^k S_j = 1.$$

Finalmente, podríamos obtener las elasticidades parciales de sustitución de Allen, basadas en la forma funcional translog. Las expresiones serían calculadas a partir de las siguientes fórmulas:

$$\sigma_{jl} = \frac{\gamma_{jl} + S_j S_l}{S_j S_l}, \quad j \neq l$$

$$\sigma_{jj} = \frac{\gamma_{jj} + S_j^2 - S_j}{S_j^2}$$

Por otro lado, las elasticidades precio de los factores pueden construirse a partir de la expresión:  $\varepsilon_{jl} = S_j \sigma_{jl}$ , de tal manera que:

$$\varepsilon_{jl} = \frac{\gamma_{jl} + S_j S_l}{S_j}, \quad j \neq l$$

$$\varepsilon_{jj} = \frac{\gamma_{jj} + S_j^2 - S_j}{S_j}$$

Con esta expresión podemos determinar qué inputs son sustitutivos y qué inputs son complementarios. Los inputs sustitutivos son aquellos que poseen elasticidades cruzadas,  $\varepsilon_{jl}$ ,  $j \neq l$ , en las que los signos son positivos. Los inputs complementarios son aquellos que poseen elasticidades cruzadas,  $\varepsilon_{jl}$ ,  $j \neq l$ , en las que los signos son negativos.

Siguiendo el ejemplo de la Sección 2.5.1, el sistema de ecuaciones de demanda de factores en la base de las participaciones de cada factor, que representa las condiciones de primer orden del problema de costes, quedaría especificado sin restricciones para  $j, l=K, L, E, M$ ; igual a:

$$\begin{aligned} S_K &= \alpha_K + \gamma_{KK} \log w_K + \gamma_{KL} \log w_L + \gamma_{KE} \log w_E + \gamma_{KM} \log w_M + \gamma_{Ky} \log y \\ S_L &= \alpha_L + \gamma_{LK} \log w_K + \gamma_{LL} \log w_L + \gamma_{LE} \log w_E + \gamma_{LM} \log w_M + \gamma_{Ly} \log y \\ S_E &= \alpha_E + \gamma_{EK} \log w_K + \gamma_{EL} \log w_L + \gamma_{EE} \log w_E + \gamma_{EM} \log w_M + \gamma_{Ey} \log y \\ S_M &= \alpha_M + \gamma_{MK} \log w_K + \gamma_{ML} \log w_L + \gamma_{ME} \log w_E + \gamma_{MM} \log w_M + \gamma_{My} \log y \end{aligned}$$

donde existirán también 24 parámetros a estimar. Sin embargo, también a este sistema pueden imponérseles las restricciones 1) y 2), a)-d). Así, podemos considerar los siguientes aspectos:

a) Restricciones de simetría:

$$\gamma_{KL} = \gamma_{LK}, \gamma_{KE} = \gamma_{EK}, \gamma_{KM} = \gamma_{MK}, \gamma_{LE} = \gamma_{EL}, \gamma_{LM} = \gamma_{ML}, \gamma_{ME} = \gamma_{EM}$$

Imponiendo sólo estas restricciones el número de parámetros a estimar se reduce de 24 a 18.

b) Restricciones de homogeneidad de grado 1 en precios. Concretamente, y al igual que para la ecuación de costes, las restricciones de homogeneidad de grado uno en precios son:

$$\begin{aligned}
 \alpha_K + \alpha_L + \alpha_E + \alpha_M &= 1 \\
 \gamma_{KK} + \gamma_{KL} + \gamma_{KE} + \gamma_{KM} &= 0 \\
 \gamma_{LK} + \gamma_{LL} + \gamma_{LE} + \gamma_{LM} &= 0 \\
 \gamma_{EK} + \gamma_{EL} + \gamma_{EE} + \gamma_{EM} &= 0 \\
 \gamma_{MK} + \gamma_{ML} + \gamma_{ME} + \gamma_{MM} &= 0 \\
 \gamma_{Ky} + \gamma_{Ly} + \gamma_{Ey} + \gamma_{My} &= 0
 \end{aligned}$$

Pues bien, incorporando las dos restricciones: simetría y homogeneidad de grado 1 en precios, y, además, eliminando la ecuación correspondiente a  $M$ , el sistema de ecuaciones de demanda de factores se reduce a:

$$\begin{aligned}
 S_K &= \alpha_K + \gamma_{KK} \log\left(\frac{w_K}{w_M}\right) + \gamma_{KL} \log\left(\frac{w_L}{w_M}\right) + \gamma_{KE} \log\left(\frac{w_E}{w_M}\right) + \gamma_{Ky} \log y \\
 S_L &= \alpha_L + \gamma_{KL} \log\left(\frac{w_K}{w_M}\right) + \gamma_{LL} \log\left(\frac{w_L}{w_M}\right) + \gamma_{LE} \log\left(\frac{w_E}{w_M}\right) + \gamma_{Ly} \log y \\
 S_E &= \alpha_E + \gamma_{KE} \log\left(\frac{w_K}{w_M}\right) + \gamma_{LE} \log\left(\frac{w_L}{w_M}\right) + \gamma_{EE} \log\left(\frac{w_E}{w_M}\right) + \gamma_{Ey} \log y
 \end{aligned}$$

Y, luego, para estimar los coeficientes de la ecuación  $S_M$  se utilizan las restricciones de homogeneidad de grado uno en precios, despejando los parámetros desconocidos en cada una de ellas. En este sentido, el número de parámetros libres para ser estimados será de 12.

## 2.6. METAFRONTERAS DE PRODUCCIÓN

Los conjuntos de tecnología,  $T$ , entre diferentes industrias, regiones o países difieren debido a las diferencias existentes en la disponibilidad de los inputs (capital, trabajo o capital financiero), la infraestructura económica, la dotación de recursos y cualquier otra característica del entorno físico, social y económico en el que tiene lugar la producción.

Estas diferencias pueden ser estudiadas obteniendo funciones de frontera separadas para cada grupo de empresas en la industria, regiones o países. Pero, también puede medirse el grado de eficiencia entre los diferentes grupos.

En esta sección proporcionamos algunos conceptos teóricos útiles para la comparación de grupos de empresas con diferentes tecnologías.<sup>6</sup> Para ello, se definen dos conceptos generales: primero, la eficiencia relativa a una metafrontera común como el límite de un conjunto de tecnología sin restringir y, segundo, las fronteras de grupo como los límites del conjunto de la tecnología restringida, donde las restricciones se derivan de la falta de infraestructura económica y otras características de la producción. Cabe resaltar que la metafrontera envuelve a las fronteras de grupo.

Los antecedentes del enfoque de metafronteras se relacionan con el concepto de función de metaproducción definidos por Hayami (1969) y Hayami y Ruttan (1970, 1971). La función de metaproducción está basada en la idea que todos los productores tienen potencial acceso a una variedad de tecnologías de producción, pero cada uno elige una tecnología particular dependiendo de sus circunstancias específicas, tales como la regulación, los recursos para la producción o los precios relativos de los inputs. Estas condiciones dificultan la elección de la mejor tecnología de entre las potenciales tecnologías del conjunto de tecnologías para algunas empresas en algunos grupos. El gap de tecnologías de producción es la diferencia entre la mejor tecnología y la tecnología elegida (es decir, la frontera del grupo).

La eficiencia relativa a la metafrontera puede descomponerse en dos componentes:

---

<sup>6</sup> La descripción de los conceptos utilizados en esta sección está basada en los documentos y artículos escritos por Rao et al. (2003), Battese et al. (2004) y O'Donnell et al. (2008).



- 1) Uno que mide la distancia desde el punto que define la combinación input-output a la frontera de grupo (es la medida común de eficiencia técnica vista en la Sección 3).
- 2) El otro mide la distancia entre la frontera de grupo y la metafrontera (representando la naturaleza restrictiva de la producción).

A continuación, definiremos la metafrontera y la frontera específica de grupo en términos del conjunto de output y las funciones distancia con orientación output, las cuales pueden emplearse también para definir la eficiencia técnica y las ratios de metatecnología.

### 2.6.1. METAFRONTERAS

En la Sección 2 se definieron los conjuntos de tecnología ( $T$ ) y producción ( $P$ ), y en la Sección 2.3.3 se definió la función distancia ( $D$ ) (es decir, dados los inputs, la máxima cantidad en que la empresa puede extender su output). Estos conceptos también serán utilizados en la metafrontera por lo que, para distinguirlos de las fronteras de grupo, incluiremos el superíndice  $M$  en las expresiones de metafrontera.

Así, por ejemplo, el conjunto de tecnología de la metafrontera se define por:

$$T^M = \{(y, x) : x \text{ puede producir } y\}$$

y el conjunto de metaproducción se define por:

$$P^M(x) = \{y : (y, x) \in T^M\}$$

La función metadistancia con orientación output es definida como:

$$D_o^M(x, y) = \min \{ \mu : y/\mu \in P^M(x) \}$$

Una combinación  $(x, y)$  puede considerarse técnicamente eficiente con respecto a la metafrontera sí y solo sí  $D_o^M(x, y) = 1$ .

### **2.6.2. FRONTERAS ESPECÍFICAS DE GRUPO**

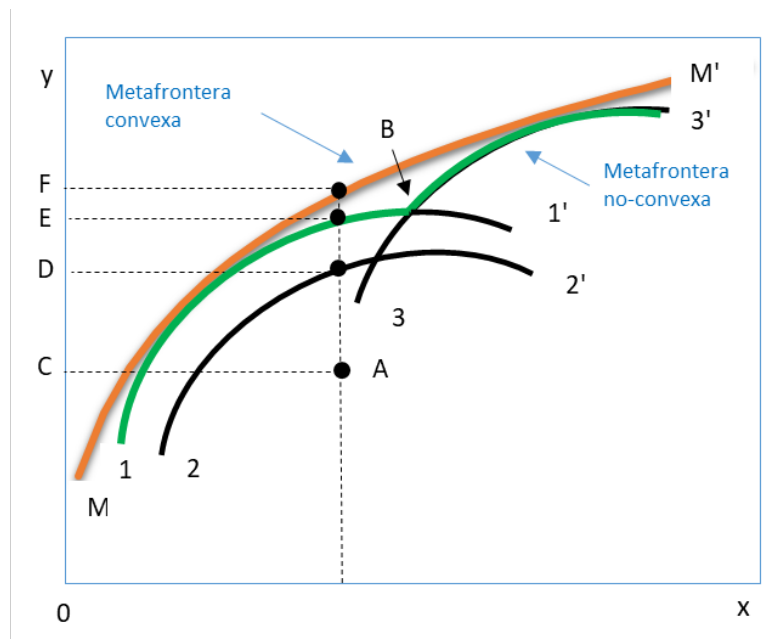
En esta sección se presentan los conceptos que definen la existencia de sub-tecnologías los cuales representan las posibilidades de producción de grupos de empresas.

Supongamos que existen  $J$  grupos. Los conjuntos de sub-tecnología y producción y la función distancia que son específicas de los grupos pueden representarse como:

$$\begin{aligned} T^j &= \{ (y, x) : x \text{ puede producir } y \text{ en el grupo } j \}, j = 1, \dots, J \\ P^j(x) &= \{ y : (y, x) \in T^j \} \\ D_o^j(x, y) &= \min \{ \mu : y/\mu \in P^j(x) \} \end{aligned}$$

La Figura 2.4 muestra las posibilidades de producción disponibles para un solo output y un solo input. En ella pueden apreciarse las metafronteras convexa ( $MM'$ ) y no convexa ( $1-B-1'$ ) así como las fronteras específicas de grupo ( $1-1'$ ,  $2-2'$  y  $3-3'$ ) que son los límites del conjunto de output del grupo.

**Figura 2.4.** Metafronteras.



Las siguientes reglas caracterizan el comportamiento de las combinaciones input-output, la tecnología, la distancia y la convexidad. Estas son:

- 1) Si  $(x, y) \in T^j$  para cualquier  $j$ , entonces  $(x, y) \in T^M$ .
- 2) Si  $(x, y) \in T^M$ , entonces  $(x, y) \in T^j$  para algún  $j$ .
- 3)  $T = \{T^1 \cup T^2 \cup \dots \cup T^J\}$ .
- 4)  $D_o^j(x, y) \geq D_o^M(x, y), \forall j = 1, \dots, J$ .
- 5) La convexidad de  $P^M(x)$  no implica convexidad de  $P^j(x)$ , y viceversa.

Las reglas 1) a 4) implican que los conjuntos de producción  $P^j(x)$ ,  $j = 1, \dots, J$  son subconjuntos del conjunto no restringido  $P^M(x)$ . Se asume que las tres fronteras de grupo son convexas, pudiendo ser los grupos exhaustivos o no. Así, puede ocurrir que:

- En el caso de que los tres grupos sean exhaustivos, la metafrontera abarcaría todas las combinaciones input-ouput que podrían producirse por cualquier empresa. Esto implicaría que la metafrontera no es convexa (1-B-1').
- Si los tres grupos no son exhaustivos, entonces pueden ocurrir combinaciones de output-input y la función metafrontera puede ser una curva convexa de  $M$  a  $M'$ .

### 2.6.3. EFICIENCIA TÉCNICA Y RATIO DE GAP TECNOLÓGICO

Suponiendo la orientación-output, como hemos comentado anteriormente, una combinación  $(x, y)$  es técnicamente eficiente con respecto a la metafrontera sí y solo sí  $D_o^M(x, y) = 1$ . De esta forma, podemos escribir que la eficiencia técnica para la metafrontera es igual a la función distancia, tal que:

$$TE_o^M(x, y) = D_o^M(x, y),$$

mientras que para las fronteras específicas de grupo puede escribirse:

$$TE_o^j(x, y) = D_o^j(x, y), j = 1, \dots, J.$$

La ratio de gap tecnológico ( $TGR$ ), que representa la distancia desde la frontera específica de grupo a la metafrontera, se escribe como:

$$TGR^j = \frac{D_o^M(x, y)}{D_o^j(x, y)} = \frac{TE_o^M(x, y)}{TE_o^j(x, y)} \leq 1, j = 1, \dots, J$$

es decir, dicho ratio puede construirse o bien con las funciones de distancia de orientación output o bien con la eficiencia técnica de la misma orientación pues, como hemos visto, son iguales. Cabe destacar que tanto las eficiencias técnicas como los ratios de gap tecnológico son menores o iguales a la unidad.

La expresión anterior permite construir la ratio de eficiencia técnica de la empresa con respecto a la metafrontera de la tecnología de producción, tal que:

$$TE_o^M(x, y) = TE_o^j(x, y) \times TGR^j \leq 1, j = 1, \dots, J$$

Esta expresión muestra que la eficiencia medida respecto de la metafrontera se puede descomponer en el producto entre la eficiencia técnica del subgrupo frontera y la ratio metatecnológica. Esta descomposición es muy útil ya que permite a los gestores de las empresas estimar el potencial de sus diferentes programas a la hora de fijar y conseguir objetivos.

Para disponer de una idea sobre el cálculo de las medidas anteriormente citadas, podemos utilizar la información de la Figura 2.4 que distingue varias metafronteras: la convexa y no convexa, respectivamente. A partir de dicha figura podemos construir las medidas de eficiencia técnica y ratio de gap tecnológico considerando el punto A, que es una unidad productiva perteneciente al grupo que genera la frontera 2. Atendiendo a este punto, la medida de eficiencia y la ratio metatecnológica, tanto para la metafrontera convexa como para la no-convexa son los siguientes.

a) *Metafrontera no-convexa*

La eficiencia técnica puede construirse de la siguiente manera:

$$TE_o^M(A) = \frac{OC}{OE}$$
$$TE_o^2(A) = \frac{OC}{OD}$$

donde,

$$TGR^2 = \frac{TE_o^M(A)}{TE_o^2(A)} = \frac{\frac{0C}{0E}}{\frac{0C}{0D}}$$

b) *Metafrontera convexa*

Las ratios de eficiencia técnica son:

$$TE_o^M(A) = \frac{0C}{0F}$$

$$TE_o^2(A) = \frac{0C}{0D}$$

y,

$$TGR^2 = \frac{TE_o^M(A)}{TE_o^2(A)} = \frac{\frac{0C}{0F}}{\frac{0C}{0D}}$$

Como conclusión, podríamos decir que los resultados obtenidos para la metafrontera no-convexa deberían ser más ajustados que si los proyectásemos en la convexa. Esto es así debido a que la metafrontera convexa es la envolvente de la no-convexa.

Finalmente, resaltar que las estimaciones de las eficiencias  $TE_o^j(x, y)$  y la ratio tecnológica  $TGR^j, j=1, \dots, J$  serán explicadas en los Capítulos 3 y 4, pues dichas estimaciones pueden construirse a partir de modelos DEA pero también en el ámbito de los modelos de frontera estocástica.

## 2.7. MÉTODOS USADOS EN EL ANÁLISIS DE LA EFICIENCIA

En esta sección destacamos los distintos modelos de análisis de la eficiencia atendiendo la consideración de la frontera, es decir, si es determinista o estocástica, y

distinguiendo el uso de métodos paramétricos, semiparamétricos y no paramétricos. Dichos modelos serán explicados en los Capítulos 3 y 4 de esta Tesis Doctoral.

Cabe destacar que los métodos paramétricos y no paramétricos usan distintas aproximaciones a la estimación de la eficiencia. Por un lado, los métodos paramétricos estiman la frontera con métodos econométricos. Los métodos paramétricos imponen una forma funcional explícita tanto para la frontera como para las desviaciones de ella (es decir, la ineficiencia). Por otro lado, los métodos no paramétricos usan la programación lineal para calcular segmentos lineales por partes (a trozos) de la frontera eficiente. Por el contrario, los métodos no paramétricos, no imponen ninguna hipótesis sobre la forma funcional de la frontera, ni sobre la ineficiencia. El principal inconveniente de los métodos no paramétricos es que no incluyen el error aleatorio en la estimación de la eficiencia, por lo que la distancia a la frontera eficiente se mide en su totalidad como ineficiencia.

**Tabla 2.2** Métodos usados en los modelos de fronteras eficientes.

<b>Métodos</b>	<b>Paramétrico</b>	<b>Semi-paramétrico</b>	<b>No-paramétrico</b>
<b>Determinista</b>	OLS, <i>Goal programming</i> , COLS, MOLS	--	DEA, CNLS
<b>Estocásticos</b>	SFA (frecuentista y Bayesiano)	SFA	StoNED

La Tabla 2.2 clasifica los métodos atendiendo a los modelos de frontera determinista y estocástica.

Sea  $f(x)$  una función de producción que representa la tecnología de producción, donde  $x$  es un vector de inputs  $k$  dimensional. La frontera  $f(x)$  indica el output máximo que puede producirse con el vector de inputs  $x$ , y por lo tanto la función  $f(x)$  caracteriza el límite del conjunto de posibilidades de producción. Suponemos que la función  $f$  pertenece a la clase de funciones continuas, monótona creciente y globalmente cóncava que pueden ser no diferenciables.

Por ejemplo, si deseamos obtener la eficiencia en modelos deterministas, podemos usar los métodos no paramétricos como el Análisis Envolvente de Datos (DEA) de Charnes et al. (1978) o los mínimos cuadrados no lineales convexos (CNLS) propuestos por Hildreth (1954) y Hanson y Pledger (1976). Por el contrario, si deseamos la obtención de fronteras estocásticas podemos utilizar distintas aproximaciones. Por ejemplo, mediante mínimos cuadrados ordinarios (OLS), como hicieron Cobb y Douglas (1928) sin asumir comportamiento distribucional alguno del error, la programación por objetivos (*goal programming*) de Aigner y Chu (1968), los mínimos cuadrados ordinarios corregidos (COLS) de Winsten (1957) y Greene (1980a), y los mínimos cuadrados ordinarios modificados (MOLS) de Afriat (1972) y Richmond (1974). O bien, también podemos asumir comportamiento distribucional para los errores del modelo, que es el caso de los métodos de máxima verosimilitud (ML) empleados por la metodología SFA asumiendo diferentes mezclas de distribuciones, tanto para el caso de independencia (o modelo clásico) como dependencia (modelos recientes basados en cópulas o distribuciones bivariadas) de los términos de ineficiencia y ruido, que veremos más adelante. Por último, también cabría comentar el enfoque semiparamétrico de la estimación de frontera (Fan et al., 1996; entre otros) y el enfoque no paramétrico del Análisis Envolvente no Paramétrico Estocástico de Datos



(*Stochastic Nonparametric Envelopment of Data*, StoNED) que mezcla aspectos de la metodología SFA con el enfoque no paramétrico DEA. También aludiremos a la metodología Bayesiana.

En resumen, teniendo en cuenta lo anterior, dedicaremos los Capítulos 3 y 4 a describir los modelos deterministas y estocásticos de frontera eficiente más relevantes, tanto desde un punto de vista no paramétrico: el DEA – que puede emplearse tanto para funciones de producción como para costes -, el cálculo de los índices de Malmquist para evaluar el cambio de productividad en el tiempo, y la metafrontera para analizar el gap tecnológico, como desde el punto de vista de los métodos paramétricos basados en el cálculo/estimación de funciones de producción/coste eficientes. En cuanto a los modelos estocásticos, distinguimos el enfoque paramétrico del SFA, que usa modelos econométricos en los que el error aleatorio es descompuesto en dos partes: el componente idiosincrático y la ineficiencia, y el enfoque semiparamétrico y también el enfoque no paramétrico del StoNED. También ilustraremos los modelos Bayesianos de fronteras estocásticas.

# CAPÍTULO 3

## MEDICIÓN DE LA EFICIENCIA – FRONTERAS DETERMINISTAS PARAMÉTRICAS Y NO PARAMÉTRICAS

---

### 3.1. INTRODUCCIÓN

Este Capítulo se dedica a la descripción de los modelos de frontera más utilizados en la literatura sobre eficiencia para el cálculo y estimación tanto de las fronteras eficientes deterministas paramétricas como no paramétricas.<sup>7</sup> Dejaremos para el Capítulo 4 la descripción de los modelos y métodos en el ámbito de la frontera estocástica paramétrica, semiparamétrica y no paramétrica.

### 3.2. MODELOS DE REGRESIÓN DE FRONTERA DETERMINISTAS

Consideremos la siguiente función de producción de un output y un input,  $f(x_i; \beta)$ , donde  $x_i$  es el input  $i$ -ésimo, y  $\beta$  es un parámetro desconocido que caracteriza la tecnología, tal que:

---

<sup>7</sup> Cabe destacar, en este punto, que Kumbhakar y Lovell (2000, página 67) distinguen entre las expresiones “cálculo” y “estimación”, indicando que la primera se refiere a la obtención de la frontera vía programación matemática, mientras que la segunda se refiere al uso de técnicas de regresión que permiten la inferencia estadística (estimación y contraste de los parámetros de los modelos). Sin embargo, en esta Tesis Doctoral usaremos indistintamente las expresiones “cálculo” y “estimación”.

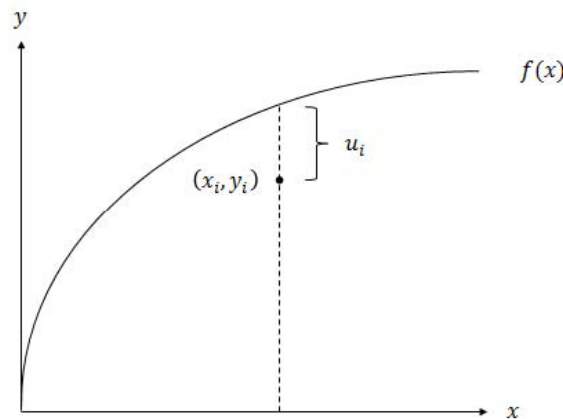
$$y_i = f(x_i; \beta) \exp(-u_i), \quad i = 1, 2, \dots, n, \quad [3.1]$$

siendo  $y_i$  la observación del producto  $i$ -ésimo y  $u_i$  el error positivo que caracteriza la ineficiencia, sobre el que no se establece ningún supuesto distribucional. En este caso, la eficiencia técnica,  $TE$ , puede calcularse fácilmente a través de la expresión:

$$TE_i = \frac{y_i}{f(x_i; \beta)} = \exp(-u_i), \quad i = 1, 2, \dots, n,$$

donde  $u_i \geq 0$ , garantiza que  $y_i \leq f(x_i; \beta)$ , y que  $TE_i \leq 1$ .

**Figura 3.1.** Eficiencia técnica con una función de producción determinista.



La Figura 3.1 muestra una representación de la ineficiencia en un contexto de frontera determinista y con una orientación hacia el output usando solamente un input y un solo output. La medida Debreu-Farrell se obtiene mediante la proyección de la observación a la frontera de producción y formando la ratio entre la producción observada y la producción máxima. Expresado de otra forma, sería la distancia entre el punto que define la coordenada de la combinación input y output,  $(x_i, y_i)$ , y el valor

que alcanzaría en la frontera para una cantidad de input  $x_i, f(x_i)$ . Una medida orientada hacia los inputs también podría construirse de una manera similar.

El principal objetivo en este enfoque de frontera determinista es obtener estimaciones de  $\beta$  y estimaciones de  $u_i$ , lo que luego permitirá obtener estimaciones de  $TE_i$ .

Hay tres formas de hacerlo (véase Kumbhakar y Lovell, 2000).

La primera es la propuesta por Aigner y Chu (1968), quienes desarrollan un enfoque determinístico basado en la estimación de funciones de producción paramétricas para calcular la ineficiencia proponiendo técnicas de programación lineal y cuadrática (Kumbhakar y Lovell (2000) denominan a este método *goal programming*).

Suponiendo que  $f(x_i; \beta)$  es lineal como, por ejemplo, la función log-lineal de Cobb-Douglas:

$$\log y_i = \beta_0 + \sum_{j=1}^k \beta_j \log x_{ij} - u_i, \quad i = 1, 2, \dots, n \quad [3.2]$$

entonces el problema de optimización lineal se escribe como:

$$\begin{aligned} & \min \sum_{i=1}^n u_i \\ & \text{sujeto a :} \\ & \beta_0 + \sum_{j=1}^k \beta_j \log x_{ij} \geq \log y_i, \quad i = 1, 2, \dots, n \end{aligned}$$

mientras que el modelo de programación cuadrática que minimiza la suma del cuadrado de los errores es, tal que:

$$\begin{aligned} & \min \sum_{i=1}^n u_i^2 \\ & \text{sujeto a :} \\ & \beta_0 + \sum_{j=1}^k \beta_j \log x_{ij} \geq \log y_i, \quad i = 1, 2, \dots, n \end{aligned}$$

La estimación del error se obtendría calculando:  $\hat{u}_i = \left( \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j \log x_{ij} \right) - \log y_i$ , a partir de la cual se podrá luego estimar la eficiencia técnica,  $TE_i = \exp(-\hat{u}_i)$ .

Este enfoque considera que la desviación de una observación específica de la frontera se interpreta enteramente como ineficiencia. La principal limitación de este método es que no permite hacer inferencias sobre los parámetros estimados ni tampoco sobre la eficiencia, salvo que se establezcan supuestos distribucionales sobre la ineficiencia.<sup>8</sup>

El segundo de los métodos es el estimador de mínimos cuadrados ordinarios corregidos (COLS). Este método fue propuesto por Winsten (1957) para estimar las funciones de producción deterministas. Supone la aplicación de dos etapas, la primera es la estimación del modelo lineal por OLS y, la segunda, es la corrección del término constante debido a que está sesgado tratando de asegurar con ello que la frontera estimada limita los datos desde arriba. Es por ello por lo que a este método se denomina OLS corregido. Sin describir el método OLS, que es estándar y puede consultarse en cualquier manual de texto sobre la disciplina de la Econometría, el término constante del modelo [3.2] estimado por el método COLS es igual a:

$$\hat{\beta}_0^{COLS} = \hat{\beta}_0 + \max \{ \hat{u}_i \}$$

siendo  $\hat{u}_i$  los residuos OLS. Los residuos OLS se corrigen en la dirección opuesta a través de la expresión:

---

<sup>8</sup> En el caso de que se establezca algún supuesto y este sea que  $u_i \geq 0$  sigue la distribución exponencial, las estimaciones de programación lineal son estimaciones ML (Schmidt, 1976). Si la distribución es la half-normal, las estimaciones de programación cuadrática son estimaciones ML. Sin embargo, a pesar de esta relación, la obtención de errores estándar para poder realizar contrastes de hipótesis es imposible en el caso half-normal (Greene, 1980), y bastante compleja en el caso exponencial.

$$-\hat{u}_i^{COLS} = \hat{u}_i - \max \{ \hat{u}_i \}$$

Estos residuos pueden usarse para obtener la eficiencia técnica,  $TE_i = \exp(-\hat{u}_i^{COLS})$ . Este método es un enfoque ligeramente distinto de MCO puesto que implica desplazar la curva hacia la empresa con mejor desempeño o mejores prácticas. El procedimiento COLS desplaza la frontera en la cantidad de los errores más elevados, generando así una frontera que realmente envuelve los datos. Sin embargo, a pesar de su simplicidad, este método tiene un problema y es que, al corregir solamente al coeficiente de la constante, la función de producción estimada es paralela a la obtenida por OLS haciendo que la estructura de la mejor tecnología de producción sea la misma que las que no son tan buenas.

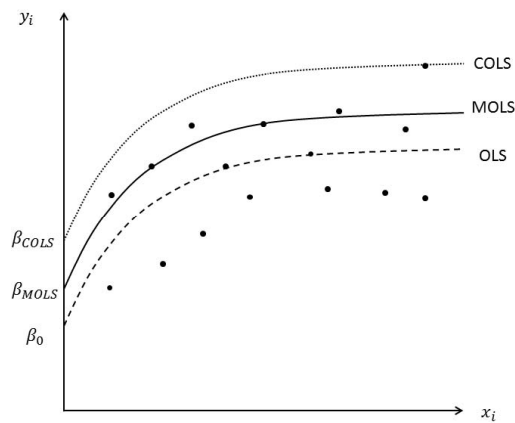
Debido a esto, se planteó un tercer método, denominado OLS modificado (MOLS). Desarrollado por Afriat (1972) y Richmond (1974) este método propone una variación de COLS. Este enfoque requiere una suposición acerca de la distribución del término de ineficiencia: exponencial, half-normal, o normal-truncada. Dado que la media de los residuos MCO es cero por construcción, lo que ofrece poca información, usar un momento centrado de los residuos puede utilizarse para obtener un estimador consistente de la media de la ineficiencia,  $E[u_i]$ . Después de la obtención de esta estimación, la frontera se desplaza de la misma manera que en COLS, pero ahora usando  $E[u_i]$  como el elemento de cambio.

Básicamente, MOLS posee dos etapas que implican, primero, estimar el modelo de producción por OLS asumiendo una distribución de una sola cola para la ineficiencia (por ejemplo, exponencial o half-normal) y luego realizar la corrección de la constante usando la media de los errores obtenidos, tal que:  $\hat{\beta}_0^{MOLS} = \hat{\beta}_0 + E[\hat{u}_i]$ . Y, segundo, redefinir los residuos OLS a través de la expresión:  $-\hat{u}_i^{MOLS} = \hat{u}_i - E[\hat{u}_i]$ . Así, la

eficiencia técnica puede calcularse según:  $TE_i = \exp(-\hat{u}_i^{MOLS})$ . Este es otro método fácil de aplicar. Este método también implica que la frontera es paralela a la estimación OLS y también tiene el problema que si  $\hat{u}_i - E[\hat{u}_i] > 0$ , la puntuación de eficiencia podría ser superior a la unidad.

La Figura 3.2 muestra gráficamente el comportamiento de los tres métodos expuestos para el caso de un output y un input, donde puede verse el desplazamiento de las curvas debido a los diferentes tratamientos del término constante.

**Figura 3.2.** Fronteras de producción obtenidas por OLS, COLS y MOLS.



### **3.3. MÉTODOS NO PARAMÉTRICOS**

#### **3.3.1. ANÁLISIS ENVOLVENTE DE DATOS (DEA)**

Los métodos de estimación de la eficiencia anteriormente descritos entraron en desuso hacia finales de los años 70 del siglo XX y fueron rápidamente reemplazados por otros métodos, como fueron el Análisis Envolverte de Datos (*Data Envelopment Analysis* o DEA) y el Análisis de Fronteras Estocásticas (*Stochastic Frontier Analysis* o SFA), actualmente muy utilizados.

En esta Sección nos ocuparemos del método DEA (Farrell, 1957; Charnes et al., 1978), dejando para el Capítulo 4 la descripción de los modelos SFA. Para una revisión reciente de los métodos DEA y su base estadística puede consultarse Simar y Wilson (2013).<sup>9</sup>

El método DEA es un enfoque de programación matemática axiomática para el análisis de la eficiencia. La principal ventaja de DEA, en comparación con las herramientas econométricas basadas en regresión, es el tratamiento no paramétrico de la frontera basándose en axiomas de la teoría de la producción como, por ejemplo, la monotonidad fuerte y convexidad. El DEA no asume ninguna forma funcional particular para la frontera o la distribución de la ineficiencia. Es un enfoque directo basado en datos y es útil para comunicar rápidamente los resultados del análisis de eficiencia a los tomadores de decisiones. Sin embargo, la principal deficiencia del DEA es que atribuye todas las desviaciones de la frontera a la ineficiencia.

---

<sup>9</sup> Para una revisión completa de la literatura sobre los desarrollos teóricos y las aplicaciones de los modelos DEA, véase Charnes et al. (1995), Seiford (1996), Coelli et al. (2005), entre otros.



El cálculo de la frontera eficiente se realiza a partir del casco convexo lineal a trozos (*piecewise-linear convex hull*) propuesto por Farrell (1957). Diversos autores como Boles (1966) y Afriat (1972) sugirieron el uso de métodos de programación matemática. Sin embargo, es en el trabajo de Charnes et al. (1978, CCR) - quienes acuñaron el nombre de Análisis Envolvente de Datos -, donde se propuso estimar la frontera propuesta por Farrell (1957) a través de un modelo que tiene una orientación-input y asume una tecnología donde los rendimientos son constantes a escala (CRS). En un trabajo posterior, Banker et al. (1984, BCC) desarrollaron un modelo en el que se asumía que la tecnología era la de los rendimientos variables a escala (VRS) que desarrolla la localización de la escala más productiva al comparar cada unidad analizada con las que tienen similar tamaño. No obstante, cabe resaltar que también se desarrollaron otros enfoques considerando la orientación output.

Entre las ventajas del enfoque DEA está que no se requiere especificar una forma funcional y no realiza supuestos sobre la distribución del término ineficiencia. Entre las limitaciones están que es sensible al número de variables, errores de medida y observaciones anómalas (*outliers*).

Antes de iniciar la explicación de los modelos DEA más utilizados, consideremos la siguiente notación general para este apartado, que cambia ligeramente respecto de la empleada en la ecuación [2.1].

La notación es la siguiente:

- Consideremos que disponemos de  $n$  observaciones de unidades de toma de decisión (DMU), por ejemplo, oficinas bancarias, tal que  $i = 1, 2, \dots, n$ .
- Sea  $x_{ik}$  el nivel observado del input  $k$ -ésimo para la DMU  $i$  y sea  $(x_{i1}, \dots, x_{iK})$  un vector de  $K$  inputs para la  $i$ -ésima unidad.

- Sea  $y_{iq}$  el nivel observado de output  $q$ -ésimo de la DMU  $i$ , siendo  $(y_{i1}, \dots, y_{iQ})$  un vector de  $Q$  outputs para la  $i$ -ésima unidad.
- Sea  $\theta$  (o  $\phi$ ) el score de la eficiencia radial. En general, este valor está comprendido entre 0 y 1 para los modelos no paramétricos de frontera determinista, de tal manera que un valor igual a 1 indica que la DMU cae en la frontera y es totalmente eficiente; 0 indicaría lo contrario, es decir, es totalmente ineficiente. En los casos de frontera parcial, este parámetro puede ser superior a la unidad.
- $\lambda_j$  son los pesos óptimos de las unidades  $j$ .
- $s_j^-$  es el exceso para el  $j$ -ésimo input y  $s_r^+$  es el déficit el  $r$ -ésimo output, y
- $\varepsilon$  es un infinitesimal no-Archimedeano.

El modelo DEA permite obtener la eficiencia técnica y de escala, la eficiencia de costes y asignativa, y también el cambio de productividad en el tiempo a través de los índices de Malmquist, permitiendo éste la descomposición en cambio técnico, cambio de eficiencia técnica y cambio de escala.

A continuación, describiremos cómo obtener cada uno de ellos.

### **3.3.2. EFICIENCIA TÉCNICA Y DE ESCALA**

En esta sección definiremos el modelo que permite estimar la eficiencia técnica y de escala. El concepto “técnicamente eficiente” refleja la capacidad de una unidad para obtener el máximo output a partir de unos inputs dados. Eficiente en escala (o escala

óptima) indica que el tamaño de sus operaciones es óptimo, de modo que cualquier modificación en su tamaño hará que la unidad sea menos eficiente.

Para calcular ambas eficiencias, deben plantearse dos modelos. El primero, es el modelo CCR, que permite obtener la eficiencia técnica (o eficiencia técnica global). En la orientación-input y considerando que la tecnología de rendimientos constantes a escala (CRS), puede escribirse como DEA-CRS tal que:

$$\begin{aligned} \min_{\theta, \lambda} \theta - \varepsilon \left( \sum_{q=1}^Q s_q^+ + \sum_{k=1}^K s_k^- \right) \\ \text{sujeto a :} \\ y_{0q} = \sum_{i=1}^n y_{iq} \lambda_i - s_q^+ \\ \theta x_{0k} - \sum_{i=1}^n x_{ik} \lambda_i - s_k^- = 0 \\ \lambda_i, s_k^-, s_q^+ \geq 0, \forall i, k, q \end{aligned} \quad [3.3]$$

donde el sub-índice 0 representa la DMU que está siendo evaluada en el conjunto  $i = 1, 2, \dots, n$ . En este modelo,  $\theta$  representa la eficiencia técnica global (GTE). Es decir,  $TE = \theta$ . Cabe destacar que, si la orientación fuese output, el modelo debería re-especificarse de la siguiente manera:

$$\begin{aligned} \min_{\theta, \lambda} \phi - \varepsilon \left( \sum_{q=1}^Q s_q^+ + \sum_{k=1}^K s_k^- \right) \\ \text{sujeto a :} \\ \phi y_{0q} = \sum_{i=1}^n y_{iq} \lambda_i - s_q^+ \\ x_{0k} - \sum_{i=1}^n x_{ik} \lambda_i - s_k^- = 0 \\ \lambda_i, s_k^-, s_q^+ \geq 0, \forall i, k, q \end{aligned} \quad [3.4]$$

es decir, considerando que  $\phi$  aparece en la restricción de los outputs.

Los modelos definidos por [3.3] y [3.4] (sean de orientación-input u orientación-output) son apropiados si todas las DMU están operando en una escala óptima. Sin embargo, no son apropiados si no operan en dicha escala. Por ejemplo, la competencia imperfecta o las restricciones financieras, entre otras cuestiones, pueden provocar que la DMU no esté operando en la escala óptima. Por ello, Banker et al. (1984) sugirieron la extensión del modelo DEA-CRS (o CCR) para tener en cuenta los rendimientos variables a escala (VRS).

El segundo de los modelos que comentaremos brevemente es el BCC (o DEA-VRS). Este modelo, que usa una tecnología VRS, puede expresarse en su orientación-input como:

$$\min_{\theta, \lambda} \theta - \varepsilon \left( \sum_{q=1}^Q s_q^+ + \sum_{k=1}^K s_k^- \right)$$

*sujeto a :*

$$y_{0q} = \sum_{i=1}^n y_{iq} \lambda_i - s_q^+ \tag{3.5}$$

$$\theta x_{0k} - \sum_{i=1}^n x_{ik} \lambda_i - s_k^- = 0$$

$$\sum_{i=1}^n \lambda_i = 1$$

$$\lambda_i, s_k^-, s_q^+ \geq 0, \forall i, k, q$$

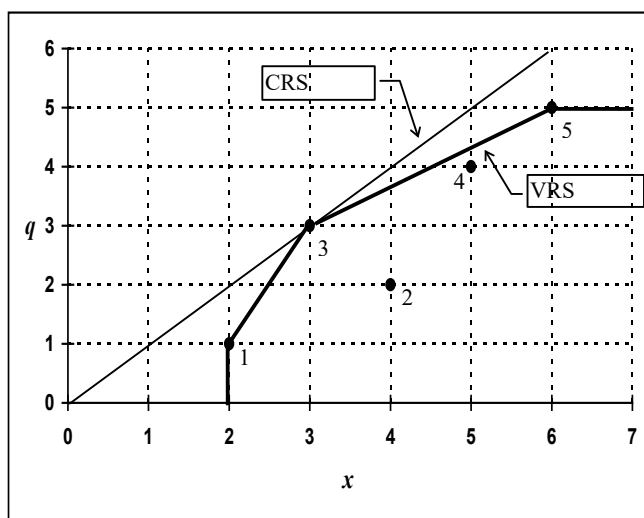
La principal diferencia con el modelo CCR es que el modelo BCC (ecuación [3.5]) incluye la restricción de convexidad  $\sum_{i=1}^n \lambda_i = 1$ , que asegura que una unidad sólo se compara con aquellas de tamaño similar. De esta forma,  $\theta$  representa una eficiencia técnica “pura” (PTE). El modelo con VRS crea la frontera como una envolvente de planos convexos que se cortan en contraste con el modelo con CRS, que forma un casco cónico. Así, el modelo VRS envuelve los datos con más fuerza y proporciona puntuaciones eficientes que son iguales o mayores que las del modelo CRS.

Usando los resultados de las eficiencias técnicas estimadas en cada modelo, la eficiencia de escala (SE) puede computarse fácilmente a través del cociente de la eficiencia técnica estimada por el modelo CCR y la eficiencia técnica estimada por el modelo BCC, es decir,

$$SE = \frac{GTE}{PTE},$$

que mide la proporción de ineficiencia debida exclusivamente a problemas de tamaño. Es decir, la puntuación de la eficiencia de escala indica si una DMU opera al tamaño de escala más productivo.

**Figura 3.3.** Fronteras CRS y VRS DEA input orientadas (Coelli, 1996).



La Figura 3.3 (Coelli, 1996) ilustra las fronteras eficientes para el caso de un solo input y output bajo CRS y VRS, respectivamente. Los puntos 2 y 4 son claramente ineficientes, tanto para CRS como DEA.

Sin embargo, una limitación de la medición de las economías de escala es que el valor obtenido de la eficiencia no indica la naturaleza de las ineficiencias de escala, es

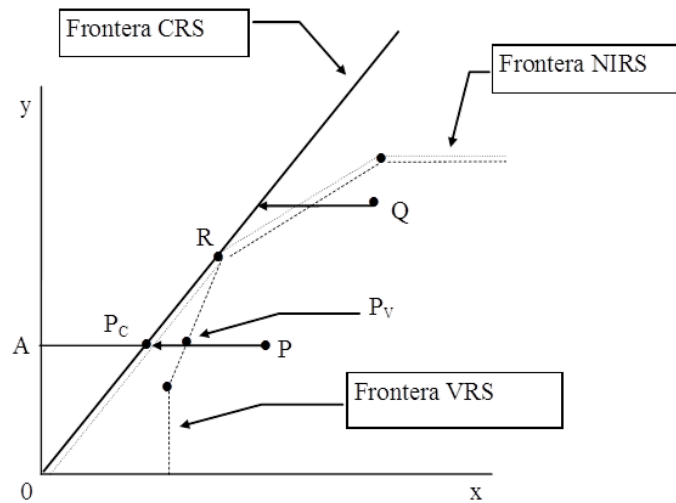
decir, si la DMU está operando en un área de rendimientos a escala crecientes (IRS) o decrecientes (DRS). Para poder identificar si una unidad puede ser DRS o IRS, debe revisarse el modelo BCC considerando un modelo DEA adicional con rendimientos a escala no crecientes (*non-increasing returns to scale*, NIRS). Para ello, simplemente deberá sustituirse la restricción de convexidad  $\sum_{i=1}^n \lambda_i = 1$  por  $\sum_{i=1}^n \lambda_i \leq 1$ . Así, para una DMU particular, también podemos evaluar la eficiencia técnica a través del modelo NIRS, e identificar si las unidades se encuentran en DRS o IRS.

La formulación del modelo NIRS es como sigue:

$$\begin{aligned} \min_{\theta, \lambda} \theta - \varepsilon \left( \sum_{q=1}^Q s_q^+ + \sum_{k=1}^K s_k^- \right) \\ \text{sujeto a :} \\ y_{0q} = \sum_{i=1}^n y_{iq} \lambda_i - s_q^+ \\ \theta x_{0k} - \sum_{i=1}^n x_{ik} \lambda_i - s_k^- = 0 \\ \sum_{i=1}^n \lambda_i \leq 1 \\ \lambda_i, s_k^-, s_q^+ \geq 0, \forall i, k, q \end{aligned} \quad [3.6]$$

La puntuación de eficiencia obtenida debe evaluarse si es igual a la obtenida por el modelo BCC (DEA-VRS). Si esta puntuación no es igual, entonces existen rendimientos a escala crecientes (IRS) para la DMU. Si estas son iguales, entonces existen rendimientos a escala decrecientes (DRS) (véase Coelli, 1996; página 30).

**Figura 3.4.** Fronteras DEA-CRS, VRS y NIRS (Coelli, 1996).



La Figura 3.4 (Coelli, 1996) recoge las diferentes fronteras que hemos considerado en este epígrafe usando la técnica DEA. Usando la información de dicha Figura pueden calcularse las eficiencias técnicas y de escala a partir de las siguientes expresiones:

- Eficiencia técnica con tecnología CRS:  $TE_{CRS} = AP_C/AP$
- Eficiencia técnica con tecnología VRS:  $TE_{VRS} = AP_V/AP$
- Eficiencia de escala:  $SE = AP_C/AP_V$

A partir de las expresiones anteriores puede obtenerse fácilmente la relación entre las tres eficiencias, tal que:  $TE_{CRS} = TE_{VRS} \times SE$ .

### 3.3.2.1. MODELOS DEA DE SUPEREFICIENCIA

Un problema importante en la literatura de la DEA es el de clasificar aquellas DMU consideradas eficientes por el modelo, las cuales tienen una puntuación igual a la unidad. Un enfoque que resuelve el problema de la clasificación es el proporcionado por

el modelo de supereficiencia de Andersen y Petersen (1993). Véase también Banker et al. (1989). El modelo de supereficiencia implica ejecutar los modelos DEA estándar (CRS o VRS), pero bajo el supuesto de que la DMU que se evalúa está excluida del conjunto de referencia. El modelo DEA de supereficiencia orientado a inputs se puede formular cuando se evalúa una unidad de toma de decisiones (DMU) ( $j \neq 0$ ) es excluida del conjunto de referencia.

Partiendo del modelo DEA input orientado de Banker et al. (1984, BCC), con rentabilidades a escala variables, escrito como:

$$\begin{aligned} & \min_{\phi, \lambda} \phi \\ & \text{sujeto a:} \\ & y_{0q} = \sum_{j=1}^n y_{qj} \lambda_j \\ & \phi x_{0k} - \sum_{j=1}^n x_{kj} \lambda_j = 0 \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j \geq 0, \forall j \end{aligned}$$

puede formularse un modelo DEA de supereficiencia, tal que:

$$\begin{aligned} & \min_{\gamma, \lambda} \gamma \\ & \text{sujeto a:} \\ & y_{0q} = \sum_{j=1, j \neq 0}^n y_{qj} \lambda_j \\ & \gamma x_{0k} - \sum_{j=1, j \neq 0}^n x_{kj} \lambda_j = 0 \\ & \sum_{j=1, j \neq 0}^n \lambda_j = 1 \\ & \lambda_j \geq 0, \forall j \end{aligned}$$

Cuando la DMU<sub>0</sub> es eficiente y es factible,  $\gamma > 1$ , indicando que los inputs de la DMU<sub>0</sub> se incrementan para llegar a la frontera formada por el resto de DMUs. Es decir, la supereficiencia se expresa en términos de aumentos de inputs.



### 3.3.2.2. Modelos SBM y SBM de supereficiencia

DEA es un método de programación lineal para medir la eficiencia relativa de la DMU y construir fronteras de producción con varios inputs y outputs. Los modelos más básicos de DEA son CCR y BCC, pero estos no tienen en cuenta el efecto de la holgura (*slack*) de los inputs o outputs en la estimación del modelo, por lo que sus estimaciones de eficiencia tienen desviaciones significativas.

En general, DEA utiliza medidas radiales o no radiales para evaluar la eficiencia de las DMU. La suposición para los modelos radiales es que hay un cambio proporcional de inputs u outputs y, generalmente, estos cambios ignoran las holguras en las puntuaciones de eficiencia. Las medidas no radiales son las que tienen en cuenta las holguras de cada input y output, y permiten que las variaciones de estos sean no proporcionales.

Tone (2001) recomendó emplear el modelo SBM (*slacks-based measure*) basado en los *slacks* para evaluar la eficiencia relativa de las DMUs. Este es un modelo DEA no radial. Concretamente, este modelo estima las eficiencias no radiales a diferencia de las versiones DEA clásicas como CCR y BCC. Es conveniente comentar que la diferencia entre una y otra eficiencia tiene que ver con el hecho de que se asume que los inputs y outputs cambian con sus proporciones, ignorando las holguras que existen en la eficiencia. Esto permite considerar la variación no proporcional de factores productivos, a diferencia del modelo DEA clásico, donde dichas variaciones (en inputs u outputs) son proporcionales.

El modelo SBM basado en la tecnología VRS puede escribirse como:

$$\rho^* = \min_{\rho, \lambda} \rho = \frac{1 - \frac{1}{K} \sum_{k=1}^K \frac{s_k^-}{x_{0k}}}{1 + \frac{1}{Q} \sum_{q=1}^Q \frac{s_q^+}{y_{0q}}}$$

sujeto a :

$$\begin{aligned} y_{0q} &= \sum_{j=1}^n y_{qj} \lambda_j - s_q^+ \\ x_{0k} - \sum_{j=1}^n x_{kj} \lambda_j - s_k^- &= 0 \\ \sum_{j=1}^n \lambda_j &= 1 \\ \lambda_j, s_k^-, s_q^+ &\geq 0, \forall j, k, q \end{aligned}$$

El valor óptimo  $\rho^*$  es la eficiencia del modelo SBM,  $0 \leq \rho^* \leq 1$ . Cuando  $\rho^* = 1$  la DMU es totalmente eficiente, porque  $s_k^- = s_q^+ = 0$ , mientras que cuando  $0 \leq \rho^* < 1$  la DMU es ineficiente.

Tone (2002) propuso el modelo SBM de supereficiencia para comparar y distinguir las unidades de eficiencia. El modelo SBM de supereficiencia se obtiene cuando una unidad de toma de decisiones (DMU) en evaluación ( $j \neq 0$ ) se excluye del conjunto de referencia. El modelo supereiciente no angular se puede escribir como:

$$\theta^* = \min_{\theta, \lambda} \theta = \frac{\frac{1}{K} \sum_{k=1}^K \bar{x}_k}{\frac{1}{Q} \sum_{q=1}^Q \bar{y}_q}$$

sujeto a :

$$\begin{aligned} \bar{y} &\leq \sum_{j=1, j \neq 0}^n y_j \lambda_j \\ \bar{x} &\geq \sum_{j=1, j \neq 0}^n x_j \lambda_j \\ \sum_{j=1, j \neq 0}^n \lambda_j &= 1 \\ \lambda_j, \bar{x}, \bar{y} &\geq 0, \bar{y} \leq y_0 \end{aligned}$$

El valor óptimo  $\theta^*$  es el valor de supereficiencia del modelo SBM, que puede ser mayor que 1 para las DMUs supereficientes. Cuando  $\theta^* = 1$ , la DMU es totalmente eficiente, porque  $s_k^- = s_q^+ = 0$ , mientras que cuando  $0 \leq \theta^* < 1$  la DMU es ineficiente.

### 3.3.2.3. TRATANDO OUTPUTS INDESEABLES EN DEA

El tratamiento de los resultados no deseados en el análisis envolvente de datos (DEA) ha recibido recientemente gran atención en la investigación. Halkos y Petrou (2019) discuten la existencia de cuatro posibles opciones para tratar con ellos:

1. Ignorarlos de la función de producción,
2. Tratándolos como inputs regulares,
3. Tratándolos como outputs normales, y
4. Realizando transformaciones necesarias para tenerlas en cuenta.

También nuevas propuestas de modelización han sido realizadas para su tratamiento. Cada método trae consigo ventajas e inconvenientes que cada investigador deberá tener en cuenta en cada etapa de su investigación, evaluando qué método es más apropiado para ser utilizado.

Barros et al. (2012) comentan que la literatura sobre los modelos no radiales puede dividirse en tres grupos. Primero, aquellos que utilizan la medida de Russell (presenta por vez primera en Färe y Lovell (1978) con orientación hacia los inputs). Segundo, el modelo aditivo desarrollado por Charnes et al. (1978). Sin embargo, este modelo no proporciona directamente una medida de eficiencia. Pastor et al. (1999) revisaron el

modelo de Charnes et al. (1978) introduciendo una nueva medida denominada "medida de gráfico Russell mejorada" (ERGM), que a su vez combina la medida de Russell de input y output en forma de ratio. Y, tercero, el modelo basado en holguras (SBM), propuesto por Tone (2001), cuyo objetivo es maximizar todas las holguras de los inputs y outputs usando programación fraccional. Cooper et al. (2007) demostraron que SBM es equivalente a ERGM.

En esta sección describiremos el modelo de Chen et al. (2010), que es utilizado por Barros et al. (2012). A continuación, haremos una breve explicación siguiendo a Barros et al. (2010).

Sea  $n$  el número de observaciones o DMUs,  $i = 1, 2, \dots, n$ . Sea  $x_{ik}$  el nivel observado del  $k$ -ésimo input para la DMU  $i$ , siendo  $x = (x_{i1}, \dots, x_{iK}) \in \mathfrak{R}_+^K$  el vector de  $K$  inputs para la  $i$ -ésima unidad;  $y_{iq}$  el nivel observado de output deseable  $q$ -ésimo de la DMU  $i$ , siendo  $(y_{i1}, \dots, y_{iQ}) \in \mathfrak{R}_+^Q$  un vector de  $Q$  outputs para la  $i$ -ésima unidad; y  $b_{im}$  el nivel observado de output no deseable  $m$ -ésimo de la DMU  $i$ , siendo  $(b_{i1}, \dots, b_{iM}) \in \mathfrak{R}_+^M$  un vector de  $M$  outputs no deseados para la  $i$ -ésima unidad. El conjunto de tecnología puede definirse como:  $T = \{(x, y, b) : x \text{ puede producir } (y, b)\}$ , y satisface los supuestos de rendimientos constantes de escala, fuerte disponibilidad de outputs e inputs deseables, y débil disponibilidad de outputs no deseados.

La función de distancia direccional, que busca aumentar los outputs deseables y disminuir los inputs y outputs indeseables direccionalmente, puede escribirse como:  $\bar{D}(x, y, b; g)$ , donde  $g(-g_x, g_y, -g_b)$  es un vector no nulo que determina las direcciones en que los inputs, y los outputs deseables y no deseados, son escalados (expansión en los outputs buenos y contracción en los inputs y outputs no deseados).

El modelo de distancia direccional de Russell ponderado (*weighted Russell directional distance model, WRDDM*) introducido por Chen et al. (2010) puede escribirse como sigue:

$$\begin{aligned} \bar{D}(x, y, b; g) &= \rho \\ &= \max w_y \left( \sum_{q=1}^Q \omega_q \alpha_q \right) + w_x \left( \sum_{k=1}^K \omega_k \beta_k \right) + w_b \left( \sum_{m=1}^M \omega_m \gamma_m \right) \end{aligned}$$

sujeto a :

$$\sum_{i=1}^n z_i y_{qi} \geq y_{q0} + \alpha_q g_{y_q}, \quad q = 1, \dots, Q$$

$$\sum_{i=1}^n z_i b_{mi} = b_{m0} - \beta_m g_{b_m}, \quad m = 1, \dots, M$$

$$\sum_{i=1}^n z_i x_{ki} \leq x_{k0} - \gamma_k g_{x_k}, \quad k = 1, \dots, K$$

$$z_i \geq 0, \quad i = 1, \dots, n$$

donde  $z_i$  son las variables de intensidad para expandir o reducir las actividades individuales observadas de DMU  $i$  para construir combinaciones convexas de las inputs y outputs observados, donde  $\alpha_q$ ,  $\beta_m$  y  $\gamma_k$  son las medidas de ineficiencia individual para cada output deseable  $y_q$ , cada output no deseable  $b_m$  y cada input  $x_k$ . Esto implica que esta especificación no solo permite que la ineficiencia técnica asociada con el output deseable, el output no deseable y el input sean diferentes, sino también permite que la ineficiencia entre cada uno de los outputs deseables, los no deseables y los inputs sean diferentes.

La función de distancia direccional ponderada proporciona la expansión en los outputs buenos y la contracción en los outputs malos y en los inputs, todos juntos. Cuando  $\bar{D}(x, y, b; g) = 0$ , la DMU  $i$  es técnicamente eficiente. Esto se debe a que no hay mejoras adicionales en los outputs buenos, y los outputs malos y los inputs son viables.  $\bar{D}(x, y, b; g) > 0$  indica ineficiencia técnica. Los coeficientes  $w_y$ ,  $w_b$  y  $w_x$  están

asociados con las prioridades o preferencias gerenciales dados los outputs (buenos y malos) y los inputs, y su suma se normaliza a la unidad. Las mejoras de los outputs deseables, no deseados e inputs puede medirse por  $\alpha$ ,  $\beta$  y  $\gamma$ , de tal forma que luego pueden usarse para computar  $\rho$  (la puntuación ponderada de ineficiencia).

#### **3.3.2.4. DEA CON DATOS NEGATIVOS**

En la presencia de datos negativos, los modelos radiales tradicionales para la evaluación de la eficiencia no se pueden utilizar sin transformar los datos, ya que mueven los inputs/outputs negativos en la dirección incorrecta.

El tratamiento de los datos negativos es, en cierto modo, similar al tratamiento de outputs y/o inputs no deseables, ya que tanto los datos negativos como los outputs no deseables deben moverse en una dirección que es contraria a la dirección utilizada en los modelos DEA tradicionales (Silva Portela et al., 2004).

Existen varios enfoques para hacer frente a los resultados no deseados (véanse Allen (1999) y Dyckhoff y Allen (2001), entre otros, para una revisión general). Uno de estos enfoques se basa en la función de distancia direccional, y fue propuesto por primera vez por Chung et al. (1997).

Siguiendo a Silva Portela et al. (2004), el modelo aditivo estándar es la principal herramienta de evaluación de la eficiencia que se ha utilizado en estos casos, debido a sus propiedades invariantes en la traslación (y unidades invariantes en algunos casos). Sin embargo, el modelo aditivo tiene dos desventajas principales: tiende a proyectar unidades en los puntos más lejanos de la frontera, lo que implica esfuerzos innecesarios por parte de las unidades de producción, y no proporciona una medida de eficiencia final mediante la cual se pueden hacer comparaciones y clasificaciones.

En esta sección se formula el modelo de Silva Portela et al. (2004), basado en el modelo de distancia direccional, que extiende el modelo de Chambers et al. (1996, 1998), para tratar datos negativos en el modelo DEA, y supera las limitaciones antes comentadas para el modelo aditivo.

Sea  $n$  el número de DMUs,  $i=1,2,\dots,n$ . Sea  $x_{ik}$  el nivel observado del  $k$ -ésimo input para la DMU  $i$  y sea  $x=(x_{i1},\dots,x_{iK})\in\mathfrak{R}^K$  el vector de  $K$  inputs para la  $i$ -ésima unidad;  $y_{iq}$  es el nivel observado del  $q$ -ésimo output para la DMU  $i$ , donde  $y=(y_{i1},\dots,y_{iQ})\in\mathfrak{R}^Q$  es un vector de  $Q$  outputs.

A continuación, escribimos la expresión del modelo general para una función direccional de distancia como la propuesta por Silva Portela et al. (2004), que extiende el modelo de Chambers et al. (1996, 1998) para el caso de la tecnología VRS y un vector de inpus y outputs definidos en  $\mathfrak{R}^{K+Q}$ .

El modelo se denomina RDM (*range directional measure*) y se escribe como sigue:

$$\begin{aligned} & \max \beta_0 \\ & \text{sujeto a :} \\ & \sum_{i=1}^n \lambda_i y_{qi} \geq y_{q0} + \beta_0 R_{q0}, q = 1, \dots, Q \\ & \sum_{i=1}^n \lambda_i x_{ki} \leq x_{k0} - \beta_0 R_{k0}, k = 1, \dots, K \\ & \sum_{i=1}^n \lambda_i = 1 \\ & \lambda_i \geq 0, \forall i \end{aligned}$$

donde  $R_{q0} = \max_i \{y_{qi}\} - y_{q0}$ ,  $q = 1, \dots, Q$  y  $R_{k0} = x_{k0} - \min_i \{x_{ki}\}$ ,  $k = 1, \dots, K$ , que son definidos como *el rango de posibles mejoras de la unidad 0*. La eficiencia RDM puede calcularse sencillamente haciendo  $1 - \beta_0$ , de tal forma que cuando  $\beta_0 = 0$ , la eficiencia

es 1. Este resultado solo será alcanzado cuando los valores observados son también los valores objetivo para los inputs y outputs de la unidad 0, como es el caso del modelo DEA tradicional.

Algunas limitaciones de esta aproximación son que para que la solución sea Pareto-eficiente, deben cumplirse dos condiciones: a)  $\beta_0 = 0$ , y b) todas las restricciones del modelo anterior deben cumplirse en sentido de igualdad (o sea, no existen holguras). Por tanto, en caso de cumplirse la primera, pero no la segunda, entonces el resultado en el modelo RDM implicaría eficiencia débil (véase Silva Portela et al. (2004, página 1115) para más detalles).

### **3.3.3. EFICIENCIA DE COSTES Y ASIGNATIVA**

La eficiencia de costes (CE) indica la habilidad de la unidad de producir outputs al mínimo coste.

Tanto el modelo CCR como BCC se centran en los aspectos técnico-físicos de la producción. Ambos son adecuados cuando no se pueden hacer suposiciones sobre el comportamiento de los objetivos de las empresas como, por ejemplo, la minimización de costes o la maximización de beneficios. Además, ambos modelos pueden resultar útiles si la información del precio y el coste unitarios no están disponibles o su calidad es cuestionable debido, por ejemplo, a importantes errores de medición. Por otro lado, si las funciones objetivo son razonables y si la información sobre precios fiables está disponible, el modelo DEA puede utilizarse para identificar la asignación de recursos eficientes (Cooper et al., 2000).

Antes de obtener la eficiencia de costes para una DMU con múltiples inputs y outputs, primero debemos calcular el coste mínimo de la producción de un output



corriente de una DMU con precios conocidos de los inputs. Para ello, se resolverá siguiente problema de programación lineal, que fue inicialmente formulado por Färe et al. (1985).

La minimización de coste-DEA puede definirse, en una tecnología de rendimientos constantes a escala (CRS), como:

$$\begin{aligned} & \min \sum_{k=1}^K w_{i0k} x_k^0 \\ & \text{sujeto a :} \\ & x_{0k} = \sum_{i=1}^n x_{ik} \lambda_i, k = 1, 2, \dots, K \\ & \sum_{i=1}^n y_{iq} \lambda_i \geq y_{i0q}, q = 1, 2, \dots, Q \\ & \lambda_i \geq 0, i = 1, 2, \dots, n \\ & x_k^0 \geq 0, k = 1, 2, \dots, K \end{aligned} \quad [3.7]$$

donde  $w_{i0k}$  es el precio del input  $k$  para la DMU  $i0$  bajo evaluación.  $x_k^0$  es una variable que, en la solución óptima, proporciona la cantidad de input  $k$  que será empleado por la DMU  $i0$  con la finalidad de producir los outputs actuales al mínimo coste, sujeto a las restricciones tecnológicas impuestas por el conjunto de posibilidades de producción existente.

A partir de la solución óptima  $(x^*, \lambda^*)$  del problema anterior, la eficiencia de costes (CE) o eficiencia económica de la DMU  $i0$  puede ser calculada por la ratio:

$$CE_{i0} = \frac{\sum_{k=1}^K w_{i0k} x_k^{0*}}{\sum_{k=1}^K w_{i0k} x_{i0k}}$$

donde el mínimo coste a los precios actuales  $\left( \sum_{k=1}^K w_{i0k} x_k^{0*} \right)$  se divide por el coste observado  $\left( \sum_{k=1}^K w_{i0k} x_{i0k} \right)$ . La interpretación de CE significa que podemos producir el mismo vector de outputs ahorrando  $(1-CE) \times 100$  del coste efectivamente incurrido.

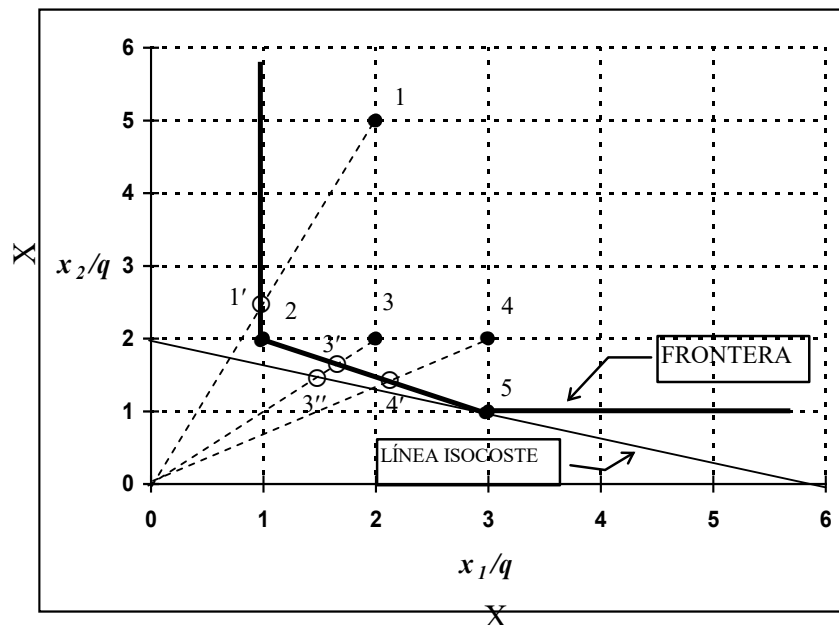
La eficiencia asignativa (AE) - que refleja la habilidad de una unidad para usar el conjunto de inputs en proporciones óptimas, dados sus respectivos precios -, puede obtenerse dividiendo la eficiencia de costes por la eficiencia técnica. Es decir,  $AE = CE / TE$ .

Finalmente, cabe destacar que los rendimientos variables a escala (VRS) pueden considerarse en el problema de minimización de costes añadiendo simplemente la restricción  $\sum_{i=1}^n \lambda_i = 1$ .

La Figura 3.5 (Coelli, 1996) muestra un ejemplo de la frontera de costes de una tecnología que usa dos inputs  $(x_1, x_2)$  para producir un output  $(y)$ , usando para todas las empresas el mismo precio de los inputs:  $w_1 = 1$  y  $w_2 = 3$ . Así, la pendiente de la línea isocoste es  $-1/3$ . Solo el producto 5 es únicamente coste eficiente. Las eficiencias pueden calcularse a partir de las siguientes relaciones:

- La eficiencia técnica de la empresa 3 es la ratio definida por la distancia desde 0 hasta 3' respecto de la distancia desde 0 hasta 3.
- La eficiencia asignativa será la ratio entre la distancia entre 0 y 3'' y la distancia entre 0 y 3'.
- La eficiencia de costes será la ratio entre la distancia de 0 a 3'' respecto de la distancia desde 0 hasta 3.

**Figura 3.5.** Eficiencia de costes en el modelo DEA-CRS usando la isocuanta convexa lineal a trozos y la línea de isocostes (Coelli, 1996).



### 3.3.3.1. UNA VARIACIÓN DEL MODELO DEA DE COSTES

En los modelos DEA tradicionales de eficiencia de costes (Färe et al., 1985), se supone que los precios de los inputs son los mismos en todas las unidades de toma de decisión (DMU). Sin embargo, los mercados no funcionan necesariamente bajo competencia perfecta y los precios de los inputs pueden no ser idénticos en todas las DMU. Así, como señaló Tone (2002), el modelo tradicional DEA de coste-eficiencia definido por la ecuación [3.7] no tiene en cuenta el hecho de que los costes pueden obviamente disminuirse, reduciendo los precios de los factores de los inputs.

La idea, siguiendo las explicaciones de Dong et al. (2014) y considerando que las DMU son bancos, es que, si dos bancos tienen los mismos inputs y outputs, mientras que los precios unitarios de los inputs para un banco son el doble que los del otro banco, entonces los costes totales del banco con los mayores precios unitarios de inputs serán mayores que los del banco con los menores precios unitarios de inputs. Sin embargo,

según el modelo tradicional DEA, la función de costes es homogénea de grado uno en los precios de los inputs y el factor de escala se anula en el coeficiente de eficiencia de costes y, por tanto, se asignará a los dos bancos la misma medida de eficiencia de costes independientemente del hecho de que tienen precios de inputs significativamente diferentes. Esto representa un grave inconveniente para evaluar los niveles de eficiencia relativa bajo el modelo tradicional DEA y es causado por la estructura peculiar del modelo DEA que se centra exclusivamente en la eficiencia técnica de los dos bancos y no puede tener en cuenta las variaciones en los precios unitarios de inputs entre los bancos. Por lo tanto, para evitar esta deficiencia, Tone (2002) propuso un nuevo esquema para evaluar la eficiencia de costes bajo el cual la tecnología de producción es homogénea de grado uno en los costes totales, a diferencia del modelo tradicional DEA que implica homogeneidad de grado uno en los precios de los inputs. Esto significa que bajo el nuevo modelo DEA, los bancos con diferentes precios de inputs tendrán diferentes medidas de eficiencia de costes.

Por ello, Tone (2002) planteó el siguiente problema de programación lineal para rendimientos constantes a escala con el que determinar los niveles de eficiencia de costes, tal que:

$$\begin{aligned}
 & \min e\bar{x}_{0k} \\
 & \text{sujeto a :} \\
 & \sum_{i=1}^n \bar{x}_{ik} \lambda_i \leq e\bar{x}_{0k}, k = 1, 2, \dots, K \\
 & \sum_{i=1}^n y_{iq} \lambda_i \geq y_{i0q}, q = 1, 2, \dots, Q \\
 & \lambda_i \geq 0, i = 1, 2, \dots, n \\
 & x_k^0 \geq 0, k = 1, 2, \dots, K
 \end{aligned}$$

donde  $e \in \mathfrak{R}^K$  es un vector fila de unos y  $\bar{x}_{ik} = (w_{i1}x_{i1}, \dots, w_{ik}x_{ik})'$  es un vector de coste de los inputs en el que los elementos están en unidades homogéneas en términos monetarios. Ahora, la eficiencia de costes se calcularía como:

$$CE = \frac{e\bar{x}^{0*}}{e\bar{x}}$$

siendo  $\bar{x}^{0*}$  la combinación óptima de costes para producir  $y_0$ .

### 3.3.4. CAMBIO DE PRODUCTIVIDAD EN EL TIEMPO

Las medidas globales de productividad han gozado de gran interés entre los investigadores que analizan el desempeño de las empresas y, en los últimos años, las aplicaciones que utilizan los índices Malmquist se han hecho muy comunes en el análisis de la productividad (Orea, 2000).

Este índice, que fue propuesto por primera vez en Caves et al. (1982), tiene la ventaja sobre los índices tradicionales de productividad de Fisher y Törnqvist que puede descomponerse en cambios técnicos y cambios de eficiencia técnica usando sólo información de cantidades. Así, el cambio en el índice de productividad total de los factores (TFP) de Malmquist puede descomponerse en dos componentes: uno midiendo el cambio técnico y otro midiendo el cambio de la frontera. Esto es, el índice de Malmquist permite la separación del efecto debido al cambio de eficiencia (TEC, es decir, los cambios en el tiempo de la eficiencia técnica de cada DMU en el tiempo con respecto a la frontera de mejores prácticas (*best practice frontier*), del cambio

tecnológico (TC) (por ejemplo, el cambio de la frontera de mejores prácticas en el tiempo debido al progreso tecnológico).<sup>10</sup>

El modelo DEA basado en el índice TFP de Malmquist mide el cambio de productividad en el tiempo.

### 3.3.4.1. ÍNDICE DE PRODUCTIVIDAD DE MALMQUIST

El índice de productividad de Malmquist hace uso de las funciones distancia,  $D(y, x)$  - donde  $y$  es un vector de outputs y  $x$  es un vector de inputs -, para medir el cambio de productividad. Éste índice se puede definir utilizando funciones de distancia orientadas hacia el input u output. Así, en una medida con orientación input:

$D_I(y, x) = \min \{ \theta : x/\theta \in L(y) \}$ , mientras que en una medida con orientación output:

$D_O(y, x) = \max \{ \phi : y/\phi \in P(x) \}$ .

Siguiendo a Färe et al. (1994), el índice de cambio de productividad de Malmquist se puede definir en el modelo de orientación-output como:

$$M_O(x_{t+1}, y_{t+1}, x_t, y_t) = \sqrt{\frac{D_O^t(x_{t+1}, y_{t+1})}{D_O^t(x_t, y_t)} \times \frac{D_O^{t+1}(x_{t+1}, y_{t+1})}{D_O^{t+1}(x_t, y_t)}} \quad [3.8]$$

donde  $M_O(x_{t+1}, y_{t+1}, x_t, y_t)$  representa la productividad en el punto de producción  $(x_{t+1}, y_{t+1})$  que usa la tecnología en  $t+1$ , relativo al punto de producción  $(x_t, y_t)$  que usa la tecnología en  $t$ , y donde  $D_O^t(x_t, y_t)$ ,  $D_O^{t+1}(x_t, y_t)$ ,  $D_O^t(x_{t+1}, y_{t+1})$  y  $D_O^{t+1}(x_{t+1}, y_{t+1})$  son

---

<sup>10</sup> Recientemente, Diewert y Fox (2010) extienden los resultados del trabajo de Caves et al. (1982) para demostrar cómo un índice de productividad estándar de Törnqvist, derivado de un índice teórico de Malmquist, se puede descomponer en cambios técnicos y componentes de los rendimientos a escala bajo competencia imperfecta.

cuatro expresiones de las denominadas funciones distancia (que son también eficiencias técnicas,  $\theta$ ), y que se calculan para distintos periodos de tiempo. Un valor mayor que la unidad del índice TFP indicará un crecimiento positivo desde el periodo  $t$  al periodo  $t+1$ . Una breve reseña sobre estas funciones puede verse en el epígrafe 3.1 del Capítulo 2.

La expresión [3.9] permite construir dichos índices en una orientación-input.<sup>11</sup> Así,

$$M_O(y_{t+1}, x_{t+1}, y_t, x_t) = \sqrt{\frac{D_{IO}^t(y_{t+1}, x_{t+1})}{D_O^t(y_t, x_t)} \times \frac{D_O^{t+1}(y_{t+1}, x_{t+1})}{D_O^{t+1}(y_t, x_t)}} \quad [3.9]$$

La expresión [3.9] puede descomponerse como sigue:

$$M_O(y_{t+1}, x_{t+1}, y_t, x_t) = \frac{D_O^{t+1}(y_{t+1}, x_{t+1})}{D_O^t(y_t, x_t)} \sqrt{\frac{D_O^t(y_{t+1}, x_{t+1})}{D_O^{t+1}(y_{t+1}, x_{t+1})} \times \frac{D_I^t(y_t, x_t)}{D_I^{t+1}(y_t, x_t)}} \quad [3.10]$$

donde:

- $EFFCH = \frac{D_O^{t+1}(y_{t+1}, x_{t+1})}{D_O^t(y_t, x_t)}$ , representa el cambio de eficiencia, y
- $TECHCH = \sqrt{\frac{D_O^t(y_{t+1}, x_{t+1})}{D_O^{t+1}(y_{t+1}, x_{t+1})} \times \frac{D_I^t(y_t, x_t)}{D_I^{t+1}(y_t, x_t)}}$ , representa el cambio técnico.

Recordando que  $M$  indica el grado de cambio de productividad, entonces si  $M > 1$  se producen ganancias de productividad, mientras que si  $M < 1$  entonces se producen pérdidas de productividad. La eficiencia técnica aumenta (disminuye) sí y sólo sí  $EFFCH$  es más grande (menor) que uno. Una interpretación del cambio técnico es que el progreso técnico (regresión) se ha producido si  $TECHCH$  es mayor (menor) que uno.

También se puede hacer una evaluación de las principales fuentes de ganancias/pérdidas de productividad comparando los valores de  $EFFCH$  y  $TECHCH$ .

---

<sup>11</sup> Berg et al. (1992), Fukuyama (1995) y Worthington (1999) han empleado una especificación idéntica para medir la eficiencia y productividad de la banca.

Si  $EFFCH > TECHCH$ , los aumentos de productividad son en gran medida el resultado de mejoras en la eficiencia, mientras que si  $EFFCH < TECHCH$  entonces las ganancias de productividad son principalmente el resultado del progreso tecnológico.

La expresión [3.10] puede reformularse considerando que el cambio de eficiencia puede descomponerse en cambio de eficiencia pura y cambio de escala. Färe et al. (1994) reformulan la expresión [3.9] considerando el cambio técnico, pero, también, el cambio de eficiencia pura y cambio de escala. Concretamente, siguiendo a Färe et al. (1994, página 75), y adaptándolo a la orientación-output, tenemos que:

$$M_o(x_{t+1}, y_{t+1}, x_t, y_t) = TECHCH \times PEFFCH \times SCH \quad [3.11]$$

donde  $PEFFCH$  representa el cambio puro de eficiencia y  $SCH$  representa el cambio de escala. Los componentes de cambio de escala y de cambio de eficiencia pura son descomposiciones del cambio de eficiencia utilizando la tecnología de rendimientos variables a escala (VRS), que supone incluir la restricción de convexidad,  $\sum_{j=1}^n \lambda_j = 1$ , en las expresiones de los cuatro problemas formulados en el epígrafe siguiente.

Así, para descomponer el cambio de eficiencia técnica, tenemos que:  $EFFCH = PEFFCH \times SCH$ . Recordemos que el término  $EFFCH$  se refiere al cambio de eficiencia calculado bajo rendimientos constantes a escala (CRS), y  $PEFFCH$  es el cambio de eficiencia calculado bajo rendimientos variables a escala (VRS).<sup>12</sup> En este caso, las expresiones de los cambios en la eficiencia pura y de escala pueden escribirse como sigue:

---

<sup>12</sup> Por ejemplo, el software DEAP (Coelli, 1996), permite la obtención del índice relativo al cambio técnico de eficiencia ( $TEC$ , respectivo a la tecnología CRS), el cambio tecnológico (TC), el cambio de eficiencia técnica pura ( $PTEC$ , obtenido en relación con la tecnología VRS), el cambio de eficiencia de escala ( $SEC$ ). Finalmente, también puede obtenerse el índice de cambio de productividad total de factores ( $TFPC$ ).



$$\begin{aligned}
 - \quad PEFFCH &= \frac{D_{O,VRS}^{t+1}(y_{t+1}, x_{t+1})}{D_{O,CRS}^t(y_t, x_t)}. \\
 - \quad SCH &= \frac{D_{O,VRS}^{t+1}(y_{t+1}, x_{t+1})/D_{O,CRS}^{t+1}(y_{t+1}, x_{t+1})}{D_{O,VRS}^{t+1}(y_t, x_t)/D_{O,CRS}^{t+1}(y_t, x_t)} \times \frac{D_{O,VRS}^t(y_{t+1}, x_{t+1})/D_{O,CRS}^t(y_{t+1}, x_{t+1})}{D_{O,VRS}^t(y_t, x_t)/D_{O,CRS}^t(y_t, x_t)}.
 \end{aligned}$$

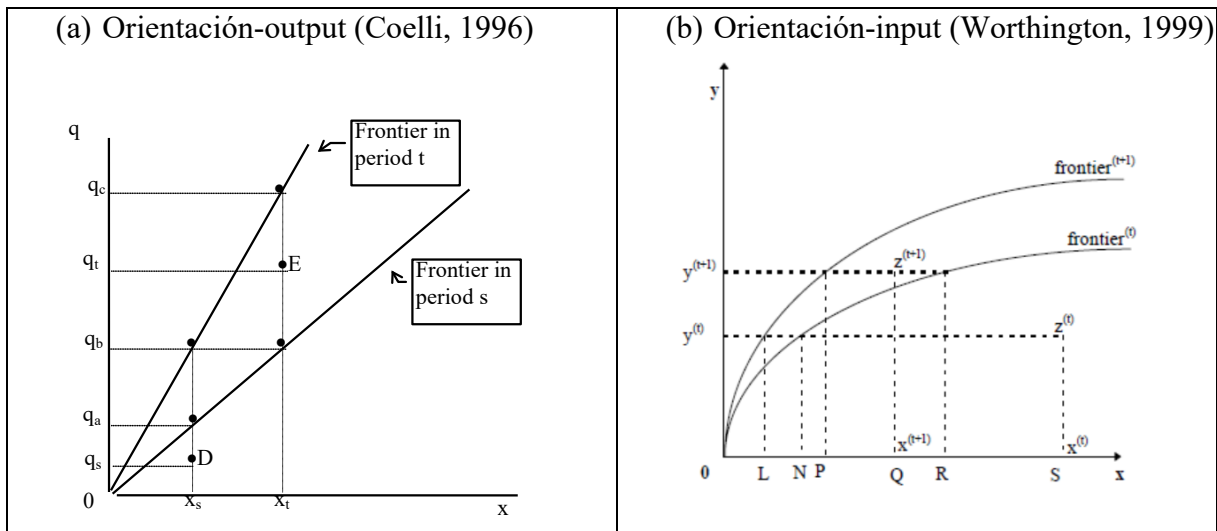
En resumen, las distancias (o eficiencias técnicas) necesarias para calcular el índice de Malmquist corresponden a la frontera DEA-CRS de períodos anteriores, la frontera DEA-CRS de los períodos actuales, la frontera DEA-CRS de los periodos siguientes y la frontera DEA-VRS de los períodos actuales.

