

EVALUACIÓN DEL DESEMPEÑO DE LOS MÉTODOS DE REDUCCIÓN DE DIMENSIÓN APLICADO A
MODELOS DE FRONTERA ESTOCÁSTICA DE PRODUCCIÓN, EN VISIBILIDAD CIENTÍFICA

SOL MERY ÁLVAREZ ARGAEZ

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
MEDELLÍN
2013

EVALUACIÓN DEL DESEMPEÑO DE LOS MÉTODOS DE REDUCCIÓN DE DIMENSIÓN APLICADO A
MODELOS DE FRONTERA ESTOCÁSTICA DE PRODUCCIÓN, EN VISIBILIDAD CIENTÍFICA

SOL MERY ÁLVAREZ ARGAEZ

Trabajo de investigación para optar al título de “Magíster en Ingeniería”

Director: Ms Juan Delgado Lastra

UNIVERSIDAD DE ANTIOQUIA
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
MEDELLÍN
Junio, 2013

Agradecimientos

Al profesor Juan Delgado, por su paciencia, inspiración, disciplina, inteligencia, por la pasión con la que realiza su trabajo y que contagia, por ser un profesor en todo el sentido de la palabra, por darme una y mil oportunidades, por confiar en mí, por esto y por mucho más infinitas gracias.

Al profesor Bernardo Calderón por ser un gran maestro, por hacer los arreglos necesarios para dictarme clase en la maestría.

Al profesor Elkín Castaño por su asesoría y valiosos aportes a este trabajo.

A Natalia Gómez, por incentivarme y motivarme a que iniciáramos este viaje que se convierte en un eslabón más en mi formación.

A mis padres por todo el amor que me han brindado, por sus consejos, por enseñarme a trazarme metas importantes y esforzarme por lograrlas y por su apoyo incondicional en todos los aspectos de mi vida.

A Jonathan Ortiz, por su amor, apoyo y paciencia en todo este proceso.

Tabla de contenido

INTRODUCCIÓN.....	6
Capítulo 1	7
Capital intelectual	7
Capital interno	8
Capital externo.....	8
Capital humano.....	8
Visibilidad Científica.....	8
Capítulo 2	11
Eficiencia	11
Frontera estocástica.....	11
Análisis multivariado.....	13
Componentes principales	13
Análisis factorial	15
Diferencias entre análisis factorial y componentes principales	19
Capítulo 3	20
Simulaciones	20
Resultados.....	23
Caso 1: Distribución del término de ineficiencia Exponencial ($\lambda=0.5$).....	23
Caso 2: Distribución del término de ineficiencia Half-Normal $\mu = 3, \sigma^2 = 6$	25
Resultados con unidades eficientes técnicamente.....	27
Caso 3: Simulaciones con unidades eficientes técnicamente y distribución Exponencial en el término de ineficiencia U_i	27
Caso 4: Simulaciones con unidades eficientes técnicamente y distribución Half-Normal en el término de ineficiencia U_i	29
Caso de estudio.....	31
Reducción de dimensión.....	31
Frontera estocástica.....	31
CONCLUSIONES.....	34
BIBLIOGRAFÍA.....	35

Listado de tablas

Tabla 1. Estimación de la frontera estocástica de producción. Caso 1: Con distribución exponencial en el término de ineficiencia U_i	23
Tabla 2. Estimación de la frontera estocástica de producción. Caso 2: Con distribución half-normal en el término de ineficiencia U_i	25
Tabla 3. Estimación de la frontera estocástica de producción. Caso 3: Con distribución Exponencial y 1% de las unidades con cero en el término de ineficiencia U_i	27
Tabla 4. Estimación de la frontera estocástica de producción. Caso 4: Con distribución half-normal y 1% de las unidades con cero en el término de ineficiencia U_i	29
Tabla 5. Proporción de información explicada por el factor y componente principal	31
Tabla 6. Correlación entre las nuevas variables capacidad y resultados con ambos métodos de reducción de dimensión	32
Tabla 7. Estimación de la frontera estocástica de producción para el caso de estudio, con distribución en el término de ineficiencia U_i Half-normal y Exponencial.	32

INTRODUCCIÓN

Medir el desempeño de las instituciones se hace imperante en un contexto donde su vigencia depende del uso adecuado de las capacidades para maximizar los resultados, es decir, la medición de su eficiencia, este concepto puramente económico no sólo es aplicado a instituciones de esta índole, también se aplica a entidades cuyo objetivo principal es la generación de conocimiento tales como universidades y/o institutos de investigación, los cuales no sólo deben generar valor sino visibilizarlo en la comunidad científica a fin de garantizar su validez, para esta clase particular de instituciones existen teorías como la del capital intelectual que explica la generación de valor de una institución a través de tres dimensiones: capital humano, estructural e interno sus relaciones y la intersección de ellas, al definir los indicadores para estas mediciones se generan bases de datos de alta dimensión y con estructuras de correlación y dependencia por lo que se hace necesario el uso de métodos multivariados que permitan transformar un conjunto de datos grandes y complejos, en modelos de menor complejidad y de fácil interpretación, tales como análisis factorial y componentes principales, estos métodos se caracterizan por ser un punto intermedio entre la complejidad de los datos y el uso de modelos de orden superior como regresión lineal y frontera estocástica.

En el presente trabajo se evaluará el desempeño de los métodos de reducción de dimensión aplicados a frontera estocástica en un periodo de tiempo como medición del desempeño de las unidades experimentales inicialmente mediante simulaciones Monte Carlo, para luego extenderlo al caso particular de las universidades que componen el Sistema Universitario Estatal Colombiano –SUE el cual viene desarrollando una dinámica de medición de desempeño en el sistema desde el año 2003, mediante la definición de la batería de indicadores, en los tres ejes misionales de la universidad, investigación, docencia y extensión así como en el componente de bienestar bajo la dicotomía capacidad-resultados

El capítulo uno de este trabajo plantea el problema de la medición de desempeño para instituciones que generan conocimiento a partir de la teoría del capital intelectual en el segundo capítulo se presenta la teoría usada para la evaluación del desempeño de los métodos de reducción de dimensión aplicados a frontera estocástica y en el último capítulo se ilustran los resultados de los experimentos Monte Carlo y la aplicación al caso de estudio.

Capítulo 1

Las organizaciones buscan constantemente alcanzar la misión y visión propuesta; el entorno exige cada vez más de ellas por lo que están siendo constantemente evaluadas de manera directa o indirecta por los clientes, proveedores y por entidades de características similares o superiores, es por esta razón, entre otras, que cobra sentido medir su desempeño, el cual se define como: el esfuerzo sistemático que realiza una organización para evaluar su gestión orientada al cumplimiento de su misión, a partir de la optimización de sus procesos. Arriagada R. (2002). Derivado de la visión y misión, las organizaciones realizan un plan estratégico que permite definir las metas que deberán seguirse para el logro de aquellas; este plan puede denominarse como la ruta a seguir y es precisamente allí donde se definen los indicadores que facilitan el seguimiento al cumplimiento de los objetivos trazados, además de sistematizar la información que luego se convertirá en insumo esencial en la medición. Con ellos se consolida una base de datos de alta dimensión de la forma: p variables y n unidades experimentales y, con presencia de estructuras de correlación y dependencia multivariadas.

Sin embargo y pese a lo determinante de la medición de desempeño en la toma de decisiones estratégicas acertadas, éste constructo no siempre es de fácil medición debido a:

1. las particularidades propias de las entidades, sea por el lugar donde se ubica, el sector económico a que pertenece o la actividad que desarrolla.
2. no se cuenta con las herramientas o conocimientos apropiados para ello.
3. su carácter multivariado, entre otras.

Este escenario se torna un tanto diferente cuando las entidades que desean medir su desempeño no tienen como fin principal la generación de riqueza; es el caso de las instituciones donde se genera conocimiento tales como las Universidades, allí la complejidad de las mismas dificulta aún más la medición. Sin embargo, teorías como la del capital intelectual permiten definir indicadores en dimensiones que representan de manera muy acertada la complejidad de los sistemas universitarios, aunque no sea el único campo de aplicación.

Capital intelectual

El capital intelectual es el conocimiento de una institución, no es visible pero permite el desarrollo de capacidades propias y características de cada organización; en general contiene las siguientes clasificaciones: capital interno, capital externo y capital humano, estos tres ejes del capital intelectual y sus relaciones permiten formular la estrategia de la organización, mediante la utilización de nueve dimensiones que formulan la creación de valor. para la organización Sveiby K (2006) Figura 1.

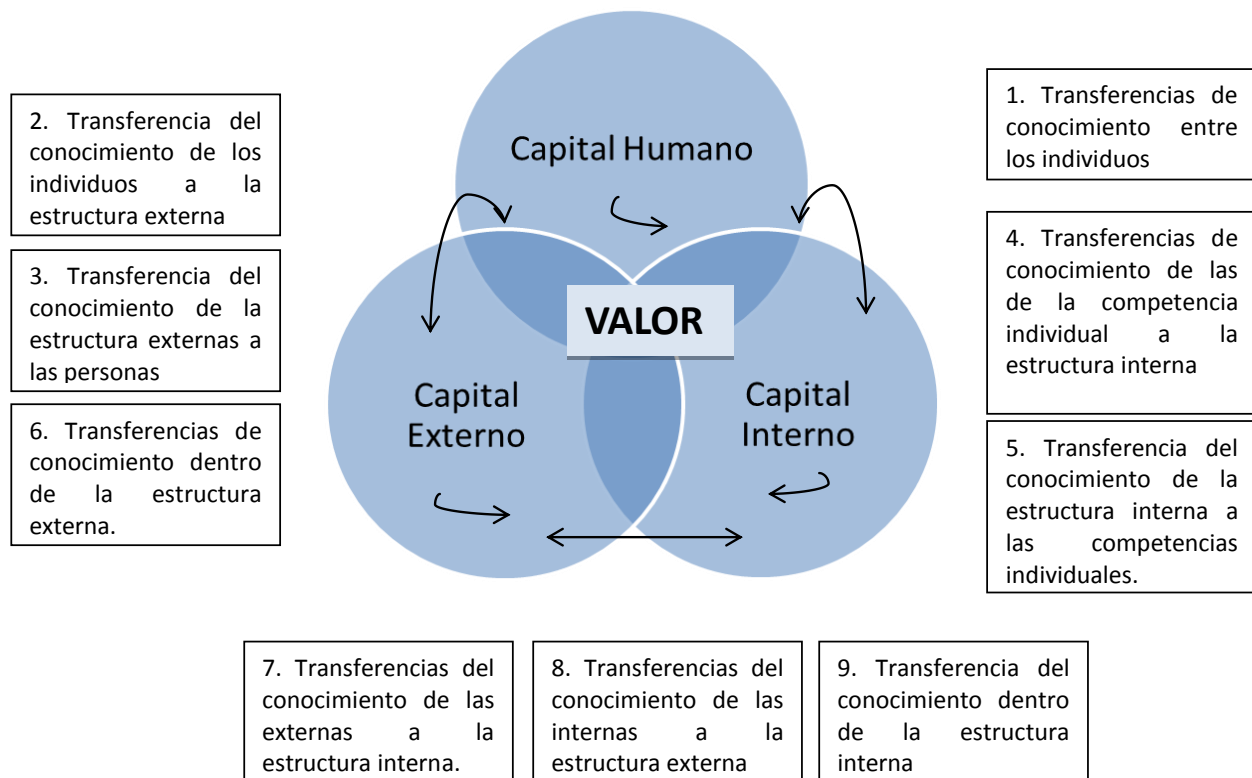


Figura 1. Esquema de capital intelectual, construido a partir de Sveiby K (2006)

Capital interno

Son todas aquellas herramientas físicas, tecnológicas con las que cuenta la empresa y que han sido desarrolladas por los empleados pero que en general son propiedad de la organización; dicho de otra manera, es todo aquello que queda en la empresa una vez los empleados salen de ella.

