Revista Española de Economía, Vol. 15, nº 2, 1998 **265-271**

Fronteras de producción eficiente y convergencia regional: un comentario al trabajo de Maudos, Pastor y Serrano

Ángel de la Fuente
Instituto de Análisis Económico, CSIC.

Hace unos meses recibí el encargo de evaluar el trabajo de Maudos, Pastor y Serrano que aparece en el presente número de *Revista Española de Economía*. Mi reacción al artículo fue bastante ambivalente. Por un lado, se trataba de un trabajo razonablemente bien escrito, técnicamente competente y muy interesante por cuanto utilizaba una técnica relativamente nueva en el campo de crecimiento (al menos a nivel español) y ofrecía una visión del proceso de convergencia regional muy diferente de la que se desprende de la mayor parte de la literatura preexistente. Por otro lado, pensaba (y sigo haciéndolo) que la metodología empleada por Maudos y otros, no es la más indicada para analizar los determinantes del crecimiento, y que esto puede explicar la diferencia de resultados en relación a contribuciones anteriores. Aun así, soy consciente de que mis objecciones entran dentro del terreno de lo opinable. Por eso no me he opuesto a la publicación del trabajo y lo que era inicialmente un informe de evaluación se ha convertido en este comentario. Espero que la discusión abierta y razonada de las ventajas e inconvenientes de enfoques alternativos contribuya al progreso técnico (o a la reducción de la ineficiencia X) en este área de investigación.

En esencia, mis objecciones al trabajo de Maudos y otros, son dos. La primera es que, cuando trabajamos con agregados regionales o nacionales, no es buena idea estimar una frontera de producción eficiente en vez de una función de producción tradicional —y menos aún si lo hacemos utilizando una técnica de programación lineal no estocástica. La segunda es que, dada la muestra con que estamos trabajando, la técnica de estimación empleada no genera una descripción razonable de la tecnología— lo que contribuye a reforzar mi opinión de que el punto de partida no es el correcto.

1. ¿Eficiencia técnica o ineficiencia X?

Controlando por las dotaciones de factores, la productividad de dos regiones puede diferir por al menos dos razones. La primera es que una de ellas podría tener acceso a una tecnología más avanzada, lo que le permitiría extraer un mayor *output* de un volumen dado de recursos. La segunda es que, incluso teniendo acceso a la misma tecnología, las dos economías podrían diferir en la eficacia de sus procesos productivos o en la calidad de la gestión de sus empresas (esto es, en sus niveles de eficiencia X). Parece razonable pensar que ambos tipos de diferencias pueden presentarse en la práctica. Creo que estamos todos de acuerdo, además, en que lo ideal sería poder cuantificar ambos factores y aislar sus contribuciones respectivas a los diferenciales de productividad que observamos entre países o regiones.

El problema es que es prácticamente imposible separar las dos cosas. En consecuencia, las estimaciones existentes en la literatura parten de uno de dos supuestos extremos. Maudos y otros (como todos los que optan por estimar fronteras de producción eficiente) suponen que las regiones difieren en su nivel de eficiencia X pero tienen acceso a una tecnología común en cada momento del tiempo. Esto les permite estimar una única frontera, que refleja la producción máxima alcanzable en función de las dotaciones de factores productivos, y medir la ineficiencia por la distancia a ella. La alternativa más común en la literatura es la opuesta: se supone que las diferencias de eficiencia entre regiones son poco significativas e incorreladas con otras variables explicativas (lo que permite subsumirlas en el término de error) y se permiten diferencias de nivel entre las distintas funciones de producción regionales que se interpretan como indicadores del nivel de desarrollo tecnológico de cada economía.

La elección entre estos dos supuestos es más importante de lo que podría parecer a primera vista porque no resulta neutral a la hora de realizar la estimación. No se trata tan sólo de decidir que etiqueta vamos a ponerle a los diferenciales de productividad no explicables en términos de dotaciones de recursos, sino también de cómo determinar los coeficientes de estos factores. La figura 1