### 3.3.4.2. ILUSTRACIÓN GRÁFICA

La Figura 3.6 muestra una representación de dos tecnologías según sea la orientación-input u output. Una de las tecnologías se encuentra en el periodo  $t$  y otra en el periodo  $s$  ( $t+1$ ). En el eje de ordenadas aparece la cantidad de producto ( $q$  o  $y$ ) y en el eje de abscisas aparece la cantidad de input. Atendiendo a la orientación-output, Figura 3.6(a), las tecnologías son CRS en ambos casos. A partir de dicho gráfico pueden computarse el cambio en eficiencia y el cambio técnico, tal que:

$$\begin{aligned}
 - \quad \text{Cambio en eficiencia} &= EFFCH = \frac{q_t/q_c}{q_s/q_a}. \\
 - \quad \text{Cambio técnico} &= TECHCH = \left[ \frac{q_t/q_b}{q_t/q_c} \times \frac{q_s/q_a}{q_s/q_b} \right]^{0.5}.
 \end{aligned}$$

**Figura 3.6.** Cambio en eficiencia y cambio técnico según orientación-output e input.



En la Figura 3.6(b), se muestra el cambio en una orientación-input. Considerando las combinaciones de input/output  $z(t)$  y  $z(t+1)$  en dos periodos de tiempo distintos, puede comprobarse como la ineficiencia en la tecnología del periodo  $t$  es  $0N/0S$ , mientras que la ineficiencia en el periodo  $t+1$  es  $0P/0Q$ . Si comparamos la combinación producida con la tecnología del periodo  $t$  con la de  $t+1$ , la ineficiencia comparable con la del periodo  $t$  es  $0R/0Q$ , que es mayor que 1.

### 3.3.4.3. OBTENCIÓN DE LAS EFICIENCIAS

Atendiendo a la orientación-output con rendimientos constantes a escala (CRS), se pueden resolver los siguientes problemas DEA:

#### **Problema 1**

$$\left[ D_O^t(x_t, y_t) \right]^{-1} = \max_{\phi, \lambda} \phi$$

sujeto a :

$$-\phi y_{0q,t} + \sum_{i=1}^n y_{iq,t} \lambda_i \geq 0$$

$$x_{0k,t} - \sum_{i=1}^n x_{ik,t} \lambda_i \geq 0$$

$$\lambda_i \geq 0, \forall i$$

#### **Problema 2**

$$\left[ D_O^{t+1}(x_{t+1}, y_{t+1}) \right]^{-1} = \max_{\phi, \lambda} \phi$$

sujeto a :

$$-\phi y_{0q,t+1} + \sum_{i=1}^n y_{iq,t+1} \lambda_i \geq 0$$

$$x_{0k,t+1} - \sum_{i=1}^n x_{ik,t+1} \lambda_j \geq 0$$

$$\lambda_i \geq 0, \forall i$$

#### **Problema 3**

$$\left[ D_O^t(x_{t+1}, y_{t+1}) \right]^{-1} = \max_{\phi, \lambda} \phi$$

sujeto a :

$$-\phi y_{0q,t+1} + \sum_{i=1}^n y_{iq,t} \lambda_i \geq 0$$

$$x_{0k,t+1} - \sum_{i=1}^n x_{ik,t} \lambda_i = 0$$

$$\lambda_i \geq 0, \forall i$$

#### **Problema 4**

$$\left[ D_O^{t+1}(x_t, y_t) \right]^{-1} = \max_{\phi, \lambda} \phi$$

sujeto a :

$$-\phi y_{0q,t} + \sum_{i=1}^n y_{iq,t+1} \lambda_i \geq 0$$

$$x_{0k,t} - \sum_{i=1}^n x_{ik,t+1} \lambda_i \geq 0$$

$$\lambda_i \geq 0, \forall i$$

donde todos los sub-índices 0,  $t$  indican que la DMU está siendo evaluada en el periodo  $t$ , y  $\phi$  es la eficiencia. Los valores de  $\phi$  y  $\lambda$  pueden ser diferentes.

Por último, cabría destacar que, bajo determinadas condiciones restrictivas, el índice de productividad de Malmquist puede también calcularse a partir de datos de precios y cantidades. Caves et al. (1982) muestran, bajo determinadas condiciones, que el índice de Törnqvist es equivalente a la media geométrica de dos índices de productividad de

Malmquist para la tecnología translog. Färe y Grosskopf (1992) muestran que el índice de Fisher es igual a un índice de Malmquist bajo rendimientos constantes a escala y minimización de costes competitiva. Grifell y Lovell (1999), considerando rendimientos constantes a escala, demuestran que la media geométrica de dos índices generalizados de productividad de Malmquist es igual al índice de productividad de Törnqvist.

En este punto, conviene destacar algunas limitaciones del método. Primero, como destaca Orea (2000), Grifell and Lovell (1995) muestran que el índice de productividad de Malmquist introducido por Caves et al. (1982) no provee una medida precisa del cambio de productividad porque ignora la potencial contribución de las economías de escala al cambio de productividad. Para evitar este problema, Balk (1999) propone varias medidas de cambio de productividad, abarcando no solo al cambio técnico y el cambio de eficiencia técnica, sino que también el cambio en la eficiencia de escala para una empresa con múltiples inputs y outputs. Grifell y Lovell (1999) sugieren usar un índice generalizado de productividad de Malmquist, que puede ser construido como el producto del índice de productividad de Malmquist convencional y un índice de escala.

Y, segundo, O'Donnell (2009) demuestra que los índices completos satisfacen importantes axiomas de la teoría de los números índice. Él define formalmente lo que se entiende por completitud y demuestra que la clase de índices TFP completos incluye los índices Törnqvist, Konus y Moorsteen-Bjurek, y que el índice TFP de Caves et al. (1982) es incompleto, no satisfaciendo por tanto todos los axiomas de los números índice.

### 3.3.5. METAFRONTERAS DE PRODUCCIÓN USANDO DEA

Siguiendo a O'Donnell et al (2008), la estimación de la metafrontera puede realizarse recurriendo al siguiente modelo DEA que, en la orientación-ouput y un panel de datos, puede escribirse como:

$$\max_{\phi_{it}, \lambda_{it}} \phi_{it}, i = 1, \dots, n, t = 1, \dots, T$$

sujeto a :

$$-\phi y_{it} + Y_{it} \lambda_{it} \geq 0$$

$$x_{it} - X_{it} \lambda_{it} \geq 0$$

$$\lambda_{it} \geq 0, \forall it$$

donde  $\phi_{it}$  es un escalar para la  $i$ -ésima empresa en el  $t$ -ésimo periodo que representa la eficiencia técnica respecto de la metafrontera ( $\hat{TE}_{it}$ ),  $y_{it}$  es un vector  $Q \times 1$  de outputs,  $x_{it}$  es un vector  $K \times 1$  de inputs,  $Y_{it}$  es una matriz de outputs de orden  $Q \times n$ ,  $X_{it}$  es una matriz de inputs de orden  $K \times n$ , y  $\lambda_{it}$  es un vector de ponderaciones  $n \times 1$ .

Para estimar las fronteras específicas de grupo, simplemente debemos considerar la expresión del modelo anterior, pero incluyendo solamente las observaciones del grupo  $j$ -ésimo. Este proceso debería repetirse  $J$  veces, con los  $J$  grupos. Dichas eficiencias podemos escribirlas como  $\hat{TE}_{it}^j, j = 1, \dots, J$ .

Una vez hemos obtenido las eficiencias técnicas específicas de grupo y las correspondientes a la metafrontera, podremos calcular la ratio del gap tecnológico a través de la siguiente expresión:

$$TGR_{it}^j = \frac{\hat{TE}_{it}}{\hat{TE}_{it}^j}$$

### 3.4. RELAJANDO EL SUPUESTO DE CONVEXIDAD: *FREE DISPOSAL HULL*

#### (FDH)

Farrell (1959) indicó como posibles violaciones de la convexidad la indivisibilidad de los inputs y outputs, las economías de escala y la especialización.

El modelo *Free Disposal Hull* (FDH) es un método no paramétrico determinista para medir la eficiencia de las unidades de producción (DMUs) que fue conceptualizado, formulado y desarrollado por Deprins et al. (1984) y ampliado por Lovell et al. (1994). El aspecto básico del modelo FDH es que relaja el supuesto de convexidad del modelo DEA, de tal manera que el único supuesto que se requiere es la monotonía de la tecnología. Mientras que DEA asume una tecnología convexa y emplea programación lineal para envolver los datos, FDH se basa en el principio de dominancia débil y se aleja de la suposición de convexidad inherente en DEA. Es decir, FDH envuelve los datos con un casco en escalera (*staircase-hull*) no convexo (ver Cooper et al. (2000) para una discusión exhaustiva de DEA y FDH).

La ventaja que la tecnología definida por FDH no sea convexa es la flexibilidad (véase Deprins et al., 1984). Esta parece una propiedad atractiva de FDH ya que a menudo es difícil encontrar una buena justificación teórica o empírica para postular una producción convexa en el análisis de eficiencia.

A partir de la expresión [3.6] correspondiente al supuesto VRS en la orientación-input - eliminando los *slacks* de la formulación por simplicidad -, las eficiencias FDH pueden estimarse resolviendo el siguiente problema de programación entera, tal que:

$$\min_{\theta, \lambda} \theta$$

sujeto a :

$$y_{0q} \leq \sum_{i=1}^n y_{iq} \lambda_i$$

$$\theta x_{0k} - \sum_{i=1}^n x_{ik} \lambda_i \geq 0$$

$$\sum_{i=1}^n \lambda_i = 1, \lambda_i \in \{0,1\}, \forall i$$

En el método FDH se añade la restricción  $\lambda_i \in \{0,1\}$ , es decir,  $\lambda_i$  solamente puede tomar dos valores, o el cero o el uno. Esto hace que el problema se convierta en un programa de optimización entera.

Otra forma de plantear el problema de optimización es considerando la maximización de las holguras, tal que:

$$\max \sum_{q=1}^Q s_q^+ + \sum_{k=1}^K s_k^-$$

sujeto a :

$$y_{0q} \leq \sum_{i=1}^n y_{iq} \lambda_i$$

$$\theta x_{0k} - \sum_{i=1}^n x_{ik} \lambda_i \geq 0$$

$$\sum_{i=1}^n \lambda_i = 1, \lambda_i \in \{0,1\}, \forall i$$

Cabe destacar que si el conjunto de producción verdadero es convexo, entonces los métodos DEA y FDH proporcionan estimadores consistentes. Sin embargo, FDH muestra una menor tasa de convergencia (debido a los menores supuestos que requiere) con respecto a DEA. Por el contrario, si el verdadero conjunto de producción no es convexo, entonces DEA no es un estimador consistente del conjunto de producción, mientras que FDH es consistente.

FDH puede ser económicamente más significativo que el caso monótono convexo DEA, también en condiciones económicas alternativas de trascendencia. Por lo tanto,

las medidas de eficiencia técnica de FDH siguen siendo significativas para las teorías de la empresa que permiten una competencia imperfecta o incertidumbre.

### 3.4.1. LA PRÁCTICA COMPUTACIONAL DE FDH

En la práctica, el estimador FDH se calcula mediante un procedimiento simple de comparación vectorial que equivale a un algoritmo de enumeración completo propuesto en Tulkens (1993).

Tulkens (1993) desarrolló un algoritmo que facilita la rápida obtención de la solución óptima. Formalmente, en el caso de la orientación-input, la eficiencia se estima comparando cada DMU,  $i = 1, \dots, n$ , con todas las otras DMU,  $j = 1, \dots, n$ , en los datos que producen al menos tanto de cualquier output como la DMU  $i$ . El conjunto de DMU iguales en la muestra que satisface la condición  $y_{jl} \geq y_{il}, \forall l$  podemos denotarlo como  $D_i$ . Entre las DMUs iguales, la única que exhibe el consumo mínimo de inputs sirve como referencia para  $i$ , y el estimador  $\hat{\theta}_i^{FDH}$  se calcula como input relativo tal que:

$$\hat{\theta}_i^{FDH} = \min_{i=1, \dots, n / y_{jl} \geq y_{il}, \forall l} \left\{ \max_{k=1, \dots, K} \left( \frac{x_{jk}}{x_{ik}} \right) \right\}$$

Es un procedimiento maximin (para la orientación input) basado en el principio de dominancia débil: la parte "max" del algoritmo identifica las DMU más dominantes con respecto a las cuales se evalúa una DMU determinada. Una vez que se identifican las DMU más dominantes, las holguras se calculan a partir de la parte "min" del algoritmo.

En general, una de las principales limitaciones de este estimador es que es altamente sensible a la existencia de outliers y errores de medida, al igual que el estimador DEA.



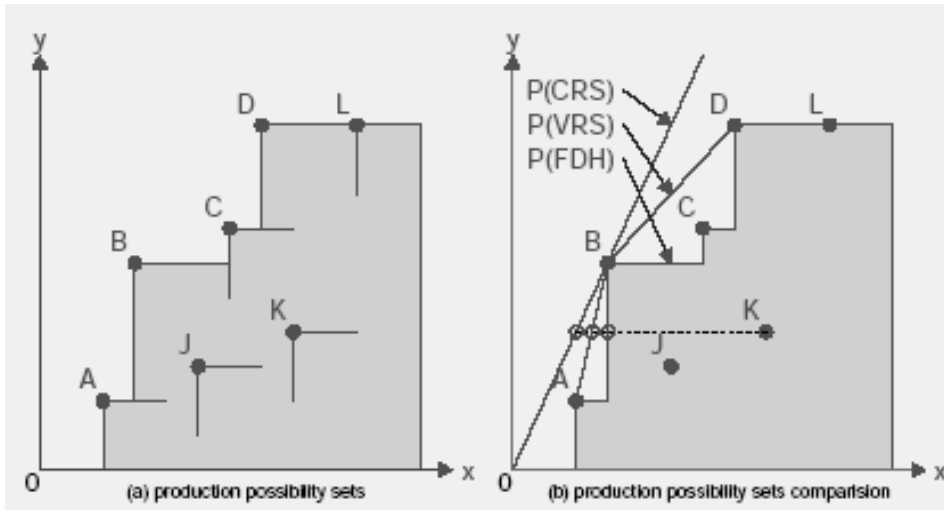
Más concretamente, a la influencia de valores atípicos "supereficientes". Esto es una consecuencia del hecho de que la frontera eficiente está determinada por observaciones muestrales que son puntos extremos. Simar (1996) señala la necesidad de identificar y eliminar valores atípicos al usar modelos no paramétricos. Si no se pueden identificar, se recomienda el uso de modelos de frontera estocástica.

### 3.4.2. UNA ILUSTRACIÓN GRÁFICA

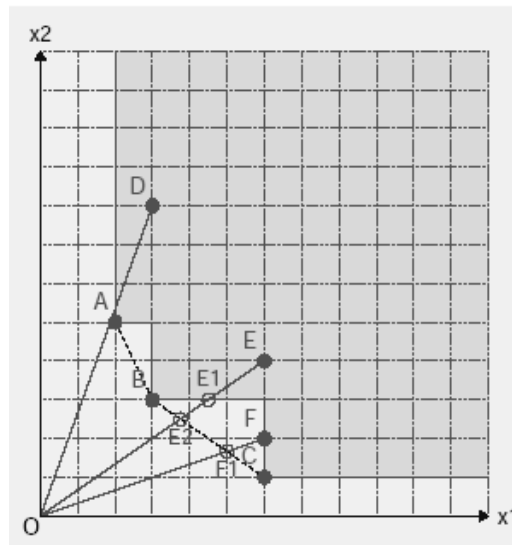
Las Figuras 3.7 y 3.8 muestran la representación de las fronteras FDH en comparación con las obtenidas por DEA. Siguiendo a Ji y Lee (2010), la Figura 3.6 indica que si cualquier par  $(x, y)$  pertenece al conjunto de posibilidades de producción ( $P$ ), entonces el par  $(tx, ty)$  pertenece a  $P$  para cualquier escalar positivo  $t$ . Esta propiedad representa los rendimientos constantes a escala (CRS), supuesto que puede modificarse permitiendo la existencia de rendimientos variables a escala (VRS). El conjunto de posibilidades de producción del modelo FDH se obtiene definiéndolo de forma diferente a los modelos CRS y VRS. En los modelos CRS y VRS, si  $(x_1, y_1)$  y  $(x_2, y_2)$  pertenecen al conjunto de posibilidades de producción, entonces  $(a(x_1 + x_2), b(y_1 + y_2))$  con cualquier escalar positivo  $a, b$  también se considera que están en el mismo conjunto de posibilidades de producción. Este axioma se denomina convexidad.

La disposición libre significa que, si se puede conseguir un par específico de inputs y outputs, también se pueden conseguir pares de más inputs y menos outputs. El modelo FDH permite la disponibilidad libre para construir el conjunto de posibilidades de producción. En consecuencia, la línea de frontera para el modelo FDH se construye a partir de los inputs y outputs observados que permiten dicha disponibilidad. Esta frontera es no convexa.

**Figura 3.7.** Conjunto de posibilidades de producción en CRS, VRS y FDH (Ji and Lee, 2010).



**Figura 3.8.** Medidas de eficiencia FDH en la orientación-input (Ji and Lee, 2010).



En la Figura 3.8, la forma del conjunto de posibilidades de producción en FDH es escalonada. Las fronteras DEA-VRS y FDH se presentan considerando dos inputs y un output para 6 unidades de producción etiquetadas de A a F. En el modelo DEA-VRS (segmentos AB y BC), las DMU A, B y C son eficientes, pero en el modelo FDH (área escalonada y sombreada) son eficientes A, B, C y F. Si quisiéramos obtener la eficiencia

de la observación E en ambos modelos tendríamos que hacer lo siguiente. En el modelo DEA-VRS se define como  $\theta E$ , es decir,  $OE_2 / OE$ , mientras que en el modelo FDH se define como  $\theta E$ , es decir,  $OE_1 / OE$ .

Las puntuaciones de eficiencia en el modelo FDH también están entre 0 y 1. Sin embargo, respecto de las estimaciones DEA-CRS y DEA-VRS existen diferencias. Por ejemplo, bajo la orientación-input, las puntuaciones de la eficiencia del modelo FDH son siempre mayores que las puntuaciones obtenidas en el modelo VRS. Además, las puntuaciones de la eficiencia del modelo VRS son siempre mayores que las del modelo CRS. De esta manera, puede escribirse que:

$$\theta_{FDH,input} > \theta_{VRS,input} > \theta_{CRS,input}$$

En otras palabras, el conjunto de posibilidades de producción del modelo FDH es un subconjunto del modelo VRS y del modelo CRS.

### 3.5. MÉTODOS NO PARAMÉTRICOS ROBUSTOS

En esta subsección ilustraremos brevemente los enfoques de frontera parcial que generalizan al método FDH y que son métodos no paramétricos estocásticos de fronteras eficientes.

En general, los enfoques no paramétricos han sido criticados por los econométricos por ser enfoques deterministas, carecer de un proceso de generación de datos bien definido y, más importante, por ser extremadamente vulnerables a los valores atípicos y al error de medición.

Sin embargo, la última objeción antes comentada sobre la medición de la eficiencia no paramétrica se ha abordado recientemente mediante los llamados enfoques de frontera parcial.

Dentro de las propuestas de frontera parcial se incluyen los estimadores *order-m* (Cazals et al., 2002) y *order-a* (Aragon et al., 2005; Daouia y Simar, 2007). Ambas aproximaciones se basan en la definición probabilística de la tecnología  $T = \{(y, x) : x \text{ puede producir } y\}$ . Por ejemplo, en la orientación-output, y para el caso de un vector de inputs  $x_i$  y un vector de outputs  $y_i$  en la DMU  $i$ , la probabilidad de ser dominado es  $H(y_i, x_i) = P\{Y \geq y_i, X \leq x_i\}$ , siendo  $(X, Y)$  variables aleatorias, la cual representa la probabilidad de producir más outputs con menos inputs que la DMU  $i$ .

Las características más relevantes de ambos métodos son varias. Primero, que ambas propuestas proporcionan estimadores robustos a la existencia de valores extremos, cosa que no hacen los métodos DEA o FDH que hemos comentado. Segundo, estas metodologías permiten localizar a unidades super-eficientes detrás de la frontera de posibilidades de producción. Y, tercero, estas fronteras parciales envuelven una submuestra de los datos. No obstante, aunque estos métodos recientemente están ganando una aceptación más amplia, también han presentado algunas críticas (Krüger, 2012).

### **3.5.1. ESTIMADORES ORDER-M**

Los estimadores *order-m* se caracterizan porque agregan una "capa de aleatoriedad" a la estimación de las puntuaciones de eficiencia FDH.

Cazals et al. (2002) introducen el concepto de "frontera de inputs mínimos esperada de *order-m*" (orientación-input) y "frontera de output máximo esperada de *order-m*" (orientación-output); donde la suposición de convexidad puede relajarse y también puede existir ruido con valor esperado cero (este último no permitido en DEA o FDH). Por ejemplo, en la orientación-output y para cualquier  $y$ , el nivel de output máximo esperado se puede definir como  $y_m^{\hat{o}}(x) = \phi_m y$ , donde  $\phi_m$  es la puntuación de eficiencia media condicionada a  $X \leq x_i$  en dicha orientación. De esta forma, la versión *order-m* del conjunto de posibilidades de producción:  $T = \{(y, x) : x \text{ puede producir } y\}$ , definido en el Capítulo 2, se convierte en  $T_m = \{(y, x) / (y, x) \in T, y \leq y_m^{\hat{o}}(x)\}$ , la cual es la frontera *order-m*. En el caso de la orientación-input, la frontera *order-m* es  $T_m = \{(y, x) / (y, x) \in T, x \leq x_m^{\hat{o}}(y)\}$ , siendo  $x_m^{\hat{o}}(y) = \theta_m x$  donde  $\theta_m$  es la eficiencia media condicionada a  $Y \geq y_i$ .

Para obtener dichas fronteras, el método *order-m* requiere dibujar  $m$  variables aleatorias i.i.d. a partir de una distribución condicional a los valores de la DMU  $i$ . Cazals et al. (2002) sugieren que puede usarse una técnica simple de Monte Carlo para obtener estimaciones no paramétricas de las medias condicionadas. Cabe resaltar que Cazals et al. (2002) definen la función de supervivencia conjunta para obtener dichas eficiencias. Por ejemplo, en la orientación-input, definen  $S_{X/Y}(x_i/y_i) = P(X \geq x_i, Y \geq y_i) / P(Y \geq y_i)$ . Para el caso de la orientación-output, también puede construirse una función condicionada  $q$ -variada tal como:  $F_{Y/X}(y_i/x_i) = P(Y \geq y_i / X \leq x_i)$ .

Atendiendo a los valores que puede tomar  $m$ , podemos decir lo siguiente. Para  $m \rightarrow \infty$ , el estimador *order-m* proporcionaría el mismo resultado que el estimador FDH estándar; en este caso, ambos enfoques generan resultados idénticos. Para valores finitos de  $m$ , sin embargo, el estimador *order-m* no abarca todos los puntos y, por lo tanto, es más robusto para los valores extremos o valores atípicos en los datos.

Siguiendo a Tauchmann (2012), y en términos computacionales, el estimador *order-m* puede aplicarse como sigue (Daraio and Simar 2007, página 72):

1) Del conjunto  $D_i$  se obtienen aleatoriamente  $m$  DMUs iguales con reemplazamiento.

2) En la orientación-input, el procedimiento requiere estimar la pseudo eficiencia

$$\text{definida por } \hat{\theta}_{mi}^{FDH_b} = \min_{i=1, \dots, n / y_{jl} \geq y_{il}, \forall l} \left\{ \max_{k=1, \dots, K} \left( \frac{x_{jk}^b}{x_{ik}} \right) \right\}, \text{ usando la muestra artificialmente}$$

creada para cada  $b$ .

3) Los pasos 1 y 2 se repiten  $B$  veces,  $b=1, \dots, B$ .

4) La eficiencia *order-m* se calcula como media de las puntuaciones pseudo-FDH, tal que:

$$\begin{aligned} \hat{\theta}_{mi}^{Order-m} &= \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{mi}^{FDH_b} \\ &= \frac{1}{B} \sum_{b=1}^B \min_{i=1, \dots, n / y_{jl} \geq y_{il}, \forall l} \left\{ \max_{k=1, \dots, K} \left( \frac{x_{jk}^b}{x_{ik}} \right) \right\} \end{aligned}$$

A causa del remuestreo aleatorio, la DMU  $i$  puede o no estar disponible con su par en cada réplica  $d$ . Por esta razón, la eficiencia *order-m* excede el valor de 1 en el caso de orientación-input. En consecuencia, *order-m* permite DMU supereficientes ubicadas más allá de la frontera estimada de posibilidades de producción. Esta es la diferencia

clave con FDH, donde para una DMU se descarta que el consumo relativo de inputs exceda la unidad.

Calcular la eficiencia *order-m* requiere elegir valores para dos parámetros,  $B$  y  $m$ . Si bien la elección de  $B$  es una pura cuestión de precisión, donde mejorar la precisión se logra a expensas del tiempo de computación prolongado, la elección de  $m$  es crítica. Cuanto más pequeño sea el valor de  $m$ , mayor será la proporción de DMU supereficientes. Como hemos comentado antes, cuando  $m \rightarrow \infty$ , el estimador *order-m* coincide con FDH, y cuando  $m = n$ , aún pueden ocurrir DMU supereficientes.

Las limitaciones son que el uso del remuestro puede llevar mucho tiempo (para grandes conjuntos de datos), y que determinar el valor apropiado para  $m$  puede requerir probar numerosos valores. También, puede ocurrir que no exista una referencia que sirva como única referencia para la DMU  $i$ -ésima. En tal caso, puede determinarse una DMU de pseudo referencia,  $j_i^{ref}$ , como:

$$j_i^{ref} = \arg \min_{i=1, \dots, n / y_{ji} \geq y_{il}, \forall l} \left\{ \max_{k=1, \dots, K} \left( \frac{x_{jk}}{x_{ik}} \right) - \hat{\theta}_{mi}^{order-m} \right\}$$

### 3.5.2. CAMBIO DE PRODUCTIVIDAD EN EL TIEMPO USANDO EL MÉTODO

#### **ORDER-M**

Wheelock y Wilson (2003) señalan que el estimador *order-m* puede ser útil cuando se aplica a los datos del panel para evaluar el cambio de productividad a lo largo del tiempo utilizando el índice de Malmquist.

Como sabemos, el cambio de productividad puede descomponerse en un progreso tecnológico (es decir, el cambio de la frontera), y el *catching-up* o cambio en la

eficiencia individual a lo largo del tiempo. Así, el índice de Malmquist *order-m* para dos periodos ( $t$  y  $t+1$ ) puede obtenerse en su orientación-input como sigue:

$$M_{I,m}(y_{t+1}, x_{t+1}, y_t, x_t) = \underbrace{\frac{D_I(y_{t+1}, x_{t+1}/C(T_m^{t+1}))}{D_I(y_t, x_t/C(T_m^t))}}_{\Delta \text{eficiencia}} \sqrt{\underbrace{\frac{D_I(y_{t+1}, x_{t+1}/C(T_m^t))}{D_I(y_{t+1}, x_{t+1}/C(T_m^{t+1}))} \times \frac{D_I(y_t, x_t/C(T_m^t))}{D_I(y_t, x_t/C(T_m^{t+1}))}}_{\Delta \text{tecnología}}}}$$

donde  $C(T_m^t)$  representa el cono convexo de la tecnología *order-m* en el periodo  $t$ . El

método *order-m* puede emplearse para estimar la distancia de cada unidad a la frontera,

tal que:  $\hat{D}_{I,i} = D_{I,i}(y_i, x_i/C(T_m^t)) = \frac{1}{B} \sum_{b=1}^B \min_{i=1, \dots, n/y_{jl} \geq y_{il}, \forall l} \left\{ \max_{k=1, \dots, K} \left( \frac{x_{jk}^b}{x_{ik}} \right) \right\}$ . Esta expresión

considera convenientemente la comparación de las distancias de los inputs y outputs de las DMU al cono convexo de la tecnología en cada periodo. Cabe destacar que esta estimación no incorpora la restricción VRS.

En cuanto a la interpretación, cabe resaltar que los valores que exceden la unidad implican progreso en productividad, tecnología o eficiencia, mientras que valores menores que la unidad implican retroceso.

Por último, la expresión del cambio en la productividad de *order-m*,  $M_{I,m}$ , puede descomponerse en cuatro partes: cambios en la eficiencia técnica, cambios en la eficiencia de escala, cambios o desplazamientos de la frontera y cambios de eficiencia de escala debidos a desplazamientos en la frontera (véase Wheelock y Wilson, 2003: página 13).



### 3.5.3. ESTIMADORES *ORDER-ALPHA*

La aproximación *order- $\alpha$*  considera la frontera de producción como un cuantil.

Los estimadores *order- $\alpha$*  se caracterizan porque se elige el percentil  $[100-\alpha]$  ( $\alpha$ -ésimo), con  $0 \leq \alpha \leq 100$ , en lugar de un mínimo (máximo) como punto de referencia de eficiencia. No se minimiza el consumo de inputs, pero el percentil  $[100-\alpha]$ -ésimo (entre pares-DMU) sirve como referencia. El modelo FDH representa un caso especial del modelo *order- $\alpha$*  cuando  $\alpha=100$ . Cuando  $\alpha < 100$ , algunas DMU pueden clasificarse como supereficientes y no estar abarcadas por la frontera estimada de posibilidades de producción. Al igual que  $m$  para la eficiencia del modelo *order- $m$* ,  $\alpha$  puede considerarse como un parámetro de ajuste que determina el número de DMU supereficientes. Los valores de  $\alpha$  suelen elegirse en el intervalo  $[90,99]$

En el método *order- $\alpha$*  no es necesario el remuestreo, lo que supone menos consumo de tiempo que *order- $m$* . En este caso, y en la orientación-input, el minimax se construye como:

$$\hat{\theta}_{i,\alpha} = P_{100-\alpha}^{i=1,\dots,n/y_{jl} \geq y_{il}, \forall l} \left\{ \max_{k=1,\dots,K} \left( \frac{x_{jk}}{x_{ik}} \right) \right\}$$

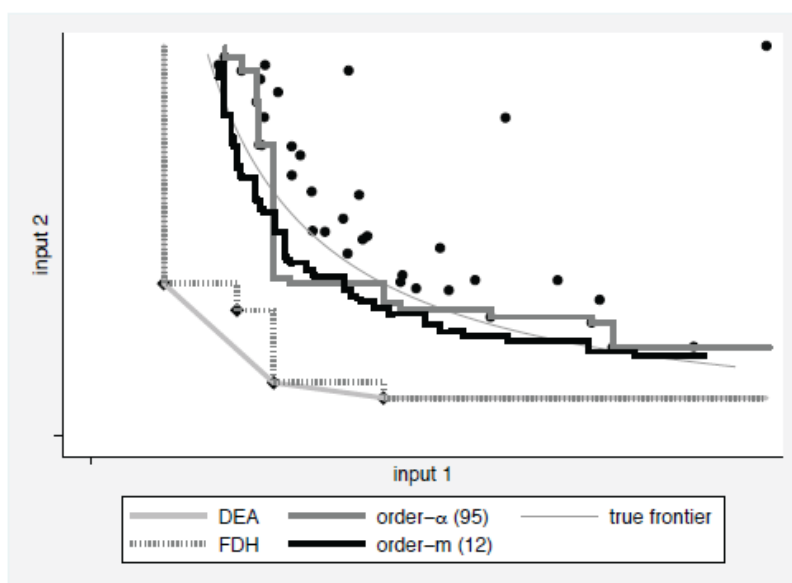
En el caso de la orientación-output, el problema sería:

$$\hat{\theta}_{i,\alpha} = P_{\alpha}^{i=1,\dots,n/x_{jk} \leq x_{ik}, \forall k} \left\{ \max_{q=1,\dots,Q} \left( \frac{y_{jq}}{y_{iq}} \right) \right\}$$

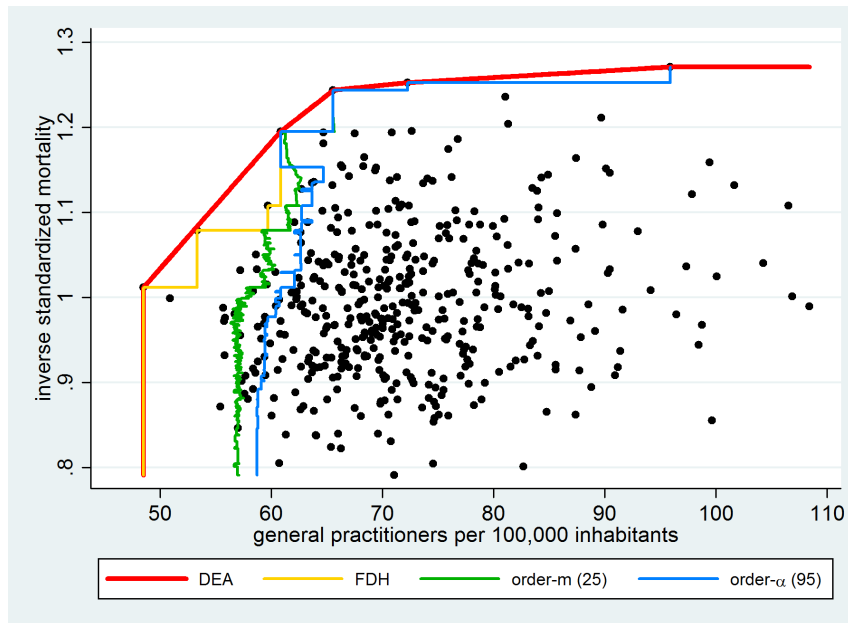
### 3.5.4. ILUSTRACIÓN GRÁFICA DE LOS ESTIMADORES

Las Figuras 3.9 y 3.10 muestran distintas fronteras. La Figura 3.9 muestra las fronteras estimadas por DEA, FDH, *orden- $\alpha$*  y *orden- $m$*  construidas por Tauchmann (2012). Los puntos en las fronteras estimadas se construyen como el consumo de inputs observados escalados por la puntuación de eficiencia estimada. Para DEA y FDH, las DMU irregulares abarcan las fronteras estimadas, lo que hace que el resto de las DMU sean muy ineficientes. De hecho, las observaciones regulares no afectan en absoluto a las fronteras estimadas por DEA y FDH. Por el contrario, los modelos *orden- $\alpha$*  ( $\alpha = 95$ ) y *orden- $m$*  ( $m = 12$ ) permiten que las DMU anormales sean ubicadas fuera de las fronteras estimadas de posibilidades de producción. Para lograr esto, *orden- $\alpha$*  y *orden- $m$*  usan la información en las DMU regulares para estimar la frontera, que a su vez se comparan con un punto de referencia más apropiado.

**Figura 3.9.** Fronteras de posibilidades de producción (isocuantas) estimadas no paramétricamente (Tauchmann, 2012).



**Figura 3.10.** Producción de salud a nivel de distrito en Alemania en 2004. Ilustración DEA, FDH, *order-m* y *order- $\alpha$*  para el caso de un solo input y output (Tauchmann, 2011).



La Figura 3.10 recoge un caso de estudio sobre la relación entre la inversa de la mortalidad estandarizada y el número de médicos por 100.000 habitantes en Bavaria (Alemania) en 2004 (Tauchmann, 2011). Las DMUs son los distritos en Bavaria en el año 2004 (96 observaciones).

### 3.5.5. DETECCIÓN DE OUTLIERS BASADA EN LOS MODELOS DE FRONTERA

#### PARCIAL

El análisis de frontera parcial se puede utilizar para detectar posibles valores atípicos en datos destinados a analizar la eficiencia no paramétrica usando posteriormente DEA o FDH (véase Daraio y Simar (2007), página 79). El enfoque sugerido por Daraio y Simar (2007) se basa en: 1) llevar a cabo una serie de análisis de frontera parciales para diferentes valores de  $\alpha$  o  $m$ ; 2) graficar la proporción de DMU supereficientes contra  $\alpha$  o  $m$ ; y 3) identificar discontinuidades en la curva resultante. Dichas discontinuidades

apuntan a que las DMU son valores atípicos que se clasifican como supereficientes para los valores correspondientes de  $\alpha$  y  $m$ . Este procedimiento también se puede usar para determinar las elecciones apropiadas para  $\alpha$  y  $m$ . El próximo comando Stata8 implementó la detección de valores atípicos orden- $\alpha$ .

Tauchmann (2012), que usa los estimadores *order- $\alpha$* , implementa métodos para detectar outliers siguiendo el procedimiento de Daraio y Simar (2007), y luego estimar la frontera DEA y/o FDH en STATA.

La idea principal que subyace en este procedimiento es que aumentando el valor de  $\alpha$  se reduce el número de DMUs clasificadas como supereficientes, así que en ausencia de valores atípicos la proporción de DMUs supereficientes debería disminuir sin problema.

El procedimiento requiere obtener los puntos de discontinuidad en presencia de valores atípicos. Así, las DMUs clasificadas como supereficientes para  $\alpha \geq \alpha^{disc}$  (punto de discontinuidad) son los outliers o valores atípicos más probables. Con esta información, tales valores atípicos pueden ser excluidos del análisis de eficiencia aplicando FDH o DEA.

El procedimiento sugerido por Daraio y Simar (2007) se basa en los siguientes pasos:

1. Llevar a cabo análisis de una serie de fronteras parciales para diferentes valores de  $\alpha$  o  $m$ ;
2. Dibujar la proporción de DMU supereficientes contra  $\alpha$  o  $m$ ; e
3. Identificar discontinuidades en la curva resultante. Dichas discontinuidades apuntan a que las DMU son valores atípicos que se clasifican como supereficientes para los valores correspondientes de  $\alpha$  y  $m$ . Este

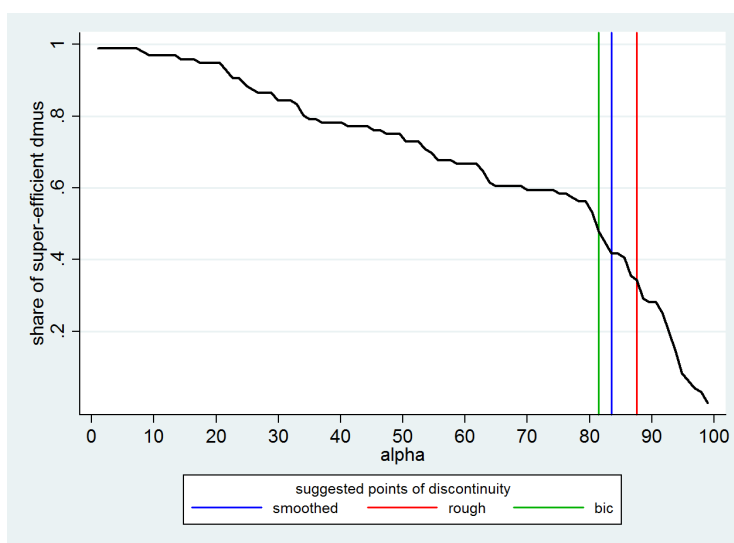
procedimiento también se puede usar para determinar las elecciones apropiadas para  $\alpha$  y  $m$ .

Daraio y Simar (2007) proponen una serie de reglas: dos locales y una global para detectar discontinuidades. Por ejemplo,

- a)  $\alpha$  para el cual la serie dos veces diferenciada toma valor mínimo (siguiendo uno no negativo).
- b) Valores de  $\alpha$  para los cuales los valores negativos persisten después de suavizar repetidamente las series dos veces diferenciadas ejecutando suavizadores medianos espaciados (*odd-spaced median smoothers*) (*smooth*).
- c)  $\alpha$  que minimiza BIC para dividir la serie en dos partes y ajustar funciones lineales (cuadráticas) a cada una.

Gráficamente, la Figura 3.11 muestra un ejemplo de cómo observar los puntos de discontinuidad.

**Figura 3.11.** Puntos de discontinuidad en el modelo de producción regional de salud en Bavaria en 2004 (Tauchmann, 2011).



No obstante, cabría destacar en este punto que una crítica a estos últimos métodos ha sido elaborada por Krüger (2012). Este autor presenta los resultados de un extenso experimento de Monte Carlo para comparar diferentes métodos de análisis de eficiencia. Además de los tradicionales métodos paramétricos deterministas (COLS), no paramétricos (DEA y FDH), paramétrico-estocásticos (SFA), y también métodos no paramétricos-robustos (*order-m* y *order-α*). El diseño experimental comprendió una amplia variedad de situaciones con diferentes regímenes de los rendimientos a escala, elasticidades de sustitución y errores de medida. Sus resultados mostraron que el *ranking* de métodos de eficiencia obtenido sería:

$$\text{SFA} > \text{DEA-CRS} > \text{DEA-VRS} > \text{COLS} > \text{FDH} > \textit{order-m} > \textit{order-}\alpha.$$

es decir, los nuevos métodos robustos no paramétrico-estocásticos no deberían utilizarse sin verificación cruzada con otros métodos como el análisis de frontera estocástica o el análisis envolvente de datos. De hecho, estos últimos métodos parecen bastante sólidos en los experimentos.

### **3.6. DETERMINANTES DE LA EFICIENCIA. ¿CÓMO EXPLICAR LOS**

#### **DIFERENCIALES DE EFICIENCIA?**

El método DEA se centra, por ejemplo, en medir la eficiencia, pero no en explicar las diferencias de eficiencia entre las unidades analizadas.

Para explicar tales diferencias, se emplean habitualmente métodos en dos etapas. La primera fase consiste en la obtención de las eficiencias (por ejemplo, técnicas o de coste mediante DEA o FDH), mientras que la segunda trata de regresar las eficiencias estimadas frente a covariables exógenas que representan a características propias (sean

corporativas o propias de las unidades estudiadas) o características específicas de la región o país en que está la DMU, entre otras. La estimación de los coeficientes del modelo suele realizarse o bien por mínimos cuadrados ordinarios (OLS) o bien mediante estimaciones de máxima verosimilitud usando el modelo truncado como, por ejemplo, el modelo Tobit.

Sin embargo, Simar y Wilson (2007) mostraron que la aproximación en dos etapas es inapropiada, básicamente por dos cuestiones. Primero, porque al procedimiento en dos etapas le falta un proceso generador de datos (PGD) correctamente definido (la regresión censurada o truncada no es apropiada y existen problemas de muestras finitas). Y, segundo, DEA genera complejos patrones de correlación entre las puntuaciones de eficiencia estimadas. Es decir, las eficiencias técnicas no son independientes entre sí, dada la naturaleza del DEA, y la inferencia basada en *bootstrap naive* en un procedimiento en dos etapas es errónea.

Para superar estas limitaciones, Simar y Wilson (2007) proponen la siguiente aproximación:

- a) Construir y simular un proceso generador de datos (DGP).
- b) Generar muestras mediante *bootstrap* con observaciones independiente e idénticamente distribuidas (i.i.d.) artificiales de un DGP artificial.
- c) Construir los errores estándar e intervalos de confianza a partir de la simulación *bootstrap*.

Más concretamente, el procedimiento de Simar y Wilson (2007) es como sigue:

- 1) Estimar  $TE_i$  usando DEA, para las DMUs:  $i = 1, \dots, n$ .

2) Ajustar  $TE_i^{\hat{}} = \beta' z_i + \varepsilon_i$ , donde  $\beta$  es un vector de parámetros,  $z_i$  es un vector de variables explicativas, y  $\varepsilon_i$  es un término de error; usando una regresión truncada estimada por máxima verosimilitud para obtener  $\hat{\beta}$  y  $\hat{\sigma}_\varepsilon$ . Para ello, deben hacerse las siguientes fases:

a. Excluir las DMUs eficientes, es decir, aquellas donde  $TE_i^{\hat{}} = 1$ .

Consideremos que se excluyen  $m$  DMUs,  $j = 1, \dots, m$ .

b. Construir  $\xi_i = \varepsilon_i + \zeta_i$  con  $\zeta_i = TE_i^{\hat{}} - TE_i$ .

c.  $TE_i^{\hat{in}} \in (0, 1]$  - orientación-input: truncamiento a la derecha en 1.

d.  $TE_i^{\hat{out}} \in [0, \infty)$  - orientación-output: truncamiento a la izquierda en 1.

3) Realizar un bucle durante los siguientes tres pasos  $B$  veces ( $b = 1, \dots, B$ ). Esto permitirá:

a. Dibujar  $\varepsilon_i^b$  a partir de  $N(0, \hat{\sigma}_\varepsilon)$  con truncamiento a la izquierda (si es de orientación-output) o truncamiento a la derecha si es de orientación-input en el punto  $(1 - \hat{\beta}' z_i)$  para  $i = m + 1, \dots, n$ .

b. Computar  $TE_i^b = \hat{\beta}' z_i + \varepsilon_i^b$  para  $i = m + 1, \dots, n$ .

c. Estimar  $\hat{\beta}^b$  y  $\hat{\sigma}_\varepsilon^b$  mediante regresión truncada usando el score  $TE_i^b$  como variable del lado izquierdo (variable endógena).

4) Construir los errores estándar para  $\hat{\beta}'$  y  $\hat{\sigma}_\varepsilon$  (también los intervalos de confianza) a partir de la distribución simulada de  $\hat{\beta}'$  y  $\hat{\sigma}_\varepsilon$ .



Simar y Wilson (2007) proporcionaron un modelo estadístico que puede racionalizar la estimación en dos etapas de la eficiencia técnica en entornos no paramétricos. La estimación en dos etapas ha sido ampliamente utilizada, pero requiere un supuesto fuerte: las variables de la segunda etapa no pueden afectar las variables de inputs y output en la primera etapa. Recientemente, Daraio et al. (2011) proporcionan una prueba completamente no paramétrica de este supuesto. Además, proporcionan un enlace teórico con los resultados obtenidos por Politis et al. (2001), lo que permite estimar los valores críticos para sus estadísticos de prueba utilizando *bootstrap subsampling*, mientras que se optimiza la elección del tamaño de la sub-muestra, minimizando una medida de la volatilidad. Sus resultados de simulación indican que los contrastes funcionan bien en términos de tamaño y potencia.

## MEDICIÓN DE LA EFICIENCIA – FRONTERAS ESTOCÁSTICAS

---

### 4.1. INTRODUCCIÓN

Los métodos de estimación de la frontera determinista vistos en el Capítulo 3 fueron también sustituidos, además de por la metodología DEA, por los modelos de frontera estocástica.

Aigner et al. (1977), Battese and Corra (1977) y Meeusen y van den Broeck (1977) introdujeron simultáneamente los modelos de frontera estocástica comenzando así con el Análisis de Fronteras Estocásticas o SFA. Estos modelos simplemente añadieron a la frontera estimada otro elemento estocástico distinto de la ineficiencia ya considerada. Con ello, se consideraba que la producción pudiera fluctuar debido a otras fuentes que no fuesen estrictamente la ineficiencia (por ejemplo, aquellos elementos que están al margen del control de la empresa como factores omitidos relacionados con la heterogeneidad no observada de las empresas y sus entornos operativos, restricciones en el suministro de factores, elementos meteorológicos, pero también los que corresponden al ruido estadístico como los errores de medida y procesamiento de datos o los errores de especificación del modelo).

Así pues, la verdadera innovación en SFA provino de la especificación del término de error, que ahora es una convolución de dos variables aleatorias representadas por la ineficiencia y el ruido. Dicho de otra manera, la estimación de mínimos cuadrados

ordinarios (MCO) de un modelo de producción/costes solo permite estimar un único término de error, que se considera ineficiencia, mientras que el término de error compuesto en SFA permite la separación del ruido de la ineficiencia.

En este capítulo hablaremos de los modelos de frontera estocástica distinguiendo varias aproximaciones. Por un lado, el enfoque frecuentista paramétrico, tanto en datos de corte transversal (sección cruzada) - considerando la independencia y la dependencia de los errores aleatorios (ineficiencia y ruido) -, como también en el entorno de los datos de panel. Por otro lado, también consideramos el enfoque frecuentista con perspectiva semiparamétrica, así como el enfoque no paramétrico del modelo StoNED. Finalmente, se comentan los modelos paramétricos de fronteras estocásticas usando la metodología Bayesiana, tanto en corte transversal como panel de datos.

## **4.2. ANÁLISIS PARAMÉTRICO DE FRONTERAS ESTOCÁSTICAS (SFA)**

En esta sección distinguimos los enfoques SFA frecuentistas y Bayesiano, tanto para corte transversal como panel de datos.

### **4.2.1. ENFOQUE FRECUENTISTA**

#### **4.2.1.1. MODELOS DE FRONTERA ESTOCÁSTICA EN CORTE TRANSVERSAL**

En general, partiendo de un modelo lineal, el modelo de frontera estocástica en el contexto de los datos de corte transversal puede escribirse tal como:

$$\begin{aligned}
 y_i &= f(x_i; \beta) + v_i \pm u_i, \quad i = 1, 2, \dots, n \\
 &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + v_i \pm u_i \\
 &= \beta_0 + x_i' \beta + v_i \pm u_i \tag{4.1} \\
 v_i &\sim N(0, \sigma_v^2) \\
 u_i &\sim \mathfrak{I}
 \end{aligned}$$

donde  $x_i'$  es un vector fila de orden  $1 \times k$  de inputs y  $\beta$  es un vector  $k \times 1$  de parámetros de la tecnología.  $v_i$  y  $u_i$  son el ruido idiosincrático y la ineficiencia. El signo,  $\pm$ , indica si la función es de costes (positiva) o producción (negativa).

Atendiendo a los elementos contenidos en la expresión [4.1] podemos hacer las siguientes consideraciones:

- **Respecto de la tecnología,  $f(x_i; \beta)$ .**

A menudo, la elección de la tecnología de producción es una cuestión de preferencia del investigador. En el trabajo aplicado una amplia mayoría de investigadores utilizan la función de Cobb-Douglas<sup>13</sup> o la función flexible de producción/costes logarítmica trascendental de segundo grado (translog) (véase Tabla 2.1, Capítulo 2), como ha señalado Greene (2008, página 98).

Por ejemplo, si asumimos que  $f(x_i; \beta)$  es una función de producción log-lineal de Cobb-Douglas (véase la Tabla 2.1, Capítulo 2), entonces el modelo de frontera estocástica de producción se escribe como:

---

<sup>13</sup> La función de producción Cobb-Douglas reúne algunas propiedades útiles, tales como la interpretación práctica de los parámetros estimados y la sencilla inclusión del progreso técnico neutral de Hicks en las fronteras de producción estimadas. En el progreso técnico neutral de Hicks el capital y el trabajo son afectados por el progreso tecnológico. La cantidad de factores utilizados disminuye. Aumenta la eficiencia y la productividad de todos los factores productivos utilizados.

$$\log y_i = \beta_0 + \sum_{j=1}^k \beta_j \log x_{ij} + v_i - u_i, \quad i = 1, 2, \dots, n,$$

donde  $\log y_i$  es el logaritmo natural de la producción de la  $i$ -ésima empresa; y  $\log x_i$  es un vector  $k \times 1$  of transformaciones logarítmicas de las cantidades de input de la  $i$ -ésima empresa. Usando la frontera estocástica de costes de tipo Cobb-Douglas para  $\log c_i$  ( $c_i$  es el coste de la empresa  $i$ -ésima), el modelo sería en el caso de un solo output y los precios de los factores,  $w_{ij}, i = 1, \dots, n; j = 1, \dots, k$ , igual a:

$$\log c_i = \beta_0 + \sum_{j=1}^k \beta_j \log w_{ij} + \gamma_1 \log y_i + v_i + u_i, \quad i = 1, 2, \dots, n$$

Introduciendo el supuesto de homogeneidad de grado +1 en inputs, y dividiendo por un numerario, por ejemplo,  $w_{ik}, i = 1, \dots, n$ , tanto los costes como los precios de los factores restantes, el modelo resultante se convierte en:

$$\log \left( \frac{c_i}{w_{ik}} \right) = \beta_0 + \sum_{j=1}^{k-1} \beta_j \log \left( \frac{w_{ij}}{w_{ik}} \right) + \gamma_1 \log y_i + v_i + u_i, \quad i = 1, 2, \dots, n$$

Sin embargo, un problema observado con las funciones de producción/costes más flexibles (por ejemplo, aproximaciones de  $f$  que sean semiparamétricas), es el hecho de que tales especificaciones pueden producir estimaciones que no satisfacen los axiomas básicos de la producción (Kumbhakar y Lovell, 2000). Más específicamente, una forma funcional más flexible puede crear problemas en la estimación econométrica, puesto que la función necesita restricciones adicionales sobre los parámetros estimados para satisfacer, por ejemplo, la monotonicidad y concavidad. Un análisis exhaustivo de este tema fue presentado por Gong y Sickles (1992), que investigaron la sensibilidad de los resultados en los modelos de frontera estocástica a la elección de la forma funcional, llegando a la conclusión de que la elección de la tecnología de producción correcta (en

el sentido de que la tecnología de producción debe describir con mayor precisión la verdadera función de producción), es imprescindible para obtener resultados insesgados.

- **Respecto de la distribución de los errores:**  $v_i$  y  $u_i$ .

La literatura clásica asume que ambos errores son independientes, pero además que  $v_i$  (que es denominado error idiosincrático o ruido) tiene una distribución simétrica i.i.d.  $N(0, \sigma_v^2)$ , siendo  $\sigma_v^2$  su varianza. Por otro lado,  $u_i$  (conocido como término de ineficiencia) en general se distribuye  $u_i \sim \mathfrak{J}$ , siendo  $\mathfrak{J}$  cualquier distribución de probabilidad con soporte no negativo.

Cabe señalar que las ineficiencias estimadas pueden cambiar cuando se modifican los supuestos de distribución utilizados. Independientemente de cómo se especifique la distribución de los términos de ineficiencia, la lógica común detrás de todas las formulaciones es que las ineficiencias deben distribuirse de tal forma que la media sea cero o suficientemente cercana a cero. El razonamiento es que con las empresas que operan en mercados competitivos, debemos esperar que la ineficiencia deba ser bastante baja, y así la masa de la mayor parte de la probabilidad debería estar en la parte "eficiente" de la distribución (cola derecha).

Por ejemplo, Meeusen y van den Broeck (1977) asignaron la distribución exponencial:  $u_i \sim Exp$  con parámetro  $\sigma_u^2 > 0$ , Battese and Corra (1977) asumieron la distribución half-normal:  $u_i \sim N^+(0, \sigma_u^2)$  y Aigner et al. (1977) consideraron ambas distribuciones. Sin embargo, ya que la half-normal y exponencial son especificaciones de un solo parámetro con modas iguales a cero, esto produjo cierto escepticismo acerca de su generalidad. En este sentido, Stevenson (1980) sugirió la distribución truncada-

normal y gamma. Greene (1980a,b) propuso la distribución gamma y Lee (1983) propuso una familia de distribuciones de Pearson de cuatro parámetros. Más recientemente, Greene (1990, 2003) ha propuesto la distribución de densidad gamma de dos parámetros como una alternativa más general.

Llegados a este punto, cabría destacar que la estimación es paramétrica y se realiza por máxima verosimilitud (ML), donde los supuestos distribucionales para  $v_i$  y  $u_i$  son importantes porque de ellos depende la estimación de los parámetros  $\beta$ , el término de ineficiencia  $u_i$ , la varianza de los errores ( $\sigma_v^2, \sigma_u^2$ ), y la eficiencia técnica (o de costes) de cada empresa, obtenida a partir de la distribución de  $u_i/\varepsilon_i$ .

Sin embargo, la literatura sobre SFA también ha prestado atención a la hipótesis de dependencia de los componentes del error y al hecho de que el ruido no posea una distribución normal. En cuanto al primer caso, Smith (2008) ha propuesto representar la estructura de dependencia entre ambas variables aleatorias por medio de funciones cópula.<sup>14</sup> Este enfoque permite determinar si existe o no alguna dependencia estadística entre los componentes del error del modelo a través de una prueba de hipótesis estándar. Smith (2008, página 73) justifica esta aproximación argumentando que en un sistema productivo no es fácil entender que el error debe estar compuesto por partes

---

<sup>14</sup> La técnica de modelado utilizada por Smith (2008) se basa en el teorema de representación debido a Sklar (ver Sklar, 1959, 1973) en el que la distribución conjunta de variables aleatorias puede expresarse en función de sus marginales univariantes y la función que determina la dependencia. Los modelos basados en cópulas proporcionan una gran flexibilidad en el modelado de las distribuciones multivariadas, especialmente cuando las distribuciones marginales son diferentes de la normal, permitiendo al investigador especificar modelos para las distribuciones marginales separadamente de la estructura de dependencia (cópula) que se unen para formar una distribución conjunta. Smith (2008) utiliza varios enfoques flexibles basados en diferentes familias de cópulas bivariadas como Farlie (1960), Gumbel (1960) y Morgenstern (1956), más conocidos como cópula FGM, o Ali et al. (1978, AMH), Plackett (1965) o Frank (1979). Pero, existen muchas otras, incluidas aquellas que modelan la dinámica de la dependencia (véase Patton, 2006a,b; 2012).

independientes. Por ejemplo, dice que, en la agricultura, el ruido o las fluctuaciones estacionales a menudo afectan las decisiones productivas. Otro ejemplo, es el posible efecto adverso sobre la eficiencia de la propia fuerza de trabajo resultante del derrame incontrolado de productos peligrosos por parte de contaminantes en la industria pesada. Algunos trabajos recientes que analizan la dependencia a través de cópulas son El Mehdi y Hafner (2014), Wiboonpongse et al. (2015), y también Gómez-Déniz y Pérez-Rodríguez (2015, 2017). Estos últimos autores usan distribuciones bivariadas a partir de las cuales obtienen soluciones cerradas de las funciones de verosimilitud necesarias para estimar los parámetros del modelo, anidando las distribuciones clásicas normal-half normal y normal-exponencial, respectivamente.

En cuanto al segundo caso, la teoría clásica sobre la frontera estocástica supone que los *shocks* aleatorios que afectan la producción o el coste se consideran casi siempre como una variable distribuida normalmente. La pregunta es: ¿Es crucial el supuesto de normalidad para el ruido aleatorio en la frontera estocástica? La respuesta es que podría relajarse. De hecho, el supuesto de normalidad, aunque probablemente no afecte, es una suposición más que uno podría preferir no hacer. En este sentido, existe una literatura, aún escasa, sobre supuestos diferentes de la distribución normal. Es el caso de Smith (2008), quien propone una distribución logística (modelo logístico-exponencial) o Miller (2005) quien propone un enfoque teórico para construir clases más flexibles de modelos paramétricos de frontera estocástica para sistemas de producción de un solo output que anidan todos los métodos paramétricos de uso común. Por otro lado, otros autores simplemente utilizan alternativas no paramétricas basadas en estimadores *kernel*, como son los casos de Baccouche y Kouki (2003), Fan et al. (1996), Kumbhakar et al. (2007) o Martins-Filho y Yao (2010), entre otros.



- **Respecto de la estimación de la eficiencia**

Para la obtención de los índices de eficiencia (técnica o de costes) específicos de las empresas pueden usarse varias medidas.

Por un lado, la expectativa o la moda condicionales de  $u$  dado  $\varepsilon$  usando las expresiones de Jondrow et al. (1982, JLMS). Por ejemplo, para el cálculo de la eficiencia técnica y de costes, respectivamente, tenemos las siguientes expresiones:

$$TE_i = \exp(-\hat{u}_i),$$

$$CE_i = \exp(\hat{u}_i),$$

donde  $\hat{u}_i = E[u_i/\varepsilon_i]$  (media condicional) o  $\hat{u}_i = M[u_i/\varepsilon_i]$  (moda condicional), que toman valores entre 0 (totalmente ineficiente) y 1 (totalmente eficiente). Cabe destacar que JLMS no genera estimaciones consistentes de la eficiencia de los productores en un corte transversal, a diferencia del panel de datos, donde sí las obtiene. El estimador de eficiencia JLMS es inconsistente cuando  $n \rightarrow \infty$ . Esto no es sorprendente por dos razones. En primer lugar, en una sección transversal, como  $n \rightarrow \infty$ , se añaden nuevas empresas a la muestra con su propio nivel de ineficiencia en lugar de nuevas observaciones para ayudar a obtener un determinado nivel específico de ineficiencia de las empresas. En segundo lugar, el estimador de eficiencia JLMS no está diseñado para estimar la ineficiencia no condicional, sino que está diseñado para estimar la ineficiencia condicional a  $\varepsilon_i$ , para lo cual es un estimador coherente. Por otra parte, el estimador de ineficiencia JLMS se conoce como un estimador *shrinkage*. En promedio, sobreestima el nivel de ineficiencia de una empresa con  $u_i$  pequeño, mientras que subestima la ineficiencia para una empresa con  $u_i$  grande.

Por otro lado, otro método para obtener el estimador puntual de la eficiencia técnica (o de costes) es el propuesto por Battese y Coelli (1988). Para estos autores, la eficiencia técnica puede obtenerse a partir de la expresión:

$$TE_i = E[\exp(-\hat{u}_i)/\varepsilon_i].$$

mientras que la eficiencia de costes puede obtenerse a partir de la expresión:

$$CE_i = E[\exp(\hat{u}_i)/\varepsilon_i].$$

Cabe destacar que los estimadores puntuales JLMS y Battese y Coelli (1988) pueden proporcionar distintos resultados. Por ejemplo, para el caso de la eficiencia técnica, se obtiene que  $\exp(-\hat{u}_i) \neq E[\exp(-\hat{u}_i)/\varepsilon_i]$ .

Finalmente, en aquellos casos en los que se conoce la distribución condicional de  $u$  dado  $\varepsilon$ , pueden obtenerse los intervalos de confianza de la estimación puntual de las eficiencias. Taube (1988), Hjalmarsson et al. (1996), Horrace y Schmidt (1996) y Bera y Sharma (1999) obtienen expresiones de los intervalos basándose en la función condicionada  $f(u_i/\varepsilon_i)$ .

Por ejemplo, Horrace y Schmidt (1996) derivan el sistema de ecuaciones que permiten obtener los límites del intervalo de confianza de las eficiencias para un nivel de confianza del  $(1-\alpha)\%$  de la distribución condicional de  $u_i/\varepsilon_i$ , tal que:

$$\begin{aligned}\sigma^2 &= \sigma_v^2 + \sigma_u^2 \\ \lambda &= \sigma_u / \sigma_v \\ \mu_i^* &= -\varepsilon_i \sigma_u^2 / \sigma^2 = -\varepsilon_i \lambda^2 / (1 + \lambda^2) \\ \sigma^* &= \sigma_u \sigma_v / \sigma = \sigma \lambda / (1 + \lambda^2) \\ LB_i &= \mu_i^* + \sigma^* \Phi^{-1} \left[ 1 - \left( 1 - \frac{\alpha}{2} \right) \Phi(\mu_i^* / \sigma^*) \right] \\ UB_i &= \mu_i^* + \sigma^* \Phi^{-1} \left[ 1 - \frac{\alpha}{2} \Phi(\mu_i^* / \sigma^*) \right]\end{aligned}$$

donde  $LB_i$  y  $UB_i$  son los límites inferior y superior, respectivamente. Si los elementos contenidos en dichas fórmulas fuesen los verdaderos parámetros, entonces el intervalo  $[LB_i, UB_i]$  abarcaría o englobaría al  $(1 - \alpha)\%$  de la distribución de  $u_i / \varepsilon_i$ . Cabe destacar que para construir los intervalos de confianza de la eficiencia técnica, es necesario calcular los límites como  $[\exp(-LB)_i, \exp(-UB)_i]$ .

Recientemente, Wheat et al. (2014) han obtenido los intervalos de predicción de anchura mínima. Estos autores demuestran cuál es el intervalo mínimo de predicción para  $u_i / \varepsilon_i$ .

A continuación, describiremos someramente algunos de los modelos y métodos más generales comentados anteriormente distinguiendo entre la independencia y dependencia del ruido y la ineficiencia.

#### 4.2.1.2. INDEPENDENCIA DE LOS ERRORES

En esta sección, describiremos los modelos clásicos utilizados bajo esta consideración. Básicamente, son los modelos denominados: normal-exponencial,

normal-half-normal y normal-truncada normal, respectivamente. Cabe destacar que, en general, hemos utilizado la notación de Kumbhakar y Lovell (2000).

### 1) Normal-exponencial

El modelo clásico de frontera estocástica con el supuesto de que la convolución del error es normal-exponencial está descrito por la siguiente representación estocástica:

$$(i) v_i \sim iid N(0, \sigma_v^2);$$

$$(ii) u_i \sim iid \text{ exponencial con parámetro } \sigma_u > 0; \text{ y}$$

(iii)  $u_i$  y  $v_i$  están distribuidos independientemente el uno del otro y de los regresores.

A partir de los supuestos anteriores, las funciones de densidad de probabilidad (pdf) de  $v_i$  y  $u_i$  son, respectivamente:

$$f_{\sigma_v}(v_i) = \frac{1}{\sigma_v \sqrt{2\pi}} e^{-\frac{v_i^2}{2\sigma_v^2}} \quad [4.2]$$

$$f_{\sigma_u}(u_i) = \frac{1}{\sigma_u} e^{-\frac{u_i}{\sigma_u}}$$

donde  $-\infty < v_i < \infty$ ,  $\sigma_v > 0$ ,  $u_i > 0$  y  $\sigma_u > 0$ .

Sabiendo que  $\varepsilon_i = y_i - x_i' \beta$ , la distribución marginal conjunta para  $\varepsilon_i$  está descrita por:

$$f_{\sigma_u, \sigma_v}(\varepsilon_i) = \frac{1}{\sigma_u} \Phi\left(-\frac{\varepsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right) \exp\left\{\frac{\varepsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right\} \quad [4.3]$$

que está asimétricamente distribuida con media:  $E[\varepsilon_i] = -\sigma_u$  y varianza dada por  $var(\varepsilon_i) \equiv \sigma = \sigma_u^2 + \sigma_v^2$ .  $\Phi(\cdot)$  es la función de distribución acumulada (cdf) de la normal estándar.

Al igual que en Aigner et al. (1977), puede emplearse la re-parametrización  $\lambda = \sigma_u / \sigma_v$ , pues ésta permite una interpretación útil de la existencia o no de ineficiencia en el modelo de frontera estocástica. Esta ratio indica la relación entre la desviación estándar atribuible a la ineficiencia respecto de la desviación estándar debida al ruido aleatorio. Así,

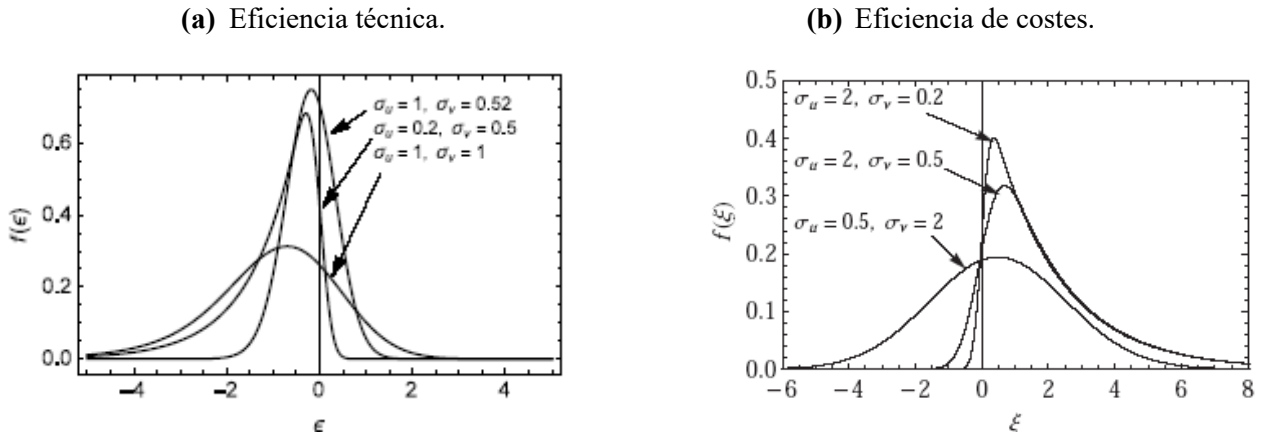
- Una estimación insignificante de  $\lambda$  significa que la ineficiencia no existe.

Claramente, podemos comprobar como cuando  $\lambda \rightarrow 0$ ,  $\sigma_u^2$  va a cero o  $\sigma_v^2$  va al infinito. Por lo tanto, no existe la ineficiencia o, lo que es lo mismo, todas las desviaciones se deben a ruido aleatorio.

- Del mismo modo, cuando  $\lambda \rightarrow \infty$  observamos que  $\sigma_u^2 \rightarrow \infty$  o  $\sigma_v^2 \rightarrow 0$ , lo que implica que toda desviación se explica por la ineficiencia. Entonces, la ineficiencia es “determinista” y se asemeja a los enfoques con exclusión de ruido aleatorio tales como el DEA.

La Figura 4.1 muestra diferentes gráficos tanto para la eficiencia técnica como de costes en el caso de la densidad del error compuesto.

**Figura 4.1.** Distribuciones marginales  $f(\varepsilon)$  y  $f(\xi)$ , según sean de producción o costes, para distintos valores de los parámetros.



La derivación de la función de verosimilitud está basada en el supuesto de independencia entre el ruido e ineficiencia y el uso de la función  $f_{\sigma_u, \sigma_v}(\varepsilon_i)$ . La estimación máximo verosímil usando todas las observaciones de la muestra de  $n$  empresas, puede escribirse como:

$$\begin{aligned} \log L(\beta, \sigma_u, \sigma_v) &= \sum_{i=1}^n \log f_{\sigma_u, \sigma_v}(\varepsilon_i) = \\ &= \sum_{i=1}^n \log \left[ \frac{1}{\sigma_u} \Phi \left( -\frac{\varepsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u} \right) \exp \left\{ \frac{\varepsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2} \right\} \right] \\ &= -n \log \sigma_u + n \frac{\sigma_v^2}{2\sigma_u^2} + \sum_{i=1}^n \log \Phi(\tilde{\mu}_i / \sigma_v) + \sum_{i=1}^n \frac{\varepsilon_i}{\sigma_u} \end{aligned}$$

donde  $\tilde{\mu}_i = -\varepsilon_i - \sigma_v^2 / \sigma_u$ .

Dada la no linealidad de dicha función, pueden emplearse algoritmos de optimización numérica para estimar los parámetros del modelo, como Berndt-Hall-Hausman (BHHH) o Broyden-Fletcher-Goldfarb-Shanno (BFGS), entre otros. Otra vía es usar directamente las condiciones de primer y segundo orden del problema de maximización y emplear métodos de resolución de sistemas de ecuaciones como Gauss-Seidel. Por simplicidad, omitiremos ambos desarrollos.

Para la estimación puntual de la eficiencia técnica es necesario construir previamente la función de densidad de  $u_i$  condicionada a  $\varepsilon_i$ ,  $f(u_i/\varepsilon_i)$ . Esta función sigue una distribución half-normal,  $N^+(\tilde{\mu}_i, \sigma_v^2)$ , cuya expresión está dada por:

$$f(u_i/\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma_v\Phi(\tilde{\mu}_i/\sigma_v)} \exp\left\{-\frac{1}{2\sigma_v^2}(u_i - \tilde{\mu}_i)^2\right\} \quad [4.4]$$

A partir de la expresión [4.4] puede obtenerse la media condicionada de la ineficiencia frente al ruido, que es igual a:

$$\begin{aligned} E(u_i/\varepsilon_i) &= \tilde{\mu}_i + \sigma_v \frac{\phi(-\tilde{\mu}_i/\sigma_v)}{\Phi(-\tilde{\mu}_i/\sigma_v)} \\ &= \sigma_v \left( \frac{\phi(A_i)}{\Phi(-A_i)} - A_i \right) \end{aligned}$$

donde  $\phi(\cdot)$  es la pdf de la normal estándar y  $A_i = -\tilde{\mu}_i/\sigma_v$ .

A partir de la anterior expresión y usando la fórmula de JLMS, la eficiencia técnica se calcularía como:

$$TE_i = \exp(-E(u_i/\varepsilon_i)) = \exp\left(-\sigma_v \left( \frac{\phi(A_i)}{\Phi(-A_i)} - A_i \right)\right)$$

Por otro lado, en el caso de emplear expresión de Battese y Coelli (1988), la eficiencia técnica se calcularía a través de la expresión:

$$\begin{aligned} TE_i &= E[\exp(-u_i)/\varepsilon_i] \\ &= \left[ \frac{1 - \Phi(\sigma_v - \tilde{\mu}_i/\sigma_v)}{1 - \Phi(-\tilde{\mu}_i/\sigma_v)} \right] \exp\left(-\tilde{\mu}_i + \frac{1}{2}\sigma_v^2\right) \end{aligned}$$

que podemos comprobar no coinciden.

Las expresiones de las distribuciones anteriores obtenidas para las funciones de producción son similares a las obtenidas en el caso de las funciones de coste, simplemente cambiando de signo de los errores. Omitiendo escribir todas las

expresiones para las funciones de coste por este motivo, las cuales pueden consultarse en Kumbhakar y Lovell (2000), escribiremos una expresión genérica del logaritmo de verosimilitud que anida tanto a los resultados de la función de costes como de producción en el caso normal-exponencial. Esta expresión general puede escribirse como:

$$\log L(\beta, \sigma_u, \sigma_v) = -n \log \sigma_u + n \frac{\sigma_v^2}{2\sigma_u^2} + \sum_{i=1}^n \log \Phi(\tilde{\mu}_{*i}/\sigma_v) + \sum_{i=1}^n \frac{s\varepsilon_i}{\sigma_u}$$

donde  $\tilde{\mu}_{*i} = -s\varepsilon_i - \sigma_v^2/\sigma_u$ , siendo  $s = 1$  si la función es de producción y  $s = -1$  si es de costes.

Para obtener la estimación de  $u_i$ , puesto que se necesita obtener la media condicional a través de  $f(u_i/\varepsilon_i)$ , tenemos que una expresión, llamémosle “anidada”, para las funciones de producción y coste es:

$$E(u_i/\varepsilon_i) = \tilde{\mu}_{*i} + \sigma_v \frac{\phi(-\tilde{\mu}_{*i}/\sigma_v)}{\Phi(\tilde{\mu}_{*i}/\sigma_v)}$$

de tal manera que, usando la anterior expresión, una fórmula general que permite obtener las eficiencias de producción y coste usando la expresión de Battese y Coelli (1988), que denominamos genéricamente como  $E_i$ , puede escribirse como:

$$E_i = E \left[ \exp(-su_i)/\varepsilon_i \right] = \left[ \frac{1 - \Phi(s\sigma_v - \mu_{*i}/\sigma_v)}{1 - \Phi(-\mu_{*i}/\sigma_v)} \right] \exp \left( -s\mu_{*i} + \frac{1}{2}\sigma_v^2 \right)$$

Así, por ejemplo, en el caso de la estimación de la eficiencia de costes,  $s = -1$ , podemos escribir la siguiente expresión:



$$E_i = E \left[ \exp(u_i) / \varepsilon_i \right] \\ = \left[ \frac{1 - \Phi(-\sigma_v - \tilde{\mu}_{*i} / \sigma_v)}{1 - \Phi(-\tilde{\mu}_{*i} / \sigma_v)} \right] \exp \left( \tilde{\mu}_{*i} + \frac{1}{2} \sigma_v^2 \right)$$

donde  $\tilde{\mu}_{*i} = \varepsilon_i - \sigma_v^2 / \sigma_u$ . En esta ocasión, podríamos renombrar  $E_i$  por  $CE_i$ . En el caso de que fuera la eficiencia técnica podríamos sustituir  $E_i$  por  $TE_i$ .

## 2) Normal-half normal

El modelo de frontera estocástica con el supuesto de que la convolución del error es normal-half normal está descrito por la siguiente representación estocástica:

(i)  $v_i \sim iid N(0, \sigma_v^2)$ ;

(ii)  $u_i \sim iid N^+(0, \sigma_u^2)$  con parámetro  $\sigma_u > 0$ .

(iii)  $u_i$  y  $v_i$  están distribuidos independientemente el uno del otro y de los regresores.

A partir de los supuestos anteriores, las funciones de densidad de probabilidad (pdf) de  $v_i$  está dada por [4.2] y la de  $u_i$  es:

$$f_{\sigma_u}(u_i) = \frac{2}{2\pi\sigma_u\sigma_v} e^{\left( -\frac{u_i^2}{2\sigma_u^2} - \frac{v_i^2}{2\sigma_v^2} \right)}$$

mientras que la densidad marginal de error conjunto para el caso de la producción es:

$$f_{\sigma_u\sigma_v}(\varepsilon_i) = \frac{2}{2\sigma} \phi\left(\frac{\varepsilon_i}{\sigma}\right) \Phi\left(\frac{-\varepsilon_i\lambda}{\sigma}\right)$$

donde  $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$ . En este caso,  $E[\varepsilon_i] = -\sigma_u \sqrt{\frac{2}{\pi}}$  y la varianza es

$$\text{var}(\varepsilon_i) = \left(\frac{\pi-2}{\pi}\right) \sigma_u^2 + \sigma_v^2.$$

La función de densidad condicionada es:

$$f(u_i/\varepsilon_i) = \frac{1}{\sqrt{2\pi}\sigma_*} \exp\left\{-\frac{(u_i - \check{\mu}_i)^2}{2\sigma_*^2}\right\} / [1 - \Phi(-\check{\mu}_i/\sigma_*)] \quad [4.5]$$

donde  $\check{\mu}_i = -\varepsilon_i \sigma_u^2 / \sigma^2$  y  $\sigma_* = \sigma_u \sigma_v / \sigma$ .

Siguiendo la misma idea que para el caso normal-exponencial, escribiremos a continuación la función del logaritmo de verosimilitud que “anida” tanto a los resultados obtenidos para el caso de la función de producción como costes. Esta función general puede escribirse como:

$$\log L(\beta, \sigma_u, \sigma_v) = \frac{n}{2} \log\left(\frac{2}{\pi}\right) - n \log \sigma + \sum_{i=1}^n \log \Phi(-s\varepsilon_i \lambda / \sigma) - \sum_{i=1}^n \frac{\varepsilon_i^2}{2\sigma^2}$$

donde  $\lambda = \sigma_u / \sigma_v$ .

La media condicional es igual a:

$$E(u_i/\varepsilon_i) = \check{\mu}_{*i} + \sigma_* \frac{\phi(-\check{\mu}_{*i}/\sigma_*)}{\Phi(\check{\mu}_{*i}/\sigma_*)}$$

donde  $\check{\mu}_{*i} = -s\varepsilon_i \sigma_u^2 / \sigma^2$ , siendo  $s=1$  si la función es de producción y  $s=-1$  si es de costes.

La eficiencia técnica y de costes, en este modelo, se podría obtener usando la medida de Battese y Coelli (1988), a partir de la siguiente expresión:

$$\begin{aligned} E_i &= E[\exp(-su_i)/\varepsilon_i] \\ &= \left[ \frac{1 - \Phi(s\sigma_* - \check{\mu}_{*i}/\sigma_*)}{1 - \Phi(-\check{\mu}_{*i}/\sigma_*)} \right] \exp\left(-s\check{\mu}_{*i} + \frac{1}{2}\sigma_*^2\right) \end{aligned}$$

### 3) Normal-truncada normal

Aunque las variantes normal-half normal y normal-exponencial son los más extendidas en el análisis empírico, también pueden considerarse otras distribuciones para el término de ineficiencia, como es el caso del modelo normal-normal truncada, propuesto por Stevenson (1980).

La distribución normal truncada tiene la ventaja de permitir que la media de la ineficiencia sea distinta de cero, lo que puede ser preferible, por ejemplo, en situaciones en las que es sostenible que las empresas operen en un entorno donde pueda estar presente una ineficiencia significativa.

El modelo de frontera estocástica con el supuesto de que la convolución del error es normal-normal truncada está descrito por la siguiente representación estocástica:

$$(i) \quad v_i \sim iid N(0, \sigma_v^2);$$

(ii)  $u_i \sim iid TN(\mu_i, \sigma_u^2)$  con parámetros  $\mu_i$  y  $\sigma_u > 0$ , siendo  $\mu_i$  una función de las variables explicativas de la ineficiencia, como se discutirá más adelante.

(iii)  $u_i$  y  $v_i$  están distribuidos independientemente el uno del otro y de los regresores.

Las distribución de probabilidad de  $v_i$  está dada por [4.2], mientras que la de  $u_i$  es:

$$f_{\sigma_u}(u_i) = \frac{1}{\sigma_u \sqrt{2\pi} \Phi\left(-\frac{\mu}{\sigma_u}\right)} \exp\left[-\frac{1}{2}\left(\frac{u_i - \mu}{\sigma_u}\right)^2\right], \quad u_i > 0, \quad [4.6]$$

donde  $\mu \geq 0$ ,  $\sigma_u > 0$ ,  $\sigma_v > 0$  y  $\Phi(\cdot)$  representa la función de distribución normal estándar. Asumiendo que el error en la función de producción está dado por  $\varepsilon_i = v_i - u_i$ , la distribución marginal, asumiendo independencia entre  $u$  y  $v$ , es:

$$f_{\mu, \sigma_u, \sigma_v}(\varepsilon_i) = \frac{1}{\sigma} \phi\left(\frac{\varepsilon_i + \mu}{\sigma}\right) \Phi\left(\frac{\mu}{\lambda\sigma} - \frac{\varepsilon_i \lambda}{\sigma}\right) \left[\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1}, \quad -\infty < \varepsilon_i < \infty,$$

donde  $\phi(\cdot)$  representa la función de densidad de probabilidad de la normal estándar,  $\lambda = \sigma_u / \sigma_v$  y  $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$ , como los epígrafes anteriores.