Capital externo

Son las relaciones intangibles que desarrolla la empresa con sus clientes y/o proveedores, las cuales se rigen bajo el concepto de calidad, confianza, cumplimiento y todos los valores que desarrolla la empresa que constituyen su reputación o imagen, que puede ser positiva o negativa y que recae en el subjetivismo del individuo con el que se relaciona.

Capital humano

Incluye a todas las personas que trabajan en la organización sin importar el lugar que ocupe, por lo que va desde el personal de apoyo hasta el personal de investigación y desarrollo, si bien cada uno dentro de la organización tiene una manera diferente de relacionarse con el exterior.

Aunque la teoría del capital intelectual define claramente las dimensiones existentes que permiten la generación de valor, conocimiento para el caso de las universidades, el papel de éstas no se queda sólo allí, una vez lo alcanza deberá hacerlo visible.

Visibilidad Científica

La visibilidad científica puede definirse como las acciones que permiten que una institución de conocimiento sea perceptible en su comunidad de pares; esta visibilidad se logra a través de la

publicación de resultados de investigación, los cuales son manifiestos en su gran mayoría mediante artículos. Este concepto de tanto auge en la última década no sólo cobra sentido para las universidades de rango mundial, también lo es para universidades que se están esforzando por alcanzarlo. La dinámica de estas instituciones y su impacto en el crecimiento económico y la competitividad global han volcado las miradas del mundo hacia las mismas, ubicándolas como una pieza clave en ese contexto; prueba de ello es el acelerado crecimiento del número de rankings Salmi (2009), los cuales incluyen en su metodología desde indicadores muy selectos como número de premios Nobel de la institución hasta el impacto de su sitio web y los documentos alojados allí. La disparidad en conceptos y múltiples formas de medición permiten ubicar la misma universidad en puestos muy diferentes tanto así que la idea de un ranking de rankings se convierte en una constante.

La Universidad colombiana no es ajena a estas tendencias, incluso a fin de actualizar, mejorar y medir el desempeño de sus procesos de formación, investigación y extensión, se creó el Sistema Universitario Estatal - SUE- el cual contiene una batería de 43 indicadores de capacidad y resultados para medir el desempeño de las 32¹ universidades estatales; aunque los indicadores definidos no dan cuenta de la visibilidad internacional de las mismas, algunos indicadores del eje misional investigación apuntan aunque de manera tímida a este propósito.

Como se mencionó inicialmente, en la consolidación de esta información se presentan nuevos retos en la medición de desempeño, ya que los datos recolectados crean bases de datos de alta dimensión en tres direcciones: p variables, n unidades y T periodos, como se muestra en la Figura 2.

¹ No se tiene en cuenta la Universidad Naval por ser una universidad con características muy especiales que difieren notablemente del conjunto.

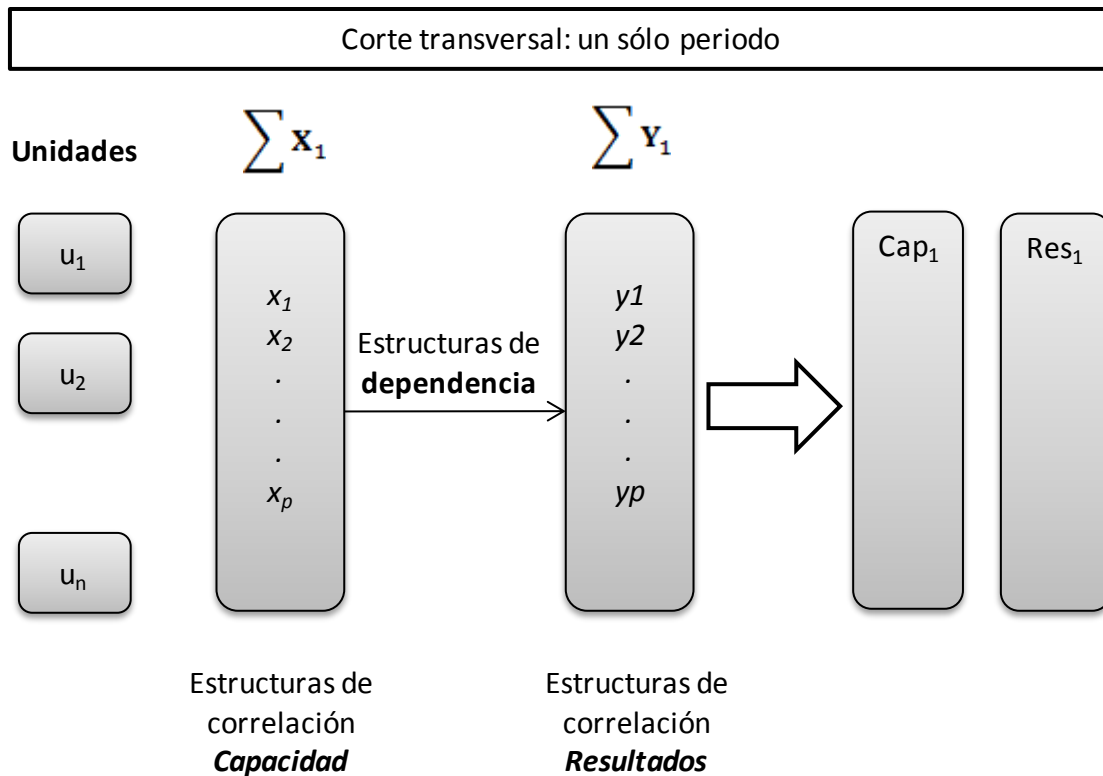


Figura 2. Medición de desempeño estructura del problema

Esta situación hace muy complejo el análisis de la información en la caracterización de todo el sistema y de cada una de las universidades que lo conforman, para llevar a cabo la correspondiente planeación estratégica que las conduzca a la excelencia académica de visibilidad internacional, en este sentido se hace necesario el uso de métodos estadísticos, tales como componentes principales (PCA) y análisis factorial (FA), para la reducción de dimensión de los datos y los modelos de regresión y/o frontera estocástica, para medir el desempeño. El análisis de componentes principales transforma un conjunto de variables correlacionadas en un nuevo conjunto de menos variables no correlacionadas llamadas componentes principales, las cuales explican gran parte de la varianza total de las variables. De manera similar el análisis factorial permite reducir las p variables originales a m pocas variables nuevas, ni correlacionadas ni observables, llamadas factores adyacentes o factores comunes (variables latentes), basándose en la estructura de covarianza de los datos. Ambos métodos simplifican la dimensionalidad de los datos, con una mínima pérdida de información, por lo que en muchas ocasiones se usan indistintamente, sin embargo existen diferencias entre ambos que podrían afectar significativamente los resultados de modelos de orden superior. El presente trabajo desea evaluar el desempeño de éstos cuando son aplicados al modelo de frontera estocástica como método para la medición de la eficiencia de cada una de las universidades en un período de tiempo.

Capítulo 2

Eficiencia

La eficiencia se define como el máximo producto alcanzado dados unos insumos. Coelli (1998), este concepto aunque puramente económico ha trascendido a muchos campos, no solo a las entidades productivas, tal es el caso de las universidades, la necesidad de medir la eficiencia se viene desarrollando desde los años 60. Farrell (1957) eliminó la idea de la eficiencia absoluta propuso la idea de eficiencia relativa como una alternativa más real. Cada organización o unidad productiva individual es puesta en relación con aquéllas consideradas más eficaces, comparación de la que se desprenderá el grado de eficiencia de cada una de ellas. Allí propuso tres tipos de eficiencia:

- Eficiencia técnica: Supone utilizar correctamente los factores de producción, es decir, dados unos recursos obtener con ellos la máxima producción posible. Es por tanto, un concepto técnico y no económico, está asociada a la función de producción.
- Eficiencia asignativa, está asociada a la función de costos.
- Eficiencia de escala, está asociada a la función de beneficio.

Las fronteras expresan los valores límites que pueden alcanzar las empresas, por lo que es en relación a esa frontera que debe medirse la eficiencia de ellas. Usando como base la definición de Farrell, se han usado dos métodos para la estimación de las fronteras; las paramétricas y las no paramétricas, en este estudio se analiza la frontera estocástica incluida en las paramétricas. Ver Figura 2.

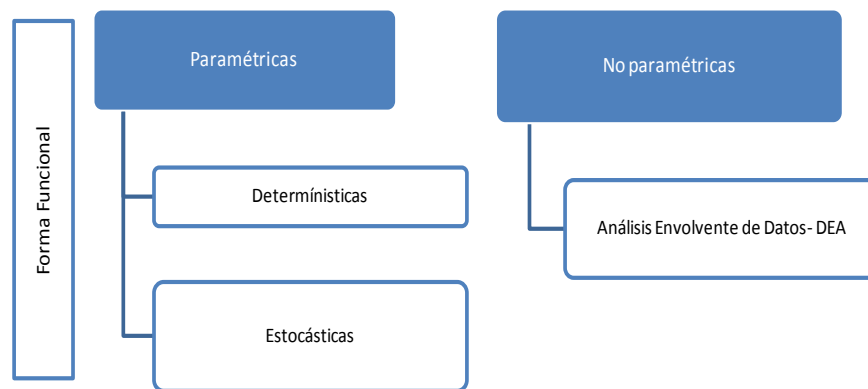


Figura 3. Tipos de fronteras de producción

Frontera estocástica

El modelo de frontera estocástica se planteó por Aigner, Lovell y Schmidt (1977). En este modelo las ineficiencias de las entidades pueden deberse no solamente a un mal uso de los inputs sino también a otras variables no conocidas, por lo que se incluye el término de error de tipo aleatorio que recoge información de todo aquello que no es controlable.