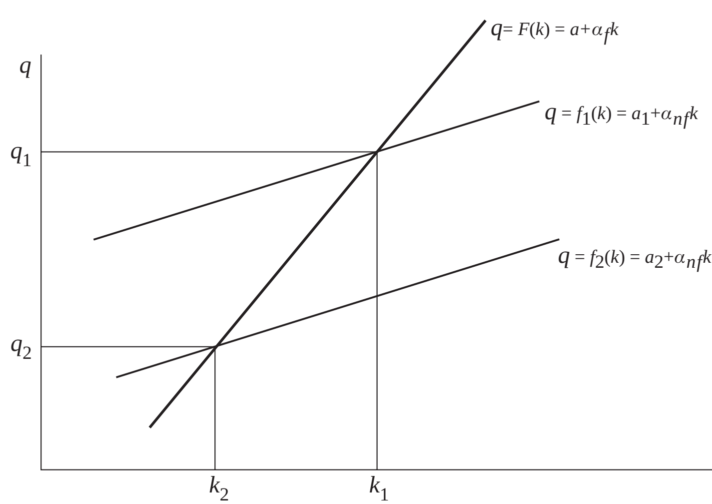

Figura 1

ilustra el problema. Supongamos que tenemos dos regiones con distintas ratios capital/trabajo (k_1 y k_2, medidas en logaritmos) y diferentes niveles de *output* por trabajador (q_1 y q_2, también en logaritmos), con $q_1 > q_2$ y $k_1 > k_2$. Supongamos también que la tecnología es Cobb-Douglas, con lo que la relación entre q y k es de la forma $q = a + \alpha k$. Claramente, la hipótesis de que ambas estimaciones se encuentran en una frontera común (llamada $f(\cdot)$ en la figura 1) generará un valor estimado de α que puede ser muy superior al que obtendríamos bajo el supuesto contrario (o, en el presente contexto no paramétrico, puede hacer que la frontera sea excesivamente «empinada»). Sospecho que este sesgo podría explicar por qué la contribución de la acumulación de capital a la convergencia es mayor en el presente trabajo que en otros que utilizan técnicas más convencionales —aunque la ausencia de parámetros hace difícil comprobar si éste es el caso.

Como indican Maudos y otros, si el mundo es de una cierta forma y suponemos lo contrario, estaremos introduciendo sesgos potencialmente importantes en la estimación de los parámetros de la función de producción, o en los de la frontera eficiente. A los autores, sin embargo, sólo parece preocuparles el posible sesgo que se derivaría de ignorar diferencias interregionales en niveles de eficiencia X. Pero al intentar remediar este problema podrían introducir uno peor. Parece, por lo tanto, sensato comenzar preguntándonos qué técnica es mejor (esto es, genera el sesgo más pequeño). En mi opinión, la respuesta a esta pregunta depende fundamentalmente de la muestra. Si trabajamos con datos a nivel de establecimientos dentro de un sector relativamente homogéneo y con una tecnología bien establecida, la estimación de una frontera de producción eficiente podría ser lo más indicado. Pero dudo mucho que éste sea el caso con una muestra de agregados regionales, por no hablar ya de los estudios (que los hay) con agregados nacionales.

De acuerdo con los resultados de nuestros autores, un trabajador extremeño sería capaz de producir tan solo un 60% del *output* que obtendría un madrileño con la misma dotación de capital. Maudos y otros, se inclinan por la explicación de que esta menor productividad no se debe a diferencias tecnológicas entre las dos regiones sino a la menor eficiencia de los trabajadores y empresarios extremeños —esto es, a que éstos son más torpes, menos trabajadores o peor organizados que los madrileños—. Francamente, ésta no parece una explicación atractiva de las diferencias de productividad entre dos regiones. De hecho, los propios autores se desmarcan implícitamente de esta interpretación cuando, en una de las secciones finales del trabajo, intentan explicar las diferencias interregionales en niveles de eficiencia en términos de dotaciones de capital humano e infraestructuras entre otros factores. Con ello, sin embargo, incurren en una cierta inconsistencia, ya que no parece razonable interpretar como «ineficiencia» los bajos niveles de productividad derivados de deficiencias en las dotaciones de ciertos recursos productivos. Sería más adecuado incluir estos factores como argumentos de la frontera. Dada la naturaleza agregada de los datos, el hecho de que las empresas no controlen algunas de estas variables no parece un buen argumento para justificar un procedimiento en dos etapas.

2. Las «ventajas» de una frontera no estocástica

Puestos a estimar una frontera de producción eficiente, ¿por qué hacerlo empleando una técnica determinista y no paramétrica? Maudos y otros, señalan tres posibles «ventajas» de este enfoque. La primera es que una técnica no paramétrica ofrece una mayor flexibilidad porque no supone una forma funcional específica, la segunda es que este procedimiento permite obtener estimaciones individualizadas de las tasas de progreso técnico de las diferentes regiones, y la tercera que de esta forma se evita la necesidad de hacer supuestos sobre la distribución de las perturbaciones. Las dos primeras afirmaciones son ciertas, aunque ambas ventajas tienen su precio, pero la tercera claramente no lo es. El supuesto de que la perturbación es siempre cero es ciertamente una hipótesis distribucional. De hecho, se trata de un supuesto extremadamente fuerte y que hace que los resultados puedan ser muy sensibles a observaciones atípicas o errores de medición.

