Neiva, 24 de julio de 2020

Señores
CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN
UNIVERSIDAD SURCOLOMBIANA
Ciudad

Los suscritos, Jhon Jairo Cabrera Carrasco con C.C. No. 1075227528 y Augusto Fernando Medina Rivas con C.C. No. 12.234.137. Autores del trabajo de grado titulado: ***Análisis de la eficiencia del recaudo tributario de los municipios del departamento del Huila entre el 2008-2019, mediante la técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica*.** Presentado y aprobado en el año 2020 como requisito para optar al título de Especialista en Estadística. Autorizamos al CENTRO DE INFORMACIÓN Y DOCUMENTACIÓN de la Universidad Surcolombiana para que, con fines académicos, muestre al país y el exterior la producción intelectual de la Universidad Surcolombiana, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios puedan consultar el contenido de este trabajo de grado en los sitios web que administra la Universidad, en bases de datos, repositorio digital, catálogos y en otros sitios web, redes y sistemas de información nacionales e internacionales “open access” y en las redes de información con las cuales tenga convenio la Institución.

- Permita la consulta, la reproducción y préstamo a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato Cd-Rom o digital desde internet, intranet, etc., y en general para cualquier formato conocido o por conocer, dentro de los términos establecidos en la Ley 23 de 1982, Ley 44 de 1993, Decisión Andina 351 de 1993, Decreto 460 de 1995 y demás normas generales sobre la materia.

- Continúo conservando los correspondientes derechos sin modificación o restricción alguna; puesto que, de acuerdo con la legislación colombiana aplicable, el presente es un acuerdo jurídico que en ningún caso conlleva la enajenación del derecho de autor y sus conexos.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, “Los derechos morales sobre el trabajo son propiedad de los autores”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

EL AUTOR/ESTUDIANTE: AUGUSTO FERNANDO MEDINA RIVAS

Firma:

EL AUTOR/ESTUDIANTE: JHON JAIRO CABRERA CARRASCO

Firma: Jhon Jairo Cabrera C.

TÍTULO COMPLETO DEL TRABAJO: ***Análisis de la eficiencia del recaudo tributario de los municipios del departamento del Huila entre el 2008-2019, mediante la técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica.***

AUTOR O AUTORES:

Primero y Segundo Apellido	**Primero y Segundo Nombre**
CABRERA CARRASCO	JHON JAIRO
MEDINA RIVAS	AUGUSTO FERNANDO

DIRECTOR Y CODIRECTOR TESIS:

Primero y Segundo Apellido	Primero y Segundo Nombre
SÁNCHEZ HERNÁNDEZ	ALFONSO

ASESOR (ES):

Primero y Segundo Apellido	Primero y Segundo Nombre

PARA OPTAR AL TÍTULO DE: ESPECIALISTA EN ESTADÍSTICA

FACULTAD: CIENCIAS NATURALES Y EXACTAS

PROGRAMA O POSGRADO: ESPECIALIZACIÓN EN ESTADISTICA

CIUDAD: NEIVA **AÑO DE PRESENTACIÓN:** 2020 **NÚMERO DE PÁGINAS: 56**

TIPO DE ILUSTRACIONES (Marcar con una **X**)**:**

Diagramas___ Fotografías___ Grabaciones en discos___ Ilustraciones en general___ Grabados___ Láminas___ Litografías___ Mapas___ Música impresa___ Planos___ Retratos___ Sin ilustraciones___ Tablas o Cuadros X

SOFTWARE requerido y/o especializado para la lectura del documento: PARA LA LECTURA DEL SCRIPT LENGUAJE R

MATERIAL ANEXO: SCRIPT LENGUAJE R

PREMIO O DISTINCIÓN *(En caso de ser LAUREADAS o Meritoria)*:

PALABRAS CLAVES EN ESPAÑOL E INGLÉS:

<u>**Español**</u>	<u>**Inglés**</u>
1. EFICIENCIA	EFFICIENCY
2. FISCALIDAD	TAXATION
3. ANALISIS ENVOLVENTE DE DATOS	DATA ENVELOPMENT ANALYSIS
4. FRONTERA DE PRODUCCIÓN ESTOCÁSTICA	STOCHASTIC PRODUCTION FRONTIER

RESUMEN DEL CONTENIDO: (Máximo 250 palabras)

El objetivo del trabajo es analizar la eficiencia del recaudo tributario de los municipios del departamento del Huila utilizando la técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica en el periodo 2008 2018.

ABSTRACT: (Máximo 250 palabras)

The objective of the work is to analyze the efficiency of the tax collection of the municipalities of the department of Huila using the Data Envelopment Analysis technique and the Stochastic Frontier Analysis technique in the period 2008-2018.

APROBACION DE LA TESIS

Nombre Jurado: **YINARE CUELLAR LOSADA**

Firma:

Nombre Jurado: **DIEGO MAURICIO ECHEVERRI SUAZA**

Firma:

Análisis de la Eficiencia del Recaudo Tributario de los Municipios del Departamento del Huila entre el 2008-2019, Mediante la Técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica

Jhon Jairo Cabrera
Augusto Fernando Medina

Especialización en Estadística

Facultad de Ciencias

UNIVERSIDAD SURCOLOMBIANA

NEIVA

2020

Análisis de la Eficiencia del Recaudo Tributario de los Municipios del Departamento del Huila entre el 2008-2019, Mediante la Técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica

Jhon Jairo Cabrera
Augusto Fernando Medina

Trabajo de grado para optar al títulode
Especialista en Estadística

Director:
Alfonso Sánchez Hernández
MSc. en Investigación Operativa y Estadística

UNIVERSIDAD SURCOLOMBIANA

FACULTAD DE CIENCIAS

NEIVA
2020

UNIVERSIDAD
SURCOLOMBIANA
NIT: 891180084-2

FACULTAD DE CIENCIAS EXACTAS Y NATURALES

ESPECIALIZACIÓN EN ESTADÍSTICA

CARTA DE ACEPTACIÓN

En calidad de Coordinador del Posgrado Especialización en Estadística, programa reconocido por el Ministerio de Educación Nacional mediante Resolución de Registro Calificado No. 3683 del 2 de marzo de 2018 y adscrito a la Facultad de Ciencias Exactas y Naturales de la Universidad Surcolombiana, me permito informar que el trabajo de investigación titulado: **"ANÁLISIS DE LA EFICIENCIA DEL RECAUDO TRIBUTARIO DE LOS MUNICIPIOS DEL DEPARTAMENTO DEL HUILA ENTRE EL 2008-2019, MEDIANTE LA TÉCNICA DE ANÁLISIS DE LA ENVOLVENTE DE DATOS Y LA TÉCNICA DE ANÁLISIS DE LA FRONTERA ESTOCÁSTICA"** presentado por los estudiantes Jhon Jairo Cabrera Carrasco y Augusto Fernando Medina Rivas; es ACEPTADO como trabajo de grado para optar el título de Especialista en Estadística.

Para constancia se firma en la Ciudad de Neiva, a los veinte (20) días del mes de julio del año 2020.

JOSE MIGUEL CRISTANCHO FIERRO
Coordinador

Sede Central / Av. Pastrana Borrero - Cra. 1
Sede Administrativa / Cra. 5 No. 23 - 40
www.usco.edu.co / Neiva - Huila
PBX: 875 4753
PBX: 875 3686
Línea Gratuita Nacional: 018000 968722

Índice General

Índice de tablas

Lista de Figuras

1. Introducción

La Matemática, como área fundamental de las Ciencias Puras, base del ambiente científico y de muchas aplicaciones en diferentes áreas del conocimiento: Economía, Sociología, Ciencias de la Salud, Ecología e Ingeniería, inducen a temáticas como la Matemática Aplicada y dentro de esta se encuentra la Optimización Matemática. Dentro de esta última ramificación se tiene la Investigación de Operaciones, constituída por un conjunto de técnicas matemáticas, las cuales ayudan a fortalecer y adecuar la toma de decisiones, en cualquier ambiente del mundo que nos rodea.

El Análisis Envolvente de datos (DEA), por su parte, es una herramienta no paramétrica, la cual se fundamenta en la programación lineal, y que trata de medir la *eficiencia relativa* de una serie de unidades económicas, las cuales se desenvuelven en condiciones homogéneas de producción, insumos y entornos, (2001(1), pp. 9) metodología propia de DEA, surge la necesidad de efectuar análisis con apoyo estadístico o estocástico, para obtener *eficiencias relativas* óptimas. por esta razón se tiene el método de Análisis de la Frontera Estocástica (SFA).

Sin dejar de lado la importancia de la Estadística, área de gran impacto con evolución histórica propia y de gran interés para la humanidad, la cual usa como herramienta el análisis de regresión. El presente trabajo tiene como objetivo primordial estudiar la eficiencia del recaudo tributario, en los municipios del departamento del Huila, mediante DEA y SFA.

2. Planteamiento del Problema

Al igual que las empresas privadas, las instituciones públicas utilizan recursos para atender las demandas sociales de bienes y o servicios públicos, es decir, generan un o unos productos a partir de la utilización de unos insumos. Por ejemplo, en el caso de las administraciones locales, una parte importante de los ingresos municipales provienen de la recaudación tributaria, lo cual no sería posible si el municipio no dispusiera de los recursos y el personal especializado para su funcionamiento. La población estará más dispuesta a pagar impuestos si perciben que su esfuerzo económico es compensado con la cantidad y calidad de los bienes y servicios públicos recibidos, por lo tanto, es un deber de las administraciones ser eficientes en el manejo de los recursos públicos entregados; entendiendo la eficiencia como la relación entre productos obtenidos e insumos utilizados. Sin embargo, por diversos motivos, no siempre ocurre así. Un informe del Departamento Nacional de Planeación DNP, (2003), sobre la eficiencia del recaudo tributario municipal, concluye que el 79 % de los municipios son ineficientes. Una situación de ineficiencia en el recaudo tributario municipal se convierte en una oportunidad ante una situación de déficit en las finanzas de los gobiernos municipales; en el informe del DNP (2015) sobre el desempeño fiscal de los departamentos y municipios, estimó el déficit para el año 2015 en un valor equivalente al 0.85 % del Producto Interno Bruto nacional. Por tanto, estimar la eficiencia del recaudo tributario municipal es importante para las finanzas públicas municipales. El trabajo de investigación que se propone pretende analizar la eficiencia del recaudo tributario de los municipios del departamento del Huila en el periodo 2008 - 2019.

¿Cuál es la eficiencia del recaudo tributario de los municipios del departamento del Huila en el periodo 2008 - 2019?

3. Referentes Teóricos

3.1. Caracterización de los Sistemas Impositivos en los Municipios Huilenses. 2008-2019

Los tributos ocupan un lugar importante en la organización y funcionamiento de los municipios. Al ser una de las fuentes principales del presupuesto municipal, influyen en las decisiones económicas de los habitantes. Un sistema tributario acorde a las características socioeconómicas del municipio coadyuva al desarrollo del mismo; por el contrario, un sistema tributario distorsionado priva al municipio de los recursos necesarios para atender las demandas de bienes y servicios de la comunidad. La asignación eficiente de los recursos económicos por parte de los agentes privados, evita el caos y la perturbación. Son numerosos los estudios sobre los sistemas impositivos de los países amparados por organizaciones internacionales como el Fondo Monetario Internacional y el Banco Mundial. A nivel nacional, ha sido el Departamento Nacional de Planeación, DNP, la institución oficial más comprometida con el estudio del sistema impositivo, y sus efectos en la eficiencia, en la equidad y la competitividad de la economía. Sin embargo, la mayoría de los estudios no descienden hasta el nivel local, se concentran en las principales ciudades del país o departamentos. Relativamente, son pocos los estudios de los sistemas tributarios municipales y si consideramos los que existen, éstos están concentrados en los municipios más desarrollados del país. No es de extrañar entonces que no se encuentren muchos estudios de los sistemas tributarios de los municipios huilenses. Tal vez, la principal dificultad en la apropiación de los estudios de los sistemas impositivos al nivel municipal, es originada por la escasez y la precariedad de la información estadística de que se dispone. No es fácil conseguir información suficiente, amplia y fiable, sobre las características físicas, económicas y sociales de los municipios, por ejemplo, algo tan básico como la consecución del estatuto tributario actualizado del municipio, se convierte en todo un desafío. Es evidente el atraso en materia estadística que tiene el país. Con esa limitante, este trabajo se nutre de diversas fuentes de información principalmente: las alcaldías municipales, el Consolidador de Hacienda e Información Pública CHIP del Ministerio de Hacienda, el Instituto Geográfico Agustín Codazzi IGAC, el Departamento Administrativo Nacional de Estadística DANE y el Departamento Nacional de Planeación DNP; con las cuales se intenta hacer un acercamiento a algunas de las características de los sistemas impositivos de los municipios huilenses. Este es un trabajo de alcance descriptivo, que pretende ser la primera de una secuencia de trabajos de caracterización, análisis y diagnóstico, de los sistemas impositivos de los municipios del departamento del Huila, con el objetivo principal de aportar a la reflexión política de los alcaldes y concejales y de esta forma, enriquecer la función pública.