El modelo de frontera estocástica en su forma funcional se define como:

$$Y_i = f(X_i; \beta) e^{V_i - U_i}, \quad i = 1, 2, 3, \dots, N$$

Particularizando en la forma funcional más usada Cobb-Douglas:

$$\ln(Y_i) = \beta_0 + \beta \ln(X_i) + (V_i - U_i)$$

Donde Y_i , representa la producción o resultados para cada una de las entidades, f es la forma funcional con la que queremos comparar, X_i , las entradas o capacidad, β el vector de parámetros desconocido, V_i el término de error aleatorio que recoge los factores no controlables de la entidad, se distribuye como una $N(0, \sigma^2)$ e independiente del error U_i que es el término al que se le atribuye la ineficiencia, en general se distribuye como una Exponencial, Half-Normal o Normal truncada.

En este modelo la eficiencia se define como el cociente entre los resultados alcanzados y los resultados potenciales.

$$ET = \frac{f(X_i; \beta)e^{V_i - U_i}}{f(X_i; \beta)e^{V_i}} = e^{-U_i}$$

El modelo de frontera estocástica puede representarse gráficamente de la siguiente manera:

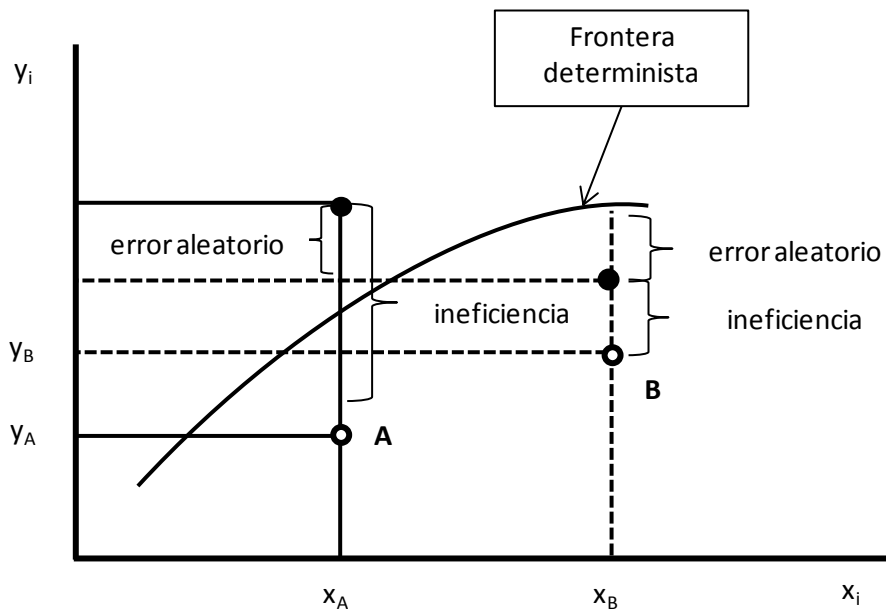


Figura 4. Esquema de la frontera estocástica de producción a partir de lo propuesto en Coelli (1998)

En la figura se representan dos entidades A y B, A usa el nivel de input x_A para obtener un nivel de outputs y_A de igual manera lo hace la entidad B, los resultados que obtienen ambas se dan ante la presencia de ineficiencia, representados por el círculo sin rellenar, sin embargo si el término de ineficiencia fuera cero, su producción alcanzaría niveles más altos, representado por el círculo sólido, estando así por encima o por debajo de la frontera de producción debido solamente al signo que tome el término de error V_i . Por lo que para la unidad B la frontera pasará por debajo de la función de producción debido a que el error V_B es negativo, por el contrario para la unidad A la frontera estará por encima de la función de producción dado que el error V_A , es positivo. Coelli (1998)

Para la estimación de la frontera estocástica de producción, se usan indicadores que permitan medir la capacidad y los resultados, dada la dicotomía de el modelo y la gran dimensión de los datos se hace necesario el uso de métodos multivariados, a fin de reducir las variables de resultados en una sola y eliminar la multicolinealidad existente en las variables de capacidad.

Análisis multivariado

El análisis multivariado consiste en un conjunto de métodos que permiten al investigador transformar un conjunto de datos grandes y complejos, donde existen estructuras de correlación y dependencia, en modelos de menor complejidad y de fácil interpretación, con los cuales se puede predecir el comportamiento de las unidades experimentales, evaluar el desempeño y evolución de las mismas, entre otros. Al trabajar con bases de datos tan robustas los métodos de reducción de dimensión, tales como componentes principales y análisis factorial, permiten reducir la dimensionalidad al igual que obtener una mejor interpretación de los modelos usados.

Componentes principales

“El análisis de componentes principales comprende un procedimiento matemático que transforma un conjunto de variables correlacionadas de respuesta en un conjunto menor de variables no correlacionadas llamadas componentes principales ” Jonson (2000). Este análisis busca maximizar la varianza de la combinación lineal de las variables Rencher (2012), por lo que un conjunto grande de datos puede explicarse a través de unas pocas combinaciones lineales de estas variables, lo que resulta útil para decantar los datos y reducir dimensión, a través de estas nuevas variables también se pueden realizar agrupaciones de las unidades experimentales en subgrupos semejantes, usar como variables de entrada para el análisis discriminante, resolver problemas de multicolinealidad y hacer comparaciones de subgrupos (poblaciones) a través de ANOVAS. El método de componentes principales se constituye en el paso inicial para la identificación de variables “latentes” o no observadas que generan la variabilidad de los datos. Peña (2003)

El problema que se desea resolver es cómo encontrar un espacio de dimensión más reducida que represente adecuadamente los datos. El problema puede abordarse desde tres perspectivas equivalentes: enfoque descriptivo, enfoque estadístico, enfoque geométrico. Peña (2003)

Enfoque descriptivo

Se desea encontrar un subespacio de dimensión menor que p tal que al proyectar sobre él los puntos conserven su estructura con la menor distorsión posible. Para un subespacio de dimensión uno, una recta. Se desea que las proyecciones de los puntos sobre esta recta mantengan, en la medida que sea posible, sus posiciones relativas.

Enfoque estadístico

Representar puntos p dimensionales con la mínima pérdida de información en un espacio de dimensión uno es equivalente a sustituir las p variables originales por una nueva variable Y_1 ,

que resuma óptimamente la información. Esto supone que la nueva variable debe tener globalmente máxima correlación con las originales o, en otros términos, debe permitir prever las variables originales con la máxima precisión.

Se debe encontrar una nueva dirección definida por un vector unitario, a_2 , que, sin pérdida de generalidad, puede tomarse ortogonal a a_1 , y que verifique la condición de que la proyección de un punto sobre este eje maximice las distancias entre los puntos proyectados. Estadísticamente esto equivale a encontrar una segunda variable Y_2 , incorrelada con la anterior, y que tenga varianza máxima. En general la componente $Y_r (r < p)$ tendrá varianza máxima entre todas las combinaciones lineales de las p variables X originales. Y estará condicionada a estar incorrelacionadas con las Y_1, Y_2, \dots, Y_{r-1} previamente obtenidas.

Enfoque geométrico

Para un subespacio de dimensión dos, la nube de puntos se sitúa siguiendo una elipse y se puede describir su orientación dando la dirección del eje mayor de la elipse y la posición de los puntos por su proyección sobre esta dirección. Este eje es la recta que minimiza las distancias ortogonales, al tener varias dimensiones se tendrán elipsoides y la mejor aproximación a los datos es la proporcionada por el eje mayor del elipsoide, considerar los ejes de éste como nuevas variables originales supone pasar de variables correlacionadas a variables ortogonales.

Modelo

Jonhson (2000) y Wichern(2007) definen el modelo de componentes principales como sigue:

Sea el vector aleatorio $X' : [X_1, X_2, \dots, X_p]$ con matriz de varianzas y covarianzas Σ , que tiene valores propios $\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots \geq \lambda_p \geq 0$

Entonces:

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

.

.

.

$$Y_p = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

$$\text{Con: } Var(Y_i) = a_i' \Sigma a_i \text{ para } i = 1, 2, 3, \dots, p$$

$$\text{y } Cov(Y_i Y_k) = a_i' \Sigma a_k \text{ para } i, k = 1, 2, 3, \dots, p$$

Las componentes principales Y_1, Y_2, \dots, Y_p , son aquellas combinaciones lineales no correlacionadas

La primera componente principal es la combinación lineal que maximiza:

$$Var(a_1' X), \text{ sujeto a: } a_1' a_1 = 1.$$

La segunda componente principal es la combinación lineal que maximiza:

$$Var(a_2'X), \text{ sujeto a: } a_2'a_2 = 1 \text{ y } Cov(y_1, y_2) = 0$$

Como resultado se tiene:

Sea Σ la matriz de varianzas y covarianzas del vector aleatorio $X': [X_1, X_2, \dots, X_p]$. Entonces Σ tiene de valores y vectores $(\lambda_1, a_1), (\lambda_2, a_2), \dots, (\lambda_p, a_p)$ os donde: $\lambda_1, \lambda_2, \lambda_3 \dots \geq \lambda_p \geq 0$

Entonces la i-ésima componente principal está dada por:

$$y_i = a_i'X = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \text{ para } i = 1, 2, \dots, p$$

$$Var(Y_i) = \lambda_i = a_i' \Sigma a_i$$

$$Cov(Y_i Y_k) = 0 = a_i' \Sigma a_k$$

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_i Var(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_i Var(Y_i)$$

Por lo que el total de la variabilidad: $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p$, y por tanto la proporción de la variabilidad total explicada por la componente k-ésima es:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_k}$$

Este método siempre debe ser usado como un primer paso, en el análisis multivariado, como una técnica exploratoria, para el uso de otras metodologías, entro otras razones por las propiedades de las nuevas variables.

