

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS

DEPARTAMENTO DE MATEMÁTICAS

TESIS DE GRADO

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN INVESTIGACIÓN MATEMÁTICA”

TEMA:

**“ANÁLISIS COMPARATIVO DE REGRESIÓN ORTOGONAL COMO ALTERNATIVA A LA
REGRESIÓN ORDINARIA”**

AUTOR:

COLÓN MARIO CELLERI MUJICA

GUAYAQUIL – ECUADOR

2014

DEDICATORIA

A Dios que me ha bendecido durante todo este tiempo. A mi familia por su apoyo incondicional en todo el proceso de mi progreso académico.

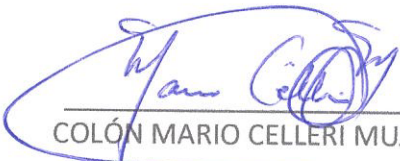
AGRADECIMIENTOS

A DIOS, a JESUS y a mis madres Isabel Hidalgo y Marjorie Mujica. A mis hermanas Raquel Plaza Mujica, Jessica y Marjorie Culcay Mujica. A Jemi por ser tan especial para mi vida y mis creencias; y a mis profesores y compañeros por todas sus enseñanzas, consejos y sugerencias.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en esta TESIS DE MAESTRÍA, me corresponde exclusivamente; el patrimonio intelectual del mismo, corresponde exclusivamente a la **Facultad de Ciencias Naturales y Matemáticas (FCNM), Departamento de Matemáticas** de la Escuela Superior Politécnica del Litoral (ESPOL).

(Reglamento de Graduación de la ESPOL)


COLÓN MARIO CELLERI MUJICA

TRIBUNAL DE GRADUACIÓN



M. Sc. Gaudencio Zurita
PRESIDENTE DEL TRIBUNAL



Doctor Francisco Vera
DIRECTOR DE TESIS



Doctor Omar Ruíz
VOCAL DEL TRIBUNAL

AUTOR DE LA TESIS



COLÓN MARIO CELLERI MUJICA

INTRODUCCIÓN

El método de regresión lineal es un método que permite ajustar una recta, la “mejor” recta, a una colección de puntos. El término “mejor” se refiere a que se minimiza la suma total de los errores o suma de los errores individuales. El método de regresión más comúnmente utilizado es el llamado de los “mínimos cuadrados” (Gillard, 2006), el cual consiste en medir la distancia existente entre el punto que corresponde a la observación y el punto sobre la recta (o subespacio) hipotético, aquella recta (o subespacio) que produzca la mínima suma es la elegida.

El concepto de regresión lineal, incluyendo el método de mínimos cuadrados tiene ya una larga historia. Los avances en los aspectos matemáticos, estadísticos y científicos durante el siglo 18 permitieron el desarrollo de lo que hoy conocemos como regresión lineal y el método de mínimos cuadrados en los albores del siglo 19.

Los principales problemas científicos del siglo 18 estaban muy relacionados con el método de mínimos cuadrados: el modelar matemáticamente el movimiento de la Luna, los movimientos aparentemente no periódicos de Júpiter y Saturno, y los esfuerzos por determinar la forma de la Tierra (Gillard, 2006). Todos estos problemas contemplaban la existencia de observaciones astronómicas, los cuales presentaban retos intelectuales para los más reconocidos y hábiles matemáticos contemporáneos.

Los trabajos de Euler y Mayer (Markovsky & Van Huffel, 2007), este último un astrónomo que en 1750, a partir de numerosas observaciones de la Luna trataba de explicar las excentricidades de la órbita lunar. El método de Mayer consistía en utilizar un sistema de ecuaciones e identidades trigonométricas sobre sus observaciones, Mayer construía 27 ecuaciones que correspondían a las 27 observaciones que realizaba de la Luna, y definía tres incógnitas, agrupaba las ecuaciones en tres grupos de 9 ecuaciones, las sumaba y resolvía el sistema con las tres ecuaciones resultantes. Aunque este método presenta gran simplicidad, era considerado revolucionario para su época.

El método de Mayer, a diferencia del método de mínimos cuadrados, no entregaba la “mejor” solución, pero era de fácil aplicación y se volvió de amplia aplicación, siendo un método *ad hoc* dependía mucho de la experiencia e intuición del investigador.

En 1787 Simon Laplace (Gillard, 2006) propuso un método muy similar al de Mayer, Laplace tomaba 24 ecuaciones y las agrupaba en grupos de 4 ecuaciones y definía 4 incógnitas, aunque Laplace nunca hizo una explicación razonable de cómo elegir las ecuaciones en cada grupo.

En 1805 el matemático francés Adrien Marie Legendre (Gillard, 2006) publicó el método de mínimos cuadrados en un apéndice de nueve páginas titulado: “*Sur la méthode des moindres quarré*” y en su presentación, en palabras de Legendre, era el método más claro y elegante de los métodos que aportan a la historia de la Estadística.

Las publicaciones de Legendre atrajeron la atención de otro gran matemático, Carl Friedrich Gauss, que en una publicación de 1809 menciona la utilización del método de mínimos cuadrados desde 1795, cuando Gauss contaba con tan sólo 18 años.

Durante un tiempo atrás estuvo en discusión si Gauss o Legendre había sido la persona que en forma inicial utilizó el método que en la actualidad se denomina de mínimos cuadrados. Inclusive existe una misiva que Gauss envió a Laplace, para que de forma diplomática intervenga como mediador entre Gauss y Legendre para resolver el impase.

Posterior investigaciones mostrarían (De Groen, 1996) que si en verdad Gauss había utilizado el método de mínimos cuadrados, no lo había hecho conocer de manera pública, ya que sólo era conocido por amigos del genio matemático, mientras que Legendre fue el primero que publicó el método, y por tanto lo hizo conocer a la comunidad científica, posteriormente en 1823, Gauss presenta el método, su utilización y la demostración con el rigor matemático del caso.

Posteriormente en 1877 Adcock (Gillard, 2006), utiliza un método de ajuste que se conoce con el nombre de “regresión ortogonal”, modelo con error en las variables” o

también llamado con “errores en la medición”, pero que en la actualidad también se conoce con el nombre de método de los “mínimos cuadrados totales”.

Los trabajos de Adcock (Gillard, 2006) se refieren al modelo univariado, pero el método de regresión ortogonal ha sido redescubierto muchas veces, por diferentes investigadores y de forma independiente, y hace unos 30 años, Sprent y Gleser (Gleser, 1981) trabajaron la técnica para problemas multivariados.

Más recientemente, el método de mínimos cuadrados totales también ha despertado el interés fuera de las estadísticas. En el campo del análisis numérico, este problema se estudió por primera vez por Golub y Van Loan (Golub & Van Loan, 1980). Su análisis, así como su algoritmo, se basa en el procedimiento de descomposición de valores singulares. Una visión geométrica de las propiedades de la descomposición de valores singulares se presentó por Staar (Golub & Van Loan, 1980) de forma independiente.

Van Huffel y Vandewalle (Van Huffel, 2004) generalizaron el algoritmo de Golub y Van Loan para todos los casos en que dicho algoritmo fallaba al producir una solución, además describieron las propiedades de estos problemas a los que llamaron problemas de mínimos cuadrados totales no genéricos, y demostraron que con el uso de restricciones adicionales al espacio solución se puede encontrar la solución general. Este método era semejante al estudiado por Gleser (Gleser, 1981) en la resolución de los problemas de errores en variables multivariado. El método de resolución de Gleser se basa en el análisis de valores y vectores propios, a diferencia que en el método de mínimos cuadrados totales se utiliza la descomposición de valores singulares que es un método más robusto en el sentido de su implementación algorítmica.

En los campos de ingeniería, por ejemplo, en el análisis experimental modal de estructuras, la técnica de mínimos cuadrados totales más comúnmente conocida como la técnica H_v se introdujo hace unos 20 años por Leuridan (Golub & Van Loan, 1980). En el campo de los sistemas de identificación, Levin fue el primero que estudió el problema. Su método, llamado método de vector propio o método Koopmans-Levin (De Groen, 1996),

resuelve el problema de mínimos cuadrados totales siempre que el problema tenga solución única.

Método de mínimos cuadrados compensados ha sido otro nombre que surge en esta área, Stoica (Golub & Van Loan, 1980) demostró que es asintóticamente equivalente al método de los mínimos cuadrados totales. Además, en el área de procesamiento de señales, el método de la norma mínima, introducido por Kumaresan y Tufts, mostró ser equivalente al de los mínimos cuadrados totales.

En el campo de la quimiometría, el método de los mínimos cuadrados totales está fuertemente relacionado con el método de las componentes principales presentado por Wentzell (Gleser, 1981).

En otro campo de la ciencia, la medicina, Deming (Markovsky & Van Huffel, 2007) diseñó un método de regresión que permite comparar dos procedimientos de medición de parámetros, en que se suponen las variables medidas con error (por ejemplo medición de un indicador sanguíneo), bajo ciertos supuestos también es un problema de mínimos cuadrados totales.

El papel clave del método de los mínimos cuadrados en el análisis de regresión lineal es el mismo que el del método de los mínimos cuadrados totales en los modelos de error en variables (Golub & Van Loan, 1980).

Sin embargo, hay una gran cantidad de confusión en los campos del análisis numérico y estadísticas sobre el principio de mínimos cuadrados totales y su relación con los modelos con error en variables. Las ventajas computacionales de los algoritmos de los mínimos cuadrados totales aún no se conocen en la comunidad estadística, mientras que el concepto de modelado con error en variables no penetra lo suficientemente en el campo de las matemáticas computacionales e ingeniería. Una descripción general dedicada a los mínimos cuadrados totales se presentan en los textos de Van Huffel y Vanderwalle (Van Huffel, 2004).

El objetivo de este trabajo es comparar los dos procedimientos de regresión, ordinaria y ortogonal o también llamada total, y mostrar, que de existir variables regresoras no fijas, el método de regresión ordinaria no es el adecuado para determinar los valores de la variable de respuesta a partir de las variables regresoras, más bien, el método de mínimos cuadrados totales o regresión ortogonal es el más indicado. Por lo tanto la hipótesis de este trabajo, que desea probar es:

El mejor desempeño del método de regresión ortogonal o total en comparación con el método tradicional de regresión ordinaria cuando el problema de regresión presenta variables no fijas, generalmente observacionales.

Para ello se ha utilizado como método de verificación, además de la respectiva derivación teórica analítica, la simulación de Montecarlo para la generación de los datos contemplando el hecho que se ha considerado a las variables como no fijas. Se mostrará además el comportamiento de los resultados para el coeficiente de la recta de regresión (β_2), y se han obtenido intervalos de confianza para el parámetro de estudio, tanto para el método de regresión ordinaria como el método de regresión ortogonal.

ÍNDICE DE CONTENIDOS

<u>CONTENIDO</u>	<u>PÁGINA</u>
Introducción	
1. MÉTODO DE REGRESIÓN ORTOGONAL (MÍNIMOS CUADRADOS TOTALES)	1
2. ANÁLISIS DEL MÉTODO DE MÍNIMOS CUADRADOS TOTALES (TLS) MULTIVARIABLE	13
3. ANÁLISIS DE LA SIMULACIÓN Y SUS RESULTADOS	25
3.1 ANÁLISIS DEL MODELO	25
3.2 CASOS DE LA SIMULACIÓN Y SUS RESULTADOS	30
CONCLUSIONES	49
REFERENCIAS BIBLIOGRAFICAS	51
ANEXOS	55

ÍNDICE DE FIGURAS

<u>FIGURA</u>	<u>PÁGINA</u>
1 MÉTODO DE REGRESIÓN ORTOGONAL (MÍNIMOS CUADRADOS TOTALES)	
1.1 Comparación entre mínimos cuadrados ordinarios y totales	1
1.1 Distancia en la regresión ortogonal	3
1.2 Vectores en la regresión ortogonal	3
1.3 Función de verosimilitud perfil del parámetro θ	5
 3 ANALISIS DE LA SIMULACION Y SUS RESULTADOS PROMEDIOS Y ERROR DE ESTIMACION PARA β_2	
3.2.1.1 Estimación parámetro (β_2) caso 1	31
3.2.1.2 Estimación parámetro (β_2) caso 1	31
3.2.1.3 Estimación parámetro (β_2) caso 1	32
3.2.2.1 Estimación parámetro (β_2) caso 2	34
3.2.2.2 Estimación parámetro (β_2) caso 2	35
3.2.2.3 Estimación parámetro (β_2) caso 2	35
3.2.3.1 Estimación parámetro (β_2) caso 3	37
3.2.3.2 Estimación parámetro (β_2) caso 3	37
3.2.3.3 Estimación parámetro (β_2) caso 3	38
3.2.4.1 Estimación parámetro (β_2) caso 4	40
3.2.4.2 Estimación parámetro (β_2) caso 4	40
3.2.4.3 Estimación parámetro (β_2) caso 4	41
3.2.5.1 Estimación parámetro (β_2) caso 5	43
3.2.5.2 Estimación parámetro (β_2) caso 5	43

3.2.5.3 Estimación parámetro (β_2) caso 5	44
3.2.6.1 Estimación parámetro (β_2) caso 6	46
3.2.6.2 Estimación parámetro (β_2) caso 6	47
3.2.6.3 Estimación parámetro (β_2) caso 6	47

ÍNDICE DE TABLAS

TABLA

PÁGINA

3. ANALISIS DE LA SIMULACION Y SUS RESULTADOS

PROMEDIOS Y ERROR DE ESTIMACION PARA β_2

3.1 CASO 1: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$	30
3.2 INTERVALO CONFIANZA CASO 1	32
3.3 CASO 2: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$	33
3.4 INTERVALO CONFIANZA CASO 2	35
3.5 CASO 3: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$	36
3.6 INTERVALO CONFIANZA CASO 3	38
3.7 CASO 4: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 4, \sigma_{\varepsilon_2}^2 = 4, \sigma_y^2 = 1$	39
3.8 INTERVALO CONFIANZA CASO 4	41
3.9 CASO 5: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 4, \sigma_{z_2}^2 = 4, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$	42
3.10 INTERVALO CONFIANZA CASO 5	44
3.11 CASO 6: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 4, \sigma_{z_2}^2 = 4, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 4$	45
3.12 INTERVALO CONFIANZA CASO 6	48