En este caso,  $E[\varepsilon_i] = \left[\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1} \left[-\frac{\mu}{2} - \frac{\sigma_u}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(-\frac{\mu}{\sigma_u}\right)^2\right\}\right]$  y la varianza es

$$\text{var}(\varepsilon_i) = \mu^2 \frac{a}{2} \left(1 - \frac{a}{2}\right) \frac{a}{2} \left(\frac{\pi - a}{\pi}\right) \sigma_u^2 + \sigma_v^2, \text{ donde } a = \left[\Phi\left(-\frac{\mu}{\sigma_u}\right)\right]^{-1}.$$

La función de densidad condicionada es:

$$f(u_i / \varepsilon_i) = \frac{1}{\sqrt{2\pi} \sigma_* (1 - \Phi(-\hat{\mu}_i / \sigma_*))} \exp\left\{-\frac{(u_i - \hat{\mu}_i)^2}{2\sigma_*^2}\right\} \quad [4.7]$$

donde  $\hat{\mu}_i = (-\varepsilon_i \sigma_u^2 + \mu \sigma_v^2) / \sigma^2$  y  $\sigma_* = \sigma_u \sigma_v / \sigma$ .

Al igual que en los modelos anteriores, escribimos la función del logaritmo de verosimilitud que “anida” tanto a los resultados obtenidos para el caso de la función de producción como costes, la cual es igual a:

$$\begin{aligned} \log L(\beta, \sigma_u, \sigma_v, \mu) = & -\frac{n}{2} \log 2\pi - n \log \sigma - n \log \Phi\left(-\frac{\mu}{\sigma\sqrt{\gamma}}\right) + \sum_{i=1}^n \left(\frac{\varepsilon_i + s\mu}{\sigma}\right)^2 \\ & + \sum_{i=1}^n \log \Phi\left(\frac{(1-\gamma)\mu - s\gamma\varepsilon_i}{(\sigma^2\gamma(1-\gamma))^{1/2}}\right) \end{aligned}$$

donde  $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$  y  $\gamma = \sigma_u^2 / \sigma$ , siendo  $s = 1$  si la función es de producción y  $s = -1$  si es de costes.

En este caso, para construir las medidas de eficiencia de costes y producción, la media condicional es igual a:

$$E(u_i / \varepsilon_i) = \hat{\mu}_{*i} + \sigma_* \frac{\phi(-\hat{\mu}_{*i} / \sigma_*)}{\Phi(\hat{\mu}_{*i} / \sigma_*)}$$

donde  $\hat{\mu}_{*i} = (-s\varepsilon_i\sigma_u^2 + \mu\sigma_v^2) / \sigma^2$ .

Al igual que en las distribuciones anteriores, la eficiencia técnica y de costes en este modelo también se podrá obtener usando la medida de Battese y Coelli (1988) a partir de la siguiente expresión:

$$\begin{aligned} E_i &= E[\exp(-su_i) / \varepsilon_i] \\ &= \left[ \frac{1 - \Phi(s\sigma_* - \hat{\mu}_{*i} / \sigma_*)}{1 - \Phi(-\hat{\mu}_{*i} / \sigma_*)} \right] \exp\left(-s\hat{\mu}_{*i} + \frac{1}{2}\sigma_*^2\right) \end{aligned}$$

Nótese que, aunque las expresiones son parecidas, la notación cambia.

#### 4) Determinantes exógenos de la ineficiencia y la heterocedasticidad

Una cuestión importante que puede tratarse en la metodología SFA es la consideración de factores o variables que afectan a la ineficiencia. Estas variables, que usualmente no son ni los inputs ni los outputs del proceso de producción, sí afectan al desempeño de la unidad productiva y deberían ser tenidas en cuenta.

La incorporación de dichas variables al modelo de frontera estocástica definida por la ecuación [4.1] se ha hecho de varias maneras. Por ejemplo, éstas pueden aparecer en la función frontera pero también afectar a la distribución de la eficiencia. Kumbhakar y Lovell (2000) destacan que la presencia de heterogeneidad observable no controlada en

$u_i$  y/o  $v_i$  puede afectar la inferencia en los modelos de frontera estocástica. De hecho, mientras que la heteroscedasticidad no tratada en  $v_i$  no produce ningún sesgo para las estimaciones de los parámetros de la frontera, sí conduce a estimaciones sesgadas de la ineficiencia (Belotti et al., 2012).

En esta sección presentamos algunos enfoques que permiten introducir la heterogeneidad en el comportamiento del parámetro de localización de la distribución de ineficiencia (media), y la existencia de heteroscedasticidad en la ineficiencia, así como del término de error idiosincrático.

El punto de partida más simple para introducir factores exógenos en el modelo de ineficiencia (por ejemplo, la productividad, la experiencia gerencial, entre otras) está en la media de la distribución. Los primeros enfoques que consideraron la inclusión de variables explicativas en la ineficiencia fueron sugeridos por Kumbhakar et al. (1991) y Reifschneider y Stevenson (1991). Por ejemplo, considerando que  $u_i = z_i' \psi + \mathcal{G}_i$ , es decir, la ineficiencia posee un término sistemático  $z_i' \psi$ , siendo  $z_i'$  un vector fila formado por observaciones de variables que pueden afectar la ineficiencia (incluida una constante) y  $\psi$  es un vector de parámetros desconocidos; y un elemento aleatorio,  $\mathcal{G}_i$ .

También autores como Kumbhakar et al. (1991) y Huang y Liu (1994) propusieron parametrizar la media de la distribución de ineficiencia truncada normal. Básicamente,

$$u_i \sim iid TN(\mu_i, \sigma_u^2)$$

$$\mu_i = z_i' \psi$$

Una característica interesante de estas aproximaciones es que  $z_i$  puede contener interacciones con variables input (factores) permitiendo contrastar hipótesis sobre si la ineficiencia es neutral con respecto a su impacto como factores.

Otra alternativa para analizar el efecto de determinantes exógenos sobre la ineficiencia es considerar que la varianza depende de variables explicativas. Caudill y Ford (1993), Caudill et al. (1995) y Hadri (1999) propusieron modelos para la varianza de la distribución (heterocedasticidad) de  $u_i$ , tal como:

$$u_i \sim iid TN(0, \sigma_{ui}^2)$$

$$\sigma_{ui}^2 = \exp(z_i' \psi)$$

Hadri (1999) también consideró que la varianza del error idiosincrático pudiera ser heterocedástica, tal que:

$$v_i \sim iid N(0, \sigma_{vi}^2)$$

$$\sigma_{vi}^2 = \exp(h_i' \xi)$$

donde las variables en  $h_i$  no necesariamente están contenidas en  $z_i$ .  $\xi$  es un vector de parámetros desconocidos en la ecuación de la varianza del ruido idiosincrático. Otros autores, sin embargo, han considerado que las expresiones de la media y varianza pudieran tener la misma función (Wang y Schmitdt, 2002). Por ejemplo,  $u_i \sim iid TN(\mu h(w_i' \delta), \sigma_{ui}^2 h(w_i' \delta)^2)$ . En este caso,  $w_i$  es un vector de variables explicativas que no contiene un término constante, y  $\delta$  es el vector de parámetros.

El logaritmo de verosimilitud para la distribución normal-truncada normal con media y varianza de las ineficiencias iguales a  $\mu_i = w_i' \delta$  y  $\sigma_{ui}^2 = \exp(w_i' \delta)$ , respectivamente, puede escribirse como:

$$\log L(\beta, \sigma_u, \sigma_v, \delta) = -n \log 2\pi - n \log \sigma - \sum_{i=1}^n \log \Phi \left( \frac{w_i' \delta}{\sqrt{\gamma \sigma^2}} \right)$$

$$+ \sum_{i=1}^n \log \Phi \left( \frac{(1-\gamma) w_i' \delta - s \gamma \varepsilon_i}{\sqrt{\gamma(1-\gamma) \sigma^2}} \right) - \frac{1}{2} \sum_{i=1}^n \left( \frac{\varepsilon_i + s w_i' \delta}{\sigma} \right)^2$$

donde  $\gamma = \sigma_u^2 / \sigma^2$  y  $s = 1$  (si la función es de producción y  $s = -1$  (si la función es de costes).

### 4.2.1.3. DEPENDENCIA ENTRE LOS ERRORES

Los modelos clásicos de frontera estocástica consideran el supuesto de independencia de los errores. Sin embargo, este supuesto habitual puede ser relajado. Por ejemplo, Smith (2008) fue el primer autor que propuso un modelo de dependencia de los errores basado en cópulas.

A continuación, expondremos brevemente su planteamiento.

A partir del Teorema de Sklar, la función de distribución acumulada (cdf) conjunta se define por:

$$H(u, v) = P(U \leq u, V \leq v) = C_\theta(F(u), F(v))$$

donde  $C_\theta(F(u), F(v))$  es la función cópula bivariada de la distribución conjunta de  $U$  y  $V$ ,  $F(u)$  y  $F(v)$  son las cdf univariantes del ruido y la ineficiencia, respectivamente, y  $\theta$  es el parámetro de dependencia. Si  $U$  y  $V$  son variables aleatorias independientes,  $C_\theta(F(u), F(v)) = uv$ .

La función de densidad de probabilidad (pdf) de  $U$  y  $V$  es:

$$h(u, v) = \frac{\partial^2}{\partial u \partial v} H(u, v) = f(u) f(v) c_\theta(F(u), F(v))$$

donde  $f(u)$  y  $f(v)$  son las funciones de densidad univariantes y  $c_\theta(F(u), F(v))$  es la densidad de la cópula. Si  $U$  y  $V$  son independientes,  $h(u, v) = f(u) f(v)$ .



Partiendo del modelo de frontera estocástica de producción, donde si  $U$  y  $V$  son independientes,  $\varepsilon = v - u$ , puede escribirse la pdf como:

$$h(u, \varepsilon) = f(u) f(u + \varepsilon) c_{\theta}(F(u), F(u + \varepsilon))$$

Y a partir de aquí, pueden formularse expresiones de la pdf de  $\varepsilon$ , la función de densidad condicional de  $\varepsilon$  con respecto a  $u$ , y medidas de la eficiencia técnica. Smith (2008) usa diferentes cópulas como la normal, FGM, Plackett, entre otras (véase el Apéndice I para idea general de los casos de funciones cópula Normal, Clayton, Gumbel, o SJC).

Pero, también se pueden emplear otras cópulas como la definida por Sarmanov (1966). Este es el caso de los modelos desarrollados por Gómez-Déniz y Pérez-Rodríguez (2015, 2017). Sus modelos tienen la propiedad que permiten anidar a los modelos de independencia de los errores vistos anteriormente. Dado lo extensa que resulta la formulación de estos modelos, circunscribiremos la descripción al modelo normal-exponencial, dejando al lector la lectura del artículo de Gómez-Déniz y Pérez-Rodríguez (2015) para las expresiones en el caso normal-half-normal.

### 1) Normal-exponencial

Siguiendo a Gómez-Déniz y Pérez-Rodríguez (2017) expondremos en esta sección algunas de las expresiones de interés para la densidad conjunta de los errores, la ineficiencia condicionada al ruido, y la eficiencia técnica basándonos en la mezcla (mixtura) de distribuciones normal (ruido) y exponencial (ineficiencia), y considerando la dependencia de dichos errores, como aspecto novedoso en el tratamiento del modelo de fronteras estocásticas.

Estos autores utilizan la familia flexible de distribuciones bivariadas de Sarmanov, proponiendo un nuevo modelo con rango de dependencia más general y flexible que permite contrastar la independencia entre el término de ruido y la ineficiencia.

Esta familia de distribuciones se construye asumiendo que  $f_1(x_1)$  y  $f_2(x_2)$  son funciones de densidad de probabilidad marginales. Sea  $\varphi_s(t)$ ,  $s=1,2$ , funciones acotadas no constantes tales que  $\int_{-\infty}^{\infty} \varphi_s(t) f_s(t) dt = 0$ , entonces la función definida por

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)[1 + \omega\varphi_1(x_1)\varphi_2(x_2)] \quad [4.8]$$

es una función de densidad conjunta con marginales  $f_1(x_1)$  y  $f_2(x_2)$ , siendo  $\omega$  un número real que satisface la condición  $1 + \omega\varphi_1(x_1)\varphi_2(x_2) \geq 0$ , para todo  $x_1$  y  $x_2$ . La familia de cópulas Farlie–Gumbel–Morgernstern (FGM) puede verse como un caso especial del modelo [4.8], haciendo  $\varphi(x_s) = 1 - 2F(x_s)$ ,  $s=1,2$  similar a uno de los modelos propuestos por Smith (2008).  $F(\cdot)$  representa la cdf de variables aleatorias con pdf  $f(\cdot)$ .

A continuación, escribiremos las expresiones analíticas para la pdf marginal de la variable aleatoria  $\varepsilon$ , la distribución condicional de  $u | \varepsilon$ , y la eficiencia técnica. El modelo clásico normal-exponencial se obtiene como caso particular cuando  $\omega = 0$  mientras que  $\omega \neq 0$  indica la existencia de una estructura de dependencia entre los errores.

Tal y como sucede en el modelo clásico, sea  $v = u + \varepsilon$ . Usando [4.8] obtenemos que la densidad bivariada conjunta es igual a:

$$f_{\sigma_u, \sigma_v, \omega}(u_i, \varepsilon_i) = f_{\sigma_u}(u_i)f_{\sigma_v}(u_i + \varepsilon_i) \left[ 1 + \omega\varphi_{\sigma_u}(u_i)\varphi_{\sigma_v}(u_i + \varepsilon_i) \right], \quad [4.9]$$

y tomando  $\varphi_{\sigma_u}(u) = e^{-u} - \delta_u(\sigma_u)$ ,  $\varphi_{\sigma_v}(v) = e^{-v^2-2v} - \delta_v(\sigma_v)$ , con

$$\delta_u(\sigma_u) = \frac{1}{1 + \sigma_u}, \quad [4.10]$$

$$\delta_v(\sigma_v) = \frac{1}{\sqrt{1 + 2\sigma_v^2}} \exp\left\{\frac{2\sigma_v^2}{1 + 2\sigma_v^2}\right\}, \quad [4.11]$$

define una distribución bivariada de  $(u, v)$  con distribuciones marginales  $f_{\sigma_v}(v_i)$  y  $f_{\sigma_u}(u_i)$  definidas en el modelo clásico y donde  $\omega_1 \leq \omega \leq \omega_2$ , siendo

$$\omega_1 = \max\left\{\frac{-1}{\delta_u(\sigma_u)\delta_v(\sigma_v)}, \frac{-1}{(1 - \delta_u(\sigma_u))(e - \delta_v(\sigma_v))}\right\}, \quad [4.12]$$

$$\omega_2 = \min\left\{\frac{1}{\delta_u(\sigma_u)(e - \delta_v(\sigma_v))}, \frac{1}{(1 - \delta_u(\sigma_u))\delta_v(\sigma_v)}\right\}, \quad [4.13]$$

En el caso del modelo de producción (también puede consultarse el artículo citado para el caso del modelo de costes), la distribución pdf marginal de  $\varepsilon$  está dada por:

$$\begin{aligned} f_{\sigma_u, \sigma_v, \omega}(\varepsilon_i) = & \frac{\Upsilon_{\sigma_u, \sigma_v}^1(\varepsilon_i)}{\sigma_u \sqrt{1 + 2\sigma_v^2}} \left\{ \Phi\left(\frac{\tilde{\mu}_i}{\sigma_v}\right) \left[ \sqrt{1 + 2\sigma_v^2} + \frac{\omega e^{\frac{2\sigma_v^2}{1 + 2\sigma_v^2}}}{1 + \sigma_u} \right] \right. \\ & - \omega \Upsilon_{\sigma_u, \sigma_v}^2(\varepsilon_i) \Phi\left(\frac{\mu_i}{\sigma_v} - \sigma_v\right) \\ & - \frac{\omega}{1 + \sigma_u} \Upsilon_{\sigma_u, \sigma_v}^3 \Phi\left(-\frac{\sigma_v}{\sigma_u} \frac{1 + 2\sigma_u}{\sqrt{1 + 2\sigma_v^2}} - \frac{\varepsilon_i}{\sigma_v} \sqrt{1 + 2\sigma_v^2}\right) \\ & \left. + \omega \Upsilon_{\sigma_u, \sigma_v}^4(\varepsilon_i) \Phi\left(-\frac{\sigma_v}{\sigma_u} \frac{1 + 3\sigma_u}{\sqrt{1 + 2\sigma_v^2}} - \frac{\varepsilon_i}{\sigma_v} \sqrt{1 + 2\sigma_v^2}\right) \right\}, \end{aligned} \quad [4.14]$$

donde  $\tilde{\mu}_i$  está definido en la sección 2.1.1 y,

$$\Upsilon_{\sigma_u, \sigma_v}^1(\varepsilon_i) = \exp\left\{\frac{\varepsilon_i}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right\},$$

$$Y_{\sigma_u, \sigma_v}^2(\varepsilon_i) = \exp\left\{\varepsilon_i + \frac{\sigma_v^2}{\sigma_u} + \frac{\sigma_v^2}{2} + \frac{2\sigma_v^2}{1+2\sigma_v^2}\right\},$$

$$Y_{\sigma_u, \sigma_v}^3 = \exp\left\{\frac{\sigma_v^2}{2\sigma_u^2} \frac{4\sigma_u - 2\sigma_v^2}{1+2\sigma_v^2} + \frac{2\sigma_v^2}{1+2\sigma_v^2}\right\},$$

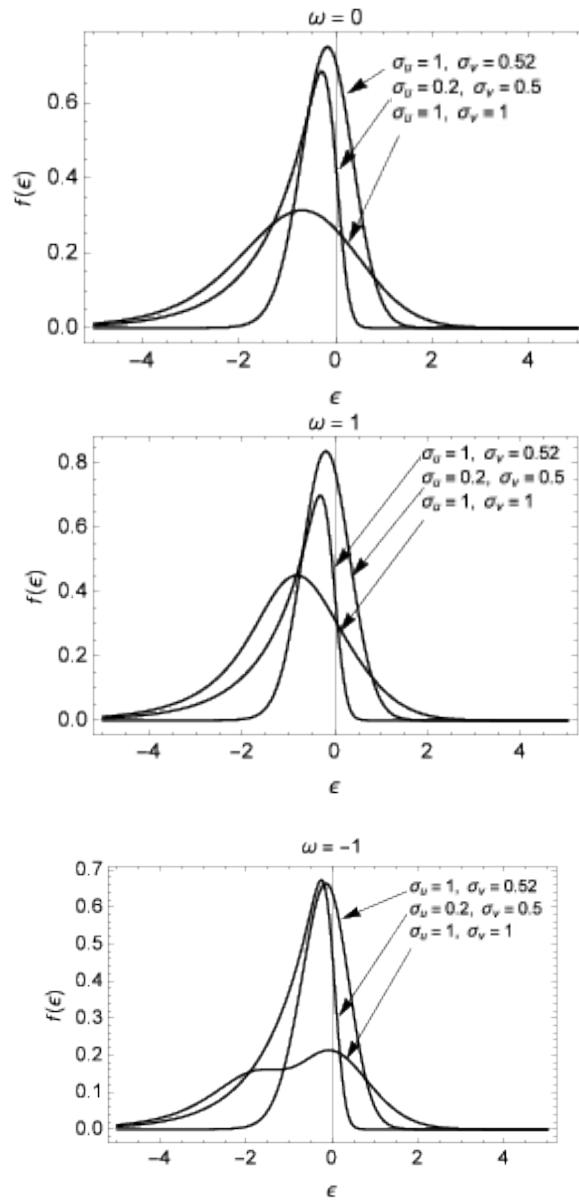
$$Y_{\sigma_u, \sigma_v}^4(\varepsilon_i) = \exp\left\{\varepsilon_i + \frac{\sigma_v^2}{2\sigma_u^2} \frac{5\sigma_u^2 + 6\sigma_u - 2\sigma_v^2}{1+2\sigma_v^2} + \frac{2\sigma_v^2}{1+2\sigma_v^2}\right\}.$$

La media de la pdf marginal [4.14] es igual a  $E(\varepsilon_i) = -\sigma_u$  mientras que su varianza es:

$$\text{var}(\varepsilon_i) = \sigma_u^2 + \sigma_v^2 \left[ 1 - \frac{4\omega\sigma_u^2}{(1+\sigma_u^2)(1+2\sigma_v^2)^{3/2}} e^{\frac{2\sigma_v^2}{1+2\sigma_v^2}} \right].$$

La Figura 4.2 muestra algunos ejemplos de la pdf marginal para diferentes valores de los parámetros. Puede apreciarse como las curvas coinciden con el caso de independencia de la Figura 4.1 para valores de  $\omega = 0$ .

**Figura 4.2.** Funciones de densidad conjuntas para diferentes valores de los parámetros.



La función del logaritmo de verosimilitud puede escribirse como:

$$\begin{aligned}
 \text{Log}L(\beta, \sigma_u, \sigma_v, \omega) &= \sum_{i=1}^n \log f_{\sigma_u, \sigma_v, \omega}(\varepsilon_i) \\
 &= -n \log \left( \sigma_u \sqrt{1+2\sigma_v^2} \right) + \sum_{i=1}^n \log \Upsilon_{\sigma_u, \sigma_v}^1(\varepsilon_i) \\
 &\quad + \sum_{i=1}^n \log \left\{ \Phi \left( \frac{\tilde{\mu}_i}{\sigma_v} \right) \left[ \sqrt{1+2\sigma_v^2} + \frac{\omega e^{\frac{2\sigma_v^2}{1+2\sigma_v^2}}}{1+\sigma_u} \right] \right. \\
 &\quad \left. - \omega \Upsilon_{\sigma_u, \sigma_v}^2(\varepsilon_i) \Phi \left( \frac{\mu_i - \sigma_v}{\sigma_v} \right) \right. \\
 &\quad \left. - \frac{\omega}{1+\sigma_u} \Upsilon_{\sigma_u, \sigma_v}^3 \Phi \left( -\frac{\sigma_v}{\sigma_u} \frac{1+2\sigma_u}{\sqrt{1+2\sigma_v^2}} - \frac{\varepsilon_i}{\sigma_v} \sqrt{1+2\sigma_v^2} \right) \right. \\
 &\quad \left. + \omega \Upsilon_{\sigma_u, \sigma_v}^4(\varepsilon_i) \Phi \left( -\frac{\sigma_v}{\sigma_u} \frac{1+3\sigma_u}{\sqrt{1+2\sigma_v^2}} - \frac{\varepsilon_i}{\sigma_v} \sqrt{1+2\sigma_v^2} \right) \right\},
 \end{aligned}$$

La obtención de la eficiencia técnica puede conseguirse usando las expresiones de Gómez-Déniz y Pérez-Rodríguez (2017), quienes calculan la distribución condicionada de  $u$  dado  $\varepsilon$ , la cual está dada por:

$$f_{\sigma_u, \sigma_v, \omega}(u_i | \varepsilon_i) = \frac{f_{\sigma_u, \sigma_v}(u_i | \varepsilon_i) + \omega \Psi_{\sigma_u, \sigma_v}^0(u_i | \varepsilon_i)}{1 + \omega \Psi_{\sigma_u, \sigma_v}^1(\varepsilon_i)}, \quad [4.15]$$

para  $u > 0$ , donde  $f_{\sigma_u, \sigma_v}(u_i | \varepsilon_i)$  está dado por [4.4] y,

$$\begin{aligned}
 \Psi_{\sigma_u, \sigma_v}^0(u_i | \varepsilon_i) &= \varphi_{\sigma_u}(u_i) \varphi_{\sigma_v}(u_i + \varepsilon_i) f_{\sigma_u, \sigma_v}(u_i | \varepsilon_i), \\
 \Psi_{\sigma_u, \sigma_v}^1(\varepsilon_i) &= \frac{1}{\sqrt{1+2\sigma_v^2}} \frac{1}{\Phi \left( \frac{\tilde{\mu}_i}{\sigma_v} \right)} \left[ \Phi \left( \frac{\tilde{\mu}_i}{\sigma_v} \right) \frac{e^{\frac{2\sigma_v^2}{1+2\sigma_v^2}}}{1+\sigma_u} - \Upsilon_{\sigma_u, \sigma_v}^2(\varepsilon_i) \Phi \left( \frac{\tilde{\mu}_i - \sigma_v}{\sigma_v} \right) \right. \\
 &\quad \left. - \frac{1}{1+\sigma_u} \Upsilon_{\sigma_u, \sigma_v}^3 \Phi \left( -\frac{\sigma_v}{\sigma_u} \frac{1+2\sigma_u}{\sqrt{1+2\sigma_v^2}} - \frac{\varepsilon_i}{\sigma_v} \sqrt{1+2\sigma_v^2} \right) \right. \\
 &\quad \left. + \Upsilon_{\sigma_u, \sigma_v}^4(\varepsilon_i) \Phi \left( -\frac{\sigma_v}{\sigma_u} \frac{1+3\sigma_u}{\sqrt{1+2\sigma_v^2}} - \frac{\varepsilon_i}{\sigma_v} \sqrt{1+2\sigma_v^2} \right) \right] \quad [4.16]
 \end{aligned}$$

A partir de las anteriores expresiones podemos obtener la fórmula del estimador puntual de eficiencia técnica dado por Battese y Coelli (1988), que está relacionado con la expectativa condicionada de  $e^{-u_i}$ , esto es,  $E(e^{-u_i} | \varepsilon = \hat{\varepsilon})$ . Así, la eficiencia técnica es:

$$TE_i = \frac{\Psi_{\sigma_u, \sigma_v, \omega}^{01}(\varepsilon_i) + \omega \Psi_{\sigma_u, \sigma_v}^{02}(\varepsilon_i)}{1 + \omega \Psi_{\sigma_u, \sigma_v}^1(\varepsilon_i)}, \quad i = 1, 2, \dots, n,$$

donde

$$\Psi_{\sigma_u, \sigma_v, \omega}^{01}(\varepsilon_i) = [1 + \omega \delta_u(\sigma_u) \delta_v(\sigma_v)] \frac{\Phi(-A_i - \sigma_v) e^{\sigma_v(A_i + \sigma_v/2)}}{\Phi(-A_i)},$$

y

$$\begin{aligned} \Psi_{\sigma_u, \sigma_v}^{02}(\varepsilon_i) = & \frac{1}{\Phi(-A_i)} \left[ \frac{1}{\sqrt{1 + 2\sigma_v^2}} \Upsilon_{\sigma_u, \sigma_v}^{11}(\varepsilon_i) \Phi\left(-\frac{A_i + 2(2 + \varepsilon_i)\sigma_v}{\sqrt{1 + 2\sigma_v^2}}\right) \right. \\ & - \delta_v(\sigma_v) \Upsilon_{\sigma_u, \sigma_v}^{22}(\varepsilon_i) \Phi(-A_i - 2\sigma_v) \\ & \left. - \frac{1}{\sqrt{1 + 2\sigma_v^2}} \delta_u(\sigma_u) \Upsilon_{\sigma_u, \sigma_v}^{33}(\varepsilon_i) \Phi\left(-\frac{A_i + \sigma_v(3 + 2\varepsilon_i)}{\sqrt{1 + 2\sigma_v^2}}\right) \right], \end{aligned}$$

con

$$\Upsilon_{\sigma_u, \sigma_v}^{11}(\varepsilon_i) = \exp\left\{-\frac{\varepsilon_i^2 - A_i \sigma_v (4 - A_i \sigma_v) - 8\sigma_v^2 + 2\varepsilon_i(1 - \sigma_v(A_i + 2\sigma_v))}{1 + 2\sigma_v^2}\right\},$$

$$\Upsilon_{\sigma_u, \sigma_v}^{22}(\varepsilon_i) = \exp\{2\sigma_v(A_i + \sigma_v)\},$$

$$\Upsilon_{\sigma_u, \sigma_v}^{33}(\varepsilon_i) = \exp\left\{-\frac{\varepsilon_i^2 - A_i \sigma_v (3 - A_i \sigma_v) - 9/2\sigma_v^2 + 2\varepsilon_i(1 - \sigma_v(A_i + \sigma_v))}{1 + 2\sigma_v^2}\right\},$$

donde  $A_i = -\tilde{\mu}_i/\sigma_v$ ,  $\tilde{\mu}_i = -\varepsilon_i - \sigma_v^2/\sigma_u$  y las expresiones  $\delta_u(\sigma_u)$ ,  $\delta_v(\sigma_v)$  y

$\Psi_{\sigma_u, \sigma_v}^1(\varepsilon_i)$  están dadas en [4.10], [4.11] y [4.16], respectivamente.

#### **4.2.1.4. MODELOS DE FRONTERA ESTOCÁSTICA EN DATOS DE PANEL**

Tal como señalan Kumbhakar y Lovell (2000) y Greene (2008), las principales cuestiones a las que los datos de panel proporcionan una solución en el contexto de los modelos de frontera estocástica, son:

- La relajación de los supuestos fuertes sobre la distribución del término de error compuesto ( $u$  y  $v$ ).
- La asignación de la correlación entre el término de ineficiencia e inputs.
- La estimación consistente de la ineficiencia condicional específica de la empresa con el estimador puntual JLMS, aunque éste no es consistente en los datos de corte transversal.

Los modelos de frontera estocástica de datos de panel se clasifican en dos grandes categorías: los modelos de efectos fijos y los modelos de efectos aleatorios, con los que se permite la estimación de la eficiencia, ya sea variable o invariante en el tiempo.

#### **4.2.1.5. MODELOS DE EFECTOS FIJOS Y ALEATORIOS**

Los manuales de texto clásicos sobre microeconometría (por ejemplo, Greene, 1998; Cameron y Trivedi, 2005) ilustran los modelos de efectos fijos y los modelos de efectos aleatorios en el ámbito de fronteras estocásticas. Sin embargo, algunos supuestos planteados en los modelos de efectos fijos para estimar las eficiencias pueden ser insostenibles.<sup>15</sup>

---

<sup>15</sup> Existen dos procedimientos para estimar un modelo de datos de panel: uno de ellos implica el



La decisión acerca de la estructura apropiada para el análisis, es decir, efectos fijos versus efectos aleatorios depende en parte de los siguientes aspectos:

- El sesgo de selección en los paneles no balanceados. En este contexto, los modelos de efectos fijos se pueden utilizar en paneles no balanceados si no hay una correlación entre la observación de una unidad en el panel y el término de error (Hayashi, 2000). Sin embargo, en SFA, donde el término de error incluye el término de ineficiencia, este problema se agrava en los paneles grandes, donde el sesgo de selección puede ser considerable.
- Los modelos de efectos fijos pueden capturar otros efectos invariables en el tiempo que no están incluidos en el modelo y, por construcción, se atribuyen a la ineficiencia (Greene, 2005b).

Estos dos argumentos apoyan la utilización de un modelo de efectos aleatorios si se desea obtener precisión en la estimación de la eficiencia.

Kumbhakar y Lovell (2000, p.106) afirman que un enfoque de estimación basado en efectos aleatorios con métodos de máxima verosimilitud es preferible al tratamiento de efectos fijos para los paneles relativamente cortos con un número importante de observaciones. Kim y Schmidt (2000) comparan varias especificaciones de frontera estocástica en el contexto de los métodos de panel, y observan que la estimación de efectos fijos se realiza generalmente de forma deficiente.

---

reconocimiento de que las variables omitidas pueden generar cambios en la media ya sea a través del tiempo o entre unidades de corte transversal, en este caso, el modelo de efectos fijos trata de aproximar estos cambios con variables *dummy*; el otro modelo es el de efectos aleatorios, que trata de capturar estas diferencias a través del componente aleatorio del modelo. La técnica de datos de panel permite contemplar la existencia de efectos individuales específicos a cada unidad de corte transversal, invariables en el tiempo que afectan la manera en que cada unidad de corte transversal toma sus decisiones.

Visto lo anterior, los modelos de frontera estocástica que presentaremos a continuación serán fundamentalmente modelos de efectos aleatorios, tanto de producción como costes.

### 1) Eficiencia invariante temporalmente

Pitt y Lee (1981) fueron los primeros autores en extender el modelo de frontera estocástica de corte transversal hacia una estructura de panel de datos.

Ellos propusieron la estimación máximo verosímil de dicho modelo considerando que la ineficiencia era invariante en el tiempo, la convolución del error era normal-half normal y el modelo es de efectos aleatorios. El modelo lineal puede escribirse como:

$$\begin{aligned} y_{it} &= f(x_{it}; \beta) + v_{it} \pm u_i, \quad i = 1, 2, \dots, n; t = 1, 2, \dots, T_i \\ &= \beta_0 + x'_{it} \beta + v_{it} \pm u_i \\ v_{it} &\sim N(0, \sigma_v^2) \\ u_i &\sim N^+(0, \sigma_u^2) \end{aligned}$$

donde  $\varepsilon_{it} = v_{it} \pm u_i$ ,  $T_i$  es el tamaño muestral temporal para la empresa  $i$ -ésima, que puede ser igual para todas las empresas (panel balanceado) o distinto (panel no balanceado). Se asume también que  $u_i$  y  $v_{it}$  se distribuyen de forma independiente el uno del otro y de las variables exógenas contenidas en el vector  $x_{it}$ .

La generalización de este modelo al caso normal-truncada normal ha sido propuesta por Battese y Coelli (1988). Cabe resaltar que, como ha señalado Schmidt y Sickles (1984), la estimación de un modelo de frontera estocástica con ineficiencia invariante en el tiempo también puede ser realizada mediante el estimador de efectos fijos, lo que permite correlacionar la ineficiencia con los regresores y evitar los supuestos distributivos sobre  $u_i$ .

Después de obtener las estimaciones máximo versímiles de todos los parámetros, el estimador JLMS puede utilizarse para obtener estimaciones de las eficiencias de cada empresa reemplazando  $\varepsilon_{it}$  con  $\bar{\varepsilon}_i$  y  $\sigma^2$  con  $\sigma^2/T$  (Kim and Schmidt, 2000).

## 2) Eficiencia variable en el tiempo

La naturaleza invariable en el tiempo del término de ineficiencia ha sido cuestionada, especialmente, en aplicaciones empíricas basadas en conjuntos grandes de datos de panel. Para relajar esta restricción, Cornwell et al. (1990) propusieron el siguiente modelo de frontera estocástica con ineficiencia variable temporalmente, tal que:

$$\begin{aligned} y_{it} &= \beta_0 + x'_{it}\beta + v_{it} \pm u_i, \quad i = 1, 2, \dots, n; t = 1, 2, \dots, T_i \\ v_{it} &\sim N(0, \sigma_v^2) \\ u_{it} &= \omega_i + \omega_{i1}t + \omega_{i2}t^2 \end{aligned}$$

donde la ineficiencia,  $u_{it}$ , posee parámetros individuales específicos desconocidos para cada unidad,  $\omega_i, \omega_{i1}, \omega_{i2}$ , y depende del tiempo ( $t$ ) en forma cuadrática. Sin embargo, el problema es que requiere la estimación de un gran número de parámetros para la ineficiencia:  $n \times 3$ .

Siguiendo una estrategia diferente, Lee y Schmidt (1993) propusieron una estimación alternativa en la que  $u_{it}$  está especificado por:

$$u_{it} = g(t)u_i$$

donde  $g(t)$  está representado por un conjunto de variables *dummy* temporales. Esta especificación es más parsimoniosa y no impone ninguna forma paramétrica, pero es menos flexible ya que restringe el patrón temporal de  $u_{it}$  a ser el mismo para todas las

unidades productivas. Kumbhakar (1990) fue el primero en proponer la estimación máximo verosímil de un modelo de frontera estocástica donde la ineficiencia varía en el tiempo. Así,  $g(t)$  puede especificarse como:

$$g(t) = [1 + \exp(\gamma t + \delta t^2)]^{-1}$$

Este modelo contiene solo dos parámetros para estimar,  $\gamma$  y  $\delta$ , y la hipótesis de ineficiencia invariante temporalmente es  $\gamma = \delta = 0$ .

Un modelo similar al de Kumbhakar (1990) fue propuesto por Battese y Coelli (1992), que lo denominaron modelo de decaimiento temporal (*time decay*), y que puede expresarse como:

$$g(t) = \exp(-\eta(t - T_i))$$

En este modelo solamente se requiere la estimación de un único parámetro,  $\eta$ , el cual representa al patrón de ineficiencia temporal. Battese y Coelli (1992) asumen que el modelo para el error compuesto es normal-truncada normal, de tal manera que la ineficiencia está definida por:

$$\begin{aligned} u_{it} &= g(t)u_i \\ &= \exp(-\eta(t - T_i))u_i, u_i \sim N^+(\mu, \sigma_u^2) \end{aligned}$$

Si  $\eta > 0$ , entonces la eficiencia aumenta con el tiempo (a un ritmo creciente) y por el contrario si  $\eta < 0$ , entonces la eficiencia técnica está disminuyendo con el tiempo. El modelo anida el modelo invariante en el tiempo de Pitt y Lee (1981), cuando  $\eta = \mu = 0$  (es decir, no hay variación temporal de la eficiencia), y el componente de ineficiencia sigue una distribución half-normal. Así, un contraste adecuado para el supuesto de invarianza temporal es analizar la hipótesis nula de que el parámetro  $\eta$  es cero. Si se rechaza dicha hipótesis nula, éste es a la vez significativo y distinto de cero, por lo que la eficiencia puede decirse que varía con el tiempo.

Un aspecto adicional que cabe hacer es que la ineficiencia puede estar relacionada con variables explicativas como, por ejemplo, la experiencia gerencial. Así, para identificar algunas de las razones de las diferencias en la eficiencia estimada entre las empresas de una industria, puede introducirse un procedimiento de estimación en donde la ineficiencia dependa de variables explicativas (véase el epígrafe 2.1.1 para algunas aproximaciones utilizadas en el ámbito del corte transversal).

Battese y Coelli (1995) proponen un modelo que es equivalente a la especificación de Huang y Liu (1994)<sup>16</sup> pero que difiere de Reifschneider y Stevenson (1991) en que los errores son independientes, pero no idénticamente distribuidos. Básicamente, Battese y Coelli (1995) consideran que  $u_{it} = \delta'z_{it} + g_i$ , siendo  $z_{it}$  un vector  $p \times 1$  de variables que pueden influenciar la eficiencia de la empresa,  $\delta$  es un vector  $1 \times p$  de parámetros desconocidos, y  $g_i \sim N(0, \sigma_g^2)$  estando limitada por debajo por la variable de truncamiento  $-\delta'z_{it}$ , es equivalente a  $u_{it} \sim N^+(\mu_{it}, \sigma_u^2)$ , donde  $\mu_{it} = \delta'z_{it}$ . Esto se debe a que el supuesto distribucional sobre  $g_i$  es consistente con que  $u_{it} \sim N^+(\mu_{it}, \sigma_u^2)$ . De esta forma, el modelo propuesto es normal-truncado normal. En este escenario, recordemos que  $\sigma^2 = \sigma_v^2 + \sigma_u^2$  y  $\gamma = \sigma_u^2 / \sigma^2$  (este parámetro debe estar comprendido entre 0 y 1) y, por lo tanto, contrastar la hipótesis de que la ineficiencia no está presente en el modelo es equivalente a evaluar  $H_o : \delta_1 = \dots = \delta_p = 0, \sigma^2 = 0, \gamma = 0$ . No rechazar la hipótesis nula implica que  $u_{it}$  debería ser eliminado del modelo.

---

<sup>16</sup> Battese y Coelli (1995) formularon un modelo que es esencialmente el mismo que Huang y Liu (1994), con dos excepciones: primero, el modelo es formulado en un entorno de panel de datos y, segundo, no incluyen inputs en su especificación de  $\delta'z_{it}$ .

La característica común de todos los modelos anteriores es que el término constante es el mismo para todas las unidades productivas, generando así un sesgo de especificación errónea en presencia de factores no observables invariantes en el tiempo, que no están relacionados con el proceso de producción pero que afectan a la producción. Como resultado, el efecto de estos factores puede ser capturado por el término de ineficiencia, produciendo resultados sesgados. En este sentido, Greene (2005a) desarrolla un modelo que permite estimar la ineficiencia variable en el tiempo sin imponer estructuras rígidas para la ineficiencia como hacen los modelos anteriores y considerando constantes específicas para cada unidad. Su modelo permite tener en cuenta la heterogeneidad no observable (factores no observados) que en otros modelos son, por construcción, incluidos en el término de ineficiencia, y que son constantes en el tiempo.

Los métodos propuestos por Greene (2005a), comparados con otros modelos y métodos, separan la eficiencia variable temporalmente y la heterogeneidad no observada que es constante en el tiempo introduciendo constantes específicas para cada unidad, tanto desde el punto de vista de los efectos fijos (TFE, *true fixed effects*) como aleatorios (TRE, *true random effects*), pero también en el modelo de frontera estocástico de parámetros aleatorios (*random parameters SF model*), que permite englobar a las formulaciones sobre efectos fijos y aleatorios introduciendo aleatoriedad en todos los parámetros de la tecnología, y no solo en el término constante.

La heterogeneidad no observada específica de la unidad se puede interpretar como una ineficiencia persistente (Kumbhakar, 1991; Kumbhakar y Heshmati, 1995 y Kumbhakar y Hjalmarsson, 1993 y 1995) o efectos firmes. Por ejemplo, Pitt y Lee (1981) desarrollaron específicamente la parte persistente de la ineficiencia, como Schmidt y Sickles (1984). El efecto transitorio de la ineficiencia está en Kumbhakar

(1990), Cornwell et al. (1990), Battese y Coelli (1992, 1995) o Greene (2005a, b). Recientemente Colombi et al. (2014) y Tsionas y Kumbhakar (2014) dan un paso más en esta literatura y dividen el término de error invariante en dos términos, uno para capturar la heterogeneidad latente de las empresas (Greene, 2005a, b) y el otro componente captura largo -función (persistente) de ineficiencia (Kumbhakar y Heshmati, 1995 y Kumbhakar y Hjalmarsson, 1993 y 1995). Filippini y Greene (2016) explican que la eficiencia productiva de una empresa puede verse como compuesta de dos partes, una persistente y otra transitoria. La parte persistente está relacionada con la presencia de problemas estructurales en la organización del proceso de producción de una empresa o la presencia de deficiencias sistemáticas en las capacidades gerenciales. La parte transitoria puede deberse a la presencia de problemas de gestión no sistemática que pueden resolverse a corto plazo.

Más recientemente, Filippini y Greene (2016) desarrollaron un estimador de probabilidad simulado máximo de información completa práctica para el modelo de Colombi et al. (2014) para estimar los términos de ineficiencia transitoria y persistente, y que amplía el modelo de TRE. Según lo sugerido por Tsionas y Kumbhakar (2014), este modelo se denomina modelo TRE generalizado (GTRE).

A continuación, comentaremos brevemente dichos modelos.

## 2.1) **El modelo TFE**

La especificación utilizada por Greene (2005a) permite aislar la ineficiencia variable en el tiempo de la heterogeneidad no observada invariante en el tiempo, que es específica de la unidad. Así, la siguiente especificación permitirá estimar la ineficiencia

que varía en el tiempo usando un modelo normal-half normal de frontera estocástica con constantes específicas de las oficinas, tal que:

$$\begin{aligned}
 y_{it} &= \alpha_i + x'_{it}\beta + \varepsilon_{it}, i = 1, 2, \dots, n, t = 1, 2, \dots, T \\
 \varepsilon_{it} &= v_{it} \pm u_{it} \\
 v_{it} &\sim N(0, \sigma_v^2) \\
 u_{it} &= |U_{it}|, U_{it} \sim N(0, \sigma_u^2)
 \end{aligned}$$

donde las variables observadas son  $y_{it}$ , que representa el coste (producción), y  $x_{it} = (x_{1,it}, \dots, x_{k,it})'$ , que es un vector de covariables que no incluyen la constante.  $\alpha_i$  es la heterogeneidad específica de cada unidad y  $\beta$  es un vector de parámetros desconocidos que deben estimarse. En nuestro modelo, el error compuesto es  $\varepsilon_{it} = v_{it} \pm u_{it}$ , el cual es asimétrico (el signo  $\pm$  dependerá de si la frontera es de costes (positiva) o de producción (negativa)), y donde  $v_{it}$  se asume que está distribuido i.i.d.  $N(0, \sigma_v^2)$  y  $u_{it}$  es el componente de error de una sola cola, que es no negativo y que es denominado usualmente término de ineficiencia. Este último término de error se assume que es una variable aleatoria i.i.d. siguiendo una distribución no Gaussiana (por ejemplo, half-normal, exponencial o truncada-normal). En nuestro caso, consideramos el caso normal-half normal donde  $u_{it} = |U_{it}|, U_{it} \sim N(0, \sigma_u^2)$ , lo que implica que la ineficiencia varía en el tiempo.

La estimación por máxima verosimilitud del modelo TFE requiere la solución de dos cuestiones principales relacionadas con la estimación de modelos de datos de panel no lineales. La primera es puramente computacional debido a la gran dimensión del espacio de parámetros. Sin embargo, Greene (2005a) mostró que el método de máxima verosimilitud con variables dummy (MLDV) es factible desde el punto de vista computacional también en presencia de un gran número de parámetros molestos



(*nuisance parameters*),  $\alpha_i$  ( $n > 1000$ ). El segundo, el llamado problema de los parámetros incidentales (Lancaster, 2002), es una cuestión inferencial que surge cuando el número de unidades es relativamente grande en comparación con la longitud del panel. En estos casos, las constantes específicas de cada unidad se calculan de forma inconsistente como para un  $T$  fijado, ya que sólo se utilizan  $T_i$  observaciones para estimar cada parámetro específico de la unidad. Tal y como han mostrado Belotti e Ilardi (2012), esta inconsistencia contamina los parámetros de la varianza, los cuales representan el ingrediente clave en la estimación de las ineficiencias, y de esta manera la estimación MLDV solo es apropiada cuando el panel es suficientemente grande ( $T \geq 10$ ).

El logaritmo de la función de verosimilitud para el modelo de efectos fijos de una función de costes puede escribirse como:

$$\log L(\alpha_1, \dots, \alpha_n, \beta, \lambda, \sigma) = \sum_{i=1}^n \sum_{t=1}^T \log \left[ \frac{1}{\Phi(0)} \Phi \left( -\frac{\varepsilon_{it} \lambda}{\sigma} \right) \phi \left( \frac{\varepsilon_{it}}{\sigma} \right) \right]$$

donde  $\varepsilon_{it} = v_{it} + u_{it}$ ,  $\phi$  y  $\Phi$  son la densidad y cdf de la distribución normal estándar,  $\lambda = \frac{\sigma_u}{\sigma_v}$ ,  $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$  y  $\varepsilon_{it} = y_{it} - \alpha_i - x'_{it} \beta$ . Greene (2005a) indica que ninguna operación de transformación producirá una función de verosimilitud libre de efectos fijos, por lo que es necesario estimar todos los  $n+k+2$  parámetros simultáneamente.

## 2.2) El modelo TRE y sus extensiones

Greene (2005a, página 287) especificó e introdujo un modelo de parámetros aleatorios de frontera estocástica.

La forma general de este modelo se escribe teniendo en cuenta las expresiones de frontera estocástica, distribución de ineficiencia y heterogeneidad de parámetros. Siguiendo a Filippini y Greene (2016), la especificación del modelo TRE puede escribirse como:

$$\begin{aligned} y_{it} &= \alpha_i + x'_{it}\beta + \varepsilon_{it}, i = 1, 2, \dots, n, t = 1, 2, \dots, T \\ \alpha_i &= \alpha + \omega_i, \omega_i \sim N(0, \sigma_\omega^2) \\ \varepsilon_{it} &= v_{it} \pm u_{it} \\ v_{it} &\sim N(0, \sigma_v^2) \\ u_{it} &= |U_{it}|, U_{it} \sim N(0, \sigma_u^2) \end{aligned}$$

donde  $\alpha_i = \alpha + \omega_i$ ,  $\omega_i \sim N(0, \sigma_\omega^2)$ , siendo  $\alpha$  un término común constante para todas las unidades. El término  $\omega_i \sim N(0, \sigma_\omega^2)$  es un término constante específico aleatorio que captura la heterogeneidad persistente no observada específica de la unidad (no la ineficiencia), y se distribuye normalmente con la media cero y la varianza subyacente  $\sigma_\omega^2$ .

En el marco del modelo TRE, cualquier componente invariante o persistente de la ineficiencia es completamente absorbido en el término constante individual-específico. La estimación del modelo TRE puede hacerse por máxima verosimilitud. Sin embargo, debido al hecho de que no hay una forma cerrada para la densidad del término de perturbación compuesto, la estimación se propone que sea realizada mediante la máxima verosimilitud simulada (MSL). En este caso, el logaritmo de verosimilitud puede escribirse como:

$$\log(\alpha, \beta, \lambda, \sigma, \sigma_\omega) = \sum_{i=1}^n \log \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^T \left[ \frac{2}{\sigma} \phi \left( \frac{y_{it} - \alpha - x'_{it}\beta - \sigma_\omega W_i}{\sigma} \right) \times \Phi \left( -\frac{(y_{it} - \alpha - x'_{it}\beta - \sigma_\omega W_i)\lambda}{\sigma} \right) \right] \right\} \phi(W_i) dW_i$$

donde  $\phi$  es la densidad normal estándar y  $\Phi$  es la cdf de la normal estándar - normal,

$\lambda = \frac{\sigma_u}{\sigma_v}$ ,  $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$  y  $\omega_i = \sigma_\omega W_i$  donde  $W_i$  está normalmente distribuido con

media cero y varianza unitaria. Debido a que la integral no tiene una solución en forma cerrada, dicha expresión puede evaluarse mediante simulación.

La función del logaritmo de verosimilitud simulada puede expresarse mediante la expresión:

$$\log(\alpha, \beta, \lambda, \sigma, \sigma_\omega) = \sum_{i=1}^n \log \frac{1}{R} \sum_{r=1}^R \left\{ \prod_{t=1}^T \left[ \frac{2}{\sigma} \phi \left( \frac{y_{it} - \alpha - x'_{it} \beta - \sigma_\omega W_{ir}}{\sigma} \right) \times \Phi \left( - \frac{(y_{it} - \alpha - x'_{it} \beta - \sigma_\omega W_{ir}) \lambda}{\sigma} \right) \right] \right\}$$

donde  $W_{ir}$  son  $R$  simulaciones de observaciones (*draws*) pertenecientes a una variable normal estándar. Estas *draws* pueden ser obtenidas mediante secuencias de Halton o secuencias generalizadas de Halton más que a través de pseudo-números aleatorios. También se simulan todas las derivadas para la optimización basada en gradiente y para calcular el estimador de la matriz de covarianza asintótica. El modelo es convencional y satisface las condiciones de regularidad en el marco de máxima verosimilitud.

El modelo clásico de efectos aleatorios (RE) puede fácilmente ser obtenido simplemente haciendo que  $\omega_i$  no exista en el modelo. De esta forma, el logaritmo de verosimilitud es:

$$\log(\alpha, \beta, \lambda, \sigma) = \sum_{i=1}^n \log \left[ \frac{2}{\sigma} \phi \left( \frac{\varepsilon_{it}}{\sigma} \right) \Phi \left( - \frac{\varepsilon_{it} \lambda}{\sigma} \right) \right]$$

Cabe señalar que el modelo TRE también puede sufrir el "sesgo de las variables omitidas", porque las variables no observadas pueden estar correlacionadas con los

regresores. Farsi et al. (2005) propusieron utilizar en el modelo TRE una ecuación auxiliar que permite considerar el problema econométrico del sesgo de heterogeneidad no observada.

Recientemente, Colombi (2010), Colombi et al. (2011, 2014) y Tsionas y Kumbhakar (2014) amplían la metodología de la frontera estocástica proponiendo un modelo de datos de panel que proporciona estimaciones separadas de los dos componentes de la eficiencia: la que es invariante en el tiempo o permanente y la que varía con el tiempo o es transitoria.<sup>17</sup> Este modelo puede ser llamado modelo generalizado TRE (GTRE). El modelo GTRE es simplemente un modelo TRE en el que el efecto invariante en el tiempo tiene una distribución *skew* normal, en lugar de una distribución normal. Filippini y Greene (2016), debido a que el modelo de Colombi et al. (2014) es computacionalmente más complicado, proporcionan un enfoque econométrico alternativo para la estimación del GTRE basado en la máxima verosimilitud simulada que permite la distinción entre la parte persistente y transitoria de la eficiencia.

Siguiendo a Filippini y Greene (2016), el modelo GTRE puede especificarse como sigue:

---

<sup>17</sup> Filippini y Greene (2016) exponen que la eficiencia productiva de una empresa puede considerarse compuesta de dos partes, una persistente y otra transitoria. La parte persistente está relacionada con la presencia de problemas estructurales en la organización del proceso productivo de una empresa o la presencia de deficiencias sistemáticas en las capacidades de gestión. La parte transitoria puede deberse a la presencia de problemas de gestión no sistemáticos que pueden resolverse a corto plazo.

$$\begin{aligned}
 y_{it} &= \alpha + (\omega_i - h_i) + x'_{it}\beta + \varepsilon_{it}, i = 1, 2, \dots, n, t = 1, 2, \dots, T \\
 \omega_i &\sim N(0, \sigma_\omega^2) \\
 h_i &= |H_i|, H_i \sim N(0, \sigma_h^2) \\
 \varepsilon_{it} &= v_{it} \pm u_{it} \\
 v_{it} &\sim N(0, \sigma_v^2) \\
 u_{it} &= |U_{it}|, U_{it} \sim N(0, \sigma_u^2)
 \end{aligned} \tag{4.17}$$

donde  $h_i$  es un efecto aleatorio que representa la ineficiencia persistente invariante temporalmente de la unidad  $i$ -ésima, siendo  $h_i = |H_i|$  y  $H_i$  distribuida normalmente con media cero y varianza  $\sigma_h^2$ .

El logaritmo de versomilitud puede ser escrito como:

$$\log(\alpha, \beta, \lambda, \sigma, \gamma, \theta) = \sum_{i=1}^n \log \int_{-\infty}^{\infty} \left\{ \prod_{t=1}^T \left[ \frac{2}{\sigma} \phi \left( \frac{y_{it} - \alpha - x'_{it}\beta - \delta_i}{\sigma} \right) \times \Phi \left( -\frac{(y_{it} - \alpha - x'_{it}\beta - \delta_i)\lambda}{\sigma} \right) \right] \right\} \frac{2}{\theta} \phi \left( \frac{\delta_i}{\theta} \right) \Phi \left( \frac{\delta_i}{\theta} \right) d\delta_i$$

donde  $\theta = (\sigma_\omega^2 + \sigma_h^2)^{0.5}$  y  $\delta_i = \sigma_\omega W_i - \sigma_h |H_i|$ . El logaritmo de máxima verosimilitud simulada está definido por:

$$\log(\alpha, \beta, \lambda, \sigma, \sigma_\omega, \sigma_h) = \sum_{i=1}^n \log \frac{1}{R} \sum_{r=1}^R \left\{ \prod_{t=1}^T \left[ \frac{2}{\sigma} \phi \left( \frac{y_{it} - \alpha - x'_{it}\beta - \delta_{ir}}{\sigma} \right) \times \Phi \left( -\frac{(y_{it} - \alpha - x'_{it}\beta - \delta_{ir})\lambda}{\sigma} \right) \right] \right\}$$

donde  $\delta_{ir} = \sigma_\omega W_{ir} - \sigma_h |H_{ir}|$ , siendo tanto  $W_{ir}$  como  $H_{ir}$  distribuidas normalmente con media cero y varianza unitaria.  $W_{ir}$  y  $H_{ir}$  son  $R$  observaciones simuladas normales usando la secuencia de Halton.

Las eficiencias técnicas individuales se pueden calcular por medio de una ligera modificación del estimador JLMS, como en Greene (2005b), donde la adaptación debe

hacerse con respecto tanto al error compuesto como  $w_i$ . Este modelo permite variar con el tiempo a la ineficiencia de una manera más flexible que en el modelo del Battese y Coelli (1992), incluyendo persistencia de la ineficiencia y heterogeneidad no observable, para evitar el sesgo en las estimaciones de los parámetros del modelo.

Por último, cabe destacar que los modelos SFA de parámetros aleatorios (por ejemplo, Tsionas, 2002; Greene, 2005a) permiten introducir la heterogeneidad entre las empresas mediante la inclusión de coeficientes específicos de las mismas. Concretamente, Greene (2005a) desarrolla un modelo de frontera estocástico de parámetros aleatorios (*random parameters SF model*), que engloba a los modelos TFE y TRE, y sus extensiones.

### **2.3) Modelos de parámetros aleatorios**

Esta sección se dedica a presentar la formulación de parámetros aleatorios que modela la heterogeneidad de la empresa en forma de variación continua de parámetros. Otro modelo que no presentaremos, pero que puede consultarse en Greene (2005a), es el modelo de clase latente que también puede verse como una aproximación a esto, en que la variación se genera por una distribución discreta.<sup>18</sup> Las formulaciones Bayesianas del modelo de frontera estocástico pueden ser vistas también como modelos de parámetros aleatorios. A este tipo de modelos Bayesianos dedicaremos la sección 2.2.

Greene (2005a) propone la siguiente representación general del modelo de parámetros aleatorios, tal que:

---

<sup>18</sup> Kumbhakar et al. (2013) han propuesto un modelo de frontera estocástico en un entorno en el que una fracción desconocida de la población opera sin ineficiencia. En el modelo de frontera estocástica, correspondería a un conjunto de empresas para las cuales  $\sigma_u = 0$ . Dado que no hay separación observable de los dos tipos de empresas, esto podría corresponder a una especificación de clase latente.

1) Modelo de frontera estocástica:

$$y_{it} = \beta_i' x_{it} + \varepsilon_{it}, \varepsilon_{it} = v_{it} \pm u_{it}; i = 1, 2, \dots, n, t = 1, 2, \dots, T$$

$$v_{it} \sim N(0, \sigma_v^2)$$

2) Distribución de la ineficiencia:

$$u_{it} = |U_{it}|, U_{it} \sim N(\mu_{it}, \sigma_{u,it}^2)$$

$$\mu_{it} = \delta_i' m_{it}$$

$$\sigma_{u,it}^2 = \sigma_u^2 \exp(\gamma_i' w_{it})$$

3) Heterogenidad de los parámetros

$$\beta_i = \beta + \Delta_\beta q_i + \Gamma_\beta v_{\beta i}$$

$$\delta_i = \delta + \Delta_\delta q_i + \Gamma_\delta v_{\delta i}$$

$$\gamma_i = \gamma + \Delta_\gamma q_i + \Gamma_\gamma v_{\gamma i}$$

En la estructura anterior, es preciso definir los elementos que están incorporados en dichas expresiones:

- $x_{it}, m_{it}$  y  $w_{it}$  son vectores de variables explicativas de orden  $k$ ,  $r_1$  y  $r_2$ , respectivamente
- $\beta, \delta, \gamma$  son vectores de orden  $k$ ,  $r_1$  y  $r_2$  de parámetros aleatorios, respectivamente.
- $\beta, \delta, \gamma$  son constantes fijas.
- $\Delta_j, j = \beta, \delta, \gamma$  es una matriz de coeficientes de las variables observadas que no varían en el tiempo,  $q_i$ , que es un vector de orden  $M \times 1$ . Si  $\Delta_\beta$ , la matriz es de orden  $k \times M$ .
- $\Gamma_j, j = \beta, \delta, \gamma$  es una matriz triangular inferior o diagonal que produce la matriz de covarianzas de los parámetros aleatorios:  $\Omega_j = \Gamma_j A \Gamma_j'$  siendo  $A$  la matriz de

covarianzas conocidas de  $v_{it}$ . Esta matriz puede reescribirse en el caso de que exista un proceso autorregresivo de orden 1 para el error:  $v_{it} = Rv_{it-1} + z_{it}$ , siendo  $R$  una matriz diagonal de coeficientes específicos de autocorrelación. Así,  $\Omega_j = \Gamma_j A^{0.5} (1 - R^2) A^{0.5} \Gamma_j'$ .

-  $v_{jt}$ ,  $j = \beta, \delta, \gamma$  es un vector de orden  $k$ ,  $r_1$  y  $r_2$ , respectivamente.

Considerando que  $\Theta$  es un vector de parámetros formado por todos los coeficientes de  $\beta_i, \delta_i, \gamma_i$  y también por los denominados *ancillary parameters*:  $\sigma_v^2$  y  $\sigma_u^2$ , la función del logaritmo de verosimilitud condicionada es:

$$\text{Log } L | v_1, \dots, v_n = \sum_{i=1}^n \sum_{t=1}^T \log f(\Theta / y_{it}, x_{it}, q_i, m_{it}, w_{it}, v_i)$$

Para estimar los parámetros es necesario integrar la heterogeneidad fuera de la verosimilitud. Al igual que los modelos TRE y GTRE, la expresión del logaritmo de verosimilitud no condicionado es:

$$\text{Log } L_S = \sum_{i=1}^n \int_{v_i} \sum_{t=1}^T \log f(\Theta / y_{it}, x_{it}, q_i, m_{it}, w_{it}, v_i)$$

no tiene una solución cerrada que permita la estimación sencilla de los parámetros. Por ello, se usa la máxima verosimilitud simulada, cuya expresión general es:

$$\text{Log } L_S = \sum_{i=1}^n \frac{1}{R} \sum_{r=1}^R \left[ \sum_{t=1}^T \log f(\Theta / y_{it}, x_{it}, q_i, m_{it}, w_{it}, v_i) \right]$$

y que en el modelo de fronteras estocásticas de parámetros aleatorios es igual a:



$$\text{Log } L_S = \sum_{i=1}^n \frac{1}{R} \sum_{r=1}^R \left\{ \begin{array}{l} \sum_{t=1}^T \log \Phi \left( \frac{\mu_{ir} / (\sigma_{uir} / \sigma_v) \pm [(y_{it} - \beta'_i x_{it}) (\sigma_{uir} / \sigma_v)]}{\sqrt{\sigma_{uir}^2 + \sigma_v^2}} \right) \\ - \frac{1}{2} \left( \frac{\mu_{ir} \pm (y_{it} - \beta'_i x_{it})}{\sqrt{\sigma_{uir}^2 + \sigma_v^2}} \right)^2 + \log \left( \frac{1}{\sqrt{2\pi}} \right) - \log \Phi \left( \frac{\mu_{ir}}{\sigma_{uir}} \right) - \log \sqrt{\sigma_{uir}^2 + \sigma_v^2} \end{array} \right\}$$

#### 4.2.2. ENFOQUE BAYESIANO

La aproximación Bayesiana a la frontera estocástica fue introducida por Van den Broeck et al. (1994). Estos autores presentan diversas formulaciones Bayesianas considerando diferentes distribuciones como, por ejemplo, la normal, gamma, exponencial y troncada normal. Las principales ventajas en este contexto son la inferencia exacta sobre las ineficiencias, especialmente con muestras pequeñas, la incorporación directa de restricciones, la especificación formal de la incertidumbre sobre los parámetros y el modelo, obteniendo una distribución de la ineficiencia para cada empresa, y el cálculo directo de las ineficiencias medias a través de la distribución predictiva a posteriori de  $u_i$ .

Los métodos de estimación usados para los modelos de fronteras estocásticas en entorno bayesiano requieren del uso de métodos de integración numérica como *Monte Carlo Markov Change* (MCMC). En particular, el algoritmo de muestreo de Gibbs con el aumento de datos introducido en Koop et al. (1995) se utiliza con bastante frecuencia en la literatura de fronteras estocásticas bayesianas. Griffin y Steel (2007) implementan

métodos MCMC para el análisis bayesiano de modelos de fronteras estocásticas utilizando el paquete WinBUGS.<sup>19</sup>

#### **4.2.2.1. MODELO BAYESIANO PARA CORTE TRANSVERSAL**

A continuación, especificaremos un modelo bayesiano para datos de corte transversal.

Partiendo del modelo de producción de un solo output con una función de producción  $f(x_i, \beta)$ , que es dependiente de un vector de inputs  $x_i$  siendo  $\beta$  el vector de parámetros de la tecnología de orden  $k \times 1$ , y en la que pueden añadirse un término de ruido  $v_i$  y también de ineficiencia  $u_i$ , entonces puede escribirse:

$$y_i = f(x_i, \beta) + v_i - u_i$$

Para construir el modelo de regresión Bayesiano, debemos definir el conjunto de distribuciones a priori para el vector de parámetros desconocido,  $\theta = [\beta, \sigma^2, \lambda]$ , donde siendo  $\sigma^2$  la varianza de  $\varepsilon_i = v_i - u_i$ , y  $\lambda$  es el parámetro de la distribución exponencial. Atendiendo a este vector, la estructura para las distribuciones a priori puede ser como sigue:

$$\begin{aligned} p(\theta) &= p(\beta, \sigma^2, \lambda) \\ &= p(\beta/\sigma^2) p(\sigma^2) p(\lambda) \end{aligned}$$

donde:

---

<sup>19</sup> Estos autores presentan códigos generales para datos transversales y de panel y discuten diversas maneras de resumir la inferencia a posteriori. Aunque WinBUGS puede no ser tan eficiente para modelos complicados, el artículo de Griffin y Steel (2007) facilita y hace accesible a los investigadores aplicados los principales métodos de inferencia bayesiana con modelos de fronteras estocásticas.

- La distribución a priori de  $\beta/\sigma^2$  es  $p(\beta/\sigma^2) \sim N(\beta_0, \sigma^2 W)$ , donde  $\beta_0 = 0$  y  $W = d_0 I_k$ , siendo  $I_k$  una matriz identidad de orden  $k \times k$ . El hiperparámetro  $d_0 = 10^4$  es un valor de referencia. Este valor puede cambiarse y elegirse, por ejemplo, puede emplearse un valor mayor. Este cambio no producirá un efecto significativo en la inferencia a posteriori.
- La distribución a priori de  $\sigma^2$  es  $p(\sigma^2) \sim IG(a_0/2, b_0/2)$ , siendo IG la Gamma inversa. En el contexto de panel de datos, Fernández et al. (1997) relajan dicha elección usan una distribución a priori no informativa.
- La distribución a priori de  $\lambda$  se basa en la distribución Gamma. En particular, la distribución a priori es  $p(\lambda^{-1}/\phi) \sim G(\phi, -\log r^*)$ , donde  $r^*$  es la eficiencia media implícita de la distribución a priori. Puede fijarse  $\phi = 1$  o puede asumirse que  $\phi \sim G(1,1)$  que está centrada a través de la media a priori sobre el valor que conduce a la distribución exponencial, y tiene una varianza a priori razonable para  $\phi$  igual a la unidad.

La densidad conjunta de  $(y, u)$ , donde  $y = (y_1, \dots, y_n)$  y  $u = (u_1, \dots, u_n)$ , está dada por:

$$f(y, u) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f(x_i, \beta) - u_i)^2}{2\sigma^2}\right) \times \frac{\lambda^{-\nu}}{\Gamma(\nu)} u_i^{\nu-1} \exp\left(-\frac{u_i}{\lambda}\right)$$

Después de marginalizar la función  $f(y, u)$  sobre  $u$ , la función de verosimilitud puede expresarse como:

$$L(\theta/y) \propto \prod_{i=1}^n \frac{\lambda^{-\nu}}{\Gamma(\nu)} \exp\left(\frac{\sigma^2}{2\lambda^2} + \lambda^{-1}(y_i - f(x_i, \beta))\right) \times \int_0^\infty u_i^{\nu-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - m_i)^2}{2\sigma^2}\right) du_i$$

donde  $m_i = y_i - f(x_i, \beta) - \sigma^2/\lambda$ .

La distribución a posteriori es igual al producto de las distribuciones a priori y la verosimilitud, de tal manera que:

$$\begin{aligned} p(\theta/y) &\propto L(\theta/y) \times p(\theta) \\ &= \prod_{i=1}^n \frac{\lambda^{-\nu}}{\Gamma(\nu)} \exp\left(\frac{\sigma^2}{2\lambda^2} + \lambda^{-1}(y_i - f(x_i, \beta))\right) \\ &\quad \times \int_0^\infty u_i^{\nu-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - m_i)^2}{2\sigma^2}\right) du_i \\ &\quad \times p(\beta/\sigma^2) p(\sigma^2) p(\lambda) \end{aligned}$$

Esta distribución es analíticamente intratable por lo que se hace necesario el uso de algoritmos sofisticados para simular, como es el caso del método de *Monte Carlo Markov Chain* (MCMC). En este método, el muestreo de Gibbs permite dibujar las observaciones aleatorias que caracterizan la distribución a posteriori. En el caso de las fronteras estocásticas, además, se hace necesario incrementar el número de datos para tratar las ineficiencias  $u$  como un vector de variables aleatorias latentes que son parámetros desconocidos que deben estimarse.

Por ejemplo, la distribución condicional para las ineficiencias puede escribirse como:

$$p(u_i/y_i, x_i, \theta) = \Phi^{-1}(m_i/\sigma) f_N(u_i/m_i, \sigma^2) I(u_i \geq 0)$$

donde  $I(u_i \geq 0)$  es una función indicadora que vale 1 cuando se cumple  $u_i \geq 0$ , y cero en caso contrario. Así, las distribuciones a posteriori de  $u_i$  y la eficiencia  $r_i = \exp(u_i)$  (eficiencia absoluta) pueden representarse mediante las siguientes expresiones, respectivamente:

$$p(u_i/y_i, x_i) = \int p(u_i/y_i, x_i, \theta) p(\theta/y_i, x_i) d\theta$$

$$p(r_i/y_i, x_i) = TE_i^{-1} \int p(u_i/y_i, x_i) du_i$$

La ineficiencia media, por otro lado, puede calcularse mediante la expresión:

$$p(u_f/y, x) = \int f_G(u_f/\lambda^{-1}) p(\lambda/y, x) d\lambda$$

Cabe resaltar que si los modelos son log-lineales, el muestreo de Gibb puede emplearse sin problema, por ejemplo, suponiendo que  $f(x_i, \beta) = x_i' \beta$ , es decir, sea una función lineal. En el caso de que los modelos no sean lineales, la inferencia Bayesiana a través de métodos de simulación de las funciones a posteriori es más compleja.

Los modelos de fronteras estocásticas también pueden extenderse al ámbito de los datos de panel, considerando los enfoques de efectos fijos y aleatorios, algunos de los cuales han sido vistos.

#### 4.2.2.2. MODELO BAYESIANO PARA PANEL DE DATOS

Algunos de los problemas estadísticos del análisis clásico (por ejemplo, inconsistencia de las estimaciones puntuales de la eficiencia específica de la empresa) pueden ser aliviadas con datos de panel y el supuesto de una forma distribucional particular para la distribución de la ineficiencia.

Siguiendo a Koop y Steel (1999), supongamos  $i=1, \dots, n$  empresas y  $t=1, \dots, T$  periodos de tiempo.  $y_i$  y  $v_i$  son  $T \times 1$ ,  $x_i$  es  $T \times k$  y no contiene un vector de unos.  $u_i$  es la eficiencia constante en el tiempo, con  $u$  es  $n \times 1$ .  $y = (y_1', \dots, y_n')'$  y  $v = (v_1', \dots, v_n')'$  son  $nT \times 1$ ,  $x = (x_1', \dots, x_n')'$  es  $nT \times k$ .

El modelo de datos de panel, considerando el modelo lineal de producción, puede escribirse como:

$$y_i = \beta_0 t_T + x_i \delta + v_i - u_i$$

donde  $\beta_0$  es una constante,  $\delta$  es un vector  $k \times 1$ ,  $t_T$  es un vector  $T \times 1$  de unos, y  $v_i \sim f_N^T(v_i/0, h^{-1}I_T)$ , de donde  $p(h) \propto h^{-1}$  es una distribución a priori no informativa para la varianza.

A continuación, distinguiremos las expresiones más generales utilizadas en el modelo de efectos fijos y aleatorios.

#### 4.2.2.3. EFECTOS FIJOS

El efecto específico individual puede escribirse como:  $\alpha_i = \beta_0 - u_i$ , siendo  $\alpha = (\alpha_1, \dots, \alpha_n)'$ . Luego, el modelo puede reescribirse como:

$$y_i = \alpha_i t_T + x_i \delta + v_i$$

con la distribución a priori:  $p(\alpha, \delta, h) \propto p(h) p(\delta) = h^{-1} p(\delta)$ .

Puesto que la inferencia directa sobre  $u_i$  no puede hacerse, y así tampoco sobre la eficiencia absoluta:  $\tau_i = \exp(-u_i)$ . Sin embargo, definiendo la ineficiencia relativa:

$u_i^{rel} = u_i - \min(u_j) = \max(\alpha_j) - \alpha_i$  puede obtenerse la eficiencia relativa

$r_i^{rel} = \exp(-u_i^{rel})$ . La empresa más eficiente es  $r_i^{rel} = 1$ . Se considera que

$$p(r_i^{rel}) \propto \frac{1}{r_i^{rel}}, r_i^{rel} \in (0,1).$$

La inferencia Bayesiana en el modelo de efectos fijos es fácil. Por ejemplo, a) puede asumirse que  $p(\delta)$  sea Uniforme, b) el modelo es un modelo de regresión lineal

Normal con distribuciones a priori de Jeffreys, c) el vector  $(\alpha', \delta')$  es  $t$ -Student con  $(T-1)-k$  grados de libertad; pero, si  $n$ ,  $T$  y  $k$  son muy grandes, es válida la aproximación Normal.

Las distribuciones a posteriori de los parámetros y la eficiencia relativa pueden construirse como sigue:

$$i) \quad p(r_i^{rel}/y, x) = \sum_{\substack{j=1 \\ j \neq i}}^n p(r_i^{rel}/y, x, r_j^{rel} = 1) \times p(r_j^{rel} = 1/y, x) \quad \text{siendo}$$

$p(r_j^{rel} = 1/y, x)$  la probabilidad de que la empresa  $j$ -ésima sea la más eficiente. Usando los resultados del modelo de regresión estándar Bayesiano en Judge et al. (1985, capítulo 4),  $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i \hat{\delta}$ .

$$ii) \quad \text{La marginal a posteriori de } \delta \text{ es } p(\delta/y, x) = f_n^k(\delta/\hat{\delta}, \hat{h}^{-1}\hat{s}^{-1})p(\delta).$$

#### 4.2.2.4. EFECTOS ALEATORIOS

El modelo Bayesiano de efectos aleatorios implica que:

$$y_i = \beta_i t_T + x_i \delta + v_i - u_i t_T$$

con distribución a priori,

$$p(\beta_0, \delta, h, u, \lambda^{-1}) \propto h^{-1} p(\delta) f_G(\lambda^{-1}/1, -\log(-\tau^*)) \times \prod_{i=1}^n f_G(u_i/1, \lambda^{-1})$$

con a prioris no informativas para  $\beta_0$  y  $h$  y las ineficiencias se distribuyen exponencialmente con media  $\lambda$ , tal que  $p(\lambda) = f_G(\lambda^{-1}/a_\lambda, b_\lambda)$ , pudiendo ser  $a_\lambda = 1$  y  $b_\lambda = -\log(\tau^*)$ , siendo  $\tau^*$  la mediana de la distribución a priori de las ineficiencias.

Definiendo  $\beta = (\beta_0, \delta')$  y  $X = (i_{nt} : x)$ , las distribuciones condicionadas a posteriori pueden representarse de la siguiente manera:

a) La precisión del error de medida es:

$$p(h/y, x, u, \beta, \lambda^{-1}) = f_G \left( h/\frac{nT}{2}, \frac{1}{2} [y - X\beta + (I_n \otimes i_T)u]' [y - X\beta + (I_n \otimes i_T)u] \right)$$

El vector de parámetros es:  $p(\beta/y, x, u, h, \lambda^{-1}) = f_N^{k+1} \left( \bar{\beta}/\bar{\beta}, h^{-1} (X'X)^{-1} \right) p(\delta)$

donde  $\bar{\beta} = (X'X)^{-1} [y + (I_n \otimes i_T)u]$

b) Las ineficiencias son:

$$p(u/y, x, \beta, h, \lambda^{-1}) = f_N^n \left( u/(i_n : \tilde{x}), \beta - \tilde{y} - (Th\lambda)^{-1} i_n, (Th)^{-1} I_n \right) \times \prod_{i=1}^n (u_i \geq 0)$$

donde  $\tilde{y} = (\bar{y}_1, \dots, \bar{y}_n)'$  y  $\tilde{x} = (\bar{x}_1, \dots, \bar{x}_n)'$ .

c) Finalmente, la distribución de  $\lambda^{-1}$  puede escribirse como:

$$p(\lambda^{-1}/y, x, u, \beta, h) = f_G \left( \lambda^{-1}/n + 1, u' i_n - \log(\tau^*) \right).$$

### 4.3. MODELOS DE FRONTERA ESTOCÁSTICA SEMIPARAMÉTRICOS Y NO PARAMÉTRICOS

#### 4.3.1. ENFOQUE SEMIPARAMÉTRICO

Park y Simar (1994), y Park et al. (1998, 2003) han explorado la estimación semiparamétrica de los modelos de fronteras estocásticas en el contexto de los datos de panel. Fan et al. (1996), Henderson y Simar (2005) y Kumbhakar et al. (2007) han explorado la estimación no paramétrica de fronteras estocásticas mediante la regresión kernel y técnicas locales de máxima verosimilitud. Sin embargo, una de las críticas a



estos enfoques en que la imposición de condiciones de regularidad (monotonidad, concavidad / convexidad) es muy difícil en este tipo de modelización.

A continuación, describiremos el modelo de Fan et al. (1996), con la finalidad de mostrar cómo es la estrategia de estimación en este tipo de enfoques.

El modelo semiparamétrico puede escribirse como:

$$\begin{aligned} y_i &= g(x_i) + \varepsilon_i, \quad \varepsilon_i = v_i \pm u_i, \quad i = 1, 2, \dots, n \\ v_i &\sim N(0, \sigma_v^2) \\ u_i &\sim N^+(0, \sigma_u^2) \end{aligned}$$

donde  $g(x_i)$  es una función alisada (*smooth*) desconocida por el investigador.

La función de verosimilitud del modelo anterior puede escribirse por:

$$\log L(\lambda, \sigma) = \frac{n}{2} \log \left( \frac{2}{\pi} \right) - n \log \sigma + \sum_{i=1}^n \log [1 - \Phi(\varepsilon_i \lambda / \sigma)] - \sum_{i=1}^n \frac{\varepsilon_i^2}{2\sigma^2}$$

donde  $\varepsilon_i = y_i - g(x_i)$ .

Si  $g(x_i)$  fuese una función completamente conocida, podrían obtenerse las condiciones de primer orden del problema de maximización del logaritmo de verosimilitud para  $\lambda$  y  $\sigma^2$ . Sin embargo, si  $g(x_i)$  no se conociese completamente entonces la obtención de  $\lambda$  y  $\sigma^2$  sería más compleja. Esto es debido a que  $E[y_i/x_i] = g(x_i) - E[u_i/x_i] \neq g(x_i)$ , por lo que no puede separarse  $g(x_i)$  de  $E[y_i/x_i]$ .

La solución que Fan et al. (1996) proponen es asumir que  $g(x_i) = E[y_i/x_i] + \mu$ ,

donde  $\mu = \frac{\sigma_u \sqrt{2}}{\sqrt{\pi}} = \frac{\sigma \lambda \sqrt{2}}{[\pi(1 + \lambda^2)]^{0.5}}$ . Así,  $E[y_i/x_i]$  puede estimarse consistentemente por

métodos no paramétricos.

El método empleado por Fan et al. (1996) consta de dos etapas:

1. La primera es computar la estimación kernel de la expectativa condicional de  $E[y_i/x_i]$  mediante la expresión:

$$\hat{E}[y_i/x_i] = \frac{\sum_{j=1}^n y_j K\left(\frac{(x_i - x_j)}{h}\right)}{\sum_{j=1}^n K\left(\frac{(x_i - x_j)}{h}\right)},$$

siendo  $K$  el kernel (por ejemplo, del tipo de Epanechnikov, Gaussiano, rectangular, o Nadaraya-Watson, entre otros) y el valor  $h$ , denominado ventana, representa el área de influencia que se le pretende dar a cada observación muestral. Su elección suele depender de la dispersión de los valores.

2. Y, la segunda etapa supone construir la pseudo-verosimilitud semiparamétrica para estimar  $\lambda$ , tal que:

$$\log L(\lambda) = -n \log \hat{\sigma} + \sum_{i=1}^n \log [1 - \Phi(\hat{\varepsilon}_i \lambda / \hat{\sigma})] - \sum_{i=1}^n \frac{\hat{\varepsilon}_i^2}{2\hat{\sigma}^2}$$

donde el error estimado es  $\hat{\varepsilon}_i = y_i - \hat{E}[y_i/x_i] - \mu(\hat{\sigma}^2, \lambda)$  y donde la desviación es igual

$$a \hat{\sigma} = \left[ \frac{\frac{1}{n} \sum_{i=1}^n [y_i - \hat{E}[y_i/x_i]]^2}{\left(1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}\right)} \right]^{0.5}. \text{ Así, podrá obtenerse } \hat{\lambda}.$$

De esta forma, la frontera estocástica puede ser consistentemente estimada mediante

$$\hat{g}(x_i) = \hat{E}[y_i/x_i] + \hat{\mu}, \text{ donde } \hat{\mu} = \frac{\hat{\sigma} \hat{\lambda} \sqrt{2}}{\left[\pi(1+\hat{\lambda}^2)\right]^{0.5}}.$$

### **4.3.2. ANÁLISIS ESTOCÁSTICO NO PARAMÉTRICO ENVOLVENTE DE DATOS (STONED)**

La estimación no paramétrica desarrollada por Kuosmanen (2006) y Kuosmanen y Kortelainen (2012), llamada Análisis Estocástico No Paramétrico Envolverte de Datos (StoNED), puede verse como una síntesis de las técnicas DEA y SFA, incluyendo tanto la estimación frontera no paramétrica inherente al método DEA y el elemento estocástico incluido en la metodología o enfoque SFA.