Características de los componentes principales

- Las nuevas variables no están correlacionadas, son ortogonales
- La primera componente explica tanta variabilidad como sea posible. Jonhson (2000).
- La suma de las varianzas de los componentes es igual a la suma de las varianzas de las variables originales, y la varianza generalizada de los componentes es igual a la original. Peña (2003)
- Son un medio más que un fin, en general se convierten en el punto intermedio para la aplicación de otros modelos.
- Se ven afectadas por la estandarización de las variables, por lo que sólo se debería usar en caso de que las unidades de las variables tengan rangos muy diferentes. Wichern (2007)

Análisis factorial

El propósito esencial del análisis de factorial es describir, si es posible, las relaciones de covarianza entre muchas variables en términos de unas pocas, no correlacionadas, llamadas

factores subyacentes, no observables éstos recogen gran parte de la variabilidad total de los datos originales con una pérdida mínima de información, también deben proporcionar una mejor y más fácil comprensión y análisis de los datos originales que permita evaluar las unidades experimentales. Wichern (2007), menciona que es útil para estructuras donde se buscan constructos o latentes.

Wichern (2007), define el modelo de factor análisis así: sea el vector aleatorio \mathbf{X} , con p componentes, media $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$. El modelo de factores supone que \mathbf{X} es linealmente dependiente de unas pocas variables aleatorias no observables $F_1, F_2, F_3, \dots, F_m$, llamados factores subyacentes o **factores comunes** y de una fuente de variación adicional $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_p$, llamados errores o **factores específicos** tales que:

$$\begin{aligned}(X_1 - \mu_1) &= \ell_{11} F_1 + \ell_{12} F_2 + \dots + \ell_{1m} F_m + \varepsilon_1 \\(X_2 - \mu_2) &= \ell_{21} F_1 + \ell_{22} F_2 + \dots + \ell_{2m} F_m + \varepsilon_2 \\&\vdots \\(X_p - \mu_p) &= \ell_{p1} F_1 + \ell_{p2} F_2 + \dots + \ell_{pm} F_m + \varepsilon_p\end{aligned}$$

En su forma matricial así:

$$(\mathbf{X} - \boldsymbol{\mu})_{p \times 1} = \mathbf{L}_{p \times m} \mathbf{F}_{m \times 1} + \boldsymbol{\varepsilon}_{p \times 1}$$

donde:

ℓ_{ij} son llamadas las cargas de la i -ésima variable en el j -ésimo factor,

\mathbf{L} es la matriz de cargas o pesos del factor

el i -ésimo factor específico ε_i está asociado sólo con la i -ésima variable respuesta X_i .

Las P desviaciones $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ están expresadas en términos m variables aleatorias $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ que son no observables.

Supuestos

1. F Sigue una distribución $N(0, I)$.
2. $E[F] = \mathbf{0}_{m \times 1}$, $Cov(F) = E[FF'] = I_{m \times m}$, donde, $I_{m \times m}$ Es la matriz identidad.
3. $E[\boldsymbol{\varepsilon}] = \mathbf{0}_{p \times 1}$, $Cov(\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\psi}_{p \times p}$, donde, $\boldsymbol{\psi}_{p \times p}$: Es una matriz diagonal.

$$Con \boldsymbol{\psi} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & \dots & \dots & \psi_p \end{bmatrix}$$

donde, F y E son independientes, $Cov[\varepsilon, F] = E[EF'] = O_{pxm}$

Entonces el modelo $(X - \mu)_{px1} = L_{pxm} F_{mx1} + \varepsilon_{px1}$, bajo estos supuestos se constituye en el modelo de factor ortogonal, el cual define la estructura de covarianza como:

- $Cov(X) = \Sigma LL' + \psi$, o

$$Var(X_i) = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 + \psi_i$$

$$Cov(X_i, X_k) = \ell_{i1} * \ell_{k1} + \dots + \ell_{im} * \ell_{km}$$

- $Cov(X, F) = L$, o

$$Cov(X_i, F_j) = \ell_{ij}$$

La varianza de la i-ésima variable contribuida por los m factores comunes es llamada la i-ésima *comunalidad*. La componente de $var(x_i) = \sigma_{ii}$ debida al factor específico es llamada la varianza específica (única). Se demuestra por h_i^2 la i-ésima comunalidad.

$$\underbrace{\sigma_{ii}}_{Var(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2}_{comunalidad} + \underbrace{\psi_i}_{varianza\ específica}$$

Donde, $h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2$ y $\sigma_{ii} = h_i^2 + \psi_i, i = 1, 2, 3, \dots, p$

La i-ésima comunalidad es la suma de cuadrados de las cargas de la i-ésima variable sobre los m factores comunes.

La razón $(\sum_k^m \ell_{ik}^2) / \sigma_{ii}$ es la proporción de la varianza de X_i que es explicada por los factores comunes.

Unicidad del modelo

En el modelo $(X - U) = LF + \varepsilon$, la matriz de cargas L y los factores F son no observables, lo cual plantea un problema de indeterminación, que se resuelve imponiendo restricciones sobre la matriz de cargas L. Existen dos criterios de normalización que son utilizados en los métodos de estimación:

1. Exigir que $L'L = D$: Diagonal mediante el cual se obtienen factores sobre las P variables observadas ortogonales.
2. Exigir que $L'\psi^{-1}L = D$: Diagonal, así los factores se hacen incorrelacionados

Dado lo anterior se establecen los métodos de estimación.

Método del factor principal

En este método se estima la matriz de carga basada en componentes principales, donde Σ que tiene las parejas de valores y vectores propios $((\lambda_i, e_i))$ con $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, entonces Σ puede descomponerse en:

$$\begin{aligned} \Sigma &= \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \dots + \lambda_p e_p e_p' \\ &= \left[\sqrt{\lambda_1} e_1 : \sqrt{\lambda_2} e_2 : \dots : \sqrt{\lambda_p} e_p \right] * \begin{bmatrix} \sqrt{\lambda_1} e_1' \\ \dots \\ \sqrt{\lambda_2} e_2' \\ \dots \\ \dots \\ \dots \\ \sqrt{\lambda_p} e_p' \end{bmatrix} \\ &= LL' + 0 = LL' \end{aligned}$$

Esta expresión supone que la varianza del error específico puede ser ignorado, sin embargo, si se quiere estimar a ψ_i podría hacerse mediante $\widehat{\psi}_i = S_{ii} - \sum_j^p \widehat{\ell}_{ij}^2$ de $\psi = \Sigma - LL'$ con $\widetilde{h}_i^2 = \widetilde{\ell}_{i1}^2 + \widetilde{\ell}_{i2}^2 + \dots + \widetilde{\ell}_{im}^2$.

En este método los estimadores son consistentes pero no eficientes y no son invariantes a transformaciones lineales.

Método de máxima verosimilitud

En este caso las observaciones $(X - U) = LF + \varepsilon$ tienen distribución $N(\mu, \Sigma)$, y al maximizar la función de máximo – verosimilitud $L(\mu, \Sigma)$ es necesario normalizar con $L' \psi^{-1} L = D$, siendo D una matriz diagonal.

Donde,

$$L(\mu, \Sigma) = (2\pi)^{\frac{-np}{2}} |\Sigma|^{\frac{-n}{2}} e^{\frac{-1}{2} \text{tr} \left[\Sigma^{-1} ((X_j - \bar{X})(X_j - \bar{X})' + n(\bar{X} - \mu)(\bar{X} - \mu)') \right]}$$

La cual depende de L y ψ a través de $\Sigma = LL' + \psi$ obteniéndose:

$$\widehat{L}_z = \widehat{V}^{-1/2} \widehat{L} \quad \text{y} \quad \widehat{\psi}_z = \widehat{V}^{-1/2} \psi V^{-1/2} \quad \text{con} \quad \widehat{\mu} = \bar{x}.$$

Donde \hat{L}_z y $\hat{\psi}_z$ son los estimadores a partir de los datos estandarizados y, a partir de estos se hace la re-transformación para obtener \hat{L} y $\hat{\psi}$.

Rotación de los ejes

La rotación significa que se giran los ejes en el origen hasta alcanzar una determinada posición. Existen dos tipos de rotación, la ortogonal y la oblicua. En la transformación ortogonal el eje se mantienen formando un ángulo de 90°, manteniendo la significancia de los factores y simplificando por filas y columnas de la matriz de factores, facilitando la interpretación de los factores. Mientras que la rotación oblicua no limita el ángulo de rotación, permitiendo la existencia de factores correlacionados en lugar de mantener la independencia entre los factores rotados.

Varimax

Entre los métodos de rotación ortogonal se encuentra el varimax, que minimiza la suma de las varianzas de las cargas requeridas de la matriz de factores, haciendo que las cargas factoriales sean muy altas y algunas cargas cercanas a cero, indicando para valores altos una clara asociación positiva o negativa entre las variables y el factor. Cargas cercanas a cero, señalan una clara ausencia de asociación, así si el objetivo es reducir el número de variables originales, con independencia de la significancia de los factores, facilitando su interpretación, la rotación sugerida sería la ortogonal (Varimax).

Características de análisis factorial

- Permite explicar un conjunto grande de variables en uno menor a través de variables latentes o no observadas.
- Permite reducir la redundancia entre las variables usando un número pequeño de factores. Rencher (2012).
- Reducir dimensionalidad en los datos mediante el método de análisis factorial es más complicado que usar componentes principales, sin embargo la rotación VARIMAX, que maximiza la varianza, y que es permitida cuando se usa análisis factorial, mejora la interpretación.
- Para la interpretación de las variables que determinan el factor debe considerarse aquellas que están por encima de 0.4. Ahora en el número de factores o nuevas variables determinadas corresponde al número de valores propios diferentes de cero productos del proceso de estimación. Wichern (2007).

Diferencias entre análisis factorial y componentes principales

Ambos métodos de reducción de dimensión suelen relacionarse y usarse indistintamente debido a que los dos cumplen con el objetivo de reducir dimensión sin embargo existen algunas diferencias:

- Los componentes principales son definidos como combinaciones lineales de las variables originales, mientras que análisis factorial expresa las variables originales como combinaciones lineales de los factores. Rencher (2012)

- Análisis factorial busca explicar las covarianzas entre las variables, de otro lado componentes principales explica gran parte de la varianza total de las variables. Rencher (2012)
- La posibilidad de rotar los ejes para mejorar la interpretación ofrece una ventaja única de análisis factorial sobre componentes principales.
- Los componentes principales son una herramienta descriptiva, mientras que el análisis factorial propone un modelo estadístico formal. Peña (2003)
- Cuando $\widehat{\psi}_i \approx 0$, es lo mismo tomar m componentes principales que estimar m factores, la diferencia será menor cuanto menor sea $\widehat{\psi}_i$. Peña (2003)

Capítulo 3

Simulaciones

El objetivo de este estudio es evaluar el desempeño de los métodos de reducción de dimensión, aplicados al modelo de frontera estocástica de producción, el cual permite medir la eficiencia de una unidad experimental como la combinación entre varios insumos (capacidad) que dan como resultado un único producto (resultado), este modelo exige una forma funcional para su estimación, un supuesto de distribución para los errores de ineficiencia y para los aleatorios.