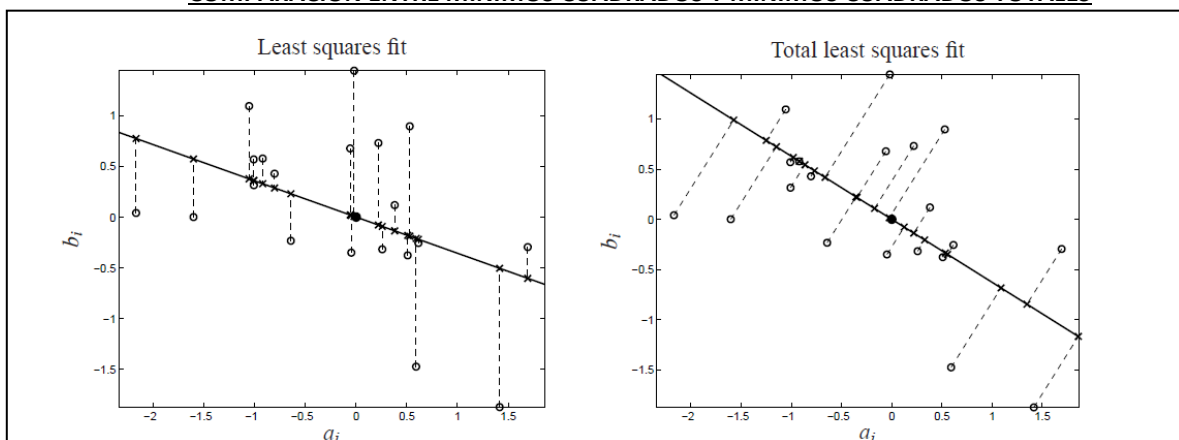
1. MÉTODO DE REGRESIÓN ORTOGONAL (MÍNIMOS CUADRADOS TOTALES)

En muchos problemas que se presentan en ciencias e ingeniería se obtiene un conjunto de datos asociados en parejas, y si se considera que, la relación entre las variables que representan los datos, es lineal, suele usualmente ajustarse mediante el método de los mínimos cuadrados, donde se busca estimar los parámetros de regresión, que en estos casos son los elementos de la recta.

Para este método se considera la desviación vertical entre la ordenada del dato y la estimación de la ordenada mediante la recta de regresión obtenida (Freund, Miller, & Miller, 2000), y que se considera es la más “cercana” o que mejor “ajuste” presenta.

En el método de regresión ortogonal (véase figura 1) se cambia la definición del error (distancia desde el punto observado a la recta de regresión), de tal modo que la distancia se mide de manera perpendicular del punto a la recta estimada.

FIGURA 1
COMPARACION ENTRE MINIMOS CUADRADOS Y MINIMOS CUADRADOS TOTALES



Fuente: Overview of total least squares methods

Elaboración: Markovsky & Van Huffel.

Supongamos que tenemos una recta con un vector paralelo $(1,m)$ y perpendicular al vector $(m,-1)$, siendo m la pendiente de la recta, además suponga que la ecuación de la recta estimada es:

$$y = mx + b$$

A continuación se presenta la derivación de los estimadores de regresión ortogonal según (Martin, 1998). Sea:

\vec{n} : el vector normal a la recta

(x_i, y_i) : un punto de observación (dato)

(x_o, y_o) : un punto de la recta por lo que satisface la ecuación de la recta (refiérase figura 2 y 3). Así se obtiene que:

$$d_i = \| \text{proy}_{\vec{n}} \vec{a} \| = \frac{(\vec{a} \cdot \vec{n})}{\|\vec{n}\|}$$

Se observa además que:

$$\vec{a} = (x_i - x_o, y_i - y_o)$$

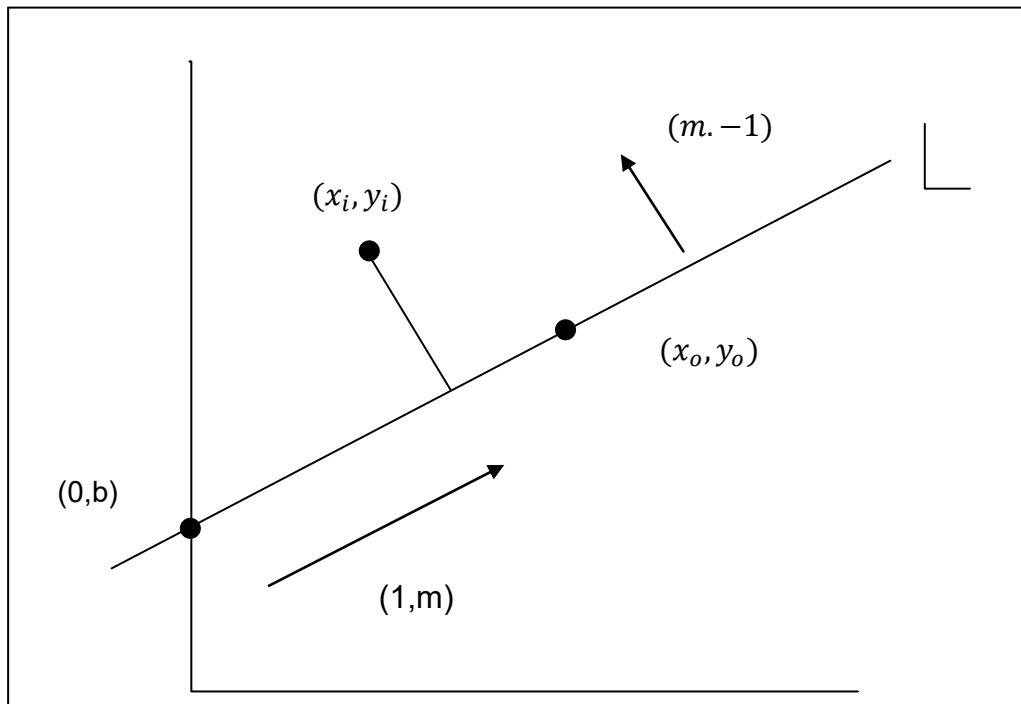
Entonces, reemplazando estos valores:

$$d_i = \frac{(x_i - x_o, y_i - y_o) \cdot (m, -1)}{\sqrt{m^2 + 1}}$$

y que resulta en:

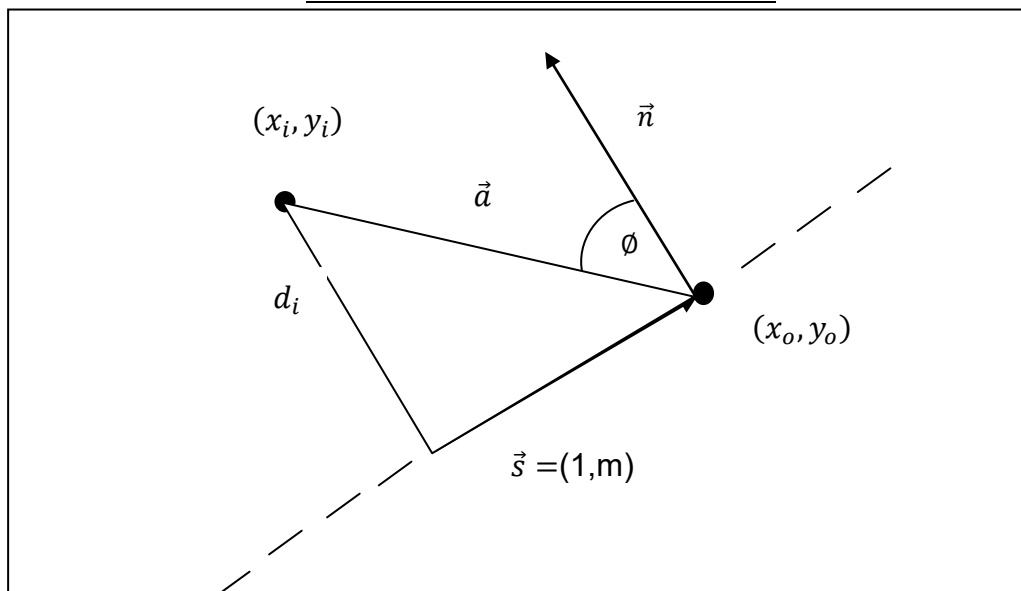
$$d_i = \frac{m(x_i - x_o) - (y_i - y_o)}{\sqrt{m^2 + 1}}$$

FIGURA 2
DISTANCIA EN LA REGRESION ORTOGONAL



Fuente: Casella & otros: Statistical Inference
Elaboración: autor.

FIGURA 3
VECTORES EN LA REGRESION ORTOGONAL



Fuente: Casella & otros: Statistical Inference
Elaboración: autor.

Como en el método de regresión lineal tradicional tomamos la distancia al cuadrado d_i^2 , para poder determinar la suma de los errores cuadráticos S , o de la línea de regresión del error total:

$$d_i^2 = \frac{(m(x_i - x_o) - (y_i - y_o))^2}{m^2 + 1}$$

Como se conoce el valor de b :

$$\begin{aligned} & \text{b} \\ & \text{-----} \\ & = \frac{(m x_i - y_i + (y_o - m x_o))^2}{m^2 + 1} \\ & = \frac{(m x_i - y_i + b)^2}{m^2 + 1} \end{aligned}$$

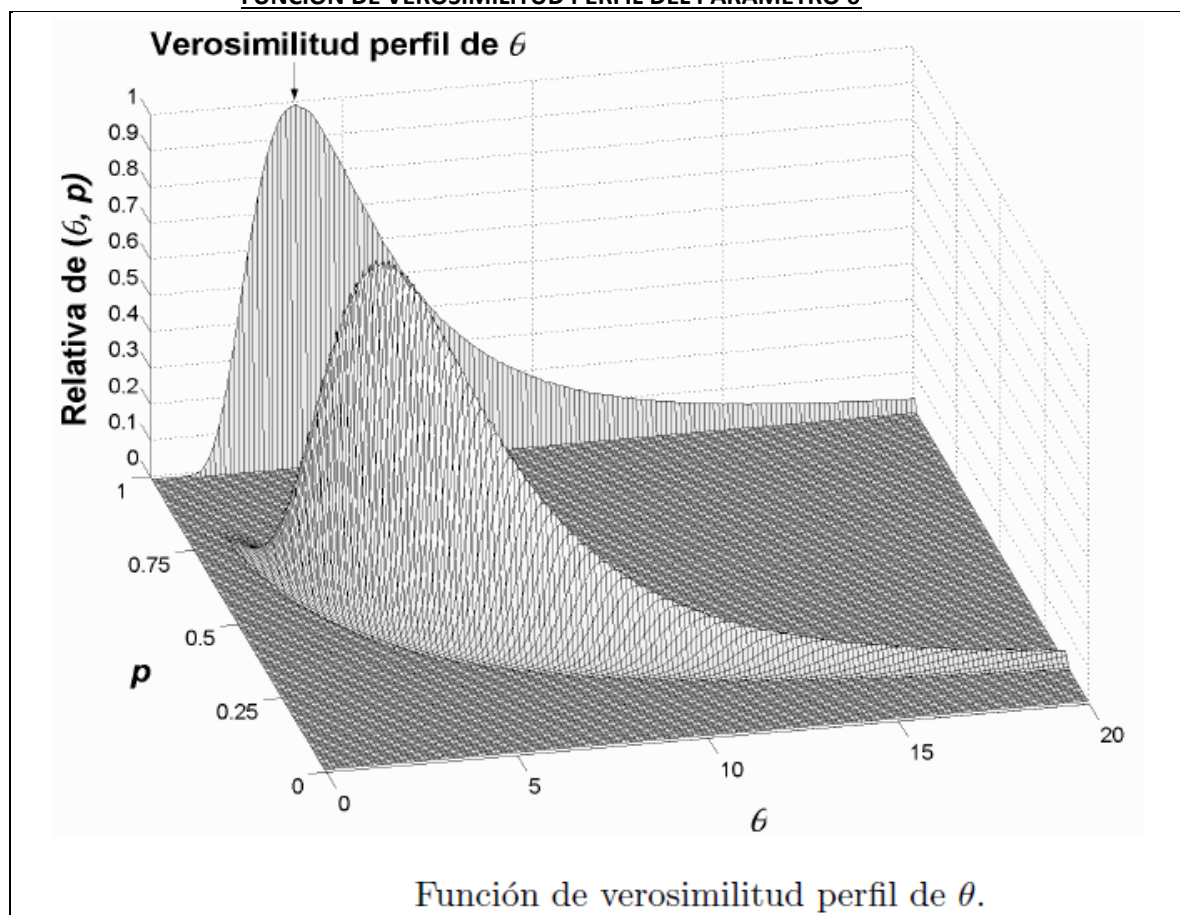
Así la suma de los errores se expresa como:

$$\begin{aligned} S &= \sum_{i=1}^n d_i^2 = \sum_{i=1}^n \left(\frac{(m x_i - y_i + b)^2}{m^2 + 1} \right) \\ &= \frac{1}{m^2 + 1} \sum_{i=1}^n (m x_i - y_i + b)^2 \end{aligned}$$

Para minimizar la suma de los errores se deriva en forma parcial respecto a m y b . Para ello se utiliza el procedimiento denominado de **la función de verosimilitud maximizada o perfil** el cual es un método estadístico muy simple y poderoso que sirve para estimar por separado un parámetro de interés en presencia de parámetros de estorbo (parámetros que no son de interés) (Davidson, 2014) y (Montoya, 2008). Aunque ya se había usado este método antes en la literatura estadística por Hood y Koopmans en el contexto de modelos econométricos denominándola ‘verosimilitud concentrada’ y Box y Cox utilizaron el método para ayuda en la selección del parámetro de forma de sus

modelos de transformaciones. Sin embargo, Sprott y Kalbfleisch, por vez primera le dieron el nombre de maximizada y la presentaron formalmente como un método general para eliminar parámetros de estorbo (parámetros que no son de interés al momento de maximizar una función objetivo) en Kalbfleisch y Sprott (Di Ciccio & Tibshirani, 1991). Para este propósito, la función de verosimilitud maximizada o perfil es mucho más general que otras verosimilitudes tales como la condicional, la marginal o la integrada.