3.2. El Sistema Impositivo Municipal

Las principales reformas que definirían los nuevos derechos y deberes de las entidades territoriales en materia fiscal se presentaron en la década de los ochenta y comienzos de los noventa, expresadas, entre otras, en la Ley 14 de 1983, Ley 12 de 1986, Ley 44 de 1990, y la inclusión en la Constitución Política de 1991. Este hecho histórico, en un país caracterizado por una fuerte centralización, fue una consecuencia, entre otras razones, de los problemas de ineficiencia del gasto público y el deterioro de las finanzas públicas. (Iregui, A., Ramos, J. y Saavedra, L. 2001, (2)). Respecto al primer punto, la eficiencia del gasto público, cada vez existe más consenso sobre los beneficios de la descentralización en materia del gasto y su control. Permitir que fuesen los propios gobiernos territoriales y su población los que decidieran la mejor manera de utilizar los recursos y, simultáneamente, diseñar los mecanismos de control y veeduría, supondría una mayor eficiencia de los recursos públicos. (Restrepo, 2012 (3)). En cuanto a las finanzas públicas, su deterioro se acentúo con la crisis del petróleo ocurrida a comienzos de la década de los 70. Al multiplicarse por cinco el precio del barril de petróleo, como consecuencia del control de la oferta a partir de la creación de la Organización de Países Exportadores de Petróleo OPEP, el mundo presenció la irrupción de un fenómeno económico desconocido hasta ese momento por la literatura económica, y que luego fue conocido como estanflación, consistente en procesos inflacionarios importantes y desaceleración económica, tuvo efectos colaterales en la economía de varios países latinoamericanos que vieron una fuerte caída de su producto interno bruto (PIB) y un aumento exagerado en el servicio de la deuda externa. Este periodo, década de los ochenta, llamado "la Década Perdida", se caracterizó por la caída del PIB y de los ingresos fiscales, y las declaratorias de impago de la deuda externa. (Ocampo, 1994 (4)). El proceso de descentralización que inició entonces el país desde el siglo pasado y que se acentuó e institucionalizó en la Constitución política de 1991, transfirió a las entidades territoriales una serie de obligaciones y derechos en materia administrativa y fiscal. En el caso fiscal, los entes territoriales fueron autorizados, con ciertas limitaciones, para administrar y recaudar algunos impuestos antes de competencia del gobierno central. (Iregui, Ramos y Saavedra, 2001(2)). La Ley aumentó la autonomía de los entes territoriales transfiriéndoles derechos y deberes. La concesión de autonomía fiscal, junto con la participación en las rentas del Estado, les permitió a los gobiernos locales atender las demandas de bienes y servicios públicos de la población que antes eran atendidas por el gobierno central. Sin embargo, el proceso no ha sido sincrónico, en algunos casos se transfirieron funciones sin el debido respaldo presupuestal y en otros casos, los recursos se transfirieron más no así las funciones. (Iregui, Ramos y Saavedra, 2001(2)). La mayor autonomía en materia fiscal cedida a los entes territoriales ha ido acompañada de una serie de limitaciones contempladas en las diferentes leyes y la propia Constitución Política. En su mayor proporción, estas restricciones son de carácter presupuestario, destinación de las partidas, control del gasto y potestad tributaria. Son estas condiciones las que le permiten a Iregui, Ramos y Saavedra (2001) aseverar que el proceso de descentralización en Colombia ha sido esencialmente una desconcentración en la ejecución del gasto público nacional, antes que una mayor autonomía de los entes locales. La trascendencia que tiene el proceso de descentralización administrativa y fiscal es tal vez, junto con la elección popular de mandatarios locales, el hecho político de mayor relevancia en la historia reciente de las entidades territoriales. Una de las razones que explicaría la atención de diferentes instituciones académicas de carácter público y privado en estudiar la evolución y desempeño de este proceso, especialmente el concerniente a la descentralización fiscal. El esquema tributario municipal actual es el resultado de un largo e irregular proceso que viene desde la colonia, con momentos de avances importantes y otros de estancamiento, e incluso, retroceso. De la mano del proceso de descentralización de las finanzas públicas iniciado a mediados de la década de los 80 y posteriormente la Constitución Política de 1991, se le dio un impulso importante a la capacidad tributaria de los entes territoriales, aunque mantuvo los lineamientos generales en materia impositiva. Se les reconoce a los departamentos y municipios su derecho a establecer sus propios tributos y fijar los tipos impositivos dentro de los parámetros establecidos por la ley. (Fino, 2012(5)). A continuación se describen dos de los impuestos más significativos por su antigüedad, por su objeto y por el aporte económico dentro de la estructura tributaria de los municipios, son el Impuesto de Industria, Comercio y Complementario Avisos y Tableros ICA y el Impuesto Predial Unificado IPU.

3.3. El Impuesto de Industria, Comercio y Complementario Avisos y Tableros ICA

Tal vez el registro histórico más antiguo que se tiene de la existencia de un impuesto de este tipo se encuentra en la Edad Media. Una persona que se encontrara dentro de los dominios del señor feudal y realizara algún tipo de actividad comercial o elaboración de algún tipo de bien destinado para la venta, estaba en el deber de pagarle un porcentaje del fruto de su venta (Mario Posada García-Peña). Este impuesto llegó a Colombia probablemente de la mano de la conquista española en donde existía para ese tiempo un tributo conocido como "Servicio ordinario de monedas" el cual consistía en un impuesto sobre las ganancias de las fincas y el comercio. (Ibíd.) La herencia colonial en materia impositiva y su propia evolución en España devino en el año de 1826 en la creación en el país de una contribución industrial que otorgaba una licencia de funcionamiento para quienes ejercieran actividades industriales, comerciales, artes y oficios. Esta contribución sería recogida de manera formal en el art. 1 literal f de la Ley 97 de 1913 que otorgaba atribuciones especiales en materia impositiva al Consejo municipal de Bogotá. Textualmente dice el literal: "Impuesto de patentes sobre carruajes de todas clases y vehículos en general, incluidos los automóviles y velocípedos; sobre establecimientos industriales en que se usen máquinas de vapor o de electricidad, gas y gasolina; sobre clubs, teatros, cafés cantantes, cinematógrafos, billares, circos, juegos y diversiones de cualquier clase, casas de préstamo y empeño, pesebreras, establos, corrales, depósitos, almacenes y tiendas de expendio de cualquier clase" (Posada, 2008).

En poco menos de dos años el Congreso de la República se pronunció de nuevo y por medio de la Ley 84 de 1915, art. 1 literal a., extendió el derecho concedido al Consejo municipal de Bogotá a través de la Ley 97 de 1913 al resto de consejos del país, en cualquier previa autorización de sus respectivas asambleas departamentales. Finalmente, el 6 de julio de 1983 se expidió la que sería la Ley marco del Impuesto de Industria y Comercio, Ley 14 de 1983 "Por la cual se fortalecen los fiscos de las entidades territoriales y se dictan otras disposiciones." Los artículos 32 – 48 abordan el impuesto iniciando con su definición hasta la descripción particular de la base impositiva para la cuantificación del impuesto al sector financiero. El art. 32 empieza por definirlo: "El Impuesto de Industria y Comercio recaerá, en cuanto a materia imponible, sobre todas las actividades comerciales, industriales y de servicio que ejerzan o realicen en las respectivas jurisdicciones municipales, directa o indirectamente, por personas naturales, jurídicas o por sociedades de hecho, ya sea que se cumplan en forma permanente u ocasional, en inmuebles determinados, con establecimientos de comercio o sin ellos." Y a continuación, el art. 33, se refiere a la base gravable del impuesto: "El Impuesto de Industria y Comercio se liquidará sobre el promedio mensual de ingresos brutos del año anterior, expresados en moneda nacional y obtenidos por las personas y sociedades de hecho indicadas en el artículo anterior, con exclusión de: devoluciones -ingresos provenientes de venta de activos fijos y de exportaciones- recaudo de impuestos de aquellos productos fijos cuyo precio esté regulado por el Estado y precepción de subsidios." En el mismo artículo, 33, se establecen los límites de las tarifas posibles que podrán aplicar los Consejos Municipales: Del dos al siete por mil (2-7 x 1000) mensual para actividades industriales y del dos al diez por mil (2-10 x 1000) mensual para actividades comerciales y de servicios. Estas tarifas no invalidan aquellas que hayan sido fijadas por los municipios con anterioridad y que superen los límites establecidos. Posteriormente, el Decreto 1333 de 1986 que da vida al Código de Régimen Municipal, incluye en su articulado, arts. 195-213, la normativa expedida del Impuesto de Industria y Comercio procedente de la Ley 14 de 1983.

Algunas características del impuesto son: en cuanto al hecho generador, todas aquellas actividades industriales, comerciales o de servicios, presuntas de rendimientos económicos en cabeza de un titular quien actúa como el sujeto pasivo del gravamen. Respecto a la territorialidad del impuesto, el art. 32 de la Ley 14 de 1983, antes citado, indica que todas aquellas actividades económicas que se ejerzan o realicen en las respectivas jurisdicciones municipales. Los sujetos de la obligación tributaria principal son definidos en la norma como "... por personas naturales y jurídicas o por sociedades de hecho" (Ley 14/83, art.32 (6)).Por último, la base gravable será la constituida por el "... promedio mensual de ingresos brutos del año anterior con exclusión de: devoluciones -ingresos provenientes de venta de activos fijos y de exportaciones- recaudo de impuestos de aquellos productos fijos cuyo precio esté regulado por el Estado y precepción de subsidios (Ley 14/83, art.33 (6))."

3.4. El Impuesto Predial Unificado IPU

Al parecer, el impuesto predial tendría más de seis mil años de existencia, desde cuando se elaboró el primer catastro con fines impositivos, que según algunos autores, ocurrió entre los caldeos hacia el año 4.000 A.C. y que recogía los linderos de los predios. (Camacho, 2011 (7)) En el caso colombiano, el impuesto predial existe prácticamente desde la llegada de los españoles que concuerda con la fecha en el que el Rey "Felipe V implanta el sistema de la única para los territorios de la Corona de Aragón" (Camacho, ibíd., como se citó en Parra, 2002). El impuesto era de plena autonomía de municipios y departamentos quienes además de organizar su propio catastro, fijaban los tipos impositivos y se encargaban de su recaudo. A finales del siglo XIX, en el gobierno del presidente Rafael Núñez se expidió la Ley 48 de 1887, la cual le daba al impuesto predial el carácter de impuesto departamental. Luego, a principios del siglo XX, se expediría la Ley 20 de 1908 la cual le otorgó al impuesto su carácter municipal. Las leyes siguientes, Ley 14 de 1983, 333 de 1986 y la Ley 44 de 1990, definieron respectivamente: la potestad de los Consejos Municipales para establecer los tipos impositivos dentro de los límites fijados por el gobierno; incorporó el impuesto al Código municipal y; compiló el Impuesto Predial, el impuesto de parques y arborización, el impuesto de estratificación socioeconómica y, la sobretasa de levantamiento catastral, en un impuesto que denominó el Impuesto Predial Unificado. Las leyes más recientes corresponden a la Ley 1430 de 2010, respecto al control tributario y competitividad, en su artículo 54 conceptúa sobre los sujetos pasivos de los impuestos territoriales; la Ley 1450 de 2011, por medio de la cual se aprueba el Plan Nacional de Desarrollo 2010-2014, en su artículo 23 incrementa la tarifa mínima del Impuesto Predial Unificado modificando en esa parte la Ley 44 de 1990. El profesor Camacho (2011) define el impuesto predial en general como "... un tributo real, cedular y directo, que recae sobre los predios o bienes raíces que se encuentran ubicados en áreas urbanas o rurales, con o sin edificaciones, en el perímetro del respectivo ente territorial" (p.22). Por cedular se entiende, que el impuesto recae sobre el bien en cuestión sin tener en cuenta la capacidad adquisitiva del contribuyente y por directo, que aplica sobre quien figura como sujeto pasivo del gravamen (propietario, usufructuario, poseedor, etcétera). El capítulo II del Título X del Código de Régimen Municipal, Decreto 1333 de 1986, describe los diferentes impuestos que conforman los ingresos tributarios de los municipios y señala los de posible creación. Entre los impuestos especificados en el Decreto figuran:

- El impuesto predial. Como su nombre lo indica, es un impuesto que grava la propiedad inmueble; con una tarifa impositiva fijada por el respectivo Consejo Municipal que oscila entre el 1 por mil y el 16 por mil del avalúo catastral del inmueble a excepción, de aquellos predios urbanizables no urbanizados o urbanizados no edificados cuya tarifa podrá ser mayor sin exceder el 33 por mil.

  También, el artículo 4, Ley 1450 de 2011, establece que a partir del 2012 la tarifa mínima será del 3 por mil, en el 2013 del 4 por mil y en el 2014 del 5 por mil. A excepción, de los inmuebles urbanos con destino habitacional o rurales con destino económico agropecuario que pertenezcan a los estratos 1, 2 y 3, cuyo precio sea inferior a los ciento treinta y cinco salarios mínimos legales mensuales vigentes (135 SMLMV).