A fin de garantizar que las variables de entrada y salida procedieran de un modelo de producción con presencia de ambos errores; ineficiencia y aleatorio en el diseño se obtuvieron los datos mediante simulación Monte Carlo, para los cuales se fijo la distribución de las ineficiencias, el error aleatorio y la función de producción. Se generaron las entradas, como cinco variables provenientes de una distribución *Normal multivariada* ($\mu = \mathbf{0}, \Sigma$), la matriz sigma se estimó como la matriz \mathbf{R} , con correlaciones aleatorias entre 0.7 y 0.9, y las salidas se generaron como un modelo de frontera estocástica, particularmente con la forma funcional Cobb-Douglas:

$$\ln(Y_i) = \beta \ln(X_i) + (V_i - U_i),$$

los β como una distribución uniforme entre 0.5 y 1, lo cual indica que los rendimientos a escala son crecientes, V_i como una distribución *Normal multivariada* ($\mu = \mathbf{0}, \Sigma$), con correlaciones aleatorias entre 0.7 y 0.9, a fin de garantizar las correlaciones en la matriz \mathbf{Y} , U_i como una distribución exponencial ($\lambda=0.5$) o como una distribución Half-Normal ($\mu = 3, \sigma^2 = 6$). En ambos casos se usaron tamaños de muestra $n=30,80$ y 200 .

Posteriormente para evaluar el desempeño de los métodos se asignó que un 1% de las unidades fueran eficientes técnicamente, para este caso se tomó un tamaño de muestra igual a 1000, manteniendo lo demás constante.

Luego de obtener los datos se reduce dimensión en las \mathbf{Y} y en las \mathbf{X} , por los métodos a comparar; análisis factorial y componentes principales, obteniendo de esta manera dos nuevos vectores que serán respectivamente la capacidad (insumos) y los resultados (productos), con ellos se aplica el modelo de frontera estocástica y se estiman las eficiencias de las unidades. La

siguiente figura ilustra la primera parte del procedimiento. De igual manera se realiza para $n=1000$ con unidades eficientes técnicamente.

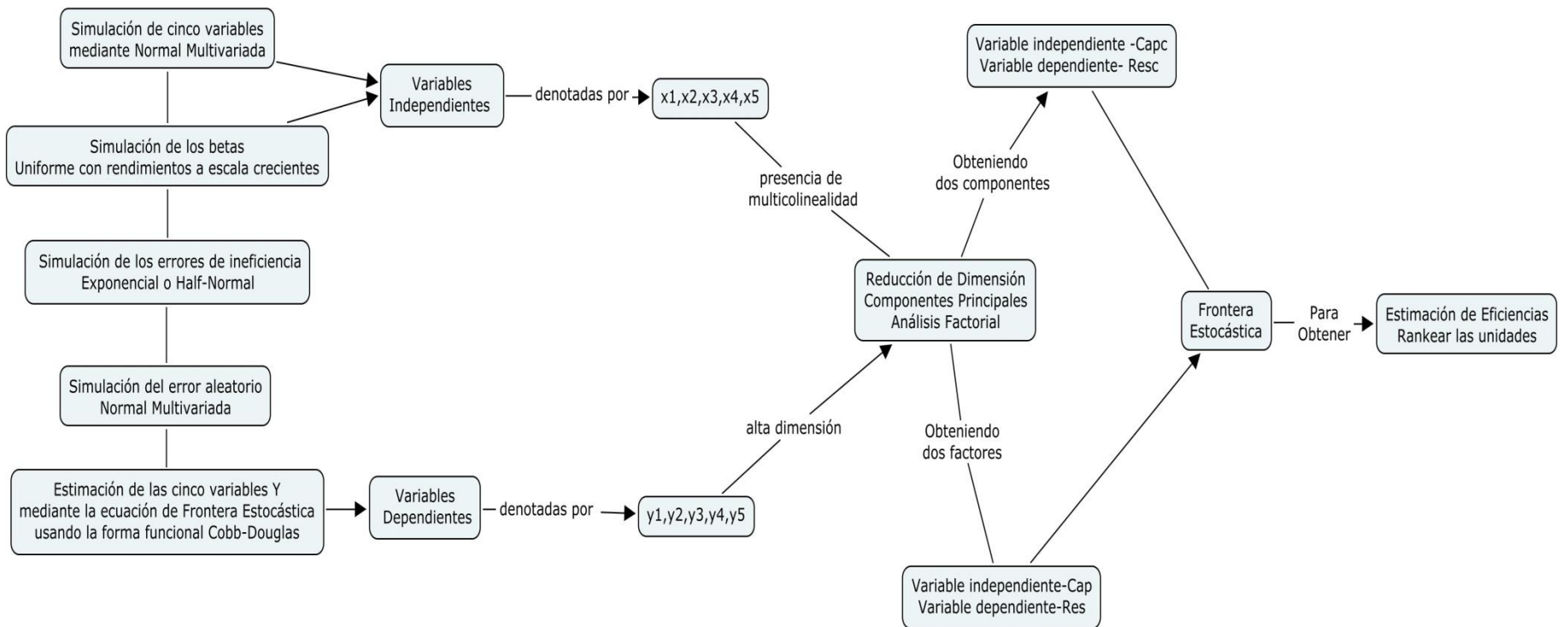


Figura 5. Procedimiento para la estimación de frontera estocástica a partir de métodos de reducción de dimensión Componentes Principales y Análisis Factorial

Resultados

Las estimaciones obtenidas en el modelo de frontera estocástica de producción para cada tamaño de muestra, método de reducción de dimensión y distribución del término de ineficiencia se muestran en las siguientes tablas, el software usado fue Stata 12[®]. Para la lectura adecuada de los resultados se deberá tener presente lo siguiente:

1. Ineficiencia igual a cero

$$H_0: \sigma_U = 0 \quad H_1: \sigma_U \neq 0$$

Prueba para validar la hipótesis Chibar2(01)

2. Proporción de la varianza del término de ineficiencia sobre la varianza del término de error aleatorio

$$\lambda = \frac{\sigma_u}{\sigma_v}$$

Caso 1: Distribución del término de ineficiencia Exponencial ($\lambda=0.5$)

		n=30		n=80		n=200	
		Análisis Factorial	Componentes Principales	Análisis Factorial	Componentes Principales	Análisis Factorial	Componentes Principales
Correlación		0.8048	0.8163	0.8253	0.8207	0.6862	0.6869
x		0.7874	0.6896	0.7482	0.6387	0.5693	0.4775
valor P		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
cons		0.4383	0.6537	0.3623	0.7042	0.4611	0.8716
valor P		0.0100	0.0530	0.0000	0.0000	0.0000	0.0000
sigma u		0.4383	0.6537	0.3623	0.7042	0.4611	0.8716
sigma v		0.3583	0.8316	0.3539	0.6863	0.4162	0.7733
sigma 2		0.3205	1.1190	0.2565	0.9668	0.3859	1.3578
Lambda		1.2230	0.7861	1.0236	1.0261	1.1080	1.1272
chibar2(01)		2.4000	1.0000	4.5900	4.9200	10.3900	10.7200
valor P		0.0600	0.1580	0.0160	0.0130	0.0010	0.0010
Wald chi2(1)		83.1700	70.3600	220.2300	214.1300	212.2600	213.8600
valor P		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Eficiencias	Mínimo	0.1796	No aplica ²	0.2722	0.0650	0.1385	0.0281
	Máximo	0.8935		0.9210	0.8585	0.8866	0.8073
	Media	0.6942		0.7342	0.5878	0.6828	0.5322
	Desviación estándar	0.1710		0.1480	0.1801	0.1660	0.1936

Tabla 1. Estimación de la frontera estocástica de producción. Caso 1: Con distribución exponencial en el término de ineficiencia U_i

² No se realiza el análisis univariado para las eficiencias resultantes dado que este caso el valor p de la prueba χ es mayor que el nivel de significancia 0.1, por lo que no se puede rechazar $H_0: \sigma_u = 0$.

Se puede observar que cuando el tamaño de muestra es pequeño, el modelo de frontera estocástica aplica al usar el método de reducción de dimensión análisis factorial, lo que inclinaría la balanza así este, sin embargo a medida que aumenta el tamaño de muestra, no existe diferencia entre los métodos de reducción de dimensión, salvo que la eficiencia media obtenida por el método de análisis factorial siempre es mayor, como se puede ver en las siguientes gráficas, para cada uno de los tamaños de muestra donde el modelo de frontera fue significativo.