StoNED no es el primer intento que combina las características del modelo DEA y del SFA, sino que contribuye a una larga serie de estudios previos que han perseguido objetivos similares (Park y Simar, 1994; Fan et al., 1996; Park et al., 1998, 2003; Henderson y Simar, 2005; y Kumbhakar et al., 2007). Sin embargo, la mayoría de estos modelos han demostrado ser computacionalmente muy difíciles o, de otro modo, muy complejos para trabajar con ellos. Como también se ha señalado por Kuosmanen (2006), aparte de los problemas computacionales, ninguno de los modelos propuestos produce resultados que satisfacen los axiomas básicos de producción. Siguiendo a Kuosmanen et al. (2015), estos autores dicen que el marco unificado de StoNED ofrece una plataforma general y flexible para el análisis de la eficiencia y temas relacionados como la estimación de la frontera y el análisis de la producción, permitiendo combinar las herramientas existentes de análisis de eficiencia en nuevas formas a través del espectro DEA-SFA, facilitando nuevas oportunidades de desarrollos metodológicos.

El modelo más próximo a StoNED es el modelo de máxima verosimilitud restringida desarrollado por Banker y Maindiratta (1992), donde la frontera no paramétrica de estilo DEA se combina con la inclusión del ruido y los términos de

ineficiencia estilo SFA. En la práctica, sin embargo, el problema de máxima verosimilitud restringida de Banker y Maindiratta es extremadamente difícil de resolver.

El método StoNED se aplica tanto a los datos de corte transversal como a los de panel. En los datos de panel, los componentes de ineficiencia invariables en el tiempo se pueden estimar de una manera completamente no paramétrica recurriendo a los efectos fijos o efectos aleatorios. En el corte transversal, son necesarios algunos supuestos distributivos adicionales para identificar la ineficiencia del ruido.

El modelo StoNED se estima mediante la regresión no paramétrica de los mínimos cuadrados no lineales convexos (CNLS) (Hildreth, 1954; Hanson y Pledger, 1976; Groeneboom et al., 2001), que no requiere de ningún parámetro de suavizado.

El método StoNED consta de dos etapas. En la primera etapa, se estima la forma de la función de producción/costes por CNLS, sin hacer ninguna suposición sobre la forma funcional, distributiva o suavidad del alisado. CNLS proporciona un estimador para la forma de la frontera de producción/costes, pero la ineficiencia y los términos de ruido siguen siendo indistinguibles. Por lo tanto, en la segunda etapa se usa el método de Fan et al. (1996) imponiendo algunas suposiciones distribucionales estándar adoptadas de la literatura SFA, y se estima el valor esperado condicional del término de ineficiencia utilizando el método de momentos o técnicas de pseudo-verosimilitud.

Formalmente, partiendo del modelo de frontera estocástica sigue:

$$y_i = f(x_i) + \varepsilon_i, \varepsilon_i = v_i \pm u_i, i = 1, 2, \dots, n$$

la estimación en dos etapas se realizaría como:

**1ª etapa:** Estimar por CNLS para obtener los residuos,  $v_i = y_i - g(x_i)$ . Dado que  $E[\varepsilon_i] = [-u_i] = -\mu < 0$  es necesario reescribir el modelo para que se cumplan las

condiciones de Gauss-Markov, tal que:  $y_i = [f(x_i) - \mu] + [\varepsilon_i - \mu] = g(x_i) + v_i$ , siendo  $g(x_i)$  la función de producción frontera de “práctica promedio” (*average practice*) en vez de la función de producción frontera de la “mejor práctica” (*best practice*).

El siguiente problema de programación cuadrática en un corte transversal, que se formula como sigue:

$$\min_{v, \alpha, \beta} \sum_{i=1}^n v_i^2$$

sujeto a:

$$y_i = \alpha_i + \beta_i' x_i + v_i$$

$$\alpha_i + \beta_i' x_i + v_i \leq \alpha_h + \beta_h' x_h + v_h, h = 1, \dots, n$$

$$\beta_i \geq 0, i = 1, \dots, n$$

permite la estimación de los residuos  $v_i$ . La primera restricción puede interpretarse como la ecuación de regresión lineal. El coeficiente  $\alpha_i$  y el vector de parámetros  $\beta_i$  son específicos de cada ecuación y no son los parámetros de  $g(x_i)$ . Por otro lado, la restricción de desigualdad puede interpretarse como el sistema de desigualdades de Afriat (1972).

**2ª etapa:** Dados los residuos  $\hat{v}_1, \dots, \hat{v}_n$  y asumiendo una distribución half-normal para la ineficiencia,  $u_i \sim N^+(0, \sigma_u^2)$ , y normal para el ruido,  $v_i \sim N(0, \sigma_v^2)$ , pueden emplearse tanto el método de los momentos como el método de pseudo máxima verosimilitud de Fan et al. (1996) para estimar  $\sigma_u^2$  y  $\sigma_v^2$ .

En el caso del método propuesto por Fan et al. (1996), el logaritmo de verosimilitud para estimar  $\lambda$  es:

$$\log L(\lambda) = -n \log \hat{\sigma} + \sum_{i=1}^n \log [1 - \Phi(\hat{\varepsilon}_i \lambda / \hat{\sigma})] - \sum_{i=1}^n \frac{\hat{\varepsilon}_i^2}{2\hat{\sigma}^2}$$

donde  $\hat{\varepsilon}_i = \hat{v}_i - (\hat{\sigma} \lambda \sqrt{2}) / \left[ 1 - \frac{2\lambda^2}{\pi(1+\lambda^2)} \right]^{0.5}$  y  $\hat{\sigma} = \left[ \left( \frac{1}{n} \sum_{i=1}^n \hat{v}_i^2 \right) / \left( 1 - \frac{2\lambda^2}{\pi(1+\lambda^2)} \right) \right]^{0.5}$

A partir de las estimaciones de  $\hat{\lambda}$  y  $\hat{\sigma}$  pueden obtenerse  $\hat{\sigma}_u^2 = \hat{\sigma} \hat{\lambda} / (1 + \hat{\lambda})$  y  $\hat{\sigma}_v^2 = \hat{\sigma} / (1 + \hat{\lambda})$ , y a partir de ellas obtener las eficiencias construyendo  $\hat{E}[u_i / \hat{\varepsilon}_i]$ , tal que:

$$\hat{E}[u_i / \hat{\varepsilon}_i] = -\frac{\hat{\varepsilon}_i \hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} + \frac{\hat{\sigma}_u^2 \hat{\sigma}_v^2}{\hat{\sigma}_u^2 + \hat{\sigma}_v^2} \left[ \frac{\phi(\hat{\varepsilon}_i / \hat{\sigma}_v^2)}{1 - \Phi(\hat{\varepsilon}_i / \hat{\sigma}_v^2)} \right]$$

donde  $\hat{\varepsilon}_i = \hat{v}_i - \hat{\sigma}_u^2 \sqrt{2/\pi}$  es el estimador del error compuesto. Este estimador es insesgado, pero inconsistente, cualquiera que sea el tamaño  $n$ .

Kuosmanen and Kortelainen (2012) extienden este método también al caso del panel de datos con ineficiencia no variable en el tiempo, reformulando el modelo de fronteras estocásticas en términos de desviaciones de la variable endógena, para evitar problemas de autocorrelación y heterocedasticidad (véase la página 19).

El método StoNED difiere de los tratamientos paramétricos, semiparamétricos y no paramétricos en el enfoque SFA en que no se hace ninguna suposición de forma funcional o suavidad de la función, sino que se basa en las limitaciones globales de la forma funcional (monotonidad, concavidad). En comparación con DEA, el método StoNED difiere en su robustez a los valores extremos y observaciones extremas y en su tratamiento probabilístico de la ineficiencia y el ruido. Mientras que la frontera DEA es una envolvente no estocástica de datos, StoNED conecta SFA y DEA mediante el desarrollo de un marco que abarca ambos métodos, y también casos especiales.

Una comparación del método StoNED contra otras técnicas clásicas ha sido realizado por Andor y Hesse (2014), quienes concluyen que el modelo funciona muy bien bajo distintos tipos de simulaciones de Monte Carlo, en particular son superiores en situaciones en las que los datos en cuestión están sujetos a un ruido importante.

Por último, cabe destacar que StoNED ha sido adoptado como el método de análisis de la eficiencia por la Autoridad Finlandesa de Energía (EMVI, Energiamarkkinavirasto) para el periodo regulatorio 2012-2015 (Kuosmanen et al., 2010).

#### **4.4. METAFRONTAS**

Las metafronteras también pueden estimarse en el ámbito de las fronteras estocásticas, al igual que se ha hecho con los modelos DEA.

Battese et al. (2004) y O'Donnell et al. (2008) propusieron un método basado en un procedimiento en dos etapas para estimar la metafrontera. Estos autores combinan una primera etapa que usa el método basado en las fronteras estocásticas con una segunda etapa basada en técnicas de programación matemática para estimar la metafrontera. Sin embargo, la principal dificultad de su procedimiento es que no existen propiedades estadísticas de los estimadores de la metafrontera. Los gaps de tecnología pueden estar contaminados de *shocks* aleatorios en la programación matemática. Por este motivo, Huang et al. (2014) propusieron un método para estimar la metafrontera que difiere del presentado por Battese et al. (2004) y O'Donnell et al. (2008) considerando el siguiente procedimiento que modifica la segunda etapa de los anteriores autores incluyendo la estimación en el ámbito de las fronteras estocásticas.

Antes de describir el procedimiento de estimación propuesto por Huang et al. (2014) se describirá la formulación del marco teórico que caracterizará el entorno para aplicar el procedimiento de estimación en dos etapas.

#### 4.4.1. FORMULACIÓN DEL MARCO TEÓRICO

La frontera de producción específica de grupo puede representarse por la siguiente ecuación:

$$Y_{jit} = f_t^j(X_{jit}) \exp(v_{jit} - u_{jit}), i = 1, \dots, n, t = 1, \dots, T$$

donde  $Y_{jit}$  es un output de la  $i$ -ésima unidad del periodo  $t$  en el  $j$ -ésimo grupo,  $X_{jit}$  es un vector de inputs la  $i$ -ésima unidad del periodo  $t$  en el  $j$ -ésimo grupo,  $f_t^j(X_{jit})$  es la función frontera de producción específica de cada grupo dependiente del vector de inputs,  $v_{jit}$  es el ruido que se distribuye  $N(0, \sigma_v^{j2})$  y  $u_{jit}$  es el término de ineficiencia que se distribuye truncada normal tal que  $TN(\mu^j(Z_{jit}), \sigma_u^{j2}(Z_{jit}))$  siendo  $Z_{jit}$  un vector de variables explicativas.

Visto lo anterior, la eficiencia técnica puede escribirse como:

$$TE_{it}^j = \frac{Y_{jit}}{f_t^j(X_{jit}) \exp(v_{jit})} = \exp(-u_{jit}) \leq 1$$

La función de producción metafrontera es  $f_t^M(X_{jit})$  y abarca todas las fronteras específicas de grupo o individuales. De esta forma, podemos escribir la siguiente relación:

$$f_t^j(X_{jit}) = f_t^M(X_{jit}) \exp(-u_{jit})$$



donde  $u_{jit} \geq 0$  implica que  $f_t^M(X_{jit}) > f_t^j(X_{jit})$ . Así, la ratio de la  $j$ -ésima función de producción a la frontera, que se ha definido como ratio del gap tecnológico en el Capítulo 2, puede escribirse como:

$$TGR_{it}^j = \frac{f_t^j(X_{jit})}{f_t^M(X_{jit})} = \exp(-u_{jit}) \leq 1$$

De esta forma, la ratio de eficiencia técnica de la empresa con respecto a la metafrontera de la tecnología de producción es:

$$MTE_{it}^j = \frac{Y_{jit}}{f_t^M(X_{jit}) \exp(v_{jit})} = TGR_{it}^j \times TE_{it}^j \leq 1$$

que podemos relacionarlo con la expresión del apartado 6.3 del Capítulo 2,  $TE_O^M(x, y)$ .

#### 4.4.2. ESTIMACIÓN

La estimación de la metafrontera se realiza en dos pasos. Estos son los siguientes:

**Etapa 1<sup>a</sup>**: Estimar el modelo log-lineal definido por:

$$\log Y_{jit} = \log f_t^j(X_{jit}) + v_{jit} - u_{jit}, i = 1, \dots, n, t = 1, \dots, T$$

para cada  $j = 1, \dots, J$ , con la finalidad de obtener la estimación específica de grupo,  $\log Y_{jit}$ . Por tanto, deben estimarse  $J$  regresiones.

En esta etapa, se emplea el modelo de Battese y Coelli (1995) para datos de panel con eficiencias variables temporalmente, con la intención de estimar las eficiencias técnicas,

$$TE_{it}^j = \hat{E} \left[ \exp(-u_{jit}) / \hat{\varepsilon}_{jit} \right] \leq 1,$$

donde  $\hat{\varepsilon}_{jit} = \log Y_{jit} - \log f_t^j(X_{jit})$ .

**Etapa 2<sup>a</sup>:** Construir un pool de datos con las observaciones estimadas en la etapa 1<sup>a</sup> para cada grupo para regresar:

$$\log f_t^j(X_{jit}) = \log f_t^M(X_{jit}) + v_{jit}^M - u_{jit}^M, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad j = 1, \dots, J$$

donde,

$$TGR_{it}^j = \hat{E} \left[ \exp(-u_{jit}^M) / \hat{\varepsilon}_{jit}^M \right] \leq 1$$

siendo  $\hat{\varepsilon}_{jit}^M = \log f_t^j(X_{jit}) - \log f_t^M(X_{jit})$ .

Así, combinando los resultados de la etapa 1<sup>a</sup> y 2<sup>a</sup>, la ratio de eficiencia técnica de la empresa a la metafrontera puede calcularse como:

$$MTE_{it}^j = TGR_{it}^j \times TE_{it}^j \leq 1$$

## 4.5. UN MODELO DE FRONTERA DE PRODUCCIÓN ESTOCÁSTICA CON

### FUNCIÓN DE DISTANCIA

Las funciones de producción usando la función distancia se utilizan especialmente cuando hay varios outputs e inputs.

Siguiendo a Assaf y Magnini (2012), asumiendo que la tecnología satisface los axiomas de Färe et al. (1994), y de acuerdo con Färe and Primont (1995), la tecnología definida para  $M$  outputs y  $K$  inputs, puede escribirse como:

$$\begin{aligned}\log y_{M,it} &= \alpha + \sum_{k=1}^K \alpha_k \log x_{k,it} + (v_{it} - u_{it}) \\ v_{it} &\sim N(0, \sigma_v^2) \\ u_{it} &\sim N^+(\mu_{it}, \sigma_u^2)\end{aligned}$$

donde  $\log$  es the logaritmo natural,  $i = 1, 2, \dots, n$  es el número de DMUs,  $t = 1, 2, \dots, T_i$  es el número de períodos temporales (que pueden ser distintos para cada DMU).  $\alpha$  es una constante común para todas las DMUs.  $v_{it}$  es un proceso i.i.d.  $N(0, \sigma_v^2)$ , y  $u_{it}$  es un componente de error no-negativo y de una sola cola (término de ineficiencia) que se asume i.i.d. por una distribución Normal truncada.:  $u_{it}$  es distribuida como  $N^+(\mu_{it}, \sigma_u^2)$ , donde  $\mu_{it} = \delta' z_{it}$  es la media del término de ineficiencia que depende de un vector de variables  $z_{it}'$  y un vector de parámetros desconocidos  $\delta$ . Este hecho implica que la eficiencia sea variable en el tiempo.

Dicha expresión puede representarse alternativamente usando la función distancia con respecto al output, tal que:  $D_0(x, y) = \min \{ \theta : (y/\theta) \in P(x) \}$ . Esta función es no decreciente, positiva y linealmente homogénea, y convexa en  $y$ . Toma el valor menor o igual a 1 si el vector de outputs,  $y$ , es un elemento del conjunto de producción factible,  $P(x)$ . Es decir,  $D_0(x, y) \leq 1$  si  $y \in P(x)$ . Aún más,  $D_0(x, y) = 1$  si  $(x, y)$  pertenece a la “frontera” del conjunto de posibilidades de producción.

Si elegimos la función de producción translog, para la estimación de la función distancia, entonces el modelo de fronteras estocásticas para dicha función puede escribirse como:

$$\log D_{0,it} = \alpha + \sum_{m=1}^M \gamma_m \log(y_{m,it}) + \sum_{k=1}^K \alpha_k \log x_{k,it} + (v_{it} - u_{it})$$

La restricción requerida para imponer la homogeneidad de grado 1 en outputs es:

$\sum_{m=1}^M \gamma_m = 1$ . Las condiciones de homogeneidad y simetría pueden expresarse mediante la

normalización de la función distancia dividiendo por uno de los outputs. Si el  $M$ -ésimo output es arbitrariamente elegido, entonces la formulación adopta la siguiente forma:

$$\log(D_{0,it}/y_{M,it}) = \alpha + \sum_{m=1}^{M-1} \gamma_m \log(y_{m,it}/y_{M,it}) + \sum_{k=1}^K \alpha_k \log x_{k,it} + (v_{it} - u_{it})$$

que puede reescribirse alternativamente como:

$$-\log(y_{M,it}) = \alpha + \sum_{m=1}^{M-1} \gamma_m \log(y_{m,it}/y_{M,it}) + \sum_{k=1}^K \alpha_k \log x_{k,it} + (v_{it} + u_{it})$$

donde  $u_{it} = -\log(D_{0,it})$ . Para más detalles, ver Coelli et al. (2005).

Extendiendo la especificación anterior a una función de producción tranlog similar a Elin y Luhnén (2010) en la industria de los seguros, o Assaf y Magnini (2012), y Assaf y Barros (2013) para la industria hotelera, podemos escribir el siguiente modelo que también incluye el progreso tecnológico junto a varias variables explicativas (inputs y outputs). El modelo puede escribirse como:

$$\begin{aligned} -\log y_{M,it} = & \alpha + \sum_{m=1}^{M-1} \gamma_m \log(y_{m,it}/y_{M,it}) + \frac{1}{2} \sum_{j=1}^{M-1} \sum_{m=1}^{M-1} \gamma_{jm} \log(y_{j,it}/y_{M,it}) \log(y_{m,it}/y_{M,it}) \\ & + \sum_{k=1}^K \alpha_k \log x_{k,it} + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^K \beta_{kj} \log x_{k,it} \log x_{j,it} + \sum_{k=1}^K \sum_{m=1}^{M-1} \beta_{km} \log x_{k,it} \log(y_{m,it}/y_{M,it}) \\ & + \kappa_1 t + \kappa_2 t^2 + (v_{it} - u_{it}) \\ v_{it} \sim & N(0, \sigma_v^2) \\ u_{it} \sim & N^+(\mu_{it}, \sigma_u^2) \end{aligned}$$

La estimación de modelo puede realizarse por máxima verosimilitud.



# CAPÍTULO 5

## REVISIÓN DE LA LITERATURA SOBRE EFICIENCIA BANCARIA Y SUCURSALES

---

### 5.1. INTRODUCCIÓN

Los estudios sobre la eficiencia bancaria son abundantes, tanto usando los métodos DEA como el análisis de frontera estocástica (SFA), bien realizando el análisis por separado, comparando los resultados obtenidos por ambas técnicas, estudiando la eficiencia en cada país o realizando comparaciones internacionales entre diferentes países.

Los enfoques utilizados para evaluar el desempeño de los bancos o sucursales bancarias desde un punto de vista metodológico suelen realizarse, en general, a partir del estudio de tres dimensiones del desempeño: la producción, la intermediación y la rentabilidad (véase el Capítulo 6 para más detalles). En cada uno de esos enfoques, la eficiencia analizada ha sido, generalmente, la de costes, si bien también se han analizado, aunque en menor medida, la eficiencia técnica y la eficiencia de beneficios.

En este capítulo haremos una breve revisión de la literatura de los estudios sobre la eficiencia de los bancos (que es la mayormente analizada) y también sobre las oficinas bancarias.

## **5.2. EFICIENCIA BANCARIA**

En esta sección describiremos los diferentes estudios realizados para estimar la eficiencia bancaria no solamente desde el punto de vista de la técnica utilizada sino de la comparación nacional o internacional de bancos.

### **5.2.1. LA EFICIENCIA A TRAVÉS DE DISTINTAS TÉCNICAS**

En esta sección describimos algunos trabajos que han analizado la comparación de diversas técnicas para estimar las eficiencias usando datos de un país.

En relación con los artículos que han comparado diversas técnicas para la obtención de las eficiencias, el primer estudio que compara técnicas alternativas de frontera es Ferrier y Lovell (1990). Estos autores analizaron la estructura de costes de 575 bancos de Estados Unidos para el año 1984 utilizando tanto las metodologías SFA como DEA. Ellos encontraron puntuaciones de eficiencia superiores con DEA en relación con SFA (80% y 74%, respectivamente), llegando a la conclusión que DEA es suficientemente flexible para envolver los datos más de lo que lo hace la frontera de costes translog. No obstante, las puntuaciones de eficiencia no están suficientemente correlacionadas, lo que indicaría la existencia de otros factores no controlados.

Sheldon (1994) analizó la eficiencia de costes de los bancos suizos con fronteras estocásticas y DEA en el período de 1987 a 1991. Si bien los resultados del DEA indican que el grado medio de la eficiencia de costes está alrededor del 56%, SFA produce sólo un 3,9% de eficiencia media. Esta desviación sustancial de las magnitudes obtenidas arroja algunas dudas en cuanto a una especificación adecuada de la función de costes (Amel et al., 2004). Además, el estudio también encuentra que la correlación de

rangos es del 1%, lo que indica que prácticamente no existe relación entre las eficiencias DEA y SFA estimadas.

Resti (1997) ofreció resultados muy diferentes. Este autor analizó la eficiencia de costes de 270 bancos italianos durante el período de 1988 a 1992, comparando las puntuaciones de eficiencia paramétricas y no paramétricas y, encontró que los resultados de ambas no difirieron sustancialmente. Por otra parte, este autor encontró puntuaciones de eficiencia más altas que Ferrier y Lovell (1990) y Sheldon (1994): entre 81% y 92% para SFA y entre un 60% y 78% para DEA. Además, la correlación de rangos entre SFA y DEA oscila entre el 44% y el 58%, y es estadísticamente significativa al nivel del 1%.

Berger y Humphrey (1997) realizaron una revisión de 130 estudios que estimaban las eficiencias bancarias en 21 países hasta 1996. En su estudio se evaluaron críticamente tanto los métodos de estimación como sus resultados. En general, encontraron dos metodologías ampliamente utilizadas: 1) los métodos paramétricos (69 artículos) y, 2) los métodos no paramétricos (60 artículos). Berger y Humphrey (1997) encontraron diferencias sustanciales en la eficiencia de los costes de los países, aunque ninguno de ellos considera la eficiencia de los beneficios. Dado que se centran exclusivamente en el lado de los costes, los resultados deben interpretarse con cautela. De hecho, muchos factores pueden influir en la eficiencia, como los diferentes entornos reguladores, la intensidad de la competencia, la especialización de la producción, la calidad de los inputs, etc. Por consiguiente, la estimación de la eficiencia de los beneficios alternativos, que tiene en cuenta el diferente grado de competencia, así como el efecto de la calidad de la producción sobre los ingresos, parece ser una forma más apropiada de hacer comparaciones internacionales. Además, estos autores proporcionan información sobre los niveles de eficiencia promedio estimada por métodos no



paramétricos en la industria bancaria norteamericana, que oscilaban entre 72% a 74%, es decir, una ineficiencia del 28% a 26%, respectivamente.

El estudio de Bauer et al. (1998) se encuentra entre los más significativos, dada la aplicación de cuatro enfoques: SFA, DEA, TFA (*Thick Frontier Analysis*) y DFA (*Distribution Free Analysis*) a un conjunto de datos formado por 683 bancos de Estados Unidos durante el período comprendido entre 1977 y 1988. Estos autores sugirieron seis condiciones de consistencia para analizar la robustez de las diferentes medidas de frontera eficiente analizadas, a saber, una comparación de las distribuciones de eficiencia, la correlación de rango entre las distribuciones de eficiencia, la correspondencia entre las mejores y peores prácticas de los bancos, la estabilidad de las medidas de eficiencia en el tiempo, la consistencia de la eficiencia con las condiciones de competencia del mercado y con las medidas de evaluación del rendimiento. Para cada enfoque se calcularon una medida de la eficiencia anual y una medida de la eficiencia total de varios años. La eficiencia media obtenida mediante técnicas paramétricas promedia el 83% mientras que la eficiencia media de los enfoques no paramétricos está solamente alrededor del 30%. Las técnicas no paramétricas y paramétricas poseen una consistencia muy débil en la clasificación entre sí: el orden de la correlación de rangos es del 10%. Todos los métodos son estables en el tiempo a pesar de que DEA, generalmente, muestra mejor estabilidad que los métodos paramétricos. Por otro lado, los índices de eficiencia son generalmente consistentes con las medidas de rendimiento estándar, aunque las puntuaciones de la eficiencia en el DEA son mucho menores. En resumen, Bauer et al. (1998) concluyen que no existe un único enfoque correcto para especificar una frontera eficiente. Casu y Girardone (2002) evaluaron los costes, la ganancia de la eficiencia y la variación de la productividad de los conglomerados financieros italianos durante la década de los 90 utilizando SFA,

DFA y DEA. Las medidas de eficiencia obtenidas por los métodos de fronteras estocásticas y deterministas son razonablemente similares en magnitud, y también muestran una variación similar en el nivel de eficiencia. A pesar de estas similitudes en rango y varianza de la puntuación de eficiencia, la tendencia de la eficiencia de costes en el DEA aumentó entre 1996 y 1998, y mostró una marcada disminución en el año 1999. A su vez, las estimaciones de SFA muestran una mejora constante de la eficiencia de costes.

Xiaogang et al. (2005) examinaron el coste, la eficiencia técnica y de asignación de 43 bancos chinos durante el período 1993-2000. En cuanto a la eficiencia bancaria posterior a la desregulación desde 1995, encuentran que los grandes bancos estatales y los pequeños son más eficientes que los bancos chinos medianos. Además, la eficiencia técnica domina sistemáticamente la eficiencia de asignación entre los bancos chinos. Se encuentra que la desregulación mejora los niveles de rentabilidad.

Delis y Papanikolaou (2009) midieron la eficiencia de costes y la eficiencia de los beneficios para 28 bancos comerciales griegos durante el período 1993-2005. Sus conclusiones fueron que: 1) las eficiencias de coste obtenidas por DEA son mucho más bajas que las de SFA, 2) ambos enfoques indicaron que existe una relación positiva entre la eficiencia de costes y el tamaño del banco y, 3) las conclusiones sobre el efecto de la propiedad en la eficiencia son contradictorias entre los dos enfoques.

Yung-Ho y Chen (2009) analizan la eficiencia bancaria de los bancos taiwaneses durante el período 2002-2004 y encuentran que la puntuación media de eficiencia de los bancos estatales es mayor que la de los bancos privados.

Barros et al. (2011) aplicaron un modelo DEA novedoso, el modelo B-convexo inverso, para investigar la eficiencia técnica de una muestra de bancos chinos de 1998 a

2008. En general, se muestra que la eficiencia general ha mejorado después de que China se uniera a la Organización Mundial del Comercio (OMC).

Kumbhakar et al. (2013) introdujeron un modelo que tiene en cuenta la posibilidad de distinguir los bancos totalmente eficientes de los ineficientes (es el modelo ZISF o de frontera estocástica inflado de ceros). El ejemplo empírico utilizado corresponde con información detallada de bancos comerciales estadounidenses (también usados por Kumbhakar y Tsionas (2005)), y usan un panel equilibrado de 4985 bancos durante el período 1989-2000.

Wang et al. (2014) usaron un modelo DEA de dos etapas para evaluar 16 bancos chinos durante el período 2003-2011 (la tercera ronda de reforma bancaria china). Los resultados sugieren que las reformas generalmente mejoran la eficiencia bancaria; los bancos de propiedad estatal son más eficientes que los bancos comerciales en el período pre-reforma, mientras que la diferencia se ha reducido durante el período posterior a la reforma; y la reforma del capital conjunto de los bancos estatales aumentó la eficiencia.

Dong et al. (2014) investigaron la consistencia de las puntuaciones de eficiencia obtenidas por el SFA y DEA. Estimaron la eficiencia de costes y las economías de escala sobre la base de un conjunto de datos no balanceados de bancos chinos durante el período de 1994 a 2007. Los resultados sugieren una moderada consistencia entre los métodos paramétricos y no paramétricos en las clasificaciones de eficiencia. La estabilidad de las puntuaciones de eficiencia en el tiempo y la correlación entre la eficiencia y las medidas de desempeño contable. En resumen, sus hallazgos, concluyen que se recomienda fuertemente el uso de técnicas de fronteras múltiples para el análisis de eficiencia y que este análisis metodológico de verificación cruzada resultará en evaluaciones más robustas y convincentes del desempeño de los bancos.

La conclusión es que, en general, los métodos DEA y SFA ofrecen resultados parecidos, salvo excepciones.

### **5.2.2. CAMBIOS DE PRODUCTIVIDAD**

Diversos estudios empíricos han analizado los cambios de productividad utilizando modelos de eficiencia no paramétricos y paramétricos.

Por un lado, en cuanto a las técnicas no paramétricas, se ha analizado la evolución de la productividad mediante el análisis envolvente de datos (DEA) y el índice de productividad total de factores de Malmquist (Malmquist-TFP). Este índice de productividad permite a los investigadores capturar con precisión los diferentes cambios de productividad reflejados en las operaciones de los bancos (ver Fethi y Pasiouras (2010) para una descripción general). Este método evalúa el cambio de eficiencia a lo largo del tiempo (cambio técnico en eficiencia) y el cambio de frontera (cambio tecnológico o innovación). Refleja el progreso o la regresión en la eficiencia junto con el progreso o la regresión de la tecnología de frontera a lo largo del tiempo en el marco de múltiples insumos y múltiples productos.

Por ejemplo, Casu et al. (2004) investigaron el cambio de productividad en la banca europea durante el período 1994-2000, Asmild et al. (2004) analizaron la industria bancaria en Canadá. Otros artículos han investigado el crecimiento de la productividad en períodos regulatorios. Por ejemplo, Berg et al. (1993) examinó la productividad del sector bancario noruego durante la desregulación de los años ochenta. Encontraron una regresión de la productividad en el banco promedio antes de la desregulación, pero un rápido crecimiento cuando tuvo lugar la desregulación. Fukuyama (1995) midió el crecimiento de la eficiencia y la productividad en la industria bancaria japonesa durante

los períodos 1989-1991. Noulas (1997) comparó las diferencias de eficiencia y productividad entre los bancos estatales y privados en Grecia durante el período 1991-1992. Encontró que, aunque la productividad ha aumentado tanto para los bancos estatales como para los privados, las fuentes de este crecimiento fueron diferentes. Por ejemplo, el crecimiento de la productividad de los bancos estatales proviene del progreso tecnológico, mientras que el de los bancos privados, de una mayor eficiencia. Leightner y Lovell (1998) analizaron la productividad de los bancos tailandeses para el período 1989-1994. Descubrieron que cuando se utilizaron los objetivos del Banco de Tailandia, se encontraron niveles similares de crecimiento rápido en relación con las mejores prácticas, pero la productividad total de los factores disminuyó para los bancos tailandeses y aumentó para los bancos extranjeros. Gilbert y Wilson (1998) estudiaron los efectos de la desregulación sobre la productividad de los bancos coreanos durante el período 1991-1994. Descubrieron que los bancos coreanos respondieron a la privatización y la desregulación alterando sustancialmente su combinación de insumos y productos, generando grandes cambios en la productividad. En general, la mayoría de los trabajos encontraron que el crecimiento se atribuye principalmente al progreso tecnológico.

Mukherjee et al. (2001) analizaron el crecimiento de la productividad de 201 bancos comerciales estadounidenses durante el período inicial posterior a la desregulación entre 1984 y 1990, utilizando DEA-Malmquist. Estos autores analizaron las diferencias en los cambios de productividad utilizando un método de regresión de datos de panel y concluyeron que la presencia de bases de activos más grandes y la especialización de combinaciones de productos se asociaron positivamente con la productividad, y que una mayor proporción de capital a activo indicó la probabilidad de un menor crecimiento de la productividad.

Canhoto y Dermine (2003) examinaron la eficiencia y la productividad bancarias en Portugal durante el período de desregulación 1990-1995. El índice de productividad de Malmquist para los nuevos bancos es mayor que el de los antiguos, un indicador de una mejora superior en la eficiencia a lo largo del tiempo. Por tanto, la eficiencia de los nuevos bancos es superior a la de los antiguos. Isik y Hassan (2003) analizaron el crecimiento de la productividad, el cambio de eficiencia y el progreso técnico en los bancos comerciales turcos durante el período de desregulación de los mercados financieros (1981-1990), y concluyeron que la productividad del sector bancario había aumentado significativamente en este período, impulsada por la eficiencia más que por el progreso técnico. Además, la mayor parte de este aumento de eficiencia se debió a mejores prácticas de gestión, no a la eficiencia de la escala. Finalmente, los autores observaron que en el nuevo entorno los bancos privados se desempeñaron mejor que los bancos públicos. Rezitis (2006) encontró un mayor crecimiento de la productividad después de la desregulación, que está de acuerdo con los resultados de la industria bancaria obtenidos en otros estudios.

Otros autores han analizado la relación entre el marco regulatorio y de supervisión y la productividad. Delis et al. (2011) examinaron bancos en 22 países, para el período 1999-2006, y mostraron que las regulaciones e incentivos que promovían el monitoreo privado y las restricciones a las actividades de los bancos (valores, seguros, bienes raíces y propiedad de firmas no financieras) tenían un efecto positivo en la eficiencia. Por otro lado, los requisitos de capital y el poder de supervisión oficial no tuvieron un impacto significativo en la productividad. Sánchez et al. (2013) concluyeron que la liberalización financiera había conducido a aumentos de la productividad como consecuencia del progreso tecnológico en lugar de un mayor progreso técnico en América Latina durante el período 1996-2007.

Recientemente, Kumar et al. (2016) evaluaron el desempeño del sector bancario indio para el período posterior a la reforma y la crisis financiera global entre 1995 y 2009. Estos autores concluyeron que las reformas habían producido un cambio positivo en la PTF del sector bancario, y lo habían hecho a través de cambios tecnológicos.

Finalmente, podemos destacar los trabajos que explican las diferencias en las ganancias de los bancos en función de las diferencias en la productividad. Estos artículos tienden a utilizar medidas de productividad estimadas a partir de modelos de frontera estocástica (SFA) como Berger y Mester (2003), Almanidis (2013) y Caiazza et al. (2016), o con la aplicación de DEA como Grifell-Tatje y Lovell (1999) y Fiordelisi y Molyneux (2010). Trabajos recientes han considerado la productividad como un parámetro de la función de producción de bancos como Castro y Galán (2019), Llorens et al. (2020) para España, entre otros. Por ejemplo, Llorens et al. (2020) han demostrado que la evolución de la productividad bancaria a lo largo del tiempo podría determinar los costes de intermediación y el proceso de reestructuración de la industria bancaria en la Gran Recesión. Han descubierto que los bancos menos productivos tienen más probabilidades de salir que los bancos más productivos, y que los bancos sobrevivientes adquieren banco objetivo para expandir su red de sucursales en los mercados locales donde están subrepresentados.

### **5.3. DETERMINANTES DE LA EFICIENCIA**

El error idiosincrático de un modelo SFA recoge los elementos no constitutivos de la ineficiencia como, por ejemplo, aquellos que están al margen del control de la empresa como restricciones en el suministro de factores, elementos meteorológicos, pero también los que corresponden al ruido estadístico como los errores de medida o los

errores de especificación del modelo. Sin embargo, no hay una teoría establecida sobre los factores que explican la ineficiencia.

El estudio de los determinantes de la eficiencia se ha desarrollado, generalmente, de forma *ad hoc*, analizando el modelo básico de frontera estocástica y usando un procedimiento de estimación en dos etapas para evaluar el efecto de diversos factores sobre la eficiencia. Algunos ejemplos son Allen and Rai (1996), Rai (1996), Berger y DeYoung (1997), Berger y Mester (1997), Resti (1997), DeYoung y Hasan, (1998), Bos y Kolari (2005), Lieu et al. (2005), Bonin et al. (2005), Yildirim y Philippatos (2007), Berger et al. (2009), Lin y Zhang (2009), entre otros, quienes estudian los efectos de variables macroeconómicas y los efectos específicos de los bancos tales como el tamaño, la propiedad, y las sucursales bancarias sobre la eficiencia de costes. En la primera etapa, se estima la frontera de costes y, en la segunda etapa, se utiliza un modelo de regresión relacionando la eficiencia estimada con las variables macroeconómicas y las variables específicas de los bancos.

Por ejemplo, las fusiones se han vinculado a la eficiencia de los bancos. De los abundantes estudios sobre fusiones y nivel de desempeño en la banca estadounidense en varios momentos, parece que las fusiones, en promedio, no tienen un efecto significativo y predecible sobre la eficiencia de los costes (Berger y Humphrey, 1994). Algunos estudios indican que la eficiencia de costes mejoró en la mayoría de las fusiones, aunque las ganancias fueron pequeñas (Berger y Humphrey, 1992; Boyd y Graham, 1998; DeYoung, 1997; Fried et al., 1999), mientras que otros consideran que las fusiones aumentan los costes (Berger y Humphrey, 1994) y reducen la eficiencia  $X$  (Peristiani, 1997).

Otros estudios también encontraron otros factores determinantes de la eficiencia entre los bancos como son la propiedad y la localización (Cebenoyan et al., 1993), la



antigüedad del banco, ya sean bancos nacionales o estatales, sociedad de cartera y número de sucursales (Mester, 1997), control de gestión y/o estructura organizacional (Berger y DeYoung, 2000; Valverde et al., 2007), y el desempeño económico (Christopoulos et al., 2002; Ferrier, 2001).

Sin embargo, también podemos encontrar artículos que, analizando el modelo básico de frontera estocástica, usaron un procedimiento de estimación en una etapa para evaluar el efecto de diversos factores sobre la eficiencia. Este es el caso de Battese y Coelli (1995) quienes diseñaron un modelo de panel de datos que permite incorporar factores explicativos en la media de la distribución de la ineficiencia. Recientemente, Mohanty et al. (2013) desarrollaron un modelo que tiene en cuenta la heteroscedasticidad en el componente de ineficiencia y adoptan una aproximación en una etapa siguiendo a Battese y Coelli (1995). Estos autores analizaron la eficiencia de costes de los bancos comerciales en Taiwan usando datos anuales para el periodo 1996–2011, y siguiendo a Caudill et al. (1995) y Caudill y Ford (1993).

### **5.3.1. DETERMINANTES DE LA VARIANZA**

Algunos estudios sobre eficiencia bancaria integran los riesgos asociados a las operaciones de los bancos en sus análisis con el argumento de que los estudios de eficiencia por sí solos ignoran el impacto de las características de riesgo en el desempeño general de los bancos. Dado que el riesgo no se explica explícitamente en el modelo estándar de producción y costes, la utilidad del modelo estándar puede verse socavada cuando se aplica al sector bancario donde el riesgo tiene un papel económico significativo (Hughes et al., 1995). Como consecuencia, los bancos que operan de una manera más aversa al riesgo que otros podrían ser mal etiquetados como ineficientes,

mientras que los bancos que producen una producción de menor calidad que otros podrían ser etiquetados erróneamente como eficientes (Mester, 1996). Debido a este posible defecto, varios autores amplían el modelo para acomodar el riesgo bancario y la ineficiencia (Altunbas et al., 2000; Altunbas et al., 2004; Chang y Chiu, 2006; Clark, 1996; Girardone et al., 2004; Laeven, 1999; McAllister y McManus, 1993; Pastor y Serrano, 2005).

Manglañit (2011) examinó la eficiencia de los bancos comerciales filipinos utilizando análisis de fronteras estocásticas e incorporando específicamente medidas de riesgo y de calidad de activos en la estimación. Sus resultados mostraron ineficiencias importantes en los bancos nacionales y que el riesgo y la calidad de los activos afectaron la eficiencia de los bancos. El aumento sustancial de la ineficiencia de costes podría atribuirse a los efectos adversos de la crisis financiera asiática de 1997 y los costes subsiguientes de las reformas bancarias y los cambios regulatorios que se promulgaron para estabilizar y fortalecer el sector. Las inestabilidades macroeconómicas, en particular la crisis bancaria, podrían haber distorsionado la estructura de incentivos para los bancos, haciendo más difícil la asignación de recursos para lograr la eficiencia. Desde el punto de vista de las políticas, su estudio destacó la importancia económica de fomentar una mayor eficiencia en el sector bancario aprovechando el potencial de mejoras significativas en la eficiencia de los costes de los bancos.

## 5.4. ESTUDIOS SOBRE LA BANCA ESPAÑOLA Y COMPARACIONES

### INTERNACIONALES

#### 5.4.1. BANCA ESPAÑOLA

La eficiencia bancaria en España ha sido analizada en múltiples trabajos, incluyendo el análisis de bancos comerciales (privados), cajas de ahorro (propiedad del Estado) y cooperativas de ahorro y crédito. Muchos investigadores han empleado procedimientos no paramétricos como DEA o FDH (una variante no convexa de DEA), pero también procedimientos paramétricos como el análisis de fronteras estocásticas (SFA).<sup>20</sup>

Algunos trabajos han utilizado el marco de la DEA para estimar la eficiencia de los bancos. Por ejemplo, Pastor (1995), Grifell y Lovell, (1997), Maudos y Pastor (2003), Guzmán y Reverte (2008), Färe et al. (1997), Orea (2000), Maudos y otros (2002), Cuesta y Orea (2002) y San-José et al. (2014). A este respecto, Prior et al. (2016) realizó un análisis de eficiencia para la banca española, adoptando el enfoque de la FDH para abordar tanto los años previos a la crisis como los de crisis de 2007. Estos autores examinaron bancos comerciales, cajas de ahorro y cooperativas de ahorro y crédito (la distinción es relevante porque la crisis tuvo efectos diferentes en cada una de estas formas de organización) y calcularon tanto la eficiencia de costes como la eficiencia de los ingresos.<sup>21</sup> IMPLICA

---

<sup>20</sup> El enfoque de frontera no paramétrico que utiliza DEA o FDH requiere supuestos mínimos con respecto a la estructura de la producción y no impone restricciones a la forma funcional que relaciona los insumos y los productos. Debido a que no toma en cuenta el ruido en los datos, supone implícitamente que cualquier desviación de la frontera puede ser considerada como ineficiencia. Sin embargo, en la SFA paramétrica, se hacen suposiciones tanto sobre la forma funcional como sobre la distribución de los dos tipos de error: un término de error idiosincrásico y el término de error de ineficiencia (es decir, la desviación de la frontera).

Otros autores han estudiado los cambios de productividad a través del modelo DEA-Malmquist, como Guzmán y Reverte (2008). Estos autores analizaron si aquellos bancos españoles caracterizados por mayores niveles de eficiencia y cambio de productividad tienen un mayor valor para el accionista. Estos autores informaron ligeros aumentos de la PTF, alrededor del 2%). Finalmente, también cabe mencionar a Vallespín et al. (2016), quienes analizaron la evolución de la eficiencia productiva de las entidades bancarias españolas en el periodo 2009-13. Sus resultados indicaron que las diferencias en la eficiencia de las entidades no estaban relacionadas con el tamaño y naturaleza de las mismas. Estos autores encontraron que la mayor parte de las entidades mejoraban su rendimiento a lo largo del período, lo que sugería que las reestructuraciones llevadas a cabo resultaron positivas en la recuperación del sistema.

En un enfoque alternativo, la SFA se ha utilizado en algunos estudios, como Maudos (1996), Maudos (2001), Maudos y Pastor (2001a, 2001b) y Orea y Kumbhakar (2004).

La conclusión es que los bancos españoles son eficientes en la mayoría de los estudios, superando esta eficiencia el 80%.

#### **5.4.2. COMPARACIÓN INTERNACIONAL**

Existe un número importante de comparaciones internacionales de la eficiencia de los bancos.

---

<sup>21</sup> Sus resultados indicaron que los bancos comerciales eran más eficientes que las cajas de ahorro y las cooperativas de ahorro y crédito. Sin embargo, durante los años de crisis, las diferencias entre estas instituciones se redujeron drásticamente, especialmente en lo que respecta a la eficiencia. Otro estudio reciente sobre los bancos griegos, considerando los efectos producidos en las sucursales bancarias antes y durante el período de crisis, fue realizado por Aggelopoulos et al. (2013).

Usando la clasificación de Brack y Jimborean (2009) se pueden distinguir dos categorías de comparaciones internacionales: las que estiman una frontera nacional para cada país, y las que estiman fronteras comunes a varios países en su conjunto.

A continuación, describiremos someramente algunas de ellas siguiendo a Brack y Jimborean (2009), añadiendo alguna referencia más reciente.

#### **5.4.2.1. ESTIMACIÓN DE UNA FRONTERA NACIONAL PARA CADA PAÍS**

Una primera referencia es el análisis de Berger et al. (2000). Los autores usan el enfoque SFA y estiman los costes y las fronteras de producción para cinco países (Francia, Alemania, España, Reino Unido y Estados Unidos), por separado y para cada país. La eficiencia de los bancos nacionales y extranjeros se estima en el período 1993-1998 para los Estados Unidos y también para las economías europeas entre 1992-1997. Los resultados muestran una eficiencia media de 70.9% en Francia, 79.3% en Alemania, 91.5% en España, 79.1% en el Reino Unido y 77.4% en Estados Unidos. El resultado principal es que los bancos nacionales presentan un mayor coste y mayores puntuaciones de eficiencia en la producción, en comparación con los bancos extranjeros que operan en estos países.

Lozano-Vivas et al. (2002) investigaron las diferencias de eficiencia operativa de una muestra de bancos comerciales en 10 países europeos. Primero, analizaron la eficiencia técnica de cada muestra en cada país siguiendo el modelo “básico” de Análisis Envolvente de Datos (DEA) que incorporó solo variables bancarias. Luego, se introdujeron un modelo DEA “completo”, incorporando factores ambientales junto con las variables bancarias del modelo básico. La comparación entre los dos modelos

muestra que las condiciones ambientales específicas de cada país ejercieron una fuerte influencia sobre el comportamiento de la industria bancaria de cada país.

Weill (2004) midió la eficiencia de los bancos de cinco países europeos (Francia, Alemania, Italia, España y Suiza) durante el período 1992-1998 y utilizó tres enfoques: SFA, DFA y DEA. El análisis se basó en la eficiencia de las fronteras nacionales (y no en una frontera común), mostrando la coherencia de las fronteras de eficiencia técnica en cinco marcos diferentes. Weill (2004) mostró que si bien las puntuaciones de eficiencia promedio entre los métodos SFA, DFA y DEA son ampliamente comparables, las correlaciones de rangos de eficiencia entre los enfoques paramétricos y no paramétricos son generalmente pobres. Además, también investigó dos importantes cuestiones de política; a saber, la relación entre la eficiencia de costes y el tamaño de los bancos, y el vínculo entre la eficiencia de costes y la especialización bancaria (por ejemplo, bancos cooperativos, cajas de ahorros, etc.). Aquí los resultados mixtos fueron reportados con cierto grado de consistencia entre la eficiencia de costes y los tres enfoques de frontera (DEA, SFA y DFA). En contraste, el vínculo entre la eficiencia de costes y la especialización depende de la metodología utilizada.

Al utilizar enfoques paramétricos y no paramétricos para examinar los cambios de productividad en los bancos europeos durante los años noventa, Casu et al. (2004) encontraron que los dos enfoques no producen resultados muy significativos en la identificación de los principales componentes del crecimiento de la productividad. Beccalli et al. (2006) midieron la eficiencia de costes de los bancos europeos entre 1999 y 2000. Estos autores investigaron la relación entre las medidas de eficiencia y el funcionamiento del mercado de las instituciones financieras por medio de SFA y DEA, y encontraron que las variaciones porcentuales de los precios de las acciones reflejaron cambios porcentuales en la eficiencia de costes, en particular las derivadas por el DEA.

Es más, las puntuaciones de eficiencia obtenidas por SFA son ligeramente superiores a las puntuaciones DEA, es decir, 85% frente a 83%. Además, las puntuaciones de eficiencia DEA fueron más dispersas en comparación con SFA.

Weill (2008) calculó la eficiencia de los bancos de diez países de la UE (Austria, Bélgica, Dinamarca, Francia, Alemania, Italia, Luxemburgo, Portugal, España y Reino Unido) durante el período 1994-2005 y utilizó SFA. Sus resultados mostraron una mejora en la eficiencia de todos los sistemas bancarios de la UE y la existencia de convergencia en la eficiencia de todos los países de la UE.

Molnar y Holló (2011) evaluaron la eficiencia de los bancos en Hungría en comparación con los bancos de la UE durante el período 2004-2008. Compararon su eficiencia utilizando el enfoque de frontera estocástica. Gallizo et al. (2017) analizaron la eficiencia de los bancos en los países bálticos (Estonia, Letonia, Lituania) durante el período 2000-2013 utilizando los modelos de frontera estocástica bayesiana.

#### **5.4.2.2. ESTIMACIÓN DE UNA FRONTERA COMÚN PARA VARIOS PAÍSES**

La segunda categoría de estudios a la que hacen referencia Brack y Jimborean (2009) estima una frontera común que permite comparar la eficiencia bancaria en diferentes países. Todos los bancos son comparados con los mejores bancos de todos los países analizados. El principal supuesto de estos estudios es que todos los bancos tienen la misma tecnología, lo que puede conducir a lagunas de eficiencia que resultan de la existencia real de diferentes tecnologías entre bancos de diferentes países. En decir, las brechas de eficiencia entre los países pueden ser causadas por diferentes entornos

económicos y no necesariamente por diferencias en el desempeño gerencial (Dietsch y Lozano-Vivas, 2000).

Allen y Rai (1996) estiman la función global de costes de 194 bancos internacionales (de 15 países), durante el período 1988-1992, para determinar la ineficiencia de los inputs y outputs. Según su análisis, las ineficiencias de los inputs son mayores que las de los outputs. Otro resultado es que el enfoque de DFA sobrestima el tamaño de las puntuaciones de ineficiencia, en comparación con el enfoque de SFA. Los grandes bancos tienen el mayor valor de ineficiencia de los inputs (27.5% del coste) y niveles significativos de diseconomías de escala. Para los otros bancos, la ineficiencia es del 15% del coste, con economías de escala reducidas para los bancos pequeños.

Pastor et al. (1997) compararon la eficiencia de varios bancos europeos (España, Austria, Alemania, Reino Unido, Italia, Bélgica y Francia) con la de los bancos americanos en 1992. Bajo la hipótesis de rendimientos constantes, los bancos franceses son los más eficientes (con una eficiencia media del 95%), seguidos por los bancos españoles, belgas, italianos, alemanes, americanos, austriacos e ingleses. Por otro lado, se subraya la menor productividad de los bancos franceses (se encuentran en la segunda posición, frente a los bancos españoles).

Chaffai y Dietsch (1999) proponen la descomposición de la ineficiencia de costes en la ineficiencia técnica y de asignación, basada en la metodología de las distancias en los inputs. Los autores utilizan el enfoque SFA. La muestra analizada contiene bancos europeos de 11 países (Austria, Bélgica, Alemania, Dinamarca, España, Francia, Reino Unido, Italia, Luxemburgo, Países Bajos y Portugal). El estudio mostró que, en promedio, la ineficiencia de asignación incrementa los costes bancarios en un 25%, lo mismo que la ineficiencia técnica. Otro resultado es la existencia de una correlación negativa entre las ineficiencias técnicas y de asignación.



Dietsch y Lozano-Vivas (2000) analizaron, por el enfoque DFA, el efecto de las condiciones del entorno sobre la eficiencia de las industrias bancarias francesa y española durante el período 1988-1992. Los resultados son los siguientes. Sin tener en cuenta las variables ambientales, la eficiencia de los bancos españoles es más reducida que la de los bancos franceses; La introducción de las variables ambientales en el modelo reduce las diferencias entre las dos industrias bancarias.

Altunbas et al. (2001) realizaron un análisis de una gran muestra de bancos europeos (de 15 países) durante el período 1989-1997, usando el enfoque SFA. Los resultados muestran que en promedio los bancos de UK y suecos son más ineficaces que otros bancos europeos. Los sistemas bancarios más eficaces son los de Austria, Dinamarca, Alemania e Italia. Otro resultado importante es el aumento del impacto del progreso técnico en la reducción de los costes bancarios, con el tamaño de los bancos. Durante todo el período de análisis, se observa un aumento en la eficiencia de los bancos.

Chaffai et al. (2001) proponen un índice de Malmquist que permite medir las diferencias de productividad entre bancos de distintos países y distinguir dos componentes: las diferencias causadas por efectos puramente tecnológicos y las diferencias causadas por los efectos del entorno. Este índice se utiliza para explicar las diferencias en la productividad entre los bancos de cuatro países de la zona del euro (Francia, Alemania, Italia y España), durante el período 1993-1997. Los resultados muestran que, en promedio, las diferencias causadas por las condiciones del entorno son mayores en comparación con las diferencias en la tecnología de los bancos.

Hasta aquí, y en general, podemos decir que los estudios no muestran diferencias sustanciales entre países. Sin embargo, los siguientes estudios concluyen sobre la

existencia de diferencias importantes en la eficiencia de los bancos entre los Estados miembros de la UE.

Por ejemplo, Lozano-Vivas et al. (2001) estimaron la frontera de producción de una muestra de 612 bancos de diez países de la UE (Bélgica, Dinamarca, Francia, Alemania, Italia, Luxemburgo, Países Bajos, España y Reino Unido). Los autores usaron el enfoque DEA con variables bancarias. Luego, construyeron un modelo DEA incluyendo algunos factores del entorno (PIB per capita, salario per capita, densidad de población, densidad de demanda, razón de capital y rentabilidad) y variables bancarias, para normalizarlo teniendo en cuenta las condiciones específicas de cada país. Los resultados mostraron que las condiciones adversas (favorables) son un factor positivo (negativo) para la industria bancaria del país de origen. El hecho de ser técnicamente eficiente es un elemento disuasivo de la competencia extranjera. A nivel mundial, los bancos de España, Portugal y Dinamarca son relativamente más eficientes y poseen altas puntuaciones de eficiencia si deciden trasladarse e instalarse en otro país europeo de la muestra. Al mismo tiempo, sería más difícil que bancos de otros países establecieran redes rentables en España, Portugal y Dinamarca debido a condiciones adversas. Además, los bancos italianos y franceses son los menos eficientes en el extranjero.

Van der Venet (2002) analizó los costes y las capacidades de producción en los conglomerados financieros y los bancos europeos. El análisis usó una muestra de 2375 bancos de 17 países de la UE (Austria, Bélgica, Dinamarca, Finlandia, Francia, Alemania, Grecia, Irlanda, Italia, Luxemburgo, Países Bajos, Noruega, Portugal, España, Suecia, Suiza y Reino Unido), para 1995 y 1996 y utilizó el enfoque SFA. Los resultados mostraron que los conglomerados financieros son más eficientes en términos de ingresos que sus competidores especializados, mientras que los bancos universales son los más eficientes en términos de coste y producción.

Brack y Jimborean (2009) abordaron la cuestión de la credibilidad de los bancos franceses en comparación con sus homólogos de Europa y Estados Unidos. El análisis se realizó sobre una muestra formada por los diez mayores bancos de Francia, Alemania, Italia, España, Reino Unido y Estados Unidos, durante el período 1994-2006, y se utilizó el enfoque DEA. Los resultados mostraron una mejora en la rentabilidad de los bancos franceses y españoles, mientras que, en los otros países, se observó una disminución en la eficiencia de costes. Sus resultados mostraron que los bancos ineficientes han reducido la brecha durante el periodo 1994-2006, y que los factores explicativos de la eficiencia son variables específicas del banco, del entorno macro, del régimen regulador y del desarrollo del sector no bancario. Utilizando un modelo Tobit mostraron que los bancos recién creados, con proporciones más estrictas de capital de nivel 1 y que operan en un país con un PIB per cápita más bajo, registran las puntuaciones más altas de eficiencia de costes.

Sun y Chang (2011) examinaron el papel del riesgo en la determinación de la rentabilidad de bancos internacionales en ocho mercados asiáticos emergentes. Este estudio considera ocho medidas de riesgo que cubren el riesgo de crédito, el riesgo operativo y el riesgo de mercado. Los resultados indican que las medidas de riesgo representan un efecto significativo tanto en el nivel como en la variabilidad de eficiencia bancaria, y estos efectos dependen del tiempo y la región.

Lin et al. (2016) investigaron el efecto de los cambios en la propiedad de los bancos sobre la eficiencia de costes en 12 economías en desarrollo asiáticas. Durante el período 2003-2012, la presencia extranjera mejora la eficiencia bancaria, principalmente en países con mayor libertad financiera. Además, el aumento de la propiedad del banco por parte del gobierno (nacional) parece mejorar (impedir) la eficiencia bancaria en países con más libertad financiera después de la crisis financiera global (CFG) de 2008.

Meslier et al. (2014) investigaron el impacto de la diversificación de los ingresos bancarios en el desempeño de los bancos en Filipinas. Los resultados muestran que, a diferencia de las economías occidentales, un cambio hacia actividades sin intereses aumenta las ganancias bancarias y las ganancias ajustadas al riesgo cuando los bancos están más involucrados en la negociación de valores gubernamentales y tienen menos exposición a las pymes. Los bancos extranjeros se benefician más de tal cambio.

Gulati y Kumar (2016) midieron el impacto de CFG en la eficiencia bancaria en India, centrándose en la evolución de la eficiencia de los beneficios antes, durante y después de la crisis. Este estudio emplea un marco de frontera de meta-beneficio basado en DEA que tiene en cuenta la heterogeneidad tecnológica entre grupos. Los resultados muestran que la eficiencia de los beneficios se redujo levemente durante la crisis y posteriormente se recuperó. La propiedad de los bancos genera un impacto diferente por parte de CFG, y los nuevos bancos privados experimentaron el mayor lastre en las ganancias durante el año de crisis. Los bancos extranjeros parecen ser los líderes tecnológicos. En general, en la India no existió ningún efecto adverso duradero de CFG sobre la eficiencia de los beneficios debido a la adopción de políticas macroeconómicas acomodaticias destinadas a inyectar suficiente liquidez.

Nguyen (2018) examinó el impacto de la diversificación en la eficiencia de costes y ganancias de los bancos comerciales de seis países de la ASEAN entre 2007 y 2014. Entre los múltiples resultados obtenidos, estos autores encuentran que: los bancos diversificados por ingresos tienen una menor eficiencia de costes, mientras que los bancos diversificados por activos solo tienen una menor eficiencia persistente de costes. Por otro lado, los bancos con una mayor diversificación de fondos tienen una mayor eficiencia de ganancias. Los bancos con mayor financiación diversificada y propiedad mayoritaria del gobierno demuestran una mayor eficiencia de costes, pero menor

eficiencia de ganancias que otros bancos, mientras que los bancos extranjeros tanto con diversificación de financiación como activos son menos eficientes en beneficios.

## **5.5. EFICIENCIA DE LAS SUCURSALES BANCARIAS**

La comprensión de la eficiencia de sucursales bancarias puede ayudar a resolver una serie de cuestiones de política relacionadas con la eficiencia a nivel del Banco.

En este campo de estudio, los análisis a nivel de sucursal bancaria son más útiles que los de nivel institucional. Los datos de una sucursal se recogen de manera uniforme y comparable en un banco único y, por lo tanto, no crean serios problemas para la interpretación de la eficiencia de escala. Por otra parte, la homogeneidad de las sucursales es un requisito básico de la modelización DEA para minimizar el sesgo de estimación (Coelli et al., 2005; Chortareas et al., 2009).

Berger et al. (1997) indicaron que la información sobre el desempeño de las sucursales puede ayudar a mejorar nuestra comprensión de la eficiencia del banco y contribuir a resolver los problemas de medición en este contexto. En este sentido, Paradi y Zhu (2013) indicaron que la red de sucursales suele ser la mayor fuente de gastos operativos para un banco y que los clientes realizan una gran parte de su banca de mayor valor añadido a través de sucursales.

A continuación, en los siguientes epígrafes, describiré brevemente algunos de los estudios más relevantes usando la metodología DEA o la metodología de fronteras estocásticas, fundamentalmente, la de costes.

## **5.6. ESTUDIOS SOBRE SUCURSALES BANCARIAS USANDO MÉTODOS NO**

### **PARAMÉTRICOS**

Existe una extensa literatura sobre la eficiencia de las instituciones bancarias en su conjunto (ver Fethi y Pasiouras (2010), y sus referencias), y también sobre la de las sucursales bancarias. En este último aspecto, en particular, destacan dos surveys: en el primero, Fethi y Pasiouras (2010) discutieron la eficiencia de bancos y sucursales, analizando 30 estudios, realizados entre 1998 y 2009, que utilizaron técnicas DEA para estimar la eficiencia de las sucursales. Posteriormente, Paradi y Zhu (2013) identificaron 80 estudios en los que se aplicó DEA a los bancos y sus sucursales, durante el período 1985-2011.

Desde que Sherman y Gold (1985) escribieron el primer artículo usando DEA sobre sucursales bancarias, varios artículos han propuesto mejoras metodológicas, como incluir los juicios de valor de los gerentes de sucursales (Athanasopoulos, 1997), introducir la formulación de lógica difusa (Wu et al., 2006), utilizando DEA multicapa (Alirezaee y Afsharian, 2007) o un modelo DEA “culturalmente ajustado” (Paradi et al., 2010), evaluando el desempeño de las sucursales para múltiples dimensiones (Paradi et al., 2011), y evaluando la gestión eficiencia y proporcionar pautas de mejora (Camanho y Dyson, 1999: entre otros).

Muchos de los artículos que analizan las oficinas asumen rendimientos constantes a escala, y las medidas de eficiencia radial siguen siendo los enfoques más utilizados. Sin embargo, la eficiencia en las operaciones generalmente se evalúa en diferentes áreas o dimensiones, porque ningún modelo único puede capturar completamente la naturaleza multifuncional de las sucursales bancarias. Los métodos que se han adoptado incluyen el enfoque de producción (centrado en la actividad operativa), el enfoque de

intermediación (con respecto a la eficiencia con la que la sucursal recauda depósitos y otros fondos) y el enfoque de rentabilidad (qué tan bien la sucursal genera ganancias). La calidad (o eficiencia) del servicio es otra dimensión importante del desempeño de las sucursales bancarias. Para cada una de estas dimensiones, existe una diversidad significativa entre los estudios anteriores en términos de la selección de entrada / salida realizada (ver Tabla A1 en Paradi y Zhu, 2013). Los métodos DEA se han utilizado para estimar las eficiencias técnicas y de escala, las eficiencias de asignación y de costes y, la evolución de la productividad de las sucursales bancarias.

La eficiencia técnica se ha analizado utilizando una o varias dimensiones. Paradi y Zhu (2013) también indicaron que los aspectos más estudiados en el análisis de la eficiencia de las oficinas bancarias fueron tres enfoques:<sup>22</sup> el denominado enfoque de producción, que enfatiza la actividad operativa de la sucursal (Sherman y Gold, 1985; Parkan, 1987; Vassiloglou y Giokas, 1990; Giokas, 1991; Drake, 1994; Schaffnit et al, 1997; Drake and Howcroft (1994, 2002); Camanho y Dyson, 2005a; Cook y Zhu, 2006; Portela y Thanassoulis, 2007; Giokas, 2008a; Paradi et al., 2011); el enfoque de rentabilidad (*profitability*), que evalúa la habilidad de la oficina para convertir los gastos en ingresos (McEachern y Paradi, 2007; Manandhar y Tang, 2002; Al-Tamimi y Lootah, 2007; Paradi et al., 2010; Paradi et al., 2011), y el enfoque de intermediación, que enfatiza en cómo las instituciones financieras son vistas como intermediarios entre ahorradores e inversores (Camanho y Dyson, 2005a; Paradi et al., 2011; Giokas, 2008a; Camanho y Dyson, 2005a,b; Tsolas y Giokas, 2012).

Con respecto a una sola dimensión, Parkan (1987) examinó la eficiencia de producción de 35 sucursales de un importante banco comercial en Canadá e informó

---

<sup>22</sup> Estos tres enfoques son descritos con más detalle en el Capítulo 6, epígrafe 4.

que solo 11 eran relativamente eficientes. Camanho y Dyson (2005b) analizaron 144 sucursales de un banco comercial portugués y observaron considerables ineficiencias en esta red. Pastor et al. (2006) utilizaron DEA para examinar la eficiencia de los beneficios de 573 sucursales de una gran caja de ahorros europea e informaron que las mejoras de eficiencia en las sucursales con peores resultados (en las que se podrían obtener ahorros de hasta un 45% en el uso de recursos) generar un aumento sustancial de las ganancias para el banco. Con respecto a varias dimensiones, Oral y Yolalan (1990) evaluaron las eficiencias de utilidad y servicio de 20 sucursales minoristas de un banco comercial turco, encontrando una relación positiva entre estos parámetros, pero que solo el 20% de las sucursales eran eficientes. Portela y Thanassoulis (2007) utilizaron un análisis post-hoc para comparar la calidad del servicio y diferentes medidas de eficiencia, e informaron que la calidad del servicio se relacionó positivamente con la eficiencia operativa y de los beneficios. Giokas (2008a) evaluó las eficiencias de producción, transacción y rentabilidad obtenidas en operaciones dentro de una red de 44 sucursales durante el ejercicio 2002. De acuerdo con los resultados empíricos obtenidos, había margen para mejoras sustanciales en la eficiencia (la ineficiencia media fue del 12%), y por lo tanto para mayores ganancias bancarias. Paradi y col. (2011) analizaron las tres dimensiones del desempeño de las sucursales relacionadas con la eficiencia técnica para las 816 sucursales de un importante banco canadiense y concluyeron que estos enfoques mejoraron el método de evaluación del desempeño y permitieron a los gerentes de sucursales identificar claramente las fortalezas y debilidades de sus operaciones. Finalmente, Aggelopoulos et al. (2013) utilizaron dimensiones de producción y beneficio para estudiar los bancos griegos, considerando los efectos producidos en las sucursales bancarias antes y durante el período de crisis.



También se han utilizado métodos DEA para evaluar la rentabilidad de las sucursales bancarias. Por ejemplo, Giokas (2008b), Camanho y Dyson (2005a, b) y Tsolas y Giokas (2012), entre otros, han evaluado la capacidad de una sucursal para producir productos a costes mínimos, mientras que Camanho y Dyson (2005a) estimaron los límites superior e inferior de las medidas de eficiencia de costes de las sucursales bajo diversos escenarios de incertidumbre de precios.

Finalmente, se han empleado métodos DEA para evaluar el crecimiento de la productividad, según el índice de Malmquist. Así, Camanho y Dyson (2006) evaluaron el desempeño de 144 sucursales portuguesas para determinar el margen de eficiencia dentro del grupo, las eficiencias gerenciales internas y la productividad de frontera, reflejando el impacto de los factores ambientales y de las políticas gerenciales regionales. También a este respecto, Asmild y Tam (2007) calcularon las diferencias de frontera entre las sucursales bancarias de diferentes países.

Otro método no paramétrico, FDH, ha sido utilizado para investigar la eficiencia de las sucursales bancarias, por Respaut (1989) y Tulkens (1993), mientras que uno de los pocos estudios para considerar medidas de productividad basadas en FDH fue realizado por Tulkens y Malnero (1996).

El trabajo pionero en la aplicación del método DEA a las oficinas bancarias fue Sherman y Gold (1985).

Entre estos trabajos, destacaron Athanassopoulos (1997, 1998), Camanho y Dyson (1999), Dekker y Post (2001), Wu et al. (2006), Alirezade y Afsharian (2007), Azizi y Ajirlu (2010), Paradi et al., (2010) y Paradi et al. (2011), entre otros.

Otros estudios han analizado los efectos producidos en la eficiencia de las sucursales por las diferencias en las características regionales (Cook et al., 2004) y el efecto de las

actividades de comercio electrónico sobre el desempeño bancario (Das et al., 2009; Deville, 2009).

Por último, Paradi y Zhu (2013) mostraron la existencia de diferencias importantes entre los distintos estudios para selección de inputs / outputs. También encontraron que el supuesto de rendimientos constantes a escala y las medidas de eficiencia radial siguen siendo los enfoques más ampliamente utilizados.

En España, se han realizado pocos estudios sobre la eficiencia de las redes de oficinas bancarias, ya sea empleando DEA u otros métodos no paramétricos. De hecho, hasta donde sabemos, sólo conocemos el artículo de Lovell y Pastor (1997), citado en la introducción. Estos autores utilizaron un marco DEA y rendimientos variables a escala, con una única entrada constante o, de manera equivalente, sin entradas. El problema abordado en este caso fue maximizar la expansión radial sobre el vector de 17 indicadores de éxito para las 545 sucursales bancarias evaluadas.

En resumen, la evidencia empírica actual sobre las diversas dimensiones de la eficiencia de las sucursales indica que el DEA no paramétrico es el enfoque más utilizado y que se han realizado pocas investigaciones para analizar la eficiencia de las sucursales en el sistema bancario español. Por lo tanto, el estudio de la eficiencia es adecuado para proporcionar una ilustración empírica de los efectos de los factores ambientales en las puntuaciones de eficiencia de las oficinas, utilizando métodos no paramétricos y no paramétricos robustos. Precisamente, esta será la aportación del Capítulo 7.

## **5.7. CAMBIO DE PRODUCTIVIDAD DE LAS SUCURSALES BANCARIAS**

Aunque existe un extenso cuerpo de literatura internacional teórica y empírica sobre las sucursales bancarias (ver Fethi y Pasiouras, 2010; Paradi y Zhu, 2013, para una descripción general), solo unos pocos artículos han estudiado el crecimiento de la productividad en las sucursales bancarias. Por ejemplo, se han empleado métodos DEA para evaluar el crecimiento de la productividad de las sucursales bancarias, según el índice de Malmquist, de Camanho y Dyson (2006) y Asmild y Tam (2007), entre otros. Más específicamente, Camanho y Dyson (2006) evaluaron el desempeño de 144 sucursales bancarias en Portugal para determinar su diferencial de eficiencia dentro del grupo. Este estudio también examinó la eficiencia de la gestión interna y la productividad de la frontera, que refleja el impacto de los factores ambientales y las políticas de gestión regionales. Finalmente, Asmild y Tam (2007), calcularon las diferencias de frontera entre 115 sucursales bancarias de un mismo banco ubicadas en 6 países diferentes como caso de estudio para probar su metodología. El uso de índices globales para calcular las distancias entre las fronteras nacionales proporciona información adicional y valiosa.

Precisamente, el Capítulo 9 analizará el efecto de la crisis y rescate del sistema bancario español sobre las sucursales bancarias.

## **5.8. ESTUDIOS USANDO FRONTERAS ESTOCÁSTICAS**

A diferencia de la abundante presencia de literatura sobre el uso del modelado de DEA para analizar el desempeño de las sucursales bancarias (ver Fethi y Pasiouras (2010) y Paradi y Zhu (2013) para dos excelentes estudios), pocos trabajos han descrito

el uso de los modelos estocásticos de frontera en este contexto. De hecho, hasta donde sabemos, sólo tres estudios han adoptado este enfoque.

Por lo que sabemos, muy pocos estudios han analizado la eficiencia de las sucursales bancarias utilizando modelos de frontera de costes estocásticos (SCF) y datos de panel. Uno de ellos es el artículo de Berger et al. (1997), quienes utilizaron una función de coste translog y términos trigonométricos de Fourier para reducir el número de términos. Su ecuación se estimó por separado para cada período de tiempo, lo que permite que los parámetros de la función de costes varíen con el tiempo. En otras palabras, no estimaron ningún modelo de datos de panel utilizando el método SCF. Estos autores mostraron la existencia de aproximadamente el doble de sucursales de las que se necesitarían para minimizar los costes, pero observaron que esta "ineficiencia" podría de hecho ser óptima desde el punto de vista de la rentabilidad porque la "ramificación excesiva" aumenta los ingresos al mejorar la conveniencia del cliente. Los valores de ineficiencia  $X$  calculados fueron bastante grandes, más del 20% de los costes operativos. Estos hallazgos pueden explicar algunos de los resultados de eficiencia que se encuentran comúnmente en el análisis a nivel bancario y tienen importantes implicaciones con respecto a fusiones y adquisiciones y, sucursales interestatales en el sector bancario.

Otros estudios al respecto fueron realizados por Osiewalski y Marzec (1998) y Marzec y Osiewalski (2008), quienes analizaron la rentabilidad de las sucursales bancarias utilizando un marco bayesiano.

Además, Marzec y Osiewalski (2008) analizaron la rentabilidad a corto plazo de las sucursales bancarias utilizando una metodología bayesiana en un marco de datos de panel. Estos autores presentaron estadísticas bayesianas y muestreo de Gibbs, empleadas en una técnica de simulación MCMC, como herramientas para hacer

inferencias en modelos SCF para datos de panel sobre las sucursales de un banco comercial polaco. Se utilizó una función de coste translog y la ineficiencia se trató como un efecto individual aleatorio, asumiendo la especificación de distribución de eficiencia variable (VED) propuesta por Koop et al. (1997).

Algunos trabajos empíricos no bayesianos en el campo del análisis de eficiencia bancaria han utilizado un método de dos pasos mediante el cual las estimaciones de eficiencia obtenidas en la primera etapa fueron regresadas (en la segunda) sobre variables explicativas adicionales. En este enfoque, el modelo VED bayesiano utilizado proporcionó un marco coherente para estimar y probar la influencia de factores exógenos en la eficiencia individual. Luego, se aplicó la metodología bayesiana para obtener inferencias sobre la tecnología y los niveles de eficiencia de costes a corto plazo de 58 sucursales del mismo banco. Los resultados obtenidos, basados en datos de panel de los cuatro trimestres de un solo año y en una frontera de coste variable translog, indican rendimientos crecientes a escala (que varían con la producción de la sucursal) y no hay diferencias sistemáticas en la eficiencia que pudieran explicarse por las tres variables ficticias bajo consideración. El ejemplo presentado también muestra que la cooperación con econométricos bayesianos puede proporcionar importantes conocimientos sobre el funcionamiento económico de un banco, proporcionando a la administración información no solo sobre las características microeconómicas básicas de cada sucursal, sino también sobre su eficiencia y los posibles determinantes de esta eficiencia.

Más recientemente, Gómez-Déniz y Pérez-Rodríguez (2017) ampliaron los resultados obtenidos por Gómez-Déniz y Pérez-Rodríguez (2015) a la distribución normal-exponencial con dependencia del error en el modelo de frontera estocástica. Este modelo asume una estructura de dependencia entre el término de ineficiencia y el error

aleatorio (componente idiosincrásico), y anida el modelo clásico normal-exponencial cuando el parámetro de dependencia es cero. Estos autores utilizaron una función de coste log-lineal Cobb-Douglas de las sucursales de un único gran banco comercial español para mostrar la aplicabilidad de su modelo propuesto, analizando datos mensuales para el período 2011-2014. Las variables utilizadas fueron los costes, el precio de la mano de obra y del capital y los ingresos de cada sucursal bancaria, todos ellos transformados logarítmicamente. Los resultados obtenidos mostraron que el modelo clásico normal-exponencial dio un ajuste ligeramente peor que el modelo con dependencia entre errores, y que no se rechazó el supuesto de dependencia. Este estudio utilizó datos de las sucursales bancarias solo para ilustrar el ajuste del modelo, no para interpretar los resultados relacionados con las actividades de los gerentes de sucursales, cuestiones de política o las fortalezas y debilidades de las operaciones de la sucursal.

Sin embargo, y para nuestro conocimiento, ningún estudio ha analizado la relación entre el desempeño de las sucursales bancarias y la eficiencia dentro del marco de SCF. Precisamente, esta será la aportación del Capítulo 8.



## **MODELO DE DIRECCIÓN, DATOS Y DIMENSIONES**

---

### **6.1. INTRODUCCIÓN**

La oficina bancaria es un elemento importante en la captación de clientes y generación de ingresos de los bancos. Conocer sus fases operativas es de interés para elaborar un modelo de eficiencia apropiado, dado el carácter multi-objetivo que éstas poseen: oficina captadora de depósitos de clientes, prestamista y maximizadora de ingresos y beneficios.

A continuación, describiremos un modelo de eficiencia de la red de oficinas bancarias que sirve para ilustrar las distintas fases de actividad de una oficina.

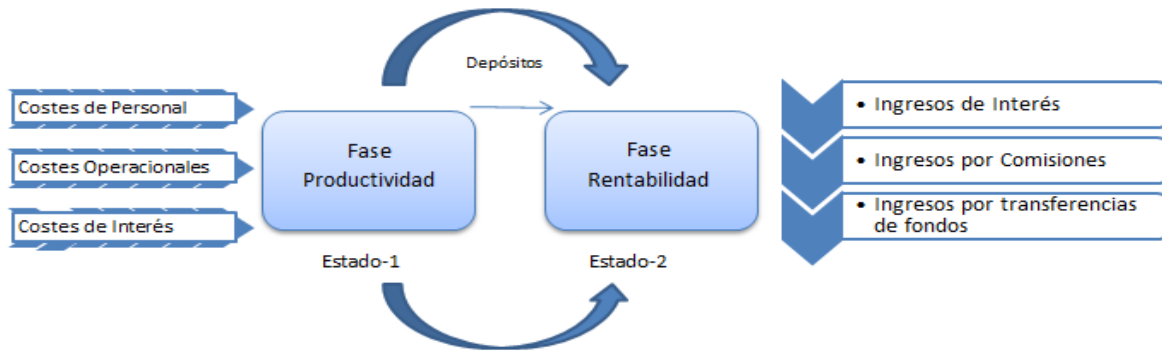
Siguiendo a Yang y Liu (2012, página 453), la Figura 6.1 ilustra dos fases de la actividad de una oficina que deberían tenerse en cuenta para analizar la eficiencia de su grado de desempeño (por ejemplo, producir, intermediar, obtener rentabilidad). La primera fase es la fase operativa o de producción y la segunda es la de rentabilidad.

Atendiendo a dicho gráfico, y para cada oficina, podríamos interpretarlo de la siguiente manera. La etapa operativa considera la producción de outputs intermediados (por ejemplo, depósitos, préstamos) utilizando para ello diversos inputs que pueden ser unidades físicas (cantidad de trabajadores, superficie de las oficinas) o unidades monetarias (costes de personal, costes operativos o coste de intereses). Los outputs obtenidos en la primera etapa (generalmente unidades monetarias) se utilizan como inputs en la etapa de obtención de rentabilidad con la finalidad de generar nuevos



outputs que son los ingresos finales (por ejemplo, renta de intereses, comisiones, o ingresos por transferencia de fondos, entre otras).

**Figura 6.1.** Un modelo de eficiencia de la red de oficinas bancarias (Yang y Liu, 2012).



En este sentido, las actividades de producción e intermediación pueden estudiarse por separado o pueden analizarse conjuntamente en un marco integrado. Por un lado, en el caso de realizar el análisis de cada etapa, la eficiencia de la producción se estudiaría en la primera etapa. En ésta se indicaría la eficiencia en la captación de recursos de los clientes. Como veremos, esto se puede analizar usando el denominado enfoque de producción o el enfoque de intermediación. Por otro lado, la segunda etapa permitiría analizar la eficiencia de la oficina para generar ingresos/beneficios por las actividades de intermediación. Cabe resaltar que, en el ámbito empírico, la eficiencia de ambas etapas se evalúa usando tanto los modelos DEA como los SFA. Por otro lado, en lo que respecta al análisis conjunto, menos estudiado, destacaríamos la programación multiobjetivo conjunta usando los métodos DEA y de subconjuntos borrosos (*fuzzy*) (propuesta por Yang y Liu (2012), por ejemplo).

Atendiendo a lo anterior, en este Capítulo describiremos sucintamente el modelo de dirección/producción de las oficinas del banco analizado, así como los datos empleados

para el análisis empírico y las diferentes dimensiones para el estudio de la eficiencia de las oficinas bancarias.

## **6.2. MODELO DE PRODUCCIÓN/DIRECCIÓN DEL BANCO ANALIZADO**

En esta sección describimos el tipo de negocio y objetivos del banco, así como, las características de las oficinas en términos de personal e infraestructuras, transacciones, ingresos y la evaluación de las oficinas.

### **6.2.1. TIPO DE NEGOCIO Y OBJETIVOS**

El banco que analizamos es uno de los tres grandes bancos españoles por cifra de negocios y posee tres tipos de actividades comerciales (negocio). Por una parte, se encuentra la banca minorista - de particulares -, comúnmente llamada **Banca Comercial**, que es el pilar fundamental de cualquier entidad financiera. En este rango de la base de la pirámide se encuentran personas físicas con una capacidad patrimonial limitada. En la entidad, los gestores especializados se encargan de asesorar a través de la **Banca Privada** a clientes con una situación patrimonial más avanzada. Y en el caso de empresas de mayor tamaño que una pyme, la **Banca de Empresas y Corporativa**, que gestiona los segmentos de Empresas, Corporativa e Instituciones.

El modelo de dirección tiene como objetivo orientar la actividad a la consecución de resultados sostenibles como palanca para desarrollar la visión cliente, a la vez que, motivar y desarrollar de forma continuada a los equipos.

Desde el punto de vista de la orientación al desarrollo del máximo potencial del negocio, lo que llaman el **eje estratégico**, se centraliza en la detección de oportunidades

en base al potencial de mercado y desempeño interno, así como, la concreción de objetivos y líneas de actuación a largo plazo. Desde el otro punto de vista, la orientación a la consecución de retos a medio plazo, **eje táctico**, se centra en focalizar la actividad de la red a la consecución de los objetivos a medio y corto plazo, asegurar la realización de la actividad mediante la aplicación del modelo de dirección.