En cuanto a las dos primeras ventajas, coincido con Maudos y otros, en que la flexibilidad del enfoque no paramétrico puede resultar atractiva cuando no conocemos la forma funcional de la relación a estimar, pero dudo que en el caso que nos ocupa esta consideración sea crucial. Por otro lado, la ausencia de parámetros hace difícil juzgar la plausibilidad de la función estimada. (Una posible solución —a la que los autores se resisten y yo volveré más adelante— es dibujar la función y ver qué pinta tiene). Por lo que se refiere a la posibilidad de estimar distintas tasas de progreso técnico para cada región, esto es posible porque la frontera de producción se estima de forma independiente para cada año y, al no imponer restricciones sobre su forma funcional, el desplazamiento vertical de la frontera entre un año y otro no tiene por qué ser uniforme en todo su dominio. El mismo resultado se conseguiría estimando por métodos más habituales una frontera distinta para cada año de la muestra con datos de corte transversal. Sin embargo, pienso que esto no sería buena idea puesto que se ignoraría la dimensión temporal de los datos, esto es, la principal ventaja de los datos de panel que otros estudios sobre el tema han intentado explotar.

3. ¿Una frontera razonable?

Para hacernos una idea más clara de las posibles virtudes y defectos del enfoque no paramétrico conviene intentar desarrollar una cierta intuición sobre cómo funciona este método de estimación. Con este fin, consideremos el caso más sencillo posible, el de una frontera de la forma $q = f(k)$, donde q es *output* y k el único *input*, al que llamaremos capital. Siguiendo la metodología empleada por Maudos y otros, habría que resolver dos problemas de programación lineal. El primero de ellos, en el que se impone la existencia de rendimientos constantes a escala, sería de la forma

$$(PC \cdot k_i) \;\; f^c(k_i) = \underset{\lambda_n;\; n=1\ldots N}{Max} \;\; q_i^f$$

$$\text{s.a} \quad q_i^f \leq \sum_{n=1}^{N} \lambda_n q_n \, , \; \sum_{n=1}^{N} \lambda_n k_n \leq k_i \, , \lambda_n \geq 0 \;\; \text{para todo } n$$

donde q_n es el producto por trabajador observado en la región n, k_n su *stock* de capital y $f^c(k_i)$ = Max q_i^f el producto eficiente en la región i. El grado de ineficiencia total vendría dado después por $\theta_i^T = f^c(k_i)/q_i$.

Seguidamente, se resuelve un problema similar al anterior tras añadir la restricción

$$\sum_{n=1}^{N} \lambda_n = 1$$

con el fin de imponer la existencia de rendimientos no-crecientes a escala. Tenemos ahora

$$(PD \cdot k_i) \;\; f^d(k_i) = \underset{\lambda_n;\; n=1\ldots N}{Max} \;\; q_i^f$$

$$\text{s.a} \quad q_i^f \leq \sum_{n=1}^{N} \lambda_n q_n \,,\; \sum_{n=1}^{N} \lambda_n k_n \leq k_i \,, \lambda_n \geq 0 \quad \text{para todo } n, \text{ y } \sum_{n=1}^{N} \lambda_n = 1$$

y el grado de ineficiencia técnica pura se calcula como $\theta_i^P = f^d(k_i)/q_i$. Finalmente, el ratio entre θ_i^P y θ_i^T se interpreta como un indicador del nivel de ineficiencias de escala.

En el caso de un único *input*, la solución del primer problema es trivial. Puesto que el supuesto de rendimientos constantes a escala en un único factor implica que la relación entre *input* y *output* ha de ser lineal, basta con identificar la región m en la que el ratio q/k es más elevado. La frontera eficiente (f^c) corresponderá después al rayo que, partiendo del origen, pasa por la observación correspondiente a esta región. Esto es, tendremos $f^c(k) = Mk$, con $M = q_m/k_m$.