  Los consejos fijaran las tarifas del impuesto de acuerdo a los siguientes factores:

- Estrato Socioeconómicos
- Los usos del suelo en el sector urbano
- La antigüedad de la formación o actualización del catastro
- El rango de área
- El avalúo catastral

3.5. El Sistema Impositivo de los municipios Huilenses

Los sistemas impositivos de los municipios están reglamentados por los Acuerdos expedidos por los diferentes Concejos municipales, quienes por ley son los encargados de aprobar el estatuto tributario del municipio. Y aunque se les reconoce autonomía tributaria a los entes locales, la normativa aprobada está sujeta a los lineamientos del Congreso lo que impide que los municipios gocen de las potestades requeridas para crear nuevos impuestos y las tarifas deben encontrarse dentro de los parámetros legales establecidos. De ahí la similitud en la estructura tributaria de los municipios. Con pequeñas diferencias, comparten tributos y tarifas municipales similares; se distinguen en variaciones de las nominaciones de algunos de los tributos y en las tarifas que aplican, aunque las diferencias en general son mínimas. Entre impuestos, tasas y contribuciones, los municipios huilenses cuentan en promedio entre 20 y 25 tributos, siendo los más comunes: Impuesto de Industria y Comercio, Impuesto Predial, Impuesto sobre vehículos automotores, Impuesto a la publicidad exterior, Impuesto de delineación, Estampillas pro cultura, pro adulto mayor, pro deporte, y por supuesto, la Sobretasa a la gasolina.

3.6. Antecedentes

La construcción de los presentes antecedentes está conformada por dos técnicas utilizadas para medir la eficiencia: Análisis de la Envolvente de Datos (DEA) y Aproximación de fronteras Estocásticas. Existen pocas investigaciones sobre eficiencia en las entidades gubernamentales por ello hemos incursionado esta investigación y se ha llegado a las siguientes conclusiones.

El pionero en la medición de la eficiencia fue el profesor de la Universidad de Cambridge, Michael James Farrell, con su celebre trabajo, "The Measurement of Productive Efficiency", Medición de la Eficiencia Productiva, en 1957.

Una de las consecuencias del trabajo de Farrell, fue el trabajo de los profesores Abraham Charnes, William.W. Cooper y Edwardo Rhodes, "Measuring the Efficiency of Decision Making Units", Medición de la Eficiencia de las Unidades de Toma de Decisiones, en 1978. En el cual introdujeron el concepto de "Análisis Envolvente de Datos" (DEA por sus siglas en inglés).

En cuanto a los modelos de producción de frontera estocástica, (SFA por sus siglas en inglés) fueron propuestos por los profesores Denis Aigner, Knox Lovell y Peter Schmidt, su famoso trabajo "Formulation and estimation of stochastic frontier production function models", Formulación y estimación de modelos de función de producción de frontera estocástica, en 1977.

Los métodos de medición de la eficiencia como DEA y SFP se popularizaron desde finales del siglo pasado a raíz de las múltiples privatizaciones de empresas de servicios públicos y su necesidad de regularlas. Famosos los trabajos de los profesores Timothy Coelly y George Battese.

(Fernández Santos Flórez López, 2006 (8)) en su trabajo, Aplicación del Modelo DEA en la Gestión Pública, afirma que “la eficiencia global de una entidad no puede aproximarse mediante la utilización de indicadores de eficiencia parcial, dado que éstos proporcionan información aislada y no considera las interrelaciones entre las variables”.

(Ordoñez Santo, 2012 (9)) habla de la descentralización en Colombia y afirma que, desde la constitución política de Colombia de 1991, mediante actos legislativos el gobierno ha realizado cambios en su estructura que no fortalece las bases tributarias territoriales, y aunque el Departamento Nacional de Planeación aplica la técnica del análisis de la envolvente de datos (DEA), no hay estudios que garanticen la aplicación de dicha técnica.

(Valencia-Cárdenas Restrepo-Morales, 2016 (10)) comentan en su trabajo nombrado como “Evolución de la Gestión Financiera usando Variables Latentes en Modelos Estocásticos de Fronteras Eficientes” cuya base de datos fue tomada de Benchmark que proporciona variables financieras por empresa, la cual se realizó una análisis descriptivo en 109 empresas que por diferencias significativas se aplicó a 105 empresas del sector financiero en Colombia; afirman que los indicadores tradicionales financieros no son factibles para dimensionar la calidad del Magnament, para esto estimaron indicadores mediante el modelo

“CAMEL, (C, Capital Adequacy; A, Assets Quality; M, Management; E, Earnings; L, Liquidity) como variables latentes, y estimadar a partir de una técnica de Análisis Multivariado de Datos conocida como Análisis del Factor Confirmatorio (CFA)” posteriormente aplican DEA que arroja una eficiencia de 0.0265 y SFA con una eficiencia de 0.38641 para analizar la calidad del Magnament de un sector financiero Colombiano, por ultimo mediante un análisis Clúster comparan las clasificaciones de los grupos empresariales mediante niveles bajo y alto en relación a la eficiencia, estos niveles comprueban una relación directa con las rentabilidades, activos, ROE y ROA los crecimientos en activos entre otras, la cual agrupan el 68 % de las empresas y existe una probabilidad de mejorar la eficiencia en el sector. No es muy claro el análisis en DEA dado que no explica si la eficiencia es bajo CRS o VRS. (Efficiency et al., n.d.) Damara S. Alvarez Gonzales y Antonio Ruiz Porras de la Universidad de Guadalajara, aplicaron DEA, SFA por máxima verosimilitud y el índice de Malmquist (IPM) en los bancos de desarrollo mexicano desde el 2011 a 2017 las mediciones de eficiencia se realizaron en base a orientaciones input y output, los rankings de eficiencia fueron consistentes independientemente del método utilizado. El análisis de los resultados en DEA, mediciones determinísticas evidencia la eficiencia relativa puesto que las DMU (Bancos) se comparan entre ellos mismos, los puntajes obtenidos mediante el enfoque de intermediación son superiores a los de enfoque de producción, el resultado obtenido indica que hay ineficiencia en los bancos. Al analizar la frontera estocástica SFA, se puede apreciar que los modelos calculados fueron los siguientes: half-normal, normal exponencial y normal truncado, y mediante máxima verosimilitud se indica el modelo con mayor bondad de ajuste con que se estiman los índices de eficiencia, obteniendo como resultados similares a los de DEA y evidenciando bajo esta técnica la ineficiencia de los bancos puede ser por factores endógenos y exógenos. Dado que nos interesa las técnicas aplicadas en DEA las DMU utilizadas fueron los bancos Banobras, Nafin, Banjercito, Bancomext, Bansefi y SHF bajo el enfoque de intermediación el banco mas eficiente es Banobras y el mas ineficiente es SHF, mediante el enfoque de producción este resultado fue inverso la cual concuerda con la literatura, en el enfoque de producción los bancos mas ineficientes fueron Bancomext bajo orientación input y Bansefi bajo la orientación output; y en SFA con el enfoque de intermediación, Nafin y Banobras resultaron mas eficientes y en el enfoque de producción SHF se mantuvo como el banco mas eficiente. Por último, podemos evidenciar que las dos técnicas nos arrojan resultados muy consistentes dado que esto depende de la calidad de los datos.

(Ramiro Rodríguez Sperat, María Pía Brugiafreddo y Erica Raña, 2017 (11)), aplicaron DEA Y SFA para estudiar la eficiencia técnica en la agricultura familiar de la provincia de Santiago del Estero (Argentina), obteniendo resultados con índices de eficiencia técnica promedio bajo rendimiento de escala constantes (CRS) de 0.567 y bajo rendimiento de escala variables (VRS) de 0.693 y aplicando SFA 0.745, al analizar los resultados se evidencia que DEA (VRS) y SFA no tienen una diferencia significativa, asegura que esto se debe al tipo de información que se tiene, lo que desea el investigador y como se aplican las técnicas en cada caso en particular.

3.7. Análisis Envolvente de Datos DEA

Siguiendo a Soto y Arenas (2010, (12), pp. 11), el Análisis Envolvente de Datos, que en inglés se escribe Data Envelopment Analysis, de aquí su contracción DEA, ha sido utilizado en los últimos años en la valoración de la eficiencia de diferentes procesos y entidades gubernamentales, por ejemplo: Municipios, Policía Nacional, Contraloría y el Departamento Nacional de Planeación, entre muchas otras.

El cálculo de la eficiencia, siguiendo (4), es:

$$EFICIENCIA = \frac{PRODUCTOS}{ENTRADAS}$$

Aparentemente es una definición simple, si se tiene en cuenta que en un proceso o entidad entran en juego una gran cantidad de entradas y salidas, las cuales están relacionadas a un sin número de recursos, actividades y factores, lo cual hace un tanto complejo el análisis matemático de la situación o entidad.

Se consideran más eficientes aquellas entidades que usan convenientemente sus entradas, obteniendo mejores salidas. En una empresa de producción, por ejemplo, al momento de obtener sus salidas o productos, le convendría cuestionarse acerca de la importancia ó el peso de algunas de sus salidas, lo mismo ocurriría con sus insumos ó entradas. Por esta razón el Análisis Envolvente de Datos permite resolver problemas de comparación de eficiencias, cuando los procesos involucran distintos tipos de entradas (insumos) y salidas (productos).

Siguiendo la notación de Charnes, Cooper y Ferguson (1955), el DEA calcula la eficiencia utilizando la siguiente notación:

$$h_{j_0} = \frac{\sum_{r=1}^{m} u_r\, y_{rj_0}}{\sum_{i=1}^{n} v_i\, x_{ij_0}}; \quad j = 1, \cdots, n \tag{1}$$

El subíndice j_0 representa la unidad de decisión a la que se le está calculando la eficiencia, u_r representa el peso del producto, mientras que v_r representa el peso del insumo o entrada, h_{j_0} es la eficiencia de la unidad de decisión que se está calculando.

La eficiencia h_{j_0} es utilizada como función objetivo con restricciones en un problema de programación lineal:

$$\begin{aligned} \frac{\sum_{r=1}^{m} u_r\, y_{rj_0}}{\sum_{i=1}^{n} v_i\, x_{ij_0}} &\leq 1 \\ u_r,\, v_i &\geq \epsilon \end{aligned} \tag{2}$$

el parámetro de perturbación ϵ permite que los pesos tanto de insumos como de productos sean mayores de 0.

Es evidente que la ecuación (2) no representa un modelo lineal, por tanto se hace necesario realizar modificaciones para que se pueda procesar como un problema de programación lineal. existen dos casos:

1. Tomar el numerador y considerar constante el denominador.

2. Tomar el denominador y considerar el numerador como constante.

Para el primer caso el modelo presentado en las ecuaciones (1) y (2) es:

$$max \qquad h_0 = \sum_{r=1}^{m} u_r\, y_{rj_0} \tag{3}$$

$$\begin{aligned} sujeto\ a: \quad & \textstyle\sum_{i=1}^{n} v_i\, x_{ij_0} = 100 & (4) \\ \sum_{r=1}^{m} u_r\, y_{rj_0} - \sum_{i=1}^{n} v_i\, x_{ij_0} \quad & \leq 0 & (5) \\ u_r\, v_i \quad & \geq \epsilon & (6) \end{aligned}$$

La eficiencia de cada unidad de decisión se obtiene cuando se resuelve el problema de programación lineal de las ecuaciones $(3)-(6)$, además se garantizan los pesos u_r y v_i de los productos e insumos respectivamente, conforme a la conveniencia de cada DMU (Decision Making Unit). Algunas DMU tendrán eficiencias del 100 %, las cuales se denominan DMU's de frontera. Es decir, son las mejores en la comparación relativa. El DEA fija en la frontera algunas DMU's, para evitar que se ubiquen en la frontera demasiadas.

3.7.1. Productividad

En el año 1883, Littre (citado por Soto etal.) define *productividad* como la "facultad de producir", mientras que en la primera década del siglo XX, Early presenta una de las definiciones más citadas en el día de hoy: "Relación entre la producción y los medios para lograrla".

Roa (2003, (13), pp. 70) presenta la expresión matemática introducida por Farrell (1957, (14)), el cual reduce la expresión al cociente:

$$Productividad = \frac{Produccin\ Creada}{Recurso\ Consumido} = \frac{Salida}{Entrada} \tag{7}$$

A la unidad productiva, se le añade el calificativo de **decisora**, Soto et.al (2010, (12), pp. 15), la cual aparece en la literatura anglosajona como *Decision Making Unity*. El término DMU fué acuñado por Charnes et.al (1978)[1] en su importante trabajo acerca de DEA.