FIGURA 4
FUNCION DE VEROSIMILITUD PERFIL DEL PARAMETRO θ



Fuente: Montoya Laos, José: La verosimilitud perfil en la Inferencia Estadística

Autor: Montoya Laos, José

La función de verosimilitud maximizada o perfil del parámetro de interés θ , $L_p(\theta)$, se define (Montoya, 2008), como:

$$L_p(\theta; y) = \max_{\lambda/\theta} L(\theta, \lambda; y) = L[\theta, \hat{\lambda}(\theta, y); y]$$

Donde $\hat{\lambda}(\theta, y)$ es el estimador de máxima verosimilitud restringido (emvr) de λ para un valor especificado de θ . El emvr $\hat{\lambda}(\theta, y)$ es el valor de λ que tiene mayor plausibilidad para ese valor fijo de θ dada la muestra observada y . Es decir, la verosimilitud perfil de θ , $L_p(\theta; y)$, se obtiene maximizando la función de verosimilitud $L(\theta, \lambda; y)$, sobre λ pero fijando θ .

Aplicando este procedimiento al problema planteado, se tiene:

Se fija el valor de m , así resulta derivando S respecto a b :

$$\frac{\partial S}{\partial b} = \frac{1}{m^2 + 1} \sum_{i=1}^n 2(mx_i - y_i + b)$$

Igualando a cero:

$$\frac{1}{m^2 + 1} \sum_{i=1}^n 2(mx_i - y_i + b) = 0$$

Lo que resulta en:

$$\sum_{i=1}^n mx_i - \sum_{i=1}^n y_i + nb = 0$$

Reescribiendo la ecuación apropiadamente:

$$b = -\frac{m}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n y_i$$

Finalmente:

$$b = \bar{y} - m\bar{x}$$

Verificando que se trata de un mínimo, se calcula la segunda derivada y se obtiene:

$$\frac{\partial^2 S}{\partial b^2} = \frac{1}{m^2 + 1} \sum_{i=1}^n 2$$

Resultado que es claramente positivo, y se concluye que es un mínimo.

Para el valor de m, siguiendo el procedimiento descrito:

$$S(m) = \min_b S(m, b) = S(m, b = \bar{y} - m\bar{x})$$

Remplazando en la ecuación de S:

$$S(m) = \frac{1}{m^2 + 1} \sum_{i=1}^n (mx_i - y_i + \bar{y} - m\bar{x})^2$$

Reescribiendo la ecuación, se obtiene:

$$= \frac{1}{m^2 + 1} \sum_{i=1}^n (m(x_i - \bar{x}) - (y_i - \bar{y}))^2$$

Resolviendo el cuadrado:

$$= \frac{1}{m^2 + 1} \sum_{i=1}^n [m^2(x_i - \bar{x})^2 - 2m(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2]$$

Para simplificar la expresión se opta por la notación:

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

Con lo que se escribe:

$$S(m) = \frac{1}{m^2 + 1} [m^2 S_{xx} - 2m S_{xy} + S_{yy}]$$

Ahora sí derivando S con respecto a m:

$$S'(m) = \frac{(m^2 + 1)(2m S_{xx} - 2S_{xy}) - 2m[m^2 S_{xx} - 2m S_{xy} + S_{yy}]}{(m^2 + 1)^2}$$

Realizando los respectivos productos, se tiene:

$$S'(m) = \frac{2m^3 S_{xx} + 2m S_{xx} - 2m^2 S_{xy} - 2S_{xy} - 2m^3 S_{xx} + 4m^2 S_{xy} - 2m S_{yy}}{(m^2 + 1)^2}$$

Reduciendo los términos semejantes, resulta en:

$$S'(m) = \frac{2m^2 S_{xy} - 2m(S_{yy} - S_{xx}) - 2S_{xy}}{(m^2 + 1)^2}$$

Igualando a cero la derivada y simplificando:

$$m^2 S_{xy} - m(S_{yy} - S_{xx}) - S_{xy} = 0$$

Como se observa la expresión representa una ecuación cuadrática en términos de la variable ***m*** y cuya solución es:

$$m = \frac{(S_{yy} - S_{xx}) \pm \sqrt{[-(S_{yy} - S_{xx})]^2 - 4S_{xy}(-S_{xy})}}{2S_{xy}}$$

Que finalmente resulta como:

$$m = \frac{S_{yy} - S_{xx} \pm \sqrt{(S_{yy} - S_{xx})^2 + 4 S_{xy}^2}}{2S_{xy}}$$

Puede observarse que al utilizar la regla general de la ecuación de segundo grado se obtienen dos posibles respuestas debido al doble signo, así que se debe verificar cuál de ellos corresponde al mínimo, para ello utilizando el criterio de la segunda derivada:

$$S''(m) = \frac{(m^2 + 1)^2 [4mS_{xy} - 2(S_{yy} - S_{xx})] - 2[m^2S_{xy} - m(S_{yy} - S_{xx}) - S_{xy}](2m)2(m^2 + 1)}{(m^2 + 1)^4}$$

Realizando las operaciones indicadas y simplificando los términos semejantes resulta que la segunda derivada de S con respecto a m es:

$$S''(m) = \frac{[4mS_{xy} - 2(S_{yy} - S_{xx})]}{(m^2 + 1)^2}$$

Para que se trate de un mínimo este resultado debe ser positivo (criterio de la segunda derivada) para el valor de **m** encontrado, es decir:

$$S''(m) > 0 \quad (\text{mínimo})$$

Así entonces, ya que el denominador es siempre positivo, se analiza el numerador:

$$4mS_{xy} - 2(S_{yy} - S_{xx}) > 0$$

Remplazando el valor de m ya calculado:

$$4 \left(\frac{S_{yy} - S_{xx} \pm \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}} \right) S_{xy} - 2(S_{yy} - S_{xx}) > 0$$

Lo cual simplificando resulta en:

$$2(S_{yy} - S_{xx}) \pm 2\sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2} - 2(S_{yy} - S_{xx}) > 0$$

Lo que finalmente queda:

$$\pm \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2} > 0$$

Lo que resulta tomar el valor con el signo positivo, para ser considerado un mínimo, el otro resultado (con signo negativo) resultaría el valor donde la función se convierte en un máximo.

Una propiedad interesante, que es conveniente presentar comparando la regresión ordinaria y la regresión ortogonal, es referente a la decisión que un investigador debe tomar cuando se le presenta un conjunto pareado de datos, requiriendo elegir la variable regresora de la que representará la variable de respuesta.

Esto muestra en el análisis de la regresión ordinaria, que las dos rectas son diferentes, esto es:

Cuando se toma:

$$y = mx + b$$

El valor estimado para la pendiente m es: $\widehat{m} = \frac{S_{xy}}{S_{xx}}$

Pero si se decide por:

$$x = m'y + b'$$

Se esperaría que la pendiente m' sea el recíproco del valor de m , es decir: $m' = \frac{1}{m}$.

Pero, $\widehat{m}' = \frac{S_{xy}}{S_{yy}}$. Con lo cual, se espera que el producto sea de valor uno, pero está claro que, el producto de las pendientes no es uno, es decir:

$$\widehat{m} \widehat{m}' = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

Lo cual corresponde al valor de R^2 , el coeficiente de determinación, el cual sólo sería uno si los puntos coincidieran todos en la misma recta. Al contrario analizando esto en la regresión ortogonal se obtiene:

Para:

$$y = mx + b$$

Se ha encontrado que el valor de la pendiente es:

$$m = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4 S_{xy}^2}}{2S_{xy}}$$

Mientras que para la recta donde se expresa:

$$x = m'y + b'$$

El valor de la pendiente viene dada por:

$$m' = \frac{S_{xx} - S_{yy} + \sqrt{(S_{xx} - S_{yy})^2 + 4 S_{xy}^2}}{2S_{xy}}$$

Así el producto de las dos pendientes m y m' tiene como valor uno. Esto es: $m m' = 1$.

Lo cual indica que no tiene importancia qué variable sea considerada la variable regresora ya que se obtendrá la misma recta de regresión.

2. ANÁLISIS DEL MÉTODO DE MÍNIMOS CUADRADOS TOTALES (TLS) MULTIVARIABLE

Aunque el nombre de mínimos cuadrados totales fue dado por Golub, (1980), el método originalmente fue interpretado como un procedimiento de resolución numérica, pero en el campo de la estadística es conocido como regresión ortogonal, método para error en variables e incluso método para error en la medida de las variables.

Esto ha originado un conflicto dentro de los investigadores ya que el problema de error en variables es bien conocido en el campo de las estadísticas, mientras los algoritmos desarrollados por Golub (1980) fueron dirigidos al campo del análisis numérico, sin análisis de carácter estadístico.

El método de mínimos cuadrados totales es una de las varias técnicas de estimación mediante el uso de parámetros lineales en el que se pretenden compensar los efectos de errores en las variables.

El algoritmo de resolución que se presenta en este trabajo está basado en la descomposición de valores singulares (Singular Value Decomposition) (SVD) (Van Huffel, 2004) en sus siglas inglesas, el cual es el más aceptado por su eficiencia, versatilidad y robustez. Desde un punto de vista práctico el método de mínimos cuadrados totales se considera una aplicación orientada o dirigida a casi la mayoría de aplicaciones ingenieriles o científicas donde los datos se ven contaminados por ruido.

En este aspecto el método de mínimos cuadrados totales como el modelado de problemas con error en la medición de las variables juegan un rol similar y las ventajas computacionales del método de mínimos cuadrados totales son muy desconocidos en la comunidad estadística, mientras que el concepto de modelo de errores en variables no ha

penetrado lo suficiente en el campo de las matemáticas computacionales y de la ingeniería.

En los modelos de regresión lineal múltiple ($Ax = b$) un supuesto en el que se enmarca, es el que establece que todos los errores son confinados al vector de respuesta b o y .

Desafortunadamente este supuesto frecuentemente no es real, ya que en la medición pueden aparecer errores de muestreo, humanos, de modelado, instrumentales, entre estos que influyen sobre la matriz de los datos.

El problema en la regresión ordinaria puede plantearse como:

$$\text{Minimizar } \|\underline{r}\|_2$$

$$\text{Sujeto a: } (\underline{b} + \underline{r}) \in \text{imagen}(A)$$

Si $\|\underline{r}\|_2$ es mínimo y $\underline{b} + \underline{r} = A\underline{x}$, siendo $A \in M_{m \times n}$, $\underline{b} \in \mathbb{R}^m$, entonces $\underline{x} \in \mathbb{R}^n$ es la solución del problema de mínimos cuadrados. Esto significa que las perturbaciones mínimas sobre el vector de respuesta b en una cantidad r (vector de error) pueden ser “predichas” mediante las columnas de la matriz A .

De aquí es más sencillo observar la idea detrás del método de mínimos cuadrados totales, ahora hay que considerar las perturbaciones tanto en el vector de respuesta b , como en la matriz de datos A , esto es:

$$\text{Minimizar } \|[\tilde{E}|\tilde{r}]\|_F$$

$$E, \underline{r}$$

$$\text{Sujeto a: } (\underline{b} + \underline{r}) \in \text{Imagen}(A + E)$$

Una vez que $[\tilde{E}|\tilde{r}]$ (matriz de error) ha sido encontrado y es el valor mínimo, entonces cualquier valor x que satisfaga:

$$(A + \tilde{E})\underline{x} = \underline{b} + \underline{\tilde{r}}$$

Se dice que es una solución del problema de mínimos cuadrados totales.

Se observa que el problema de mínimos cuadrados totales es equivalente a resolver un problema cercanamente compatible con mínimos cuadrados ordinarios en el que se desea:

$$\text{Minimizar } \|(A + \tilde{E})\underline{\tilde{x}} - (b + \underline{\tilde{r}})\|_2$$

Aquí se presenta la solución del problema de mínimos cuadrados totales haciendo uso de la descomposición de valores singulares (SVD) propuesto por Golub (1980).