En cuanto al segundo problema, $(PD \cdot k_i)$ exige que el par (q_i^f, k_i) sea una combinación convexa de los valores observados de (q_n, k_n) para las distintas regiones de la muestra siempre que las dos primeras restricciones del problema sean activas. La primera de ellas siempre ha de saturarse en un óptimo, pero la segunda podría no hacerlo. En este caso, existirá una región m que produce un *output* superior a q_i con menos recursos, y tendremos $\lambda_m = 1$, $q_i^f = q_m$ y $k_m < k_i$. Para resolver el problema, por tanto, podemos proceder en dos etapas. En primer lugar, se trataría de construir el *convex hull* del conjunto de pares observados de *outputs* y *stocks* de capital y luego elegir la frontera superior de este conjunto. La función resultante será lineal a trozos puesto que será el resultado de unir las observaciones (q_n, k_n) no «dominadas verticalmente» por otras observaciones o por combinaciones convexas de ellas, pero podría ser decreciente a partir de un punto dado. En este caso, la segunda restricción del problema original no se satura para algunas regiones y, para completar la frontera eficiente, debemos reemplazar el segmento decreciente de la función obtenida en la primera etapa por una recta horizontal que parte de la última observación eficiente hacia la derecha (esto es, q_i^f sería el máximo *output* factible con $k < k_i$).

Resulta ilustrativo aplicar el método precedente al caso de las regiones españolas trabajando con la función de producción por trabajador. Así pues, interpre-

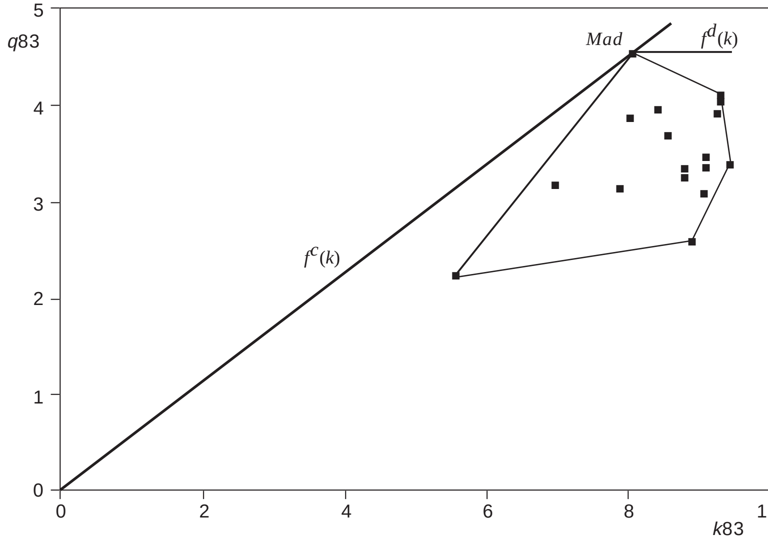

Figura 2. Frontera de producción intensiva para las regiones españolas, 1983.

taremos q_n como el producto por trabajador ocupado en la región n y k_n como el ratio capital/trabajo. La figura 2 muestra las dos fronteras (f^c y f^d) obtenidas con los datos de las regiones españolas correspondientes a 1983. Bajo el supuesto de rendimientos constantes, la frontera f^c es un rayo a través del origen que pasa por Madrid. Con rendimientos decrecientes, la frontera f^d está formada por dos segmentos: el que une las observaciones correspondientes a Galicia y Madrid, y otro horizontal que parte de Madrid hacia la derecha.

El ejercicio que acabamos de realizar no es el que proponen los autores, porque ellos no estiman la frontera de producción en términos intensivos, pero debería de ser equivalente (al menos conceptualmente). Si, como hacen ellos[1], partimos de una frontera de producción de la forma $Q = F(K, L)$ con rendimientos constantes a escala, podemos definir la función de producción intensiva $q = f(k) = F(k, 1)$, con $q = Q/L$ y $k = K/L$. La función $f()$, que presumiblemente presentará rendimientos decrecientes en k, debería coincidir con la frontera $f^d(k)$ que acabamos de estimar. Esta última función, que es lineal en buena parte de su dominio y constante en el resto, no parece una descripción razonable de la tecnología productiva.

1. Véase el Apéndice A.1 El problema de programación lineal utilizado para calcular la eficiencia total supone rendimientos constantes a escala. El segundo problema, donde este supuesto no se impone, parece utilizarse únicamente para descomponer el nivel de ineficiencia total en un componente de eficiencia técnica pura y otro que captura deseconomías de escala.

Por alguna razón que no acierto a comprender, la frontera estimada por Maudos y otros, es aún menos razonable que $f^d(k)$ —y parece de hecho provenir de $f^c(k)$ en vez de $f^d(k)$. De acuerdo con su tabla 2, en concreto, la única región «totalmente eficiente» en la mayor parte de los años considerados es Madrid. Puesto que la frontera se construye a partir de combinaciones lineales de observaciones eficientes, esto plantea además un problema obvio: ¿cómo podemos construir la frontera completa a partir de un único punto?