3.7.2. Eficiencia Relativa

Según Roa (2003, (13), pp. 73, 74), la eficiencia relativa se define como:

$$Eficiencia_j = \frac{Productividad_j}{Productividad_0} = \frac{Salidas\ Virtuales_j / Entradas\ Virtuales_j}{Salidas\ Virtuales_0 / Entradas\ Virtuales_0} \tag{8}$$

así la eficiencia relativa es:

$$Eficiencia_j = \frac{\sum_{r=1}^{s} u_{ri}\, y_{rj} / \sum_{i=1}^{m} v_{ij}\, y_{ij}}{\sum_{r=1}^{s} u_{ri}\, y_{rj} / \sum_{i=1}^{m} v_{ij}\, y_{ij}} \tag{9}$$

Es importante resaltar que existen infinitos pesos que dan la misma eficiencia. Con los pares de pesos v_{ij} y u_{rj} y un múltiplo de ellos, mediante el uso de los números reales α y β, un múltiplo de ellos: $\alpha * v_{ij}$ y $\beta * u_{rj}$ se obtiene la misma eficiencia.

3.7.3. Eficiencia de Pareto

1. *Con orientación a las entradas:* Una DMU se dice Pareto Eficiente si no es posible disminuír ninguno de sus niveles de entrada sin tener que

[1] Charnes, A.; Cooper, W.W. and Rhodes, E. Measuring efficiency of decision Making units. European Journal of Operational research. 2, 429-444.

incrementar al menos uno de sus otros niveles de entrada o disminuír al menos uno de sus niveles de salida. En términos matemáticos: Si y_{rj} con $r = 1, 2, , s$ son los niveles de salida de la DMU$_j$ y x_{ij} con $i = 1, 2, \cdots, m$ los niveles de entrada que ella usa. Una DMU$_{j_0}$ es Pareto Eficiente si no existe una DMU $j \neq j_0$ tal que $x_{i'j} < x_{i'j_0}$ para algún i' y $x_{ij} < x_{ij_0}$, para todo r.

2. *Con orientación a las salidas:* de manera similar, una DMU$_j$ es Pareto Eficiente si no es posible aumentar ninguno de sus niveles de salida, sin tener que disminuír al menos uno de sus otros niveles de salida o aumentar al menos uno de sus niveles de entrada. Sean y_{rj} con $r = 1, \cdots, s$ los niveles de salida alcanzados por la DMU$_j$ y X_{ij} con $i = 1, \cdots, m$ los niveles de recurso o entrada que ella usa. Una DMU$_{j0}$ es Pareto Eficiente si no existe una DMU $j \neq j_0$ tal que $y_{rj} > y_{r'j_0}$ para algún r' y $y_{rj} \geq y_{rj_0}$, para todo $r \neq r'$ mientras que $x_{ij} \leq x_{ij_0}$, para todo i.

La medida de eficiencia en las salidas refleja hasta que punto los niveles de las salidas de la DMU que está siendo considerada, pueden aumentarse a través de las mejoras en el desempeño sin utilización de recursos adicionales, mientras se mantiene la mezcla de salidas (expansión radial de las salidas). La medida de eficiencia en las entradas refleja hasta que punto pueden disminuirse los niveles de las entradas de la DMU bajo consideración, a través de un mejor desempeño sin reducción en las salidas mientras se mantiene su mezcla de entradas. La mezcla de entradas de una DMU se refleja en la proporción en la que se encuentran sus niveles unos con respecto a los otros. La mezcla en las salidas, se define de manera similar.

3.8. Fronteras Paramétricas Determinísticas

Son modelos en los cuales se minimiza la diferencia existente entre las observaciones y las predicciones obtenidas de una forma funcional, por ejemplo una función de producción, costos, etc. considérese por ejemplo la función de producción de Cobb-Douglas ((15), pp. 44).

$$
\begin{aligned}
Y_i = {} & e^{f(x_{i1},x_{i2},\cdots,x_{ik};\beta)}\, e^{u_i} \\
log(Y_i) = {} & \quad \beta_0 + \mathbf{X}\,\beta + u_i
\end{aligned}
\tag{10}
$$

Las estimaciones de fronteras determinísticas utilizan un término de error, lo cual implica que es posible definir de manera exacta la máxima cantidad de producción, dados los factores. Por tanto el nivel de producción observado es el producto máximo mas un término de ineficiencia.

3.8.1. Estimación por Mínimos Cuadrados Corregidos

Cuando se calculan los parámetros, resulta importante efectuar este procedimiento por medio de Mínimos Cuadrados Corregidos Clásicos, realizar los dos pasos necesarios para obtener una estimación consistente para la constante del modelo. una vez hecho esto, se debe descomponer el error compuesto para obtener una estimación de la ineficiencia de la empresa.

El método MCC es el empleado para estimar *fronteras paramétricas determinísticas*

$$lnY_i = \underbrace{[\beta_0 + E(u_i)]}_{\alpha^*} + X\beta + \underbrace{v_i + [u_i - E(u_i)]}_{\epsilon_i^*} \tag{11}$$

En la ecuación anterior el error tiene esperanza nula y se pueden aplicar Mínimos Cuadrados Ordinarios para la estimación de los parámetros.

El segundo paso del proceso de estimación, permite obtener β_0, ${\sigma_u}^2$ y ${\sigma_v}^2$. Además se debe verificar algún supuesto distribucional acerca de u_i. si se supone como distribución para u_i una Media Normal, los momentos de orden 2 y 3 son:

$$E({\epsilon_i}^2) = \frac{\pi - 2}{\pi} \quad y \quad E({\epsilon_i}^3) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left(1 - \frac{4}{\pi}\right) {\sigma_u}^3 \tag{12}$$

No es dificil observar que ${\epsilon_i}^*$ posee los mismos momentos dos y tres que ϵ_i, gracias a que $E(u_i)$ es constante. Si se utilizan los residuos de MCO, se estiman las varianzas ${\sigma_u}^2$ y ${\sigma_v}^2$. Una vez hecho esto se obtienen el parámetros:

$$\hat{\beta_0} = \hat{\alpha^*} - \hat{E}(u_i) \tag{13}$$

Cabe anotar que la esperanza del error multiplicativo de Cobb-Douglas es 1 y no cero.

3.9. Fronteras Paramétricas Estocásticas

Fueron propuestas en forma independiente por Meeusen y Broeck (1977, (15)), Aigner et.al (1977, (16)) y por Battese et.al (1977, (17)). Los anteriores autores emplearon fornteras paramétricas estocásticas en funciones de producción.

En este tipo de problemas, las desviaciones con respecto a la frontera pueden llegar a no estar totalmente bajo el control de la empresa. Meeusen et.al proponen que las desviaciones respecto a la frontera en una unidad productiva, se pueden deber a que la frontera en sí misma es estocástica y puede deberse a errores en las mediciones de las variables. estas fronteras emplean un error aditivo, el cual puede tener distribuciones de colas pesadas tales como la t-Student, Normal Truncada, Exponencial o Gamma.

3.9.1. Estimación de Máxima Verosimilitud

Este método requiere la maximización numérica de la función de verosimilitud y se requieren más cálculos que el método de los mínimos cuadrados. Afortunadamente han aparecido muchos algoritmos en los últimos años, que van acompañados de Software, que hacen mucho más sencillos los cálculos que hace algunos años. Para estimar la frontera de eficiencia, se necesitan algunos supuestos sobre las componentes aleatorias, desde sus primeros momentos. Se requiere la función de densidad de $\epsilon_i = v_i + u_i$. La función de densidad surge desde una Skew-Normal, Azzalini etal, (35).

$$f(\epsilon) = \frac{2}{\sigma}\,\phi\left(\frac{\epsilon}{\sigma}\right)\,\Phi\left(\frac{\epsilon\,\lambda}{\sigma}\right) \tag{14}$$

las funciones $\phi(\cdot)$ y $\Phi(\cdot)$ representan las funciones de densidad y distribución de una normal estándar. El término $\sigma^2 = {\sigma_u}^2 + {\sigma_u}^2$ representa la varianza total, además $\lambda = \frac{\sigma_u}{\sigma}$. Battese etal (1977, (17)) sugieren la reparametrización de la función de verosimilitud mediante

$$\sigma^2 = {\sigma_v}^2 + {\sigma_u}^2 \quad y \quad \gamma = \frac{{\sigma_u}^2}{{\sigma_u}^2 + {\sigma_v}^2} = \frac{{\sigma_u}^2}{\sigma^2} \tag{15}$$

El parámetro anterior permite justificar la relación entre la varianza de la normal, que es la procedencia de u_i y la varianza total, lo cual justifica que el parámetro pueda oscilar entre 0 y 1. La función logaritmo de verosimilitud para $f(\epsilon)$ es:

$$l = n\,\frac{2}{\sigma} - \frac{n}{2}\,ln(2\pi) - n\,ln(\sigma) - \frac{1}{2}\sum_{i01}^{n} {\epsilon_i}^2 + \sum_{i=1}^{n} ln\Phi\left(\frac{\epsilon_i\,\lambda}{\sigma}\right) \tag{16}$$

Finalmente la verosimilitud es la máxima expresión de (16), reemplazando λ por $\sqrt{\frac{\gamma}{1-\gamma}}$.

4. Regresión Lineal Simple y Múltiple

Un problema de regresión surge ante la necesidad de explicar una variable dependiente Y, basada en una serie de variables explicativas $X_1, X_2, \cdots, X_p$. A las variables explicativas se les denomina variables independientes, covariables, variables concomitantes y en casos muy específicos factores.

La relación funcional se establece mediante un modelo lineal

$$Y_i = \beta_0 + \beta_1\,X_{i1} + \beta_2\,X_{i1} + \cdots + \beta_p\,X_{ip} + \epsilon_i \quad i = 1, 2, \cdots, n \quad \epsilon \sim N(0, \sigma^2) \tag{17}$$

El modelo especificado anteriormente se llama modelo de regresión lineal múltiple, si la variable dependiente Y y las variables independientes $X_1, X_2, \cdots, X_p$ son simultáneamente de valor real. Cuando en el modelo sólo se tiene sólo una variable independiente entraríamos a hablar de un modelo de regresión lineal simple.

4.1. Regresión Lineal Simple

Siguiendo (3), en muchos análisis estadísticos se desea investigar cómo los cambios en una variable afectan a otra variable. Por ejemplo, altura y peso, ingreso y cantidad de alimento, por citar algunos, se puede hablar de modelo de regresiòn lineal simple.

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{18}$$

La variable respuesta es Y, X es la variable independiente. β_0 y β_1 son llamados *p*arámetros y deben ser encontrados. La ecuación (15) significa que para un X_i dado, el correspondiente Y_i consiste de $\beta_0 + \beta_1 X_i$ y un ϵ_i, para el cual una observación debe caer dentro de la línea de regresión verdadera. Sobre las bases de la información disponible, se deben encontrar β_0 y β_1. El término ϵ es una variable aleatoria y se llama *e*rror. Se puede escribir.

$$Y_i - \beta_0 - \beta_1 X_i = \epsilon_i \tag{19}$$

el procedimiento para encontrar el valor de los parámetros β_0 y β_1 se llama *e*stimación. existen diferentes métodos para obtener tales estimaciones.

4.1.1. Mínimos Cuadrados

Este método está basado en escoger β_0 y β_1 como el mínimo de las sumas de cuadrados de las desviaciones verticales entre los datos y la recta ajustada. La suma de los cuadrados de las desviaciones de la línea es:

$$SSD = \sum_{i=1}^{n} {\epsilon_i}^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{20}$$

luego de encontrar las estimaciones de β_0 y β_1, se sustituyen en la ecuación anterior y la suma de cuadrados de las desviaciones es mínima. Diferenciando la ecuación (17) con respecto a β_0 y β_1 e igualando las derivadas parciales a cero, se tiene:

$$\begin{aligned} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) &= 0 \\ \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i) X_i &= 0 \end{aligned} \tag{21}$$

después de un poco de álgebra se llega a:

$$\begin{aligned} n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i &= \textstyle\sum_{i=1}^{n} Y_i \\ \beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2 &= \textstyle\sum_{i=1}^{n} Y_i X_i \end{aligned} \tag{22}$$

Las ecuaciones en (19) son llamadas *e*cuaciones normales. al desarrollar este sistema de ecuaciones se obtienen:

$$\hat{\beta_1} = \frac{\sum_{i=1}^{n} Y_i X_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)/n}{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i/n)^2} \tag{23}$$

$$\hat{\beta_0} = \overline{Y} - \hat{\beta_1}\overline{X} \tag{24}$$

Y $\hat{\beta}_1$ y $\hat{\beta}_0$ son llamados los *e*stimadores de mínimos cuadrados de β_1 y β_0 respectivamente. Finalmente se puede escribir la ecuación de regresión estimada:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \tag{25}$$

la cual es llamada *ecuación de predicción.*

4.2. Regresión Lineal Múltiple

Los resultados obtenidos en esta sección están basados en diferentes autores dentro de los cuales se encuentran Arthanari y Dodge (1981, (23)), Draper y Smith (1981, (24)), Seber (1977, (25)), Weisberg (1985, (26)), Dodge y Birkes (1993, (27)), Graybill (1961, (28)), Graybill (1976, (29)) y Bloomfield y Steiger (1983, (30)) entre otros. Considérese el modelo:

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \ldots + \beta_1 \mathbf{X}_p + \boldsymbol{\epsilon} \tag{26}$$

donde $\mathbf{X}_1, \ldots, \mathbf{X}_p$ son conocidas y los β_j's son parámetros desconocidos a ser estimados y $\boldsymbol{\epsilon}$ es el término de error. Si n valores de $\mathbf{Y}$ son observados, se puede escribir $\mathbf{Y'} = (Y_1, Y_2, \ldots, Y_n)$, y el modelo anterior escrito en forma matricial es:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim \mathbf{N}(\mathbf{0}, \sigma^{\mathbf{2}}\,\mathbf{I}) \tag{27}$$

En donde $\mathbf{Y}^T = (Y_1, Y_2, \cdots, Y_n)$ es el vector de observaciones, el cual es de orden $n \times 1$, $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \cdots, \beta_p)$ es el vector de parámetros de orden $k \times 1$, donde $k = p+1$. La matriz $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_k})$ es de orden $n \times k$, la primera columna de $\mathbf{X_1}^T = (1, 1, \cdots, 1)$ y $\boldsymbol{\epsilon}^T = (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$ es el vector de errores independientes, con distribución normal multivariada con vector de medias $\mathbf{0}$ y matriz de varianzas - covarianzas $\sigma^2\,\mathbf{I}_n$.