Tal como se indicó anteriormente si $(\underline{b} + \underline{r}) \in \text{Im}(A + E)$ entonces hay un vector $\underline{x} \in R^n$ tal que:

$$(A + E)\underline{x} = \underline{b} + \underline{r}$$

Esto es:

$$(A + E)\underline{x} - (\underline{b} + \underline{r}) = \mathbb{O}$$

O escrito más convenientemente como:

$$\begin{bmatrix} A + E & \underline{b} + \underline{r} \end{bmatrix} \begin{bmatrix} \underline{x} \\ -1 \end{bmatrix} = \mathbb{O}$$

Desde el punto de vista de los sistemas de ecuaciones lineales:

$$\begin{bmatrix} A | \underline{b} + E | \underline{r} \end{bmatrix} \begin{bmatrix} \underline{x} \\ -1 \end{bmatrix} = \mathbb{O}$$

Esta ecuación muestra que el problema de mínimos cuadrados totales consistirá en encontrar una matriz de perturbación o de error $\Delta \in M_{m \times (n+1)}$, $\Delta = E | \underline{r}$, siendo de norma mínima, tal que $C + \Delta$ sea de rango deficiente, es decir de rango n , donde:

$$C = \begin{bmatrix} A | \underline{b} \end{bmatrix}$$

El método de descomposición de valores singulares es usado para este propósito. Entonces, cualquier matriz C se puede descomponer como:

$$U^T C V = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{n+1})$$

Siendo $U = [\underline{u}_1, \underline{u}_2, \dots, \underline{u}_m]$; $V = [\underline{v}_1, \underline{v}_2, \dots, \underline{v}_{n+1}]$; $\underline{u}_i \in \mathbb{R}^m$; $\underline{v}_j \in \mathbb{R}^{n+1}$

Además $\sigma_1 \geq \sigma_2 \dots \geq \sigma_k \geq \sigma_{k+1} \geq \dots \geq \sigma_n \geq \sigma_{n+1} \geq 0$

Son los valores singulares (raíces positivas de los valores propios de la matriz) en orden decreciente de magnitud, además sea la descomposición de valor singular de C con:

$$U^T U = I_m \quad y \quad V^T V = I_{n+1}$$

Bajo este criterio la solución básica del problema de mínimos cuadrados totales se establece como:

Calculando la descomposición de valores singulares de $[A|\underline{b}]$ y $\sigma_{\min}(A)$ el más pequeño valor singular de A. Si $\sigma_{\min}(A) > \sigma_{n+1}$, acorde al teorema de Schmidt – Eckart – Young – Mirsky (Golub & Van Loan, 1980), la mejor aproximación de $[\hat{A}|\hat{\underline{b}}]$ de rango n se obtiene cuando el más pequeño de los valores singulares se hace cero, es decir, $\sigma_{n+1} = 0$, entonces:

$$[\Delta \hat{A} \quad \Delta \hat{\underline{b}}] = [A \quad \underline{b}] - [\hat{A}|\hat{\underline{b}}] = \sigma_{n+1} \underline{u}_{n+1} \underline{v}_{n+1}^T$$

Con $[\hat{A}|\hat{\underline{b}}] = U \hat{\Sigma} V^T$ siendo $\hat{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_n, 0)$ y la solución del problema de mínimos cuadrados totales, existe, es única y viene dada por:

$$\hat{\underline{x}} = -\frac{1}{v_{n+1,n+1}} [v_{1,n+1}, \dots, v_{n,n+1}]^T$$

La solución aquí presentada esta obtenida bajo la siguiente consideración: (De Groen, 1996). Dadas las siguientes matrices: $A \in \mathbb{R}^{m \times n}$ ($m \geq n + p$) y $B \in \mathbb{R}^{m \times p}$ encontrar $X \in \mathbb{R}^{n \times p}$, que resuelve el sistema de ecuaciones $AX = B$ bajo la perspectiva

de los mínimos cuadrados totales. Por analogía, se busca encontrar la solución X de una ecuación matricial consistente (es decir que tiene solución):

$A'X = B'$ (es decir $\text{Im}(B') \subset \text{Im}(A')$) más cercana a $AX = B$; y por ello se tiene que minimizar:

$$\|A - A'\|_F^2 + \|B - B'\|_F^2 \text{ sujeto a } A' \in \mathbb{R}^{m \times n}, B' \in \mathbb{R}^{m \times p}, \text{ y } A'X = B'.$$

Dicho de otro modo, encontrar una aproximación S :

$$S = (A' | B') \in \mathbb{R}^{m \times (n+p)} \text{ de } (A | B), \text{ de manera que}$$

$$\|(A | B) - S\|_F^2 \text{ es mínima sujeto a } \text{rango}(S) = n$$

La solución se realiza haciendo la descomposición (SVD) de $(A | B)$, expresada en bloques:

$$(A | B) = U \Sigma V^T = \begin{pmatrix} (m \times n) & | & (m \times p) \\ U_1 & & U_2 \end{pmatrix} \begin{pmatrix} (n \times n) & 0 \\ \Sigma_1 & \Sigma_2 \\ 0 & (p \times p) \end{pmatrix} \begin{pmatrix} (n \times n) & (n \times p) \\ V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \\ (p \times n) & (p \times p) \end{pmatrix}^T$$

Manteniendo el supuesto que $\sigma_n > \sigma_{n+1}$ se deduce, que la mejor aproximación de $(A | B)$

de rango n en la norma de Frobenius viene dado por:

$$S := (U_1 | U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{pmatrix}^T = U_1 \Sigma_1 (V_{1,1}^T | V_{2,1}^T) = (A' | B'),$$

Donde $A' := U_1 \Sigma_1 V_{1,1}^T$ y $B' := U_1 \Sigma_1 V_{2,1}^T$. La ortogonalidad de las columnas de V implica:

$$\begin{pmatrix} V_{1,1} \\ V_{2,1} \end{pmatrix}^T \begin{pmatrix} V_{1,2} \\ V_{2,2} \end{pmatrix} = \mathbb{O} \text{ por lo tanto } S \begin{pmatrix} V_{1,2} \\ V_{2,2} \end{pmatrix} = A' V_{1,2} + B' V_{2,2} = \mathbb{O}$$

Así se obtiene:

$$A' V_{1,2} = -B' V_{2,2}$$

Y luego:

$$A'(-V_{1,2} V_{2,2}^{-1}) = B'$$

Bajo el supuesto de $\text{rango}(V_{2,2}) = p$, se puede concluir que: $X := -V_{1,2} V_{2,2}^{-1}$

resuelve la ecuación aproximada $A'X = B'$.

A continuación se muestra el algoritmo para la resolución básica del problema de mínimos cuadrados totales, basado en la demostración anterior, con $p = 1$:

Algoritmo:

Dado $A \in \mathbb{R}^{m \times n}$; $\underline{b} \in \mathbb{R}^m$

Paso 1: Calcule la descomposición de valores singulares de:

$$[A|\underline{b}] = U\Sigma V^T$$

Paso 2: Si $v_{n+1,n+1} \neq 0$

Entonces:

$$\underline{\hat{x}} = -\frac{1}{v_{n+1,n+1}} [v_{1,n+1}, \dots, v_{n,n+1}]^T \quad (1)$$

Se puede fácilmente observar que para el caso univariado ($n = 1$), se obtiene:

$$\hat{x}_1 = -\frac{v_{12}}{v_{22}}$$

Y luego se obtiene el valor de \hat{x}_0 (una vez calculado \hat{x}_1).

En el algoritmo mostrado se hace uso de la descomposición de valores singulares de una matriz, y la idea radica en la necesidad de aproximar una matriz X de rango p , por otra matriz \widehat{X}_r de rango $r < p$, en general en $p - 1$, la aproximación óptima es, tal como se muestra en (Peña, 2002):

$$\widehat{X}_r = X A_r A_r^T$$

Donde la matriz es A_r es de tamaño $p \times r$ y sus columnas son los vectores propios asociados a los r mayores valores propios de la matriz S (matriz de varianzas y covarianzas de X suponiendo el valor de la media igual a cero).

Este problema puede considerarse así:

Considere un espacio de dimensión r definido por una base U_r ortonormal donde U_r es de tamaño $p \times r$ y $U_r^T U_r = I$.

Se desea encontrar una aproximación de la matriz X utilizando esta base, esto es, se quiere “predecir” cada una de las filas (x_1, \dots, x_n) de la matriz, donde x_i es un vector $p \times 1$ de observaciones en el elemento i de la muestra, mediante los vectores U_r . La predicción de la variable \underline{x}_i será la proyección ortogonal sobre el espacio generado por estos vectores que es:

$$\hat{\underline{x}}_i = U_r U_r^T \underline{x}_i$$

Y se quiere determinar los vectores U_r tal que el error cuadrático de estas predicciones sea mínimo. El error cuadrático E para los elementos de la matriz X viene dado por:

$$E = \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \hat{x}_{ij})^2 = \sum_{i=1}^n (\underline{x}_i - \hat{\underline{x}}_i)^T (\underline{x}_i - \hat{\underline{x}}_i)$$

Lo cual se requiere sea mínimo, esta expresión puede escribirse como:

$$E = \sum_{i=1}^n \underline{x}_i^T \underline{x}_i - \sum_{i=1}^n \underline{x}_i^T U_r U_r^T \underline{x}_i$$

Minimizar el error equivale a maximizar el segundo término de esta expresión.

Sabiendo que un escalar es igual a su traza:

$$\sum_{i=1}^n \underline{x}_i^T U_r U_r^T \underline{x}_i = \text{tr} \left(\sum_{i=1}^n \underline{x}_i^T U_r U_r^T \underline{x}_i \right)$$

Reescribiendo la expresión:

$$= \sum_{i=1}^n \text{tr} (U_r U_r^T \underline{x}_i \underline{x}_i^T)$$

Lo cual es equivalente a:

$$= \text{tr} \left(U_r U_r^T \sum_{i=1}^n \underline{x}_i \underline{x}_i^T \right)$$

Aquí se introduce:

$$S = \frac{\sum_{i=1}^n \underline{x}_i \underline{x}_i^T}{n}$$

Sustituyendo, resulta:

$$n \text{tr}(U_r U_r^T S) = n \text{tr}(U_r^T S U_r)$$

Por lo que, finalmente resulta en:

$$\sum_{i=1}^n \underline{x}_i^T U_r U_r^T \underline{x}_i = n \text{tr}(U_r^T S U_r)$$

Como se observa en esta última expresión, minimizar el error implica encontrar un conjunto de vectores que maximicen la suma de los elementos diagonales de $U_r^T S U_r$, es decir que maximicen:

$$\sum_{j=1}^r u_j^T S u_j$$

Suponiendo que: $U_r = A_r$ y la aproximación óptima a la matriz X vendrá por:

$$\hat{\underline{x}}_r = X A_r A_r^t$$

Ahora el primer término de la ecuación del error cuadrático:

$$\begin{aligned}\sum_{i=1}^n \underline{x}_i^T \underline{x}_i &= \text{tr} \left(\sum_{i=1}^n \underline{x}_i^T \underline{x}_i \right) = \sum_{i=1}^n \text{tr} (\underline{x}_i^T \underline{x}_i) \\ &= \text{tr} \sum_{i=1}^n (\underline{x}_i \underline{x}_i^T) = n \text{tr} (S) = n \sum_{i=1}^p \lambda_i\end{aligned}$$

Y ya se obtuvo el resultado del segundo término, por lo que:

$$E = n \sum_{i=1}^p \lambda_i - n \sum_{i=1}^r \lambda_i = \sum_{i=r+1}^p \lambda_i$$

Lo cual es proporcionado por la descomposición de valores singulares de una matriz, es decir la mejor aproximación a la matriz X por otra matriz \hat{X}_r de rango $r < p$ es:

$$\hat{\underline{x}}_r = U_r D_r^{1/2} V^T = \sum_{i=1}^r \lambda_i \underline{u}_i \underline{v}_i^T$$

Donde:

U_r : Matriz de los r mayores vectores propios de XX^T

$D_r^{1/2}$: Contiene los r mayores valores singulares

V_r : Contiene los r vectores propios de $X^T X$

λ_i : valores propios de la matriz S matriz de varianzas y covarianzas

3. ANÁLISIS DE LA SIMULACIÓN Y SUS RESULTADOS

3.1 ANÁLISIS DEL MODELO

Para el análisis se supondrá el modelo lineal en que se tiene una variable de respuesta w con dos variables regresoras z_1 y z_2 , pero estas variables regresoras, además de la de respuesta w , no son las realmente medidas sino más bien una variable observada de respuesta y , las variables regresoras x_1 además de x_2 , las cuales son variables aleatorias que representan a las variables originales una vez que se ha introducido el error, el cual se supone normal con media cero y varianza $\sigma_y^2, \sigma_{z_1}^2, \sigma_{z_2}^2$, respectivamente. Esto es:

$$w = \beta_0 + \beta_1 z_1 + \beta_2 z_2$$

$$\text{Dónde:} \quad x_1 = z_1 + \varepsilon_1 \quad x_2 = z_2 + \varepsilon_2 \quad y = w + \varepsilon_y$$

Y además se supondrá que:

$$\begin{aligned} z_1 &\sim N(\mu_1, \sigma_{z_1}^2) \\ z_2 &\sim N(\mu_2, \sigma_{z_2}^2) \\ \varepsilon_1 &\sim N(0, \sigma_{\varepsilon_1}^2) \\ \varepsilon_2 &\sim N(0, \sigma_{\varepsilon_2}^2) \\ \varepsilon_y &\sim N(0, \sigma_y^2) \end{aligned}$$

Es decir, en forma matricial:

$$\begin{bmatrix} z_1 \\ z_2 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_y \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{z_1}^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{z_2}^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon_1}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon_2}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_y^2 \end{bmatrix} \right)$$

Así definidas las variables y el modelo a trabajar, se procede a calcular el valor esperado de y dado los valores de x_1 como de x_2 . Escribiendo en forma vectorial se tiene:

$$\begin{bmatrix} y \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \beta_1 & \beta_2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_y \end{bmatrix} + \begin{bmatrix} \beta_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

De aquí se obtiene el valor esperado y la varianza de este vector de variables:

$$E \begin{bmatrix} y \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1\mu_1 + \beta_2\mu_2 \\ \mu_1 \\ \mu_2 \end{bmatrix}$$

Cuya varianza se obtiene y resulta:

$$\text{var} \begin{pmatrix} y \\ x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} \beta_1\sigma_{z_1}^2 & \beta_2\sigma_{z_2}^2 & 0 & 0 & \sigma_y^2 \\ \sigma_{z_1}^2 & 0 & \sigma_{\varepsilon_1}^2 & 0 & 0 \\ 0 & \sigma_{z_2}^2 & 0 & \sigma_{\varepsilon_2}^2 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 & 1 & 0 \\ \beta_2 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Realizando la multiplicación nos queda:

$$= \begin{bmatrix} \beta_1^2\sigma_{z_1}^2 + \beta_2^2\sigma_{z_2}^2 + \sigma_y^2 & \beta_1\sigma_{z_1}^2 & \beta_2\sigma_{z_2}^2 \\ \beta_1\sigma_{z_1}^2 & \sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2 & 0 \\ \beta_2\sigma_{z_2}^2 & 0 & \sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2 \end{bmatrix}$$