La separación de la red comercial permite impulsar la actividad de las oficinas a través de un modelo con foco en el cliente (*customer-centric*), cuyo principal foco radica en el incremento de la vinculación y en una gestión del servicio especializado por segmento de cliente.

La **red comercial** incluye toda la red de distribución minorista, aparte de los segmentos de banca privada y patrimonial, para los cuales el banco tiene abiertas 151 oficinas especializadas y repartidas a lo largo de toda la geografía de las islas. El número de oficinas está en continua modificación en función del entorno económico y de la estrategia del Grupo.

### **6.2.2. LAS OFICINAS**

La oficina como punto de venta tiene una delegación especial de precios y volumen para las operaciones de inversión y recursos. Esto repercute directamente en el balance de la oficina que da lugar al margen de intereses que construye a su vez, los resultados.

Entre las funciones principales de la oficina está la captación de clientes, y vinculación de estos a través de productos transaccionales como seguros, tarjetas, nóminas o pensiones, fondos de inversión, etc.

### **6.2.3. PERSONAL**

Las oficinas cuentan con un personal especializado clasificado en diferentes perfiles.

Estos perfiles de las fuerzas de venta se enumeran en:

- Director de CBC: director de Centro de Banca Comercial que agrupa varias oficinas del mismo entorno geográfico.
- Director de oficina: responsable de una sucursal.
- Director comercial: abarca la parte comercial sin especialización dentro de una oficina.
- Asesor financiero: atención personalizada a personas físicas.
- Gestor de Pymes: atención personalizada a Pymes y PAES.
- Gestor comercial: atención no personalizada -pool bancario- .
- Gestor de atención al cliente: servicio de caja para pagos y cobros en efectivo.

El modelo de gestión de esta entidad se centra en el director de CBC como responsable de la gestión física con el cliente, remota y digital a través de oficinas extendidas en una zona con personal especializado. El personal tiene conocimientos para asesorar adecuadamente a los clientes en los diferentes productos o servicios que se ofrece en la oficina. También les corresponde gestionar la inversión irregular y los activos dudosos gracias a alertas de riesgos que se incorporan en las herramientas informáticas provistas en los terminales de cada fuerza de venta.

Como se vislumbra, el gestor de la oficina es un gestor de output y no de input. Ya que, el control de gastos recurrentes y de mayor volumen como los gastos de personal, alquileres y derramas operacionales y de procesos informáticos, se llevan desde las unidades centrales como Eficiencia, Control de Gestión y Talento y Cultura

(tradicionalmente denominado Recursos Humanos), en función del tipo de gasto a que nos refiramos.

La **gestión de equipo** o gestión de personal significa el diseño de estrategias encaminadas a desarrollar y aprovechar el potencial de los trabajadores. En estos reside el conocimiento, la experiencia y el don de hacer que las cosas cambien positivamente.

Un aspecto relevante para el desarrollo profesional del Departamento de Talento y Cultura es el proceso bienal, mediante el cual se identifican las fortalezas y áreas de mejora de cada uno de los profesionales. Cada profesional recibe feedback por parte de su responsable y se establecen los planes formativos y de desarrollo. Los resultados del proceso, junto con los indicadores de desempeño, son aspectos clave para la identificación de talento crítico del Grupo y para apoyar las decisiones de promoción.

Es destacable la preocupación de la empresa bancaria en valorar al capital humano, desde el proceso de reclutamiento y selección al buscar incorporar personas con potencial para aportar su experiencia y conocimientos específicos, como en dar oportunidades de crecimiento profesional a su personal, desde el de recién ingreso al más experimentado. Como garantía de crecimiento y elemento estratégico de la Política de Recursos Humanos, lo que se pretende es atraer, retener y potenciar el talento de sus empleados siempre desde el respeto a la diferencia y la inclusión. Por tanto, su objetivo es potenciar el desarrollo de todos los profesionales que laboren en el banco, valorar la permanencia de cada uno de ellos y garantizar la igualdad de oportunidades para aquellos que, contribuyan a generar valor y promuevan su diferenciación. En este sentido, la Gestión del conocimiento es un mecanismo que posibilita ofrecer una formación continua a medida, a través de diversos programas de desarrollo, cuya

intención es cuidar y fortalecer el capital humano procurando hacer compatibles los intereses profesionales y personales de cada persona con los objetivos de la empresa, potenciar las habilidades gerenciales de su capital humano y vigilar que cada puesto esté ocupado por el empleado más capacitado, esto es, identificar el talento individual, promover su crecimiento y retención.

Nos fijamos que cada día los mercados se hacen más competitivos y dado que son los trabajadores quienes se encargan de poner en movimiento a la organización es indispensable mejorar la calidad de su trabajo. En consecuencia, los responsables de la Gestión de Personal deben diseñar e implementar los procesos técnicos que favorezcan el desarrollo y la satisfacción del trabajador; lo que redundará en la satisfacción del cliente. Dada la correlación entre satisfacción del cliente y satisfacción laboral, podríamos concluir que, junto con la Gestión de Personal, constituyen facetas de una misma realidad.

Las compañías que integran la lista de *Great Place to Work*, que analiza anualmente cuáles son los mejores sitios para trabajar, "no sólo son competitivas en sueldo, sino que también gratifican al empleado con otras prácticas". Es lo que se llama el "salario emocional" y abarca desde cursos de formación, cheques guardería o iniciativas para fomentar la salud y el deporte en la empresa.

En cuanto al ámbito de la **Experiencia Cliente** se potencia la satisfacción del cliente. Tener clientes complacidos o plenamente satisfechos es uno de los factores clave para alcanzar el éxito en los negocios. En esta entidad se está potenciando la transformación para que tenga un profundo impacto en la experiencia de cliente y cuyo objetivo es ser el mejor banco para sus clientes en todas las geografías. La banca digital es una palanca para alcanzar el liderazgo en satisfacción de clientes en todas las

geografías, de acuerdo con los estándares establecidos por el indicador de satisfacción *Net Promoter Score* (NPS).

Estamos viviendo momentos en donde la banca se está transformando de forma profunda y se está redefiniendo el modelo de producción. Tradicionalmente, los bancos como instituciones financieras desempeñaban un rol clave en el sistema económico, la estructura que componen permite la transferencia de dinero entre los ahorradores e inversores y los prestatarios. Un banco acepta depósitos para luego, canalizar estos recursos y darlos en préstamo en el mercado de capitales hacia las distintas actividades económicas. Así los bancos son la conexión o intermediarios entre los individuos con déficit de capital e individuos con excedente de capital. Sin embargo, con la revolución digital la sociedad demanda funciones de servicio financiero que permitan contribuir al progreso de las personas y las empresas, apoyar el crecimiento, generar riqueza, crear puestos de trabajo e invertir en la sociedad.

#### **6.2.4. INFRAESTRUCTURAS**

El capítulo de **Operativos e Infraestructuras** hace referencia, por un lado, a los costes necesarios para obtener ingresos y, por otro, a la gestión de la capacidad instalada en el mercado.

Dentro del apartado de **Operativos** se trabaja con las funciones primordialmente administrativas que se sitúan en centros operativos desarrollándose procesos de automatización. Una vez conseguida la automatización, que supone un proceso más ágil, se devuelve a la red. Esto conlleva un ahorro de tiempo y, por consiguiente, una reducción de los costes laborales.

En el capítulo de gastos operativos, se incluyen los soportados directamente por la propia unidad, así como los gastos imputados vía contabilidad analítica de costes como alquileres o derramas operacionales. Algunos de estos gastos administrativos soportados por la unidad, podrían ser:

- Alumbrado, agua y calefacción
- Representación y desplazamiento (atenciones con clientes, kilometraje, comidas, etc.)
- Informes comerciales (verificaciones registrales)

Un ejemplo de los gastos imputados vía contabilidad analítica serían los,

- Alquileres propiedad del Banco
- Derramas operacionales
- Desarrollos informáticos
- Conservación equipos eléctricos
- Derrama Contact Center
- Derrama Servicios de Sociedades que dan servicio a las oficinas (incidencias, documentación de propuestas, ...)
- Servicios Jurídicos

En **Infraestructuras** se engloban las unidades de análisis, sistemas y ATM. Con unidades de análisis hacemos alusión a la red comercial del modelo tradicional bancario basado en los saldos de clientes. Sin embargo, la transformación que se está llevando a cabo hoy en día, se encauza hacia sistemas *freelance* con visión comercial donde el foco es el cliente. Uno de los costes más altos es el desarrollo de los sistemas dentro de los



CPD (Centro de Procesos de Datos) para procesar toda la información de la clientela y mantener la seguridad de las transacciones.

El **ATM**, siglas en inglés de *Automated Teller Machine* (es decir, cajero automático), es una computadora especializada que permite manejar el dinero de los usuarios de forma conveniente. Se trata de una máquina conectada informáticamente con un banco que permite efectuar al cliente ciertas operaciones bancarias mediante una tarjeta, libreta magnética o incluso con el móvil (los sistemas más avanzados), que tienen asignada una clave personal. Entre las funciones principales de esta máquina encontraremos la retirada de dinero, la opción de hacer depósitos, imprimir un estado de cuenta (un registro de la actividad o de las transacciones en su cuenta), verificar su saldo, transferir dinero entre cuentas e incluso pagar recibos. Ya se prueban servicios comerciales como abrir una cuenta e identificación de clientes. Normalmente, los cajeros automáticos ofrecen el máximo de servicios. Los cajeros automáticos son una manera segura y conveniente de manejar el dinero. Hay millones de cajeros automáticos en todo el mundo y se pueden usar muchos de ellos las 24 horas del día, los 7 días de la semana.

Estas máquinas se han desarrollado tras escuchar y entender las necesidades de los usuarios para crear soluciones y mejorar el servicio gracias a la innovación donde el ahorro de tiempo es clave. Al igual que para que las empresas del sector donde se mejora el rendimiento del personal al emplear menos tiempo en servicios básicos.

## **6.2.5. TRANSACCIONES Y SALDOS DE CLIENTES**

Hoy en día se pueden realizar todo tipo de operaciones bancarias en los cajeros automáticos, que han pasado de retirar efectivo o saldos, a recibir depósitos, hacer pagos con tarjeta, retiros sin tarjeta y cambio de NIP.

Además de la evolución de los cajeros, otras operaciones financieras como las transferencias han dado un giro de 360 grados gracias al mundo digital y a las aplicaciones bancarias, pues a través de éstas es posible realizar pagos, recargas o retiradas de efectivo sin tarjeta, entre otros. Estos canales que, en épocas de digitalización siguen siendo indispensables, cada vez brindan mejores funcionalidades adaptándose a nuevas funcionalidades para mantenerse vigentes y al mismo tiempo ofrecer nuevas experiencias a sus clientes, facilitando la manera de interactuar con estas máquinas. Algunas de las operaciones que se pueden realizar son las siguientes:

- Consulta de saldos.
- Transferencias y traspasos entre cuentas.
- Depósitos en efectivo a cuentas.
- Pagos de servicios con tarjeta de crédito.
- Pagos de servicios en efectivo.
- Pago de impuestos.
- Cambio de NIP.

Las transferencias habituales para las que se realiza seguimiento en las entidades son aquellas realizadas de forma automática a través de los OTP-autoservicios, línea telefónica, canal internet y canal oficina. La experiencia de estos últimos años señala

que existe una gran proporción de operaciones de disposición/ingreso de efectivo, pago de tributos por cuenta corriente o efectivo, seguidos de transferencia a otras cuentas o trasposos entre cuentas personales. Operaciones que realizadas en autoservicios evitan acudir a las sucursales reduciendo la frecuencia de visita a las entidades bancarias y, siendo un punto clave de digitalización y transformación en la relación con el cliente.

### **6.2.6. INGRESOS**

Los productos transaccionales permiten generar ingresos que se traducen en los OPOs de la cuenta de resultados de la oficina. Los OPOs engloban las comisiones y resto de Operaciones Financieras. Los OPOS hacen referencia a “*Otros Productos Ordinarios*” que se definen como los cobros que las entidades practican a sus clientes por la prestación de todo tipo de servicios (Comisiones totales + Diferencias de Cambio + Resto ROF (Resto Operaciones Financieras)).

Los ingresos los podemos clasificar:

1) Generados de manera directa por el propio negocio de la oficina:

- Riesgos de firma (Avales y Créditos Documentarios)
- Reclamación de Descubiertos
- Administración y mantenimiento de cuentas
- Tarjetas
- Efectos, cheques y órdenes de pago

2) Generados como compensaciones por acuerdos de Venta Cruzada (los resultados son compartidos por más de una unidad):

- Fondos de Inversión y Pensiones
- Seguros
- Valores
- Patrimonios
- Derivados
- Compensación por delegación de voto para la Junta General de Accionistas (JGA)
- Compensaciones por venta de activos adjudicados

Los ingresos se miden por el **margen bruto** generado en la unidad. El margen bruto se mide como la suma del margen de intereses y los OPOs.

El **margen de intereses** es la diferencia entre los intereses que se cobran por las inversiones crediticias y financieras que se realizan (Activo) menos los intereses que pagamos a la clientela y demás prestamistas por el dinero que éstos depositan en nuestra entidad (Pasivo). Este era el modelo tradicional de banca donde se cobra a los clientes por el activo y se paga por los depósitos prestados. Sin embargo, la situación actual de la curva de tipos, donde ha llegado a ser negativo, ha dado un vuelco al modelo de producción, ya que existen operaciones hipotecarias en el balance de un banco a tipos cercanos a cero y en el lado de los Recursos Líquidos (RLIs), imputándose en el balance a un coste elevado. De tal manera que, estamos viviendo una era de transformación del sector muy sesgada en la reducción de los costes ya que, el margen que nos dejan las operaciones de activo y pasivo es muy estrecho. De ahí, la creciente importancia que

están tomando el capítulo de los OPOs (ingresos por servicios, comisiones por actividad y resto operaciones financieras) en la Cuenta de Resultados de los Bancos.

### **6.2.7. EVALUACIÓN DE LAS OFICINAS**

Las unidades se evalúan a lo largo del año a través de unos indicadores consensuados centralizadamente y que tienen en cuenta el **Plan Estratégico**, tales como:

- Resultados y rentabilidad
- Volumen de negocio
- Calidad del crédito
- Base de clientes
- Satisfacción y calidad del servicio
- Operativos e infraestructura
- Localización geográfica
- Gestión de Equipos

De los capítulos señalados anteriormente, toman principal relevancia los resultados y la rentabilidad de las unidades.

Para medir el capítulo de ganancias se hace uso del balance y la cuenta de resultados de cada unidad.

En los últimos años, las funciones de red son valoradas por el margen bruto y los OPOs generados en la unidad. La valoración se realiza a través de las desviaciones presupuestarias que se presentan en estos capítulos. Los presupuestos son una de las

herramientas fundamentales para valorar los resultados obtenidos por la empresa, tanto a nivel general como por centros de responsabilidad o por productos. Es imprescindible comparar lo sucedido realmente con lo presupuestado. Se entiende que se producen desviaciones cuando los resultados reales son diferentes a los resultados previstos, que pueden ser desviaciones positivas cuando los ingresos superan los objetivos o negativos cuando el volumen de ingresos queda por debajo de los presupuestos marcados. El objetivo del control presupuestario es detectar y analizar las desviaciones, como parte esencial de la planificación empresarial. A medida que a través de la contabilidad general se obtienen los datos reales, deben compararse con los previstos. La diferencia o desviación suele calcularse en valor absoluto y en porcentajes, tales que:

- En términos absolutos:

$$\text{Desviación} = \text{cantidad prevista} - \text{cantidad real}$$

- En porcentaje:

$$\text{Desviación (\%)} = \frac{\text{Desviación}}{\text{Presupuesto}} \times 100$$

También, la entidad utiliza un modelo para medir el volumen de negocio de la red a través de la medición del valor recurrente generado con las ventas con un horizonte trimestral, ya que, una operación contratada en un determinado ejercicio generará ingresos en ejercicios posteriores, por lo que, es necesario medir el valor generado en términos de Valor Actual Neto (VAN). A este modelo se le conoce como **PBV**, Productividad Basada en Valor. La PBV pretende medir la actividad comercial en términos del valor generado por una venta u operación considerando, principalmente, variables como el volumen de esta, el precio, el plazo y el riesgo. Se mide trimestralmente durante el año a través de informes diarios de ventas y sus puntos

asignados. El informe comprende tres apartados diferenciados: Recursos, Inversión y Productos. Los capítulos de recursos e inversión tendrían su reflejo directo en el balance de la oficina, mientras que, los ingresos que nos genera la venta de los productos se percibirían en los OPOs dentro de la cuenta de resultados. Este seguimiento está disponible no sólo a nivel de unidad sino para cada gestor de la unidad con las ventas propias de éste. La PBV se calcula como un índice, por lo que se expresa en unidades porcentuales.

Por último, cabe destacar que una vez cerrado el año, se evalúan todos los indicadores y, en función de éstos, la entidad incentiva a los empleados a través de una remuneración variable (sistema DOR – Incentivación Anual). La entidad lo utiliza como incentivo y motivación a los empleados, a través de una puntuación calculada en base a la consecución de los objetivos de las distintas líneas estratégicas. Los empleados reciben anualmente un bonus monetario y variable, en base a los resultados del Grupo, de España y, de forma individualizada, al perfil del empleado y su puntuación.

### **6.3. DATOS**

Tal y como se ha destacado en el Capítulo 1, el sector bancario ha experimentado una importante reestructuración desde la crisis financiera de 2007, en lo que respecta a su regulación, necesidades de capital y liquidez, así como en relación con el comportamiento de los clientes y las relaciones con los proveedores de servicios financieros dentro de un entorno totalmente nuevo, con elementos novedosos como Internet, teléfonos inteligentes y redes sociales.

El sistema bancario español (acreditado como uno de los más sólidos y mejor equipados entre todas las economías occidentales para hacer frente a la crisis mundial de liquidez) ha sido profundamente reestructurado desde la crisis bancaria de 2010 y el rescate de 2012 provocado por la burbuja inmobiliaria, la cual ha tenido un impacto en la economía española y sus instituciones financieras.<sup>23</sup>

La información muestral utilizada para la realización del análisis empírico se obtiene en el ámbito de la banca minorista pura, que es un elemento central del sistema bancario español. Sin embargo, debido a la limitada disponibilidad de datos y a las dificultades derivadas de las obligaciones de confidencialidad de los bancos, este estudio se centra en un gran banco comercial español anónimo, que opera a través de una red regional de oficinas bancarias.

El banco en cuestión es representativo para nuestro propósito, como uno de los tres bancos más grandes de España en términos de activos, dado el alto grado de concentración en la banca de nuestro país. Este banco no se vio afectado por el rescate, aunque el escenario macroeconómico afectó ligeramente a su productividad laboral interna y a su eficiencia operativa. Esto podría deberse a que el banco tiene un modelo de cartera diversificada y equilibrado y, porque su gestión se basa en la prudencia y la anticipación.

En esta Tesis Doctoral se utiliza información mensual correspondiente a 151 sucursales bancarias de un gran banco comercial español durante el periodo que abarca desde 2011 hasta 2014. Para conseguir un panel balanceado para todo el período al

---

<sup>23</sup> Bankia, un banco importante y el mayor prestamista hipotecario del país, fue nacionalizado el 9 de mayo de 2012. El 25 de mayo de 2012 anunció que requeriría un rescate de 23.500 millones de euros para cubrir pérdidas por hipotecas fallidas. El rescate del sistema bancario español incluyó la toma de pérdidas por parte de los inversores de hasta 10.000 millones de euros, la creación de un "banco malo" para absorber hasta 45.000 millones de euros de préstamos fallidos acompañado del cierre de miles de sucursales bancarias.



producirse cierres continuos, se depura resultando un total de 135 sucursales. Este periodo incluye la segunda recesión económica española (2011-2013) y el rescate del sistema bancario español (2012) a través del paquete de rescate de 100.000 millones de euros por el Mecanismo Europeo de Estabilidad (ESM). En este período, los beneficios de los bancos disminuyeron y era una práctica común de la administración tratar de minimizar las ineficiencias en las operaciones bancarias.

Estas sucursales se ocupan de clientes (particulares) y cuentas de negocios (empresas) y sus actividades son homogéneas (se proporcionan servicios similares en ambos casos: líneas de productos, programas de entrenamiento, valores y objetivos corporativos, sistemas de tecnología de información, entre otros).

Dichas sucursales pertenecen a una región de España, y contienen oficinas de núcleos de población rurales y urbanos, algunas de ellas con una mayor concentración de oficinas que otras. Por razones de confidencialidad no proporcionamos referencias sobre ellas, pero puede solicitarse información a la autora.

Más concretamente, utilizamos las siguientes variables:

- El número total de empleados (incluidos los directores de CBC, sucursales y directores comerciales o apoderados, el personal especializado, comercial y los cajeros). El personal de la sucursal es considerado como Fuerza de Venta a excepción de los gestores de atención al cliente, cuya función principal se destina al manejo de efectivo en las labores de caja.
- El número de cajeros automáticos externos y la superficie de la sucursal (en m<sup>2</sup>), el número total de transacciones (Internet, línea y oficina).
- El número de transacciones en cajeros automáticos externos (es decir, el volumen de servicios prestados a través de medios automatizados).

- El número de préstamos totales (es decir, tanto los relacionados con el número de consumidores particulares como el número de préstamos a empresas).
- El número total de depósitos. Recordamos que un depósito a plazo fijo es un producto en el que un cliente entrega a una entidad bancaria una cantidad de dinero durante un período de tiempo, a cambio de una remuneración en forma de tipo de interés aplicado sobre el capital prestado y que se puede pagar al final del período o durante el mismo, en fracciones. Si el cliente reclama el dinero antes de que finalice el plazo pactado, es habitual que en el contrato de depósito se contemple una penalidad, de tal manera que se reduzca la rentabilidad del mismo.
- Los costes operativos (gastos totales excluidos los costes de personal).
- Los costes laborales.
- El valor total de los préstamos en miles de euros (la suma de las hipotecas, los valores de los préstamos de consumo y de empresas).
- El margen bruto (ingresos totales) en miles de euros (es decir, aquellos relacionados con el margen de intereses o intermediación más los OPOs).
- Los precios de los inputs para cada sucursal: trabajo (salario y prestaciones por empleado para cada sucursal en su mercado local) y el precio del capital (referido a los gastos de locales y equipos divididos estos por el valor de la construcción y la tierra).
- La productividad basada en valor (PBV) que mide la actividad comercial en términos del valor presente neto generado por una operación, tomando en cuenta variables como el volumen de la operación, su precio, plazo y riesgo, y el hecho

de que una transacción contratada en un año dado generará ingresos en años subsecuentes.

- Las desviaciones presupuestarias entre los resultados reales y estimados. La diferencia o desviación suele calcularse tanto en valor absoluto como en porcentaje. En el presente estudio, se utiliza el valor absoluto para reflejar el tamaño de las desviaciones,  $|D|$ .  $D^-$  representan las desviaciones negativas y  $D^+$  representan las desviaciones positivas.

Todos los datos de valor aparecen en € deflactados con base 2011 (es decir, utilizamos el índice de precios al consumidor del mercado local / regional en el período base 2011).

En los Capítulos 7 y 8 se detallarán los datos de las variables usadas en cada capítulo y en el periodo analizado.

#### **6.4. DIMENSIONES DEL ANÁLISIS DE LA EFICIENCIA TÉCNICA EN LA RED DE OFICINAS**

La Figura 6.1 muestra la complejidad de la actividad de una oficina bancaria y, por ende, lo difícil que puede ser evaluar la eficiencia de la misma. Téngase en cuenta que las oficinas, como captadoras de clientes, ofrecen complejos productos y servicios financieros (depósitos a la vista, a plazos, carteras, préstamos hipotecarios, préstamos personales, etc.), algunos de los cuales podrían estar incluso relacionados; y también ofrecen servicios que no son pagados directamente/instantáneamente por los clientes

(préstamos), además de las complejas regulaciones gubernamentales que podrían afectar a los servicios que ofrecen.

Dado estos aspectos, parece obvio que no exista un modelo único y perfecto capaz de capturar totalmente la naturaleza multi-rol de las oficinas bancarias, haciendo necesario un conjunto combinado de métricas para obtener de forma más precisa la eficiencia de las oficinas.

En general, estas métricas o enfoques permiten analizar las diferencias en el desempeño de los gerentes en el sentido de que los objetivos de los gestores persiguen incrementar el volumen de las ventas y gestionar los diversos productos ofrecidos de tal manera que, generen una rentabilidad elevada y se refleje en la Cuenta Resultados de la sucursal y por ende, del Banco.

Dicho lo anterior, y en general, la eficiencia de las oficinas bancarias puede evaluarse a través de tres dimensiones de análisis. Por ejemplo, si una oficina bancaria intenta mejorar sus resultados al prestar, entonces se evaluaría la eficiencia de intermediación. Por otro lado, también puede analizarse su eficiencia operacional, es decir, producir depósitos y préstamos, o también su eficiencia de rentabilidad, o sea, la capacidad de obtener ingresos. Estas tres dimensiones se denominan producción, intermediación y rentabilidad (*profitability*), las cuales describiremos más adelante.

Para analizar dichas aproximaciones se hace necesaria una selección adecuada de los inputs y outputs, que pueden ser distintos al ser diversos los objetivos. En general, los inputs apropiados son aquellos que la oficina debería minimizar, mientras que los outputs apropiados son aquellos que la oficina debería maximizar. Por supuesto, tanto la naturaleza como el número de inputs y outputs seleccionados para el análisis de la eficiencia están condicionados a la disponibilidad de los datos.

Cabe destacar que, aunque existe un acuerdo general sobre las principales categorías de inputs y outputs para evaluar la eficiencia de los bancos y las sucursales, no existe una lista única de los inputs/outputs específicos que son utilizados en los diversos estudios (Fethi y Pasiouras, 2010).

A continuación, se describen someramente los tres enfoques.

Estos enfoques pueden usarse tanto en el marco del modelo DEA como SFA y permiten evaluar el grado de desempeño de las unidades y sus gestores en cada dimensión. En general, describiremos los inputs y outputs más comúnmente empleados en dichos enfoques, fundamentalmente aquellos empleados para analizar la eficiencia de las oficinas bancarias usando DEA (véase Camanho y Dyson, 1999, 2006, o el survey de Paradi y Zhu, 2013, entre otros).

#### **6.4.1. ENFOQUE DE PRODUCCIÓN**

La medida de eficiencia de producción captura el nivel más eficiente de costes operativos para gestionar los datos económicos de las oficinas. En particular, esta aproximación analiza cómo cada oficina utiliza sus recursos (inputs) para producir préstamos, depósitos, y otras transacciones (garantías bancarias, comisiones sobre importaciones y exportaciones). Por lo tanto, el enfoque de producción enfatiza la actividad operativa, es decir, las oficinas bancarias son vistas como proveedoras de servicios y productos a los clientes utilizando personal especializado en tareas bancarias y otras fuentes como inputs, para obtener como outputs los depósitos, préstamos y otros productos.

En esta dimensión, los inputs y outputs utilizados pueden ser los siguientes. Por ejemplo, los inputs generalmente incluyen variables físicas relacionadas con los costes de operación (es decir, mano de obra, materiales, espacio o sistemas de información). Sin embargo, algunos autores también emplean los costes operacionales como inputs (véase Camanho y Dyson, 1999, 2006). El coste de operación refleja un rango variado de consumos realizados por una sucursal, que consisten en gastos de electricidad, teléfono, seguros, gastos de publicidad, papelería y otros suministros.

En cuanto a los outputs, generalmente estos representan el número de transacciones realizadas por las sucursales (transacciones internas y externas (ATM) de las oficinas), el número total de préstamos concedidos y el número de depósitos. La Tabla 6.1 resume dichas variables.

**Tabla 6.1.** Dimensión de producción.

<b>Inputs (#)</b>	<b>Outputs (# de transacciones en las sucursales)</b>
Número total de empleados	Número total de transacciones
Número total de ATMs	Número total de transacciones en ATMs
Superficie de la sucursal (en m <sup>2</sup> )	Número total de préstamos (consumo y empresas)
Costes operativos*	Número total de depósitos

\* Estos costes excluyen los costes laborales. Esta variable fue empleada por Camanho y Dyson (1999 y 2006) como input.

#### **6.4.2. ENFOQUE DE INTERMEDIACIÓN**

El enfoque de intermediación refleja cómo las instituciones financieras son vistas como intermediarios entre ahorradores e inversores.

El enfoque de intermediación se utiliza generalmente para examinar cómo de eficiente en organización es la oficina en la colección de depósitos y otros fondos de los

clientes (inputs) y luego, prestar el dinero en varias formas: préstamos, hipotecas y otros activos (por ejemplo, renting, leasing, factoring, entre otros). Este enfoque es un indicador sólido de la fortaleza de la oficina en su habilidad de prestar.

En general, los inputs se miden por los costes de intereses y no intereses y los resultados reflejan los flujos financieros asociados con la actividad de intermediación. En este modelo, Paradi et al. (2011) consideran los saldos de efectivo (*cash balances*) comprenden efectivo en caja en las sucursales, los activos fijos / accruals (*fixed assets / accruals*) son activos mantenidos por la sucursal, y otros pasivos son otras obligaciones devengadas. Los préstamos netos incobrables o morosos (*Net non-performing loans*) son préstamos considerados como fallidos por la sucursal (por ejemplo, 90 días de intereses no pagados). Pérdida esperada de préstamos (*loan loss experience*) son los montos registrados como dotaciones para cubrir los posibles préstamos incobrables y se incluyen para provisionar a las sucursales en función de la calidad crediticia.

En cuanto a los outputs, los productos están compuestos por las principales líneas de negocios del Banco (en euros y representan el volumen total de negocios: préstamos y depósitos).

La Tabla 6.2 resume algunas de las variables que pueden emplearse como inputs y outputs en esta dimensión.

**Tabla 6.2.** Dimensión de intermediación.

<b>Inputs (en €)</b>	<b>Outputs (en €)</b>
Saldos en efectivo	Valor total de los préstamos (particulares y empresas)
Activos fijos/devengados	Valor total de los depósitos
Resto Pasivo	
Préstamos Dudosos	
Pérdida Esperada	

### 6.4.3. ENFOQUE DE RENTABILIDAD

La medida de eficiencia de rentabilidad captura el nivel más eficiente de los costes totales para generar beneficios. Las unidades más eficientes minimizan más los costes incurridos en generar los ingresos y harán mejor la maximización de los mismos. Esta aproximación analiza cómo la oficina utiliza sus recursos para generar ingresos tales como ingresos por intereses o ingresos por comisiones. Por lo tanto, el enfoque de rentabilidad (*profitability*) pretende medir la habilidad de la oficina para convertir los gastos (los que la administración de la sucursal puede influir directamente) en ingresos.

Los inputs y outputs utilizados pueden ser los siguientes. Los gastos incluyen costes de empleados, cargos entre oficinas, gastos de alquileres, y otros gastos operacionales (papelería, gastos de viaje, gastos de formación, ...). Del lado del output, los ingresos representan: ingresos por no intereses, ingresos netos derivados de los intereses de sus líneas de negocio, y las comisiones obtenidas por la oficina.

La Tabla 6.3 recoge algunas de estas variables.

**Tabla 6.3.** Dimensión de rentabilidad.

<b>Inputs (gastos en €)</b>	<b>Outputs (ingresos en €)</b>
Costes operativos	Ingresos por intereses
Costes laborales	Ingresos por servicios y comisiones
Provisiones crediticias (Deterioro)	

### 6.5. DIMENSIONES EN EL ESTUDIO DE LA EFICIENCIA DE COSTES

En general, el enfoque empleado para estudiar la eficiencia de costes en el marco de los modelos DEA es el de producción, en el sentido que tanto inputs como outputs son descritos por unidades físicas. No obstante, cabe destacar que también son empleados los enfoques de producción e intermediación en modelos SFA (Berger et al., 1997). En



el caso del enfoque de intermediación se consideran los valores monetarios, por ejemplo, de los volúmenes de depósitos y préstamos, sean estos consumos a particulares, hipotecarios o préstamos a empresas.

En general, y a diferencia de los modelos que analizan la eficiencia técnica, las variables usadas en los modelos de eficiencia de costes no solo son los outputs e inputs, sino que también son necesarios los precios de los inputs.

Por ejemplo, Camaho y Dyson (2005b) usan como inputs al número de trabajadores (número de administradores de sucursales y cuentas, número de personal administrativo y comercial, número de cajeros) y también los costes operativos. Como output utilizan solamente el número de transacciones y como precio del factor trabajo al salario medio de los trabajadores (salario promedio y beneficios adicionales de la sucursal y gerentes de cuentas, salario promedio y beneficios adicionales de la administración / personal comercial, y salario promedio y beneficios adicionales de los cajeros). En esta Tesis emplearemos algunas variables similares a Camanho y Dyson (2005b).

## ESTIMACIÓN NO PARAMÉTRICA DE LA EFICIENCIA TÉCNICA DE LAS SUCURSALES BANCARIAS<sup>(24)</sup>

---

### 7.1. INTRODUCCIÓN

Numerosos estudios han analizado la eficiencia bancaria en España, centrándose en bancos comerciales, cajas de ahorros y uniones de crédito, entre otras áreas. En general, los métodos aplicados en esta investigación previa se basan en procedimientos no paramétricos como el análisis envolvente de datos (DEA) y el análisis *free disposal hull* (FDH, una variante no convexa de DEA). Sin embargo, también se han utilizado procedimientos estocásticos paramétricos como el análisis de frontera estocástica (SFA).<sup>25</sup> Entre los métodos no paramétricos, la mayoría de los estudios han usado el marco DEA para estimar la eficiencia del banco, ver por ejemplo Pastor (1995), Grifell y Lovell, (1997), Pastor et al. (1997), Maudos et al. (2002), Cuesta y Orea (2002), Maudos y Pastor (2003), Guzmán y Reverte (2008), Färe et al. (2010) y San-José et al. (2014).

---

<sup>24</sup> Este Capítulo es una versión en castellano del artículo titulado: *Assessing branch efficiency and managerial behaviour in a large Spanish commercial bank*, escrito por Idaira Cabrera-Suárez y Jorge V. Pérez-Rodríguez y publicado en la *Revista Española de Financiación y Contabilidad – Spanish Journal of Finance and Accounting*, 2020, 49 (1), pp. 48 - 73.

<sup>25</sup> El enfoque de la frontera no paramétrica que utiliza DEA o FDH requiere unos supuestos mínimos sobre la estructura de la producción y no impone restricciones sobre la forma funcional que relaciona los inputs y los outputs. Como no tiene en cuenta el ruido en los datos, asume implícitamente que toda desviación de la frontera puede considerarse como ineficiencia. Sin embargo, en el SFA paramétrico, se hacen suposiciones tanto sobre la forma funcional como sobre la distribución de los dos tipos de error: un término de error idiosincrático y el término de error de ineficiencia (es decir, la desviación de la frontera).

También a este respecto Prior et al. (2016) realizaron un análisis de eficiencia para la banca española, utilizando el enfoque FDH para abordar los períodos antes y durante la reciente crisis económica internacional.<sup>26</sup> En un enfoque alternativo, SFA se ha utilizado en algunos estudios, como Maudos (1996), Maudos (2001), Maudos y Pastor (2001a, 2001b) y Orea y Kumbhakar (2004). Finalmente, Weill (2004) aplicó y comparó las técnicas DEA, FDH y SFA en un análisis de la industria bancaria en varios países, entre ellos España.

Los análisis a nivel de las sucursales bancarias suelen ser más útiles que los del nivel institucional, tanto para los investigadores académicos como para los gestores del banco. Los datos de las sucursales de un solo banco se recopilan de manera uniforme y comparable y, por lo tanto, no crean problemas graves para la interpretación de la eficiencia de escala. Además, la homogeneidad de las oficinas es un requisito básico de DEA para minimizar el sesgo de estimación (Coelli et al., 2005; Chortareas et al., 2009). Berger y col. (1997) indicó que la información sobre el desempeño de las sucursales puede ayudar a mejorar nuestra comprensión de la eficiencia bancaria y contribuir a resolver los problemas de medición en este contexto. Por lo tanto, el conocimiento de la eficiencia permite a los gerentes de sucursales identificar fortalezas y debilidades en sus operaciones y minimizar las ineficiencias en las operaciones bancarias, al resaltar las desviaciones de una unidad de “mejores prácticas” para cualquier escala o combinación de productos (Berger et al., 1997). También a este respecto, Paradi y Zhu (2013) indicaron que la red de sucursales suele ser la mayor

---

<sup>26</sup> Estos autores examinaron los bancos comerciales, las cajas de ahorros y las cooperativas de crédito. Calcularon tanto la eficiencia de los costes como la de los ingresos. Sus resultados indicaron que los bancos comerciales eran más eficientes que las cajas de ahorros y las cooperativas de crédito. Sin embargo, durante los años de la crisis, las diferencias entre estas instituciones se redujeron drásticamente, especialmente en lo que respecta a la eficiencia de costes.

fuente de gastos operativos de un banco, y que los clientes realizan una gran proporción de su banca de mayor valor agregado a través de sucursales. Por ejemplo, incluso en países europeos con tecnología digital avanzada, entre el 30% y el 60% de los clientes prefieren realizar al menos parte de sus operaciones bancarias en sucursales.<sup>27</sup>

El análisis realizado en este capítulo contribuye a la literatura empírica sobre la eficiencia de las sucursales bancarias de varias maneras. En primer lugar, estimamos la eficiencia técnica teniendo en cuenta que los estimadores de eficiencia no paramétricos son muy sensibles a la presencia de valores atípicos. Para ello, utilizamos el método de frontera parcial no paramétrica de orden- $\alpha$  (Aragon et al., 2005), que es más robusto frente a errores de medición y valores atípicos que los métodos tradicionales como DEA o FDH. El enfoque de orden- $\alpha$  generaliza FDH al permitir que las observaciones supereficientes se ubiquen más allá de la frontera de posibilidad de producción estimada. En segundo lugar, identificamos sucursales problemáticas, teniendo en cuenta que la frontera estimada no estará conformada enteramente por observaciones atípicas. Por lo tanto, se realiza un análisis de frontera parcial para detectar posibles valores atípicos en los datos, después de lo cual se aplica un análisis de eficiencia no paramétrico DEA o FDH (ver Daraio y Simar, 2007). Este enfoque es un medio apropiado para hacer frente a los valores atípicos, que de otro modo podrían aumentar el ruido y distorsionar los resultados. En tercer lugar, verificamos este método robusto no paramétrico con DEA y FDH, recordando que Krüger (2012) encontró que DEA es bastante robusto en los experimentos de Monte Carlo diseñados para estudiar métodos deterministas y estocásticos no paramétricos y paramétricos. Finalmente, evaluamos los

---

<sup>27</sup> Véase el estudio titulado McKinsey Retail Banking Multichannel Survey (2016), McKinsey & Company.

determinantes de la ineficiencia de las sucursales bancarias, probando la relación entre las puntuaciones de eficiencia relativa estimadas y los factores ambientales utilizando el procedimiento de Simar y Wilson (2007). Se prefiere este método porque los enfoques convencionales en dos etapas carecen de un mecanismo de generación de datos bien definido; por ejemplo, DEA genera un patrón complejo (desconocido) de correlación entre las puntuaciones de eficiencia estimadas.

Nuestro análisis tiene en cuenta varios factores ambientales o determinantes que pueden afectar la eficiencia, una métrica crucial para la evaluación de la sucursal bancaria que podría estar relacionada con los objetivos del banco (Lovell y Pastor, 1997). Por lo tanto, nos hemos centrado en dos métricas del banco para evaluar su desempeño. Por otro lado, analizamos los efectos de factores específicos de la sucursal, como el tamaño del mercado (rural / urbano) y el efecto de la experiencia (cuantificada como el número de años que la sucursal lleva funcionando). Esta última variable nos permite evaluar la curva de aprendizaje y determinar la capacidad de la sucursal para permanecer en el mercado (Bernini y Guizzardi, 2010). Otras variables consideradas son la cualificación del personal de la sucursal (en términos de certificación de calidad financiera), provisiones para pérdidas crediticias, ubicación geográfica y especialización de la sucursal. Además, tenemos en cuenta el efecto de la crisis bancaria española (y de los consiguientes procedimientos de rescate), y de la reforma regulatoria llevada a cabo en 2012. En definitiva, analizamos el comportamiento (eficiencia) de las sucursales bancarias y sus gerentes durante un periodo en el que las entidades financieras tuvieron que adaptarse a un nuevo escenario macroeconómico, con cambios regulatorios, y en las que una nueva recesión golpeó a la economía española (en el primer trimestre de 2012), cuando se produjo una importante reestructuración del sector bancario.

Este capítulo describe un procedimiento de dos etapas utilizado para evaluar los determinantes de la eficiencia operativa y las dimensiones de rentabilidad, como una indicación del desempeño de los gerentes de sucursales en un gran banco comercial español, antes y después del rescate del sistema bancario español en 2012. Este procedimiento se basa en métodos de frontera no paramétricos clásicos y no paramétricos robustos y en la detección de outliers basados en fronteras parciales (siguiendo a Daraio y Simar) que nos permite detectar posibles valores atípicos en los datos, después de lo cual, se realiza un análisis de eficiencia no paramétrico clásico. En una primera etapa, las eficiencias se obtienen a partir de los datos anuales proporcionados por las sucursales bancarias en el período 2011-2014. Se obtuvieron puntuaciones altas en eficiencias operativas y de beneficios para los estimadores DEA (eficiencias puramente técnicas y de escala), FDH y orden- $\alpha$ . En general, se obtuvieron las siguientes jerarquías de puntajes: DEA < FDH < orden- $\alpha$ . En la segunda etapa de nuestro análisis, se aplicó un enfoque de modelo econométrico censurado (siguiendo a Simar y Wilson) para identificar los factores explicativos de la eficiencia.

El resto del capítulo se estructura de la siguiente manera. La siguiente sección revisa la literatura previa en esta área. La sección 7.3 describe el uso de modelos DEA, FDH y de frontera parcial para determinar la eficiencia de las sucursales bancarias, teniendo en cuenta las diferentes formas de rendimiento a escala y las restricciones de los modelos. También describimos el procedimiento de Simar y Wilson. La sección 7.4 describe los datos utilizados y los principales resultados empíricos obtenidos.

## 7.2. REVISIÓN DE LA LITERATURA

Existe una extensa literatura sobre la eficiencia de las instituciones bancarias en su conjunto (ver Fethi y Pasiouras (2010), y sus referencias), y también sobre la de las sucursales bancarias. En este último aspecto, en particular, destacan dos artículos: en el primero, Fethi y Pasiouras (2010) discutieron la eficiencia de bancos y sucursales, analizando 30 estudios, realizados entre 1998 y 2009, que utilizaron técnicas DEA para estimar la eficiencia de las sucursales. Posteriormente, Paradi y Zhu (2013) identificaron 80 estudios en los que se aplicó DEA a los bancos y sus sucursales, durante el período 1985-2011.<sup>28</sup>

Aunque en el Capítulo 5 se han descrito y comentado muchos trabajos, a continuación, describimos brevemente los enfoques estocásticos paramétricos y no paramétricos para estudiar la eficiencia de las sucursales bancarias.

En primer lugar, en cuanto a los métodos no paramétricos, el método más utilizado es el DEA. Desde que Sherman y Gold (1985) escribieron el primer artículo de aplicación del DEA sobre el análisis de sucursales bancarias, varios artículos han propuesto mejoras metodológicas, como incluir los juicios de valor de los gerentes de sucursales (Athanassopoulos, 1997), introducir la formulación de lógica difusa (Wu et al., 2006), utilizando DEA multicapa (Alirezadeh y Afsharian, 2007) o un modelo DEA "culturalmente ajustado" (Paradi et al., 2010), evaluando el desempeño de la sucursal para múltiples dimensiones (Paradi et al., 2011), y evaluando la gestión eficiencia y proporcionar pautas de mejora (Camanho y Dyson, 1999: entre otros).

---

<sup>28</sup> De los 80 trabajos publicados, el mayor número se centra en las sucursales bancarias de Canadá, seguidas por las de Grecia, Portugal, Estados Unidos y Reino Unido, respectivamente.

Muchos de los artículos que analizan las sucursales asumen rendimientos constantes a escala, y las medidas de eficiencia radial siguen siendo los enfoques más utilizados. Sin embargo, la eficiencia en las operaciones generalmente se evalúa en diferentes áreas o dimensiones, porque ningún modelo único puede capturar completamente la naturaleza multifuncional de las sucursales bancarias. Los métodos que se han adoptado incluyen el enfoque de producción (centrado en la actividad operativa), el enfoque de intermediación (con respecto a la eficiencia con la que la sucursal recauda depósitos y otros fondos) y el enfoque de rentabilidad (cómo la sucursal genera ganancias). La calidad (o eficiencia) del servicio es otra dimensión importante del desempeño de las sucursales bancarias. Para cada una de estas dimensiones, existe una diversidad significativa entre los estudios anteriores en términos de la selección input/output realizada (ver Tabla A1 en Paradi y Zhu, 2013). Los métodos DEA se han utilizado para estimar las eficiencias técnicas y de escala, las eficiencias de asignación y de costes y, la evolución de la productividad de las sucursales bancarias.

La eficiencia técnica se ha analizado utilizando una o varias dimensiones. Con respecto a una sola dimensión, Parkan (1987) examinó la eficiencia de producción de 35 sucursales de un banco comercial importante en Canadá e comprobó que solo 11 eran relativamente eficientes. Camanho y Dyson (2005b) analizaron 144 sucursales de un banco comercial portugués y observaron considerables ineficiencias en esta red. Pastor et al. (2006) utilizaron DEA para examinar la eficiencia de los beneficios de 573 sucursales de una gran caja de ahorros europea e informaron que las mejoras de eficiencia en las sucursales con peores resultados (en las que se podrían obtener ahorros de hasta un 45% en el uso de recursos) generar un aumento sustancial de las ganancias para el banco. Con respecto a varias dimensiones, Oral y Yolalan (1990) evaluaron las eficiencias de utilidad y servicio de 20 sucursales minoristas de un banco comercial



turco, encontrando una relación positiva entre estos parámetros, pero que solo el 20% de las sucursales eran eficientes. Portela y Thanassoulis (2007) compararon la calidad del servicio y diferentes medidas de eficiencia, y apuntaron que la calidad del servicio se relacionó positivamente con la eficiencia operativa y de los beneficios o rentabilidad. Giokas (2008a) evaluó las eficiencias de producción, transacción y rentabilidad obtenidas en operaciones dentro de una red de 44 sucursales durante el ejercicio 2002. De acuerdo con los resultados empíricos obtenidos, había margen para mejoras sustanciales en la eficiencia (la ineficiencia media fue del 12%) y, por lo tanto, para mayores ganancias bancarias. Paradi et al. (2011) analizaron las tres dimensiones del desempeño de las sucursales relacionadas con la eficiencia técnica para las 816 sucursales de un importante banco canadiense y concluyeron que estos enfoques mejoraron el método de evaluación del desempeño y permitieron a los gerentes de sucursales identificar claramente las fortalezas y debilidades de sus operaciones. Finalmente, Aggelopoulos et al. (2013) utilizaron dimensiones de producción y beneficio para estudiar los bancos griegos, considerando los efectos producidos en las sucursales bancarias antes y durante el período de crisis.

También se han utilizado métodos DEA para evaluar la eficiencia de costes de las sucursales bancarias. Por ejemplo, Giokas (2008b), Camanho y Dyson (2005a, b) y Tsolas y Giokas (2012), entre otros, han evaluado la capacidad de una sucursal para producir output a costes mínimos, mientras que Camanho y Dyson (2005a) estimaron la límites superior e inferior de las medidas de eficiencia de costes de las sucursales en diversos escenarios de incertidumbre de precios.

Finalmente, se han empleado métodos DEA para evaluar el crecimiento de la productividad, según el índice de Malmquist. Así, Camanho y Dyson (2006) evaluaron el desempeño de 144 sucursales portuguesas para determinar el margen de eficiencia

dentro del grupo, las eficiencias gerenciales internas y la productividad de frontera, reflejando el impacto de los factores ambientales y de las políticas gerenciales regionales. También a este respecto, Asmild y Tam (2007) calcularon las diferencias de frontera entre las sucursales bancarias de diferentes países.

Otro método no paramétrico, FDH, ha sido utilizado para investigar la eficiencia de las sucursales bancarias, por Respaut (1989) y Tulkens (1993), mientras que uno de los pocos estudios para considerar medidas de productividad basadas en FDH fue realizado por Tulkens y Malnero (1996).

Por otro lado, los métodos estocásticos paramétricos rara vez se han aplicado al análisis de la eficiencia de las sucursales bancarias. Hasta donde sabemos, solo se han empleado modelos de frontera de costes estocásticos (SCF) en estructuras de datos de panel. Por ejemplo, han sido empleados por Berger et al. (1997) para bancos estadounidenses; por Osiewalski y Marzec (1998) y Marzec y Osiewalski (2008) para bancos polacos en un contexto bayesiano y, más recientemente, por Cabrera-Suárez y Pérez-Rodríguez (2020) para un gran banco comercial español. Los últimos autores propusieron el análisis conjunto de las eficiencias de costes variables en el tiempo y la heterogeneidad no observada dentro de las sucursales bancarias.

La mayoría de los estudios anteriores han analizado la eficiencia, pero no sus determinantes. Para evaluar los efectos de ciertas variables ambientales sobre la eficiencia en un entorno no paramétrico, los estudios empíricos suelen adoptar un enfoque de dos etapas. Primero, se realiza un DEA con inputs y outputs, después de lo cual, las puntuaciones de eficiencia se relacionan contra las variables ambientales. Varios estudios han investigado los factores que influyen en la eficiencia de las sucursales bancarias. En su revisión, Paradi y Zhu (2013, p.65) clasificaron estos estudios en las siguientes categorías: comparación de sucursales entre países,

comparación entre bancos, comparación de sucursales entre regiones y análisis del impacto de algunos factores exógenos, como innovaciones tecnológicas.<sup>29</sup>

Por ejemplo, los efectos producidos en la eficiencia de las sucursales por diferencias en las características regionales han sido analizados por Das et al. (2009) y Deville (2009). En el primero de estos estudios, se midió la ineficiencia en el uso de mano de obra a nivel de sucursal de un solo banco, en cuatro ciudades importantes de la India. En cuanto al efecto de la innovación tecnológica, Cook et al. (2004) analizaron el efecto de las actividades de comercio electrónico en el desempeño bancario. En otro enfoque, Fethi y Paisouras (2010) consideraron la diversidad de entornos en los que operan las sucursales, mientras que Paradi y Schaffnit (2004) examinaron el papel de los parámetros ambientales que están más allá del control de la gestión, con respecto a las sucursales comerciales de un gran banco canadiense, e incorporó el riesgo y la tasa de crecimiento económico de la región como factores no discrecionales.

En España, se han realizado pocos estudios sobre la eficiencia de las redes de sucursales minoristas, ya sea empleando DEA u otros métodos no paramétricos. De hecho, hasta donde sabemos, sólo ha existido el artículo de Lovell y Pastor (1997), citado en la introducción de este Capítulo. Estos autores utilizaron un marco DEA y rendimientos variables a escala, con un único input constante o, de manera equivalente, sin inputs. El problema abordado en este caso fue maximizar la expansión radial sobre el vector de 17 indicadores de éxito para las 545 sucursales bancarias evaluadas.

---

<sup>29</sup> Los mismos autores (p.69) indicaron que, desde el punto de vista de la gestión, la literatura ha informado de dos tipos principales de factores ambientales: los específicos de la región (crecimiento económico local, tasa de desempleo local, tipos de comunidad local, diferentes horarios de apertura, entre otros) y las variables específicas de la empresa (diferentes áreas de especialización, como préstamos, hipotecas, depósitos, comercio de oro y divisas).

En resumen, la evidencia empírica actual sobre las diversas dimensiones de la eficiencia de las sucursales indica que el DEA no paramétrico es el enfoque más utilizado y que se han realizado pocas investigaciones para analizar la eficiencia de las mismas en el sistema bancario español. Por lo tanto, nuestro estudio de la eficiencia es adecuado para proporcionar una ilustración empírica de los efectos de los factores ambientales en las puntuaciones de eficiencia de las sucursales, utilizando métodos no paramétricos y no paramétricos robustos.

### **7.3. ESTIMADORES CLÁSICOS Y NO-PARAMÉTRICOS ROBUSTOS**

En este Capítulo se emplean los enfoques no-paramétricos y no-paramétricos robustos para la estimación de la frontera, en concreto, el método no paramétrico DEA y FDH, junto con el estimador no-paramétrico de orden- $\alpha$ , que es más robusto a la existencia de valores atípicos. Además, se utilizarán modelos orientados a los inputs porque la cantidad de negocios disponibles para una sucursal depende en gran medida de la demanda de los clientes por los servicios del banco, lo que está más allá del control del gerente de la sucursal (Paradi et al., 2011).<sup>30</sup>

Sean  $n$  observaciones de las sucursales bancarias (o, en general, unidades de toma de decisiones, DMUs),  $j = 1, 2, \dots, n$ . Sea  $x_{ij}$  el nivel observado del  $i$ -ésimo input en DMU  $j$  y sea  $x_j = (x_{1j}, \dots, x_{kj}) \in \mathfrak{R}_+^k$  el vector de  $k$  inputs para la  $j$ -ésima unidad;  $y_{rj}$  es el nivel observado del  $r$ -ésimo output en DMU  $j$ , donde  $y_j = (y_{1j}, \dots, y_{qj}) \in \mathfrak{R}_+^q$  es un vector de  $q$  outputs.

---

<sup>30</sup> Cabe destacar que, bajo rendimientos constantes a escala, los modelos orientados a inputs y a outputs son equivalentes. Sin embargo, bajo rendimientos variables a escala, son conceptos diferentes, y pueden llevar a determinar medidas de eficiencia diferentes para las sucursales ineficientes.

### 7.3.1. FUNCIONES DE FRONTERA COMPLETA ESTIMADAS POR MÉTODOS NO PARAMÉTRICOS

Los principales estimadores no paramétricos utilizados son el análisis envolvente de datos (DEA) y el análisis de libre disposición (FDH).

El estimador DEA del conjunto de producción, propuesto por primera vez por Farrell (1957) y luego, formalizado como un estimador de programación lineal por Charnes et al. (1978, CCR), asume la libre disponibilidad, la convexidad del conjunto de producción y rendimientos constantes a escala (CRS). En el caso de rendimientos variables a escala (VRS; Banker et al., 1984; o modelo BCC), la puntuación de eficiencia del input DEA estimada para una sucursal que opera al nivel  $(x_0, y_0)$  se obtiene resolviendo el siguiente programa lineal (omitimos los subíndices  $i$  y  $r$  por simplicidad):

$$\hat{\theta}_{j,DEA-VRS}(x_0, y_0) = \min \left\{ \theta / y_0 \leq \sum_{j=1}^n \lambda_j y_j; \theta x_0 \geq \sum_{j=1}^n \lambda_j x_j; \sum_{j=1}^n \lambda_j = 1; \lambda_j \geq 0; j = 1, \dots, n \right\}$$

donde  $\lambda_j$  representa los pesos óptimos para la unidad  $j$ ,  $\hat{\theta}_{j,DEA-VRS}$  está entre 0 and 1, y

$\hat{\theta}_{j,DEA-VRS} = 1$  si la unidad es totalmente eficiente y  $\hat{\theta}_{j,DEA-VRS} < 1$  si es ineficiente.

La eficiencia de escala de las sucursales bancarias se puede calcular utilizando la relación entre la eficiencia técnica de CCR (o eficiencia técnica global) y la eficiencia técnica de BCC (o eficiencia técnica pura), que identifica la proporción de ineficiencia debida exclusivamente a problemas de tamaño. Sin embargo, una deficiencia de esta medida de eficiencia de escala es que el valor no indica si la DMU está operando en un área de rendimientos a escala crecientes o decrecientes. Esta pregunta puede resolverse

ejecutando un problema DEA adicional con rendimientos a escala no crecientes (NIRS)

sustituyendo la restricción  $\sum_{j=1}^n \lambda_j = 1$  de convexidad en el modelo BCC por  $\sum_{j=1}^n \lambda_j \leq 1$ .<sup>31</sup>

El estimador FDH, propuesto por Deprins et al. (1984), es una versión más general del estimador DEA, que se basa únicamente en el supuesto de libre disponibilidad (*free disposal*) y, por lo tanto, no está restringido a tecnologías convexas.<sup>32</sup>

En la práctica, el estimador FDH se calcula mediante un procedimiento simple de comparación de vectores que equivale a un algoritmo de enumeración completo, como se propone en Tulkens (1993). Suponiendo la orientación de input, la eficiencia se puede calcular comparando cada DMU ( $s = 1, \dots, n$ ), y con respecto a todas las demás DMUs ( $j = 1, \dots, n$ ) en los datos que producen al menos la misma cantidad de cualquier output como DMU  $s$ . El conjunto de DMUs pares en la muestra que satisfacen la condición  $y_{ij} \geq y_{i0}, \forall i$  se denota como  $B_j$ . Entre las DMU pares, la que exhibe un consumo mínimo de inputs sirve como referencia a  $s$ , y el estimador FDH  $\hat{\theta}_{j,FDH} \leq 1$ , se calcula como el uso relativo de inputs, de manera que:

$$\hat{\theta}_{j,FDH}(x_0, y_0) = \min_{j \in B_j} \left\{ \max_{i=1, \dots, k} \left( \frac{x_{ij}}{x_{i0}} \right) \right\}$$

Este es un procedimiento máximo (para el marco "orientado al input"): la parte "max" del algoritmo identifica las DMUs más dominantes en relación a una

<sup>31</sup> Si son considerados rendimientos no-decrecientes a escala (NDRS), la restricción de convexidad es reemplazada por  $\sum_{j=1}^n \lambda_j \geq 1$ .

<sup>32</sup> Esto parece una propiedad atractiva del FDH ya que es frecuentemente difícil encontrar una buena justificación teórica y empírica para postular conjuntos de producción no convexos en el análisis de eficiencia. FDH, es a veces, económicamente más significativo que el caso monótono convexo, bajo determinadas condiciones económicas. Por lo tanto, las medidas de eficiencia técnica FDH tienen sentido para las teorías de la empresa que permiten una competencia imperfecta o incertidumbre.

determinada DMU evaluada. Una vez que se identifican las DMUs más dominantes, las holguras se calculan a partir de la parte "mínima" del algoritmo.

### **7.3.2. FUNCIONES DE FRONTERA PARCIAL ESTIMADAS POR MÉTODOS NO-PARAMÉTRICOS ROBUSTOS**

Uno de los principales inconvenientes de los modelos de frontera no-paramétricos (basados en DEA / FDH) es que son muy sensibles a la existencia de valores atípicos y errores de medición (es decir, la influencia de observaciones "supereficientes"). Esto es una consecuencia del hecho de que la frontera eficiente está determinada por observaciones muestrales que son puntos extremos. Simar (1996) señaló la necesidad de identificar y eliminar valores atípicos cuando se utilizan modelos no-paramétricos. Si estos valores atípicos no se pueden identificar, se recomienda el uso de modelos de frontera estocástica.

Como alternativa, los estudios han identificado métodos más robustos basados en estimaciones de eficiencia no-paramétricas, como los enfoques de frontera parcial.

La frontera parcial envuelve solo una submuestra de los datos y se obtiene a partir de una definición probabilística del conjunto de tecnología. Los enfoques *order-m* (Cazals et al., 2002) y *order- $\alpha$*  (Aragon et al., 2005) nos permiten estimar las eficiencias de las fronteras parciales no-paramétricas.

*Order-m* generaliza FDH agregando una capa de aleatoriedad al cálculo de las puntuaciones de eficiencia. En lugar de comparar una DMU por el par con mejor desempeño en la muestra en cuestión, *order-m* se basa en la idea de comparar la DMU por el mejor desempeño esperado en una muestra de  $m$  pares (ver Daraio y Simar (2007)

para el paso del procedimiento en términos computacionales). También Order- $\alpha$  generaliza FDH, pero en lugar de utilizar el consumo mínimo de inputs entre los pares disponibles como un punto de referencia, se utiliza el percentil  $(100-\alpha)$ -ésimo.

En este capítulo, comentamos brevemente sobre el estimador de *order- $\alpha$* , que se prefiere porque no requiere un nuevo muestreo y, por lo tanto, consume menos tiempo que el *order- $m$* .<sup>33</sup>

El percentil  $(100-\alpha)$ -ésimo con  $0 \leq \alpha \leq 100$  puede ser calculado como:

$$\hat{\theta}_{j,\alpha}(x_0, y_0) = P_{j \in B_j}^{100-\alpha} \left\{ \max_{i=1, \dots, k} \left( \frac{x_{ij}}{x_0} \right) \right\}$$

El modelo FDH es un caso particular del estimador de *order- $\alpha$*  cuando  $\alpha=100$ , que abarca todas las observaciones. Cuando  $\alpha < 100$ , algunas DMUs pueden clasificarse como supereficientes y no están envueltas por la frontera de posibilidades de producción estimada.  $\alpha$  se puede considerar como un parámetro de ajuste que determina el número de DMUs supereficientes. En este caso,  $\hat{\theta}_{j,\alpha} > 1$ . Las unidades eficientes tienen  $\hat{\theta}_{j,\alpha} = 1$  y las ineficientes,  $\hat{\theta}_{j,\alpha} < 1$ . La elección empírica de  $\alpha$  suele encontrarse dentro del intervalo  $[90, 99]$ .

---

<sup>33</sup> La principal diferencia con *order- $m$*  es que *order- $\alpha$*  no requiere remuestreo y así, el proceso de estimación es más rápido.



### 7.3.3. DETECCIÓN DE VALORES ATÍPICOS BASADOS EN FRONTERAS

#### PARCIALES

El análisis de frontera parcial se puede utilizar para detectar posibles valores atípicos en los datos, después de lo cual se podría realizar un análisis de eficiencia no-paramétrico usando DEA o FDH (ver Daraio y Simar, 2007).

Para este análisis, se deben obtener los puntos de discontinuidad en presencia de valores atípicos. Por lo tanto, las DMUs clasificadas como supereficientes para  $\alpha \geq \alpha^{disc}$  (punto de discontinuidad) son los valores atípicos más probables. Estos valores atípicos se pueden excluir del análisis de eficiencia aplicando FDH o DEA.

El enfoque sugerido por Daraio y Simar (2007) tiene los siguientes pasos: 1) realizar una serie de análisis de frontera parcial para diferentes valores de  $\alpha$ ; 2) graficar la proporción de DMUs supereficientes frente a  $\alpha$ ; 3) identificar discontinuidades en la curva resultante. Estas discontinuidades revelan las DMUs atípicas que se clasifican como supereficientes para los valores correspondientes de  $\alpha$ . Daraio y Simar (2007) sugirieron una regla global y dos locales para detectar discontinuidades: a)  $\alpha$  para la cual la serie doblemente diferenciada tiene un valor mínimo (después de una no negativa); b) valores de  $\alpha$  para los que persisten valores negativos después de suavizar repetidamente series de dos diferencias mediante la ejecución de suavizadores de mediana con espaciado impar (*odd-spaced*); y c)  $\alpha$  que minimiza BIC para dividir la serie en dos partes y ajustar una función lineal (cuadrática) a cada una.

### **7.3.4. PROCEDIMIENTO DE SIMAR Y WILSON**

Los efectos de las covariables (variables exógenas) sobre la eficiencia se pueden evaluar mediante un modelo de regresión lineal, que se escribe de la siguiente manera:

$$Eficiencia_j = \beta'z_j + e_j, j = 1, 2, \dots, n$$

donde  $Eficiencia_j$  es la puntuación de eficiencia obtenida previamente;  $z_j$  es un vector columna de variables explicativas (incluyendo el uno para el término constante),  $\beta'$  es un vector fila de parámetros desconocidos a estimar, y,  $e_i$  es un término error con desviación estándar igual a  $\sigma_e$ .

$\beta'$  y  $\sigma_e$  fueron estimados por el procedimiento Simar y Wilson (2007) basado en la regresión truncada usando máxima (algoritmo #1, páginas 110-111). Este enfoque excluye las DMUs eficientes y estima los modelos basados en ineficiencias (valores inferiores a 1). Nuestro análisis se basa en una orientación input, con truncamiento a la derecha en 1, como en el método de Simar y Wilson. Se realiza el siguiente procedimiento iterativo de múltiples pasos: 1) regresión truncada; 2) muestreo repetido de pseudo-perturbaciones de la distribución normal truncada, generando puntuaciones de eficiencia artificial y reestimando el modelo; 3) cálculo de distribuciones simuladas tanto para  $\beta'$  como para  $\sigma_e$ .

## 7.4. ANÁLISIS EMPÍRICO

### 7.4.1. DATOS

El análisis empírico que se presenta en este Capítulo se basa en la banca minorista pura, que como hemos mencionado anteriormente, es un elemento central del sistema bancario español.

Debido a la disponibilidad limitada de datos y las dificultades derivadas de las obligaciones de confidencialidad de los bancos, este estudio se centra en un gran banco comercial español anónimo, que opera a través de una red regional de oficinas bancarias. El banco en cuestión es representativo para nuestros propósitos, como uno de los tres bancos más grandes del país en términos de activos, dado el alto grado de concentración de la banca nacional en este país. Este banco no se vio afectado por el rescate,<sup>34</sup> aunque el escenario macroeconómico afectó levemente su productividad laboral interna y eficiencia operativa. Este resultado favorable puede haberse logrado porque el banco tiene un modelo de cartera diversificado y equilibrado, y porque su modelo de gestión se basa en políticas de prudencia y la anticipación.

Nuestro estudio se basa en datos anuales de las sucursales bancarias para el período 2011-2014, que incluyeron la segunda recesión económica española (2011-2013) y el rescate del sistema bancario español (2012) a través del paquete de rescate de 100 billones de euros proporcionado por el Mecanismo Europeo de Estabilidad (MEDE). En

---

<sup>34</sup> Se trata de uno de los bancos fuertes de nuestro país, que están bien gestionados, suficientemente capitalizados y aparentemente resistentes a nuevos choques. Otras entidades, más vulnerables, con una gran proporción de sus balances vinculados al sector inmobiliario requieren una mayor reestructuración y recapitalización.

este período, la rentabilidad de los bancos disminuyó y era una práctica común que la gerencia buscara la máxima eficiencia en las operaciones bancarias.

La información contable anual se recopiló a partir de los estados de pérdidas y ganancias, junto con los datos de los departamentos internos (como Control de Gestión, Ingeniería y Recursos Humanos), correspondientes a 134 sucursales que gestionaban cuentas tanto de clientes particulares como negocios, y brindaban servicios similares en cada caso (sus actividades fueron homogéneas). Las sucursales estaban ampliamente distribuidas en una región determinada de España, en áreas rurales y urbanas, algunas de las cuales tenían una alta concentración de sucursales. Por razones de confidencialidad, no se proporciona ninguna identificación adicional.

#### **7.4.2. MEDICIÓN DEL DESEMPEÑO DE LAS SUCURSALES: DIMENSIONES, INPUTS Y OUTPUTS**

En este estudio, se utilizaron enfoques de producción y ganancias para medir el desempeño de las sucursales bancarias, en términos de incrementar las ventas y administrar la mezcla de productos o servicios para generar una alta rentabilidad (Portela y Thanassoulis, 2007). La naturaleza y la cantidad de variables de input-output utilizadas en nuestra evaluación se seleccionaron después de considerar la disponibilidad de datos. En resumen, se emplearon los siguientes enfoques de producción y beneficio.

El enfoque de producción enfatiza las actividades operativas; en otras palabras, las sucursales bancarias se consideran proveedores de servicios a los clientes. En nuestro análisis de estas actividades, los inputs utilizados para cada sucursal fueron el número total de empleados (incluidos los gerentes de sucursales, personal especializado,

administrativo, gestores comerciales y cajeros), el número de ATMs (cajeros automáticos), el espacio del local (en m<sup>2</sup>) y, los costes operativos soportados (coste total de materiales, espacio y sistemas de información, pero excluidos los costes de personal). Estos inputs también han sido utilizados por Camanho y Dyson (1999, 2006). Los productos u outputs considerados fueron el valor total de los préstamos (como hipotecas y préstamos personales a consumidores y empresas) y, el valor total de los ahorros (depósitos en cuentas). Estos resultados, también fueron utilizados por Camanho y Dyson (1999, 2006) y, por Giokas (2008a). Sin embargo, nuestro estudio no incluye la cantidad de actividad (medida por el número total de transacciones), debido a que la información está incompleta (los datos solo estaban disponibles desde 2013).

La medida de eficiencia de beneficios refleja el nivel más eficiente de costes generales para generar rendimientos, analizando cómo cada sucursal usa sus recursos (inputs) para generar ingresos tales como ingresos por intereses e ingresos por comisiones. En el presente estudio, nos centramos en el estudio de la eficiencia del uso de los gastos de personal, los gastos operativos y los gastos por intereses para generar ingresos por intereses y comisiones (Oral y Yolalan, 1990). Los inputs utilizados son los gastos de personal, los costes operativos y los gastos por intereses, y la variable output es el margen bruto obtenido por la sucursal, es decir, los ingresos totales menos los ingresos netos por intereses más los ingresos por comisiones (no intereses).<sup>35</sup>

Todos los datos de valor están en euros constantes de 2011 (es decir, utilizamos el índice de precios al consumidor del mercado local con respecto al período base 2011).

---

<sup>35</sup> El margen bruto se compone de la suma de OPOs y el margen de intereses. OPOs se definen como las comisiones cobradas por el banco para la prestación de todo tipo de servicios (comisiones totales + diferencias de cambio + resto de transacciones financieras). El margen de interés es la diferencia entre el interés cobrado por el crédito y las inversiones financieras realizadas (Activo) menos el interés pagado a los clientes y otros prestamistas por el dinero depositado en el banco (Pasivo).

Las variables input-output que utilizamos han sido identificadas en muchos estudios de DEA que se centran en las sucursales bancarias (ver Camanho y Dyson, 1999, 2006; y Paradi y Zhu, 2013, entre otros). Sin embargo, Fethi y Pasiouras (2010) concluyeron que, aunque existe un acuerdo general sobre las principales categorías de inputs y outputs, hay poca coherencia con respecto a los inputs/outputs específicos que deban utilizarse.

Como se indica en el apartado 7.3, en este Capítulo se utiliza tecnología con orientación hacia los inputs para la evaluación general de la eficiencia.

### **7.4.3. FACTORES AMBIENTALES**

Se incluyeron varios factores ambientales en el análisis para analizar su efecto sobre las eficiencias operativas y de beneficios. Siguiendo la literatura empírica sobre sucursales bancarias citada en el apartado 7.2 (pero también en mayor detalle en el Capítulo 5), consideramos dos grupos de factores; uno relacionado con las métricas de desempeño propias del banco y el otro relacionado con otras características específicas de las sucursales, dependiendo de la disponibilidad de datos.

#### ***Métricas de desempeño propias del Banco***

El banco examinado utiliza varias medidas para medir el desempeño de las sucursales, teniendo en cuenta la presión competitiva y el estricto control que se ejerce al respecto en el sector bancario español. Para el presente análisis, nos centramos en las dos medidas más significativas del desempeño de la métrica bancaria para cada sucursal. Estas serían, por un lado, la magnitud de las desviaciones del margen bruto (o ingreso total) de los objetivos (calculado como el valor absoluto de las desviaciones de

los objetivos del margen bruto), y, por otro lado, la productividad basada en valor (PBV) en términos porcentuales.

En los últimos años, las funciones de la red de sucursales se han valorado calculando los ingresos netos por intereses y los ingresos sin intereses (comisiones) generados por cada sucursal (DMU). El objetivo del control presupuestario es detectar y analizar desviaciones de los objetivos financieros, como parte esencial de la planificación empresarial. Una forma de hacerlo es analizar la diferencia entre el ingreso total esperado y el real. Esta desviación generalmente se calcula tanto en valor absoluto como en porcentaje.

El indicador PBV mide la actividad comercial en términos del valor presente neto generado por una operación, tomando en cuenta variables como el volumen de la operación, su precio, plazo y riesgo, teniendo en cuenta que una transacción contratada en un año determinado generará ingresos en años subsecuentes. La PBV se mide trimestralmente a partir de los informes de ventas correspondientes. Estos contienen tres secciones (recursos, inversión y productos) y, se monitorean no solo a nivel de unidad sino también para cada gerente de unidad con responsabilidad de ventas. La PBV se calcula como un índice y, por lo tanto, se expresa en unidades porcentuales. Otros métodos, se basan en el balance y en los estados de pérdidas y ganancias de cada unidad.

### ***Otras características de la sucursal bancaria***

Se incluyeron varias variables como otros factores que pueden explicar la eficiencia de la sucursal bancaria:<sup>36</sup> primero, la experiencia (o edad) de la sucursal como un

---

<sup>36</sup> Estos factores son principalmente factores específicos de la región, específicos de la empresa y

indicador de su capacidad (factor no observado) para permanecer en el mercado. Esta variable nos permite evaluar el aprendizaje práctico o curva de aprendizaje, y se expresa como el número de años que la sucursal ha estado operativa.<sup>37</sup>

En segundo lugar, incluimos las dotaciones crediticias por pérdida esperada como una medida de la evaluación del banco del riesgo de incumplimiento crediticio. Desde otra perspectiva, es una medida de la adecuación de la gestión del riesgo de crédito. Por lo tanto, cuanto mayor sea la pérdida esperada de las dotaciones de los saldos de inversión, mayor será el riesgo.<sup>38</sup> En tercer lugar, incluimos dos variables de ubicación: la zona geográfica (zonas o áreas administrativas) en la que se encuentra la sucursal bancaria y una variable ficticia denominada tamaño del mercado, que distingue las sucursales urbanas de las rurales, independientemente de la zona. En cuarto lugar, identificamos el número de empleados de sucursales con certificación de Asesor Financiero Europeo (EFA), como una medida de la calidad de los empleados. En quinto lugar, incluimos la especialización de la sucursal en la venta de productos (productos específicos como hipotecas para particulares o productos para empresas).<sup>39</sup> Finalmente, incluimos una variable ficticia denominada rescate / reforma regulatoria, la cual tomó el valor 1 si las operaciones de la sucursal consideradas eran para el año 2012 y 0, en caso contrario.

---

otros factores específicos de la sucursal más allá del control del gerente.

<sup>37</sup> El número de años que una compañía ha existido puede ser un buen indicador de su experiencia o experiencia acumulada (Bernini and Guizzardi, 2010).

<sup>38</sup> Algunos autores utilizan el deterioro de la pérdida de valor de los préstamos como elemento de un enfoque basado en los beneficios (Giokas, 2008a), en la visión de que este factor podría reducir la eficiencia de los beneficios. Sin embargo, utilizamos la dotación por pérdida esperada de préstamos que está relacionada con una medida de riesgo para la sucursal. Más específicamente, si hay una alta probabilidad de que el préstamo se convierta en moroso, se considerará una mayor dotación que reducirá los beneficios de la sucursal.

<sup>39</sup> Estas categorías son definidas por el propio banco.



La Tabla 7.1 muestra los estadísticos descriptivos para las variables utilizadas. Algunas de estas variables representan cantidades y otras son valores monetarios. La muestra abarca el período 2011-2014. En esta tabla, distinguimos entre inputs y outputs y, los factores ambientales utilizados para explicar las eficiencias estimadas. Los estadísticos descriptivos utilizados, para todas las variables de estudio, fueron el promedio, la desviación estándar y los percentiles 10, 25, 50, 75 y 90 disponibles por cada variable, por año.

En promedio, cada sucursal tenía cuatro empleados y negociaba valores de 11.053 € para depósitos y 31.793 € para préstamos. El coste operativo medio fue de 143.897 euros, la desviación media entre el margen bruto real y el esperado fue de 0,003 y el PBV fue de 1,15. La mayoría de los empleados de las sucursales no tenían certificación EFA, aunque el número de certificaciones ha ido creciendo a lo largo de los años. La mayoría de las sucursales bancarias estaban ubicadas en ciudades y se especializaban en la venta de productos a particulares y, en menor medida, a empresas.

**Tabla 7.1.** Estadísticos descriptivos para todas las sucursales durante el período 2011-2014.

Variables	Media	Desviación Estándar	Percentiles				
			10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>
<b>Panel A. Inputs and outputs</b>							
<i>Inputs</i>							
Espacio del local de la sucursal (in m <sup>2</sup> )	329.69	297.32	130	170	210	340	708
Número de Empleados	4	2	2	2	3	4	8
Número de ATMs externos	3	1	1	2	3	4	4
Costes Operacionales	143,897.3	86,754.08	73,268.23	87,862.14	118,413.1	167,741.2	270,966.8
Gastos de Intereses (coste depósitos)	266,649.7	279,959.9	62,279.67	97,296.21	159,510.2	316,465.3	600,921.9
<i>Outputs</i>							
Valor Total Depósitos (en miles de euros)	11,053.27	11,089.4	2,659.054	4,385.394	6,500.832	13,667.91	23,906.04
Valor Total de Préstamos (en miles de euros)	31,792.67	23,456.92	9,013.103	16,234.38	23,765.23	41,920.97	64,096.21
Margen Bruto (neto ingresos por intereses + ingresos por no intereses (comisiones))	1,467,339	1,139,147	457,946.1	707,393.2	1,113,820	1,792,403	3,054,947
<b>Panel B. Factores Ambientales</b>							
Desviaciones de los objetivos de ingresos totales (margen bruto)	0.003	0.097	-0.096	-0.053	-0.003	0.049	0.106
Productividad basada en valor (PBV)	1.15	0.29	0.82	0.95	1.12	1.31	1.51
Experiencia (Años de la sucursal)	24.96	12.94	10	12	36	38	39
Área Administrativa (7 zonas geográficas)	3.54	1.05	3	3	4	4	4
Número de trabajadores con certificación EFA	0.37	0.61	0	0	0	1	1
Tamaño Mercado (=1, si urbano; =0, resto)	0.84	--	0	1	1	1	2
Dotación Pérdida Esperada	17,118.44	61,557.96	-15,941.41	-295.60	8,584.15	26,478.89	67,897.11
Especialización (=1, business)	0.04	--	0	0	0	0	0

#### 7.4.4. EFICIENCIAS ESTIMADAS

En esta sección, presentamos los resultados de las eficiencias técnicas y de escala obtenidas por la DEA, junto con las estimaciones de eficiencia obtenidas por FDH, el *order- $\alpha$*  no-paramétrico robusto y las eficiencias derivadas de la aplicación del procedimiento de Daraio y Simar (ver Sección 7.3 para una descripción de todos estos métodos). En el último caso, se utilizó un análisis de frontera parcial *order- $\alpha$*  para detectar posibles valores atípicos en los datos, después de lo cual se realizó un análisis de eficiencia clásico no-paramétrico por DEA y FDH. Este procedimiento nos permitió comparar los resultados obtenidos con los métodos clásicos con los diseñados para manejar la cuestión de los valores atípicos.

##### 7.4.4.1. ESTIMACIONES DE VALORES ATÍPICOS BASADOS EN DEA-VRS, FDH, ORDER-ALPHA Y FRONTERA PARCIAL

En esta sección, nos centramos en las estimaciones de eficiencia basadas en rendimientos variables a escala, utilizando DEA, FDH y *order- $\alpha$* , junto con el procedimiento de Daraio y Simar. Los diferentes resultados obtenidos se presentan a modo de comparación. En todos los casos, se utilizó la orientación hacia los inputs. También se estimó el DEA-CRS, pero como existen ineficiencias de escala y en aras de la brevedad, estos resultados no se muestran. En la Tabla 7.2 se muestran los estadísticos descriptivos para las puntuaciones de eficiencia, donde  $\hat{\theta}_{\text{método}}$  denota el estimador de eficiencia media según el método empleado (*método*): DEA-VRS, FDH y orden- $\alpha$ , respectivamente, y donde los valores atípicos basados en la frontera parcial se obtienen por DEA-VRS y FDH. Específicamente, la información relacionada con el

número de observaciones, la media y la desviación estándar, los valores máximo y mínimo, y los percentiles 10, 25, 50, 75 y 90. Una puntuación  $\hat{\theta}_{\text{método}}$  igual a 1 representa la eficiencia total, mientras que un valor inferior a uno representa la ineficiencia. También mostramos el porcentaje de sucursales bancarias totalmente eficientes (*Eff*) y súpereficientes (*SEff*), respectivamente,  $(\hat{\theta}_{\text{order-}\alpha} > 1)$  revelando así la distribución de eficiencias.

En general, las eficiencias obtenidas son asimétricas a la izquierda, y las de los estimadores robustos (*order- $\alpha$* ) son siempre superiores a las de estimadores no-paramétricos clásicos como DEA-VRS y FDH. Por ejemplo, las estimaciones de la eficiencia media se ordenan como  $\hat{\theta}_{\text{DEA-VRS}} < \hat{\theta}_{\text{FDH}} < \hat{\theta}_{\text{order-}\alpha}$ , donde  $\alpha$  (el parámetro de ajuste que determina el número de sucursales súpereficientes) es diferente cada año. En general, estos resultados indican que tener en cuenta los valores atípicos (sucursales súpereficientes) usando el *order- $\alpha$*  proporciona estimaciones de eficiencias más consistentes que lo que es posible con los métodos tradicionales DEA y FDH. Destacar que usando el método *order- $\alpha$* , el valor de  $\alpha$  se calcula como el punto sugerido de discontinuidad, que es obtenido mediante el uso de la detección de valores atípicos basada en *order- $\alpha$*  y usando la sugerencia principal del criterio de suavizado en  $\alpha$ , siguiendo el procedimiento de Daraio y Simar (2007). Los resultados asociados con la proporción de unidades súpereficientes pueden parecer improbables, en torno al 80% (es decir, parecería haber demasiadas DMU más allá de la frontera).<sup>40</sup> Para investigar y explicar estos sorprendentes resultados, realizamos una comparación detallada con otros

---

<sup>40</sup> Paradi y Zhu (2013) señalaron que un gran número de sucursales se encuentran en la frontera eficiente (entre el 25-50%), ya que las sucursales bancarias, en general, tienden a estar muy bien gestionadas, debido a las estrictas políticas impuestas por los bancos centrales y el papel regulador del gobierno.

métodos robustos no-paramétricos como *order-m*. Sin embargo, los resultados obtenidos (no mostrados aquí, en aras de la brevedad) indican la existencia de similitudes entre las estimaciones de *order- $\alpha$*  y *order-m* para los diferentes valores elegidos ( $m = 10, 25, 50, 100$  y  $250$ ). Nuestra conclusión, en este caso, es que los métodos no-paramétricos robustos revelan la existencia de una gran proporción de sucursales muy bien gestionadas en términos de los inputs y outputs utilizados.

