

XLVIII Coloquio Argentino de Estadística

VI JORNADA DE EDUCACIÓN ESTADÍSTICA "MARTHA DE ALIAGA"

27 al 30 oct 2020

Poster:

Estimación de eficiencia en educación media: una comparación del análisis paramétrico y no paramétrico

Luca M. Salvatierra Boscatto



Esta obra está bajo una
Licencia Creative Commons
Atribución-NoComercial 4.0
Internacional



FACULTAD
DE CIENCIAS
ECONÓMICAS



Universidad
Nacional
de Córdoba



ESTIMACIÓN DE EFICIENCIA EN EDUCACIÓN MEDIA UNA COMPARACIÓN DEL ANÁLISIS PARAMÉTRICO Y NO PARAMÉTRICO.

Luca M. Salvatierra Boscatto
Facultad de Ciencias Económicas – Universidad

XLVIII Coloquio Argentino de Estadística

VI Jornada de Educación Estadística "Martha de Aliaga"



RESUMEN

En la actualidad, los análisis de fronteras paramétricas y no paramétricas son las principales técnicas utilizadas en el estudio de la eficiencia de las unidades productivas. Generalmente, las aplicaciones empíricas se basan en una u otra. Este trabajo, sin embargo, emplea ambas metodologías en diferentes formulaciones, a los fines de obtener un coeficiente de eficiencia comparable entre los resultados obtenidos en cada caso. En específico, se aplica el análisis de frontera estocástica (SFA por sus siglas en inglés) en distintas formulaciones y el análisis de la envolvente de datos (DEA por sus siglas en inglés) comparando los resultados obtenidos con rendimientos constantes y variables a escala. Particularmente, se realiza la estimación de los niveles de eficiencia técnica de la educación media en Argentina en el período 2007-2016, tomando como unidades de decisión (DMUs por sus siglas en inglés) a las 23 provincias de nuestro país y la Ciudad Autónoma de Buenos Aires, empleando y comparando el análisis paramétrico y no paramétrico. Los insumos (*inputs*) definidos en este trabajo son: cantidad de unidades educativas, cantidad de alumnos matriculados y cantidad de docentes con funciones frente al aula. Por su parte, como producto (*output*) se emplea la tasa de egreso, calculada por el Ministerio de Educación de la Nación y definida como la cantidad de matriculados en el 1er año que logran culminar todos los años de estudio del nivel/ciclo, trabajando en todos los casos con datos correspondientes a la gestión estatal. Se emplean datos de panel, configurándose 240 observaciones (24 unidades a lo largo de 10 años). Los resultados del trabajo indican que existe una consistencia entre los niveles de eficiencia obtenidos con cada técnica, pero no así comparando ambos enfoques, siendo importante la selección del modelo apropiado según el caso de estudio.

OBJETIVOS

- Estimar los niveles de eficiencia de cada DMU a lo largo del período.
- Comparar los resultados obtenidos con cada formulación, en cada técnica.
- Comparar los resultados entre técnicas.

MARCO METODOLÓGICO

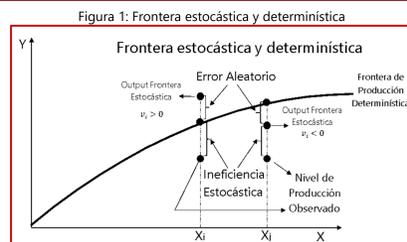
A diferencia de las primeras representaciones de fronteras de eficiencia paramétricas, el enfoque de Análisis de Frontera Estocástica (SFA por sus siglas en inglés) ha representando un gran avance respecto a las fronteras determinísticas. En tal sentido, posibilita incorporar un componente que capta los efectos exógenos que son propios de cada unidad de observación, pudiendo descomponer los desvíos de la frontera en el impacto de la ineficiencia técnica por un lado y de los efectos externos por otro.

Encontramos en Aigner (et al. 1977) la primera representación:

$$q_i = f(x_i, \beta) + \eta_i, \text{ donde } \eta_i = v_i - u_i$$

Siendo q_i la producción obtenida por cada unidad de decisión, x_i los insumos y β un vector de parámetros desconocidos que se deben estimar, mientras que η_i es la parte estocástica a estimar, el cual está compuesto por v_i (componente estocástico) y u_i (ineficiencia técnica), siendo independientes.

MARCO METODOLÓGICO



Fuente: Elaboración propia basado en Coelli et al 2005, p. 244.