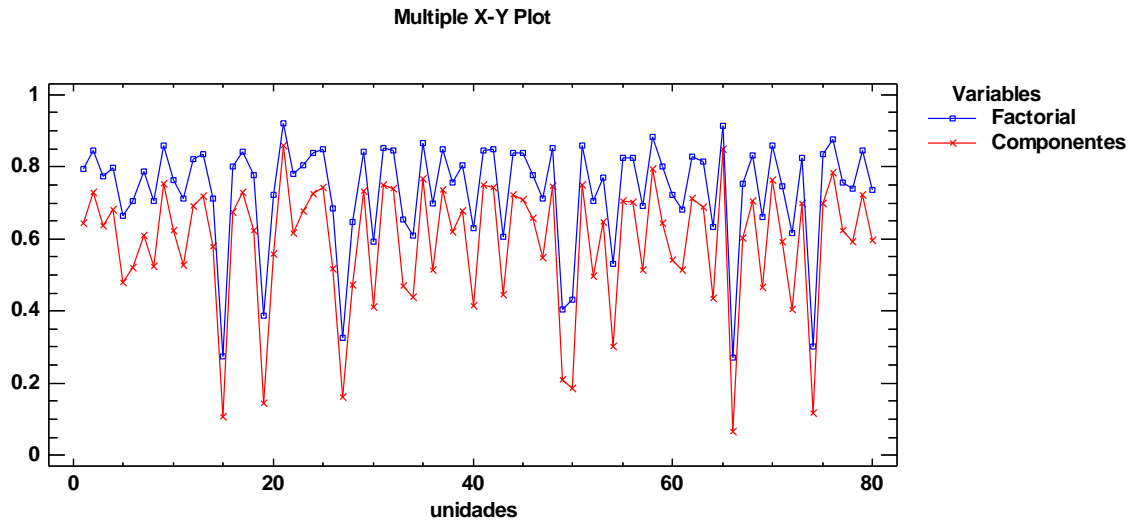


Gráfico 1. Eficiencias estimadas por ambos métodos de reducción de dimensión para $n=80$. Caso 1: Con distribución exponencial en el término de ineficiencia U_i

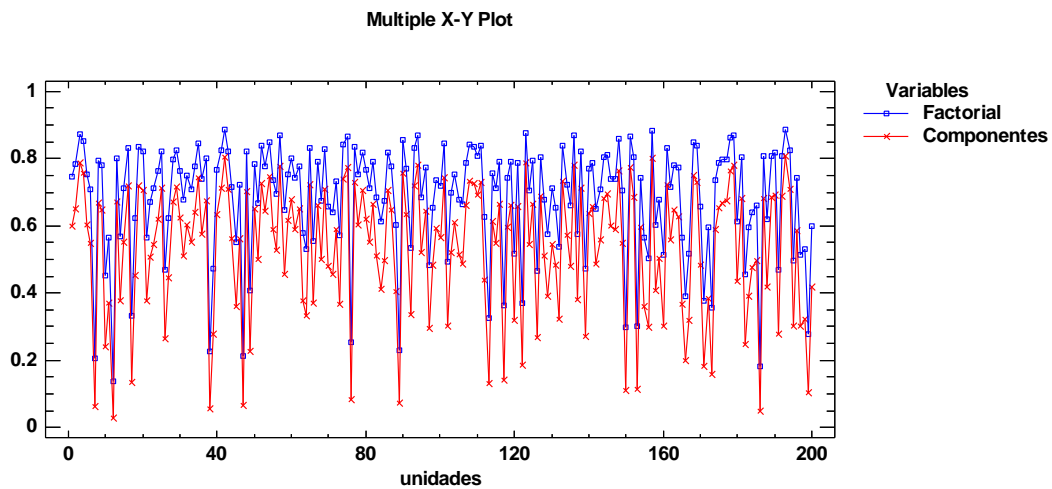


Gráfico 2. Eficiencias estimadas por ambos métodos de reducción de dimensión para $n=200$. Caso 1: Con distribución exponencial en el término de ineficiencia U_i

Caso 2: Distribución del término de ineficiencia Half-Normal ($\mu = 3, \sigma^2 = 6$).

		n=30		n=80		n=200	
		Análisis Factorial	Componentes Principales	Análisis Factorial	Componentes Principales	Análisis Factorial	Componentes Principales
Correlación		0.7779	0.7766	0.7594	0.7691	0.6325	0.6342
X		0.7642	0.6864	0.6580	0.5471	0.4770	0.3958
valor P		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
cons		0.5948	1.2453	0.6530	1.1239	0.5415	1.0111
valor P		0.0100	0.0040	0.0000	0.0000	0.0000	0.0000
sigma u		0.7449	1.5603	0.8126	1.4019	0.6792	1.2684
Sigma v		0.3714	0.6805	0.2339	0.4842	0.4256	0.7788
Sigma 2		0.6928	2.8975	0.7150	2.1998	0.6424	2.2154
Lambda		2.0054	2.2927	3.4743	2.8956	1.5958	1.6288
chibar2(01)		0.89	1.2900	6.1600	4.7600	4.0700	4.3900
valor P		0.172	0.1280	0.0070	0.0150	0.0220	0.0180
Wald chi2(1)		54.76	55.3500	139.8700	139.2400	125.2600	125.6600
valor P		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Eficiencias	Mínimo	No aplica		0.1399	0.0393	0.1634	0.0350878
	Máximo			0.8965	0.8238	0.8618	0.770476
	Media			0.5764	0.4281	0.6262	0.45776
	Desviación estándar			0.2084	0.2228	0.1469	0.172143

Tabla 2. Estimación de la frontera estocástica de producción. Caso 2: Con distribución half-normal en el término de ineficiencia U_i

Para este caso, cuando el tamaño de muestra es pequeño el modelo de frontera estocástica no aplica bajo ninguno de los métodos de reducción de dimensión, sin embargo al aumentar el tamaño de muestra, no existe diferencia entre los métodos, salvo que la eficiencia media obtenida por el método de análisis factorial siempre es mayor como se puede ver en las siguientes gráficas.

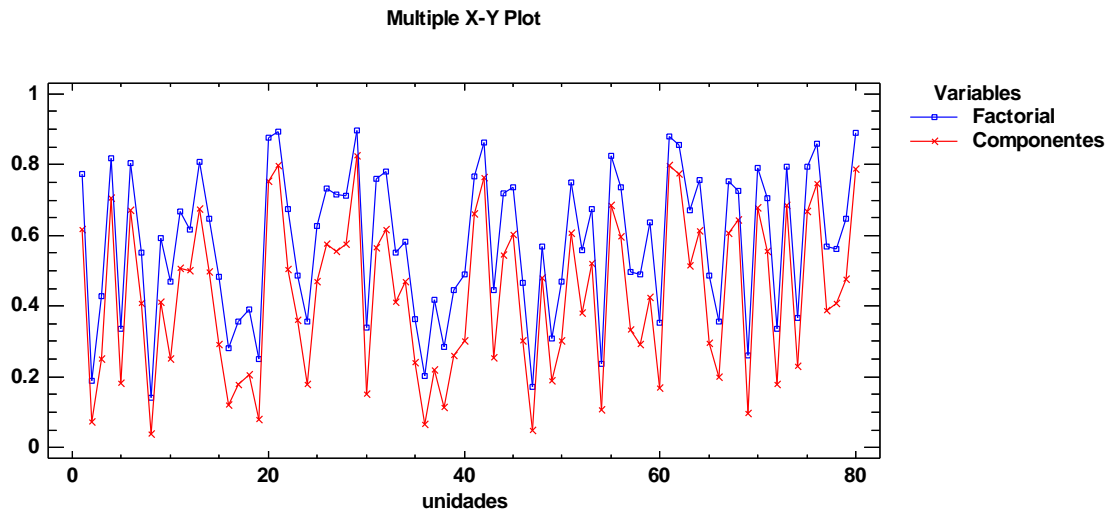


Gráfico 3. Eficiencias estimadas por ambos métodos de reducción de dimensión para $n=80$. Caso 2: Con distribución half-normal en el término de ineficiencia U_i

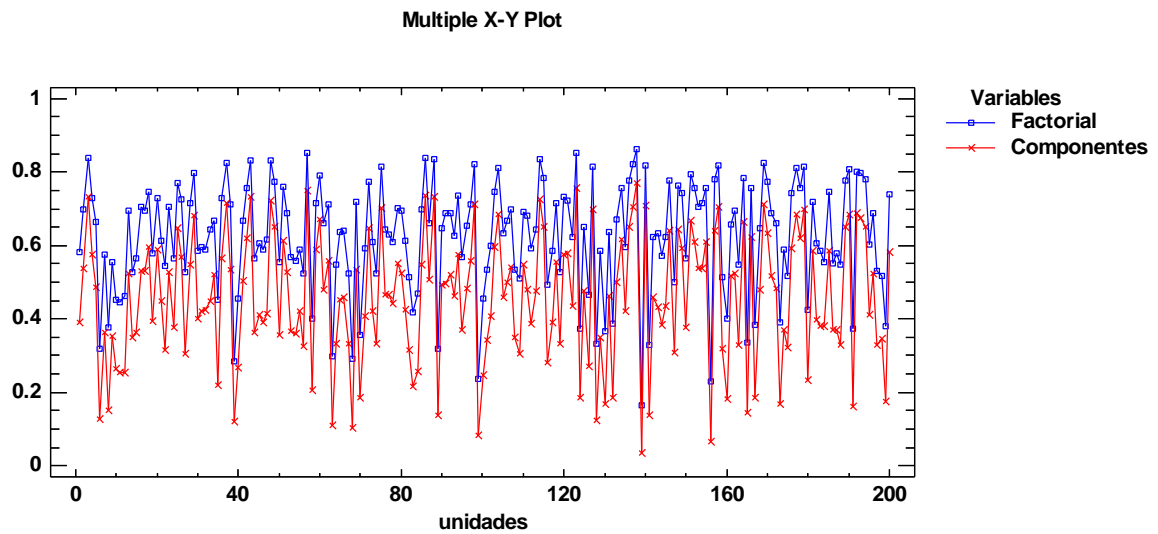


Gráfico 4. . Eficiencias estimadas por ambos métodos de reducción de dimensión para $n=200$. Caso 2: Con distribución half-normal en el término de ineficiencia U_i

Resultados con unidades eficientes técnicamente

En los resultados obtenidos se encontró que para todos los casos la eficiencia media y la de cada unidad es mayor al aplicar el método de análisis factorial, por lo que se desea comprobar cuál de los dos métodos asigna la eficiencia adecuada, cuando en los datos existen unidades eficientes técnicamente. Para esto se asignó al 1% de las unidades, cero en el componente de ineficiencia, manteniendo lo demás constante.