Siguiendo con el análisis, se calcula el valor esperado y la varianza de la variable y dado el vector $\underline{x} = (x_1, x_2)$, tomando los resultados dados por Eaton, (Eaton, 2007), los cuales son, para:

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{con tamaño} \quad \begin{pmatrix} q \times 1 \\ (N-q) \times 1 \end{pmatrix}$$

y

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{con tamaño} \quad \begin{pmatrix} q \times q & q \times (N-q) \\ (N-q) \times q & (N-q) \times (N-q) \end{pmatrix}$$

Siendo $\underline{\mu}$ un vector de medias y Σ la matriz de varianzas y covarianzas condicionadas, entonces:

$$E(y | \underline{x}) = \mu_y + \Sigma_{12}\Sigma_{22}^{-1}(\underline{x} - \underline{\mu})$$

$$var(y | \underline{x}) = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Aplicando estos resultados generales al problema en cuestión se tiene:

$$E\left[y / \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right] = \beta_0 + \beta_1\mu_1 + \beta_2\mu_2 + \begin{bmatrix} \beta_1\sigma_{z_1}^2 & \beta_2\sigma_{z_2}^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} & 0 \\ 0 & \frac{1}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

Realizando el primer producto matricial queda:

$$= \beta_0 + \beta_1\mu_1 + \beta_2\mu_2 + \begin{bmatrix} \frac{\beta_1\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} & \frac{\beta_2\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}$$

Siguiendo con el segundo producto matricial, resulta:

$$= \beta_0 + \beta_1\mu_1 + \beta_2\mu_2 - \beta_1\mu_1 \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} - \beta_2\mu_2 \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} \beta_1 x_1 \\ + \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} \beta_2 x_2$$

Reescrito de manera conveniente, finalmente se expresa el valor esperado como:

$$E(y | \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}) = \beta_0 + \beta_1\mu_1 \left(1 - \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2}\right) + \beta_2\mu_2 \left(1 - \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2}\right) + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} \beta_1 x_1 \\ + \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} \beta_2 x_2$$

Para el resultado de la varianza se procede de manera similar, entonces:

$$var\left(y/\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \sigma_y^2 + \beta_1^2 \sigma_{z_1}^2 + \beta_2^2 \sigma_{z_2}^2 - \begin{bmatrix} \beta_1 \sigma_{z_1}^2 & \beta_2 \sigma_{z_2}^2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} & 0 \\ 0 & \frac{1}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} \end{bmatrix} \begin{bmatrix} \beta_1 \sigma_{z_1}^2 \\ \beta_2 \sigma_{z_2}^2 \end{bmatrix}$$

Realizando los productos matriciales, resulta:

$$= \sigma_y^2 + \beta_1^2 \sigma_{z_1}^2 + \beta_2^2 \sigma_{z_2}^2 - \beta_1^2 \sigma_{z_1}^2 \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} - \beta_2^2 \sigma_{z_2}^2 \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2}$$

Finalmente escribiendo convenientemente la expresión se muestra como:

$$var\left(y/\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \sigma_y^2 + \beta_1^2 \sigma_{z_1}^2 \left(1 - \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2}\right) + \beta_2^2 \sigma_{z_2}^2 \left(1 - \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2}\right)$$

Un resultado interesante que permite observar muy rápidamente el comportamiento de los estimadores de regresión lineal ordinarios y ortogonales es si se supone los siguientes parámetros: $\mu_1 = \mu_2 = 0$; $\beta_0 = 0$, $\beta_1 = \beta_2 = 1$, lo que resulta reemplazando en la expresión del valor esperado como:

$$E(y/(x_1, x_2)) = \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{\varepsilon_1}^2} x_1 + \frac{\sigma_{z_2}^2}{\sigma_{z_2}^2 + \sigma_{\varepsilon_2}^2} x_2$$

Como se puede observar en el resultado obtenido los valores estimados de los regresores toman un sesgo cuando se analizan como resultado de la regresión ordinaria, puede suponerse que $\sigma_{\varepsilon_1}^2$ y $\sigma_{\varepsilon_2}^2$ son diferentes de cero con lo que los valores estimados de los regresores no resultarían en el valor de uno como se supuso. Más adelante se presentan simulaciones y sus resultados comparativos para la regresión ortogonal y ordinaria.

En el anexo se muestra el código en R, del programa que genera los datos a partir de los cuales se obtiene los estimadores tanto de regresión ordinaria y de regresión ortogonal, la variable *Nsimul* indica el número de corridas que se realizarán, mientras que *n* es el número de valores que se generan a partir de los cuales se obtiene los resultados que se muestran en las tablas a continuación.

En el programa se compara los valores de los coeficientes tanto obtenidos por el método de regresión ordinaria como los obtenidos por el método de regresión ortogonal. Es conveniente hacer notar que se utilizó la función de R “trim” que permite calcular la media acotada de los datos, ya que en algunas ocasiones durante la obtención de los resultados en la regresión ortogonal, en las diversas corridas se producían valores extremos que se alejaban demasiado del valor de la media produciendo un error cuadrático grande para el caso de *n* pequeño (*n*=10) en comparación con los resultados cuando *n* es mayor, esto se debe a la característica de cálculo que se muestra en el algoritmo en la ecuación (1) de la página 20. Como se aprecia, si el valor $v_{n+1,n+1}$ toma valores cercanos a cero, el resultado para la estimación de los elementos del vector \underline{x} se torna muy grande provocando un cierto grado de inestabilidad en los resultados.

A manera de aplicación práctica se presentan 6 casos o ejemplos de la aplicación del modelo de regresión ortogonal y ordinaria, en la que compara los resultados y la medida del error que se produce respecto al valor original utilizado.

Como se indicó al inicio del trabajo el método de verificación de los resultados se lo hace utilizando la simulación de Montecarlo, se ha elegido este procedimiento ya que en definitiva se está realizando un muestreo aleatorio de variables aleatorias y que para los fines del trabajo se ha supuesto de distribución de probabilidades normales e independientes. Por esta razón, acorde a lo que menciona Otamendi el método de Montecarlo “es un método numérico para resolver problemas matemáticos mediante el muestreo aleatorio de variables aleatorias” (Otamendi, 2006).

3.2 CASOS DE LA SIMULACIÓN Y SUS RESULTADOS:

TABLA 1

PROMEDIOS Y ERROR DE ESTIMACION PARA β_2

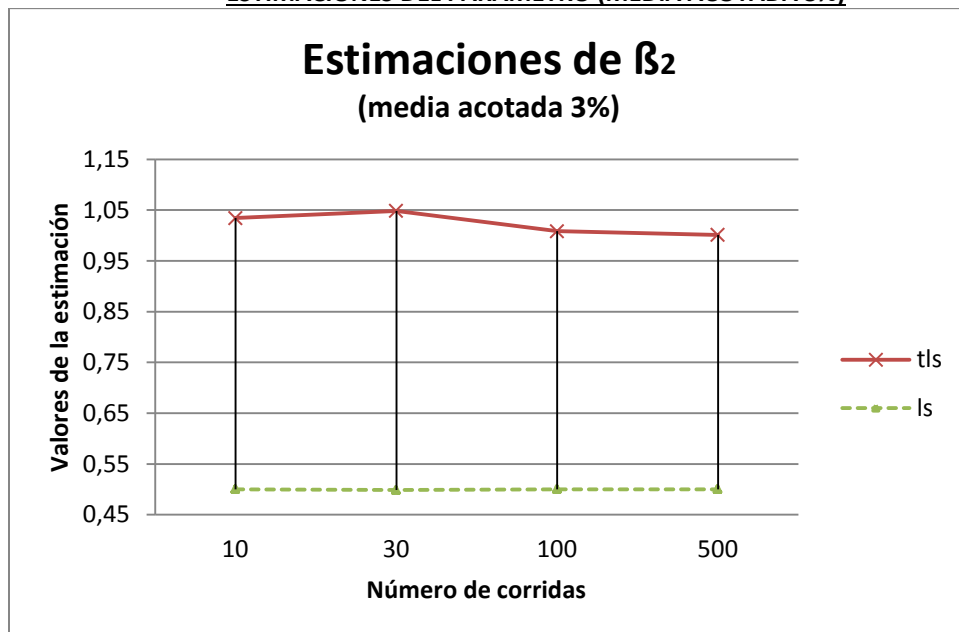
CASO 1: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$
 $\beta_0 = 0, \beta_1 = 1, \beta_2 = 1$

CASO 1:					
Nsimul = 10000					
	MEDIA ARITMETICA			ERROR CUADRATICO	
	β_2		% MEDIA	MEDIO	
	TLS	LS	ACOTADA	TLS	LS
n = 10	1,03420	0,50016	3	3,89596	0,36754
	1,02877	0,50564	5	2,28191	0,35314
	0,99314	0,49335	8	1,40568	0,33689
n = 30	1,04854	0,49842	3	0,30301	0,27901
	1,03673	0,50229	5	0,23341	0,27082
	1,02658	0,49956	8	0,19602	0,26906
n = 100	1,00845	0,50021	3	0,04865	0,25742
	1,01051	0,50183	5	0,04386	0,25467
	1,01031	0,50091	8	0,04122	0,25423
n = 500	1,00111	0,50001	3	0,00891	0,25144
	0,99945	0,49941	5	0,00808	0,25179
	1,00091	0,49962	8	0,00754	0,25141

Elaboración: Autor.

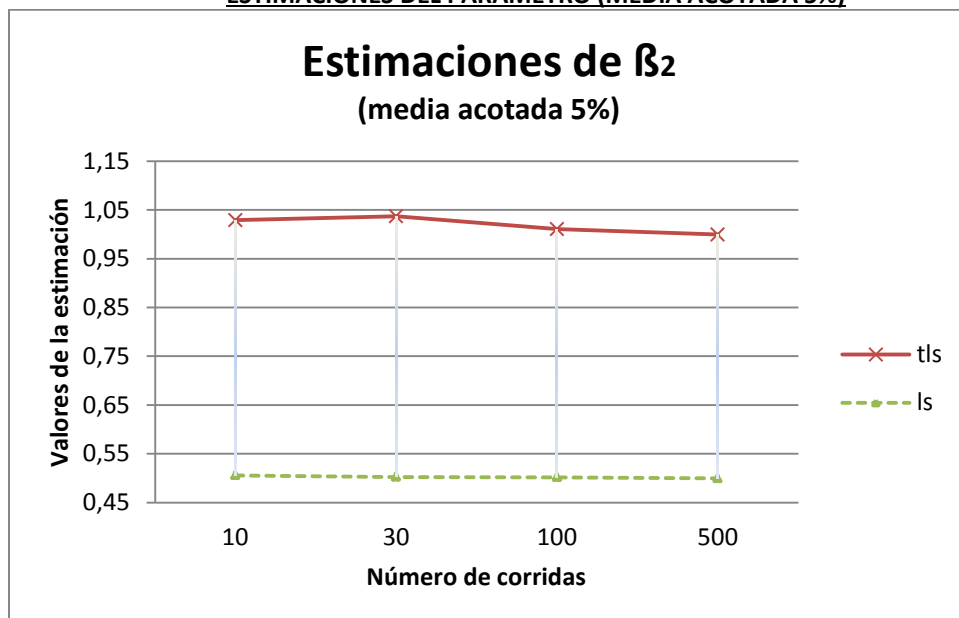
En este primer caso (tabla 1) se puede apreciar que el estimador de regresión ortogonal presenta mejor característica en la estimación de los coeficientes de las variables regresoras, además se aprecia el sesgo que involucra a los estimadores de los coeficientes de regresión ordinaria, también se puede indicar que con pocos datos ya se puede hacer una estimación muy buena en regresión ortogonal.

FIGURA 5
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 3%)



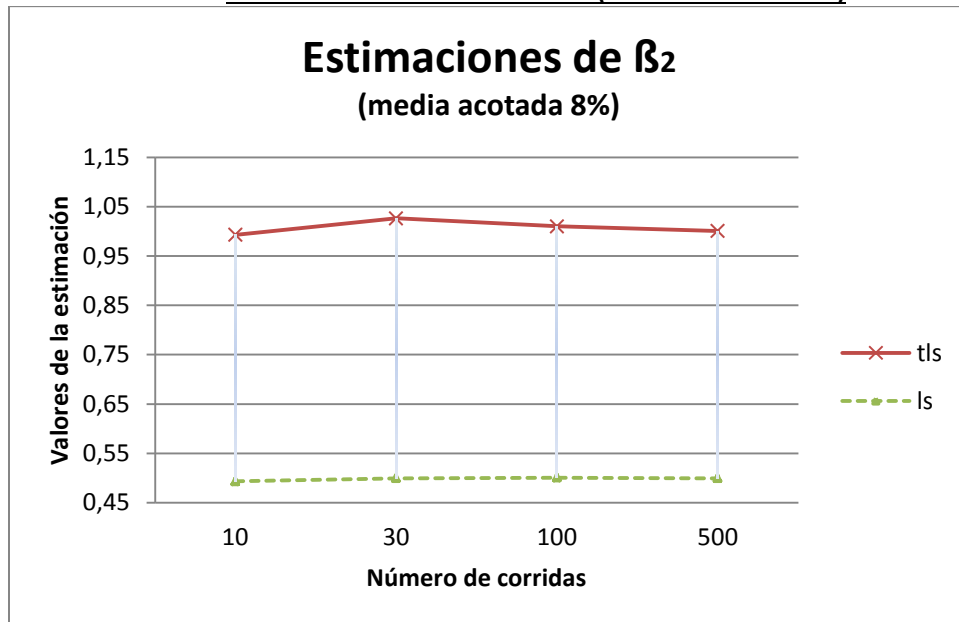
Elaboración: Autor.

FIGURA 6
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 5%)



Elaboración: Autor.

FIGURA 7
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 8%)



Elaboración: Autor.