El parámetro de eficiencia de este modelo se define como:

$$\gamma = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}$$

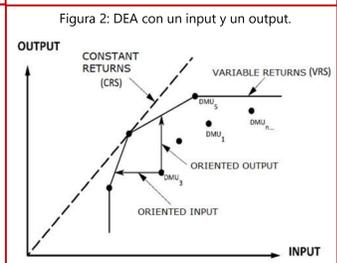
Cuando $\sigma_u^2 \rightarrow 0, \gamma = 0$, significa eficiencia o ausencia de ineficiencia técnica. Por otro lado, si $\sigma_v^2 \rightarrow \infty$ la ineficiencia podría darse por ruido en la producción. En la actualidad uno de los modelos más usados para medir la eficiencia de las unidades de producción con la técnica SFA es el propuesto por Battese y Coelli (1995), que es un modelo de distribución normal truncada que realiza la estimación mediante máxima verosimilitud. Sin embargo, los *scores* de eficiencia pueden sufrir la incidencia de efectos invariantes en el tiempo. Para corregirlo surgen los modelos de "efectos fijos/aleatorios verdaderos" (*true fixed- and random- effects*), los que permiten separar la heterogeneidad invariante en el tiempo de la ineficiencia que varía a lo largo del periodo analizado. Greene (2005) formula dos alternativas, una con efectos fijos y otra con efectos aleatorios, y en ambos casos adopta tres posibles distribuciones.

Por su parte, las técnicas no paramétricas como el Análisis de la Envolvente de Datos (DEA por sus siglas en inglés), introducido por Charnes, Cooper y Rhodes (1978) y basado en Farrell (1957), con rendimientos constantes a escala (CRS), presentan la ventaja de no tener que suponer una forma funcional que relacione insumos y productos, aunque asume homogeneidad entre las unidades de decisión (DMU) y muestra sensibilidad ante valores extremos. En este trabajo se realizan estimaciones empleando la formulación original y la de Banker, Charnes y Cooper (1984) que supone rendimientos variables a escala (VRS), en ambos casos orientados hacia el producto, identificando los desvíos en dos etapas. Podemos escribir el modelo, basado en Cano et. al. (2017), de la siguiente manera:

$$\begin{aligned} & \text{Max } \theta_j \\ & \text{Sujeto a} \\ & \sum_{j=1}^N \lambda_j y_{(r,j)} \geq \theta_j y_{(r,o)}, \text{ para } r = 1, \dots, S \\ & \sum_{j=1}^N \lambda_j x_{(i,j)} \leq x_{(i,o)}, \text{ para } i = 1, \dots, M \\ & \lambda_j \geq 0, \text{ para } j = 1, \dots, N \\ & \sum_{j=1}^N \lambda_j = 1 \rightarrow \text{Para el caso de VRS} \end{aligned}$$

Donde:
• θ_j índice de eficiencia, λ_j ponderador, $y_{(r,j)}$ cantidad de producto r y $x_{(i,j)}$ cantidad de insumo i usado de la DMU j
• r = subíndice de producto
• i = subíndice de insumo
• j = subíndice de la unidad de decisión.

Observamos que se trata de un método basado en la programación lineal. La última restricción (el sumatorio de los λ_j debe ser igual a uno) se relaciona con los rendimientos variables a escala, es decir que al aumentar la cantidad de insumos el producto puede aumentar en menor, igual o mayor proporción. En ausencia de esta restricción, el modelo es de rendimientos constantes a escala.



Fuente: CORDOVA, F. y ALBERTO, C. (2018). Modelo CRS y VRS caso de un input y un output. [Figura].

RESULTADOS

Se realizan las estimaciones de los coeficientes de eficiencia a través de siete formulaciones en el caso de SFA y dos en el caso de DEA, siendo los insumos la cantidad de estudiantes matriculados, cantidad de unidades educativas y cantidad de docentes con funciones frente al aula, mientras el output es la tasa de egreso de educación media, trabajando con un modelo log-log, para poder aplicar SFA y reducir la sensibilidad de las estimaciones ante observaciones extremas o atípicas.

Para el análisis de frontera estocástica empleamos los modelos de Battese y Coelli (1995) –BC95- y Greene (2005). En el caso de este último trabajamos con sus dos formulaciones (efectos fijos y efectos aleatorios "verdaderos") y las tres distribuciones sugeridas: exponencial, semi normal y normal truncada (TFE_E, TFE_HN, TFE_TN, TRE_E, TRE_HN y TRE_TN, respectivamente).

Los resultados obtenidos parecen a priori no diferir entre las diferentes estimaciones, sin embargo realizamos una matriz de coeficientes de correlación de Spearman, donde las celdas oscuras muestran una correlación menor a 0,7, siendo todas las estimaciones significativas a un nivel del 0,1%.

Tabla 1: Matriz de coeficientes de correlación de Spearman

	TFE_E	TFE_HN	TFE_TN	TRE_E	TRE_HN	TRE_TN	BC95	CRS	VRS
TFE_E	1								
TFE_HN	0.8793*	1							
TFE_TN	0.9028*	0.8668*	1						
TRE_E	0.9826*	0.8884*	0.9181*	1					
TRE_HN	0.9783*	0.8870*	0.9237*	0.9954*	1				
TRE_TN	0.9795*	0.8760*	0.9202*	0.9929*	0.9955*	1			
BC95	0.7329*	0.6641*	0.7558*	0.7925*	0.8169*	0.8211*	1		
CRS TE	0.2761*	0.2231*	0.3458*	0.2932*	0.3318*	0.3208*	0.3753*	1	
VRS TE	0.5445*	0.4937*	0.5874*	0.5841*	0.6167*	0.6118*	0.7601*	0.7796*	1

Se puede verificar que, excepto el caso de la relación entre SFA BC95 y DEA VRS, cuya correlación es 0.7601, el resto de los casos muestra un bajo nivel de correlación entre métodos, destacándose el hecho de que DEA CRS no tiene un coeficiente superior al 0.4 con ninguna de las modelizaciones paramétricas. Sin embargo, se puede ver que existe un alto nivel de correlación entre los modelos SFA, a excepción de BC95 con TFE_HN. En el resto de los casos, se observan coeficientes superiores a 0.7.