Caso 3: Simulaciones con unidades eficientes técnicamente y distribución Exponencial en el término de ineficiencia U_i

	Exponencial	
	Análisis Factorial	Componentes Principales
Correlación	0.6451	0.6427
X	0.5025	0.4167
valor P	0	0
cons	0.2329	0.4309
valor P	0	0
sigma u	0.2329	0.4309
sigma v	0.5454	1.0191
sigma 2	0.3517	1.2241
lambda	0.4271	0.4228
chibar2(01)	2.22	2.13
valor P	0.068	0.072
wald chi2(1)	721.56	712.31
valor P	0	0

Tabla 3. Estimación de la frontera estocástica de producción. Caso 3: Con distribución Exponencial y 1% de las unidades con cero en el término de ineficiencia U_i

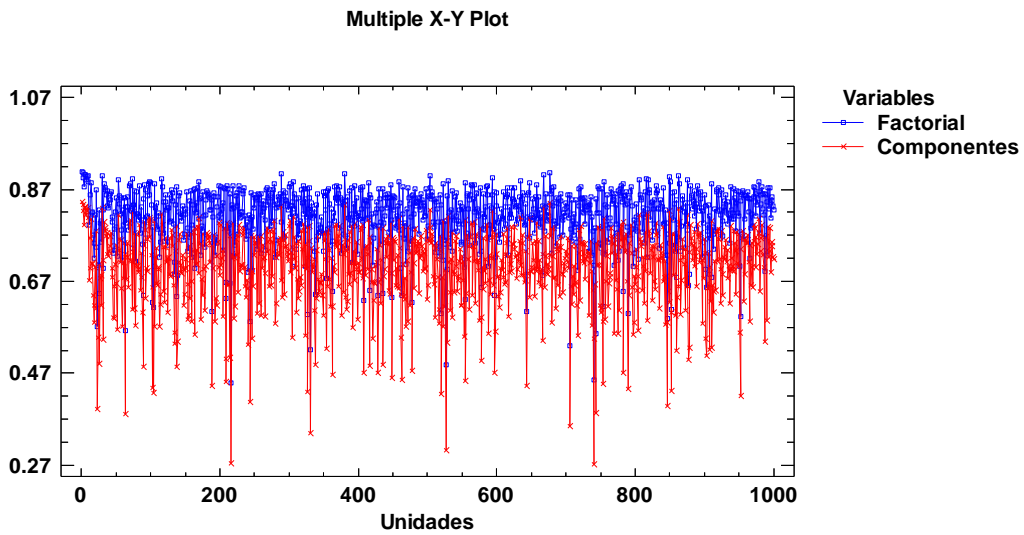


Gráfico 5. Eficiencias estimadas por ambos métodos de reducción de dimensión para $n=1000$. Caso 3: Con distribución exponencial y 1% de las unidades con cero en el término de ineficiencia U_i

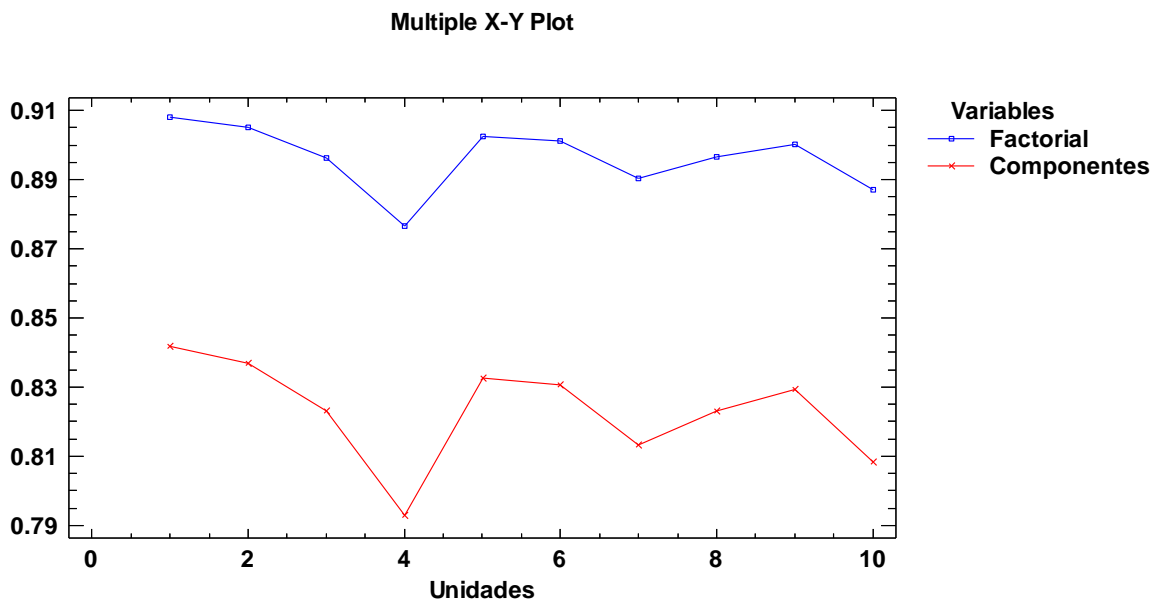


Gráfico 6. Eficiencias estimadas por ambos métodos de reducción de dimensión para las unidades con cero en el término de ineficiencia U_i

Se observa que las eficiencias son mayores en el método de reducción de dimensión, análisis factorial que en el de componentes principales, además para las diez unidades que se asignan

eficientes técnicamente, con análisis factorial obtienen eficiencias muy cercanas al 100% que difieren en siete centésimas en promedio de los valores dados por componentes principales.

Caso 4: Simulaciones con unidades eficientes técnicamente y distribución Half-Normal en el término de ineficiencia U_i

	Half-Normal	
	Análisis Factorial	Componentes Principales
Correlación	0.6451	0.6427
X	0.5025	0.4167
valor P	0	0
cons	0.4175	0.7656
valor P	0	0
sigma u	0.5232	0.9594
sigma v	0.5019	0.9426
sigma 2	0.5257	1.809
lambda	1.0424	1.0177
chibar2(01)	3.4	3.12
valor P	0.033	0.039
wald chi2(1)	722.23	712.34
valor P	0	0

Tabla 4. Estimación de la frontera estocástica de producción. Caso 4: Con distribución half-normal y 1% de las unidades con cero en el término de ineficiencia U_i

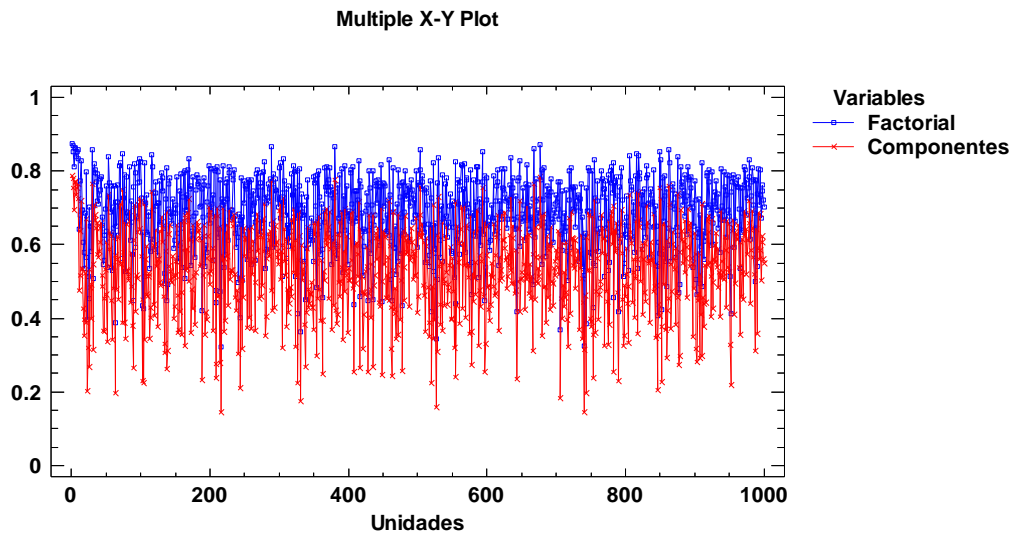


Gráfico 7. Eficiencias estimadas por ambos métodos de reducción de dimensión para $n=1000$. Caso 4: Con distribución half-normal y 1% de las unidades con cero en el término de ineficiencia U_i

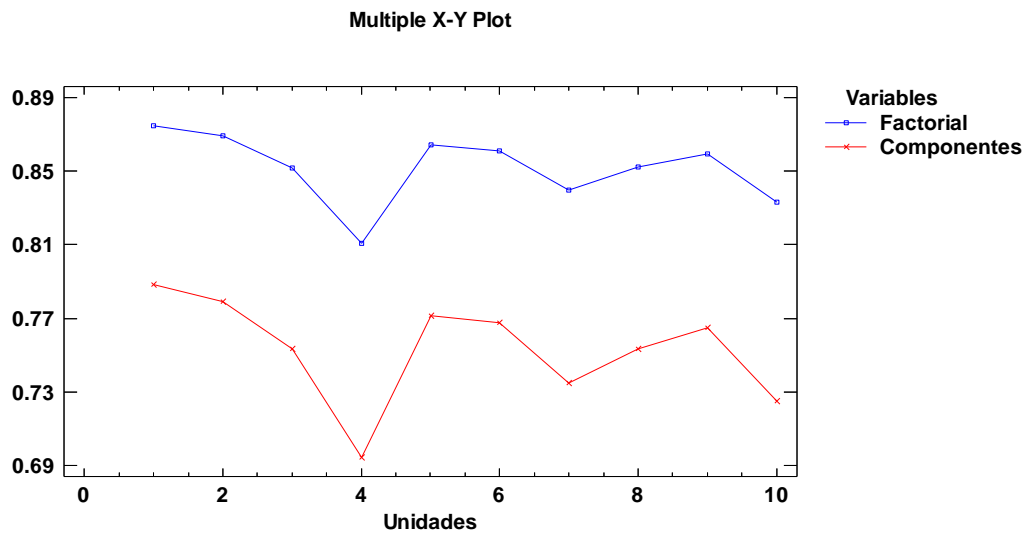


Gráfico 8. Eficiencias estimadas por ambos métodos de reducción de dimensión para las unidades con cero en el término de ineficiencia U_i

Al igual que cuando el término de ineficiencia es Exponencial, las eficiencias asignadas por el método de análisis factorial son mayores que las de componentes principales, cuando se mira en detalle los resultados para las diez unidades que se asignan eficientes técnicamente, estos valores difieren en 1 décima en promedio.

Caso de estudio

En el caso de estudio se mide la eficiencia de las 32 universidades del Sistema Universitario Estatal en el eje misional investigación aplicado a visibilidad científica, a partir de 13 indicadores que dan cuenta de este proceso, cuatro de ellos corresponden a lo que se interpreta como la capacidad y son las variables independientes, los cuales son: área, presupuesto, gasto administrativo y profesores, por otro lado las variables dependientes corresponden a grupos de investigación, artículos, citas, factor de impacto, revistas indexadas, graduados de maestría y doctorado, movilidad docente, patentes, licencias. Aunque el sistema de investigación de las Universidades públicas es complejo, el método de frontera estocástica permite obtener un modelo apropiado para la medición de su desempeño, aunque todavía se trate de un modelo de tipo dicotómico capacidad-resultados que no tiene en cuenta los procesos que se derivan al interior, así como el hecho de que existen variables que presentan doble vía, es decir, que se constituyen en capacidad y resultados, para dar cuenta de estos se tiene modelos que permiten interpretar estructuras más complejas como es el caso de los Modelos de Ecuaciones Estructurales (SEM).