4.2.1. Estimación de Mínimos Cuadrados

Tal como se propuso en la sección anterior, el método de mínimos cuadrados para estimar el vector de parámetros $\boldsymbol{\beta}$, consiste en minimizar $\sum_{i=1}^{n} {\epsilon_i}^2$ con respecto a $\boldsymbol{\beta}$ esto es minimizar la norma cuadrática $\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$. Esta forma cuadrática se puede escribir:

$$\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

realizando los productos en la anterior expresión, se tiene:

$$\mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Diferenciando parcialmente esta forma cuadrática con respecto a $\boldsymbol{\beta}$ e igualando la derivada parcial a cero, se tiene:

$$-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0$$

la cual simplificada es:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \tag{28}$$

estas son las llamadas *e*cuaciones normales. Si $\mathbf{X}$ es de rango k, en donde $k = p+1$, entonces la matriz $\mathbf{X}'\mathbf{X}$ es definida positiva y por tanto no singular, en consecuencia se tiene una solución única:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{29}$$

Entonces para cualquier $\boldsymbol{\beta}$ se tiene:

$$\begin{aligned}
(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) &= \left[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}+\mathbf{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})\right]'\left[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}+\mathbf{X}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}})\right] \\
&= (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \\
&\geq (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}).
\end{aligned}$$

lo cual muestra claramente que el mínimo de $(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$ es $(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})$ y se obtiene cuando $\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}$.

4.2.2. Modelo Estimado y Residuos

Una vez obtenido el estimador de *mínimos cuadrados* para el modelo de regresión lineal múltiple, el modelo estimado es:

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{30}$$

a la matriz $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ se le conoce como matriz hat (sombrero), se nota $\mathbf{H}$. Esta a su vez satisface una serie de propiedades importantes, a saber:

1. $\mathbf{H}$ es simétrica, esto es, $\mathbf{H}'=\mathbf{H}$.
2. $\mathbf{H}$ es idempotente, $\mathbf{H}'\mathbf{H}=\mathbf{H}\mathbf{H}'=\mathbf{H}$.
3. $Traza(\mathbf{H})=\sum_{i=1}^{n} h_{ii}=k$, el número de parámetros del modelo.
4. $(\mathbf{I}-\mathbf{H})$, donde $\mathbf{I}$ es la idéntica de orden n hereda las propiedades (1) y (2) de la matriz $\mathbf{H}$, no obstante su traza es $n-k$.
5. La matriz $(\mathbf{I}-\mathbf{H})$ es ortogonal a la matriz $\mathbf{X}$, esto es, $(\mathbf{I}-\mathbf{H}).\mathbf{X}=\mathbf{0}$

Teniendo en cuenta lo anterior, el modelo estimado se puede escribir:

$$\hat{\mathbf{Y}} = \mathbf{HY} \tag{31}$$

El vector de residuos es la diferencia entre el modelo observado y el modelo estimado:

$$\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \tag{32}$$

4.2.3. Algunos supuestos y resultados importantes

Según Graybill (1976, (29)) en el modelo matricial (24), si las variables independientes $\mathbf{X}_j$ que componen la matriz $\mathbf{X}$ son no aleatorias, además $E(\boldsymbol{\epsilon}) = 0$ y $\mathbf{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$, la relación especifica un *Modelo Lineal General*. Si por su parte las variables independientes $\mathbf{X}_j$ son aleatorias, $\mathbf{Y}$ y $\mathbf{X}$ tienen distribución conjunta y la estimación de los parámetros se realiza en la distribución condicional $(\mathbf{Y}|X = x)$, se está frente a un *Modelo de Regresión* propiamente dicho. En este modelo hay dos importantes supuestos para el error:

1. **Caso 1.** $E(\boldsymbol{\epsilon}) = 0$ y $\mathbf{cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}$.
2. **Caso 2.** $\boldsymbol{\epsilon} \sim \mathbf{N}_n(\mathbf{0}, \sigma^2\mathbf{I})$.

Sea el modelo (24) con los supuestos del Caso 2, se obtienen los siguientes resultados:

1. $\hat{\boldsymbol{\beta}}$ es el estimador de máxima verosimilitud de $\boldsymbol{\beta}$.

2. $\hat{\sigma}^2 = [1/(n-k)]\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}$ es el estimador de máxima verosimilitud de σ^2.

3. $\hat{\boldsymbol{\beta}} \sim \mathbf{N}_k\big(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\,\sigma^2\big)$.

4. $(n-k)\,\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(n-k)}$.

5. $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$ son independientes.

6. $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$ son estadísticas suficientes para $\boldsymbol{\beta}$ y σ^2.

7. $\hat{\boldsymbol{\beta}}$ y $\hat{\sigma}^2$ son estadísticas completas.

8. $\mathbf{r} \sim \mathbf{N}_n\big(\mathbf{0}, (\mathbf{I}-\mathbf{H})\sigma^2\big)$.

9. $\mathbf{r}$ y $\hat{\boldsymbol{\beta}}$ son independientes.

Es de aclarar que los resultados del teorema anterior son exactamente iguales a los resultados obtenidos mediante el método de mínimos cuadrados, para el modelo (24) bajo los supuestos Caso 2, en otras palabras, bajo el supuesto de normalidad de los errores, la estimación de parámetros y resultados posteriores, bajo los métodos de mínimos cuadrados y máxima verosimilitud coinciden.

También el modelo de regresión lineal simple, es un caso particular del modelo de regresión lineal múltiple ó el modelo de regresión lineal múltiple es un caso general del modelo de regresión lineal simple.

4.2.4. Prueba de los coeficientes de regresión

Uno de los primeros problemas en regresión consiste en ensayar si las variables independientes o regresoras contienen cualquier información explicativa significante. Se busca comparar el modelo *completo* conteniendo todas las variables explicativas, con el modelo reducido $\mathbf{Y} = \boldsymbol{\beta_0} + \boldsymbol{\epsilon}$ que no contiene variables explicativas.

Prueba Estadística. La idoneidad de un modelo puede ser juzgada por el tamaño de los residuos. Si SSR representa la suma de cuadrados de los residuos de un modelo. Se puede comparar el modelo completo con el modelo reducido, es decir comparando SSR_f con SSR_r. Específicamente la estadística para probar la hipótesis nula $H_0 : \beta_1 = \beta_2 = \cdots, \beta_p = 0$ es

$$F = \frac{SSR_r - SSR_f}{4\,\hat{\sigma}^2} \tag{33}$$

donde $\hat{\sigma}^2$ es una estimación de la varianza de los errores, estipulada en el numeral (2) del teorema anterior.

Valor P. La evidencia en contra de la hipótesis nula $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ es medida por la magnitud de la estadística F, la cual si es grande redunda en un valor pequeño de P, es decir si P es muy pequeña, por ejemplo 0,01 significa que el valor observado F es tan grande que resulta improbable que la hipótesis nula sea cierta, o es más sensato concluír que la hipótesis nula es falsa. Para calcular el valor P se hace necesario conocer la distribución de la variable F cuando la hipótesis nula es verdadera.

Prueba de $H_0 : \beta_{q+1} = \cdots = \beta_p = 0$. Al considerar el modelo $Y_i = \beta_0 +$

$\beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$. La prueba de hipótesis anterior se puede extender, para comparar el modelo completo con un modelo reducido de cualquier subconjunto de variables explicativas. Este modelo reducido puede ser $Y_i = \beta_0 + \beta_1 X_1 + \cdots + \beta_q X_q + \epsilon$ con $q < p$. Se puede notar que comparar estos dos modelos equivale a probar la hipótesis nula $H_0 : \beta_{q+1} = \cdots = \beta_p = 0$. Sea $\hat{\sigma}^2$ una estimación insesgada de σ^2 dada por:

$$\hat{\sigma}^2 = \frac{SSR_f}{n - p - 1} \tag{34}$$

la estadística para probar la hipótesis nula $H_0 : \beta_{q+1} = \cdots = \beta_p = 0$ es:

$$F = \frac{SSR_r - SSR_f}{(n - p)\hat{\sigma}^2} \tag{35}$$

Prueba de $\beta_j = 0$. El valor de $\hat{\beta}_j$ debería indicar si $\beta_j = 0$ o no. La estadística para probar $H_0 : \beta_j = 0$ en el modelo de regresión lineal simple es:

$$|t| = \frac{|\hat{\beta}_j|}{est.SD(\hat{\beta}_j)}$$

esta ecuación representa una estadística razonable para probar la hipótesis $H_0 : \beta_j = 0$. Luego se hace indispensable cómo obtener $est.SD(\hat{\beta}_j)$. En el modelo de regresión lineal simple, se cumple que:

$$Var(\hat{\beta}_j) = \sigma^2 / \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Similarmente en regresión lineal múltiple $Var(\hat{\beta}_j)$ puede ser obtenida multiplicando σ^2 por la $(j+1)$ésima entrada en $(\mathbf{X}'\mathbf{X})^{-1}$ para cualquier $j = 0, 1, 2, \ldots, p$. La estimación de σ^2 está dada por $\hat{\sigma^2} = \frac{SSR_f}{(n-k)}$, al sacar raiz cuadrada de esta expresión se obtiene $est.SD(\hat{\beta}_j)$. Cuando $\beta_j = 0$ la estadística $t \sim t_{(n-k)}$ asociada con $\hat{\sigma}^2$.

4.2.5. Coeficiente de determinación.

El *coeficiente de determinación* es una medida de cuánto las variables independientes explican la variable respuesta. Se define por:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y} - \overline{y})}{\sum_{i=1}^{n}(y_i - \overline{y})}$$

R^2 es la proporción de la variabilidad total en la variable respuesta, que ha sido explicada por las variables independientes. El coeficiente de determinación está estrechamente relacionado a la estadística F para probar la hipótesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p$. En efecto R^2 puede ser expresado como una función de F. En el caso de regresión simple, R^2 es el cuadrado de la correlación muestral entre X y Y.

4.3. Influencia Local

En Paula (2004, (32)), se presenta uno de los métodos más modernos de diagnóstico, el cual fué propuesto por Cook (1987, (33)). Consiste fundamentalmente en hallar el cambio del modelo, bajo pequeñas perturbaciones o modificaciones de los datos en el mismo modelo.

El logaritmo de la función de verosimilitud para el parámetro β es:

$$L_\delta(\beta; \mathbf{y}) = \sum_{\mathbf{j=1}}^{\mathbf{n}} \delta_{\mathbf{j}} \mathbf{L}(\beta; \mathbf{y_j}) \tag{36}$$

en donde $L(\beta; \mathbf{y_j})$ es el logaritmo de la función de verosimilitud correspondiente a la j-ésima observación y δ_j es un tipo de perturbación, definida tal que $0 \leq$

$\delta_j \leq 1$. Cuando $\delta_j = 1, \forall j$ significa que no hay perturbación en el modelo y cuando $\delta_j = 0$ significa que la j-ésima observación fué excluída.

La estimación de mínimos cuadrados, bajo la estructura (33) es dada por:

$$\hat{\beta}_\delta = (\mathbf{X^T \Delta X})^{-1}\mathbf{X^T \Delta y}$$

donde $\mathbf{\Delta} = diag\{\delta_\mathbf{1}, \delta_\mathbf{2}, \ldots, \delta_\mathbf{n}\}$. En particular cuando apenas la i-ésima observación es perturbada, esto es, cuando $\delta_i = \delta$ y $\delta_j = 1$ para $j \neq i$ se muestra que

$$\hat{\beta}_\delta = \hat{\beta} - \frac{(1-\delta)r_i}{\{1-(1-\delta)h_{ii}\}}(\mathbf{X^T X})^{-1}\mathbf{x_i}. \tag{37}$$

Para $\delta = 0$, o sea que el i-ésimo punto es excluído (34) queda expresada en forma simplificada

$$\hat{\beta_{(\mathbf{i})}} = \hat{\beta} - \frac{r_i}{(1-h_{ii})}(\mathbf{X^T X})^{-1}\mathbf{x_i} \tag{38}$$

que es bastante conocida en Regresión Lineal normal (Cook y Weisberg, (24)).