En general, la trayectoria temporal de las eficiencias entre 2011 y 2014 es estable según los métodos clásicos de estimación, en ambas dimensiones, aunque hay un ligero aumento en las eficiencias DEA y FDH eliminados los outliers, a pesar del rescate del sistema bancario español en 2012.

Centrándonos en los resultados agrupados, estos indican que la ineficiencia media gerencial de la frontera completa (DEA-VRS) fue de alrededor del 14% durante el período general, para la dimensión operativa, y del 21% para la dimensión de beneficios. El enfoque FDH arrojó valores de 1% y 7%, que son inferiores a los obtenidos por DEA-VRS. Sin embargo, si observamos los valores atípicos basados en la frontera parcial, es interesante notar que, durante el período de estudio, el 80% de las sucursales bancarias consideradas fueron supereficientes en la dimensión de producción, pero solo el 75% fueron supereficientes con respecto al beneficio. Los resultados correspondientes de DEA-VRS y FDH indican que las ineficiencias medias son menores que con los métodos no-paramétricos clásicos. Las puntuaciones obtenidas al respecto fueron del 4% y el 11%, respectivamente, con DEA-VRS, pero solo del 1% en cada caso con FDH.

En resumen, puede decirse que se registraron altos niveles de eficiencia en ambas dimensiones, aunque hubo diferencias según la tecnología empleada (especialmente

entre DEA-VRS y FDH). Cuando se utilizó el procedimiento de Daraio y Simar, basado en un análisis de frontera parcial y teniendo en cuenta los valores atípicos excluidos en los métodos DEA-VRS y FDH, las eficiencias resultantes fueron superiores a las obtenidas por los métodos clásicos en los que no se eliminaron los valores atípicos. Por tanto, podemos concluir que, cuando se permite que las observaciones supereeficientes se ubiquen más allá de la frontera de posibilidades de producción estimada, las eficiencias obtenidas son mayores que con los métodos clásicos, tanto en la dimensión de producción como en la de beneficio. Es importante señalar que estos resultados también indican que los gerentes actúan consecuentemente para alcanzar sus objetivos, en ambos enfoques.

**Tabla 7.2.** Eficiencia por métodos no-paramétricos and no-paramétricos robustos considerando las dimensiones de producción y rentabilidad.

	Frontera Completa										Frontera Parcial					Fronteras completas después de eliminar outliers										
	DEA-VRS					FDH					Order- $\alpha$					DEA-VRS					FDH					
	2011	2012	2013	2014	Pooled	2011	2012	2013	2014	Pooled	2011	2012	2013	2014	Pooled	2011	2012	2013	2014	Pooled	2011	2012	2013	2014	Pooled	
<b>Panel A. Producción.</b>																										
											$\alpha$ =92.59	$\alpha$ =85.29	$\alpha$ =94.12	$\alpha$ =80.74												
<i>N</i>	130	126	128	128	512	134	135	135	134	538	134	135	135	134	538	39	15	40	12	106	38	15	44	11	108	
<i>SEff</i>											0.72	0.89	0.67	0.92	0.80											
<i>Eff</i>	0.25	0.21	0.24	0.25	0.24	0.93	0.94	0.96	0.96	0.95	0.26	0.10	0.19	0.08	0.16	0.10	0.07	0.19	0.07	0.11	0.26	0.10	0.18	0.08	0.16	
<i>Media</i>	0.82	0.88	0.86	0.88	0.86	0.98	0.99	0.99	0.99	0.99	0.96	1.00	0.99	0.92	0.97	0.88	0.94	0.93	0.96	0.92	0.99	0.98	0.99	1.00	0.99	
<i>Sd</i>	0.14	0.12	0.12	0.12	0.13	0.07	0.06	0.05	0.05	0.06	0.16	0.00	0.02	0.29	0.12	0.12	0.10	0.11	0.08	0.11	0.04	0.06	0.04	0.00	0.04	
<i>min</i>	0.50	0.55	0.57	0.56	0.50	0.67	0.67	0.67	0.67	0.67	0.00	1.00	0.88	0.00	0.47	0.67	0.69	0.62	0.74	0.62	0.83	0.83	0.78	1.00	0.78	
<i>10<sup>th</sup></i>	0.65	0.69	0.68	0.69	0.67	1.00	1.00	1.00	1.00	1.00	0.88	1.00	1.00	1.00	0.97	0.69	0.73	0.72	0.87	0.73	1.00	0.83	1.00	1.00	1.00	
<i>25<sup>th</sup></i>	0.71	0.81	0.78	0.81	0.77	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.90	0.88	0.97	0.84	1.00	1.00	1.00	1.00	1.00	
<i>50<sup>th</sup></i>	0.82	0.90	0.89	0.92	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<i>75<sup>th</sup></i>	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<i>90<sup>th</sup></i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<i>max</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<b>Panel B. Beneficio/Rentabilidad.</b>																										
											$\alpha$ =92.59	$\alpha$ =92.65	$\alpha$ =93.38	$\alpha$ =93.33												
<i>N</i>	135	135	135	135	540	134	135	135	134	538	134	135	135	134	538	29	39	36	32	136	28	39	36	31	134	
<i>SEff</i>											0.79	0.71	0.73	0.77	0.75											
<i>Eff</i>	0.10	0.11	0.15	0.15	0.13	0.54	0.46	0.52	0.59	0.53	0.11	0.14	0.14	0.12	0.13	0.04	0.10	0.11	0.07	0.08	0.16	0.22	0.23	0.20	0.20	
<i>Media</i>	0.76	0.78	0.82	0.81	0.79	0.93	0.92	0.94	0.95	0.93	0.92	0.96	0.96	0.93	0.94	0.80	0.92	0.93	0.89	0.89	0.98	0.98	0.99	0.99	0.99	
<i>Sd</i>	0.15	0.13	0.14	0.14	0.14	0.10	0.10	0.10	0.08	0.09	0.19	0.07	0.06	0.18	0.13	0.13	0.09	0.09	0.11	0.12	0.04	0.05	0.03	0.05	0.04	
<i>min</i>	0.47	0.51	0.45	0.50	0.45	0.60	0.61	0.57	0.63	0.57	0.00	0.75	0.74	0.00	0.37	0.57	0.70	0.65	0.70	0.57	0.83	0.75	0.87	0.79	0.75	
<i>10<sup>th</sup></i>	0.57	0.61	0.63	0.63	0.61	0.77	0.76	0.78	0.84	0.79	0.83	0.80	0.87	0.82	0.83	0.62	0.76	0.81	0.71	0.71	0.95	0.90	0.96	0.99	0.95	
<i>25<sup>th</sup></i>	0.67	0.68	0.70	0.71	0.68	0.89	0.86	0.89	0.92	0.89	0.93	0.96	0.95	0.92	0.94	0.70	0.88	0.87	0.80	0.81	1.00	1.00	1.00	1.00	1.00	
<i>50<sup>th</sup></i>	0.74	0.78	0.83	0.81	0.79	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.77	0.93	0.97	0.88	0.92	1.00	1.00	1.00	1.00	1.00	
<i>75<sup>th</sup></i>	0.88	0.89	0.95	0.92	0.91	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<i>90<sup>th</sup></i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<i>max</i>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

**Notas:** Los programas para estimar las eficiencias se hicieron por separado para cada año. Los datos se agruparon para el período 2011-2014. También, se reportan los estadísticos descriptivos para la muestra combinada, es decir, 540 observaciones balanceadas de sucursales bancarias, para los años 2011 a 2014. *SEff* y *Eff* son los porcentajes de sucursales bancarias que se considera totalmente eficientes y supereficientes, respectivamente. Con respecto al orden- $\alpha$ , el valor de  $\alpha$  es el punto sugerido de discontinuidad obtenido mediante el uso de la detección de valores atípicos basado en el orden- $\alpha$  y está relacionado con la sugerencia principal por suavización de criterios en  $\alpha$  basado en el procedimiento de Daraio y Simar (2007). Anotamos que  $\alpha$  es diferente cada año. Para facilitar la comparación con DEA y FDH (que están delimitados entre 0 y 1), la media, la desviación estándar y los percentiles de la distribución de las estimaciones de orden- $\alpha$  se relacionan tanto con unidades totalmente eficientes como ineficientes en las muestras correspondientes.

#### **7.4.4.2. ESTIMACIONES DE EFICIENCIA DE ESCALA**

La evaluación de las puntuaciones de eficiencia de escala es muy importante para la administración del banco. La eficiencia de escala indica el éxito con el que una sucursal está operando a su escala óptima; una sucursal de escala eficiente funciona con el tamaño de escala más productivo. El logro de esta escala óptima contribuye a la eficiencia operativa.

En esta sección, presentamos los resultados obtenidos para las eficiencias de escala de la frontera completa y eliminando los valores atípicos utilizando DEA (Escala-DEA).

La estimación de los modelos DEA-CRS (CCR) y DEA-VRS (BCC) permitió determinar la eficiencia técnica global y la eficiencia técnica pura, respectivamente. La puntuación de eficiencia de escala se calcula dividiendo CCR por BCC. También se aplica el modelo NIRS para evaluar los rendimientos a escala obtenidos por cada sucursal bancaria. Por lo tanto, para una sucursal en particular, podemos evaluar si la puntuación de eficiencia técnica NIRS es igual a la puntuación de eficiencia técnica BCC (VRS). Si son desiguales, entonces existen rendimientos crecientes a escala (IRS) para esa DMU. Si son iguales, hay rendimientos de escala decrecientes (DRS).

Tabla 7.3 muestra una descripción numérica de estos resultados, enumerando el número de observaciones, la media y la desviación estándar, los valores máximo y mínimo y los percentiles 10, 25, 50, 75 y 90 para cada variable, por año. La tabla también muestra el porcentaje de rendimientos constantes a escala (CRS), utilizando la tecnología CRS para estimar la eficiencia técnica global, junto con rendimientos crecientes y decrecientes a escala (IRS y DRS, respectivamente) con respecto a la tecnología NIRS. Los últimos indicadores revelan si las ineficiencias de escala se deben a CRS, IRS o DRS.



**Tabla 7.3.** Estimaciones de eficiencia de escala.

	Frontera Completa					Fronteras después de eliminar outliers					
	2011	2012	2013	2014	Agrupados	2011	2012	2013	2014	Agrupados	
<b>Panel A. Producción.</b>											
<i>N</i>	130	126	128	128	512	39	15	40	12	106	
<i>Eff</i>	0.11	0.12	0.12	0.12	0.12	0.21	0.33	0.25	0.50	0.21	
<b>Media</b>	0.82	0.82	0.82	0.82	0.82	0.81	0.85	0.81	0.88	0.83	
<b>Sd</b>	0.20	0.19	0.18	0.18	0.19	0.23	0.20	0.23	0.29	0.23	
<b>Mín</b>	0.00	0.28	0.26	0.33	0.00	0.00	0.42	0.29	0.00	0.00	
<b>10<sup>th</sup></b>	0.51	0.53	0.52	0.53	0.53	0.46	0.51	0.38	0.71	0.44	
<b>25<sup>th</sup></b>	0.76	0.73	0.75	0.73	0.75	0.66	0.80	0.72	0.93	0.72	
<b>50<sup>th</sup></b>	0.86	0.87	0.87	0.88	0.87	0.91	0.93	0.93	1.00	0.94	
<b>75<sup>th</sup></b>	0.99	0.99	0.98	0.98	0.98	1.00	1.00	1.00	1.00	1.00	
<b>90<sup>th</sup></b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<b>Máx</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<b>Rendimientos a escala (%)</b>											
<b>CRS</b>	0.11	0.12	0.12	0.10	0.11	0.21	0.33	0.25	0.05	0.21	
<b>IRS</b>	0.81	0.81	0.81	0.84	0.82	0.74	0.40	0.67	0.27	0.52	
<b>DRS</b>	0.08	0.07	0.07	0.05	0.07	0.05	0.27	0.08	0.00	0.10	
<b>Panel B. Rentabilidad.</b>											
<i>N</i>	135	135	135	135	540	29	39	36	32	136	
<i>Eff</i>	0.05	0.07	0.03	0.05	0.05	0.17	0.08	0.14	0.13	0.13	
<b>Media</b>	0.93	0.88	0.87	0.89	0.89	0.96	0.83	0.84	0.87	0.87	
<b>Sd</b>	0.09	0.15	0.16	0.13	0.14	0.05	0.14	0.15	0.14	0.13	
<b>mín</b>	0.59	0.31	0.37	0.38	0.31	0.83	0.50	0.57	0.50	0.50	
<b>10<sup>th</sup></b>	0.78	0.67	0.61	0.71	0.69	0.89	0.65	0.65	0.68	0.66	
<b>25<sup>th</sup></b>	0.92	0.82	0.83	0.86	0.86	0.91	0.70	0.70	0.75	0.76	
<b>50<sup>th</sup></b>	0.97	0.94	0.94	0.95	0.95	0.98	0.86	0.89	0.93	0.91	
<b>75<sup>th</sup></b>	0.99	0.99	0.98	0.97	0.99	1.00	0.97	0.99	0.97	0.99	
<b>90<sup>th</sup></b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<b>máx</b>	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
<b>Rendimientos a escala (%)</b>											
<b>CRS</b>	0.05	0.07	0.03	0.05	0.05	0.17	0.08	0.14	0.13	0.13	
<b>IRS</b>	0.71	0.80	0.75	0.72	0.74	0.72	0.90	0.67	0.59	0.72	
<b>DRS</b>	0.24	0.13	0.22	0.23	0.21	0.10	0.03	0.19	0.28	0.15	

*Notas:* Las puntuaciones de eficiencia se estiman por separado para cada año y luego se combinaron los datos para el período 2011-2014. Se reportan los estadísticos descriptivos para cada año y la muestra combinada, es decir, 540 observaciones balanceadas de sucursales bancarias; para los años 2011 a 2014. *Eff* es el porcentaje de sucursales bancarias que se considera plenamente eficiente. La eficiencia de escala también se calcula para el caso de eliminar los outliers. Estos son detectados mediante el uso del estimador *order- $\alpha$*  y según el procedimiento de Daraio y Simar (2007).

En promedio, la trayectoria temporal de las eficiencias de escala es estable a lo largo del tiempo, con solo ligeras diferencias. Centrándonos en los datos agrupados, las estimaciones de la frontera completa y las estimaciones eliminando los valores atípicos son similares en ambas dimensiones, con eficiencias de escala promedio de 82% y 83% para la producción y 89% y 87% para la rentabilidad. Por ejemplo, dados estos

resultados de producción, las sucursales ineficientes de escala podrían reducir su tamaño en un 18% y 17%, respectivamente, sin afectar sus niveles de producción actuales, haciendo más eficiente la práctica interna.

Centrándonos en el enfoque que elimina los valores atípicos para estimar la eficiencia de escala, los resultados combinados muestran que, en la dimensión operativa, aumentar los inputs en una proporción determinada produciría un aumento en la misma proporción (rendimientos constantes a escala, CRS) de los outputs. Por lo tanto, estas sucursales estaban operando a su escala productiva óptima. Sin embargo, las ineficiencias de escala son causadas principalmente por IRS, en la mayoría de las sucursales (alrededor del 52%). Por lo tanto, aumentar los inputs en una determinada proporción aumentaría los outputs en una proporción mayor. Así, estas oficinas deben expandirse para lograr el tamaño óptimo (es decir, aquel en el que hay rendimientos constantes a escala en la relación entre inputs y outputs). DRS es de alrededor del 10%, lo que indica que las oficinas necesitan reducir su tamaño para lograr una escala óptima. En general, estos resultados son similares a los reportados por Camanho y Dyson (1999), pero contrastan con los obtenidos por Paradi y Zhu (2011), para quienes entre el 76-84% de las sucursales ineficientes fueron clasificadas como CRS, mientras que, sólo un 2% fueron IRS. Para estos autores, una disminución en los inputs de las sucursales podría permitirles operar por debajo de su escala óptima.

Finalmente, hay que comentar que se obtuvieron resultados similares utilizando los valores estimados de las eficiencias de escala mediante DEA, una vez se eliminaron los outliers. En este caso, las proporciones de sucursales fueron 13% (CRS), 72% (IRS) y 15% (DRS).

#### **7.4.5. RELACIÓN ENTRE ESTIMACIONES DE EFICIENCIA Y FACTORES DE LA SUCURSAL BANCARIA**

Esta subsección presenta las estimaciones del modelo de regresión lineal utilizando el procedimiento de Simar y Wilson, que permite examinar la relación entre la eficiencia y los factores explicativos a nivel de sucursal (descritos en las secciones 7.4.2 y 7.4.3). Específicamente, se realizaron 300 simulaciones bootstrap para las puntuaciones clásicas de eficiencia obtenidas por DEA-VRS, Escala-DEA y FDH, y también para comparar, las puntuaciones de eficiencia DEA y FDH eliminados los valores atípicos.

Para fines de gestión empresarial, las implicaciones de este estudio se pueden derivar de las Tablas 7.4 y 7.5, que muestran los coeficientes y sus errores estándar para los factores explicativos de la eficiencia técnica y de escala, respectivamente. Los resultados se presentan en dos grupos para cada dimensión (producción y rentabilidad/beneficio). El primer grupo contiene las estimaciones DEA y FDH realizadas sin eliminar las sucursales que presentaran valores atípicos en el método *order- $\alpha$* , y el segundo contiene a las estimaciones DEA y FDH de una muestra donde se han eliminado previamente las DMUs supereficientes. Aunque hay siete distritos geográficos (Tabla 7.1), usamos una variable ficticia, denominada Zona, que es igual a 1 si la sucursal está ubicada en el área administrativa número 3 (donde se encuentra la sede regional del banco), y 0 en caso contrario. Este enfoque permite analizar la influencia de la ubicación de la sede en la eficiencia.

Los resultados que se muestran en estas tablas se pueden resumir de la siguiente manera. Para todos los modelos, el coeficiente de desviación estándar de errores

estimado,  $\sigma_e$ , y el parámetro constante en cada modelo son estadísticamente significativos al 1%.

La Tabla 7.4 revela varios hechos interesantes.<sup>41</sup> En primer lugar, existen más coeficientes estadísticamente significativos para los modelos que usan todas las sucursales que aquellos que han eliminado a las sucursales supereficientes, lo que indica que los valores atípicos influyen en los resultados obtenidos. En segundo lugar, con respecto a los resultados de la muestra que elimina los valores atípicos detectados por *order- $\alpha$* , el coeficiente de la variable que representa el rescate es solo positivo y estadísticamente significativo en DEA-VRS. Por lo tanto, la eficiencia de las sucursales bancarias aumentó en 2012 a pesar del rescate bancario y las reformas regulatorias promulgadas.

En tercer lugar, el coeficiente PBV es positivo y estadísticamente significativo al 5% en DEA-VRS, tanto para el enfoque de producción como para el enfoque de beneficio. Esto significa que la propia métrica de desempeño del banco está relacionada con nuestras eficiencias estimadas y, por lo tanto, es un instrumento válido para que los gerentes de sucursales implementen políticas relacionadas con las dimensiones de producción y de rentabilidad. Sin embargo, en términos de FDH no existe tal relación; sólo el coeficiente de la variable *D* es positivo y estadísticamente significativo al 10% en FDH y en la muestra que elimina los valores atípicos. Finalmente, los resultados para otras características de la sucursal son mixtos y ambiguos. En general, pocos coeficientes son estadísticamente significativos, posiblemente debido al tamaño limitado de la muestra (alrededor de 100 observaciones). Sin embargo, los coeficientes

---

<sup>41</sup> No se muestra en esta tabla la estimación FDH para el caso de eliminar las sucursales supereficientes en el enfoque de producción, ya que la estimación no converge.

de EFA y las dotaciones crediticias por pérdida esperada nunca son estadísticamente significativos, mientras que los de ubicación y especialización son positivos y estadísticamente significativos al 5% en FDH cuando se eliminan los valores atípicos en la dimensión de rentabilidad. El coeficiente de tamaño del mercado (es decir, mercados rurales frente a urbanos) no es estadísticamente significativo, lo que indica niveles similares de eficiencia en ambos tipos de mercado.

Para los modelos clásicos, cuando el tamaño de la muestra es grande y hay más sucursales ineficientes, hay otros factores relevantes que también pueden explicar las ineficiencias, especialmente en el enfoque de rentabilidad. Sin embargo, muchos de los factores ambientales que hemos considerado no son relevantes para explicar las ineficiencias técnicas observadas, si eliminamos las observaciones supereficientes usando el procedimiento de Daraio y Simar con el método *order- $\alpha$* , con la excepción de PBV. Por lo tanto, es interesante hacer mención que la no detección de valores atípicos puede no solo distorsionar los valores de eficiencia obtenidos, sino que también podría afectar los resultados de la regresión.

**Tabla 7.4.** Factores ambientales por dimensión para eficiencias VRS y FDH. Procedimiento de Simar y Wilson.

Variables	Producción			Rentabilidad			
	VRS-DEA		FDH	VRS-DEA		FDH	
	Todas las DMUS	Eliminando DMUs supereficientes	Todas las DMUS	Todas las DMUS	Eliminado DMUs supereficientes	Todas las DMUS	Eliminado DMUs supereficientes
Constante	0.787*** (0.0441)	0.776*** (0.104)	0.806*** (0.144)	0.713*** (0.0384)	0.760** (0.356)	0.988*** (0.102)	1.058*** (0.119)
Rescate/reforma regulatoria	0.0730*** (0.0226)	0.000112 (0.0708)	0.0128 (0.0493)	-0.0134 (0.0177)	0.0975** (0.0406)	-0.0915** (0.0395)	-0.0283 (0.0528)
<i>Métricas de desempeño propias del banco</i>							
D	-0.181 (0.127)	0.450 (0.386)	0.128 (0.328)	-0.0564 (0.101)	-0.415 (0.302)	0.362 (0.271)	1.182** (0.489)
PBV	0.0696* (0.0370)	0.0624** (0.0318)	-0.0388 (0.0775)	0.0536* (0.0289)	0.162** (0.0669)	0.0780** (0.0364)	-0.107 (0.0850)
<i>Otras características de la sucursal</i>							
Experiencia (Año de la sucursal)	0.00298*** (0.000721)	-0.000930 (0.00207)	-0.00114 (0.00249)	0.00189*** (0.000601)	-0.000952 (0.00142)	-0.000130 (0.00124)	-0.00380 (0.00276)
Certificación EFA	0.0101 (0.0179)	0.0590 (0.0380)	0.0568 (0.0571)	0.0145 (0.0135)	-0.000626 (0.0232)	-0.0157 (0.0314)	0.173*** (0.0584)
Tamaño mercado (=1, Urbano)	0.0470* (0.0247)	-0.0152 (0.0777)	0.0521 (0.0502)	-0.0669*** (0.0209)	-0.0858 (0.343)	-0.0806 (0.0622)	-- --
Dotación por pérdida esperada	1.35e-07 (1.46e-07)	6.42e-08 (1.73e-07)	-1.09e-07 (4.08e-07)	1.88e-07 (1.52e-07)	-1.34e-07 (2.15e-07)	-3.10e-07 (3.21e-07)	-1.60e-08 (1.80e-07)
Zona (=1, Sede Regional)	-0.00649 (0.0194)	-0.0101 (0.0431)	0.112* (0.0614)	0.0264 (0.0171)	0.0263 (0.0321)	0.0273 (0.0307)	0.324*** (0.0890)
Especialización (=1, Negocios)	-0.0174 (0.0405)	0.0281 (0.0664)	0.208** (0.0835)	0.0457 (0.0355)	-0.0468 (0.0505)	0.0319 (0.0945)	-0.233*** (0.0801)
$\sigma_e$	0.134*** (0.00799)	0.102*** (0.0129)	0.0905*** (0.0130)	0.133*** (0.00634)	0.111*** (0.0110)	0.146*** (0.0167)	0.0624*** (0.0112)
Observaciones	465	102	489	490	127	489	126

**Notas:** Errores estándar en paréntesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. FDH basado en valores atípicos no está incluido porque la estimación no converge.

**Tabla 7.5.** Eficiencias de escala (Escala-DEA) y factores ambientales por dimensión.  
Procedimiento de Simar y Wilson.

Variables	Producción		Rentabilidad	
	Todas las DMUS	Eliminado DMUs supereficientes	Todas las DMUS	Eliminado DMUs supereficientes
Constante	0.610*** (0.171)	0.242 (0.254)	0.712*** (0.226)	0.665** (0.266)
Rescate/reforma regulatoria	0.0401 (0.0895)	-0.0679 (0.144)	-0.167 (0.144)	-0.133 (0.0902)
<b>Métricas de desempeño propias del banco</b>				
D	-0.215 (0.322)	-0.0901 (0.686)	0.279 (0.496)	0.815 (0.684)
PBV	-0.0178 (0.121)	0.100 (0.179)	-0.249 (0.189)	-0.199 (0.156)
<b>Otras características sucursal</b>				
Experiencia (Año de la sucursal)	0.0100*** (0.00377)	0.0116*** (0.00408)	0.0381** (0.0159)	0.0105*** (0.00393)
Certificación EFA	0.0253 (0.0645)	0.107 (0.0884)	0.0519 (0.103)	0.191** (0.0780)
Tamaño mercado (=1, Urbano)	0.522*** (0.106)	0.508*** (0.151)	0.961*** (0.326)	0.283 (0.194)
Dotación por pérdida esperada	1.38e-06* (7.85e-07)	-1.16e-07 (6.66e-07)	-2.15e-06 (1.50e-06)	-1.25e-06* (6.72e-07)
Zona (=1, Sede Regional)	-0.0626 (0.0821)	0.0388*** (0.0126)	0.0171 (0.136)	-0.0324 (0.0780)
Especialización (=1, Negocios)	3.948*** (1.527)	1.812* (1.018)	2.945* (1.624)	0.967** (0.449)
$\sigma_e$	0.334*** (0.0429)	0.250*** (0.0394)	0.334*** (0.0694)	0.208*** (0.0348)
Observaciones	465	101	490	127

*Notas:* Errores estándar en paréntesis. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Centrándonos en los resultados de la Tabla 7.5 (eficiencias de escala), observamos que muchos coeficientes son estadísticamente significativos para varias características de la sucursal, mientras que, los coeficientes del desempeño del propio banco y del rescate / reforma regulatoria no son estadísticamente significativos, en cualquiera de los modelos considerados. Por lo tanto, y a diferencia de la situación reflejada en la Tabla 7.4, las propias métricas del banco no afectaron la eficiencia de la escala en nuestro período de muestra. También se destaca que los modelos estimados usando toda la muestra de sucursales y la que elimina las sucursales supereficientes arrojan resultados

diferentes, aunque existen algunas similitudes en cuanto a algunas características de la sucursal, que presentan parámetros con signos similares y que son estadísticamente significativos.

Por ejemplo, los coeficientes de experiencia y especialización son siempre positivos y estadísticamente significativos para todos los modelos, y el del tamaño del mercado es estadísticamente significativo y positivo para tres modelos. Por lo tanto, hemos encontrado que las puntuaciones de eficiencia del banco son diferentes en las áreas urbanas y rurales. El coeficiente de la dotación por pérdida esperada es negativo y estadísticamente significativo para dos modelos, mientras que el de la certificación EFA es estadísticamente significativo para un modelo de rentabilidad. En conjunto, estos hallazgos indican que la eficiencia aumenta a medida que aumentan estos coeficientes, con la excepción de la pérdida esperada; en el último caso, un aumento de las provisiones crediticias se asocia a una disminución de la eficiencia.





## EL GRADO DE CUMPLIMIENTO DE LAS OFICINAS Y LA EFICIENCIA DE COSTES. UNA APROXIMACIÓN DE FRONTERAS ESTOCÁSTICAS<sup>(42)</sup>

---

### 8.1. INTRODUCCIÓN

El análisis de la eficiencia de las sucursales bancarias es importante porque ayuda a definir y resolver problemas de políticas que permiten a los gerentes de dichas sucursales identificar claramente las fortalezas y debilidades de sus operaciones.<sup>43</sup> Recordamos que la red de sucursales suele ser la mayor fuente de gastos operativos para un banco, y es a través de su sucursal que los clientes realizan una gran proporción de su banca de mayor valor agregado (Paradi y Zhu, 2013).

Los bancos utilizan varios métodos para analizar el desempeño de sus sucursales, buscando lograr una imagen general del desempeño de la red y mejorar la eficiencia mediante la racionalización de recursos, la mejora de las operaciones y la reducción de costes.

---

<sup>42</sup> Este Capítulo es una versión en castellano del artículo titulado: Bank branch performance and cost efficiency: A stochastic frontier panel data approach, escrito por Idaira Cabrera-Suárez y Jorge V. Pérez-Rodríguez, y publicado en *International Journal of Finance and Economics*, 2021, 26 (4), pp. 5850 - 5863.

<sup>43</sup> Berger et al. (1997) indicaron que el análisis de sucursales bancarias puede identificar desviaciones de las mejores prácticas para cualquier escala o producto, mejorar nuestra comprensión de los orígenes de la eficiencia y ayudar a resolver ciertos problemas de medición. El análisis a nivel de sucursal es más deseable e importante que a nivel institucional porque los datos de sucursal dentro de un determinado banco, generalmente, se recopilan de manera uniforme y comparable, por lo que no hay serios problemas con respecto a la interpretación de la eficiencia de escala.

La medición de la eficiencia puede complementar otras medidas del desempeño de las sucursales. Así, Paradi y Zhu (2013) indicaron que el desempeño debe ser validado, ya sea por referencia a otros métodos o simplemente en términos de la propia experiencia o expectativas de los gerentes. Las puntuaciones de eficiencia alcanzadas por las sucursales bancarias serán aceptables para la gerencia si están en línea con las medidas de desempeño general evaluadas para el banco.

La mayor parte de la literatura empírica dirigida a validar la eficiencia obtenida por las sucursales bancarias se ha centrado en métodos deterministas no paramétricos, por ejemplo, comparando los resultados del análisis envolvente de datos (DEA) con los producidos por las medidas de desempeño del propio banco (Lovell y Pastor, 1997; Camanho y Dyson, 1999; entre otros), o sobre métodos no paramétricos robustos (Cabrera-Suárez y Pérez-Rodríguez, 2020), o sobre comparaciones de técnicas como DEA y la frontera determinista log-lineal (Giokas, 1991). Sin embargo, hasta donde sabemos, no se han realizado estudios para analizar la relación entre el desempeño de las sucursales bancarias y la eficiencia en un marco de modelos de frontera estocástica.

En consecuencia, en el presente capítulo proponemos una función frontera de costes estocástica para las sucursales a través de los modelos de datos de panel, teniendo en cuenta que, desde el punto de vista gerencial, los costes se manejan de manera más eficiente a nivel de sucursal.<sup>44</sup> En este escenario, tomamos en cuenta varias cuestiones que no han sido consideradas previamente en la literatura sobre el desempeño de las sucursales bancarias. En primer lugar, modelamos las ineficiencias de costes variables

---

<sup>44</sup> Sin embargo, son muchos los trabajos que han estudiado la eficiencia de costes de los bancos utilizando el análisis envolvente de datos (DEA) (Maudos y Pastor, 2003; Giokas, 2008; Camanho y Dyson, 2005a,b; Fu y Heffernan, 2007; Tsolas y Giokas, 2012), o modelos de frontera de costes estocásticos (SCF) (Orea y Kumbhakar, 2004; Greene, 2005; Kumbhakar et al., 2013).

en el tiempo teniendo en cuenta la heterogeneidad específica no observada,<sup>45</sup> una heterogeneidad latente que puede reflejar la presencia de problemas de gestión no sistemáticos que pueden resolverse en el corto plazo.<sup>46</sup> En este análisis utilizamos las especificaciones propuestas por Greene (2005a,b). En segundo lugar, asumimos que la ineficiencia es función de un conjunto de variables explicativas, siguiendo a Battese y Coelli (1995). Así, asumiremos que la ineficiencia de costes media es una función del propio desempeño de la sucursal, de otras variables específicas de la sucursal y del tiempo. El procedimiento utilizado permitirá estimar simultáneamente los parámetros de la frontera estocástica y del modelo de ineficiencia correspondiente al desempeño de la sucursal.

El análisis empírico se realizó utilizando datos mensuales de las sucursales, en el período comprendido entre enero de 2013 y diciembre de 2014. Este enfoque permitirá evaluar el desempeño de las sucursales y sus gerentes.

El resto del capítulo se estructura de la siguiente manera. La siguiente sección describe brevemente la revisión de la literatura realizada, con especial énfasis en el uso de un marco de frontera estocástica para analizar el desempeño de las sucursales bancarias. La Sección 3 describe algunos modelos de frontera estocástica utilizados para determinar la eficiencia de las sucursales en el contexto de heterogeneidad no observada. La sección 4 presenta los datos obtenidos y los resultados derivados de nuestro análisis empírico.

---

<sup>45</sup> La heterogeneidad no observada podría producir un sesgo de especificación errónea en presencia de factores no observables invariantes en el tiempo en un modelo de datos de panel.

<sup>46</sup> Por ejemplo, podrían existir diferencias en la capacidad de los gerentes para fomentar el uso efectivo de nuevos canales de distribución para asegurar que el personal de la sucursal use su tiempo en actividades de valor agregado, para administrar la combinación de productos de manera que genere una alta rentabilidad o para aumentar ventas y la cartera de clientes. Cualquiera de estas diferencias puede estar presente en las unidades en estudio.

## **8.2. REVISIÓN DE LA LITERATURA**

### **8.2.1. DESEMPEÑO Y EFICIENCIA DE SUCURSALES BANCARIAS**

Se han realizado pocos estudios sobre la relación entre el rendimiento y la eficiencia de las sucursales bancarias, y la mayoría de estos utilizan métodos de frontera no paramétricos como el análisis envolvente de datos (DEA), el casco de libre disposición (*free disposal hull*, FDH) y los métodos no paramétricos robustos de tipo orden- $\alpha$ . Entre estos análisis previos, Giokas (1991) comparó la eficiencia operativa obtenida de DEA y un modelo de frontera determinista log-lineal, examinando 17 sucursales bancarias en Grecia y encontrando que los resultados de los dos métodos no presentaban diferencias significativas. Posteriormente, Lovell y Pastor (1997) utilizaron un marco DEA y tecnología de rendimientos variables a escala, con una sola entrada constante o sin entradas. Estos autores buscaron maximizar la expansión radial sobre el vector de 17 indicadores de desempeño para las 545 sucursales bancarias evaluadas. En otro estudio, Camanho y Dyson (1999) utilizaron la medida de eficiencia DEA para complementar la de rentabilidad actualmente utilizada en el banco en cuestión. Los resultados obtenidos proporcionaron información importante sobre cómo se podría mejorar el rendimiento de las sucursales bancarias. El uso de una matriz de eficiencia-rentabilidad destacó que sucursales podrían aumentar sus ganancias aumentando la eficiencia. Estos autores concluyeron que la eficiencia de las sucursales se asocia positivamente con las ganancias, pero que una alta rentabilidad no necesariamente está directamente relacionada con altos niveles de eficiencia.

Más recientemente, Cabrera-Suárez y Pérez-Rodríguez (2020) analizaron la relación entre el rendimiento de las sucursales y las puntuaciones de eficiencia presentadas por un gran banco comercial español antes y después del rescate del sistema bancario español en 2012, utilizando métodos DEA, FDH y order- $\alpha$  basados en los enfoques de producción, intermediación y rentabilidad. Además, se utilizó el análisis de frontera parcial para detectar posibles valores atípicos en los datos que posteriormente se someterían a un análisis de eficiencia no paramétrico por parte de DEA y FDH, utilizando el procedimiento descrito por Daraio y Simar (2007). Estos autores informaron que la productividad basada en el valor (PBV) era el único índice de desempeño utilizado actualmente en el banco que se relacionaba, positiva y significativamente, con los puntajes de eficiencia DEA y order- $\alpha$ .

## **8.2.2. EFICIENCIAS VARIABLES EN EL TIEMPO DE LAS SUCURSALES**

### **BANCARIAS**

#### **8.2.2.1. EFICIENCIA VARIABLE EN EL TIEMPO EN DATOS DE PANEL**

Existe una gran cantidad de literatura sobre el tema del análisis de la frontera de costes estocásticos (SCF) en un marco de datos de panel. Un enfoque importante es el de los modelos de ineficiencia invariante en el tiempo de efectos aleatorios, descritos por Pitt y Lee (1981) y Battese y Coelli (1988), en los que la ineficiencia específica de la empresa no varía con el tiempo. Otro es la versión de efectos fijos del modelo de Schmidt y Sickles (1984), caracterizado por no hacer supuestos distributivos sobre el término de ineficiencia.

Alternativamente, se ha propuesto una amplia gama de modelos de ineficiencia variables en el tiempo en el marco bayesiano, como Koop et al., (1997), y en el marco no bayesiano, por Cornwell et al. (1990) y Lee y Schmidt (1993), entre otros, junto con el modelo flexible propuesto por Kumbhakar (1990) y los modelos de efectos de decaimiento e ineficiencia descritos por Battese y Coelli (1992, 1995).

La característica común de todos estos modelos SCF variables en el tiempo es que la intersección es la misma en todas las unidades productivas, lo que genera un sesgo de especificación errónea en presencia de factores inobservables invariantes en el tiempo, que no están relacionados con el proceso de producción, pero afectan al resultado. Por lo tanto, el efecto de estos factores podría ser capturado por el término de ineficiencia, produciendo resultados sesgados. Para superar este problema, Greene (2005a,b) examinó varios enfoques alternativos para modelar la heterogeneidad en datos de panel en el modelo de frontera estocástica. Propuso especificaciones que pueden aislar la heterogeneidad de la empresa y preservar mejor el mecanismo en el modelo de frontera estocástica que produce estimaciones de ineficiencia técnica o de costes. Este autor consideró dos modelos para la eficiencia variable en el tiempo: el modelo de efectos fijos verdaderos (*true fixed effects*, TFE) y el modelo de efectos aleatorios verdaderos (*true random effects*, TRE), de acuerdo, con los supuestos realizados sobre la heterogeneidad específica de la unidad no observada.

La heterogeneidad no observada específica de la unidad puede interpretarse como ineficiencia persistente (Kumbhakar, 1991; Kumbhakar y Heshmati, 1995 y Kumbhakar y Hjalmarsson, 1993 y 1995) o relacionada con efectos de la empresa.

Colombi et al. (2014) y Tsionas y Kumbhakar (2014) dieron un paso más en este sentido al dividir el término de error invariante en el tiempo en dos términos, uno para

capturar la heterogeneidad latente de la empresa (Greene, 2005a,b) y el otro para capturar su ineficiencia persistente (Kumbhakar & Heshmati, 1995; Kumbhakar & Hjalmarsson, 1993, 1995). Esto se ha denominado modelo TRE generalizado (GTRE) (Tsionas & Kumbhakar, 2014). Con respecto a estas definiciones, Pitt y Lee (1981) desarrollaron específicamente la parte persistente de la ineficiencia, siguiendo a Schmidt y Sickles (1984), mientras que el efecto transitorio de la ineficiencia ha sido abordado por Kumbhakar (1990), Cornwell et al. (1990), Battese y Coelli (1992, 1995) y Greene (2005a, b).

El enfoque original desarrollado por Colombi et al. (2014) resultó extremadamente engorroso de implementar en la práctica. Tsionas y Kumbhakar (2014) también propusieron una solución bayesiana parcial, pero no lograron una implementación práctica completa del estimador de máxima verosimilitud (MLE). Más recientemente, Filippini y Greene (2016) desarrollaron un método MLE simulado con información completa que resultó más práctico de utilizar en el modelo propuesto por Colombi et al. (2014).

#### **8.2.2.2. EFICIENCIA DE LAS SUCURSALES EN UN MARCO DE FRONTERA ESTOCÁSTICA**

Hasta donde sabemos, muy pocos estudios han analizado la eficiencia de las sucursales bancarias utilizando modelos de fronteras estocásticas de costes (SCF) y datos de panel. Uno de ellos es el artículo de Berger et al. (1997), quienes utilizaron una función de coste translog y términos trigonométricos de Fourier para reducir el número de términos necesarios para una aproximación cercana. Su ecuación se estimó por separado para cada período de tiempo, lo que permitió que los parámetros de la función



de coste variaran con el tiempo. En otras palabras, no estimaron ningún panel de datos utilizando el método SCF. Estos autores informaron de la existencia de aproximadamente el doble de sucursales de las que se necesitarían para minimizar los costes, pero observaron que esta "ineficacia" podría ser óptima desde el punto de vista de la rentabilidad porque el "exceso de sucursales" aumentaría los ingresos al mejorar la comodidad del cliente. Los valores de ineficiencia  $X$  calculados fueron bastante grandes, más del 20% de los costes operativos. Estos hallazgos pueden explicar algunos de los resultados de eficiencia que se encuentran comúnmente en el análisis a nivel de banco y tienen implicaciones importantes con respecto a las fusiones y adquisiciones y las sucursales interestatales, o sea, dentro del territorio, en el sector bancario.

Otros estudios al respecto fueron realizados por Osiewalski y Marzec (1998) y Marzec y Osiewalski (2008), quienes analizaron la rentabilidad de las sucursales bancarias utilizando un marco bayesiano.

Además, Marzec y Osiewalski (2008) analizaron la rentabilidad a corto plazo de las sucursales bancarias utilizando una metodología bayesiana en un marco de datos de panel. Estos autores presentaron estadísticos bayesianos y el muestreo de Gibbs, empleando una técnica de simulación MCMC (*Markov Chain Monte Carlo*), como herramienta para hacer inferencias en modelos de SCF para datos de panel con respecto a las sucursales de un banco comercial polaco. Se utilizó una función de coste translog y la ineficiencia se trató como un efecto individual aleatorio, asumiendo la especificación de una distribución de eficiencia variable (VED) propuesta por Koop et al. (1997).

Algunos trabajos empíricos no bayesianos en el campo del análisis de la eficiencia bancaria han utilizado un método de dos etapas mediante el cual las estimaciones de eficiencia obtenidas en la primera etapa se utilizaron en una regresión en la segunda

etapa, relacionándola con variables explicativas adicionales. En este enfoque, el modelo bayesiano VED utilizado proporcionó un marco coherente para estimar y probar la influencia de los factores exógenos en la eficiencia individual. Luego se aplicó la metodología bayesiana para obtener inferencias sobre la tecnología y los niveles de eficiencia de costes de corto plazo de 58 sucursales del mismo banco. Los resultados obtenidos, basados en datos de panel de los cuatro trimestres de un solo año y en una frontera de costes variables translogarítmica, indicaron rendimientos crecientes a escala (que varían con la producción de la sucursal) y no hay diferencias sistemáticas en la eficiencia que puedan ser explicadas por las tres variables ficticias bajo consideración. El ejemplo presentado también muestra que la cooperación con econométricos bayesianos puede proporcionar información importante sobre el funcionamiento económico de un banco, brindando a la gerencia información no solo sobre las características microeconómicas básicas de cada sucursal, sino también sobre su eficiencia y los posibles determinantes de ésta.

Más recientemente, Gómez-Déniz y Pérez-Rodríguez (2017) utilizaron una función de coste log-lineal Cobb-Douglas de las sucursales de un gran banco comercial español para mostrar la aplicabilidad de su modelo propuesto, analizando datos mensuales para el período 2011-2014. Dichos autores extendieron los resultados obtenidos por Gómez-Déniz y Pérez-Rodríguez (2015) a la distribución normal-exponencial con dependencia del error en el modelo de frontera estocástica.<sup>47</sup> Este modelo asume una estructura de dependencia entre el término de ineficiencia y el error aleatorio (componente idiosincrásico), y anida el modelo clásico normal-exponencial cuando el parámetro de

---

<sup>47</sup> En consecuencia, el objetivo principal de este artículo es mejorar el modelado de la frontera de producción estocástica (coste) proponiendo una distribución bivariada para errores dependientes que permita anidar los modelos clásicos. Se proporcionan expresiones de forma cerrada para el término de error y la eficiencia técnica.

dependencia es cero. Las variables utilizadas fueron los costes, el precio de la mano de obra y del capital y, los ingresos de cada sucursal bancaria, todas transformadas logarítmicamente. Los resultados obtenidos mostraron que el modelo clásico normal-exponencial dio un ajuste ligeramente peor que el modelo con dependencia entre errores, y no se rechazó el supuesto de dependencia. Este estudio utilizó datos de sucursales bancarias solo para ilustrar el ajuste del modelo, no para interpretar los resultados con respecto a las actividades del gerente de la sucursal, los problemas de política o las fortalezas y debilidades de las operaciones de la sucursal.

Sin embargo, ningún estudio ha analizado la relación entre el desempeño y la eficiencia de las sucursales bancarias en el marco del SCF.

### **8.3. LA HETEROGENEIDAD LATENTE Y EFICIENCIA DE COSTES VARIABLE EN EL TIEMPO DE LAS SUCURSALES BANCARIAS**

La estimación econométrica de funciones de coste ha sido propuesta como un enfoque adecuado para determinar costes, eficiencias y elasticidades de escala, para cualquier industria.

En el presente estudio no se utiliza la especificación propuesta por Filippini y Greene (2016), basada en el modelo normal-half normal de GTRE, debido a que, este enfoque no permite incluir determinantes en la media de la ineficiencia de costes. Por lo tanto, nos centramos en los modelos TRE y TFE, que nos permiten incluir los efectos de las covariables en el término de ineficiencia media en la especificación normal truncada. Estos modelos no solo ayudan a separar los efectos invariantes en el tiempo no observados de las estimaciones de eficiencia variables en el tiempo, sino que también pueden incluir covariables en la media de un término de ineficiencia normal truncado.

Como se comentó en la Sección 2.2.1, la especificación utilizada por Greene (2005a) permite aislar la ineficiencia variable en el tiempo de la heterogeneidad no observada invariante en el tiempo específica de la unidad en el modelo translog SCF. Por lo tanto, las estimaciones de eficiencia brindan información sobre el componente transitorio (a corto plazo) de la eficiencia, mientras que el componente invariante en el tiempo (componente persistente) de la ineficiencia se absorbe completamente dentro del término constante específico de la sucursal.

La especificación general para nuestro modelo SCF, con datos de panel TRE para producción múltiple e incorporando el progreso tecnológico neutral de Hicks, expresado por la tendencia lineal y productos cruzados, junto con otras variables explicativas (outputs y precios de inputs) y tendencias al cuadrado, puede escribirse como:

$$\begin{aligned}
 \log c_{it} &= \alpha_i + \sum_{m=1}^M \beta_m \log y_{m,it} + \sum_{k=1}^K \alpha_k \log w_{k,it} \\
 &+ \frac{1}{2} \sum_{m=1}^M \sum_{k=1}^M \varphi_{mk} \log y_{m,it} \log y_{k,it} + \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^K \gamma_{kj} \log w_{k,it} \log w_{j,it} \\
 &+ \sum_{m=1}^M \sum_{k=1}^K \delta_{mk} \log w_{k,it} \log y_{m,it} \\
 &+ \kappa_1 t + \kappa_2 t^2 + \sum_{k=1}^K \theta_k \log w_{k,it} t + \sum_{m=1}^M \tau_m \log y_{m,it} t + \varepsilon_{it} \tag{8.1}
 \end{aligned}$$

$$\begin{aligned}
 \alpha_i &= \alpha + \omega_i, \omega_i \sim N(0, \sigma_\omega^2) \\
 \varepsilon_{it} &= v_{it} + u_{it} \\
 v_{it} &\sim N(0, \sigma_v^2) \\
 u_{it} &\sim N^+(\mu_{it}, \sigma_u^2)
 \end{aligned}$$

donde  $\log$  es el logaritmo natural,  $i = 1, 2, \dots, n$  es el número de sucursales,  $t = 1, 2, \dots, T_i$  es el número de períodos de tiempo (que pueden ser diferentes para cada sucursal),  $c_{it}$  es el coste de la empresa  $i$  en el período  $t$ ;  $w_{j,it}$  es el  $j$ -th precio del input y  $y_{m,it}$  es la  $m$ -th producción de la sucursal  $i$  en el período  $t$  (nivel, cuadrados y productos cruzados);  $\alpha$ ,  $\kappa_1, \kappa_2$ ,  $\beta_m, \tau_m, m = 1, \dots, M$ ,  $\alpha_k, \theta_k, k = 1, \dots, K - 1$ ,  $\varphi_{mk}$  y

$\delta_{mk}, m = 1, \dots, M, k = 1, \dots, K - 1, \gamma_{kj}, k, j = 1, \dots, K - 1$  son parámetros desconocidos. En estas expresiones,  $\alpha$  es un término constante común para todas las unidades y  $\omega_i$  es un componente aleatorio iid en el marco de efectos aleatorios. Este último es el término constante específico aleatorio que captura la heterogeneidad no observada persistente específica de la unidad (no la ineficiencia), y normalmente se distribuye con una media cero y la varianza subyacente  $\sigma_\omega^2$ . En nuestro modelo, el error compuesto  $\varepsilon_{it} = v_{it} \pm u_{it}$  es asimétrico (cuyo signo depende de si la frontera describe costes (positivo) o producción (negativo)), donde  $v_{it}$  se asume que es iid  $N(0, \sigma_v^2)$ .  $u_{it}$  es el error de componente unilateral, un componente no negativo que a menudo se denomina término de ineficiencia. Se supone que se trata de una variable aleatoria iid con una distribución no normal específica (distribución seminormal, exponencial o normal truncada). En el caso normal truncado, el término de ineficiencia  $u_{it}$  se distribuye como  $N^+(\mu_{it}, \sigma_u^2)$ , donde  $\mu_{it} = \delta' z_{it}$  es el término medio de ineficiencia en función del vector de parámetros  $z_{it}'$  el vector de parámetros desconocidos  $\delta$ . Esto implica que la ineficiencia varía con el tiempo.

En el modelo TRE, toda ineficiencia invariante en el tiempo se interpreta como la heterogeneidad específica de las sucursales y esta parte no aparece como ineficiencia. En la configuración de TRE, cualquier componente de ineficiencia persistente o invariante en el tiempo se absorbe por completo dentro del término constante específico del individuo.

La verosimilitud logarítmica simulada se expresa mediante:

$$\log(\alpha, \beta, \lambda, \sigma, \sigma_\omega) = \sum_{i=1}^n \log \frac{1}{R} \sum_{r=1}^R \left\{ \prod_{t=1}^T \left[ \begin{array}{l} \frac{2}{\sigma} \phi \left( \frac{y_{it} - \alpha - x'_{it} \beta - \sigma_\omega W_{ir}}{\sigma} \right) \\ \times \Phi \left( -\frac{(y_{it} - \alpha - x'_{it} \beta - \sigma_\omega W_{ir}) \lambda}{\sigma} \right) \end{array} \right] \right\} \quad [8.2]$$

donde  $\beta$  incluye todos los parámetros desconocidos en la ecuación de costes (es decir,  $\kappa_1, \kappa_2, \beta_m, \tau_m, m = 1, \dots, M, \alpha_k, \theta_k, k = 1, \dots, K - 1, \varphi_{mk}$  y  $\delta_{mk}, m = 1, \dots, M, k = 1, \dots, K - 1$ ) excepto  $\alpha, \phi$  es la densidad normal estándar y  $\Phi$  es la cdf estándar,  $\lambda = \frac{\sigma_u}{\sigma_v}, \sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$  y  $\varepsilon_{it} = y_{it} - \alpha_i - x'_{it} \beta$ .  $W_{ir}$  son  $R$  simulaciones de la población normal estándar. Estas simulaciones se pueden obtener utilizando Halton o secuencias de Halton generalizadas en lugar de números pseudoaleatorios. También se simulan las derivadas necesarias para la optimización basada en gradientes y para calcular el estimador de la matriz de covarianza asintótica. El modelo es convencional y satisface las condiciones de regularidad en el marco de máxima verosimilitud.

El TRE también puede estar sujeto al 'sesgo de variables omitidas', porque las variables no observadas pueden estar correlacionadas con los regresores. Farsi et al. (2005) sugirieron agregar una ecuación auxiliar a la TRE para abordar el problema econométrico del sesgo de heterogeneidad no observada.

Al asumir  $\sigma_\omega^2 = 0$  en el modelo [1], el modelo TFE se puede estimar mediante el enfoque de variable ficticia de máxima verosimilitud (MLDV), que es computacionalmente factible incluso, en presencia de una gran cantidad de parámetros molestos (*nuisance*)  $\alpha_i$  ( $n > 1000$ ) (Greene, 2005a). En este caso, la función logarítmica de verosimilitud para el modelo SCF de efectos fijos se escribe como:

$$\log(\alpha_1, \dots, \alpha_n, \beta, \lambda, \sigma) = \sum_{i=1}^n \sum_{t=1}^T \log \left[ \frac{1}{\Phi(0)} \Phi \left( -\frac{\varepsilon_{it} \lambda}{\sigma} \right) \phi \left( \frac{\varepsilon_{it}}{\sigma} \right) \right] \quad [8.3]$$

Greene (2005a) indicó que ninguna operación de transformación o condicionamiento producirá una función de verosimilitud libre de efectos fijos, por lo que es necesario estimar todos los  $n+k+2$  parámetros simultáneamente. Sin embargo, en este enfoque, las intersecciones específicas de la unidad se estiman de manera inconsistente cuando  $n \rightarrow \infty$  con  $T$  fijo, ya que sólo  $T_i$  observaciones se usan para estimar cada parámetro específico de la unidad. Como se muestra en Belotti e Ilardi (2012), esta inconsistencia contamina los parámetros de varianza, que representan los ingredientes clave en la estimación posterior de las ineficiencias, por lo que el enfoque MLDV solo es apropiado cuando el panel es suficientemente grande ( $T \geq 10$ ).

## **8.4. RESULTADOS EMPÍRICOS**

### **8.4.1. DATOS**

Para el presente estudio se recopilieron datos contables mensuales desde enero de 2013 hasta diciembre de 2014 a partir de las cuentas de pérdidas y ganancias, junto con información complementaria de departamentos internos (Control de Gestión, Recursos Humanos e Ingeniería, entre otros) de 151 sucursales de un gran banco comercial en España. Estas sucursales atienden a personas físicas y empresas, siendo el servicio recibido homogéneo (es decir, se prestan servicios similares en ambos casos). Las sucursales están ampliamente distribuidas en una región de España, y están ubicadas en áreas rurales y urbanas, algunas de las cuales, tienen una alta concentración de sucursales. Sin embargo, algunas sucursales presentaron valores extremos para ciertas variables y pueden considerarse outliers o valores atípicos. Esta circunstancia podría influir en el procedimiento de estimación y, por tanto, para evitar distorsiones por este

motivo, se eliminaron de la muestra las sucursales en cuestión. Después de filtrar esta base de datos de esta manera, quedaron para el análisis un total de 134 sucursales bancarias. Por razones de confidencialidad, no se proporciona más identificación.

#### **8.4.1.1. DOS MEDIDAS DE DESEMPEÑO PROPIAS PARA EVALUAR LAS SUCURSALES BANCARIAS**

El banco en cuestión utiliza varias medidas para evaluar el desempeño de sus sucursales, examinando las cifras de actividad de ventas de acuerdo con los seguimientos diarios, incluido el informe de productividad basada en valor (PBV), el volumen de negocios, el balance, la cuenta de pérdidas y ganancias y el proceso de digitalización. Sin embargo, este estudio se enfoca solamente en dos medidas del desempeño de las sucursales: 1) la desviación del objetivo de margen bruto (o sea, el ingreso total); 2) PBV. Estas medidas fueron elegidas para el análisis porque son los indicadores más importantes de la estabilidad de las unidades en el tiempo, y también por razones de disponibilidad de datos.

Recordamos que la PBV denominada ‘Productividad Basada en Valor’ es la medida del banco del volumen de negocio alcanzado, en términos de valor recurrente generado. Se calcula para un horizonte trimestral, ya que una operación contratada en un determinado año se estima que seguirá generando ingresos en años posteriores. La PBV mide la actividad comercial diaria en términos del valor actual neto generado por una operación, teniendo en cuenta variables como el volumen de la operación, su precio, plazo y riesgo, y el hecho de que una operación contratada generará ingresos en años siguientes. Estos contienen tres secciones: recursos, inversiones y productos, y se controlan no solo a nivel de unidad, sino también para cada gestor que compone la



unidad con responsabilidad de ventas. La PBV se presupuesta y su análisis se calcula como un índice, por lo que se expresa en unidades porcentuales. En el presente capítulo se utilizan ambas medidas.

Otras medidas se obtienen del balance y de los estados de pérdidas y ganancias de cada unidad. El objetivo del control presupuestario es detectar y analizar las desviaciones, como parte esencial de la planificación empresarial. En los últimos años, las funciones de la red de sucursales se han valorado mediante el cálculo de los ingresos netos por intereses y los ingresos no financieros (comisiones) generados por cada unidad. Esta valoración se obtiene a partir de las desviaciones presupuestarias entre los resultados reales y esperados. La diferencia o desviación suele calcularse tanto en valor como en porcentaje. En el presente estudio, una variable dummy toma el valor 1 cuando la sucursal bancaria tiene un valor negativo y, 0 en caso contrario.

#### **8.4.1.2. ESTADÍSTICOS DESCRIPTIVOS**

La Tabla 8.1 muestra los estadísticos descriptivos obtenidos para las variables utilizadas en nuestros modelos. Algunas de estas variables representan cantidades y otras son valores monetarios.

Muchas de las variables utilizadas son similares a las especificaciones del modelo consideradas en investigaciones anteriores (p. ej., Berger et al., 1997). Al igual que en este último caso, se tiene en cuenta el número total de empleados equivalentes (incluidos los directores de oficina, asesores financieros, gestores de negocios, gestores comerciales, el personal administrativo y los cajeros), la superficie de la oficina (en m<sup>2</sup>), el número total de transacciones (es decir, las relacionadas con Internet, línea y

oficinas), el número total de préstamos (préstamos de consumo y empresariales e hipotecas), los costes operativos soportados por la sucursal (gastos totales excluyendo costes de personal), y precios de inputs: mano de obra (tasa salarial, es decir, salario y beneficios por empleado, abordados por la sucursal en su mercado local) y tasa de alquiler sobre el capital (es decir, costes generales más depreciación divididos por activos tangibles)<sup>48</sup> que se imputa a la sucursal en función de su mercado local. Todos los datos de valor en € son ajustados por inflación en euros constantes de 2011 (es decir, utilizamos el índice de precios al consumidor del mercado local/regional con respecto al período base 2011).<sup>49</sup>

Otras variables utilizadas son una variable ficticia para el número de trabajadores con certificación EFA (como medida de la calidad de los empleados para el asesoramiento financiero), una variable ficticia relacionada con el tamaño del mercado (urbano o rural) y las medidas de desempeño de la sucursal utilizadas por el propio banco.

---

<sup>48</sup> Este podría ser un proxy del precio del capital, compuesto por el coste de los locales y equipos dividido por el valor del edificio y el terreno.

<sup>49</sup> Las variables deflactadas se utilizan para evitar que las fluctuaciones de los precios distorsionen nuestra evaluación de las variables económicas medidas en euros.

**Tabla 8.1.** Estadísticos descriptivos de las sucursales bancarias analizadas. Período Enero 2013 – Diciembre 2014.

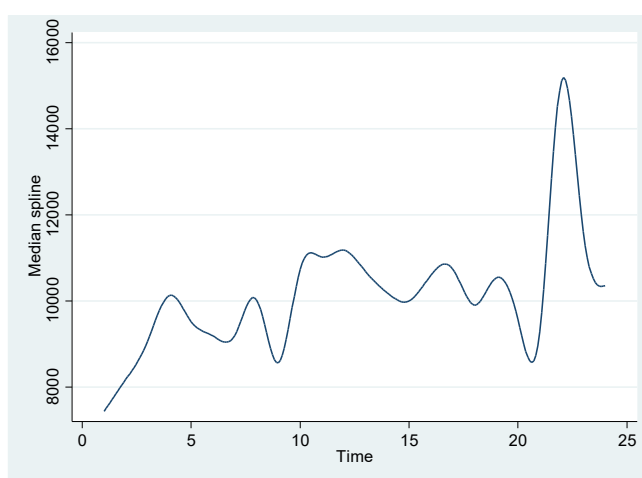
<b>VARIABLES</b>	<b>Media</b>	<b>Desviación Estándar</b>
Superficie de la sucursal (en m <sup>2</sup> )	322.56	292.26
Número de empleados	3.69	2.45
Préstamos totales (consume, préstamos empresariales e hipotecas) por sucursal/mes	15.15	16.54
Transacciones Totales (canal internet, línea y oficina)	8,830.42	4,673.65
Costes Operacionales (en euros)	13,196	10,908
Tasa salarial (salario y prestaciones por trabajador/mes, bruto) (en euros)	6,537	1771
Tasa de alquiler sobre el capital	0.15	0.40
Desviaciones de los objetivos de ingresos totales (margen bruto)	0.10	0.25
Productividad basada en valor (PBV) (%)	0.40	0.62
Trabajadores con certificación EFA; 1=Sí, 0=en caso contrario	0.35	--
Sucursal bancaria se encuentra en zona urbana (>20,000 habitantes); 1=Sí, 0=en caso contrario	0.63	--

En promedio, cada sucursal tenía cuatro empleados, negoció 15 préstamos y realizó 8.830 transacciones. El coste medio de explotación fue de 13.196 € y el salario bruto medio de 6.537 €. La desviación entre el margen bruto real y el esperado fue en promedio 0,10 mientras que el PBV fue de 0,40. La mayoría de los empleados de las sucursales no tenían certificación EFA y la mayoría de las sucursales bancarias estaban ubicadas en ciudades urbanas.

La figura 8.1 muestra la trayectoria temporal de los costes medios observados por mes utilizando un spline cúbico. Específicamente, calculamos las medianas cruzadas y las usamos como nudos para ajustar un spline cúbico. El spline resultante se representó luego como un diagrama de líneas. Como se puede observar, estos costes aumentaron con el tiempo, debido al aumento en la imputación de gastos, específicamente en lo que respecta a las facturas internas. En los últimos años se ha producido una tendencia creciente a imputar el coste de la estructura *holding* (empleados de servicios centrales

como recursos humanos, servicios jurídicos e ingeniería, entre otros, sistemas e instalaciones) de forma proporcional a las distintas regiones, lo que se denomina metodología *full costing*. Esto significa que se incrementa la partida de gastos en cuestión, así como, los derivados de impuestos, tasaciones y los gastos judiciales derivados de nueva normativa.

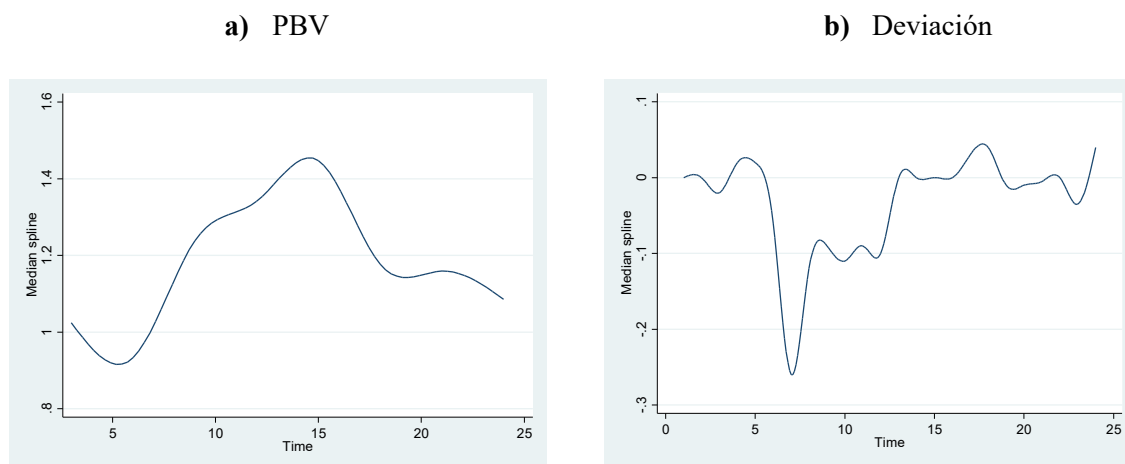
**Figura 8.1.** Trayectoria temporal de la mediana spline (cúb.) de los costes observados por mes.



La Figura 8.2 muestra la trayectoria temporal de la PBV y las desviaciones del margen bruto para cada mes. En cuanto a la PBV (Figura 2a), la productividad aumentó durante 2013 pero retrocedió a partir del segundo trimestre de 2014, posiblemente debido a una menor actividad comercial (a su vez, esto podría estar relacionado con la disminución de los valores generados por las ventas u operaciones, es decir, el volumen, precio, plazo y riesgo). Es de destacar que esta medida se mantuvo por encima de 1 durante todo el período de estudio. La Figura 2b, en general, muestra un patrón estable de las desviaciones, cercano a cero, con la excepción del segundo, tercer y cuarto trimestre de 2013. Estos resultados están en línea con las tendencias de la industria bancaria española durante este período, cuando las restricciones de crédito provocaron

una fuerte caída en el número y valor de los préstamos. Además, la crisis económica contribuyó a reducir el nivel de ahorro y el número de transacciones.<sup>50</sup>

**Figura 8.2.** Trayectoria temporal de la mediana spline (cúbica) de la PBV y la desviación del margen bruto por mes.



#### 8.4.2. ENFOQUE DE PRODUCCIÓN, RESTRICCIONES DE HOMOGENEIDAD Y EFICIENCIA DE COSTES

Para estimar el modelo [1], consideramos varios aspectos. Por un lado, para observar diferencias en el desempeño de los directores de las sucursales o en el uso que cada sucursal hace de sus recursos (en áreas tales como gastos de personal y operativos) para producir depósitos y préstamos y, transacciones, evaluamos el desempeño de las sucursales en términos de su producción, un enfoque operativo.<sup>51</sup> El enfoque de

---

<sup>50</sup> El sistema bancario español se ha reestructurado significativamente desde la crisis bancaria de 2010 y el rescate de 2012 provocado por la burbuja inmobiliaria, que tuvo un impacto severo en la economía española y sus instituciones financieras.

<sup>51</sup> Al elegir qué costes, productos y precios de inputs incluir en la especificación, en un marco SFA, Berger et al. (1997) analizaron los enfoques de 'intermediación' y 'producción'. Es de destacar que en un

producción enfatiza la actividad operativa; por lo tanto, las sucursales son vistas como proveedores de servicios para los clientes. En esta dimensión, los inputs suelen incluir variables físicas relacionadas con los costes de operación (es decir, mano de obra, materiales, espacio y sistemas de información). En este enfoque, para el caso en cuestión, usamos  $n = 134$ ,  $T_i = 24, \forall i$ ,  $M = 2$  y  $K = 2$ , con  $y_1$  y  $y_2$  que representan el número total de préstamos y transacciones, respectivamente.

Por otro lado, asumimos las restricciones habituales de simetría ( $\varphi_{mk} = \varphi_{km}$  y  $\gamma_{kj} = \gamma_{jk}$ ) e imponemos la condición de que la función de coste sea homogénea de grado uno en precios, es decir, para un nivel fijo de producción, el coste total debe aumentar proporcionalmente cuando todos los precios aumentan proporcionalmente. En otras palabras, los precios de costes e inputs son normalizados por  $w_{K,it}$  (el numerario) para imponer restricciones de homogeneidad. El numerario se considera el precio del trabajo. Por lo tanto, el precio del capital se divide por el precio del trabajo, o sea,  $w_2/w_1$ .

Finalmente, incluimos tendencias temporales y productos cruzados con el resto de las variables en la estimación de frontera. En cuanto a las características de los términos idiosincrásicos y de ruido, asumimos que el valor medio del término de ineficiencia depende del desempeño de las sucursales bancarias (PBV y la variable ficticia que representa el valor negativo de la desviación del margen bruto) y otros controles como los efectos del tiempo, superficie, la calidad de la capacitación de los empleados, tales como si la sucursal bancaria tiene empleados con certificación EFA o no, y el tipo de mercado (urbano o rural).

---

marco DEA, Oral y Yolalan (1990) analizaron varios enfoques operativos y de rentabilidad, y Paradi et al. (2011) utilizaron un análisis de eficiencia tridimensional, considerando que esto proporciona una evaluación significativamente más integral del desempeño de las sucursales bancarias; además, también es probable que sea mejor aceptado por la gerencia a nivel de sucursal.

Después de haber estimado los modelos de costes definidos en [1], la función translog de costes se puede utilizar para obtener eficiencias de costes variables en el tiempo a través del enfoque sugerido por Jondrow et al. (JMLS, 1982), y también construir los intervalos de confianza mediante el enfoque de Horrace y Schmidt (1996), junto con medidas adicionales de interés como el cambio técnico (TC) y los rendimientos a escala (RTS), que son funciones de los parámetros y datos considerados. Siguiendo a Kumbhakar et al. (2013), calculamos ambas medidas. Como son específicos de la observación, se pueden definir de la siguiente manera. Para nuestra función de coste, TC es la reducción en el coste (tasa de cambio en el coste a lo largo del tiempo, manteniendo constante todo lo demás), es decir,  $\partial \log c_{it} / \partial \log t_{it}$ . Un valor negativo sugiere un progreso técnico (reducción de costes a lo largo del tiempo, todo lo demás en igualdad de condiciones). RTS mide el aumento proporcional de los costes dado un aumento en todos los productos y se define como el recíproco de  $\sum_m \partial \log c_{it} / \partial \log y_{m,it}$ . Si RTS excede la unidad, existen economías de escala, de modo que un aumento proporcional en todos los productos produce un aumento menos que proporcional en el coste (es decir, la escala de operación está por debajo del nivel óptimo y, por lo tanto, se pueden obtener beneficios de la expansión). Lo contrario es cierto si RTS es menor que la unidad.

### 8.4.3. ESTIMACIONES DEL MODELO

En esta sección mostramos los resultados de los modelos de efectos fijos verdaderos (TFE) y efectos aleatorios verdaderos (TRE) desarrollados por Greene (2005a), en los que se consideran tanto la heterogeneidad invariable en el tiempo como la ineficiencia

empresarial variable en el tiempo. El modelo TFE se estimó mediante máxima verosimilitud con variables ficticias utilizando la distribución normal truncada para el término de ineficiencia, y el modelo TRE se estimó maximizando la verosimilitud logarítmica simulada utilizando la distribución normal truncada.<sup>52</sup> Además, las estimaciones de TRE se obtuvieron utilizando 250 secuencias de Halton aleatorias. Aunque se abordó la influencia potencial de los valores atípicos, al eliminar ciertas observaciones problemáticas, no pudimos utilizar especificaciones de frontera estocástica alternativas para tratar los posibles valores atípicos en la muestra final, debido a problemas de cálculo.<sup>53</sup> Por lo tanto, nuestra estimación podría ser sensible a la heterocedasticidad, si está presente.

La Tabla 8.2 muestra las estimaciones de los parámetros y sus p-valores para la ecuación de frontera estocástica, la ecuación de media de ineficiencia y las varianzas correspondientes. También mostramos el estadístico chi-cuadrado para la hipótesis nula de que todos los parámetros en el modelo, con la excepción del término constante, son igualmente cero, y el valor máximo del logaritmo de la probabilidad (TFE) y la probabilidad logarítmica simulada (TRE), denominado Log L. Adicionalmente, proporcionamos estadísticos descriptivos para las estimaciones JLMS de eficiencia de costes (media, mediana, máxima y mínima), junto con los valores medios de TC y RTS. Finalmente, se muestran las observaciones totales del panel, así como el número de

---

<sup>52</sup> Para realizar el cálculo de los modelos TFE y TRE hemos utilizado STATA 13 y códigos desarrollados por Belotti et al. (2012).

<sup>53</sup> Según Stead et al. (2018), existen especificaciones de frontera estocástica alternativas que son apropiadas cuando el error tiene colas grandes, como los modelos de frontera heteroscedástica (en los que la varianza está relacionada con las variables observables), el análisis de frontera gruesa o el uso de modelos de frontera estocástica gaussiana, entre otros. Teniendo en cuenta la heteroscedasticidad tanto en  $u_i$  como  $v_i$  usando covariables (ver Caudill y Ford, 1993; Caudill et al., 1995; o Hadri, 1999), intentamos estimar varias especificaciones de TFE y TRE considerando diferentes covariables en la varianza de los modelos, pero se encontraron problemas computacionales, debido a la ausencia de convergencia.



datos efectivos utilizados en las estimaciones. Los modelos estimados incluyen en la media de ineficiencia: PBV, una variable ficticia para las desviaciones negativas de los márgenes brutos, la superficie útil de la sucursal y una tendencia temporal cuadrática. Otros modelos estimados incluyeron el tamaño del mercado y las variables EFA. Sin embargo, estos resultados no se muestran por problemas de convergencia y porque sus coeficientes no son estadísticamente significativos al nivel del 5%.

En general, los resultados presentados sugieren que la eficiencia de costes varía en el tiempo y que existe una heterogeneidad no observada, lo que indica posibles problemas no sistemáticos de gestión, que pueden estar relacionados con el período de crisis y reajuste; por ejemplo, las sucursales pueden haber reajustado su fuerza laboral, a través del número de empleados y también a través de sus precios (es decir, reduciendo los costes). De hecho, en la entidad fue un período de salidas de personal a través de medidas de prejubilación, bajas incentivadas y, con reducciones de salarios en el componente variable dentro del proceso de evaluación anual conocido como incentivación. Sin embargo, como TFE conduce a estimaciones de ineficiencia sesgadas debido al problema de los parámetros incidentales, analizaremos los resultados de TRE como un procedimiento de estimación alternativo. Cabe también destacar que muchos de los coeficientes de las variables y la mayoría de los efectos fijos en el modelo TFE no son estadísticamente significativos al nivel del 5% (en aras de la brevedad, los efectos fijos no se muestran). Por lo tanto, centramos nuestras explicaciones en el modelo TRE.

Los resultados de la estimación basados en la frontera estocástica de costes translog TRE muestran que muchos parámetros de la función de costes translogarítmica son estadísticamente significativos al 5%, incluidos los parámetros de las varianzas como  $\sigma_u$ ,  $\sigma_v$  y  $\sigma_w$  ( $\sigma_v$  es mayor que  $\sigma_u$ ).

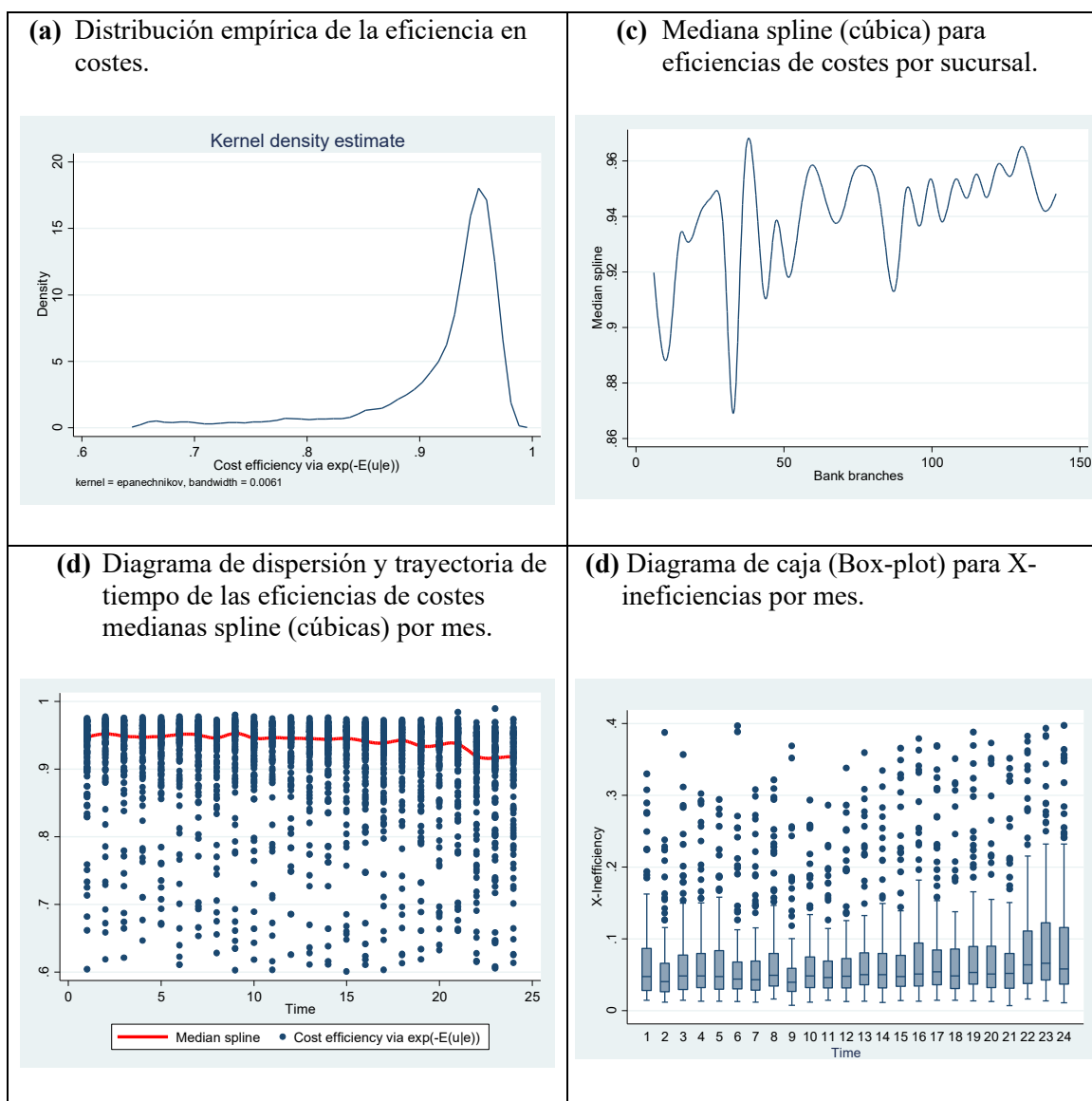
**Tabla 8.2.** Estimaciones de datos de panel del modelo de frontera estocástica de costes translogarítmicas (la ineficiencia sigue una distribución normal truncada).

Variables	TFE		TRE	
	Coef.	p-valor	Coef.	p-valor
<b>Ecuación frontera de costes</b>				
log $y_1$	0.2911	0.01	0.1403	0.00
log $y_2$	-0.1724	0.74	0.8727	0.00
log $w_2/w_1$	0.2602	0.21	1.0919	0.00
$0.5 \times (\log y_1)^2$	0.0214	0.11	0.0351	0.01
$0.5 \times (\log y_2)^2$	0.2060	0.33	-0.2302	0.00
$0.5 \times \log y_1 \times \log y_2$	0.0043	0.90	-0.0289	0.41
$0.5 \times (\log w_2/w_1)^2$	-0.0563	0.04	0.0409	0.01
log $y_1 \times \log w_2/w_1$	0.0519	0.00	0.0253	0.00
log $y_2 \times \log w_2/w_1$	-0.0450	0.11	-0.0621	0.01
tiempo	0.1018	0.00	0.0469	0.00
tiempo <sup>2</sup>	-0.0011	0.00	-0.0007	0.00
log $y_1 \times$ tiempo	0.0008	0.44	0.0003	0.77
log $y_2 \times$ tiempo	-0.0041	0.01	-0.0030	0.02
log $w_2/w_1 \times$ tiempo	0.0023	0.06	-0.0015	0.00
constante	Sin mostrar	--	3.4760	0.00
<b>Media de ineficiencia</b>				
PBV	-0.00002	0.08	$-2.76 \times 10^{-7}$	0.00
Desviaciones negativas	-1.9699	0.21	-0.0073	0.73
log( $m^2$ )	7.5917	0.06	0.8636	0.00
tiempo	-6.0505	0.04	-0.0464	0.02
tiempo <sup>2</sup>	0.0990	0.04	0.0008	0.00
constante	11.4315	0.08	-4.4183	0.00
$\sigma_u$	2.3169	0.00	0.1776	0.00
$\sigma_v$	0.0395	0.00	0.2021	0.00
$\lambda$	58.655	0.00	0.8788	0.00
$\sigma_w$	--	--	0.3550	0.00
Log L	317.36		-27.94	
Chi-square	2981.28	0.00	$1.77 \times 10^8$	0.00
<b>Promedios:</b>				
Cambio tecnológico (TC)	-0.0045		-0.0034	
	(0.02)		(0.01)	
Rendimientos a escala (RTS)	1.7316		2.4314	
	(0.40)		(0.80)	
<b>Eficiencia de costes (JLMS) e intervalos de confianza del 95%:</b>				
Media	0.9214	[0.77,0.99]	0.8486	[0.80,0.99]
Mediana	0.9398	[0.80,0.99]	0.9368	[0.82,0.99]
Mínimo	0.6107	[0.50,0.99]	0.6867	[0.58,0.99]
Máximo	0.9916	[0.91,1.00]	0.9894	[0.92,1.00]
Observaciones	3,186		3,186	
Número de sucursales	134		134	

*Nota:* Las desviaciones estándar se muestran entre paréntesis, al igual que los intervalos de confianza de Horrace y Schmidt (1996) con un nivel de confianza del 95%.

En el modelo TRE, el valor medio de RTS es 2.43 con una desviación estándar de 0.8. Como este es mayor a uno y podría considerarse estadísticamente significativo, las economías de escala están presentes en las sucursales consideradas. Por otro lado, no hay evidencia de cambio tecnológico, el cual presenta una débil relación inversa con la eficiencia en costes (-0.0034; y desviación estándar de 0.01), pero no alcanza significación estadística.