Reducción de dimensión

Para obtener un solo indicador que represente la latente capacidad y que elimine la multicolinealidad existente en los insumos y de otra parte reducir la dimensión de los resultados y así poder aplicar el modelo de frontera estocástica, se realiza la reducción de dimensión por los métodos de análisis factorial y componentes principales; el primero de ellos se estima por el método de máxima verosimilitud y usando la rotación Varimax, para explicar la máxima varianza posible, en ambos casos se usa el factor principal y el componente principal como las nuevas variables. Los resultados obtenidos se muestran a continuación.

	Capacidad		Resultados	
	Análisis Factorial	Componentes Principales	Análisis Factorial	Componentes Principales
Proporción	1.0024	0.8475	0.8722	0.5704

Tabla 5. Proporción de información explicada por el factor y componente principal

Se destaca que tanto para las nuevas variables capacidad y resultados el método de reducción de dimensión análisis factorial lleva más información que el de componentes principales, lo que es de gran importancia en el uso del modelo de frontera estocástica, dado que en éste la ineficiencia de las unidades experimentales es debida a los errores aleatorios, variables que no pueden controlar y a la combinación usada de sus entradas para obtener los resultados, por lo que el hecho de llevar más información de las variables permitirá obtener un valor de eficiencia más real.

Frontera estocástica

Con las nuevas variables se chequea la correlación entre ellas como primer indicio de la aplicación del modelo, una correlación de 0.9 o mayor sugerirá el uso de regresión lineal por

encima de frontera estocástica, dado que se supone que el error presente allí es sólo debido a errores aleatorios, además se parte del principio de parsimonia. Este análisis se hace de manera intuitiva puesto que no existen pruebas previas para la elección del modelo de frontera estocástica, ni conocimiento a priori de la distribución que siguen las ineficiencias. Las correlaciones entre las nuevas variables son:

Análisis Factorial (Cap-Res)	Componentes Principales (Capc-Resc)
0.8773	0.8673

Tabla 6. Correlación entre las nuevas variables capacidad y resultados con ambos métodos de reducción de dimensión

Se aplica el modelo de frontera estocástica para las nuevas variables bajo el supuesto de distribución de las ineficiencias; Exponencial y Half-normal dado que son los modelos más usuales Coell (1998). Donde se desea probar si existe término de ineficiencia, la hipótesis a probar es la siguiente:

$$H_0: \sigma_U = 0 \quad H_1: \sigma_U \neq 0$$

		Half-Normal		Exponencial	
		Análisis Factorial	Componentes Principales	Análisis Factorial	Componentes Principales
Correlación		0.8773	0.8673	0.8773	0.8673
Cap		0.8928	1.0754	0.9024	1.0860
valor P		0.0000	0.0000	0.0000	0.0000
cons		0.4317	0.8218	0.3059	0.6073
valor P		0.0050	0.1200	0.0110	0.0670
sigma u		0.5465	1.0341	0.3059	0.6073
sigma v		0.3118	0.9158	0.3323	0.9207
sigma 2		0.3959	1.9080	0.2040	1.2165
lambda		1.7530	1.1292	0.9206	0.6596
chibar2(01)		1.4500	0.3400	2.3500	0.8100
valor P		0.1140	0.2800	0.0630	0.1850
wald chi2(1)		147.3500	110.1000	156.6500	118.0000
valor P		0.0000	0.0000	0.0000	0.0000
Eficiencias	Mínimo	0.2568	No aplica		
	Máximo	0.9146			
	Media	0.7671			
	Desviación estándar	0.1305			
Distribución de los u					

Tabla 7. Estimación de la frontera estocástica de producción para el caso de estudio, con distribución en el término de ineficiencia U_j Half-normal y Exponencial.

Los resultados obtenidos muestran que el modelo de frontera estocástica sólo aplica cuando el método de reducción de dimensión es análisis factorial y el supuesto de distribución en el término de error Exponencial dado que el valor p , para el modelo con distribución de error Exponencial es menor que el nivel de significancia de 0.1, por lo cual se rechaza la hipótesis nula en favor de la alternativa, es decir, la ineficiencia de las universidades no sólo se debe a variables externas que no pueden controlar o errores aleatorios, también es debida al uso inadecuado de sus entradas para obtener los resultados esperados. La siguiente gráfica muestra el ranking obtenido para las universidades del Sistema Universitario Estatal.

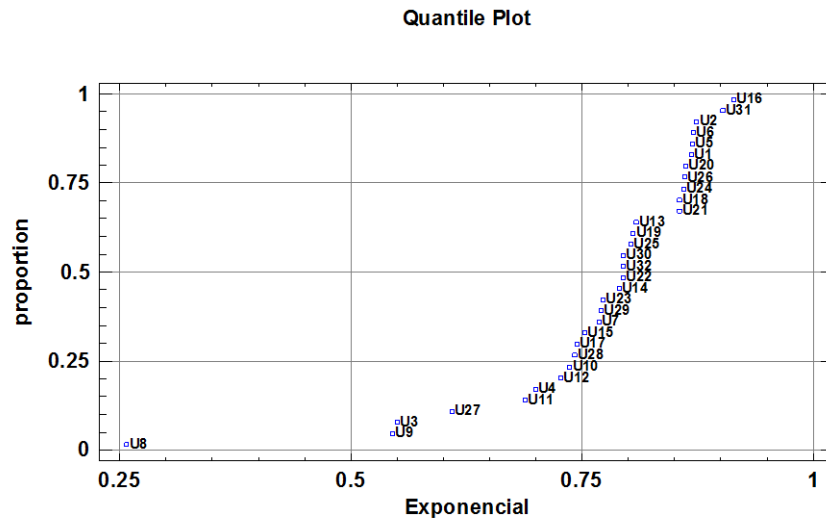


Gráfico 9. Ranking de las Universidades con supuesto de distribución en el término de ineficiencia U_i Exponencial

CONCLUSIONES

1. En el experimento Monte Carlo se usaron diferentes tamaños de muestra, ambos métodos de reducción de dimensión análisis factorial y componentes principales así como varios supuestos para el término de error de ineficiencia. En los casos cuando el supuesto de la distribución del término de error se comporta como una Half-Normal y el tamaño de muestra es grande ($n > 80$) no existe diferencia significativa entre los métodos de reducción aplicados, por otro lado cuando el supuesto de distribución del error es exponencial y n es pequeño existe mejor desempeño del método de análisis factorial que el de componentes principales, ya que las eficiencias en el primero de ellos muestran una mayor dispersión, identificando unidades experimentales eficientes cercanas al 100%, mientras que en el segundo los valores de las ineficiencias distan en un 20 ó 30% de la eficiencia óptima, determinando con ello que no existen unidades eficientes al 100%, en otras palabras, el uso equivocado de los métodos de reducción de dimensión podría subestimar el desempeño de una unidad experimental situándola por debajo del verdadero valor de su eficiencia, por lo que se recomienda el uso del análisis factorial en estos casos.
2. En los estudios realizados se encontraron casos en donde la prueba global no muestra diferencia significativa entre los métodos de reducción de dimensión, sin embargo las eficiencias asignadas a las unidades experimentales se ven afectadas por el método usado. Para todos los casos la eficiencia media asignada al usar análisis factorial fue mayor que al usar componentes principales, esto es debido a que aquel es un modelo que lleva información no sólo de la varianza que es común a los datos, sino también de la específica, lo que significa que tendrá más información de las unidades experimentales y podrá determinarse claramente que parte de la ineficiencia es debida al error aleatorio y cual a una mala combinación de la relación capacidad resultado, lo que determina un mejor desempeño del método de análisis factorial sobre el de componentes principales al aplicar el modelo de frontera estocástica.
3. Cuando existen unidades experimentales eficientes técnicamente, es el método de análisis factorial el que asigna el valor de eficiencia correspondiente a la unidad, valor que debe ser cercano al 100%.
4. En la práctica es común encontrar el uso indiscriminado de ambos métodos de reducción de dimensión, sobre todo motivada por la simplicidad que ofrece componentes principales y el no enfrentar no sólo la complejidad del análisis factorial, los supuestos de éste y la necesidad de su validación, tanto del caso de estudio como de las simulaciones se recomienda considerar ambos métodos y comparar sus resultados, aunque de este estudio se sugiere hacer uso preferencialmente del análisis factorial.
5. Para el caso de estudio el modelo de frontera estocástica sólo aplica al realizar la reducción de dimensión mediante el método de análisis factorial y bajo el supuesto de distribución en el término de ineficiencia Exponencial, lo cual muestra coherencia con los resultados de la simulación para tamaños de muestra pequeñas ($n=32$) y bajo error tipo exponencial.

BIBLIOGRAFÍA

Aigner, D.J.; Lovell, C.A.K.; Schmidt, P. (1977). "Formulation and estimation of stochastic frontier production functions models", *Journal of Econometrics*.

Arriagada, R. (2002). "Diseño de un sistema de medición de desempeño para evaluar la gestión municipal: una propuesta metodológica". Publicación de las Naciones Unidas. Chile.

Bardhan, I. R., Cooper, W. W., Kumbhakar, S C., (1998) "A Simulation Study of Joint Uses of Data Envelopment Analysis and Statistical Regressions for Production Function Estimation and Efficiency Evaluation". *Journal of Productivity Analysis*.

Camisón. C ; Cruz. S. (2008). "La medición del desempeño organizativo desde una perspectiva estratégica: creación de un instrumento de medida". *Revista Europea de Dirección y Economía de la Empresa* vol. 17, núm. 1 pp. 79-102.

Coelli, T.J; Rao, D.S.P.; Battese, G.E. (1998). "An Introduction to Efficiency and Productivity Analysis, Kluwer Academic Publishers". Springer. Boston.

Farrell, M.J. (1957). "The Measurement of Productive Efficiency". *Journal of the Royal Statistical Society, Series A*, 120(3).

Johnson, D. (2000). "Métodos Multivariados Aplicados al Análisis de Datos". International Thomson Editores.

Mutis, H. (2006), "Una aplicación del análisis de frontera estocástica: el caso de hospitales de nivel II en Colombia" *Lecturas Matemáticas*.

Peña, D. (2003). "Análisis de datos multivariantes". Mac Graw Hill.

Rencher, A. (2012). "Methods of Multivariate Analysis". John Wiley & Sons.

Salmi, J. "El desafío de crear universidades de rango mundial". Banco Mundial.

Sveiby, K. (2006). "Building a Knowledge – Based Strategy A System Dynamics Model for Allocating Value Adding".

Wichern Dean W., J. R. A. (2007). "Applied Multivariate Statistical Analysis ". Prentice Hall.

www.scimago.es. Consultado 19 de octubre de 2012