La medida de influencia más conocida está en la región de confianza para el parámetro β,

$$(\hat{\beta}-\beta)(\mathbf{X^T X}))(\hat{\beta}-\beta) \leq \mathbf{ps^2 F_{p,(n-p)}}(\alpha)$$

que para el caso $p = 2$ es un elipsoide de $\mathbf{R}^2$ centrado en $\hat{\beta}$. Tal medida, conocida como distancia de Cook es definida por:

$$D_\delta = \frac{(\hat{\beta}-\hat{\beta}_\delta)(\mathbf{X^T X}))(\hat{\beta}-\hat{\beta}_\delta)}{ps^2} \tag{39}$$

y mide cuánta perturbación $\delta = (\delta_\mathbf{1}, \delta_\mathbf{2}, \ldots, \delta_\mathbf{n})^\mathbf{T}$ aleja $\hat{\beta}_\delta$ de $\hat{\beta}$, según la métrica $(\mathbf{X^T X})$. Por ejemplo, si $D_\delta > F_{p,(n-p)}(1-\alpha)$, significa que una perturbación está distorsionando el contorno de la elipse a un nivel de significancia menor que α . En particular, cuando el i-ésimo punto es excluído, la distancia de Cook queda expresada en la forma:

$$\begin{aligned} D_i &= \frac{(\hat{\beta}-\hat{\beta}_{\mathbf{i}})(\mathbf{X^T X}))(\hat{\beta}-\hat{\beta}_{\mathbf{i}})}{ps^2} \\ &= \left\{\frac{r_i}{s(1-h_{ii})^{\frac{1}{2}}}\right\}^2 \frac{h_{ii}}{(1-h_{ii})}\frac{1}{p} \\ &= t_i^2\frac{h_{ii}}{(1-h_{ii})}\frac{1}{p} \end{aligned}$$

Por tanto D_i será grande cuando el i-ésimo punto es aberrante (t_i grande) o cuando h_{ii} es próximo a 1. La distancia D_i podrá no ser adecuada cuando r_i sea grande y h_{ii} pequeño. En este caso, s^2 podrá quedar inflado y no se tendría ninguna compensación por parte de h_{ii}, D_i puede quedar pequeño. Una medida supuestamente más apropiada fué propuesta por Belsley, Kuh y Welsch (1980):

$$\begin{aligned} DFFITS_i &= \frac{|r_i|}{s_{(i)}(1-h_{ii})^{\frac{1}{2}}}\left\{\frac{h_{ii}}{(1-h_{ii})}\right\}^{\frac{1}{2}} \\ &= |t_i^*|\left\{\frac{h_{ii}}{(1-h_{ii})}\right\}^{\frac{1}{2}} \end{aligned}$$

Como el valor esperado de h_{ii} es $\frac{p}{n}$ parece razonable dar más atención a aquellos puntos tales que

$$DFFITS_i \geq 2\left\{\frac{p}{n-p}\right\}^{\frac{1}{2}}$$

Aparentemente D_i y $DFFITS_i$ serían medidas de influencia competitivas, una vez que $DFFITS_i$ parece ser más adecuada para validar la influencia en las estimaciones de los coeficientes de un punto aberrante con h_{ii} pequeño.

4.3.1. Residuos

Siguiendo los principios de Cook (1989, (21)), se conocen tres tipos de residuos, estos son:

- **Residuos simples:** se definen como la diferencia entre el valor observado de la variable respuesta y el valor estimado por el modelo. Tomando como referencia el modelo lineal se definen:

  $$\mathbf{r} = \mathbf{Y} - \hat{\mathbf{Y}}; \quad E(\mathbf{r}) = \mathbf{0}; \quad y \quad Var(\mathbf{r}) = (\mathbf{I} - \mathbf{H})\sigma^2$$

  en donde $\mathbf{I}$ es la matriz idéntica y $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ es llamada matriz hat, tiene las propiedades de ser simétrica e idempotente, es decir $\mathbf{H}^T = \mathbf{H}$ y $\mathbf{H}^T\mathbf{H} = \mathbf{H}\mathbf{H}^T = \mathbf{H}$.

- **Residuos estandarizados:** son los mismos residuos simples estandarizados en media y varianza, para la i-ésima observación se notan y definen:

  $$t_i = \frac{r_i}{s(1 - h_{ii})} \quad con \quad s^2 = \sum_{i=1}^{n} r_i^2/(n-p) \quad para \quad i = 1, 2, \cdots, n$$

- **Residuos estudentizados:** se definen como la diferencia entre el valor observado de la variable y el valor estimado, cuando la i-ésima observación ha sido eliminada, o sea $Y_i - \hat{Y_{(i)}}$, más exactamente se definen:

  $$t_i^* = \frac{r_i}{s_{(i)}(1 - h_{ii})^{1/2}}$$

  $s_{(i)}^2$ es la varianza estimada sin la i-ésima observación.

En Rao (1973, (25), p.185) se establecen las siguientes relaciones:

$$s^2_{(i)} = s^2 \left(\frac{n-p-t_i^2}{n-p-1} \right) \quad y \quad t_i^* = \left(\frac{n-p-1}{n-p-t_i^2} \right)$$

- **Distancia de Cook:** Excluyendo la i-ésima observación del modelo se define:

$$D_{(i)} = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})^T (X^T X)(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})}{p\, s^2} = t_i^2 \frac{h_{ii}}{(1-h_{ii})} \frac{1}{p}$$

- **DFFITS:** propuestos por Belsley, Kuh y Welsch (1980, (**?**)), se definen por:

$$DFFITS = \frac{|r_i|}{s_{(i)}(1-h_{ii})^{1/2}} \left\{ \frac{h_{ii}}{(1-h_{ii})} \right\}^{1/2} = |t_i^*| \left\{ \frac{h_{ii}}{(1-h_{ii})} \right\}^{1/2}$$

- **Alternativa a los DFFITS:** Atkinson (1985, (27)) propone una alternativa a los DFFITS que se definen para la i-ésima observación como:

$$C_i = \left\{ \frac{(n-p)}{p} \frac{h_{ii}}{(1-h_{ii})} \right\}$$

4.3.2. Criterios de Información

La validez y calidad de un GLM se mide a través de la función *desvío* y los criterios de información.

- **Criterio de información de Akaike:** mide la calidad relativa del ajuste de un modelo estadístico a un conjunto de datos:

$$\mathbf{AIC} = 2\,k - 2\,log(\mathbf{L})$$

en donde k es el número de parámetros del modelo y $\mathbf{L}$ es el máximo valor de la función de verosimilitud. Se prefiere el modelo con menor **AIC**.

- **Criterio de información de Bayes:** criterio para la selección de un modelo, entre un conjunto finito de modelos,

$$\mathbf{BIC} = -2\,ln(\mathbf{L}) + k\,ln(n)$$

al tener dos modelos, se prefiere el que menor **BIC** tenga.

- **Criterio de información de Hannan - Quinn:** es un criterio alternativo al **AIC**, se define:

$$\mathbf{HQC} = n\,log\left(\frac{\mathbf{RSS}}{n}\right) + 2\,k\,log\,log(n)$$

en donde **RSS** es la reducción en sumas de cuadrados del error del modelo estimado.

5. Objetivos

5.1. Objetivo General

Analizar la eficiencia del recaudo tributario de los municipios del departamento del Huila utilizando la técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica en el periodo 2008 - 2019.

5.2. Objetivos Específicos

- Describir la evolución del recaudo tributario de los municipios del departamento del Huila
- Identificar la participación de los diferentes impuestos en los ingresos tributarios municipales
- Comparar los resultados obtenidos acerca de la eficiencia de los municipios huilenses entre la técnica de Análisis de la Envolvente de Datos y la técnica de Análisis de la Frontera Estocástica.

6. Esquema Procedimental

Como se mencionó anteriormente en la Introducción, el propósito de la presente investigación consiste en adecuar dos modelos basados en Análisis Envolvente de Datos (DEA), para la información del Recaudo Tributario en los Municipios del Departamento del Huila. Dicha información fué suministrada por el Departamento Nacional de Planeación. Como variables de análisis se utiliza el Recaudo tributario en los Municipios del departamento del Huila, utilizando como ventana de información los años 2008 - 2019, los gastos de personal y los gastos generales. Se utiliza una variable de entrada y dos de salida.

Los análisis a desarrollar en el marco del trabajo son: Análisis no paramétrico DEA y Fronteras Paramétricas Estocásticas (SFA). En esta parte del trabajo se determinan los municipios que producen mayor eficiencia.

Un tercer análisis consiste en realizar un modelo de regresión lineal múltiple, con el fin de determinar el comportamiento actual del problema y resumir una ecuación que permita desarrollar pronósticos. También se desarrolla el análisis de sensibilidad del modelo encontrado, a través de Influencia Estadística Local.

7. Análisis de Datos

7.1. Análisis Modelo DEA

Para el análisis de la informacións, una vez revisada y depurada la base de datos, se procesó mediante el Software R de libre uso, se redifinieron las variables y se procesaron los datos mediante el uso de las librerías: $Benchmarking$, GLM, $ISLR$, $XTAB$, $NPARCOMP$, con el fin de obtener los modelos DEA y SFA, que describan los datos y permitan a su vez determinar las DMU's más eficientes, con el propósito de plantear posibles planes de acción en los municipios donde se pueda comprobar ineficiencia. Así mismo se intenta estimar el modelo de predicción lineal que permita evaluar situaciones futuras que necesiten revisar posibles ajustes de tipo financiero.

Eficiencias (E)		
Rango	frecuencia	Porcentaje (%)
$0 \leq E < 0{,}1$	233	52.5
$0{,}1 \leq E < 0{,}2$	28	6.3
$0{,}2 \leq E < 0{,}3$	29	6.5
$0{,}3 \leq E < 0{,}4$	32	7.2
$0{,}4 \leq E < 0{,}5$	35	7.9
$0{,}5 \leq E < 0{,}6$	30	6.8
$0{,}6 \leq E < 0{,}7$	12	2.7
$0{,}7 \leq E < 0{,}8$	13	2.9
$0{,}8 \leq E < 0{,}9$	9	2.0
$0{,}9 \leq E < 1$	7	1.6
$E = 1$	16	3.6

Cuadro 1: Eficiencias No Paramétricas DEA

En el cuadro (1) se observa que hay 233 municipios con una eficiencia entre $0 \leq E < 0,1$, lo cual representa un $52,9\,\%$ del total de municipios, 16 municipios con una alta eficiencia, que sólo representan el $3{,}6\,\%$ del total de municipios, con una eficiencia entre $0{,}9 \leq E < 1$ se encuentran 7 municipios que sólo representan el $1{,}6\,\%$ y 9 municipios con una eficiencia $0{,}8 \leq 0{,}9$ que representan el $2\,\%$.

En los cuadros siguientes se exhiben los municipios del departamento del Huila con mayor eficiencia en el recaudo tributario, ordenados de mayor a menor eficiencia.

Municipios con Mayor Eficiencia				
Orden	Municipio	Año	Eficiencia	Porcentaje
30	Algeciras	2013	1	3.6
31	Algeciras	2014	1	3.6
69	Campoalegre	2016	1	3.6
105	El Pital	2016	1	3.6
202	La Argentina	2017	1	3.6
203	La Argentina	2018	1	3.6
220	Nátaga	2011	1	3.6
275	Palermo	2018	1	3.6
350	Suaza	2009	1	3.6
354	Suaza	2013	1	3.6
372	Tarqui	2019	1	3.6
434	Yaguará	2009	1	3.6

Cuadro 2: Municipios del Huila con Mayor Eficiencia

Municipios con Eficiencia $0{,}9 \leq E < 1$				
Orden	Municipio	Año	Eficiencia	Porcentaje
58	Baraya	2017	0.9909	1.6
61	Campoalegre	2008	0.9010	1.6
63	Campoalegre	2010	0.9715	1.6
180	Iquira	2019	0.9605	1.6
265	Palermo	2008	0.9581	1.6
383	Tello	2018	0.9906	1.6
433	Yaguará	2008	0.9232	1.6

Cuadro 3: Municipios del Huila con Eficiencia Muy Buena

Municipios con Eficiencia $0{,}8 \leq E < 0{,}9$				
Orden	Municipio	Año	Eficiencia	Porcentaje
15	Aipe	2011	0.8502	2.0
24	Aipe	2019	0.8355	2.0
57	Baraya	2016	0.8857	2.0
62	Campoalegre	2009	0.8747	2.0
150	Guadalupe	2013	0.8020	2.0
233	Neiva	2012	0.8975	2.0
276	Palermo	2019	0.8156	2.0
288	Palestina	2019	0.8531	2.0
382	Tello	2017	0.8615	2.0

Cuadro 4: Municipios del Huila con Eficiencia Buena

Algunos Municipios con Eficiencia $0{,}0 \leq E < 0{,}1$				
Orden	Municipio	Año	Eficiencia	Porcentaje
1	Acevedo	2008	0.0000	52.5
4	Acevedo	2011	0.0940	52.5
14	Aipe	2009	0.0000	52.5
39	Altamira	2010	0.0000	52.5
48	Baraya	2008	0.0000	52.5
84	Colombia	2019	0.0000	52.5
152	Guadalupe	2015	0.0000	52.5
274	Palermo	2017	0.0000	52.5
429	Villavieja	2016	0.0880	52.5

Cuadro 5: Municipios del Huila con Eficiencia Muy Baja

7.2. Análisis Modelo SFA

Para el análisis de los datos según el Análisis de Superficie Estocástica (SFA), se utilizó para los errores u_i una distribución $N(0, \sigma^2)$ y para los errores v_i se utilizó una normal truncada con media cero y varianza σ^2. Se calculó la frontera estocástica. También se determinaron las varianzas parciales de u_i y de v_i, así como la varianza total. Se probó el modelo multiplicativo de Cobb-Douglas, usando los logaritmos tanto del recaudo, como de los gastos de personal y generales.