Como se muestra en las figuras 5,6 y 7, el comportamiento del parámetro (β_2) presenta ligeras variaciones para el tamaño de muestra $n = 10$, pero a medida que se aumenta el tamaño de la muestra de valores generados se estabiliza hacia el valor previsto del parámetro en la simulación.

TABLA 2
INTERVALO CONFIANZA PARAMETRO (β_2)

CASO 1: Nsimul = 10000				
	INTERVALO DE CONFIANZA MEDIA			
	T.L.S		L.S.	
	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR	LIMITE SUPERIOR
n = 10	-3,33990	5,22651	0,49869	0,51458
n = 30	0,93600	1,11600	0,49684	0,50449
n = 100	0,97997	1,01789	0,49681	0,50082
n = 500	0,98765	1,00564	0,49910	0,50084

Elaboración: Autor.

En la tabla 2 se muestran además los intervalos de confianza del parámetro (β_2), para los diferentes valores del tamaño de muestra y para los dos métodos bajo análisis, se observa que para $n = 10$ el intervalo de confianza para (β_2) presenta en el método de regresión ortogonal tiene una mayor longitud esto como consecuencia de la alta variabilidad de los resultados a pocos valores de n . En cambio, para el método de regresión ordinaria el comportamiento es muy estable, pero como se mostró en la tabla 1, el resultado se encuentra sesgado y por ello erróneo.

TABLA 3
PROMEDIOS Y ERROR DE ESTIMACION PARA β_2

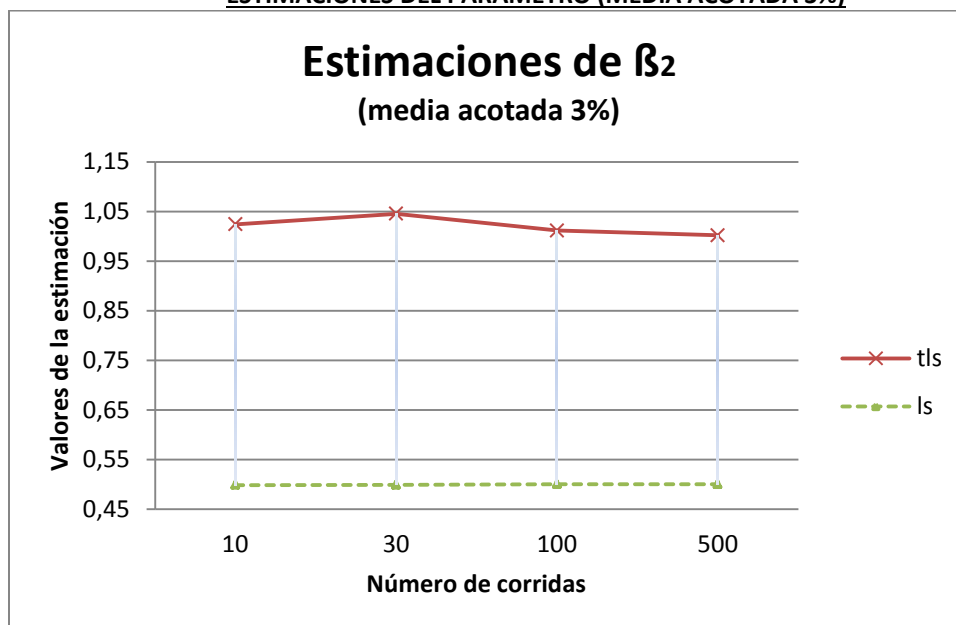
CASO 2: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$
 $\beta_0 = 1, \beta_1 = 1, \beta_2 = 1$

CASO 2:					
Nsimul = 10000					
	MEDIA ARITMETICA			ERROR CUADRATICO	
	β_2		% MEDIA	MEDIO	
	TLS	LS	ACOTADA	TLS	LS
n = 10	1,02432	0,49863	3	4,01124	0,37174
	1,03914	0,49968	5	2,13726	0,35213
	1,00635	0,49841	8	1,36104	0,33492
n = 30	1,04562	0,49917	3	0,29015	0,27792
	1,04064	0,50121	5	0,23111	0,27184
	1,02373	0,49899	8	0,18622	0,26946
n = 100	1,01167	0,50039	3	0,04995	0,25731
	1,00846	0,50111	5	0,04655	0,25535
	1,00418	0,49981	8	0,04144	0,25523
n = 500	1,00307	0,50034	3	0,00884	0,25111
	1,00289	0,50019	5	0,00883	0,25129
	1,00184	0,50014	8	0,00754	0,25084

Elaboración: Autor.

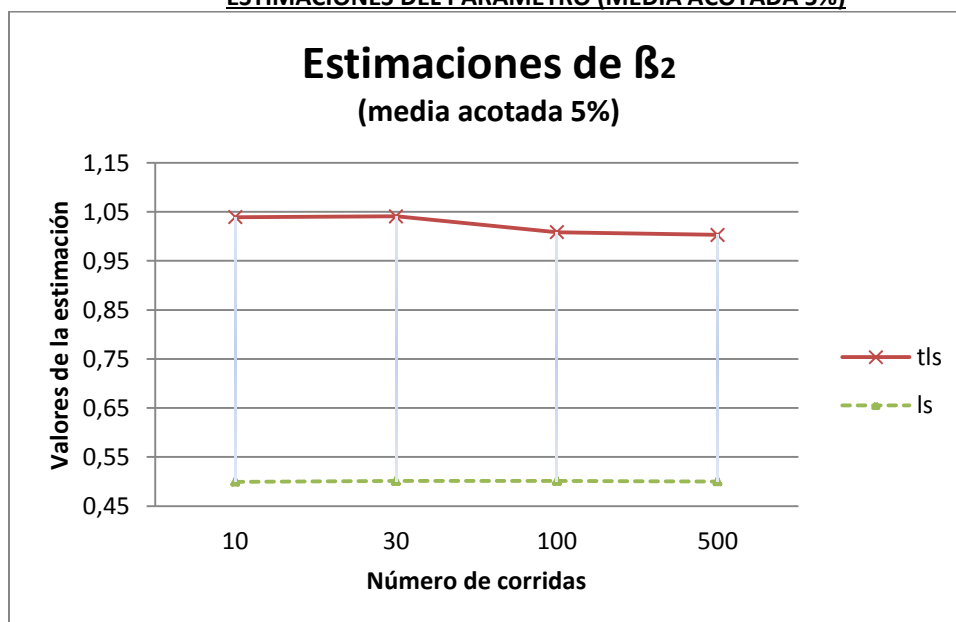
En este segundo caso (tabla 3), se ha dado un valor al coeficiente del término constante, se observa que no hay variación significativa en los resultados, comportándose el modelo similarmente como en el caso 1.

FIGURA 8
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 3%)



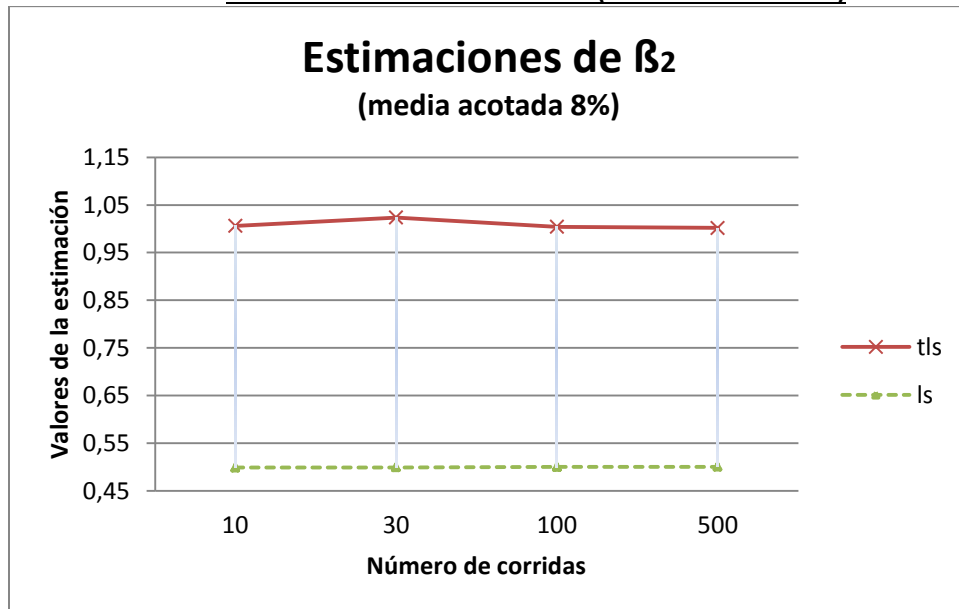
Elaboración: Autor.

FIGURA 9
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 5%)



Elaboración: Autor.

FIGURA 10
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 8%)



Elaboración: Autor.

Ahora en las figuras 8, 9 y 10 el comportamiento se presenta de manera similar al caso 1, presentando una ligera inestabilidad para el primer valor de análisis ($n = 10$), pero comportándose de manera correcta a diferencia del error que se presenta en el caso de la regresión ordinaria.

TABLA 4
INTERVALO CONFIANZA PARAMETRO (β_2)

CASO 2: Nsimul = 10000				
	INTERVALO DE CONFIANZA MEDIA			
	T.L.S		L.S.	
	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR	LIMITE SUPERIOR
n = 10	0,77069	1,64694	0,48896	0,50499
n = 30	0,46425	1,11292	0,49708	0,50471
n = 100	1,01195	1,02224	0,49749	0,50151
n = 500	0,99943	1,00345	0,49888	0,50064

Elaboración: Autor.

En el caso de los intervalos de confianza el comportamiento es similar al caso anterior (caso 1) pero como se aprecia en el intervalo de $n=100$, el valor conocido de beta es 1, y no lo contiene el intervalo de confianza del parámetro calculado, cabe destacar que los

resultados que se muestran en las tablas son de una corrida en particular los valores no se han obtenido por el cálculo de promedios de los límites del intervalo, pero aun así se aprecia el mejor desempeño del método ortogonal al ordinario.

TABLA 5

PROMEDIOS Y ERROR DE ESTIMACION PARA β_2

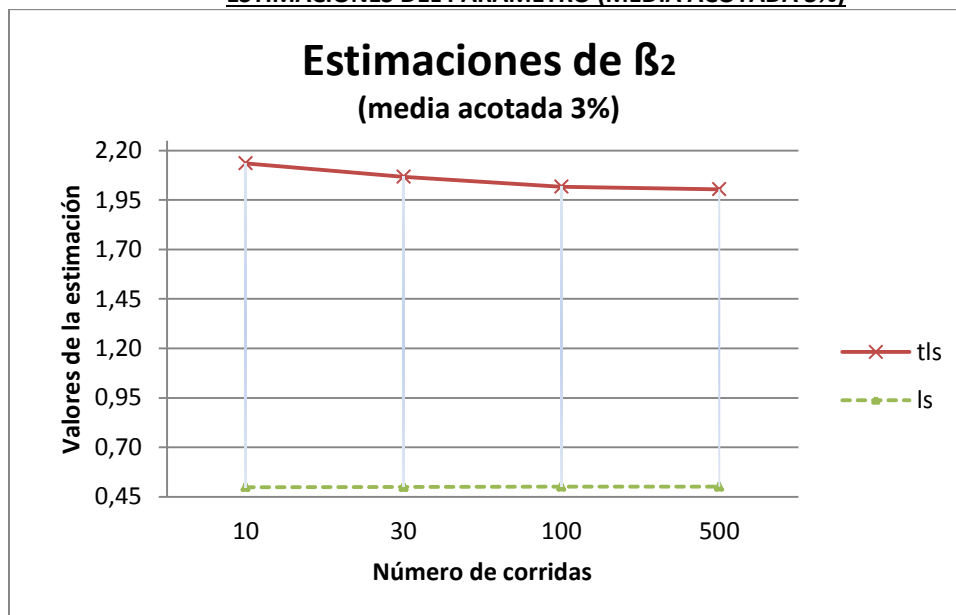
**CASO 3: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$
 $\beta_0 = 0, \beta_1 = 1, \beta_2 = 2$**

CASO 3:					
Nsimul = 10000					
	MEDIA ARITMETICA			ERROR CUADRATICO	
	β_2		% MEDIA	MEDIO	
	TLS	LS	ACOTADA	TLS	LS
n = 10	2,13515	1,00041	3	6,44619	1,19600
	2,07842	1,00191	5	3,81448	1,16066
	2,04445	0,99831	8	2,31732	1,13873
n = 30	2,06721	1,00287	3	0,40619	1,04101
	2,06294	1,00258	5	0,34006	1,03359
	2,04801	0,99588	8	0,29718	1,04061
n = 100	2,01757	1,00239	3	0,07593	1,00837
	2,01191	0,99732	5	0,07102	1,01664
	2,01424	1,00089	8	0,06486	1,00737
n = 500	2,00443	1,00108	3	0,01399	1,00039
	2,00168	1,00008	5	0,01298	1,00204
	2,00335	1,00039	8	0,01202	1,00094

Elaboración: Autor.