**Figura 8.3.** Rentabilidad en el modelo TRE.



La Figura 8.3(a) muestra las eficiencias de costes obtenidas en el modelo TRE según la medida JLMS. La Figura 3a muestra las estimaciones de la densidad kernel para las eficiencias de costes generales. Hay ineficiencia en costes, es decir, las sucursales no son totalmente eficientes (lo que ocurriría si la eficiencia en costes fuera igual a 1). Sin embargo, aunque las sucursales bancarias analizadas no son totalmente eficientes, las puntuaciones estimadas de la eficiencia de costes son altas. Estadísticamente, algunas propiedades descriptivas de la eficiencia son que es asimétrica a la izquierda, con un valor medio de 0.84, y es unimodal (mediana igual a 0.94). Estos hallazgos indican el predominio de la alta eficiencia en las sucursales. Interpretamos estos resultados de eficiencia de costes en el sentido de que se puede producir el mismo vector de productos, ahorrando  $(1 - \text{eficiencia de costes}) \times 100$  del coste efectivamente incurrido. Por lo tanto, en términos medios, las sucursales bancarias podrían lograr un ahorro del 16%.

Con respecto a la eficiencia media de las sucursales bancarias, la Figura 8.3(b) muestra el spline mediano (cúbico) para las eficiencias de costes por sucursal. La mayoría de las sucursales operan con un alto nivel de eficiencia (mediana >92%). Estos hallazgos indican que las sucursales administraron sus intereses económicos de manera eficiente durante el período de estudio.

Las puntuaciones de eficiencia de costes son aceptables para la gerencia si dichas estimaciones son consistentes con las mediciones del desempeño utilizadas por los responsables en la toma de decisiones. Para aclarar este tema, analizamos los parámetros incluidos en el valor medio de ineficiencia, encontrando que la PBV, principal indicador de las ventas de la unidad y sus fuerzas de venta, es la única variable de desempeño con un coeficiente estadísticamente significativo al 5%. Esta variable tiene signo negativo, lo que indica que la ineficiencia disminuye cuando aumenta la

PBV, lo cual es económicamente coherente. Sin embargo, el coeficiente de otra variable de desempeño bancario, el de la variable ficticia que indica desviaciones negativas (es decir, el grado de cumplimiento de los presupuestos en la cuenta de resultados), no es estadísticamente significativo y el signo no es el adecuado. Por lo tanto, solo podemos indicar que la PBV es una medida relacionada con las estimaciones de eficiencia de costes obtenidas mediante el uso del modelo SCF translog. Otro hallazgo es que una mayor superficie de la sucursal está asociada con una menor eficiencia de costes, siendo el coeficiente para esta variable estadísticamente significativo y positivo. Este resultado es inesperado debido a que las oficinas de mayor tamaño cuentan con personal con perfiles especializados en asesoramiento de inversiones, y se cree que esto agrega valor a los resultados obtenidos por la sucursal. Por lo tanto, las políticas internas podrían revisarse para poner menos énfasis en aumentar la superficie de la sucursal.

Las ineficiencias medias aumentaron ligeramente durante el período de estudio, con un efecto de ineficiencia media de  $-0.048 + 2 \times (0.0008) \times t$ . A medida que  $t$  se incrementa de 25 a 48, este efecto aumenta con el tiempo.<sup>54</sup> Ambos coeficientes son estadísticamente significativos al nivel del 5%. La figura 8.3(c) muestra el diagrama de dispersión y la trayectoria temporal de las eficiencias de costes medianos. Los patrones son constantes en el primer año, pero disminuyen en el segundo, reflejando un patrón ligeramente no estacionario. Por ejemplo, los valores medianos de eficiencia disminuyeron ligeramente, de 0.95 a 0.90. Los mayores costes (ver Figura 8.1) podrían reflejar una menor eficiencia de las unidades a medida que disminuyen sus ingresos.

---

<sup>54</sup> Esto se debe a que nuestra base de datos también contiene otros dos años de información (2011 y 2012), pero estos no fueron utilizados porque el número de préstamos y transacciones no fue computado por el banco en esos años, solo los valores de los ahorros, transacciones y préstamos se recopilaron y registraron.

Este resultado podría estar relacionado con el hecho de que, a lo largo del período, el número de sucursales bancarias se redujo de manera constante en respuesta a la percepción de sobredimensionamiento (el Banco de España hizo recomendaciones para que el sistema bancario redujese aún más su estructura). Otro aspecto relevante es el de las dificultades de pago que enfrentan las PYMES, lo que generó restricciones en la autorización de préstamos con una mayor propabilidad de impagos, por lo que se contrajo toda la parte inversora del negocio del banco, reduciendo los ingresos y, por ende, la eficiencia de las sucursales.

Finalmente, examinamos el desempeño de las sucursales en términos de mejores prácticas (o menor coste), analizando hasta qué punto los costes de las sucursales coinciden con los de las mejores prácticas, suponiendo outputs idénticos producidos en las mismas condiciones.<sup>55</sup> En otras palabras, la eficiencia-X de costes también se puede calcular a partir de estimaciones de eficiencia de costes. Formalmente, la eficiencia-X se define como la ratio entre los costes mínimos en los que se incurriría si la sucursal fuera tan eficiente como una sucursal de mejores prácticas en la muestra, dadas las mismas variables exógenas (inputs y outputs), a los costes reales pronosticados, ajustados por el error aleatorio. Siguiendo a Berger et al. (1997) y Fun y Heffernam (2007), usamos la siguiente medida de eficiencia-X específica para la sucursal:

$$X - efficiency_{it} = \frac{\hat{C}_{\min}}{\hat{C}_{it}} = \frac{\exp(\hat{u}_{\min})}{\exp(\hat{u}_{it})}$$

---

<sup>55</sup> Los directivos deben considerar en sus políticas tanto la eficiencia en la asignación como en los costes. Este último se refiere a la capacidad de una sucursal para producir productos actuales a un coste mínimo y es el resultado de la eficiencia técnica y la eficiencia en la asignación. La eficiencia asignativa refleja la capacidad de la sucursal para utilizar sus inputs en proporciones óptimas, dados los precios de sus inputs.

donde  $\hat{C}_{min}$  representa los costes mínimos previstos de la sucursal bancaria con mejores prácticas;  $\hat{u}_{min}$  el mínimo de los  $\hat{u}_{it}$  de todas las sucursales de la muestra; y  $\hat{u}_{it}$  la predicción actual de la ineficiencia de costes de una sucursal específica,  $i$ , en el período  $t$ . La ineficiencia residual  $\hat{u}_{it}$  de la regresión se utiliza para derivar las estimaciones de eficiencia-X y se calcula como  $\hat{u}_{it} = E[u_{it}/\varepsilon_{it}]$ . Teóricamente, la eficiencia-X se encuentra en el intervalo (0,1] donde los valores más altos indican una mayor eficiencia y son iguales a uno para la sucursal de mejores prácticas dentro de los datos observados. Una ineficiencia-X de 0.20 indicaría que la sucursal es el 20% menos eficiente que la sucursal de mejores prácticas que opera en las mismas condiciones.

La Figura 8.3(d) muestra el diagrama de caja (*boxplot*) mes a mes, que resume la distribución de las estimaciones de ineficiencia-X y se muestra la línea media y la dispersión de los datos. El diagrama se delimita por el primer y tercer cuartil, representando el 50% central de los datos de eficiencia de costes. La diferencia entre estos cuartiles es el rango intercuartílico y es una representación gráfica simple de una distribución de probabilidad. La mediana está representada por una línea que pasa por el centro de la caja. Los puntos que representan datos fuera del intervalo son valores atípicos.

Como se puede observar, durante la mayoría de los periodos, las ineficiencias-X rondaron el 7.7%. Se puede observar un patrón similar con respecto a las estimaciones de ineficiencia de costes. En general, puede decirse que la tendencia de las ineficiencias con respecto a las sucursales de mejores prácticas habría aumentado al final del período de estudio.

## CAPÍTULO 9

# REESTRUCTURACIÓN FINANCIERA, CRISIS Y CRECIMIENTO DE LA PRODUCTIVIDAD EN LAS OFICINAS BANCARIAS

---

### 9.1. INTRODUCCIÓN

La crisis financiera de 2007 reveló importantes deficiencias en el funcionamiento del sistema bancario. A nivel mundial, la respuesta de los gobiernos, sometidos a una gran presión social, ha sido un tsunami regulatorio, con consecuencias especialmente complejas para el sector bancario. Como resultado, el sector bancario ha experimentado una importante reestructuración con respecto a su regulación, requisitos de capital y liquidez, y también en términos de relaciones con los clientes, con los proveedores de servicios financieros e instalaciones (por ejemplo, los bancos ahora ofrecen productos y servicios dentro del contexto de un entorno completamente nuevo, con elementos novedosos como Internet, teléfonos inteligentes y redes sociales).<sup>56</sup> Este nuevo contexto regulatorio ha tenido un doble impacto en la rentabilidad. Por un lado, ha impuesto el desapalancamiento, que socava la rentabilidad de los recursos y limita las posibilidades de crecimiento. Por otro lado, los costes operativos se han incrementado, principalmente en los campos de recursos humanos y de la tecnología, éste último para implementar

---

<sup>56</sup> Un informe reciente del BIS escrito por Buch y Dages (2018) sobre las tendencias en los modelos de negocios bancarios, el desempeño y la estructura del mercado durante la última década, y sobre sus implicaciones para la estabilidad y eficiencia de los mercados bancarios. Este estudio concluyó que los cambios estructurales importantes y necesarios se implementaron en el sector bancario mundial tras la crisis.



avances en la digitalización. Vale la pena señalar que la productividad se vio fuertemente afectada durante la crisis y la regulación posterior.

Desde entonces, los bancos españoles han reevaluado y ajustado de diversas formas sus estrategias y modelos comerciales. Las nuevas sucursales se han adaptado a la presencia de una competencia mucho mayor, a las presiones sobre los márgenes y a la cambiante era digital (bajo la premisa de que las sucursales deben ser impulsadas digitalmente, deben ser eficientes, capaces de mantener una ventaja competitiva y seguir siendo rentables involucrando a los clientes en cada etapa), y estas reformas han afectado el crecimiento de la productividad.

En este contexto, se puede evaluar el resultado de los procesos regulatorios, los procesos de desregulación y liberalización, así como el impacto de las crisis financieras o bancarias en la productividad y la eficiencia que varía en el tiempo. El análisis de productividad podría ayudar a los gerentes o responsables de políticas con una herramienta para evaluar el alcance y la relevancia de las estrategias de desarrollo que han sido implementadas por los bancos. Por un lado, el resultado de una política aplicada al diseño y desarrollo de servicios bancarios se puede medir a través de cambios en la productividad y eficiencia. Por otro lado, identificar los bancos menos productivos tras la caída de los márgenes y por ende, de la productividad y la demanda, ayudando a diseñar un proceso de reestructuración de la industria bancaria (por ejemplo, fusiones y adquisiciones de bancos menos productivos y eficientes), o iniciar un proceso de reorganización dentro de cada banco (es decir, reducir el tamaño de las sucursales bancarias o cerrar sucursales menos eficientes en el tiempo), entre otros.

Varios artículos han investigado el cambio de productividad en la industria bancaria, principalmente, enfocándose en el efecto de los procesos de regulación y desregulación,

incluido el efecto de la reciente crisis financiera global (ver la sección de revisión de la literatura para una descripción general). A pesar de que la red de sucursales suele ser la mayor fuente de gastos operativos (Paradi y Zhu, 2013),<sup>57</sup> cabe señalar que, para nuestro conocimiento, ningún trabajo ha examinado el comportamiento del cambio de productividad de las sucursales bancarias en un período de crisis e iniciativas regulatorias.

Por tanto, a diferencia de otros estudios recientes sobre bancos españoles (Martín-Vallespín et al., 2016), la principal contribución de nuestro trabajo en este capítulo es analizar el comportamiento del cambio de productividad de las sucursales bancarias en un período de estrictas normas regulatorias.<sup>58</sup> Hasta donde sabemos, no existen trabajos que hayan investigado el crecimiento de la productividad de las sucursales bancarias en un período de cambios regulatorios en España. Concretamente, investigamos si la crisis y la reestructuración financiera afectaron el comportamiento de la productividad a nivel de sucursales bancarias y observamos si la tecnología de producción cambió o se mantuvo constante después del período de reestructuración. Para hacer esto, distinguimos tres tipos de sucursales en términos de su nivel de eficiencia: las sucursales supereficientes de las completamente eficientes e ineficientes. Esto lo hacemos así, porque las oficinas supereficientes son un subconjunto de observaciones que se desvían significativamente del conjunto de datos completo. Es decir, aunque

---

<sup>57</sup> Berger y col. (1997) indicaron que la información sobre el desempeño de las sucursales mejora nuestra comprensión de la eficiencia bancaria y contribuye a resolver los problemas de medición en este contexto.

<sup>58</sup> Para la revisión de la literatura, situamos al lector en el Capítulo 5, donde se comentan algunos estudios relacionados con el análisis de los cambios en la productividad bancaria, en general (Berg et al., 1992; Fukuyama, 1995; Noulas, 1997; Leightner y Lovell, 1998; Gilbert y Wilson, 1998; Mukherjee et al., 2001; Canhoto y Dermine, 2003; Casu et al., 2004; Asmild et al., 2004; Guzmán y Reverte, 2008; Delis et al., 2011; Sánchez et al., 2013; Kumar et al., 2016; o Martín-Vallespín et al., 2016, entre otros), y los cambios en la productividad de las oficinas bancarias (Camanho y Dyson, 2006; y Asmild y Tam, 2007).

estas sucursales no son anómalas en sí mismas ni en un sentido contextual ni global, podrían considerarse valores atípicos (outliers) colectivos. En cualquier caso, su tratamiento es necesario si no queremos obtener estimaciones incorrectas de las eficiencias.<sup>59</sup> Para reforzar nuestros resultados, estudiamos dos dimensiones principales para el banco: la operativa (producción) y la rentabilidad de los servicios de sucursales bancarias (también denominada, ganancias).<sup>60</sup>

El análisis empírico utiliza datos anuales de las sucursales bancarias de un único banco español.<sup>61</sup> El período analizado abarca desde 2011 a 2014, y este es un periodo en el que las entidades financieras españolas tuvieron que adaptarse a un nuevo escenario macroeconómico, a los cambios regulatorios, a los efectos de una nueva recesión en la economía española en el primer trimestre de 2012 (la segunda recesión económica española), y a la reestructuración sustancial del sector bancario que tuvo lugar tras el rescate del sistema bancario español en 2012 a través del paquete de rescate de 100.000 millones de euros proporcionado por el Mecanismo Europeo de Estabilidad (MEDE).<sup>62</sup>

---

<sup>59</sup> La naturaleza determinista de la frontera implica que las sucursales supereficientes que están más allá de la frontera podrían causar graves distorsiones en las medidas de eficiencia para toda la población porque la frontera se construye utilizando observaciones extremas y el método DEA puede ser sensible a estos puntos extremos.

<sup>60</sup> Si bien los dos enfoques más utilizados al respecto son los de producción e intermediación, Paradi y Zhu (2013) utilizaron un tercer enfoque denominado rentabilidad, el cual se basa en el análisis de la capacidad de una sucursal para convertir sus gastos en ingresos. Cabrera-Suárez y Pérez-Rodríguez (2020) utilizaron los enfoques de rentabilidad y producción para evaluar la eficiencia de las sucursales bancarias.

<sup>61</sup> Es uno de los tres bancos más grandes del país en términos de activos y cuenta con una extensa red regional. Este banco es representativo a los efectos de nuestro estudio, dado el alto grado de concentración de la banca nacional en España. Se encuentra entre el núcleo de bancos fuertes que están bien administrados, suficientemente capitalizados y aparentemente resistentes a nuevas conmociones. Otras instituciones, más vulnerables, con una gran proporción de sus balances vinculados al sector inmobiliario, requirieron de una mayor reestructuración y recapitalización.

<sup>62</sup> Aunque se le atribuía el mérito de ser una de las economías occidentales más sólidas y mejor equipadas para hacer frente a la crisis de liquidez mundial, el sector bancario español se ha reestructurado profundamente desde la crisis bancaria de 2010 y el rescate de 2012 provocado por la burbuja inmobiliaria, que produjeron un impacto severo en la economía española y sus instituciones financieras.

En este período, la rentabilidad de los bancos disminuyó y era una práctica común para la administración, tratar de minimizar las ineficiencias en las operaciones bancarias, teniendo lugar, una reordenación del modelo de negocio tradicional. En nuestro análisis, el banco no se vio afectado por el rescate, aunque el escenario macroeconómico tuvo algún impacto en su productividad laboral interna y eficiencia operativa. Entre otros factores que explican este favorable desenlace, el banco contaba con un modelo de cartera diversificado y equilibrado, y se gestionaba con énfasis en la prudencia, la anticipación, la diversificación geográfica (que le permitió beneficiarse de diferentes ciclos de crecimiento), la gestión de riesgos, el uso efectivo de la tecnología y la anticipación de la digitalización.

La metodología utilizada permite evaluar las diferencias en los patrones de cambio de productividad a lo largo del tiempo y sus componentes y, se basa en el índice de productividad DEA-Malmquist (Caves et al., 1982). Sin embargo, para evitar resultados distorsionados debido a la sensibilidad a las características de la muestra en el estimador DEA-Malmquist, como consecuencia de sucursales supereficientes (valores atípicos), también estimamos el cambio de productividad en DEA-Malmquist eliminando los valores atípicos. Para hacer esto, y por simplicidad, usamos un procedimiento de dos pasos:<sup>63</sup> 1) estimamos las eficiencias utilizando un análisis de frontera no paramétrico

---

Bankia, un banco importante y el mayor prestamista hipotecario del país, fue nacionalizado el 9 de mayo de 2012; el 25 de mayo de 2012 anunció que requeriría un rescate de 23.500 millones de euros para cubrir las pérdidas de hipotecas fallidas. El rescate del sistema bancario español incluyó la toma de pérdidas por parte de los inversores de hasta 10.000 millones de euros, la creación de un “banco malo” para absorber hasta 45.000 millones de préstamos fallidos, el cierre de miles de sucursales bancarias y despidos generalizados.

<sup>63</sup> Es de destacar que el índice de productividad de Malmquist (MPI) ha sido empleado usando estimadores robustos (versiones *order-m* y *order-a*). Por ejemplo, con base en la descomposición de MPI introducida por Wheelock y Wilson (1999), el estudio de Wheelock y Wilson (2003) fue el primer estudio que aplicó los estimadores *order-m* para crear una versión robusta de MPI. Wheelock y Wilson (2013) iniciaron por primera vez el uso de un estimador *order-a* para la creación de medidas robustas (cuantiles) del MPI.

robusto denominado *order- $\alpha$*  (Aragon et al., 2005), al igual que Cabrera-Suárez y Pérez-Rodríguez (2020), para detectar posibles sucursales supereficientes en los datos (ver Tauchmann, 2012). Este método es más robusto que los métodos tradicionales cuando hay errores de medición y valores atípicos en la muestra. 2) Una vez que se han detectado y eliminado los valores atípicos de la muestra, se puede realizar un análisis de eficiencia no paramétrico DEA-Malmquist utilizando un panel de datos balanceado formado solo por sucursales completamente eficientes e ineficientes.

El resto del capítulo se estructura de la siguiente manera. La siguiente sección describe brevemente la literatura disponible en este campo. La Sección 9.3 describe el uso de DEA-Malmquist para determinar el cambio en la productividad de las sucursales bancarias. La sección 9.4 presenta los datos y los resultados empíricos obtenidos.

## **9.2. CAMBIOS DE PRODUCTIVIDAD EN EL TIEMPO: DEA-MALMQUIST**

El índice de productividad de Malmquist basado en la DEA, formulado por Caves et al. (1982), mide el cambio de productividad a lo largo del tiempo. Tiene dos componentes: uno mide el cambio técnico y el otro, el cambio de frontera. En otras palabras, la técnica del índice de Malmquist permite separar el cambio de eficiencia técnica (TE), es decir, el cambio en el tiempo en la eficiencia técnica de cada unidad de toma de decisiones (DMU) con respecto a la frontera de mejores prácticas, del cambio tecnológico (TEC), es decir, el cambio de la frontera de las mejores prácticas a lo largo del tiempo debido al progreso tecnológico.

En este contexto, el término orientación hacia los inputs se refiere a la reducción equiproporcional de inputs, con respecto a un nivel dado de producción. El índice de

productividad de Malmquist orientado hacia los inputs también se puede utilizar para descomponer este cambio de productividad total entre los dos períodos en TEC y TE. Berg et al. (1992), Fukuyama (1995) y Worthington (1999) adoptaron este enfoque para medir la eficiencia y la productividad en la banca.

Sea  $n$  el número de observaciones de sucursales (o, en general, DMUs),  $j = 1, 2, \dots, n$ . Sea  $x_{ij}$  el nivel observado del  $i$ -ésima input para la DMU  $j$  y sea  $x = (x_{1j}, \dots, x_{kj}) \in \mathfrak{R}_+^k$  el vector de  $k$  inputs para la  $i$ -ésima unidad;  $y_{rj}$  es el nivel observado del  $r$ -th output para la DMU  $j$ , donde  $y = (y_{1j}, \dots, y_{mj}) \in \mathfrak{R}_+^m$  es el vector de  $m$  outputs.

El índice de Malmquist está basado en la función distancia,  $D(y, x)$ , donde  $y$  es un vector de outputs y  $x$  es un vector de inputs. Siguiendo a Grosskopf (1993), Färe et al. (1994) y Worthington (1999) utilizamos el índice de cambio de productividad de Malmquist orientado hacia los inputs.<sup>64</sup> En este análisis se utilizaron dichos modelos orientados hacia los inputs porque la cantidad de negocios disponibles para una sucursal depende en gran medida de la demanda de los clientes por los servicios del banco y está más allá del control del gerente de la sucursal (Paradi et al., 2011).

En la orientación input,  $D_I(y, x) = \max \{ \lambda : x/\lambda \in L(y) \}$ , donde  $L(y)$  es el conjunto de inputs de la tecnología de producción. Más aún,  $y$  puede ser producido con poca cantidad de inputs,  $x/\lambda$ ,  $\lambda > 1$  (Kumbhakar and Lovell, 2000).

El índice de productividad total de los factores (TFP) se define por:

---

<sup>64</sup> Cabe destacar que en la Sección 3.3.4.3 (Capítulo 3) se describió el modelo orientado hacia los outputs.

$$TFP_I(x_{t+1}, y_{t+1}, x_t, y_t) = \sqrt{\frac{D_I^t(x_{t+1}, y_{t+1})}{D_I^t(x_t, y_t)} \times \frac{D_I^{t+1}(x_{t+1}, y_{t+1})}{D_I^{t+1}(x_t, y_t)}}$$

donde  $TFP_I(x_{t+1}, y_{t+1}, x_t, y_t)$  representa la productividad del punto de producción  $(x_{t+1}, y_{t+1})$  relativo al punto de producción  $(x_t, y_t)$ , y donde  $D_I^t(x_t, y_t)$ ,  $D_I^t(x_{t+1}, y_{t+1})$ ,  $D_I^{t+1}(x_{t+1}, y_{t+1})$  y  $D_I^{t+1}(x_t, y_t)$  son funciones distancia en diferentes periodos temporales.

Siguiendo a Färe et al. (1993), una expresión equivalente del índice de Malmquist se define por:

$$TFP_I(y_{t+1}, x_{t+1}, y_t, x_t) = TE_I \times TEC_I$$

donde  $TE_I = \frac{D_I^{t+1}(y_{t+1}, x_{t+1})}{D_I^t(y_t, x_t)}$  representa el cambio de eficiencia técnica y

$TEC_I = \sqrt{\frac{D_I^t(y_{t+1}, x_{t+1})}{D_I^{t+1}(y_{t+1}, x_{t+1})} \times \frac{D_I^t(y_t, x_t)}{D_I^{t+1}(y_t, x_t)}}$  es el cambio tecnológico. Éste último, a su

vez, se puede dividir en cambio técnico puro (PE) y cambio de eficiencia de escala (SE). Por lo tanto, podemos escribir que

$$TFP_I(y_{t+1}, x_{t+1}, y_t, x_t) = TE_I \times TEC_I = (PE_I \times SE_I) \times TEC_I.$$

La interpretación de los resultados debemos hacerla de la siguiente manera. Si  $TFP > 1$  se produce un aumento de la productividad, mientras que si la  $TFP < 1$  se produce una pérdida de productividad. La eficiencia técnica aumenta (disminuye) si y solo si  $TE$  es mayor (menor) que uno. Una interpretación del cambio tecnológico es que el progreso técnico (regresión) ha ocurrido si  $TEC$  es mayor (menor) que uno. Si  $TE > TEC$ , las ganancias de productividad son en gran medida el resultado de mejoras en la eficiencia, mientras que si  $TE < TEC$  las ganancias de productividad son principalmente el resultado del progreso tecnológico. Dado que  $TE = PE \times SE$ , si  $PE > SE$  entonces la

principal fuente de cambio de eficiencia (tanto aumento como disminución) es la mejora en la eficiencia técnica pura, mientras que si  $PE < SE$  la principal fuente de eficiencia es una mejora en la eficiencia de la escala.

Siguiendo la estructura de problemas lineales de Coelli (1996) pero adaptándolos al marco de orientación de input siguiendo a Worthington (1999), las cantidades dentro de la fórmula se pueden calcular resolviendo cuatro programas lineales para cada sucursal bancaria, de la siguiente manera:

**Problema 1**

$$\left[ D'_I(y_t, x_t) \right]^{-1} = \min_{\theta, \lambda} \theta$$

sujeto a :

$$-y_{r0,t} + \sum_{j=1}^n y_{rj,t} \lambda_j \geq 0$$

$$\theta x_{i0,t} - \sum_{j=1}^n x_{ij,t} \lambda_j \geq 0$$

$$\lambda_j \geq 0, \forall j$$

**Problema 2**

$$\left[ D^{t+1}_I(y_{t+1}, x_{t+1}) \right]^{-1} = \min_{\theta, \lambda} \theta$$

sujeto a :

$$-y_{r0,t+1} + \sum_{j=1}^n y_{rj,t+1} \lambda_j \geq 0$$

$$\theta x_{i0,t+1} - \sum_{j=1}^n x_{ij,t+1} \lambda_j \geq 0$$

$$\lambda_j \geq 0, \forall j$$

**Problema 3**

$$\left[ D'_I(y_{t+1}, x_{t+1}) \right]^{-1} = \min_{\theta, \lambda} \theta$$

sujeto a :

$$-y_{r0,t+1} + \sum_{j=1}^n y_{rj,t} \lambda_j \geq 0$$

$$\theta x_{i0,t+1} - \sum_{j=1}^n x_{ij,t} \lambda_j \geq 0$$

$$\lambda_j \geq 0, \forall j$$

**Problema 4**

$$\left[ D^{t+1}_I(y_t, x_t) \right]^{-1} = \min_{\theta, \lambda} \theta$$

sujeto a :

$$-y_{r0,t} + \sum_{j=1}^n y_{rj,t+1} \lambda_j \geq 0$$

$$\theta x_{i0,t} - \sum_{j=1}^n x_{ij,t+1} \lambda_j \geq 0$$

$$\lambda_j \geq 0, \forall j$$

donde todos los sub-índices  $0,t$  indican que la DMU está siendo evaluada en el periodo  $t$ , y  $\theta$  es la eficiencia.



En los programas lineales anteriores, se consideran los rendimientos constantes a escala (CRS). Sin embargo, al formular retornos variables a escala (VRS), la restricción de convexidad  $\sum_{i=1}^n \lambda_j = 1$  debería ser añadida. Las distancias (o eficiencias técnicas) necesarias para el índice de Malmquist corresponden a las fronteras CRS DEA del período anterior, actual y siguiente y la frontera VRS del período actual.  $\theta$  y  $\lambda$  puede tomar diferentes valores.<sup>65</sup>

A pesar de la mayor flexibilidad del marco no paramétrico para medir la productividad, existen ciertas debilidades. Aunque el mayor inconveniente es su naturaleza determinista, las mediciones de eficiencia y productividad basadas en DEA también pueden introducir otros inconvenientes importantes. Por ejemplo, este puede introducir un sesgo atribuible al error de medición (Assaf y Tsionas, 2018), o mostrar sensibilidad a las características de la muestra, como valores extremos y valores atípicos, lo que puede influir en los índices de productividad de Malmquist estimados junto con sus componentes (Daraio y Simar, 2007). Por lo tanto, debido a que los estimadores de productividad basados en DEA tienen más probabilidades de sufrir un sesgo de medición potencial atribuido a las características de la muestra, nuestro estudio también aplica el DEA-Malmquist a la submuestra formada tras eliminar las oficinas supereficientes detectadas por el método de frontera parcial basado en el estimador *order- $\alpha$* .<sup>66</sup> Este estimador lo denominaremos DEA-Malmquist basado en outliers.

---

<sup>65</sup> Para una revisión completa de la literatura sobre los desarrollos teóricos y las aplicaciones del índice de Malmquist, puede consultarse Färe et al. (1998).

<sup>66</sup> Las expresiones utilizadas para el estimador *order- $\alpha$*  pueden consultarse en la Sección 3.5 (Capítulo 3) o en la Sección 7.3.2 (Capítulo 7).

### **9.3. ANÁLISIS EMPÍRICO**

#### **9.3.1. DATOS**

En este estudio recordamos que se recopiló información contable anual a partir de las cuentas de pérdidas y ganancias, junto con datos de departamentos internos (como Control de Gestión, Recursos Humanos e Ingeniería), para el período 2011 a 2014, correspondientes a 131 sucursales de un gran banco comercial (anónimo) en España. Estas sucursales tratan con personas físicas y personas jurídicas a través de cuentas de negocios, siendo sus actividades homogéneas, es decir, se prestan servicios similares en ambos casos. Las sucursales están ampliamente distribuidas por una región de España, y están ubicadas en áreas rurales y urbanas, algunas de las cuales tienen una alta concentración de sucursales. Por razones de confidencialidad, no se proporciona ninguna identificación adicional.

La Tabla 9.1 muestra los estadísticos descriptivos de las variables input y output utilizadas para construir el índice de Malmquist-TFP utilizando los enfoques de producción y ganancias.<sup>67</sup> Algunas de estas variables representan datos de recuento y otras son valores monetarios.

---

<sup>67</sup> Un tema importante que debe abordarse en la construcción de un modelo para evaluar la eficiencia bancaria es el de identificar inputs y outputs apropiados (Camanho y Dyson, 1999). Cabe destacar que los inputs y outputs utilizados en este estudio se emplean, generalmente, en la literatura empírica sobre bancos y sucursales bancarias.

**Tabla 9.1.** Descriptivos de todas las sucursales analizadas en el periodo 2011-2014.

<b>Variable</b>	<b>Media</b>	<b>Desviación estándar</b>
<b>Inputs</b>		
Superficie de la oficina (en m <sup>2</sup> )	322.56	292.26
Número de empleados	3.67	2.47
Número de cajeros automáticos (ATMs)	2.83	1.36
Costes operativos <sup>a</sup>	155,543	119,732
Gastos en intereses (coste de los depósitos)	266,649.7	279,959.9
Gastos laborales	267,451.4	150,539.2
<b>Outputs</b>		
Valor total de los depósitos (en miles de euros)	11,646	13,532
Valor total de los préstamos (en miles de euros)	37,189	45,674
Margen bruto (ingresos de intereses neto + comisiones)	1,467,339	1,139,147

*Nota:* (a) Usado por Camanho y Dyson (1999 y 2006) en la aproximación de producción.

Los inputs considerados fueron la superficie de la sucursal (en m<sup>2</sup>), el número total de empleados (incluidos los gerentes de sucursales y fuerzas de venta especializadas como gestores pymes, asesores financieros, gestores de banca privada, gestores comerciales, personal administrativo y cajeros), el número de cajeros automáticos externos, los costes operativos soportados por la sucursal (gastos totales excluidos los gastos de personal), gastos por intereses (gastos de depósito) y gastos laborales. Los outputs considerados fueron el valor total de los préstamos (la suma de las hipotecas, los valores de los préstamos personales de consumidores y préstamos empresariales) en miles de euros, el valor total de los ahorros (depósitos en cuentas) en miles de euros y el margen bruto (margen de intereses + ingresos sin intereses (comisiones)). Todos los datos de valor están en euros constantes de 2011 (calculados a partir del índice de precios al consumidor del mercado local con respecto al período base 2011).

En promedio, cada sucursal tenía cuatro empleados y una superficie total de 323 m<sup>2</sup>. El valor total de sus depósitos y préstamos fue de 11.646 y 37.189 miles de euros, respectivamente.

### **9.3.2. DIMENSIONES DEL DESEMPEÑO DE LA SUCURSAL BANCARIA**

Desde un punto de vista metodológico, el desempeño de las sucursales generalmente se evalúa utilizando los enfoques de producción, intermediación y ganancias/rentabilidad. Sin embargo, en nuestro caso, y dada la disponibilidad de información, usaremos los enfoques de producción y ganancias.

En el enfoque de producción, el énfasis está en la actividad operativa; en otras palabras, los bancos son vistos como proveedores de servicios a los clientes. En esta dimensión, como hemos visto anteriormente en otros modelos, los inputs que se incluyen suelen ser variables físicas relacionadas con los costes operativos (como mano de obra, materiales, espacio y sistemas de información). Por ejemplo, el presente estudio incluye el número total de empleados, el número de cajeros automáticos externos y el espacio de la oficina (en m<sup>2</sup>). Sin embargo, siguiendo a Camanho y Dyson (1999, 2006), también empleamos los costes operativos como inputs y el valor total de los ahorros y préstamos como outputs. Estas variables también fueron utilizadas por Giokas (2008). Sin embargo, como limitación, nuestro estudio no incluye la actividad de la oficina medida por el número total de transacciones, debido a la información incompleta (los datos de número sólo estaban disponibles desde 2013).

En el enfoque de ganancias, se reflejan los costes generales para generar ganancias, mostrando cómo cada sucursal usa sus recursos (inputs) para generar ingresos tales

como ingresos por intereses y otros ingresos (comisiones).<sup>68</sup> En el presente estudio, analizamos la eficiencia considerando los gastos de personal, gastos operativos y de intereses para generar ingresos por intereses y comisiones (Oral y Yolalan, 1990). La variable output utilizada es el margen bruto obtenido por la sucursal, es decir, los ingresos totales - los ingresos netos por intereses más los ingresos que no son de intereses (comisiones) -.<sup>69</sup>

### **9.3.3. PRODUCTIVIDAD TOTAL DE LOS FACTORES PARA TODAS LAS SUCURSALES EN EL PERÍODO 2011-2014**

En esta sección, se estudia el desempeño de las sucursales bancarias a lo largo del tiempo utilizando el índice DEA-Malmquist, que evalúa el crecimiento de la productividad y distingue entre cambio tecnológico y cambio de eficiencia técnica. Como se ya se indicó, en este estudio usamos la tecnología orientada hacia los inputs para la evaluación de eficiencia,<sup>70</sup> aplicando los enfoques de producción y ganancias.

Examinando el índice de TFP en el período 2011-2014 (ver Tabla 9.2), y obteniendo el desglose en cambio tecnológico (cambio de frontera) y cambio de eficiencia técnica (efecto *catch-up*), en general, los valores del índice son mayores que uno, lo que indica

---

<sup>68</sup> Este enfoque se centra en cómo las sucursales bancarias convierten los gastos (aquellos sujetos a la influencia directa de la administración de la sucursal) en ingresos.

<sup>69</sup> El margen bruto se mide como la suma de los OPOs y el margen de intermediación (margen de intereses). Los OPOs se definen como los cargos que imputa el banco por la prestación de todo tipo de servicios (comisiones totales + diferencias de cambio + resto de transacciones financieras). El margen de interés es la diferencia entre el interés cobrado por las inversiones crediticias y financieras realizadas (Activo) y el interés pagado a los clientes y otros prestamistas por el dinero depositado en el banco (Pasivo).

<sup>70</sup> También hemos calculado la eficiencia orientada hacia los outputs, pero los resultados obtenidos son ligeramente diferentes a los resultados obtenidos con orientación hacia los inputs, motivo por el que no se muestran.

crecimiento de la productividad, progreso tecnológico y cambio de eficiencia. Concretamente, la variación de la TFP se mantuvo positiva durante el período 2011-2014 y el valor medio fue superior a uno, en ambos enfoques. Por ejemplo, la productividad de las sucursales aumentó un 74,3% (producción) y un 96,9% (ganancias) en promedio. Se trata de valores elevados para las sucursales bancarias que indican un alto desempeño de las sucursales durante el período.

En cuanto a los componentes de la TFP, el cambio tecnológico es más importante que la eficiencia técnica en ambos enfoques destacando la importancia de las reformas tecnológicas realizadas por el banco. De hecho, durante el período de estudio, las sucursales bancarias experimentaron importantes e innovadoras reformas, con la introducción de ofertas multicanal en las que los servicios bancarios tradicionales se complementaron con enfoques novedosos como Internet, teléfonos inteligentes y otros servicios multimedia con orientación a la digitalización. Se reformaron los productos y servicios, con una segmentación basada en las preferencias y perfil de los clientes (requisitos bancarios, ubicación, productos, servicios preferidos y comportamiento histórico), combinados con estrategias de innovación de las sucursales (por ejemplo, permitiendo la visibilidad entre canales de los registros de los clientes para desarrollar programas de lealtad), permitiendo a los clientes especificar productos y servicios personalizados, y agregando valor a las actividades de la sucursal), y estrategias de promoción de ventas como posicionar la sucursal como un canal de ventas al acercar al personal de primera línea a los clientes. Como resultado de estas innovaciones, los índices de cambio tecnológico (TEC) fueron siempre superiores a uno en ambos enfoques, con incrementos acumulados de 53% (producción) y 78,2% (ganancias) en el período global.

En cuanto al efecto *catch-up*, fue positivo (índices superiores a uno) durante 2011-2014 y la eficiencia técnica acumulada aumentó en torno al 13,9% (producción) y el 10,5% (ganancias), aunque la tasa de crecimiento descendió, bajo ambos enfoques. Con respecto a la descomposición del cambio en TE, la eficiencia pura es más relevante que el cambio de eficiencia de escala. Durante el período de cuatro años, el cambio en la eficiencia pura (8,9%) fue mayor que en la eficiencia de escala (4,6%) en el enfoque de producción, lo que indica que la inversión del banco en factores organizativos y prácticas de gestión mejoró la eficiencia pura, en lugar de, la eficiencia de escala. Sin embargo, en el enfoque de ganancias para el período general, el cambio de eficiencia pura (-4.2%) fue menor que el de la eficiencia de escala (15.3%).

Sin embargo, ¿cuál fue el efecto en las sucursales bancarias entre subperíodos?

En general, estas cifras muestran que la regulación bancaria puede no haber afectado tanto a la actividad operativa como a los ingresos. Las sucursales bancarias mejoraron su experiencia técnica, administración y organización (y emplearon estrategias para aumentar las ventas a través de un marketing más efectivo y/o mejorando la calidad del servicio) a pesar de la crisis y la regulación bancaria de 2012, y estos esfuerzos fueron suficientes para anular la desaceleración general. De hecho, en algunos aspectos (como la eficiencia técnica) las variaciones registradas fueron positivas. Martín-Vallespín et al. (2016) mostraron que la mayoría de los bancos de su muestra (incluido nuestro Banco) mejoraron su eficiencia en comparación con 2009, lo que indicó que el proceso de integración bancaria fue beneficioso para la eficiencia del sistema financiero. Estos autores han señalado que comparando los resultados DEA para 2009 y 2013, se encuentra que el ordenamiento de los bancos es bastante similar en relación a los resultados de 2009. Este hecho indica que, a pesar de la profundidad de las reformas

realizadas en la mayoría de las instituciones financieras, el sistema financiero mantuvo un equilibrio similar al del año 2009.

**Tabla 9.2.** Índice de productividad DEA-Malmquist eliminando los valores atípicos y sus componentes por subperíodos y período general. Resumen de medias anuales.

Años	Productividad total de los factores (TFP)	Cambio tecnológico (TEC)	Cambio de eficiencia técnica (TE)	Cambio de eficiencia pura (PE)	Cambio de eficiencia de escala (SE)
<b>Panel A: Dimensión de producción</b>					
<i>Periodo de crisis y reforma</i>					
2012	1.607	1.044	1.539	1.310	1.175
<i>Periodo de post-crisis y reforma</i>					
2013	1.867	1.926	0.97	1.024	0.947
2014	1.745	1.698	1.028	0.985	1.043
Contraste de Kolmogorov-Smirnov para 2012 vs. 2013-2014	0.1642 [0.02]	0.9067 [0.00]	0.3843 [0.01]	0.2724 [0.00]	0.3358 [0.00]
<i>Periodo total</i>					
2011-2014	1.743	1.530	1.139	1.089	1.046
Número de oficinas			131		
<b>Panel B: Dimensión de ganancias</b>					
<i>Periodo de crisis y reforma</i>					
2012	2.293	1.605	1.429	0.962	1.485
<i>Periodo de post-crisis y reforma</i>					
2013	2.019	1.827	1.105	1.059	1.043
2014	1.712	1.878	0.911	0.864	1.055
Contraste de Kolmogorov-Smirnov para 2012 vs. 2013-2014	0.3694 [0.00]	0.5709 [0.00]	0.3955 [0.01]	0.1679 [0.01]	0.6418 [0.00]
<i>Periodo total</i>					
2011-2014	1.969	1.782	1.105	0.958	1.153
Númreo de oficinas			131		

**Notas:** Esta tabla muestra las medias anuales DEA-Malmquist (medias geométricas) para la TPF y sus componentes para el período general 2011-2014, y el período de crisis y reforma (año 2012) y el período posterior a la crisis y reforma (años 2013 y 2014), respectivamente. Es de destacar que el número de sucursales supereficientes es bastante elevado (alrededor del 50%). Por lo tanto, el número de sucursales eliminado los valores atípicos basados en DEA-Malmquist es bajo. La prueba de Kolmogorov-Smirnov es la prueba de Kolmogorov-Smirnov de dos muestras para la igualdad de funciones de distribución para 2012 y el período 2013-2014 (ambos inclusive). Entre paréntesis aparecen los p-valores.



### **9.3.4. EFECTOS DE LA CRISIS Y LA REESTRUCTURACIÓN FINANCIERA EN LAS SUCURSALES BANCARIAS**

En esta sección, evaluamos si la crisis y el período de reforma en 2012 afectaron el cambio de productividad y sus componentes en términos de las distribuciones empíricas para sus índices. Dado que deseamos analizar el efecto de las oficinas supereficientes; distinguiremos dos situaciones. Por un lado, consideramos todas las oficinas. Por otro lado, descartamos las sucursales supereficientes (valores atípicos u *outliers*) para evaluar el impacto en las sucursales bancarias totalmente eficientes e ineficientes.

#### **9.3.4.1. TODAS LAS SUCURSALES**

En esta subsección, analizamos el cambio de productividad de todas las oficinas analizando dos subperíodos: pre (año 2012) y post crisis y períodos de reforma (años 2013-2014) (véase también la Tabla 9.2). Los Gráficos 9.1 y 9.2 presentan una imagen completa de la TFP y sus diferentes componentes a través de funciones de densidad *kernel* tanto en el enfoque de producción como de ganancias. Estos gráficos contienen el índice de productividad total de los factores (TFP) y su desglose en cambio tecnológico (TEC) y cambio de eficiencia técnica (TE), por un lado, y el de TE en cambio de eficiencia de escala (SE) y cambio técnico puro (PE). También se utilizó la prueba de Kolmogorov-Smirnov de dos muestras para contrastar la igualdad de funciones de distribución, comparando la distribución para el período 2012 y el período 2013-2014 (ambos inclusive).

Los principales resultados son los siguientes:

- 1) Con respecto a la TFP, podemos observar ligeras diferencias entre las densidades estimadas para ambos subperíodos en los Gráficos 9.1(a) y 9.2(a). Además, la prueba de Kolmogorov-Smirnov (Tabla 9.2) indica que las estimaciones de densidad de TFP para el período de reforma y crisis, y el período posterior a la crisis y el período de reforma no son similares a un nivel de significancia del 5% para ambos enfoques. También cabe mencionar que la TFP se mueve ligeramente hacia la derecha y hacia abajo en el período posterior a la crisis y la reforma (la tendencia de la TFP está aumentando) en el enfoque de producción, pero se desplaza ligeramente hacia la izquierda y hacia arriba en el enfoque de ganancias (la tendencia de la TFP está disminuyendo). Como podemos observar en la Tabla 9.2, cuando observamos la tendencia de las medias geométricas anuales de 2012 a 2014, las ganancias de productividad disminuyeron en el enfoque de ganancias, pero aumentaron en el enfoque de producción. Por ejemplo, en el enfoque de beneficios, la TFP cayó (de 2.3 a 1.7).
- 2) Sin embargo, las figuras de densidad muestran diferencias más claras en sus componentes. Además, los resultados de la prueba de Kolmogorov-Smirnov indican que las distribuciones no son iguales. Por ejemplo,
  - a. El cambio tecnológico se desplaza hacia la derecha en el período posterior a la crisis y la reforma, aumentando las tasas de crecimiento de acuerdo con los enfoques de producción y de ganancias. Además, con el enfoque de producción, el índice de cambio tecnológico disminuyó para todos los períodos, mientras que con el enfoque de ganancias aumentó durante el período 2013-2014 (ver Tabla 9.2).
  - b. El cambio de eficiencia técnica es positivo para muchas oficinas. De acuerdo con los resultados del enfoque de producción, estas sucursales bancarias no

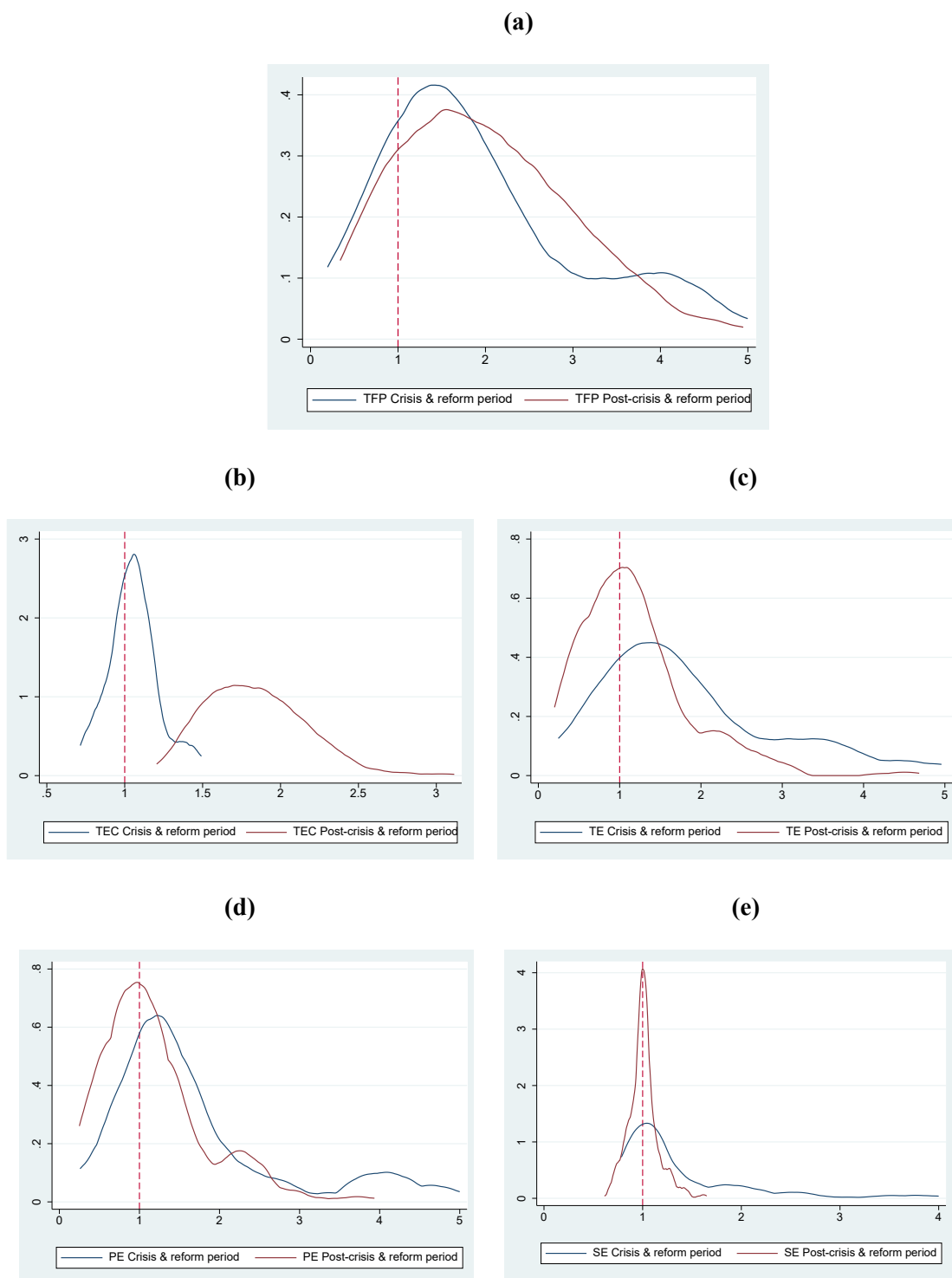
mejoraron su difusión de tecnología de mejores prácticas en 2012-2013, debido a que el cambio de TE para este período es igual a -3%. Bajo el enfoque de ganancias, el mismo índice fue menor a uno e igual a 0.91 para 2013-2014, lo que indica variaciones negativas. Nuestro desglose de los hallazgos de TE muestra los siguientes resultados de observación en la Tabla 9.2:

- i. También es de destacar que existen índices negativos como PE y SE en el período post-crisis y reforma 2013-2014 para la producción (-1,5% y -4,3%, respectivamente), y PE para el beneficio (-13,6%). Estos resultados indican que algunas sucursales bancarias no lograron una tasa positiva de cambio de eficiencia, lo que es indicativo de problemas potenciales en la gestión de la actividad de las sucursales en estos casos debido a las condiciones del mercado. Cabe señalar que en años posteriores a 2014, estos hechos fueron detonantes para que comenzara el cierre de aquellas oficinas menos eficientes en el entorno convulso donde los niveles de ingresos eran más difíciles de conseguir (por ejemplo, deterioro del margen y mayor competencia).
- ii. El cambio de eficiencia de escala muestra una reducción en su variabilidad en el segundo subperíodo como consecuencia del ajuste en la escala de sucursales.

Por tanto, podemos decir que la crisis y la reestructuración financiera afectaron, principalmente, a las sucursales bancarias tanto en términos de los cambios tecnológicos

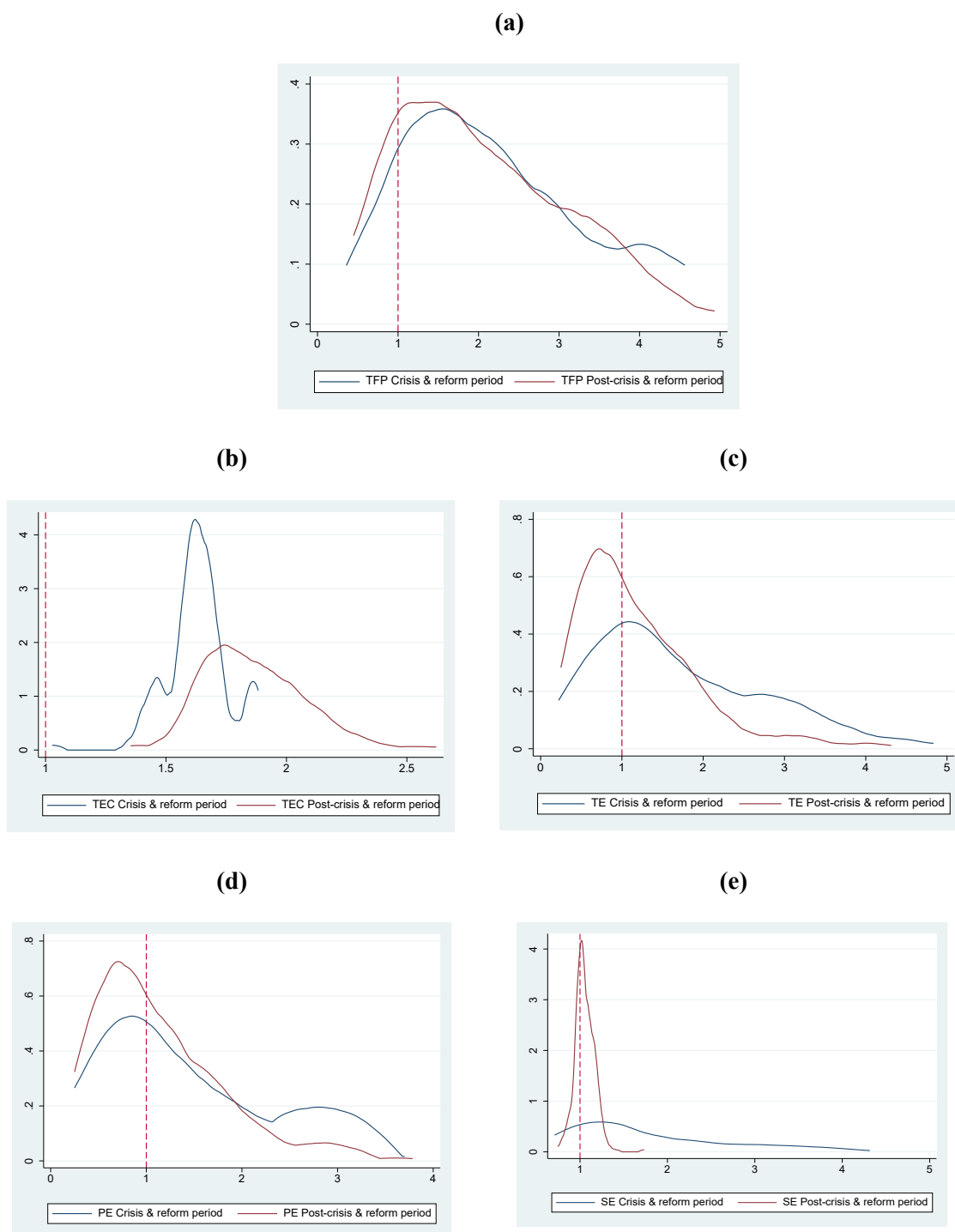
y de eficiencia con una clara reducción en las tasas de crecimiento de la eficiencia, y menos en el cambio de productividad.

**Figura 9.1.** Gráficos de densidad del índice de productividad estimado de Malmquist y sus componentes por subperiodos. Enfoque de producción.



**Nota:** La línea vertical de trazos largos de color rojo indica la unidad.

**Figura 9.2.** Gráficos de densidad del índice de productividad estimado de Malmquist y sus componentes por subperíodos. Enfoque de rentabilidad.



**Nota:** La línea vertical de trazos largos de color rojo indica la unidad.

Sin embargo, ¿estos resultados están condicionados a la presencia de sucursales muy bien gestionadas (sucursales supereficientes)?

### **9.3.4.2. SUCURSALES COMPLETAMENTE EFICIENTES E INEFICIENTES**

En esta sección, nos centramos en los valores atípicos y el modelo DEA-Malmquist para calcular TFP y sus componentes tanto en los dos subperíodos como en los dos enfoques. En este sentido, solo consideramos las oficinas plenamente eficientes y no eficientes (es decir, oficinas no supereficientes) para evitar la influencia de observaciones extremas en la estimación de los índices DEA-Malmquist.

Para distinguir las oficinas eficientes e ineficientes de las supereficientes (valores atípicos), utilizamos el método no paramétrico robusto denominado *order- $\alpha$* . Es un análisis de frontera parcial que permite detectar posibles valores atípicos en los datos. En particular, seleccionamos el valor  $\alpha = 95$  en dicho método de estimación para todos los años, a diferencia de Cabrera-Suárez y Pérez-Rodríguez (2020) quienes utilizaron un procedimiento automático para seleccionarlo. Para hacer una muestra equilibrada de sucursales para el período de cuatro años en el análisis DEA-Malmquist basado en outliers, hemos decidido hacer lo siguiente. Mantenemos la oficina en la muestra si la eficiencia *order- $\alpha$*  es menor e igual a 1 para todos los años. De lo contrario, hemos eliminado estas oficinas. Por lo tanto, la muestra de sucursales para DEA-Malmquist basado en outliers es menor que la muestra original, tanto en los enfoques de producción como de ganancias. Es de destacar que estas muestras han incluido solo oficinas que son completamente eficientes e ineficientes durante el período muestral (o sea, hemos excluido las oficinas supereficientes detectadas por el método *order- $\alpha$* ). Por ejemplo, la submuestra de 65 sucursales se utiliza en el caso del enfoque de producción y 72 sucursales en el enfoque de ganancias. Cabe señalar que alrededor del 50% de las sucursales son supereficientes en el enfoque de producción y el 45% de las sucursales

en el enfoque de ganancias, según el método robusto no paramétrico *order- $\alpha$* .<sup>71</sup> La Tabla 9.3 muestra los resultados medios anuales para los subperíodos y el período general.

Las Figuras 9.3 y 9.4 representan la densidad de la TFP y sus diferentes componentes en DEA-Malmquist para los dos subperíodos en los enfoques de producción y rentabilidad eliminando los valores atípicos. En general, podemos observar diferencias más importantes entre los dos subperíodos que en las Figuras 9.1 y 9.2 que consideran la existencia de oficinas supereficientes. Los resultados de la prueba de Kolmogorov-Smirnov (Tabla 9.3) indican, en la mayoría de los casos, el rechazo de la hipótesis nula de igualdad de distribuciones. Además, observamos los siguientes patrones. Por un lado, en todos los casos, podemos observar que las densidades *kernel* son ligeramente más leptocúrticas en el período de poscrisis y reforma que en el período de crisis. Una implicación de esta característica estadística es que menos oficinas han experimentado valores de índice más altos y se concentran alrededor de la mediana de la distribución. Por otro lado, en la mayoría de los casos, las distribuciones se han desplazado hacia la izquierda. Por tanto, podemos interpretar que la crisis y la reestructuración financiera afectaron la productividad y a sus componentes en el segundo subperíodo, provocando una menor variabilidad de resultados en las sucursales bancarias y reduciendo sus tasas de crecimiento.

---

<sup>71</sup> Estos resultados son diferentes a los de Cabrera-Suárez y Pérez-Rodríguez (2020), aunque mantienen una alta proporción de sucursales supereficientes. Estos son porcentajes inusuales de valores atípicos. Cabe destacar que también se encontraron resultados similares utilizando otros estimadores robustos como *order-m*. Una posible causa podría estar relacionada con el uso de un número elevado de inputs y outputs. De hecho, la reducción en el número de outputs conduce a porcentajes alrededor del 7-10% de los valores atípicos. Sin embargo, se ha preferido mantener estos inputs y outputs porque son los utilizados en la mayoría de los trabajos empíricos sobre sucursales bancarias.

En particular, centrándose en TFP, las Figuras 9.3(a) y 9.4(a) muestran que durante el primer subperíodo, la cantidad de sucursales que tenían niveles de TFP por encima o por debajo de uno (representado por la línea vertical) estaba bastante desequilibrada (en su mayoría, sesgadas a la derecha), predominando las oficinas con TFP por encima de la unidad en ambos enfoques. De hecho, la mayor masa de niveles de productividad de las sucursales estaba por encima de uno, lo que indica un aumento de la productividad en ambos enfoques. En general, podemos concluir que la mayoría de los niveles de productividad de las sucursales se identificaron en los dos períodos, pero la TFP se vio realmente afectada por el período posterior a la crisis y la reforma.

Las Figuras 9.3(b) y 9.4(b) también muestran que los niveles de cambio tecnológico (TEC) están por encima de la unidad. El cambio tecnológico afectó a la mayoría de las sucursales tanto en el enfoque productivo como en el de rentabilidad en los períodos de reforma y posreforma. Pero, especialmente, afectó el enfoque de ganancias que tuvo valores por encima de la unidad en el período de crisis y por debajo de la unidad en el período de poscrisis y reformas. En resumen, los niveles tecnológicos de las sucursales disminuyeron.

Las distribuciones de los niveles de cambio de eficiencia de las sucursales (*catch-up*, TE) durante los dos subperíodos mostrados en los Gráficos 9.3(c) y 9.4(c) muestran cambios importantes en las densidades del enfoque producción.

Las Figuras 9.3(d) y 9.4(d) muestran la descomposición del cambio de eficiencia técnica en cambio técnico puro. Las Figuras 9.3(e) y 9.4(e) muestran cambios en la eficiencia de la escala. Podemos observar, para estas últimas cifras con respecto al cambio de eficiencia de escala que muchas sucursales pasaron de estar por encima de la unidad en el período de crisis a estar por debajo de la unidad en el período de poscrisis y



reforma, detonante para las reestructuraciones de la red comercial que se iniciaron en los años posteriores.

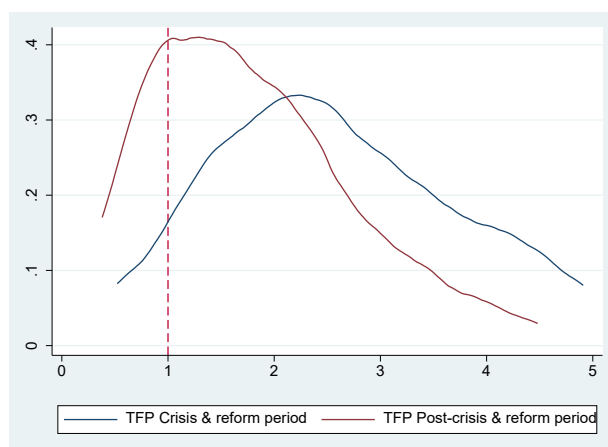
**Tabla 9.3.** Índice de productividad DEA-Malmquist eliminando los valores atípicos y sus componentes por subperíodos y período general. Resumen de medias anuales.

Años	Productividad total de los factores (TFP)	Cambio tecnológico (TEC)	Cambio de eficiencia técnica (TE)	Cambio de eficiencia pura (PE)	Cambio de eficiencia de escala (SE)
<b>Panel A: Dimensión de producción</b>					
<i>Período de crisis y reforma</i>					
2012	2.329	1.524	1.528	1.269	1.204
<i>Período de post-crisis y reforma</i>					
2013	1.765	1.344	1.313	1.180	1.113
2014	1.370	1.328	1.032	1.061	0.973
Contraste de Kolmogorov-Smirnov para 2012 vs. 2013-2014	0.3538 [0.00]	0.3846 [0.00]	0.2462 [0.01]	0.2615 [0.00]	0.3538 [0.00]
<i>Período total</i>					
2011-2014	1.779	1.396	1.274	1.167	1.092
Númreo de oficinas			65		
<b>Panel B: Dimensión de ganancias</b>					
<i>Período de crisis y reforma</i>					
2012	2.906	2.043	1.423	1.163	1.224
<i>Período de post-crisis y reforma</i>					
2013	1.537	1.309	1.174	1.178	0.996
2014	1.807	1.845	0.98	0.996	0.984
Contraste de Kolmogorov-Smirnov para 2012 vs. 2013-2014	0.3472 [0.00]	0.4868 [0.00]	0.1875 [0.07]	0.1736 [0.11]	0.3611 [0.00]
<i>Período total</i>					
2011-2014	1.986	1.702	1.178	1.109	1.063
Númreo de oficinas			72		

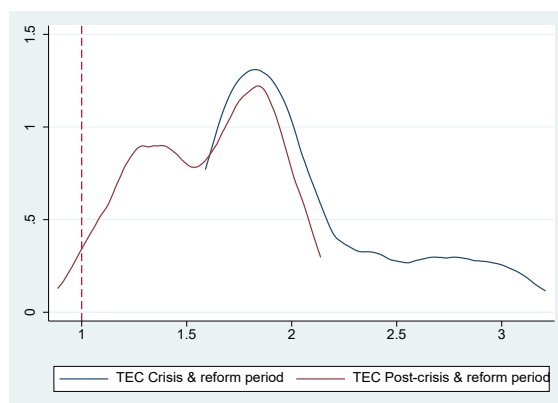
**Notas:** Esta tabla muestra las medias anuales DEA-Malmquist (medias geométricas) para la TPF y sus componentes para el período general 2011-2014, y el período de crisis y reforma (año 2012) y el período posterior a la crisis y reforma (años 2013 y 2014), respectivamente. Es de destacar que el número de sucursales supereficientes es bastante elevado (alrededor del 50%). Por lo tanto, el número de sucursales eliminado los valores atípicos basados en DEA-Malmquist es bajo. La prueba de Kolmogorov-Smirnov es la prueba de Kolmogorov-Smirnov de dos muestras para la igualdad de funciones de distribución para 2012 y el período 2013-2014 (ambos inclusive). Entre paréntesis aparecen los p-valores.

**Figura 9.3.** Gráficas de densidad del índice de productividad estimado de Malmquist y sus componentes por subperíodos. Enfoque de producción.

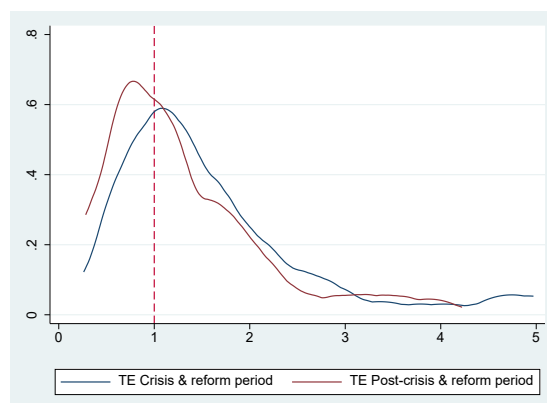
(a)



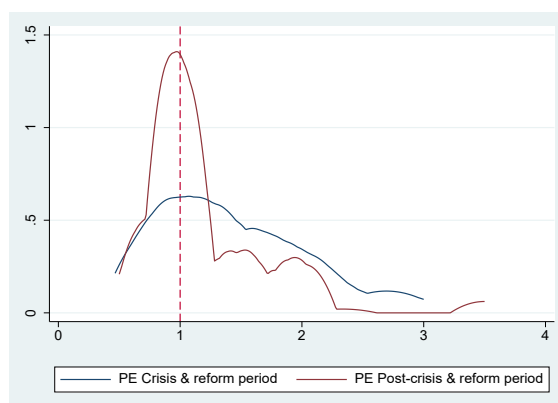
(b)



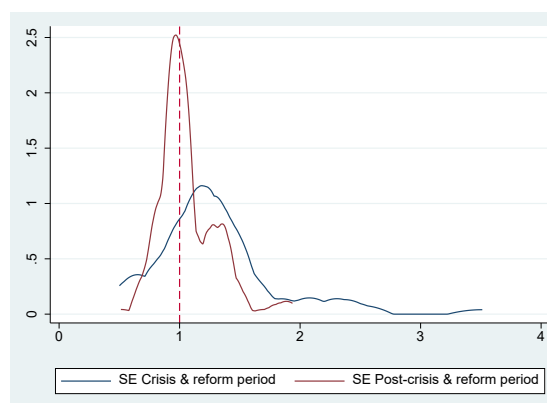
(c)



(d)

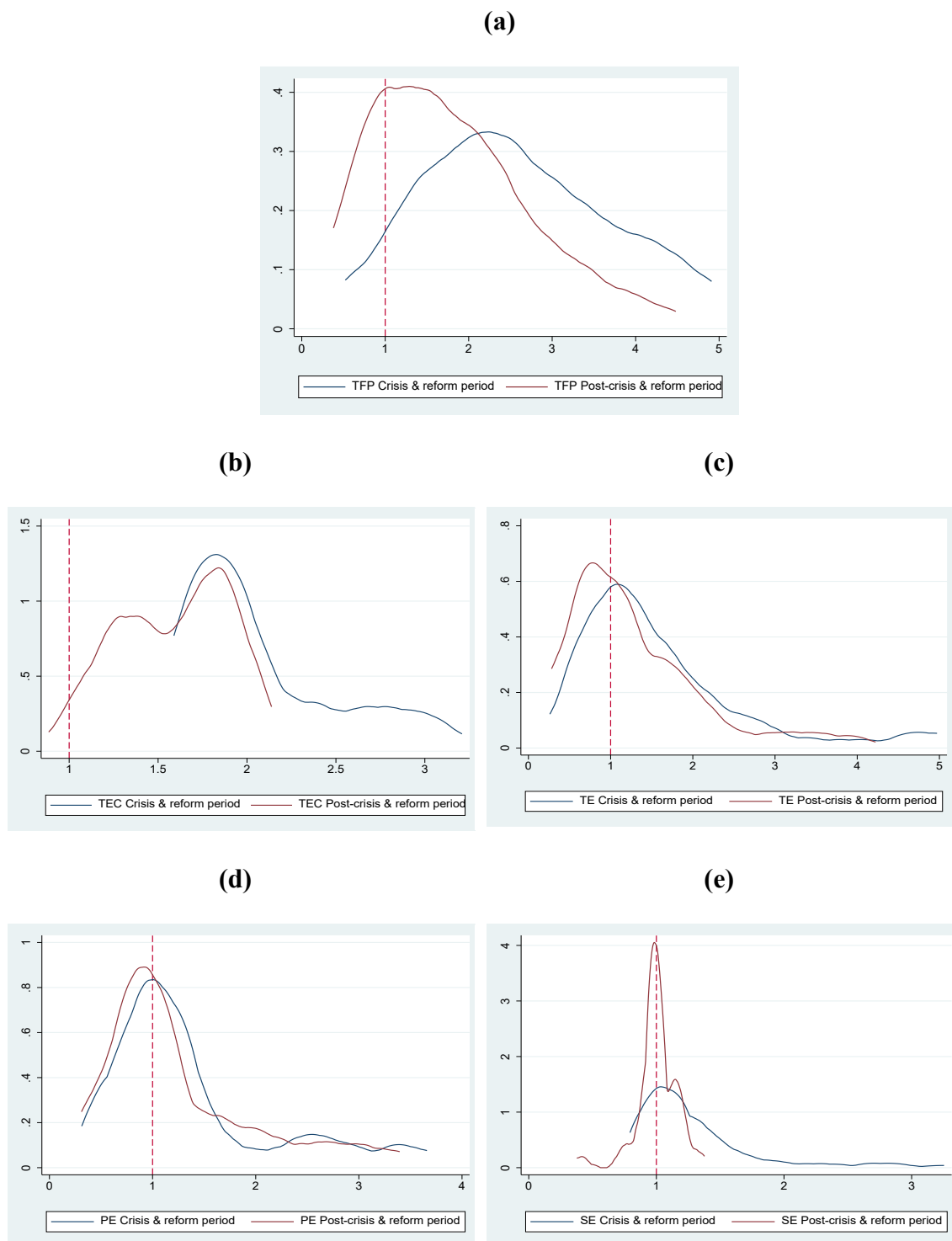


(e)



**Nota:** La línea vertical de trazos largos de color rojo indica la unidad.

**Figura 9.4.** Gráficos de densidad del índice de productividad estimado de Malmquist y sus componentes por subperíodos. Enfoque de rentabilidad.



**Nota:** La línea vertical de trazos largos de color rojo indica la unidad.

En resumen, si bien el período global muestra un aumento de la productividad de las sucursales bancarias a lo largo del tiempo, podemos decir que el efecto de la crisis y la reestructuración financiera fue evidente para varias sucursales, especialmente, las no supereficientes. El período posterior a la crisis y la reforma muestra densidades kernel más leptocúrticas y desplazadas hacia la izquierda. El período de poscrisis y reestructuración significó una contracción en el crecimiento de la productividad de las oficinas, pero también de sus componentes. En general, los resultados muestran que las reformas institucionales derivadas del nuevo escenario macroeconómico (crisis y rescate) pudieron haber afectado al Banco, provocando una disminución de sus niveles de cambio en eficiencia y productividad con respecto al año 2012.

Estos hallazgos sugieren que las oficinas no supereficientes en ambos enfoques no parecen resistentes a las condiciones del mercado establecidas por la crisis y los *shocks* de reforma en términos de niveles de productividad. Concluimos, por tanto, que tanto la regulación bancaria como la segunda recesión de la economía española podrían haber afectado al cambio de la TFP, siendo algunos directores de sucursales incapaces de explotar los cambios tecnológicos o adoptar tecnologías de mejores prácticas, probablemente debido a los *shocks* provocados por el rescate y la segunda recesión española. De ahí, la reestructuración en la red de oficinas que tendría lugar en años posteriores, cerrando sucursales y reduciendo personal.

A modo de pequeña conclusión, los principales resultados obtenidos en este estudio son tres. Primero, a pesar de los duros ajustes y la desfavorable situación económica, la productividad de las sucursales creció en el período global. En segundo lugar, el cambio de productividad se debió principalmente al cambio tecnológico, más que al cambio de eficiencia, lo que destaca la importancia de la innovación en el Banco durante este período. Y tercero, hubo una contracción en la tasa de crecimiento de la productividad y

sus componentes, y un aumento del número de oficinas que experimentaron cambios negativos, especialmente para las oficinas no supereficientes (las densidades *kernel* para el cambio de productividad y sus componentes se desplazaron hacia la izquierda y fueron más leptocúrticas en el período posterior a la crisis y la reforma).

# CAPÍTULO 10

## CONCLUSIONES

---

Esta Tesis Doctoral ha tenido como principal objetivo analizar la eficiencia en la producción y costes de las oficinas bancarias de un gran banco comercial en España, desde diversos ángulos. Para alcanzar dicho objetivo general, se han propuesto tres objetivos específicos. Primero, detectar las sucursales totalmente eficientes e ineficientes, así como las supereficientes y evaluar cuáles pueden ser los factores explicativos de las eficiencias estimadas. Segundo, estudiar los cambios en el tiempo de la productividad de las sucursales. Y, tercero, evaluar si las eficiencias de coste varían en el tiempo y cuáles pueden ser sus factores explicativos.

Obviamente, las implicaciones económicas de estos objetivos son importantes no solo para los gestores de las sucursales sino también para los gestores del propio banco.

Para tratar de responder a las cuestiones implícitas en los objetivos específicos, se han empleado diversas técnicas econométricas para estimar la eficiencia. Por un lado, empleamos diversas metodologías no paramétricas para estimar la eficiencia técnica como son el análisis envolvente de datos (DEA), la metodología FDH - que relaja el supuesto de convexidad de la tecnología implícita en DEA-, y métodos robustos para detectar la presencia de sucursales supereficientes o “outliers” basados en fronteras no paramétricas parciales como *order- $\alpha$* . Al mismo tiempo, y puesto que deseamos estudiar los cambios de productividad de las oficinas, emplearemos el método DEA junto al cálculo del índice del Malmquist. Por otro lado, también emplearemos la metodología

paramétrica de fronteras estocásticas (SFA) con la intención de investigar no solo si la eficiencia de costes es variable en el tiempo sino sus determinantes.

La justificación para estudiar un único banco se debe a la imposibilidad de disponer de los datos de las oficinas de muchos bancos que operasen en España, motivo por el que hemos construido, no exenta de dificultades - dada la confidencialidad de los datos -, una muestra de ellas atendiendo a la información de un gran banco comercial español, que aunque es de los más grandes en volumen de recursos, puede representar adecuadamente la manera de trabajar en el sector bancario español, catalogado como uno de las más importantes y más eficientes en el mundo, a pesar de la crisis sistémica que le ha afectado.

De los resultados obtenidos del análisis de los tres objetivos específicos planteados, existen una serie de conclusiones generales que pueden extraerse y que se detallan a continuación. Por ejemplo,

- a) Las estimaciones de eficiencia obtenidas usando tanto DEA como SFA indican que las eficiencias técnicas y de coste son elevadas. No obstante, cabe destacar que, solamente en DEA, encontramos algunas oficinas totalmente eficientes.
- b) La tecnología de producción no es constante en el tiempo. El cambio tecnológico afecta a dicha tecnología, al menos en el enfoque de producción.
- c) La tecnología de producción presenta economías de escala.

A continuación, comentaremos las conclusiones más específicas obtenidas de los análisis realizados en los Capítulos 7, 8 y 9, y las futuras líneas de investigación.

## **10.1. VALORACIÓN DE LA EFICIENCIA DE LAS SUCURSALES USANDO**

### **MÉTODOS NO PARAMÉTRICOS ROBUSTOS**

En el Capítulo 7 tratamos de llenar un vacío de investigación en los estudios de eficiencia de las sucursales bancarias españolas y analizar el grado de desempeño (eficiencia) de las oficinas bancarias durante un período en el que las instituciones financieras debieron adaptarse a un nuevo escenario macroeconómico, a los cambios regulatorios, a los efectos de una nueva recesión en la economía española en el primer trimestre de 2012, y a la importante reestructuración del sector bancario que tuvo lugar tras el rescate del sistema bancario español en 2012. Que tengamos conocimiento, hasta la fecha no se ha publicado ningún estudio sobre medidas de eficiencias técnicas, de escala y la evaluación de los determinantes relevantes para las sucursales bancarias en España.

Este estudio se realizó para evaluar varias medidas de eficiencia de las sucursales bancarias, con especial atención al desempeño de un importante banco comercial español. A diferencia de estudios anteriores como Oral y Yolalan (1990), Camanho y Dyson (2005b), Portela y Thanassoulis (2007), Giokas (2008a), Paradi et al. (2011) y Aggelopoulos et al. (2013), estimamos la eficiencia técnica (pura y global) y las eficiencias de escala utilizando no solo DEA, sino también FDH y *order- $\alpha$* , teniendo en cuenta que, este último enfoque permite la distinción de sucursales no eficientes, totalmente eficientes y supereficientes. Además, se estudian las relaciones entre las eficiencias estimadas y algunos determinantes económicos de la eficiencia, como la evaluación del desempeño utilizando la métrica bancaria, y otros factores específicos de la sucursal como el tamaño del mercado, la cualificación del personal, la ubicación, la



especialización en ciertas áreas de productos (clientes particulares o negocios), la dotación por la pérdida esperada en activos crediticios (como medida del riesgo crediticio) y el efecto del rescate de la banca española y la recesión secundaria que afectó a la economía española en 2012.

Las eficiencias se obtienen de los datos anuales proporcionados por las sucursales bancarias con respecto al período 2011-2014. Los resultados mostraron puntuaciones altas en eficiencias operativas y de ganancias para DEA (eficiencias puramente técnicas y de escala), FDH y usando los estimadores *order- $\alpha$* . En general, se obtuvo la siguiente jerarquía en las puntuaciones estimadas:  $DEA < FDH < order-\alpha$ .

Otros resultados importantes obtenidos fueron los siguientes. En primer lugar, con respecto a la tecnología de producción, no se prefiere CRS a VRS a nivel de sucursal, y existen eficiencias de escala. Algunos estudios empíricos previos han reportado resultados similares. Por ejemplo, Camanho y Dyson (2005a) concluyeron que la frontera debe estimarse asumiendo VRS para el enfoque de producción. En segundo lugar, debe tenerse en cuenta que nuestras estimaciones de eficiencia no paramétricas son sensibles a la presencia de valores atípicos o sucursales supereficientes. Nuestros resultados, de aplicar el procedimiento de Daraio y Simar (2007) y luego, usar DEA y FDH, muestran que las eficiencias estimadas usando este enfoque son ligeramente más altas que las que provienen por métodos de eficiencia no paramétricos clásicos usando toda la muestra. Por tanto, la existencia de sucursales supereficientes afectaron a los resultados de la eficiencia estimada. En tercer lugar, las sucursales bancarias de nuestra muestra estaban en su mayoría muy bien administradas y alcanzaron niveles altos de eficiencia técnica y de escala, generalmente, estando muy cerca de la frontera tanto en la dimensión de producción como en la de rentabilidad o ganancias, en el nuevo escenario regulatorio y también durante el período reestructuración sectorial. Concluimos, por

tanto, que la mayoría de los directores de sucursales gestionaron correctamente sus inputs y outputs, alentados en ello, por las estrictas normas que impone la Dirección Territorial del Banco y la necesidad de cumplir con la regulación y supervisión bancaria.

Nuestro estudio también reveló la existencia de causalidad entre variables de interés que podrían contribuir a una mejor comprensión del desempeño de las sucursales. Por ejemplo, con respecto a la eficiencia técnica (pura), las propias métricas de desempeño del banco, como PBV definida como la productividad basada en varlor, se pueden comparar con puntuaciones de eficiencia técnica. Por tanto, nuestros resultados validan el uso de métodos no paramétricos, lo que corrobora los hallazgos de investigaciones empíricas que comparan DEA y las propias métricas del banco (Lovell y Pastor, 1997; Berger y Humpfrey, 1997; Parkan y Wu, 1999; entre otros). Por otro lado, no existe una imagen clara de otras variables que afectan la eficiencia técnica como la cualificación del personal (EFA) o las medidas de riesgo, por lo que los gerentes no podrían realizar la segmentación en términos de estas variables. Sin embargo, en lo que respecta al tamaño de escala óptimo, los gerentes bancarios deben tener en cuenta que la eficiencia de la escala puede variar según la experiencia, el tipo de mercado y las medidas de especialización y riesgo, pero no de las propias métricas del banco utilizadas para evaluar el desempeño de las oficinas.

Finalmente, desde el punto de vista de la práctica corporativa, es de destacar que las presiones actuales que se ejercen sobre sus márgenes, debido a la feroz competencia en este sector, han obligado a las entidades a implementar nuevas estrategias para mejorar el desempeño empresarial, ofreciendo productos y servicios innovadores y buscando una mayor eficiencia y productividad. Nuestros resultados tienen varias implicaciones para la práctica empresarial, en particular, con respecto a la eficiencia de las sucursales.