Parámetro	Estimación	Error Est.	Valor z	P	Signif.
Intercepto	3.0992453	0.9670503	3.2048	0.0013514	$**$
log(gastosp)	0.6560970	0.1891246	3.4691	0.0005222	$***$
log(gastosg)	0.5802426	0.1742528	3.3299	0.0008688	$***$
log(gp)*log(gg)	-0.1279261	0.0338189	-3.7827	0.0001551	$***$
Sigma2	1.8023061	0.1634938	11.0237	2.2e-16	$***$
gamma	0.9895921	0.0059999	164.9354	2.2e-16	$***$

Cuadro 6: Modelo Análisis de la Frontera Estocástica (SFA)

Se observa en el cuadro (6) que el intercepto, las variables log(gastosp), log(gastosg), así como la interacción entre ellas resultan ser altamente significativas, esto inmediatamente nos permite calcular las eficiencias SFA, para cada una de las DMUs que son los respectivos municipios con el año correspondiente de recaudo. Se observará también que en este caso no hay eficiencias muy altas, pero gracias al análisis del modelo multiplicativo, se puede tener un poco más de credibilidad de las eficiencias SFA, frente al modelo no paramétrico DEA. Después se determinará un modelo estadístico que permitirá visualizar la mejor distribución de probabilidad, asociada al conjunto de datos ó base de datos.

También se determinan las varianzas o componentes de varianza, las cuales especifican la eficiencia o ineficiencia de las entidades involucradas en el análisis.

Orden	Municipio	Año	Eficiencia	Porcentaje
2	Acevedo	2009	0.9547	6,40
3	Acevedo	2010	0.9504	6,40
45	Altamira	2016	0.9644	6,40
102	El Pital	2013	0.9687	6,40
127	Garzón	2014	0.9219	6,40
128	Garzón	2015	0.9692	6,40
129	Garzón	2016	0.9996	6,40
130	Garzón	2017	0.9955	6,40
140	Gigante	2015	0.9138	6,40
150	Guadalupe	2013	0.9505	6,40
163	Hobo	2017	0.9626	6,40
184	Isnos	2011	0.9530	6,40
185	Isnos	2012	0.9894	6,40
199	La Argentina	2014	0.9706	6,40
200	La Argentina	2015	0.9021	6,40
232	Neiva	2011	0.9034	6,40
233	Neiva	2012	0.9721	6,40
234	Neiva	2013	0.9858	6,40
235	Neiva	2014	0.9996	6,40
236	Neiva	2015	0.9972	6,40
250	Oporapa	2017	0.9300	6,40
257	Paicol	2012	0.9754	6,40
270	Palermo	2013	0.9764	6,40
271	Palermo	2014	0.9350	6,40
285	Palestina	2016	0.9087	6,40
287	Palestina	2018	0.9100	6,40
317	Saladoblanco	2012	0.9859	6,40
320	Saladoblanco	2015	0.9074	6,40

Cuadro 7: Mejores Eficiencias SFA

7.3. Análisis Modelo Final

El modelo final consiste relacionar la variable respuesta Y : Recaudo Tributario versus las variables X_1: Gastos de Personal y X_2 : Gastos Generales. Es de aclarar que las unidades hacedoras de decisión ó DMU's están representadas por cada uno de los municipios del Departamento del Huila, durante los años 2008-2019.

Para este análisis se procede a examinar la distribución empírica de la variable log(recaudo), la cual en este análisis se usa como variable dependiente, las variables independientes se toman como log(gastosp) y log(gastosg). Se pretende encontrar hasta que punto las variables independientes, explican en forma lineal a la variable independiente. Se utilizan las librerías *glm*, *gamlss* y *gamlss.dist* del paquete R.

En el cuadro (8), se puede evidenciar que el Intercepto, así como la varia-

Parámetro	Estimació	Error Est.	Valor t	Pr($>|t|$)	Signif.
Intercepto	5.11	0.14	37.10	0.00	$***$
log.gastosg.	0.00	0.02	0.13	0.89	
log.gastosp.	0.08	0.02	4.23	0.00	$***$
σ^2	-0.43	0.18	-2.39	0.02	$***$
ν	-0.49	0.09	-5.56	0.00	$***$
τ	0.66	0.16	4.21	0.00	$***$

Cuadro 8: Mejor Modelo Lineal General

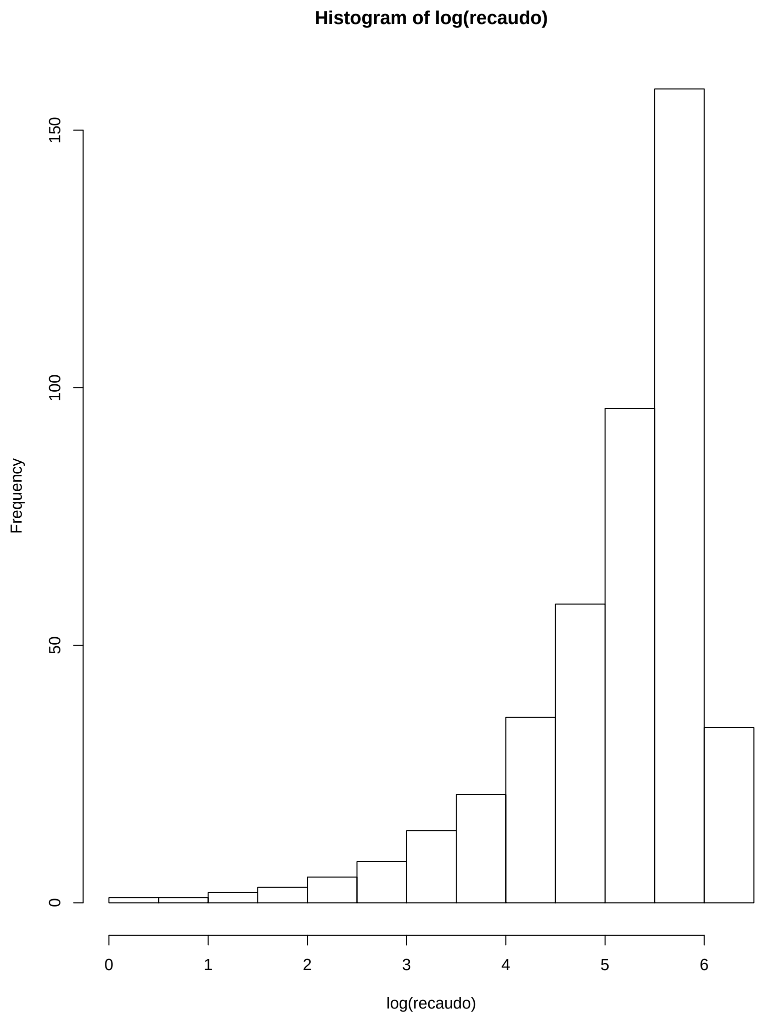

Figura 1: Distribución variable Dependiente

ble log(gastosp) son altamente significativos. El parámetro de escala es $\hat{\sigma^2} = e^{-0,43} = 0{,}6505$, el parámetro de asimetría es $\hat{\nu} = e^{-0,49} = 0{,}6126$, mientras que el parámetro de curtosis es $\hat{\tau} = e^{0,66} = 1{,}9347$, lo cual permite ver que es una distribución asimétrica negativa. Todos estos parámetros de forma son altamente significativos. Se evidencia también que el parámetro asociado con la variable independiente log(gastosg) no es significativo. Esto explica que esta variable no está directamente asociada a la variable dependiente log(recaudo). La distribución que mejor ajustó los datos es la SHASH: *Seno Hiperbólico Arco Seno Hiperbólico*, distribución con cuatro parámetros, de la cual hacemos la siguiente descripción:

7.3.1. Función de densidad de probabilidad SHSAH

$$f(y|\mu,\sigma,\nu,\tau) = \frac{c}{\sqrt{2\,\pi}\,\sigma\,(1+z^2)^{\frac{1}{2}}}\,e^{-\frac{r^2}{2}} \tag{40}$$

Los parámetros:

$$r = \frac{1}{2}\{e^{\tau\,sinh^{-1}(z)} - e^{-\nu\,sinh^{-1}(z)}\}$$

$$c = \frac{1}{2}\{\tau\,e^{\tau\,sinh^{-1}(z)} - \nu\,e^{-\nu\,sinh^{-1}(z)}\}$$

La variable $z = \frac{y-\mu}{\sigma}$, en donde $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ y $\nu > 0$.

La escogencia de esta distribución se basa en los criterios de información de Akaike, de Bayes y el Desvío residual, que después de la estimación, produjeron los valores: 940,4084, 964,888 y 928,4084, respectivamente, los menores frente a un conjunto aproximado de 20 distribuciones.

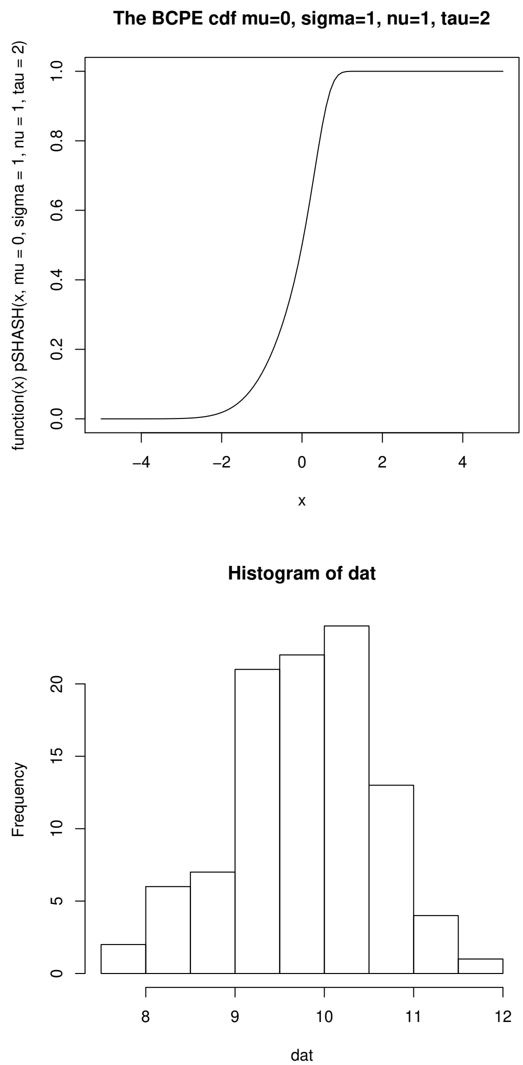

Figura 2: Distribución e Histograma SHASH

7.3.2. Influencia Estadística Local

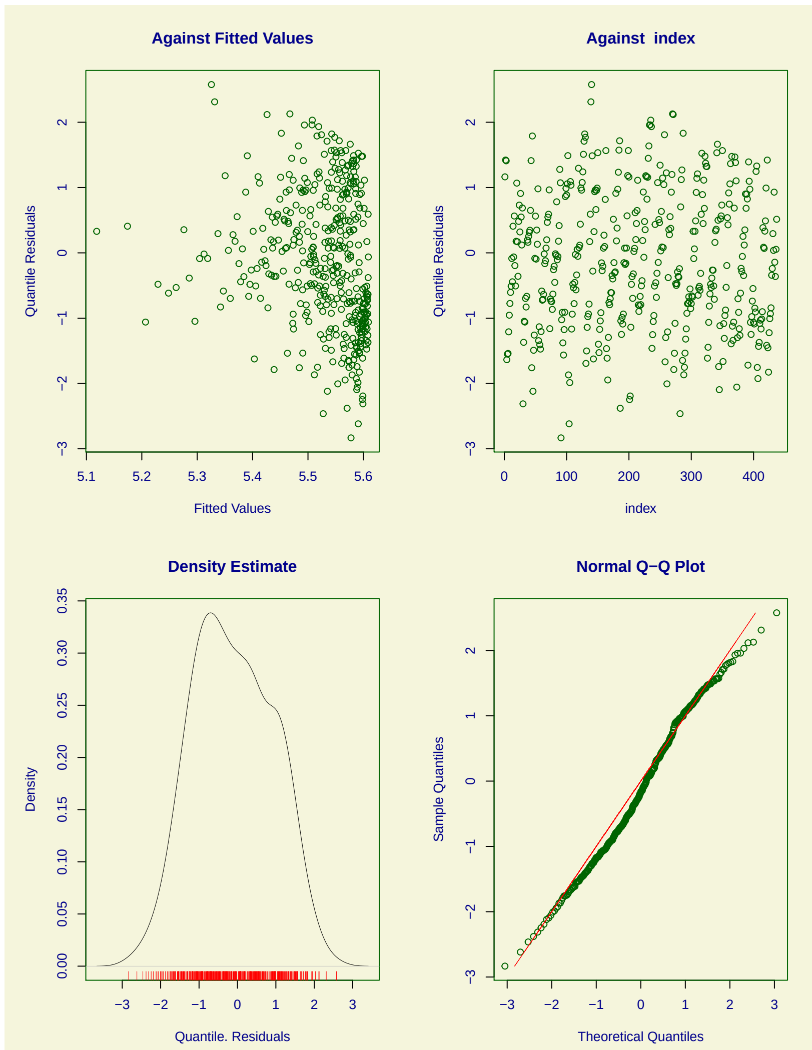

Figura 3: Influencia estadística Local, modelo SHASH

Se evidencia en la figura (3), que los cuantiles residuales oscilan en el intervalo $[-2, 2]$ en su gran mayoría, sólo alguno valores se salen de este rango.