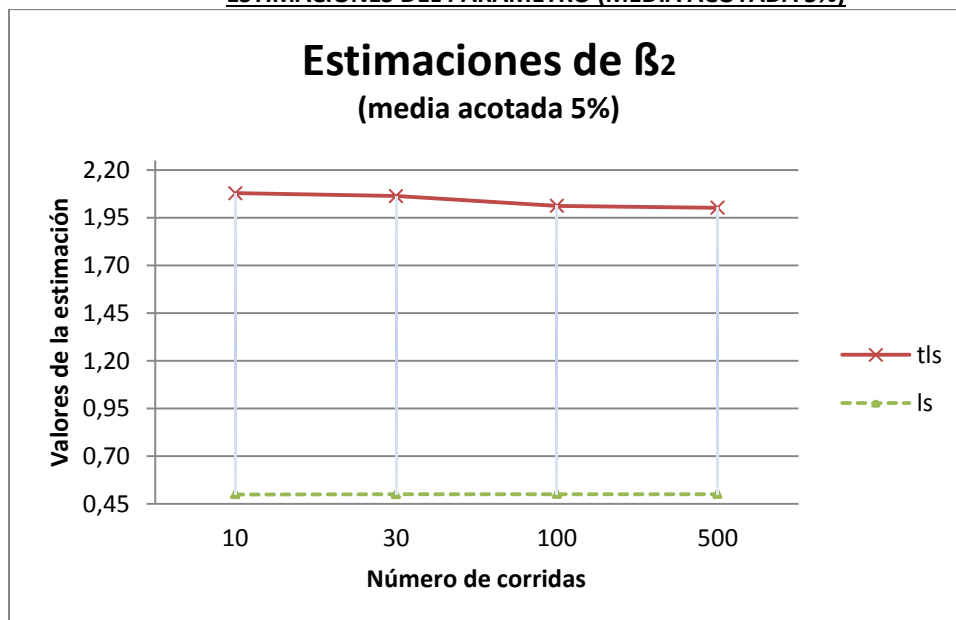
En este caso 3 (tabla 5), se ha variado el valor del coeficiente de la segunda variable regresora, el comportamiento del modelo sigue ajustándose a lo previsto, se obtiene una buena aproximación para pequeños valores del número de datos y se observa el sesgo en el coeficiente obtenido por la regresión ordinaria.

FIGURA 11
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 3%)



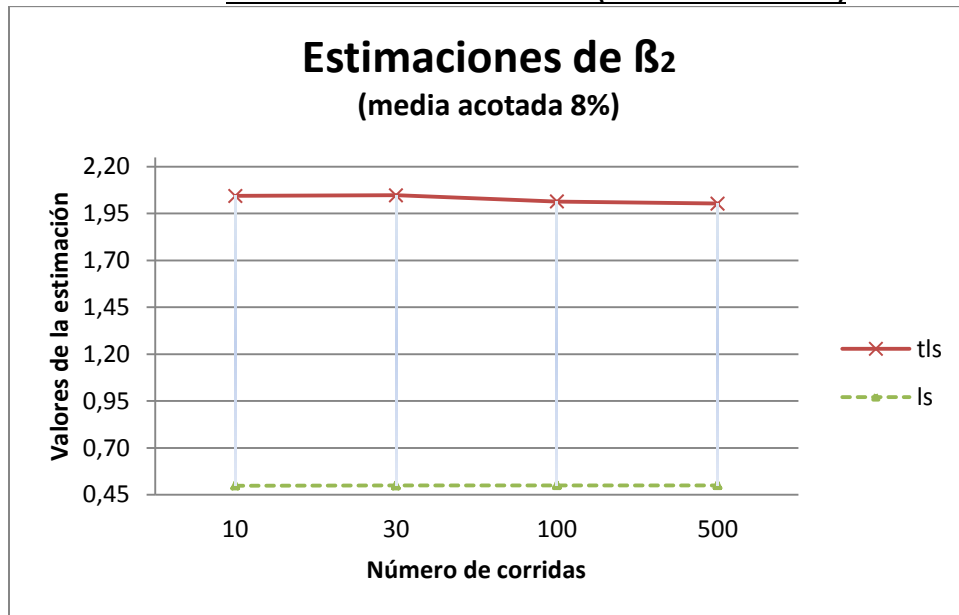
Elaboración: Autor.

FIGURA 12
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 5%)



Elaboración: Autor.

FIGURA 13
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 8%)



Elaboración: Autor.

En las figuras 11, 12 y 13, puede apreciarse este comportamiento más estable a valores mayores del tamaño de la muestra. Pero igual que en los casos anteriores el desempeño del método ortogonal es completamente correcto en comparación al de regresión ordinaria.

TABLA 6
INTERVALO CONFIANZA PARAMETRO (β_2)

CASO 3: Nsimul = 10000				
	INTERVALO DE CONFIANZA MEDIA			
	T.L.S		L.S.	
	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR	LIMITE SUPERIOR
n = 10	-0,49640	6,88284	0,98844	1,00926
n = 30	-4,42315	4,22892	0,99424	1,00456
n = 100	2,02278	2,03602	0,99866	1,00394
n = 500	2,00018	2,00527	0,99762	0,99992

Elaboración: Autor.

En lo que hace referencia a los intervalos de confianza se aprecia nuevamente el mejor desempeño de los resultados obtenidos mediante el método de regresión

ortogonal al de regresión ordinaria, aunque se observa que el valor dispuesto para (β_2) no se encuentra dentro del intervalo de confianza obtenido ($n = 100$ y $n = 500$).

TABLA 7

PROMEDIOS Y ERROR DE ESTIMACION PARA β_2

CASO 4: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 4, \sigma_{\varepsilon_2}^2 = 4, \sigma_y^2 = 1$
 $\beta_0 = 0, \beta_1 = 1, \beta_2 = 1$

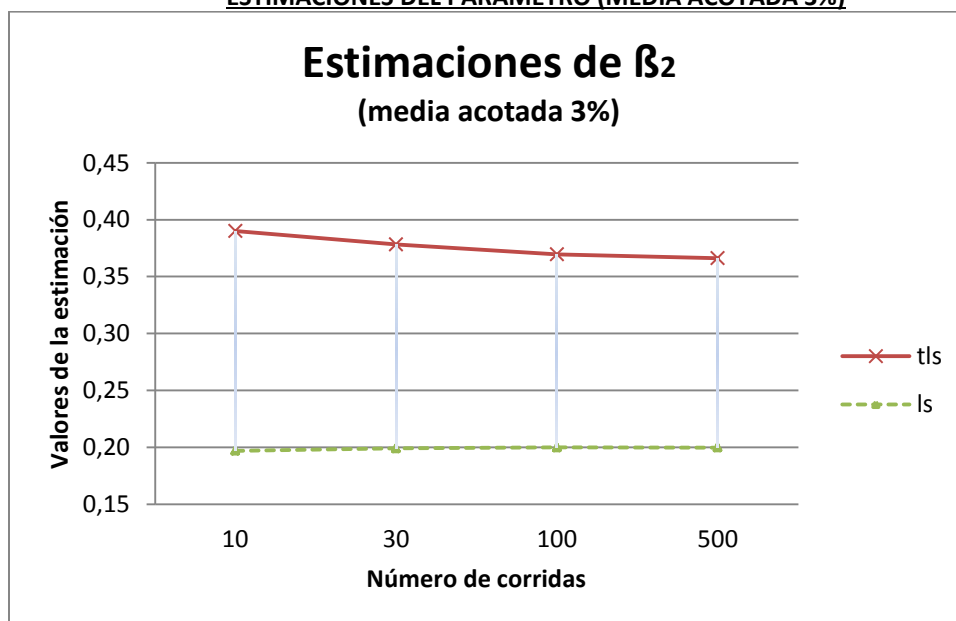
CASO 4:					
Nsimul = 10000					
	MEDIA ARITMETICA			ERROR CUADRATICO	
	β_2		% MEDIA	MEDIO	
	TLS	LS	ACOTADA	TLS	LS
n = 10	0,39007	0,19691	3	1,53897	0,70102
	0,37651	0,20331	5	1,09011	0,68118
	0,36595	0,20029	8	0,86557	0,67643
n = 30	0,37823	0,19897	3	0,47954	0,65598
	0,37908	0,20064	5	0,46695	0,65085
	0,37776	0,20097	8	0,44863	0,64802
n = 100	0,36959	0,19981	3	0,41248	0,64419
	0,36744	0,19962	5	0,41245	0,64391
	0,36986	0,19997	8	0,40742	0,64272
n = 500	0,36626	0,19974	3	0,40435	0,64091
	0,36704	0,19998	5	0,40302	0,64066
	0,36691	0,20016	8	0,40269	0,64061

Elaboración: Autor.

En este caso 4 (tabla 7) ocurre un comportamiento diferente en los resultados obtenidos anteriormente por el procedimiento de regresión ortogonal, éstos se vuelven imprecisos, sigue manteniendo el sesgo en el caso de los estimadores obtenidos por la regresión ordinaria, pero ambos se alejan de los valores que se han introducido, aunque

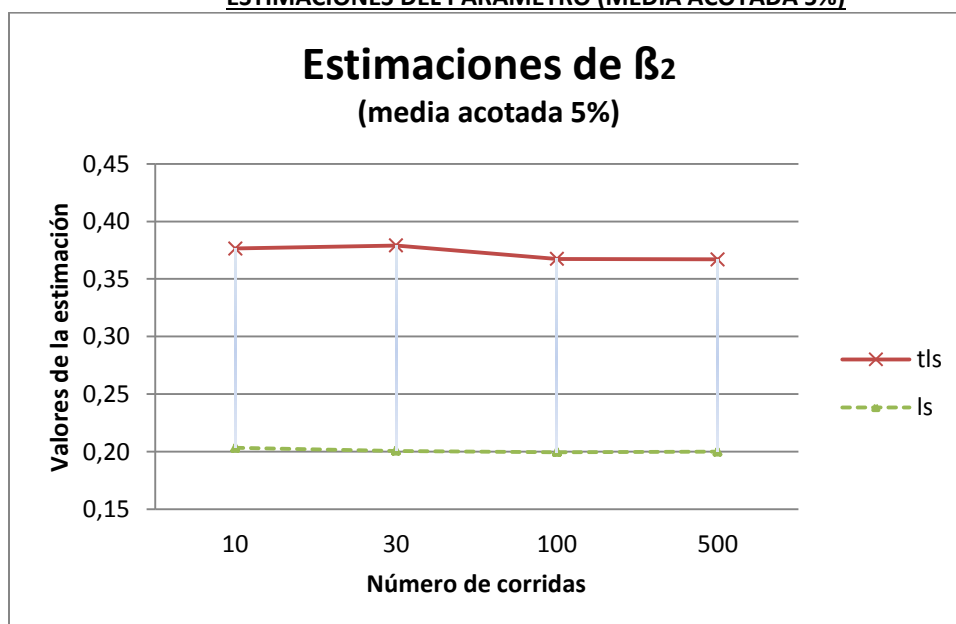
sigue siendo más cercano al verdadero valor, el resultado obtenido por medio de la regresión ortogonal. Aparentemente el introducir un ruido alto y diferente entre las variables $\sigma_{\varepsilon_1}^2 = 4, \sigma_{\varepsilon_2}^2 = 4, \sigma_y^2 = 1$ (valores de la varianza) las afecta de tal manera que ambos métodos no son aceptables para la estimación (Fuller, 1987).

FIGURA 14
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 3%)



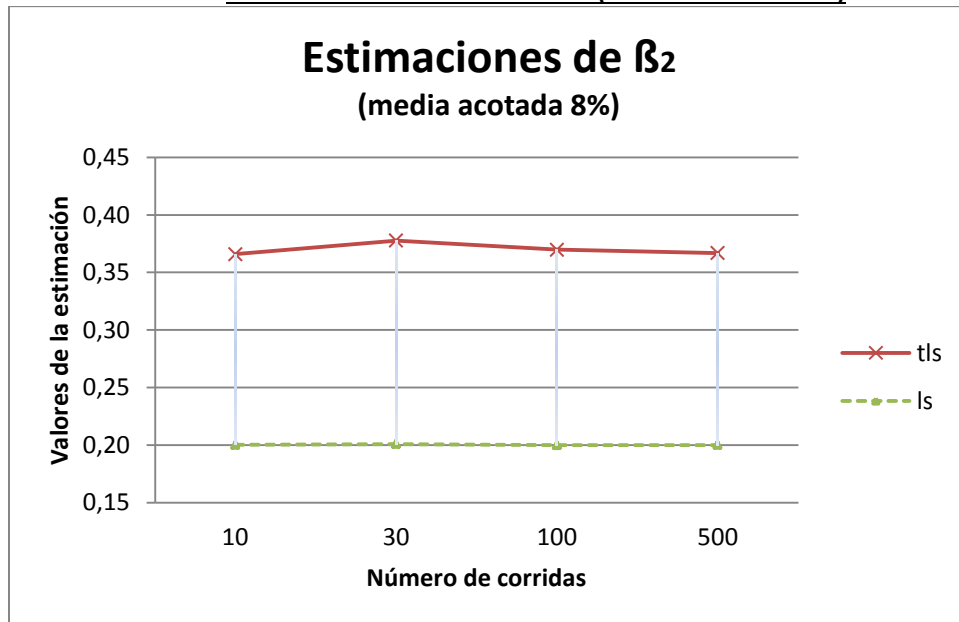
Elaboración: Autor.

FIGURA 15
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 5%)



Elaboración: Autor.

FIGURA 16
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 8%)



Elaboración: Autor.

TABLA 8
INTERVALO CONFIANZA PARAMETRO (β_2)

CASO 4: Nsimul = 10000				
	INTERVALO DE CONFIANZA MEDIA			
	T.L.S		L.S.	
	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR	LIMITE SUPERIOR
n = 10	-1,69362	0,77922	0,19773	0,20928
n = 30	0,37028	0,43558	0,19902	0,20458
n = 100	0,37021	0,37647	0,19883	0,20175
n = 500	0,36567	0,36808	0,19934	0,20061

Elaboración: Autor.

En este caso no se hará mayor comentario ya que los dos tipos de regresión muestran problemas en la estimación del parámetro bajo estudio y análisis.