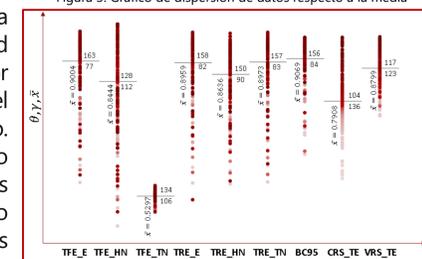
Finalmente, para el caso de las dos formulaciones de DEA encontramos que tienen correlación de 0.7796 a un nivel de significancia estadística alto (0.001). Se observa asimismo que el máximo de eficiencia obtenido por TFE_TN es de 0.5695 mientras que en el resto de los casos es superior a 0.97. Por otro lado, el resultado que presenta mayor dispersión es el de TFE_HN teniendo una desviación estándar de 0.122.

Tabla 2: Estadísticas descriptivas de los índices de eficiencia estimados

Variable	Obs	Media	DE	Min	Max
TFE_E	240	.9004	.0784	.5649	.9813
TFE_HN	240	.8444	.1220	.4459	.9999
TFE_TN	240	.5297	.0128	.4833	.5695
TRE_E	240	.8959	.0781	.5177	.9781
TRE_HN	240	.8636	.0801	.5606	.9766
TRE_TN	240	.8973	.0760	.5663	.9791
BC95	240	.9069	.0540	.6430	.9719
CRS TE	240	.7908	.0830	.5765	1
VRS TE	240	.8799	.0483	.7490	1

En esta figura se presenta la media, dispersión y cantidad de observaciones por encima y por debajo del promedio en cada caso. Vemos un comportamiento similar entre técnicas internamente, pero opuesto si comparamos ambas metodologías, al tener más unidades por encima del promedio de eficiencia en todos los casos de SFA y más por debajo en los dos casos de DEA.

Figura 3: Gráfico de dispersión de datos respecto a la media



Fuente: Elaboración propia.

CONCLUSIONES

La elección del enfoque o técnica a utilizar al momento de evaluar la eficiencia técnica de un sistema productivo no debe realizarse sin previamente analizar las características propias del mismo, la disponibilidad de datos y los objetivos de la investigación. Las estimaciones de los niveles de eficiencia con cada una de las técnicas ha demostrado tener correlación interna pero no así entre ambas. La elección específica de la modelización, una vez elegido el enfoque, dependerá de las condiciones propias del set de datos y las características que se quieran estudiar.

Sin embargo, cabe destacar que existe una divergencia en cuanto a los niveles de eficiencia medios y su relación con cada estimación. Las modelizaciones paramétricas han ubicado en todos los casos a más de la mitad de observaciones sobre la media de eficiencia obtenida en cada caso, teniendo sólo en el caso de TFE_HN una alta dispersión de datos. Por otra parte, las estimaciones por DEA en ambos casos arrojaron el resultado opuesto ubicando a más de la mitad de las observaciones por debajo de la media de eficiencia estimada.

Finalmente, cabe mencionar que para ambas metodologías encontramos resultados estadísticamente significativos por lo que la elección de uno u otro enfoque dependerá del contexto de la investigación. Luego, la metodología y modelización apropiadas surgirán de realizar pruebas de especificación a los fines de determinar la que mejor se ajuste al caso de estudio.

BIBLIOGRAFÍA

- Aigner et al. (1977). 'Formulation and estimation of stochastic frontier production model', Journal of Econometrics 6: 21-37.
- Banker, R. D., Charnes, A., y Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092.
- Battese, G. E., and T. J. Coelli. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20: 325-332.
- Charnes, A., Cooper, W.W., and Rhodes, E., (1978). Measuring the Efficiency of Decision Making Units, European Journal of Operational Research, 2, 429-444.
- Coelli T., et al. (2005). An Introduction to Efficiency and Productivity Analysis. Segunda Edición 2005. Editorial Springer.
- Córdova, F., & Alberto, C.. (2018). Measurement of efficiency in the construction industry and its relationship with working capital. Revista ingeniería de construcción, 33(1), 69-82.
- Farrell, M.J. (1957). The measurement of productive efficiency, Journal of Royal Statistical Society A 120, 253-281.
- Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics 126: 269-303
- Greene, W. (2005). Fixed and random effects in stochastic frontier models. Journal of Productivity Analysis 23: 7-32.

CONTACTO

Luca.salvatierra1@gmail.com