La distribución empírica de los cuantiles residuales parece una mezcla de tres normales y la recta de ajuste con los datos es aceptable. Por tanto el ajuste se considera bueno.

8. Conclusiones

1. El análisis no paramétrico DEA permitió calcular las eficiencias de los municipios del departamento del Huila. No obstante hubo municipios sin información que resultaron eficientes.

2. El método de la frontera Estocástica resultó ser más eficiente, usando incluso eficiencias más bajas que las obtenidas con el método no paramétrico DEA. El municipio de Neiva obtuvo una eficiencia promedio de 0,9721 con SFA, frente a un 0,8975 con DEA no paramétrico.

3. En el método no paramétrico, el municipio de Neiva sólo presentó una réplica, mientras que con el modelo SFA presentó 5 réplicas.

4. El modelo econométrico apropiado para los datos es el modelo multiplicativo de Cobb-Douglas, para las variables log(recaudo) como variable dependiente y log(gastosp) y log(gastosg) como variables independientes.

5. La distribución estadística más apropiada para la base de datos resultó ser la distribución de 4 parámetros SHASH: *Seno Hiperbólico Arco Seno Hiperbólico*, con esta distribución se puede utilizar un modelo aditivo.

6. La variable Log(gastosp) explica significativamente a la variable respuesta, en forma aditiva, usando sólo la distribución SHASH, ver cuadro (8).

9. Bibliografía

[1] Iregui, Ana María and Ramos, Jorge and Saavedra, Luz Amparo and others *Análisis de la descentralización fiscal en Colombia* Banco de la República, Subgerencia de Estudios Económicos, 2001.

[2] Iregui, A., Ramos, J. y Saavedra, L. 2001. *Análisis de la descentralización fiscal en Colombia.* Recuperado de http://www.banrep.gov.co

[3] Restrepo, J. 2012. *Hacienda Pública. (9ª ed.). Bogotá:. Universidad Externado de Colombia.*

[4] Ocampo, J. 1994 (comp.). *Historia Económica de Colombia.* Bogotá: Tercer Mundo Editores

[5] Fino. G. 2012. *Reforma tributaria municipal, una propuesta de crecimiento con equidad.* Bogotá: Editorial Temis S.A.

[6] Congreso de Colombia. (6 de julio de 1983). Por la cual se fortalecen los fiscos de las entidades territoriales y se dictan otras disposiciones(Ley 14 de 1983). Recuperado de http://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=267

[7] Camacho, A. 2011. *El Impuesto Predial: Análisis desde una perspectiva nacional. Vínculo con el catastro. Bogotá: Instituto Colombiano de Derecho Tributario.*

[8] Fernández Santos, Yolanda Flórez López, Raquel *Aplicación del modelo dea en la gestión pública. Un análisis de la eficiencia de las capitales de provincia españolas*, 2006.

[9] Ordoñez Santo, Mary.2012. *Panorama de la de descentralización administrativa en Colombia.*

[10] Valencia-Cárdenas, Marisol Restrepo - Morales, Jorge Anibal. 2016. *Evaluación de la gestión financiera usando variables latentes en modelos estocásticos de fronteras eficientes.*

[11] Sperat, Ramiro Rodríguez Brugiafreddo, María Raña, Erica., 2017. *Análisis envolvente de datos (DEA) versus aproximación de fronteras estocásticas (SFA)*

[12] Soto, J.A. Arenas, W. *Análisis Envolvente de Datos.* Facultad de Ingeniería Industrial. Universidad Tecnológica de Pereira, 2010.

[13] Roa, A.; centauro, L.; Padilla, K. y Quesada, V.M. (2003). *Productividad y Eficiencia en la Empresa: un enfoque práctico.* Universidad de Cartagena.

[14] Farrell, M.J. (1957). The Measurement of Productivity Effi-

ciency. *Journal of the Royal Statistical Society.* Series A 120(3), 253-290.

[15] MEEUSEN, W. AND BROECK, V.D 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International economic review.* 435-444.

[16] AIGNER, D.J., LOVELL, C.A.K AND SCHMIDT, P. (1977). Formulation and Estimation of Stochastics. *Review of Economics and Statistics.* 80(3), 454-465.

[17] BATTESE, G.E., AND CORRA, G.S. (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.

[18] SÁNCHEZ, A. *Comparación de Algunos Métodos de Regresión Alternativa vs. Estadística Bayesiana Usando MCMC.* Tesis de Maestría en Investigación Operativa y Estadística, Universidad tecnológica de Pereira, 2014.

[19] TAHA, H. *Investigación de Operaciones.* Prentice Hall, Mexico, 1998.

[20] THANASSOULIS, E. *Introduction to the Theory and Application of Data Envelopment Data.* Kluwer Academic Publishers, 2003.

[21] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications.* Second Edition. Wiley, New York.

[22] Belsley, D.A.; Kuh, E. and Welsch, R.E (1980). *Regression Diagnostics.* Second Edition. John Wiley, New York.

[23] Atkinson, A.C. (1985). *Plots, Transformations and Regressions.* Oxford Statistical Sciences Series. Oxford.

[24] Arthanari, T.S. and Dodge, Y. (1981). *Mathematical Programming in Statistics.* Jhon Wiley, New York.

[25] Drapper, N. and Smith, Y. (1981). *Applied Regression Analysis.* 2nd. Ed. Jhon Wiley & Sons Inc., New York.

[26] Seber, G.A. (1981). *Linear Regression Analysis.* Wiley, New York.

[27] Weisberg, S. (1985). *Applied Linear Regression.* Jhon Wiley & Sons Inc., New York.

[28] Birkes, D. and Dodge, Y. (193). *Alternative Methods of Regression.* Jhon Wiley & Sons Inc., New York.

[29] Graybill, F.A. (1961). *An Methods of Introduction to Linear Statistical Models.* McGraw-Hill, New York.

[30] Graybill, F.A. (1976). *Theory and Application of the Linear Models.* North Scituate, MA: Duxbury press.

[31] **Bloomfield, P. and Steiger, W.L. (1983).** ***Least Absolute Deviations.*** **Boston: Birkhäuser.**

[32] **Cook, R.D. (1989). Assessing influence on regression coefficients in generalized linear models.** ***Biometrika***, **Vol 76(4), 741 - 749.**

[33] **Paula, G.A.** ***Modelos de Regresión com apoio computacional***. **Instituto de Matemática e Estatística. Universidade de Sao Paulo. Versión preliminar. Sao Paulo, 2004.**

[34] **Cook, R.D.(1987) Influence assesment.** ***Journal of Applied Statistics***. **14, 117-131.**

[35] **Cook, R.D; Peña, D. and Weisberg, S. (1988). The likelihood dispacement: A unifying principle for influence measures.** ***Communications in Statistics, Theory and Methods***. **17, 623-640.**

[36] **rao, C.R. (1973).** ***Linear statistical inference and its applications.*** **Wiley, New York.**

[37] **Azzalini, Adelchi and Valle, A Dalla. (1996). The multivariate skew-normal distribution.** ***Biometrika***. **Vol. 86 47, 715-726.**

[38] **Atkinson, A. C. (1985).** ***Plots, Transformations and Regressions***. **Oxford Statistical Science Series, Oxford.**

[39] **Cooper, W.W., Lawrence M.S: and Tone, K.** ***Data Evelopment Analysis***. **Academic Publisher, 2004.**

f

A. Anexo I: Scritp en Software R

install.packages("Benchmarking")

install.packages("gamlss")

install.packages("gamlss.dist")

install.packages("frontier")

install.packages("foreign")

install.packages(readxl")

install.packages("tmvtnorm")

install.packages("htmlTable")

library(gamlss)

library(gamlss.dist)

library(Benchmarking)

library(frontier)

library(foreign)

library(readxl)

library(tmvtnorm)

library(htmlTable)

library(xtable)

datos<-read.table(Ç:/Tareas/USCO2020/JhonAM/RD.csv", header=T, sep="|")

View(datos)

dim(datos)

datos1<-which(datos$Recaudo == ”N/A”$)

$datos1 < -which(datos$Gastos.de.Personal== "N/A")

datos1<-which(datos$Gastos.Generales == "N/A")$

$datos1 < -datos[-c(98, 205, 350, 354, 373, 409, 421),]$

$dim(datos1)$

$recaudo < -as.numeric(datos1$Recaudo)

gastosp<-as.numeric(datos1$Gastos.de.Personal)$

$gastosg < -as.numeric(datos1$Gastos.Generales)

x <- cbind(log(gastosp), log(gastosg))

y <- matrix(log(recaudo))

Calculating efficiency

dea(x,y, RTS="vrs", ORIENTATION=ïn")

e <- dea(x,y, RTS="vrs", ORIENTATION=ïn")

which(e$eff == 1)$

$which(0{,}9 <= e$eff $eeff < 1)$

$which(0{,}8 <= e$eff $eeff < 0{,}9)$

$which(0{,}0 <= e$eff $eeff < 0{,}1)$

$View(datos1)$

$eff(e)$

$peers(e)$

$peers(e, NAMES = TRUE)$

$print(peers(e, NAMES = TRUE), quote = FALSE)$

$lambda(e)$

$summary(e)$

$View(datos1)Cobb - Douglasproductionfrontier$

$cobbDouglas < -sfa(log(recaudo)\ log(gastosp) * log(gastosg), data = datos1)$

$summary(cobbDouglas)$

Análisis Frontera Estocástica

sfa(x, y, beta0 = NULL, lambda0 = 1,

TRANSPOSE = FALSE, DEBUG=FALSE,

control=list(), hessian=2)

Estimate efficiency for each unit

o <- sfa(x,y)

eff(o)

te <- te.sfa(o)

teM <- teMode.sfa(o)

teJ <- teJ.sfa(o)

cbind(eff(o),te,Mode=eff(o, type="Mode"),teM,teJ)[1:10,]

sigma2.sfa(o) Estimated varians

lambda.sfa(o) Estimated lambda

which(0.9<eff(o) eff(o)<1)

Análisis DEA

Calculating efficiency

dea(x,y, RTS="vrs", ORIENTATION=ïn")

e <- dea(x,y, RTS="vrs", ORIENTATION=ïn")

eff(e)

peers(e)

peers(e, NAMES=TRUE)

print(peers(e, NAMES=TRUE), quote=FALSE)

lambda(e)

summary(e)

Calculating super efficiency

esuper <- sdea(x,y, RTS="vrs", ORIENTATION=ïn")

print(peers(esuper,NAMES=TRUE),quote=FALSE)

dea.plot(x,y, main="Basic plot of frontier")

dea(x,y, RTS="vrs", ORIENTATION=.ºut")

e <- dea(x,y, RTS="vrs", ORIENTATION=ïn")

eff(e)

peers(e)

peers(e, NAMES=TRUE)

print(peers(e, NAMES=TRUE), quote=FALSE)

lambda(e)

summary(e)

dea.plot.frontier(x,y,RTS="vrs",txt=LETTERS[1:length(x)],

xlim=c(0,20),ylim=c(0,20))

dea.plot.frontier(x,y,RTS="drs", add=TRUE, lty="dashed", lwd=2)

dea.plot.frontier(x,y,RTS=çrs", add=TRUE, lty="dotted")

Análisis modelo Lineal

fit1<-lm(log(recaudo) log(gastosp)+log(gastosg),data=datos1)

```
summary(fit1)

hist(log(recaudo))

fit2<-gamlss(log(recaudo) log(gastosg)+log(gastosp),data=datos1,family=SHASH)

summary(fit2)

fitted(fit2)

r<-residuals(fit2)

hist(r)

plot(fit2)
```