En primer lugar, los gerentes pueden aplicar simultáneamente más de un método para evaluar y comparar la eficiencia y el desempeño de las sucursales (es decir, DEA, FDH y *order- $\alpha$* , entre otras técnicas). Esto les permitiría identificar mejor las oportunidades de mejora, monitorear el patrón de operaciones y la estabilidad o no de cada sucursal a lo largo del tiempo, y construir modelos para ayudar en la toma de decisiones. En segundo lugar, debido a que ningún modelo puede capturar completamente la naturaleza multifuncional de las sucursales bancarias, como se ha demostrado en la investigación empírica, los gerentes deben decidir por sí mismos las mejores formas de medir el desempeño de sus unidades operativas, con respecto a la estrategia prescrita (por ejemplo, dimensiones de producción y/o de rentabilidad) y ajustar los procedimientos de gestión, si es necesario. En tercer lugar, las puntuaciones de eficiencia técnica se pueden utilizar para evaluar el desempeño de la sucursal, complementando las propias métricas de desempeño del banco basadas en PBV. Así, cada director debe considerar las estrategias más adecuadas para incrementar la productividad y el desempeño y, así cumplir con las metas financieras establecidas para la sucursal.

En conclusión, tanto los métodos no paramétricos clásicos como los robustos son apropiados para su uso en un entorno dinámico y pueden modificarse fácilmente para abordar las necesidades cambiantes, revelando así las fortalezas, debilidades y capacidades de mejora de las sucursales bancarias y proporcionando información útil sobre el desempeño, teniendo en cuenta que los gerentes, como prioridad empresarial, están obligados a maximizar la eficiencia de la sucursal.

Sin embargo, este trabajo presenta al menos dos limitaciones, que conviene tener en cuenta; por un lado, el volumen de datos y el período de estudio son limitados. Por otro lado, se necesita una comparación de mayor alcance con otros métodos como *order-m*. Las investigaciones futuras en este campo, por lo tanto, deben incluir un análisis más

detallado de la estimación no paramétrica en cuanto a puntuaciones de supereficiencia, que son más altas que las reportadas en estudios empíricos que utilizaron este método en otros sectores empresariales. También sería útil investigar el resultado de emplear el enfoque del índice de productividad total de los factores (TFP) de Malmquist, que podría basarse en DEA, en métodos robustos no paramétricos como *order-m* pero también en *order- $\alpha$* , y otros métodos DEA basados en eficiencia de costes (evaluar la capacidad de una sucursal para producir productos corrientes a un coste mínimo) y estudiar la cuestión de la eficiencia asignativa (es decir, la capacidad de la sucursal para utilizar sus inputs en proporciones óptimas, dados sus precios).

## **10.2. EFICIENCIA DE COSTES USANDO MODELOS DE FRONTERAS**

### **ESTOCÁSTICAS**

En el Capítulo 8 quisimos poner énfasis en la estructura de costes de las sucursales, debido a que, en los últimos años, el sector bancario ha experimentado una importante reestructuración, no solo debido a la crisis financiera de 2007 (con cambios regulatorios sobre los requisitos de capital y liquidez), sino también en respuesta a los cambios en los patrones de comportamiento de los clientes y las relaciones con los proveedores de servicios financieros, lo que ha afectado a los costes del negocio.

Así, el objetivo del estudio fue estimar el comportamiento de las eficiencias de costes utilizando datos de nuestra muestra y evaluar su relación con las medidas propias del banco sobre el desempeño de las sucursales.

El enfoque econométrico utilizado fue el del modelo costes translogarítmicos de datos de panel, que permitirá examinar la dimensión productiva del desempeño de las sucursales y, en el que se supone, que los efectos de ineficiencia de costes son una

función de diversos factores explicativos como el nivel de desempeño de la sucursal y otros determinantes.

Dado que las estimaciones de eficiencia en el modelo de frontera estocástica con datos de panel pueden estar sesgadas si se ignora la heterogeneidad no observada, utilizamos los modelos desarrollados por Greene (2005a,b) para tratar dicha heterogeneidad, como el modelo TRE o *true random effects*.

En este estudio usamos datos mensuales correspondientes al período 2013-2014, obtenidos del balance, cuentas de pérdidas y ganancias de sucursales, así como información de departamentos internos del banco en cuestión. Este periodo de estudio, aunque más corto que el del Capítulo 7, también se correspondía con el nuevo escenario macroeconómico y la reestructuración que se produjo en el sector bancario español tras la crisis bancaria de 2010 y el rescate de 2012.

Los principales resultados obtenidos muestran la siguiente información útil para los gerentes o managers bancarios:

1) La eficiencia de costes varía con el tiempo y existe una heterogeneidad no observada en las sucursales bancarias que es estadísticamente significativa. Esta heterogeneidad no observada podría estar relacionada con aspectos como la heterogeneidad tecnológica debida a diferencias de tamaño de las sucursales. No obstante, cabe destacar que este hecho no fue contrastado evaluando la relación entre los rendimientos de escala y el tamaño de la sucursal (véase, por ejemplo, Feng and Zhang, 2014).

2) Si bien las sucursales no son totalmente eficientes, su rentabilidad es alta. El uso de los recursos de las sucursales para generar préstamos y transacciones (el enfoque de producción) logra puntuaciones de eficiencia elevados (85%-99%). Por lo tanto, la red

de sucursales podría mejorar su capacidad para generar/producir productos actuales a un coste mínimo (1%-15%).

3) Se observaron patrones temporales no estacionarios en los costes medios.

4) La tecnología de producción es constante y, por lo tanto, el cambio tecnológico no fue una variable significativa durante el período de estudio.

5) La tecnología presenta economías de escala.

6) Las eficiencias de costes variables en el tiempo estimadas están asociadas positivamente con medidas del volumen de negocios (productividad basada en valor) y, negativamente asociadas con la superficie de la sucursal, y también con el tiempo. Sin embargo, no se encontró relación con el balance y los estados pérdidas y ganancias (cuenta resultados), es decir, las desviaciones negativas de los objetivos del margen bruto. Otras variables, como el número de personal cualificado y el tamaño/tipo de mercado (rural o urbano) no tuvieron efecto sobre la rentabilidad. 7) La X-ineficiencia, referida a la ineficiencia relacionada con la sucursal de mejores prácticas en el período analizado está en torno al 7,7%. Por lo tanto, se debe alentar a las sucursales bancarias ineficientes a elevar sus niveles de eficiencia a los del índice de referencia o benchmark tomando de referencia las mejores prácticas.

En general, llegamos a la conclusión que las sucursales bancarias examinadas presentan buenos niveles de desempeño en cuanto al enfoque productivo y son eficientes en la gestión de sus resultados económicos. Este resultado favorable puede provenir, por un lado, de que el banco disponga de un modelo de cartera diversificado y un sistema de gestión basado en la prudencia y por otro lado, a la monitorización constante de las cifras de negocio de las sucursales. Los resultados obtenidos de nuestra medida de eficiencia de costes variable en el tiempo son comparables con los de las propias medidas del banco sobre el desempeño de las sucursales.

Finalmente, cabe también destacar que este estudio presenta ciertas limitaciones que podrían subsanarse en futuras investigaciones. Por un lado, aunque nuestra muestra original fue revisada eliminando algunas sucursales problemáticas, que eran valores atípicos potenciales, los resultados siguen siendo sensibles a los valores atípicos porque no pudimos usar especificaciones de frontera estocástica alternativas para tratar con posibles valores atípicos en la muestra final, debido a problemas computacionales intratables. Además, no pudimos separar los dos elementos de la eficiencia de las sucursales: uno persistente y otro transitorio, utilizando modelos de tipo GTRE (*generalized true random effects*). Por lo tanto, no pudimos determinar la presencia o no de ineficiencia en costes sistemática y/o no sistemática en las sucursales bancarias. Tampoco, se tuvo en cuenta la posible presencia de un sesgo de heterogeneidad no observado debido a la correlación de las variables exógenas con el término estocástico, por ejemplo, utilizando la estimación del modelo GTRE con el ajuste de Mundlak (MGTRE), tal y como propusieron Filippini y Greene (2016).

### **10.3. CAMBIOS DE PRODUCTIVIDAD DE LAS OFICINAS**

El Capítulo 9 explora el crecimiento de la productividad de las sucursales de nuestra muestra antes y después del período afectado por la crisis de 2012 en el que apareció un nuevo escenario macroeconómico y el sistema bancario español sufrió una amplia reestructuración derivada de reformas regulatorias. En concreto, medimos el cambio de productividad en el tiempo y lo desglosamos entre el que se debe a mejoras en la eficiencia técnica y el que es debido al cambio tecnológico. Cabe destacar que, desafortunadamente, no podemos distinguir los efectos de la crisis de las caracterizadas por las reformas de reestructuración.

El análisis de los cambios de productividad se realiza utilizando datos anuales de las sucursales bancarias y calculando el índice de productividad de Malmquist usando DEA con orientación input, tanto bajo el enfoque de producción (operaciones) como el de rentabilidad (ganancias). El índice de productividad de Malmquist también permite desglosar la productividad total de los factores (TFP) en el cambio de eficiencia técnica y cambio tecnológico (progreso tecnológico). Con dicho índice podemos evaluar el crecimiento de la producción total de los factores de las sucursales.

El banco analizado en el período 2011-2014, a pesar de ser uno de los bancos con mayor solvencia de Europa y con mayor solidez en las pruebas de resistencia realizadas por el Banco Central Europeo (que ha permitido afrontar la crisis financiera de 2007 con menos necesidad de adoptar medidas drásticas), muestra algunos resultados interesantes para sus sucursales utilizando el método DEA-Malmquist basado en valores atípicos (es decir, solo considerando las sucursales totalmente eficientes e ineficientes, pues se han eliminado aquellas supereficientes).

Primero, a pesar de los duros ajustes y la desfavorable situación económica, la productividad de las sucursales creció, pero a un ritmo decreciente. En segundo lugar, en términos de las distribuciones empíricas, los resultados muestran que las densidades kernel se desplazaron hacia la izquierda y fueron más leptocúrticas en el período posterior a la crisis y la reforma, lo que indica una contracción en la tasa de crecimiento de la productividad y sus componentes (el crecimiento de la productividad operativa y los ingresos disminuyeron a lo largo del tiempo), y un aumento en el número de sucursales que experimentaron cambios negativos. Estos hechos son indicativos de que la tecnología de producción no fue constante. De hecho, esta menor productividad podría estar asociada a un cambio en el modelo productivo, orientado a mantener la sostenibilidad empresarial a través de un cambio en la relación banco-cliente con una



mayor digitalización (como consecuencia de lo cual, la mayoría de las transacciones ahora se pueden realizar con un teléfono inteligente). Actualmente, se están implementando protocolos en las rutinas del banco para maximizar la agilidad, mejorar la calidad y satisfacer las expectativas del cliente que buscan soluciones digitales. Los principales factores explicativos del cambio en la productividad total fueron la experiencia, el deterioro del crédito esperado y el tamaño del mercado (en el enfoque de ganancias) y las calificaciones profesionales del personal bancario y la especialización de sucursales (en el enfoque de producción).

Y tercero, el principal factor subyacente al efecto de la TFP fue el cambio tecnológico (cambio de frontera) más que el cambio de eficiencia (el efecto *catch-up*). Es decir, el cambio tecnológico derivado de la innovación fue un impulsor más fuerte del crecimiento de la productividad en las sucursales bancarias. Destaca la importancia de la innovación en el Banco durante este período con una fuerte inversión en tecnología.

Como áreas útiles para una investigación futura de los cambios en la productividad, podrían realizarse análisis mediante otros enfoques dimensionales (Paradi et al., 2011), por ejemplo, considerar la dimensión entre ahorradores e inversores, donde los inputs se relacionan con los intereses y costes (no relacionados con intereses) y los productos reflejan los flujos financieros asociados con la actividad de intermediación y, la rentabilidad (es decir, cómo las sucursales bancarias convierten los gastos - aquellos sujetos a la influencia directa de la dirección de la sucursal - en ingresos). Otra línea de desarrollo podría ser realizar un análisis del cambio de productividad utilizando métodos robustos no paramétricos para estimar, por ejemplo, las fronteras *order-m* con el fin de crear una versión robusta del índice de productividad de Malmquist usando el enfoque de Wheelock y Wilson (2003).

#### **10.4. FUTURAS LÍNEAS DE INVESTIGACIÓN**

Para finalizar, podríamos comentar brevemente algunas de las líneas de investigación que actualmente se están desarrollando, utilizando la información de las sucursales bancarias empleadas en esta Tesis Doctoral, y que pretenden dar continuidad al esfuerzo investigador en esta línea de trabajo sobre oficinas bancarias.

Aunque existen muchos temas interesantes en la práctica cotidiana de las sucursales que sirvan a los gestores para tomar decisiones basadas en la eficiencia y también métodos de estimación de las eficiencias (algunos expuestos en esta Tesis Doctoral en los Capítulos 3 y 4), actualmente se están investigando dos líneas de investigación prometedoras. Por un lado, una línea que pretende analizar el efecto de las provisiones crediticias por pérdida esperada sobre la eficiencia y, por otro lado, otra línea que pretende estudiar la heterogeneidad no observada de la tecnología de producción de las sucursales, que puede afectar a la estimación de sus rendimientos a escala.

Concretamente, y respecto del primer trabajo – que se encuentra en fase avanzada de investigación y cuyo título provisional es “*Expected credit loss provisions and performance: Handling negative data in the bank’s branches efficiency*”-, se analiza la idea que la calidad de los activos de los bancos es un indicador adelantado e importante de las señales de quiebra y puede influir en la eficiencia y la estabilidad de la oficina. Cabe destacar que la literatura empírica que estudia el papel de los créditos impagados sobre la eficiencia se centra en el estudio de las provisiones de pérdidas crediticias (en inglés, *loan loss provisions*) o también los préstamos “no cobrados” después de un periodo o plazo superior a 1 año (en inglés, *non-performing loans*, NPL), como medidas de los préstamos incobrables que pueden afectar negativamente no solo los ingresos del banco sino también la estabilidad de su negocio. Sin embargo, a diferencia de estos

enfoques, nuestra aportación a la literatura pretende evaluar la eficiencia técnica de las sucursales bancarias utilizando las provisiones por pérdidas esperadas (en inglés, *expected loss provisions*, en vez de las *loan loss provisions*), por ser este un concepto más amplio que afecta igualmente a la eficiencia y estabilidad del banco, y que actualmente está regulado por la normativa de Basilea y también del Banco de España a través de distintas circulares.<sup>72</sup> Hay que entender que las diferentes provisiones pueden tener un impacto sobre la ciclicidad de las ganancias de los bancos (es decir, un resultado interesante es la relación negativa entre las provisiones y el crecimiento de los préstamos y el PIB, lo que parece implicar que las provisiones son procíclicas), y el capital regulatorio.

La inclusión de esta variable como input supondría un cambio en el uso de los modelos tradicionales de eficiencia puesto que dicha variable podría presentar datos negativos y positivos. En este sentido, el enfoque utilizado para estimar la eficiencia requeriría del uso de un modelo que permitiese el uso de datos negativos y positivos, como es el modelo de rango direccional, quien tiene la propiedad de permitir manejar datos sin restricciones de positividad en un marco DEA. El modelo que hemos utilizado es el derivado por Silva Portela et al. (2004). Los resultados preliminares muestran que la eficiencia no se ha visto reducida, pero tampoco incrementada, consecuencia del uso de la pérdida esperada como variable input (tal y como la literatura empírica recomienda, al menos en el caso de las provisiones por posible impago o pérdida de los préstamos).

---

<sup>72</sup> Conviene destacar que la provisión por pérdida esperada es una variable que pretende anticipar las dotaciones en función de la probabilidad default de aquellas carteras crediticias que se preveían caer a mora. De esta manera, se utilizaría para aumentar la severidad en aquellas unidades con mayor perfil de entradas a mora y en función del producto. Si finalmente la operación cae a mora se desdota (signo +) y entra el flujo por dotación neta (según calendario-normativa), pero no se incrementa la cifra de saneamientos totales porque se había anticipado: De esa forma, los resultados de explotación no sufrirían variaciones significativas. También podremos ver una desdotación cuando el cliente sana posiciones y deja de tener una probabilidad alta de entrada en mora.

La otra línea de investigación se centraría en la investigación de la heterogeneidad tecnológica no observada en las sucursales bancarias (relacionada, por ejemplo, con el efecto del tamaño (sucursales grandes/pequeñas), el grado de especialización de las sucursales, las diferencias entre áreas rurales/urbanas), y la eficiencia variable en el tiempo. Para ello, podría utilizarse un modelo de frontera de distancia estocástica de coeficientes aleatorios que tendría en cuenta la existencia de múltiples inputs y outputs en un enfoque Bayesiano, similar al utilizado por Feng y Zhang (2014) para estudiar los rendimientos a escala de los grandes bancos de EE. UU. La principal ventaja de este modelo es que sus coeficientes pueden variar entre las sucursales, lo que permite la heterogeneidad tecnológica no observada entre las mismas. En este caso, y a diferencia de los trabajos presentados en esta Tesis Doctoral, e incluso el trabajo propuesto anteriormente, la muestra de los inputs y outputs se ha conseguido ampliar al periodo 2011-2017. El título provisional del trabajo que se está realizando es “*Technological heterogeneity and efficiency in bank branches: A Bayesian approach*”. No obstante, y a diferencia de la anterior línea, aún no disponemos de resultados preliminares del modelo propuesto que puedan avanzarse, los cuales podrían ayudar no solamente a obtener un ranking de sucursales, sino que también a conocer sus rendimientos a escala.



## REFERENCIAS BIBLIOGRÁFICAS

---

- Afriat, S. N. (1972). Estimation of production functions. *International Economic Review*, 13, 3, 568-598.
- Aggelopoulos, E., Georgopoulos, A. and Siriopoulos, C. (2013). Comparative efficiency analysis of Greek bank branches in the light of the financial crisis. CiteSeerX.psu:10.1.1.366.723
- Aigner, D. and Chu, S. F. (1968). On estimating the industry production function. *American Economic Review*, 58, 4, 826-839.
- Aigner, D. J., Lovell, C. A. K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier function models. *Journal of Econometrics*, 6, 21–37.
- Ali, M.M., Mikhail, N.N. and Haq, M.S. (1978). A class of bivariate distributions including the bivariate logistic. *Journal of Multivariate Analysis*, 8, 405-412.
- Alirezaee, M.R. and Afsharian, M. (2007). Model improvement for computational difficulties of DEA technique in the presence of special DMUs. *Applied Mathematics and Computation*, 186, 1600–1611.
- Allen, L. and Rai, A. (1996). Operating efficiency in banking: an international comparison. *Journal of Banking and Finance*, 20, 655–672.
- Allen, K. (1999). Dea in the ecological context — an overview. In: Westermann, G. (ed). *Data Envelopment Analysis in the Service Sector*. Gabler Edition Wissenschaft: Harzer, pp 203–235.

- Almanidis, P. (2013). Accounting for heterogeneous technologies in the banking industry: a time-varying stochastic frontier model with threshold effects. *Journal of Productivity Analysis*, 39, 191–205.
- Al-Tamimi, H.A.H. and Lootah, A.M. (2007). Evaluating the operational and profitability efficiency of a UAE-based commercial bank. *Journal of Financial Services Marketing*, 11, 333–348.
- Altunbas, Y., Liu, M., Molyneux, P. and Rama, S. (2000). Efficiency and risk in Japanese banking. *Journal of Banking and Finance*, 24, 1605–1628.
- Altunbas, Y., Gardener, E., Molyneux, P. and Moore, B. (2001). Efficiency in European banking. *European Economic Review*, 45, 1931-1955.
- Altunbas, Y., Bondt, G. and Marques-Ibanez, D. (2004). Bank Capital, Bank Lending, and Monetary Policy in the Euro Area. *Kredit and Kapital*, 4, 18 pages.
- Amel, D., Barnes, C., Panetta, F. and Salleo, C. (2004). Consolidation and efficiency in the financial sector: A review of the international evidence. *Journal of Banking & Finance*, 28, 2493-2519.
- Andersen, P. and Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39, 1261–1264.
- Andor, M. and Hesse, F. (2014). The StoNED age: The departure into a new era of efficiency analysis? A Monte Carlo comparison of StoNED and the “oldies” (SFA and DEA). *Journal of Productivity Analysis*, 41, 85-109.
- Apergis, N. and Rezitis, A. (2004). Cost structure, technical change and productivity growth in the Greek banking sector. *International Advances in Economic Research*, 10: 1-15.

- Aragon, Y., Daouia, A. and Thomas-Agnan, C. (2005). Nonparametric frontier estimation: a conditional quantile based approach. *Econometric Theory*, 21, 358–389.
- Asmild, M., Paradi, J., Aggarwal, V. and Schaffnit, C. (2004). Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry. *Journal of Productivity Analysis*, 21, 67–89.
- Asmild, M. and Tam, F. (2007). Estimating global frontier shifts and global Malmquist indices. *Journal of Productivity Analysis*, 27, 137–148.
- Assaf, A.G. and Magnini, V. (2012). Accounting for customer satisfaction in measuring hotel efficiency: Evidence from the US hotel industry. *International Journal of Hospitality Management*, 31, 642-647.
- Assaf, A.G. and Barros, C.P. (2013). A global benchmarking of the hotel industry. *Tourism Economics*, 19(4), 811-821.
- Assaf, A.G. and Tsionas, M. (2018). The estimation and decomposition of tourism productivity. *Tourism Management*, 65, 131-142.
- Athanassopoulos, A.D. (1997). Service quality and operating efficiency synergies for management control in the provision of financial services: Evidence from Greek bank branches. *European Journal of Operational Research*, 98, 300–313.
- Athanassopoulos, A.D. (1998). Nonparametric frontier models for assessing the market and cost efficiency of large-scale bank branch networks. *Journal of Money, Credit, and Banking*, 30, 172–192.
- Azizi, H. and Ajirlu, S.F. (2010). Measurement of overall performances of decision-making units using ideal and anti-ideal decision-making units. *Computers and Industrial Engineering*, 59, 411–418.



- Baccouche, R. and Kouki, M. (2003). Stochastic production frontier and technical inefficiency: A sensitivity analysis. *Econometric Reviews*, 22, 79-91.
- Balk, B.M. (1999). Scale efficiency and productivity change. *Mimeo*.
- Banker, R. D., Charnes, A. and Cooper, W. W. (1984). Some model for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, 1078–1092.
- Banker, R.D., Das, S. and Datar, S.M. (1989). Analysis of cost variances for management control in Hospitals. *Res Gov Nonprofit Account*, 5, 268–291.
- Banker, R.D. and Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. *Journal of Productivity Analysis*, 3, 4, 401-415.
- Barros, C.P, Chen, Z., Liang, Q.B. and Peypoch, N. (2011). Technical efficiency in the Chinese banking sector. *Economic Modelling*, 28, 5, 2083-2089.
- Barros.C.P., Managi, S. ad Matousek, R. (2012). The technical efficiency of the Japanese banks: non-radial directional performance measurement with undesirable outputs. *Omega*, 40, 1-8.
- Barth, J.R., Caprio, Jr.G., Levine, R. (2004). Bank regulation and supervision: what works best?. *Journal of Financial Intermediation*, 13, 205-248.
- Battese, G., and Corra, G. (1977). Estimation of a production frontier model: With application to the pastoral zone of Eastern Australia. *Australian Journal of Agricultural Economics*, 21, 3, 169–179.
- Battese, G. E. and Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, 38, 387-399.

- Battese, G. and Coelli, T. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, 3, 1, 153-169.
- Battese, G. and Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20, 325-332.
- Battese, G. E., Rao, D. S. P. and O'Donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, 21, 91-103.
- Bauer, P. W., Berger, A.N., Ferrier G.D. and Humphrey, D.B. (1998). Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. *Journal of Economics and Business*, 50, 85-114.
- Beccalli, E., Casu, B. and Girardone, C. (2006). Efficiency and stock performance in European banking. *The Journal of Business, Finance and Accounting* 33, 218-235.
- Belotti, F. and Ilardi, G. (2012). Consistent estimation of the “true” fixed-effects stochastic frontier model. CEIS Research Papers (231).
- Belotti, F., Daidone, S., Ilardi, G. and Atella, V. (2012). Stochastic frontier analysis using Stata. *The Stata Journal*, vv, Number ii, 1-39.
- Bera, A. K. and Sharma, S. C. (1999). Estimating production uncertainty in stochastic frontier production function models. *Journal of Productivity Analysis*, 12, 2, 187-210.

- Berg, S.A., Forsund, F.R., Hjalmarsson, L. and Suominen, M. (1992). Malmquist Indices of Productivity Growth during the Deregulation of Norwegian Banking, 1980-89. *The Scandinavian Journal of Economics*, 94, 211-228.
- Berg, S.A., Førsund, F.R., Hjalmarsson, L. and Suominen, M. (1993). Banking efficiency in Nordic countries. *Journal of Banking and Finance*, 17, 371-388.
- Berger, A. and Humphrey, D. (1992). Measurement and efficiency issues in commercial banking. In Z. Griliches (Ed.), *Z. Griliches Studies in income and wealth: Output measurement in the service sectors*. Chicago: University of Chicago Press.
- Berger, A. and Humphrey, D. (1994). Bank scale economies, mergers, concentration and efficiency: The U.S. experience. Working Paper 94-25, The Wharton School, University of Pennsylvania.
- Berger, A.N. and DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking and Finance*, 21, 849-870.
- Berger, A.N. and Humphrey, D.B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*, 98, 175-212.
- Berger, A.N. and Mester, L.J. (1997). Inside the black box: what explains differences in the efficiencies of financial institutions?. *Journal of Banking and Finance*, 21, 895-947.
- Berger, A.N., Leusner, J.H. and Mingo, J.J. (1997). The efficiency of bank branches. *Journal of Monetary Economics*, 40, 141-162.
- Berger, A., De Young, R., Genay, H. and Udell, G. (2000). Globalization of financial institutions: Evidence from cross-border banking performance. *Brookings-Wharton Papers on Financial Services*, 23-125.

- Berger, A. and Mester, L. J. (2003). Explaining the dramatic changes in performance of US banks: technological change, deregulation, and dynamic changes in competition. *Journal of Financial Intermediation*, 12, 1, 57-97.
- Berger, A.N., Hasan, I. and Zhou, M. (2009). Bank ownership and efficiency in China: what will happen in the world's largest nation? *Journal of Banking and Finance*, 33, 113-130.
- Berndt, E.R. (1991). *The practice of econometrics classic and contemporary*. Addison Wesley.
- Boles, J.N. (1966). Efficiency squared – Efficient computation of efficiency indexes. Proceedings of the 39th Annual Meeting of the Western Farm Economic Association, 137-142.
- Bonin, J.P., Hasan, I. and Wachtel, P. (2005). Bank performance, efficiency, and ownership in transition countries. *Journal of Banking and Finance*, 29 (1), 31–53.
- Bos, J.W. and Kolari, J.W. (2005). Large bank efficiency in Europe and the United States: are there economic motivations for geographic expansion in financial services?. *Journal of Business*, 78, 4, 1555-1592.
- Boyd, J. and Graham, S. (1998). Consolidation in U.S. banking: Implications for efficiency and risk. In Y. Amihud and G. Miller (Eds.), *Bank mergers and acquisitions*. Boston: Kluwer Academic Publishers.
- Brack, E. and Jimborean, R. (2009). The cost-efficiency of French banks. See [https://mpra.ub.uni-muenchen.de/23471/1/MPRA\\_paper\\_23471.pdf](https://mpra.ub.uni-muenchen.de/23471/1/MPRA_paper_23471.pdf).
- Buch, C. and Dages, B.G. (2018). Structural changes in banking after the crisis. *Committee on the Global Financial System (CGFS) Papers*, No 60. Bank for International Settlements (BIS).

- Cabrera-Suárez, I. and Pérez-Rodríguez, J.V. (2020). Assessing branch efficiency and managerial behaviour in a large Spanish commercial bank, *Spanish Journal of Finance and Accounting*, 49 (1), 48 - 73.
- Cabrera-Suárez, I. and Pérez-Rodríguez, J.V. (2021). Bank branch performance and cost efficiency: A stochastic frontier panel data approach. *International Journal of Finance and Economics*, 26 (4), 5850-5863.
- Caiazza, S., Pozzolo, A. F. and Trovato, G. (2016). Bank efficiency measures, M&A decision and heterogeneity. *Journal of Productivity Analysis*, 46, 25-41.
- Camanho, A.S. and Dyson, R.G. (1999). Efficiency, size, benchmarks and targets for bank branches: An application of data envelopment analysis. *Journal of the Operational Research Society*, 50, 903-915.
- Camanho, A.S. and Dyson, R.G. (2005a). Cost efficiency, production and value-added models in the analysis of bank branch performance. *Journal of the Operational Research Society*, 56, 483–494.
- Camanho, A.S. and Dyson, R.G. (2005b). Cost efficiency measurement with price uncertainty: A DEA application to bank branch assessments. *European Journal of Operational Research*, 161, 432-46.
- Camanho, A.S. and Dyson, R.G. (2006). Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of Productivity Analysis*, 26, 35-49.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Canhoto, A. and Dermine, J. (2003). A Note on Banking Efficiency in Portugal, New vs. Old Banks. *Journal of Banking and Finance*, 27, 2087-2098.

- Castro, C. and Galán, J. E. (2019). Drivers of Productivity in the Spanish Banking Sector: Recent Evidence. *Journal of Financial Services Research*, 55, 115–141.
- Casu, B. and Girardone, C. (2002). A comparative study of the cost efficiency of Italian bank conglomerates. *Managerial Finance*, 28, 3-23.
- Casu, B., Girardone, C. and Molyneux, P. (2004). Productivity change in European banking: A comparison of parametric and non-parametric approaches. *Journal of Banking and Finance*, 28, 2521-2540.
- Caudill, S., and Ford, J. (1993). Biases in frontier estimation due to heteroscedasticity. *Economic Letters*, 41, 17-20.
- Caudill, S., Ford, J. and Gropper, D. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business Economics and Statistics*, 13, 105-111.
- Caves, D.W., Christensen, L.R. and Diewert, W.E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica*, 50, 1393–1414.
- Cazals, C., Florens, J.P. and Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106, 1-25.
- Cebenoyan, S., Cooperman, E., Register, C. and Hudgins, S. (1993). The relative efficiency of stock versus mutual S&Ls: A stochastic cost frontier approach. *Journal of Financial Services Research*, 7, 151-170.
- Chaffai, M.E., Dietsch, M. (1999). Mesures de l'efficience technique et de l'efficience allocative par les fonctions de distance et application aux banques européennes. *Revue économique*, 50, 3, 633-644.

- Chaffai, M.E., Dietsch, M. and Lozano-Vivas, A. (2001). Technological and Environmental Differences in the European Banking Industries. *Journal of Financial Services Research*, 19, 1/2, 147-162.
- Chambers, R.G., Chung, Y. and Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory*, 70, 407–419.
- Chambers, R.G., Chung, Y. and Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, 98, 351–364.
- Chang, T. and Chiu, Y. (2006). Affecting factors on risk-adjusted efficiency in Taiwan's banking industry. *Contemporary Economic Policy*, 24, 634-648.
- Charnes, A., Cooper, W. W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.
- Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. (1995). *Data envelopment analysis: Theory, methodology and applications*. Kluwer.
- Chen, P.C., Yu, M., Managi, S. and Chang, C. (2010). Non-radial directional performance measurement with undesirable outputs. Working Paper, 2010, Tohoku University, Japan.
- Christensen, L.R. and Greene, W.H. (1976). Economies of scale in U.S. electric power generation. *Journal of Political Economy*, 84, 4 (Part 1), 655-676.
- Christopoulos, D., Lolos, S. and Tsionas, E. (2002). Efficiency of the Greek banking system in view of the EMU: A heteroscedastic stochastic frontier approach. *Journal of Policy Modeling*, 24, 813-829.
- Chortareas, G., Girardone, C. and Ventouri, A. (2009). Efficiency and productivity of Greek banks in the EMU era. *Applied Financial Economics*, 16, 1317- 1328.

- Chung, Y., Färe, R. and Grosskopf, S. (1997). Productivity and undesirable outputs: a directional distance function approach. *Journal of Environmental Management*, 51, 229–240.
- Clark, J. (1996). Economic cost, scale efficiency, and competitive viability in banking. *Journal of Money, Credit and Banking*, 28, 342-364.
- Cobb, C.W. and P.H. Douglas, P.H. (1928). A theory of production. *American Economic Review*, 18, 139-165.
- Coelli, T. (1996). A guide to DEAP version 2.1: A data envelopment analysis (computer) program. CEPA Working Papers, 8. University of New England, Australia.
- Coelli, T., Prasada Rao, D. S. and Battese, G. E. (1997). An introduction to efficiency and productivity analysis. Boston: Springer.
- Coelli, T., Prasada Rao, D.S., O'Donnell, C.J., Battese, G.E. (2005). An Introduction to Efficiency and Productivity Analysis, 2nd edition. *Kluwer Academic Publishers*, Boston.
- Colombi, R. (2010). A skew normal stochastic frontier model for panel data. Proceedings of the 45-th Scientific Meeting of the Italian Statistical Society.
- Colombi, R., Martini, G., and Vittadini, G. (2011). A stochastic frontier model with short-run and long-run inefficiency random effects. Department of Economics and Technology Management, Università Di Bergamo, Italy.
- Colombi, R., Kumbhakar, S. C., Martini, G., and Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis*, 42, 2, 123-136.



- Cook, W.D., Seiford, L.M. and Zhu, J. (2004). Models for performance benchmarking: Measuring the effect of e-business activities on banking performance. *Omega*, 32, 313–322.
- Cook, W.D. and Zhu, J. (2006). Incorporating multiprocess performance standards into the DEA framework. *Operations Research*, 54, 656–665.
- Cooper, W. W., Seiford, L. M. and Tone, K. (2000). Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-solver software. Boston: Kluwer Academic Publishers.
- Cooper, W.W., Seiford, L.M. and Tone, K. (2007). Some models and measures for evaluating performances with DEA: past accomplishments and future prospects. *Journal of Productivity Analysis*, 28, 151–63.
- Cornwell, C., Schmidt, P. and Sickles, R. (1990). Production frontiers with cross sectional and time series variation in efficiency levels. *Journal of Econometrics*, 46, 185-200.
- Cuesta, R. A. and Orea, L. (2002). Mergers and technical efficiency in Spanish savings banks: A stochastic distance function approach. *Journal of Banking & Finance*, 26(12), 2231–2247.
- Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140, 375–400.
- Das, A., Ray, S.C. and Nag, A. (2009). Labor-use efficiency in Indian banking: A branch-level analysis. *Omega*, 37, 411–425.
- Daraio, C., and Simar, L. (2007). Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications. New York: Springer.
- Daraio, C., Simar, L. and Wilson, P.W. (2011). Testing whether two-stage estimation is meaningful in non-parametric models of production. See <http://www.tse->

fr.eu/sites/default/files/medias/doc/conf/frontier/List\_of\_Communications/paul\_wilson\_dsw\_sep.pdf.

- Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19, 3, 273-292.
- Dekker, D. and Post, T. (2001). A quasi-concave DEA model with an application for bank branch performance evaluation. *European Journal of Operational Research*, 132, 296–311.
- Delis, M.D., Papanikolaou, N.I. (2009). Determinants of bank efficiency: Evidence from a semi-parametric methodology. *Managerial Finance*, 35, 3, 260-275.
- Delis, M. D., Molyneux, P. and Pasiouras, F. (2011). Regulations and Productivity Growth in Banking: Evidence from Transition Economies. *Journal of Money, Credit and Banking*, 43, 4, 735-764.
- Demircug-Kunt, A., Detragiach, E., Tressel, T. (2008). Banking on the principles: Compliance with Basel core principles and bank soundness. *Journal of Financial Intermediation*, 17, 4, 511-542.
- Deprins, D., Simar, L. and Tulkens, H. (1984). Measuring labor-efficiency in post offices. In M. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: Concepts and measurement* (Chap. 10, pp. 243–267). Amsterdam: North-Holland.
- Deville, A. (2009). Branch banking network assessment using DEA: A benchmarking analysis—a note. *Management Accounting Research*, 20, 252–261.
- DeYoung, R. (1997). Bank mergers, X-efficiency, and the market for corporate control. *Managerial Finance*, 23, 31-46.
- DeYoung, R. and Hasan, I. (1998). The performance of denovo commercial banks: a profit efficiency approach. *Journal of Banking and Finance*, 22, 565-587.

- Dietsch, M. and Lozano-Vivas, A. (2000). How the environment determines the efficiency of banks: A comparison between French and Spanish banking industry. *Journal of Banking and Finance*, 24, 6, 985-1004.
- Diewert, E.W. and Fox, K.J. (2010). Malmquist and Törnqvist productivity indexes: Returns to scale and technical progress with imperfect competition. *Journal of Economics*, 101, 73-95.
- Dong, Y., Hamilton, R. and Tippett, M. (2014). Cost efficiency of the Chinese banking sector: A comparison of stochastic frontier analysis and data envelopment analysis. *Economic Modelling*, 36, 298-308.
- Drake, L. and Howcroft, B. (1994). Relative efficiency in the branch network of a UK bank: An empirical study. *Omega*, 22, 83-90.
- Drake, L. and Howcroft, B. (2002). An insight into the size efficiency of a UK bank branch network. *Managerial Finance*, 28, 24-36.
- Dyckhoff, H. and Allen, K. (2001). Measuring ecological efficiency with data envelopment analysis. *European Journal of Operations Research*, 132, 312-325.
- El Mehdi, R. and Hafner, C.M. (2014). Inference in stochastic frontier analysis with dependent error terms. *Mathematics and Computers in Simulation*, 102, 104-116.
- Eling, M. and Luhn, M. (2010). Efficiency in the international insurance industry: A cross-country comparison. *Journal of Banking & Finance*, 34(7), 1497-1509.
- Fan, Y., Li, Q. and Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics*, 14, 460-468.

- Fang, K.T., Kotz, S. and Ng, K.W. (1990). Symmetric multivariate and related distributions. Volume 36 of *Monographs on Statistics and Applied Probability*. Chapman & Hall.
- Färe, R. and Lovell, C.A.K. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory*, 19, 150–62.
- Färe, R., Grosskopf, S. and Lovell, C.A.K. (1985). *The measurement of efficiency of production*. Kluwer Academic Publishers.
- Färe, R., Grosskopf, S., Lindgren, B., and Roos, P. (1992). Productivity changes in Swedish pharmacies 1980-1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*, 3, 85-102.
- Färe, R., Grosskopf, S., Lindgren, B. and Roos, P. (1993). Productivity developments in Swedish hospitals: A Malmquist output index approach. In: A. Charnes, W.W. Cooper, A.Y. Lewin and L.M. Seiford, eds., *Data envelopment analysis: Theory, methodology and applications* (Kluwer, Boston) 253-271.
- Färe, R., Grosskopf, S., Lovell, C.A.K. (1994). *Production Frontiers*. University Press, Cambridge.
- Färe, R., Grosskopf, S., Norris, M. and Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1), 66-83.
- Färe, R., Primont, D., 1995. *Multi-Output Production and Duality: Theory and Applications*. Kluwer Academic Publishers, Boston.
- Färe, R., Grabowski, R., Grosskopf, S. and Kraft, S. (1997). Efficiency of a fixed but allocatable input: A non-parametric approach. *Economics Letters*, 56, 2, 187-193.

- Färe, R., Grosskopf, S. and Russell, R.R. (1998). *Index numbers: Essays in honour of Sten Malmquist*. Kluwer Academic Publishers, Dordrecht.
- Färe, R., Grosskopf, S., Maudos, J. and Tortosa-Ausina, E. (2010). Efficiency and market power in Spanish banking. 23rd Australasian Finance and Banking Conference 2010 Paper. Available at SSRN: <https://ssrn.com/abstract=1662142> or <http://dx.doi.org/10.2139/ssrn.1662142>.
- Farlie, D. J. G. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47, 307-323.
- Farrell, M.J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society*, 253-290.
- Farsi, M., Filippini, M., and Kuenzle, M. (2005). Unobserved heterogeneity in stochastic cost frontier models: An application to Swiss nursing homes. *Applied Economics*, 37 (18), 2127–2141.
- Feng, G. and Zhang, X. (2014). Returns to scale at large banks in the US: A random coefficient stochastic frontier approach. *Journal of Banking & Finance*, 39, 135-145.
- Fernández, C., Osiewalski, J. and Steel, M.F.J. (1997). On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics*, 79, 169-193.
- Ferrier, G. (2001). Bank efficiency and economic growth: The case of ASEAN. Manuscript. Department of Economics, University of Arkansas.
- Ferrier, G. D. and Lovell, C.A.K. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics*, 46, 229-245.

- Fethi, M.D. and Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research*, 204, 189–198.
- Filippini, M. and Greene, W. (2016). Persistent and transient productive inefficiency: a maximum simulated likelihood approach. *Journal of Productivity Analysis*, 45, 2, 187-196.
- Fiordelisi, F. and Molyneux, P. (2010). Total factor productivity and shareholder returns in banking. *Omega*, 38, 241-253.
- Fried, H., Lovell, C. and Yaisawarng, S. (1999). The impact of mergers on credit union service provision. *Journal of Banking & Finance*, 23, 367-386.
- Fried, H. O., Lovell, K. and Schmidt, S. S. (2008). Efficiency and productivity. In Fried, H. O., Lovell, K. and Schmidt, S. S. (eds.), *The measurement of productive efficiency and productivity growth*, 3-91.
- Fiordelisi, F. and Molyneux, P. (2010). Total factor productivity and shareholder returns in banking. *Omega*, 38, 241-253.
- Fuentes, H., Grifell, E., and S. Perelman (1998). A parametric distance function approach for Malmquist index estimation. Working Paper CREPP 98/03, Université de Liège.
- Fukuyama, H. (1995). Measuring efficiency and productivity growth in Japanese banking: A nonparametric frontier approach. *Applied Financial Economics*, 5, 95–117.
- Fun X. and Heffernan, S. (2007). Cost X-efficiency in China's banking sector. *China Economic Review*, 18, 35-53.

- Gallizo, L. JL., Moreno, G. J. and Salvador, F.M. (2017). The Baltic banking system in the enlarged European Union: the effect of the financial crisis on efficiency. *Baltic Journal of Economics*, 18, 1, 1–24.
- Gilbert, R. A. and Wilson, P. W. (1998). Effects of Deregulation on the Productivity of Korean Banks. *Journal of Economics and Business*, 50, 2, 133-155.
- Giokas, D.I. (1991). Bank branch operating efficiency: A comparative application of DEA and the loglinear model. *Omega*, 19, 549–557.
- Giokas, D.I. (2008a). Assessing the efficiency in operations of a large Greek bank branch network adopting different economic behaviors. *Economic Modeling*, 25, 559–574.
- Giokas, D.I. (2008b). Cost efficiency impact of bank branch characteristics and location. *Managerial Finance*, 34, 3, 172-185.
- Girardone, C., Molyneux, P. and Gardener, E. (2004). Analysing the determinants of bank efficiency: The case of Italian banks. *Applied Economics*, 36, 215-227.
- Gómez-Déniz, E. and Pérez-Rodríguez, J.V. (2015). Closed-form solution for a bivariate distribution in stochastic frontier models with dependent errors. *Journal of Productivity Analysis*, 43, 2, 215-223.
- Gómez-Déniz, E. and Pérez-Rodríguez, J.V. (2017). Stochastic Frontier Models with Dependent Errors based on Normal and Exponential Margins. *Revista de Métodos Cuantitativos para la Economía y la Empresa*, 23, 3-23.
- Gong, B-H. and Sickles, R.C. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *Journal of Econometrics*, 51, 259-284.
- Greene, W. (1980a). Maximum likelihood estimation of econometric frontier functions. *Journal of Econometrics*, 13, 1, 27-56.

- Greene, W. (1980b). On the estimation of a flexible frontier production model. *Journal of Econometrics*, 13, 1, 101-115.
- Greene, W. (1990). A gamma distributed stochastic frontier model. *Journal of Econometrics*, 46, 1, 141-164.
- Greene, W. (1998). *Econometric analysis*. Pearson Education.
- Greene, W. (2003). Maximum simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis*, 19, 2-3, 179-190.
- Greene, W. (2005a). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126, 269-303.
- Greene, W. (2005b). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, 23, 7-32.
- Greene, W. (2008). The econometric approach to efficiency analysis. In Fried, H. O., Lovell, K. and Schmidt, S. S. (eds.), *The Measurement of Productive Efficiency and Productivity Growth*, 92-250.
- Grifell, E. and Lovell, C.A.K. (1997). The sources of productivity change in Spanish banking. *European Journal of Operational Research*. 98, 364-380.
- Grifell-Tatjé, E. and Lovell, C.A. (1999). Profits and productivity. *Management Science*, 45, 1177-1193.
- Grifell-Tatjé, E. and Lovell, C.A.K. (1999). A generalized Malmquist productivity index. *Sociedad de Estadística e Investigación Operativa*. 7, 81-101.
- Griffin, J. and Steel, M. (2007). Bayesian stochastic frontier analysis using WinBUGS. *Journal of Productivity Analysis*, 27, 163-176.
- Groeneboom, P., Jongbloed, G. and Wellner, J.A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Annals of Statistics*, 29, 1653-1698.



- Grosskopf, S. (1993). Efficiency and productivity. In Fried, H.O. and Schmidt, S.S. (eds.) *The measurement of productive efficiency: Techniques and applications*, Oxford U.K.: 160-194.
- Gulati, R. and Kumar, S. (2016). Assessing the impact of the global financial crisis on the profit efficiency of Indian banks. *Economic Modelling*, 58, 167-181.
- Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of American Statistical Association*. 55, 698-707.
- Guzmán, I. and Reverte, C. (2008). Productivity and efficiency change and shareholder value: Evidence from the Spanish banking sector. *Applied Economics*, 40, 15, 2037-2044.
- Hadri, K. (1999). Estimation of a doubly heteroscedastic stochastic frontier cost function. *Journal of Business Economics and Statistics*, 17, 3, 359-363.
- Halkos, G. and Petrou, K.N. (2019). Treating undesirable outputs in DEA: A critical review. *Economic Analysis and Policy*, 62, 97-104.
- Hanoch, G. (1975). The elasticity of scale and the shape of average costs. *American Economic Review*, 65, 3, 492-497.
- Hanson, D.L. and G. Pledger, G. (1976). Consistency in concave regression. *Annals of Statistics*, 4, 6, 1038-1050.
- Hayami, Y. (1969). Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*, 51, 564-575.
- Hayami, Y. and Ruttan, V.W. (1970). Agricultural productivity differences among countries. *American Economic Review*, 40, 895-911.
- Hayami, Y. and Ruttan, V.W. (1971). *Agricultural development: An international perspective*. Johns Hopkins University Press, Baltimore.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press.

- Henderson, D.J. and Simar, L. (2005). A fully nonparametric stochastic frontier model for panel data. Discussion Paper 0417, Institut de Statistique, Universite Catholique de Louvain.
- Hildreth, C. (1954). Point estimates of ordinates of concave functions. *Journal of the American Statistical Association*, 49, 598-619.
- Hjalmarsson, L., Kumbhakar, S.C. and Heshmati, A. (1996). DEA, DFA, and SFA: A comparison. *Journal of Productivity Analysis*, 7, 2, 303-327.
- Horrace, W. C. and Schmidt, P. (1996). Confidence statements for efficiency estimates from stochastic frontier models. *Journal of Productivity Analysis*, 7, 257-282.
- Huang, C.J. and Liu, J.T. (1994). Estimation of a nonneutral stochastic frontier production function. *Journal of Productivity Analysis* 5, 171–180.
- Huang, C.J., Huang, T-H. and Liu, N.H. (2014). A new approach to estimating the metafrontier production function based on a stochastic frontier framework. *Journal of Productivity Analysis*, 42, 241–254.
- Hughes, J., Lang, L., Mester, L. and Moon, C. (1995). Recovering technologies that account for generalized managerial preferences: An application to non-risk neutral banks. Working Paper 95-16, Center for Financial Institutions, The Wharton School, University of Pennsylvania.
- Isik, I. and Hassan, M. K. (2003). Financial deregulation and total factor productivity change: An empirical study of Turkish commercial banks. *Journal of Banking & Finance*, 27, 1455–1485.
- Ji, Y. and Lee, C. (2010). Data envelopment analysis. *Stata Journal*, 10, 2, 267-280.  
<http://www.stata-journal.com/article.html?article=st0193>.

- Jondrow, J., Lovell, C., Materov, I. and Schmidt, P. (1982). Technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19, 233-238.
- Judge, G., Griffiths, W., Hill, R., Lütkepohl, H. and Lee, T. (1985): *The theory and practice of econometrics*. John Wiley & Sons. New York. Second Edition.
- Kim, Y. and Schmidt, P. (2000). A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. *Journal of Productivity Analysis*. 14, 91-118.
- Koop, G., Steel, M. and Osiewalski, J. (1995). Posterior analysis of stochastic frontier models using Gibbs sampling. *Computational Statistics*, 10, 353-373.
- Koop, G., Osiewalski, J. and Steel, M.F.J. (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. *Journal of Econometrics*, 76, 77–105.
- Koop, G. and Steel, M. (1999). Bayesian analysis of stochastic frontier models. Manuscript, University of Warwick. El document puede consultarse en la dirección: [http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/steel/steel\\_homepage/baltfin.pdf](http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/steel/steel_homepage/baltfin.pdf)
- Koopmans, T.C. (1951). An analysis of production as an efficient combination of activities. In Koopmans, T. C. (ed.): *Activity analysis of production and allocation*, Cowles Commission for Research in Economics, Monograph No. 13, John Wiley & Sons.
- Krüger, J.J. (2012). A Monte Carlo study of old and new frontier methods for efficiency measurement. *European Journal of Operational Research*, 222, 137–148.
- Kumar, M., Charles, V. and Misha, C.S. (2016). Evaluating the performance of indian banking sector using DEA during post-reform and global financial crisis. *Journal of Business Economics and Management*, 17, 1, 156-172.

- Kumbhakar, S.C. (1990). Production frontiers, panel data and time-varying technical efficiency. *Journal of Econometrics*, 46, 201-211.
- Kumbhakar, S.C. (1991). The measurement and decomposition of cost-inefficiency: The Translog cost system. *Oxford Economic Papers, New Series*, 43(4), 6767-683.
- Kumbhakar, S.C., Ghosh, S. and McGuckin, J.T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business and Economic Statistics*, 9, 279-286.
- Kumbhakar S.C. and Hjalmarsson, L. (1993). Technical efficiency and technical progress in Swedish dairy farms. In H.O. Fried, C.K. Lovell and S.S. Schmidt (Eds.), *The measurement of productive efficiency. Techniques and applications* (pp. 226-270). Oxford University Press.
- Kumbhakar S. C. and Heshmati, A. (1995). Efficiency measurement in Swedish dairy farms: An application of rotating panel data. *American journal of Agricultural Economics*, 77, 3, 660-674.
- Kumbhakar, S. C. and Hjalmarsson, L. (1995). Labour-use efficiency in Swedish social insurance offices. *Journal of Applied Econometrics*, 10, 1, 33-47.
- Kumbhakar, S. C. and Lovell, K. (2000). *Stochastic frontier analysis*. Cambridge University Press.
- Kumbhakar, S. C., Lozano-Vivas, A., Lovell, C.A.K. and Hasan, I. (2001). The effects of deregulation on the performance of financial institutions: The case of Spanish saving banks. *Journal of Money, Credit, and Banking*, 33, 101-121.
- Kumbhakar, S.C. and Tsionas, E. (2005). Measuring technical and allocative inefficiency in the translog cost system: A Bayesian approach. *Journal of Econometrics*, 126, 355-384.

- Kumbhakar, S.C., Park, B.U., Simar, L. and Tsionas, E.G. (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics*, 137, 1, 1-27.
- Kumbhakar, S.C., Parmeter, C.F. and Tsionas, E.G. (2013). A zero inefficiency stochastic frontier model. *Journal of Econometrics*, 172, 66-76.
- Kuosmanen, T. (2006). Stochastic nonparametric envelopment of data: Combining virtues of SFA and DEA in a unified framework. MTT Discussion Paper No. 3/2006.
- Kuosmanen, T., Kortelainen, M., Kultti, K., Pursiainen, H., Saastamoinen, A. and Sipiläinen, T. (2010). Sähköverkkotoiminnan kustannustehokkuuden estimointi StoNED – menetelmällä. Sigma-Hat Economics.
- Kuosmanen, T. and Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38, 11-28.
- Kuosmanen, T., Johnson, A.L. and Saastamoinen, A. (2015). Stochastic nonparametric approach to efficiency analysis: A unified framework. In Zhu, J. (Ed.) *Handbook on Data Envelopment Analysis*, Vol. 2, pages 191-244, Springer.
- Laeven, L. (1999). Risk and efficiency in East Asian banks. Policy Research Working Paper 2255, The World Bank, Washington, D.C.
- Lancaster, T. (2002). The incidental parameters problem since 1948. *Journal of Econometrics*, 95: 391-414.
- Lee, L.-F. (1983). A test for distributional assumptions for the stochastic frontier functions. *Journal of Econometrics*, 22, 3, 245-267.
- Lee, Y. and Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical inefficiency. In Fried, H. O., Lovell, K. and Schmidt, S. S.,

- eds. *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York.
- Leightner, J. E. and Lovell, C. K. (1998). The Impact of Financial Liberalization on the Performance of Thai Banks. *Journal of Economics and Business*, 50, 115-131.
- Levine, R. (2005). The microeconomic effects of different approaches to bank supervision. Conference on Economics, Political Institutions, and Financial Markets II: Institutional Theory and Evidence from Europe, the United States, and Latin America, February 4-5, Stanford University, available at: <http://sshi.stanford.edu/Conferences/2004-2005/FinMarkets2005.htm>
- Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70, 317-341.
- Lieu, P.T., Yeh, T.-L. and Chiu, Y.-H. (2005). Off-balance sheet activities and cost inefficiency in Taiwan's Banks. *Service Industries Journal*, 25, 7, 925-944.
- Lin, X. and Zhang, Y. (2009). Bank ownership reform and bank performance in China. *Journal of Banking and Finance*, 33, 1, 20-29.
- Lin, K.L, Doan, A.T. and Doong, S.C. (2016). Changes in ownership structure and bank efficiency in Asian developing countries: The role of financial freedom. *International Review of Economics and Finance*, 43, 19-34.
- Lindskog, F., McNeil, A.J., Schmock, U., (2003). Kendall's tau for elliptical distributions. In: Bol, G., Nakhaeizadeh, G., Rachev, S. T., Ridder, T., Vollmer, K.-H. (Eds.), *Credit Risk–Measurement, Evaluation and Management*. Physica–Verlag, Heidelberg, 149-156.
- Llorens, V., Martín-Oliver, A. and Salas-Fumas, V. (2020). Productivity, competition and bank restructuring process. *SERIEs*, 11, 313-340.

- Lovell, C., Grosskopf, S., Ley, E., Pastor, J., Prior, D. and Eeckaut, P. (1994). Linear programming approaches to the measurement and analysis of productive efficiency. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 2, 2, 175-248.
- Lovell, C.A.K. and Pastor, J.T. (1997). Target setting: An application to a bank branch network. *European Journal of Operational Research*, 98, 290–299.
- Lozano-Vivas, A., Pastor, J.T. and Hasan, I. (2001). European bank performance beyond country borders: What really matters? *European Finance Review*, 5, 1-2, 141-165.
- Lozano-Vivas, A., Pastor, J.T. y Pastor, J.M. (2002). An efficiency comparison of European banking systems operating under different environmental conditions. *Journal of Productivity Analysis*, 18, 59-77.
- Manandhar, R. and Tang, J.C.S. (2002). The evaluation of bank branch performance using data envelopment analysis: A framework. *Journal of High Technology Management Research*, 13, 1–17.
- Manlagñit, M.C.V. (2011). Cost efficiency, determinants, and risk preferences in banking: A case of stochastic frontier analysis in the Philippines. *Journal of Asian Economics*, 22, 23-35.
- Martins-Filho, C. and Yao, F. (2010). Nonparametric stochastic frontier estimation via profile likelihood. URL <http://spot.colorado.edu/~martinsc/Research.html>.
- Martín-Vallespín, E., Bachiller, A. y Bachiller, P. (2016). La eficiencia del sistema bancario español: impacto de la reestructuración de entidades financieras. *Presupuesto y Gasto Público*, 83, 141-164.
- Marzec J. and Osiewalski J. (2008). Bayesian inference on technology and cost efficiency of bank branches. *Financial Markets and Institutions*, 29-43.

- Maudos, J. (1996). Eficiencia, cambio técnico y productividad en el sector bancario español: Una aproximación de frontera estocástica. *Investigaciones Económicas*, N° XX, 339-358.
- Maudos, J. (2001). Rentabilidad, estructura de mercado y eficiencia en el sector bancario español. *Revista de Economía Aplicada*, 25, 9, 193-207.
- Maudos, J. and Pastor, J.M. (2001a). La eficiencia del sistema bancario español en el contexto de la Unión Europea. *Papeles de Economía Española*, Núm. 84/85, 155-168.
- Maudos, J. and Pastor, J.M. (2001b). Cost and profit efficiency in banking: An international comparison of Europe, Japan and the USA. *Applied Economics Letters*, 8(6), 383–387.
- Maudos, J., Pastor, J.M. and Pérez, F. (2002). Competition and efficiency in the Spanish banking sector: The importance of specialisation. *Applied Financial Economics*, 12, 505-516.
- Maudos, J. and Pastor, J.M. (2003). Cost and profit efficiency in the Spanish banking sector (1985–1996): A non-parametric approach. *Applied Financial Economics*, 13, 1, 1-12.
- McAllister, P. and McManus, D. (1993). Resolving the scale efficiency puzzle in banking. *Journal of Banking and Finance*, 17, 389-405.
- McEachern, D. and Paradi, J.C. (2007). Intra- and inter-country bank branch assessment using DEA. *Journal of Productivity Analysis*, 27, 123–136.
- Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, Vol. 18, 2, 435-444.



- Meslier, C., Tacneng, R. and Tarazi, A. (2014). Is bank income diversification beneficial? Evidence from an emerging economy. *Journal of International Financial Markets, Institutions and Money*, 31, 97-126.
- Mester, L. (1996). A study of bank efficiency taking into account risk-preferences. *Journal of Banking and Finance*, 20, 1025-1045.
- Mester, L. (1997). Measuring efficiency at U.S. banks: Accounting for heterogeneity is important. *European Journal of Operational Research*, 98, 230–248.
- Miller, D. (2005). An information theoretic approach to flexible stochastic frontier models. Unpublished manuscript.
- Mohanty, S.K., Lin, W.T. and Lin, H-J. (2013). Measuring cost efficiency in presence of heteroskedasticity: The case of the banking industry in Taiwan. *Journal of International Financial Markets, Institutions and Money*, 26, 77-90.
- Molnar, M. and Holló, D. (2019). How Efficient Are Banks in Hungary? OECD Economics Department Working Papers. URL <https://doi.org/10.1787/18151973>.
- Morgenstern, D. (1956). Einfache Beispiele zweidimensionaler Verteilungen. *Mitteilungsblatt für Mathematische Statistik*, 8, 234-235.
- Mukherjee, K., Ray, S. C. and Miller, S. (2001). Productivity growth in large US commercial banks: The initial post-deregulation experience. *Journal of Banking & Finance*, 25, 913–39.
- Nelsen, R. B. (1999). *An introduction to copulas*. New York: Springer.
- Nguyen T. (2018). Diversification and bank efficiency in six ASEAN countries. *Global Finance Journal*, 37, 57-78.
- Noulas, A. G. (1997). Productivity growth in the Hellenic banking industry: State versus private banks. *Applied Financial Economics*, 7, 223-228.

- O'Donnell, C. J., Rao, D. S. P. and Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*, 34, 231-255.
- O'Donnell, C.J. (2009). An aggregate quantity-price framework for measuring and decomposing productivity and profitability change. CEPA - School of Economics, The University of Queensland.  
<http://www.uq.edu.au/economics/cepa/docs/WP/WP072008.pdf>.
- Oral, M. and Yolalan, R. (1990). An empirical study on measuring operating efficiency and profitability of bank branches. *European Journal of Operational Research*, 46, 282-294.
- Orea, L. (2000). A parametric decomposition of a generalized Malmquist-type productivity index. Documentos de trabajo, Universidad de Oviedo. Facultad de Ciencias Económicas, N°. 233, 2001, 20 págs.
- Orea, L. and Kumbhakar, S.C. (2004). Efficiency measurement using a latent class stochastic frontier model. *Empirical Economics*, 29, 1, 169–183.
- Osiewalski J. and Marzec J. (1998). Bayesian analysis of cost efficiency with an application to bank branches, in: E. Miklaszewska (ed.), *Global Tendencies and Changes in East European Banking*, Jagiellonian University, Cracow.
- Paradi, J.C., Vela, S. and Zhu, H. (2010). Adjusting for cultural differences, a new DEA model applied to a merged bank. *Journal of Productivity Analysis*, 33, 109–123.
- Paradi, J., Rouatt, S. and Zhu, H. (2011). Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega*, 29, 99-109.
- Paradi, J. and Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega*, 41, 61–79.

- Park, B., and Simar, L. (1994). Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association*, 89, 427, 929-936.
- Park, B., Sickles, R.C. and Simar, L. (1998). Stochastic panel frontiers: A semiparametric approach. *Journal of Econometrics*, 84, 273-301.
- Park, B., Sickles, R.C. and Simar, L. (2003). Semiparametric efficient estimation of AR(1) panel data models. *Journal of Econometrics*, 117, 279-309.
- Parkan, C. (1987). Measuring the efficiency of service operations: an application to bank branches. *Engineering Costs and Production Economics*, 12, 237-242.
- Pasiouras, F. (2008). International evidence on the impact of regulations and supervision on banks' technical efficiency: an application of two-stage data envelopment analysis. *Review of Quantitative Finance and Accounting*, 30, 187-223.
- Pasiouras, F., Tanna, S. and Zopounidis, C. (2009). The impact of banking regulations on banks' cost and profit efficiency: cross-country evidence. *International Review of Financial Analysis*, 18(5):294-302.
- Pastor, J.M. (1995). Eficiencia, cambio productivo y cambio técnico en los bancos y cajas de ahorros españolas: Un análisis de la frontera no paramétrico. *Revista Española de Economía*, 12, 1, 35-73.
- Pastor, J.M., Pérez, F. and Quesada, J. (1997). Efficiency analysis in banking firms: an international comparison. *European Journal of Operational Research*, 98, 395-407.
- Pastor, J.T., Ruiz, J.L. and Sirvent, I. (1999). An enhanced DEA Russell graph efficiency measure. *European Journal of Operational Research*, 115, 596-607.

- Pastor, J. and Serrano, L. (2005). Efficiency, endogenous and exogenous credit risk in the banking systems of the Euro area. *Applied Financial Economics*, 15, 631-649.
- Pastor, J., Lovell, C. and Tulkens, H. (2006). Evaluating the financial performance of bank branches. *Annals of Operations Research* 145, 321–337.
- Patton, A. (2006a). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21, 147-173.
- Patton, A. (2006b). Modelling asymmetric exchange rate dependence. *International Economic Reviews*, 47, 2, 527-556.
- Patton, A. (2012). A review of copula models for economic time series. *Journal of Multivariate Analysis*, 100, 4-18.
- Peristiani, S. (1997). Do mergers improve X-efficiency and scale efficiency of U.S. banks? *Journal of Money, Credit and Banking*, 29, 326-337.
- Pitt, M. and Lee, L. (1981). The measurement and sources of technical inefficiency in Indonesian weaving industry. *Journal of Development Economics*, 9, 43-64.
- Plackett, R.L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association*, 60, 516–522.
- Politis, D.N., Romano, J.P. and Wolf, M. (2001). On the asymptotic theory of subsampling. *Statistica Sinica*, 11, 1105–1124.
- Portela, M.C.A.S. and Thanassoulis, E. (2007). Comparative efficiency analysis of Portuguese bank branches. *European Journal of Operational Research*, 177, 1275–1288.
- Prior, D., Tortosa-Ausina, E., García-Alcober, M.P. and Illueca, M. (2016). Costs, revenues and performance in Spanish banking: A comparative analysis of pre-

- and early crisis years. *Spanish Journal of Finance and Accounting*, 45, 3, 345-364.
- Rai, A. (1996). Cost efficiency of international insurance firms. *Journal of Financial Service Research*, 10, 213-233.
- Rao, D.S.P., O'Donnell, C.J. and Battese, G.E. (2003). Metafrontier functions for the study of interregional productivity differences. Centre for Efficiency and Productivity Analysis (CEPA). Working Paper Series No. 01/2003.
- Reifschneider, D. and Stevenson, R. (1991). Systematic departures from the frontier: A framework for the analysis of firm inefficiency. *International Economic Review*, 32, 715-723.
- Respaut, B. (1989). Mesures de l'efficacité productive des 911 agences d'une banque privée Belge. Chapter 3, in H. Tulkens (ed.), *Efficacité et Management*. CIFOP, Charleroi, Belgium.
- Resti, A. (1997). Evaluating the cost efficiency of the Italian banking system: What can be learned from the joint application of parametric and non-parametric techniques. *Journal of Banking and Finance*, 21, 221-250.
- Rezitis A. N. (2006). Productivity Growth in the Greek Banking Industry: A Non-Parametric Approach. *Journal of Applied Economics*, 9, 1, 119-138.
- Richmond, J. (1974). Estimating the efficiency of production. *International Economic Review*, 15, 2, 515-521.
- Sánchez, B., Hassan, K.M. and Bartkus, J.R. (2013). Efficiency determinants and dynamic efficiency changes in Latin American banking industries. *Journal of CENTRUM Cathedra: The Business and Economics Research Journal*, 6(1), 27-52.

- San-Jose, L., Retolaza, J.L. and Pruñonosa, J.T. (2014). Efficiency in Spanish banking: A multistakeholder approach analysis. *Journal of International Financial Markets, Institutions and Money*, 32, 240–255.
- Schaffnit, C., Rosen, D. and Paradi, J.C. (1997). Best practice analysis of bank branches: An application of DEA in a large Canadian bank. *European Journal of Operational Research*, 98, 269–289.
- Schmidt, P. (1976). On the statistical estimation of parametric frontier production functions. *Review of Economics and Statistics*, 58, 2, 238-239.
- Schmidt, P. and Sickles, R. (1984). Production frontiers and panel data. *Journal of Business Economics and Statistics*, 2, 4, 367-374.
- Seiford, L.M. (1996). Data envelopment analysis: The evolution of the state of the art (1978-1995). *Journal of Productivity Analysis*, 7, 99-138.
- Sheldon, G. (1994). Economies, inefficiencies and technical progress in Swiss banking. In D. Fair and R. Raymond (Eds.), *The competitiveness of financial institutions and centers in Europe*, 115-133. Kluwer.
- Sherman, H.D. and Gold, F. (1985). Bank branch operating efficiency: Evaluation with data envelopment analysis. *Journal of Banking and Finance*, 9, 297-315.
- Silva Portela, M.C.A., Thanassoulis, E. and Simpson, G. (2004). Negative data in DEA: a directional distance approach applied to bank branches. *Journal of the Operational Research Society*, 55(10), 1111-1121.
- Simar, L. (1996). Aspects of statistical analysis in DEA-type frontier models. *Journal of Productivity Analysis*, 7, 177–185.

- Simar, L. and Wilson, P.W. (2007). Estimation and inference in two stage, semi-parametric models of productive efficiency. *Journal of Econometrics*, 136, 31–64.
- Simar, L. and Wilson, P.W. (2013). Estimation and inference in nonparametric frontier models: Recent developments and perspectives. *Foundations and Trends in Econometrics*, 5, 2, 183-337.
- Sklar, A. (1959). Fonctions de répartition à  $n$  dimensions et leurs marges”. Publications de l’Institute de Statistique de l’Universite de Paris 8, 229–31.
- Sklar, A. (1973). Random variables, joint distributions, and copulas. *Kybernetika*, 9, 449–60.
- Smith, M. (2008). Stochastic frontier models with dependent error components. *Econometrics Journal*, 11, 172-192.
- Stead, AD, Trigo, P. and Greene, W.H. (2018). Estimating efficiency in the presence of extreme outliers: A Logistic-Half Normal stochastic frontier model with application to highway maintenance costs in England. Chapter 1 in: W.H. Greene et al. (eds.), *Productivity and Inequality*, Springer Proceedings in Business and Economics, págs. 1-19.
- Stevenson, R. (1980). Likelihood functions for generalized stochastic frontier functions. *Journal of Econometrics*, 13, 57-66.
- Sun, L. and Chang, T.P. (2011). A comprehensive analysis of the effects of risk measures on bank efficiency: Evidence from emerging Asian countries. *Journal of Banking and Finance*, 35, 7, 1727-1735.
- Taube, R. (1988). Möglichkeiten der effizienzmessung von öffentlichen verwaltungen. Duncker & Humbolt GmbH, Berlin.

- Tauchmann, H. (2011). Orderalpha: Non-parametric order- $\alpha$ . Efficiency analysis for Stata. Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI). July 1st, 2011, German Stata Users Group Meeting, University of Bamberg.
- Tauchmann, H. (2012). Partial frontier efficiency analysis. *Stata Journal*, 12, 3, 46-478. <http://www.stata-journal.com/article.html?article=st0270>.
- Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130, 498–509.
- Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research*, 143, 32–41.
- Tone, K. (2002). A strange case of the cost and allocative efficiencies in DEA. *Journal of Operations Research Society*, 53, 1225-1231.
- Tsionas, E.G., and Kumbhakar, S.C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random effects model. *Journal of Applied Econometrics*, 29, 1, 110-132.
- Tsolas, I.E. and Giokas, D.I. (2012). Bank branch efficiency evaluation by means of least absolute deviations and DEA. *Managerial Finance*, 38, 8, 768-785.
- Tulkens, H. (1993). On FDH analysis: Some methodological issues and applications to retail banking, courts and urban transit. *Journal of Productivity Analysis*, 4(1), 183–210.
- Tulkens, H. and Malnero, A. (1996). Nonparametric approaches to the assessment of the relative efficiency of bank branches. In D.G. Mayes (ed.), *Sources of Productivity Growth*. Chapter 10, Cambridge University Press, New York.
- Valverde, S., Humphrey, D. and del Paso, R. (2007). Opening the black box: Finding the source of cost inefficiency. *Journal of Productivity Analysis*, 27, 209-220.



- Vallespín, E.M., Cacho A.B. and Baroja P.B. (2016) La eficiencia del sistema bancario español: impacto de la reestructuración de entidades financieras. *Presupuesto y Gasto Público*, 83/2016, 141-164.
- Van den Broeck, J., Koop, G., Osiewalski, J. and Steel, M. (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics*, 61, 273-303.
- Van der Venet, R. (2002). Cost and profit efficiency of financial conglomerates and universal banks in Europe. *Journal of Money, Credit and Banking*, 34, 1, 254-282.
- Vassiloglou, M. and Giokas, D.I. (1990). A study of the relative efficiency of bank branches: an application of data envelopment analysis. *Journal of the Operational Research Society*, 41, 591–597.
- Wang, H., and Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, 18, 2, 289-296.
- Wang, H., W, Wu J, and Liu Y-N. (2014). Efficiency measures of the Chinese commercial banking system using an additive two-stage DEA. CEEP-BIT Working Paper, 5.
- Weill, L. (2004). Measuring cost efficiency in European banking: A comparison of frontier techniques. *Journal of Productivity Analysis*, 21, 2, 133-152.
- Weill, L. (2008). Convergence in banking efficiency across European countries. Working Papers of LaRGE (Laboratoire de Recherche en Gestion et Economie) 2008-07, Louis Pasteur University, Strasbourg (France).
- Wheat, P., Greene, B. and Smith, A. (2014). Understanding prediction intervals for firm specific inefficiency scores from parametric stochastic frontier models. *Journal of Productivity Analysis*, 42, 55-65.

- Wheelock, D.C. and Wilson, P. (2003). Robust nonparametric estimation of efficiency and technical change in US commercial banking. *Federal Reserve Bank of St. Louis Working Paper Series*, (2003-037).
- Wiboonpongse, A., Liu, J., Sriboonchitta, S. and Denoeux, T. (2015). Modeling dependence between error components of the stochastic frontier model using copula: Application to intercrop coffee production in Northern Thailand. *International Journal of Approximate Reasoning*, 65, 34–44.
- Winsten, C. B. (1957). Discussion on Mr. Farrell's paper. *Journal of the Royal Statistical Society, Series A*, 120, 3, 282-284.
- Worthington, A. (1999). Malmquist indices of productivity change in Australian financial services. *Journal of International Financial Markets, Institutions and Money* 9, 3, 303-320.
- Wu, D., Yang, Z. and Liang, L. (2006). Efficiency analysis of cross-region bank branches using fuzzy data envelopment analysis. *Applied Mathematics and Computation*, 181, 271–281.
- Xiaogang, C., Skully, M. and Brown, K. (2005). Banking efficiency in China: Application of DEA to pre- and post-deregulation eras: 1993–2000. *China Economic Review*, 16, 229-245.
- Yang, C. and Liu, H-M. (2012). Managerial efficiency in Taiwan bank branches: A network DEA. *Economic Modelling*, 29, 450–461.
- Yildirim, H.S. and Philippatos, G.C. (2007). Efficiency of banks: recent evidence from the transition economies of Europe, 1993–2000. *European Journal of Finance*, 13, 2, 123-143.

Yung-Ho, C. and Chen, Y. (2009). The analysis of Taiwanese bank efficiency incorporating both external environment risk and internal risk. *Economic Modelling*, 26, 456-463.

# APÉNDICE I

## ALGUNAS CÓPULAS ELÍPTICAS Y ARCHIMEDEANAS

---

El uso de la teoría de cópulas permite una gran flexibilidad para modelar la distribución conjunta de un par de variables aleatorias, y lo hace de forma sencilla ya que solo se necesita especificar la función copula y las marginales.

Las cópulas pueden extraer la estructura de dependencia de la función de distribución conjunta de un vector de variables aleatorias y, al mismo tiempo, permiten separar la estructura de dependencia de su función marginal.

Existen muchos tipos de cópulas. Las más frecuentemente utilizadas son las cópulas elípticas y Arquimedeanas.

Una cópula elíptica es aquella que corresponde a una distribución elíptica según el teorema de Sklar. La discusión general sobre las distribuciones elípticas se puede encontrar en Fang et al., (1990). Los generadores para cópulas Arquimedeanas se pueden encontrar en Nelsen (1999) (por ejemplo, Farlie-Gumbel-Morgenstern (FMG) y la cópula de Frank).

La elección del modelo de cópula se basa en el análisis previo, la tractabilidad y la pertenencia a una clase bastante amplia que permita la agrupación de observaciones extremas y la posible asimetría.

En este apéndice, hemos elegido comentar algunas cópulas como la Gaussiana, Gumbel, Clayton y la simétrica de Joe-Clayton.

Por ejemplo, una cópula Gaussiana es una copula elíptica que implica simetría y dependencia cero en extremas colas. Asumiendo que  $u_1 = F(u)$  y  $v_1 = F(v)$ , en este caso, la cópula de densidad bivariada Gaussiana se define por:

$$c_{\rho}(F(u), F(v)) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2} X'R^{-1}X\right) \exp\left(-\frac{1}{2} \sum_{i=1}^2 x_i^2\right)^{-1},$$

donde  $R = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  es una matriz de correlaciones con  $-1 \leq \rho \leq 1$ ,

$X = [\Phi^{-1}(F(u)) \quad \Phi^{-1}(F(v))]'$ , con  $x_j, j=1,2$  siendo el elemento  $j$ -ésimo del vector

$X$ , y  $\Phi^{-1}(\cdot)$  la inversa de la cdf normal estándar. Esta cópula elíptica presenta algunas características atractivas como ser simétrica (en el sentido de que la dependencia positiva y negativa son tratadas igualmente) y también que los coeficientes de asociación miden los co-movimientos de variables en presencia de relaciones no lineales como la medida de Spearman,  $r_{\rho} \in (-1,1)$ , que puede calcularse en forma

cerrada por  $r_{\rho} = \frac{6}{\pi} \arcsin\left(\frac{\rho}{2}\right)$ , o el coeficiente de  $\tau$  de Kendall, que puede calcularse

mediante  $\tau = \frac{2}{\pi} \arcsin(\rho)$ , donde  $\rho \in (-1,1)$ .

En el caso de la cópula de  $t$  de Student, que permite eventos extremos conjuntos, la función de densidad bivariada se define como:

$$c_{\rho, \nu}(u_1, v_1) = \frac{1}{|R|^{1/2}} \frac{\Gamma\left(\frac{\nu+2}{2}\right) \Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)^2} \left(1 + \frac{X'R^{-1}X}{\nu}\right)^{-\left(\frac{\nu+2}{2}\right)} \left[ \prod_{j=1}^2 \left(1 + \frac{x_j^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \right]^{-1}$$

donde  $\Gamma(\cdot)$  es la función de Gamma y  $X = [t_{\nu}^{-1}(u_1) \quad t_{\nu}^{-1}(v_1)]'$ , con  $t_{\nu}^{-1}(\cdot)$  siendo la función cdf inversa de la  $t$ -Student con  $\nu$  grados de libertad. La estimación de la  $\tau$  de Kendall puede obtenerse usando las relaciones antes mencionadas, que son válidas para distribuciones elípticas (Lindskog et al., 2003).

En lo que respecta a las cópulas Arquimedianas, la densidad cópula de Gumbel, que permite una mayor dependencia en la cola positiva que en la cola negativa, se define por:

$$c_{\kappa}(u_1, u_2) = C_{\kappa}(u_1, v_1) (\log(u_1) \log(v_1))^{\kappa-1} (u_1 v_1)^{-1} \left[ (-\log(u_1))^{\kappa} + (-\log(v_1))^{\kappa} \right]^{\left(-2-\frac{1}{\kappa}\right)} \\ \times \left[ \left[ (-\log(u_1))^{\kappa} + (-\log(v_1))^{\kappa} \right]^{\frac{1}{\kappa}} + \kappa - 1 \right]$$

donde  $\kappa$  es el parámetro de dependencia, y

$$C_{\kappa}(u_1, u_2) = \exp \left[ - \left[ \left( (-\log(u_1))^{\kappa} + (-\log(v_1))^{\kappa} \right)^{\frac{1}{\kappa}} \right] \right],$$
 es la distribución cópula

bivariada. En este caso, el parámetro de asociación es  $\kappa = (1 - \tau)^{-1}$ , que depende de la  $\tau$  de Kendall. La cópula de Gumbel implica una mayor dependencia en la cola derecha de las distribuciones marginales.

También puede usarse la cópula de Clayton, que es otra cópula asimétrica Arquimedeana. Esta cópula permite una mayor dependencia en la cola negativa que en la positiva. Es decir, una mayor dependencia para eventos negativos conjuntos que para eventos positivos conjuntos. Su función de densidad bivariada se define como:

$$c_{\kappa}(u_1, v_1) = (1 + \kappa) \left[ (u_1)^{-\kappa} + (v_1)^{-\kappa} - 1 \right]^{\left(-2-\frac{1}{\kappa}\right)} (u_1 v_1)^{-(\kappa+1)}$$

donde  $\kappa = \frac{2\tau}{1-\tau}$ , que depende de la  $\tau$  de Kendall. La cópula de Clayton implica una

mayor dependencia en las colas izquierdas. Ambas cópulas Archimedeanas son cópulas explícitas.

Finalmente, podemos exponer brevemente la cópula de Joe-Clayton.

Esta es una modificación de la copula "BB7" de Joe (1997). Patton (2006b) se refiere a la copula BB7 como la cópula de Joe-Clayton, ya que se construye tomando una transformación particular de Laplace de la copula de Clayton.

La cópula de Joe-Clayton es:

$$C_{\tau_U, \tau_L}^{JC}(u_1, v_1) = 1 - \left( 1 - \left\{ \left[ 1 - (1 - u_1)^\kappa \right]^{-\gamma} + \left[ 1 - (1 - v_1)^\kappa \right]^{-\gamma} - 1 \right\}^{-1/\gamma} \right)^{1/\kappa}$$

donde  $\kappa = \frac{1}{\log_2(2 - \tau_U)}$ ,  $\gamma = -\frac{1}{\log_2(\tau_L)}$ ,  $\tau_U \in (0,1)$  y  $\tau_L \in (0,1)$ .

La copula de Joe-Clayton tiene dos parámetros,  $\tau_U$  y  $\tau_L$ , que son medidas de dependencia conocidas como dependencia en las colas (*tail dependence*). Patton (2006b, páginas 541-542) indica que si el límite existe, entonces la cópula exhibe menor dependencia de la cola inferior si  $\tau_L \in (0,1]$  y no existe dependencia en la cola inferior si  $\tau_L = 0$ . De forma similar, si el límite  $\tau_U$  existe, entonces la cópula exhibe dependencia en la cola superior si  $\tau_U \in (0,1]$  y si  $\tau_U = 0$ , entonces no hay dependencia en la cola superior. La dependencia en las colas captura eventos extremos en la conducta de las variables aleatorias analizadas.

Debido a que un inconveniente principal de la cópula de Joe-Clayton es que incluso cuando las dos medidas de dependencia de las colas sean iguales, todavía hay una (leve) asimetría en la cópula de Joe-Clayton, debido simplemente a la forma funcional de esta cópula. Un modelo más deseable que tendría medidas de dependencia de las colas que determinarían completamente la presencia o ausencia de asimetría es la denominada cópula simétrica de Joe-Clayton. La distribución de la cópula bivariada se expresa como sigue:

$$C_{\tau_U, \tau_L}^{SJC}(u_1, v_1) = 0.5 \left( C_{\tau_U, \tau_L}^{JC}(u_1, v_1) + C_{\tau_U, \tau_L}^{JC}(1 - u_1, 1 - v_1) + u_1 + v_1 - 1 \right)$$

La cópula simétrica de Joe-Clayton (SJC) es claramente una ligera modificación de la cópula original de Joe-Clayton, pero por construcción es simétrica cuando  $\tau_U = \tau_L$ .