TABLA 9**PROMEDIOS Y ERROR DE ESTIMACION PARA β_2**

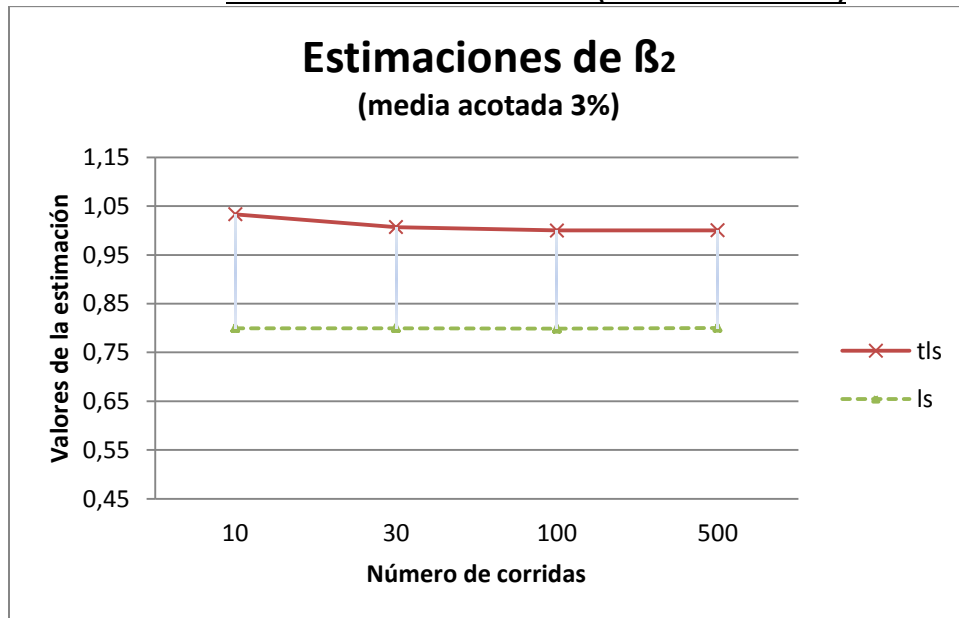
CASO 5: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 4, \sigma_{z_2}^2 = 4, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 1$
 $\beta_0 = 0, \beta_1 = 1, \beta_2 = 1$

CASO 5:					
Nsimul = 10000					
	MEDIA ARITMETICA			ERROR CUADRATICO	
	β_2		% MEDIA	MEDIO	
	TLS	LS	ACOTADA	TLS	LS
n = 10	1,03322	0,79962	3	0,16893	0,10608
	1,02135	0,79745	5	0,14457	0,09698
	1,02036	0,80286	8	0,11939	0,09625
n = 30	1,00683	0,79951	3	0,03091	0,05521
	1,00481	0,79907	5	0,02712	0,05281
	1,00388	0,79959	8	0,02508	0,05067
n = 100	1,00002	0,79863	3	0,00804	0,04445
	1,00176	0,79969	5	0,00742	0,04344
	1,00084	0,80026	8	0,00668	0,04389
n = 500	1,00029	0,80018	3	0,00149	0,04069
	1,00028	0,79974	5	0,00142	0,04075
	1,00047	0,80008	8	0,00131	0,04049

Elaboración: Autor.

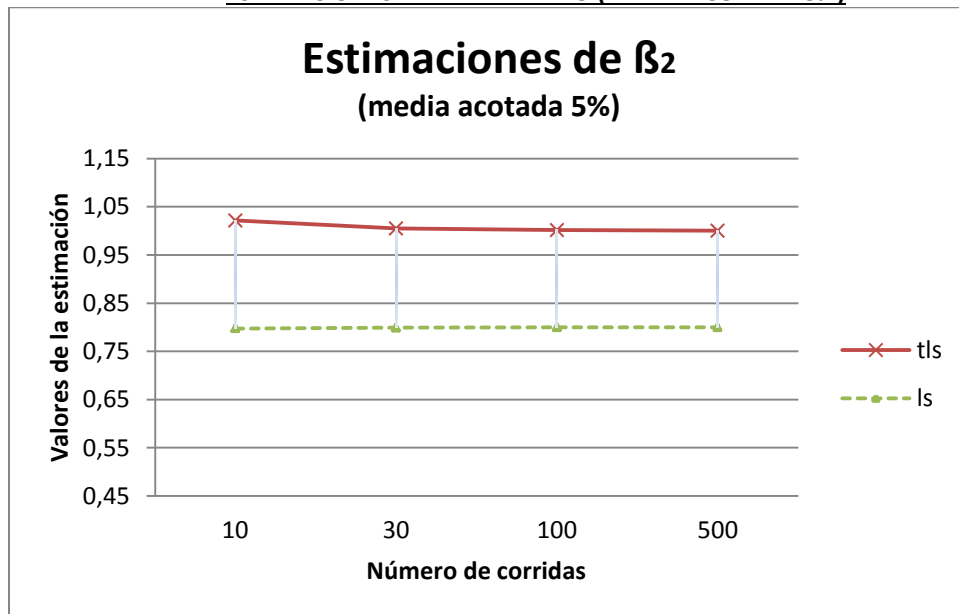
En el caso 5 (tabla 9) se ha introducido una mayor varianza en la variable original del modelo, y se puede observar que los resultados se ajustan a lo previsto, y se mantiene el sesgo en la estimación del coeficiente a partir de la regresión ordinaria, así mismo el método se ajusta con muy pocos datos, obsérvese también los valores de las varianzas del ruido introducido.

FIGURA 17
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 3%)



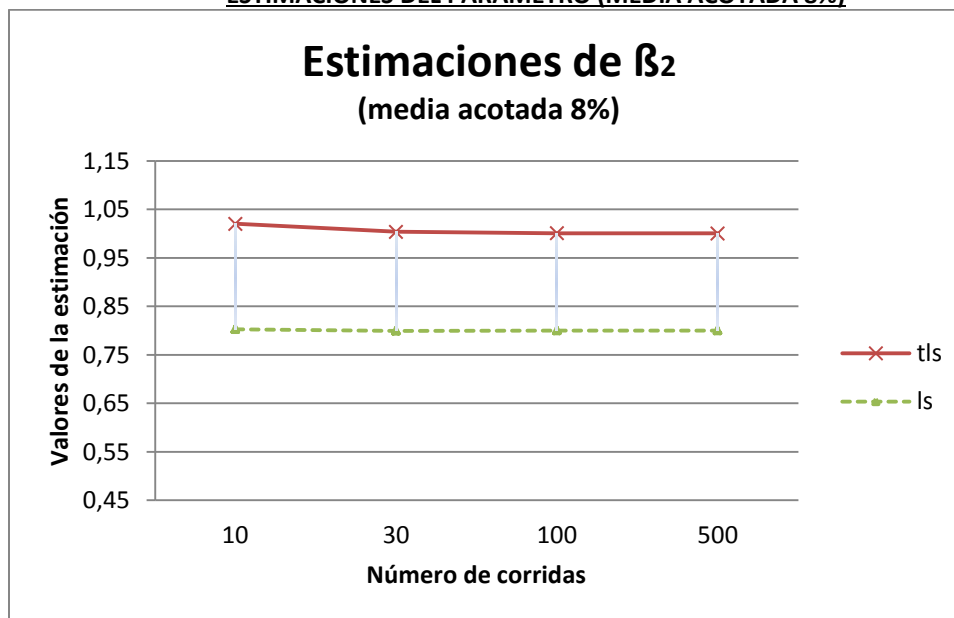
Elaboración: Autor.

FIGURA 18
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 5%)



Elaboración: Autor.

FIGURA 19
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 8%)



Elaboración: Autor.

Nuevamente el proceso de regresión ortogonal muestra mejores resultados a los diferentes valores de n utilizados en la simulación. Los valores de regresión ordinaria mantienen el comportamiento sesgado que lo hace inadecuado en este comportamiento de las variables regresoras.

TABLA 10
INTERVALO CONFIANZA PARAMETRO (β_2)

CASO 5: Nsimul = 10000				
	INTERVALO DE CONFIANZA MEDIA			
	T.L.S		L.S.	
	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR	LIMITE SUPERIOR
n = 10	0,87910	1,11201	0,80088	0,81258
n = 30	1,00531	1,01297	0,79685	0,80243
n = 100	1,00008	1,00382	0,79793	0,80083
n = 500	0,99961	1,00128	0,79919	0,80047

Elaboración: Autor.

Así mismo el comportamiento de los intervalos de confianza para la regresión ordinaria presentan el sesgo; y como ya se ha indicado en otros casos el valor del parámetro (β_2) que se ha fijado en 1, no se encuentra en los intervalos de confianza calculados ($n = 30$ y $n = 100$).

TABLA 11

PROMEDIOS Y ERROR DE ESTIMACION PARA β_2

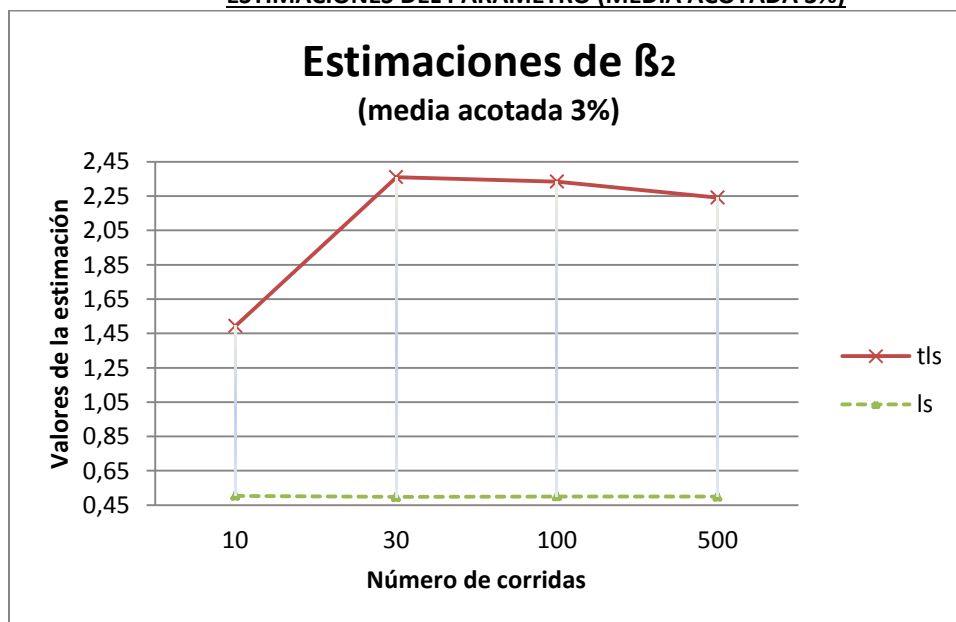
CASO 6: $\mu_1 = 0, \mu_2 = 0, \sigma_{z_1}^2 = 1, \sigma_{z_2}^2 = 1, \sigma_{\varepsilon_1}^2 = 1, \sigma_{\varepsilon_2}^2 = 1, \sigma_y^2 = 4$
 $\beta_0 = 0, \beta_1 = 1, \beta_2 = 1$

CASO 6:					
Nsimul = 10000					
	MEDIA ARITMETICA			ERROR CUADRATICO	
	β_2		% MEDIA	MEDIO	
	TLS	LS	ACOTADA	TLS	LS
n = 10	1,49421	0,50454	3	48,86889	0,54808
	1,56169	0,49703	5	28,49619	0,53616
	1,47024	0,50123	8	18,53913	0,49303
n = 30	2,36012	0,49742	3	24,62975	0,32437
	2,34683	0,49935	5	14,69175	0,31207
	2,32258	0,49562	8	10,15156	0,30201
n = 100	2,33461	0,50056	3	4,34036	0,26844
	2,32171	0,49794	5	3,68074	0,26815
	2,29311	0,50061	8	3,04532	0,26245
n = 500	2,24162	0,49966	3	1,72961	0,25402
	2,24381	0,50081	5	1,70552	0,25241
	2,22981	0,50031	8	1,63388	0,25213

Elaboración: Autor.

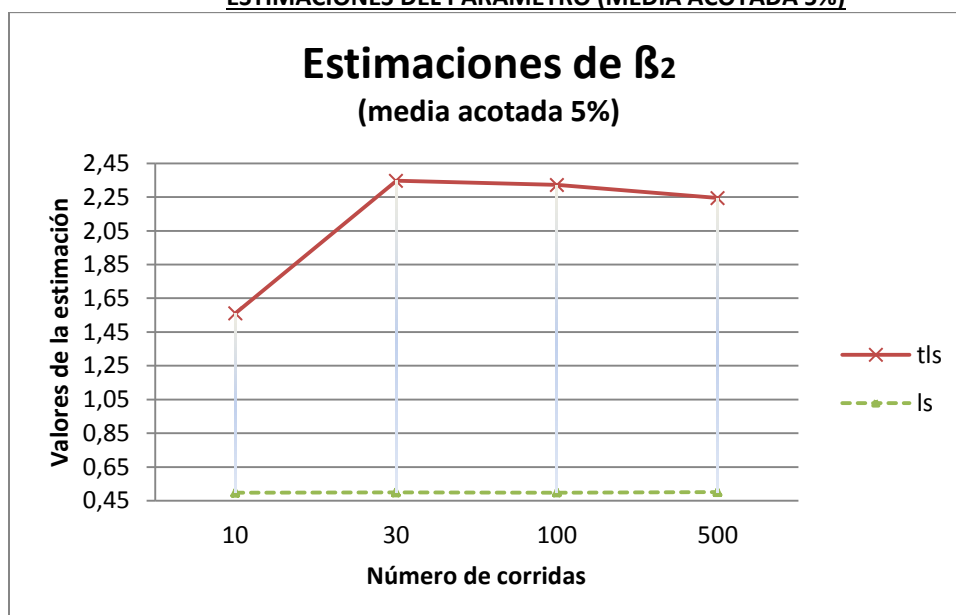
En este caso 6 (tabla 11) ocurre un comportamiento similar al ocurrido en el caso 4 (tabla 7), siendo los resultados obtenidos por el procedimiento de regresión ortogonal inexactos, se sigue manteniendo el sesgo en el caso de los estimadores obtenidos por la regresión ordinaria, pero ambos se alejan de los valores que se han introducido. Aparentemente asignar un ruido alto (valor de varianza) y diferente en las variables las afecta de tal manera que ambos métodos no son aceptables para la estimación.

FIGURA 20
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 3%)



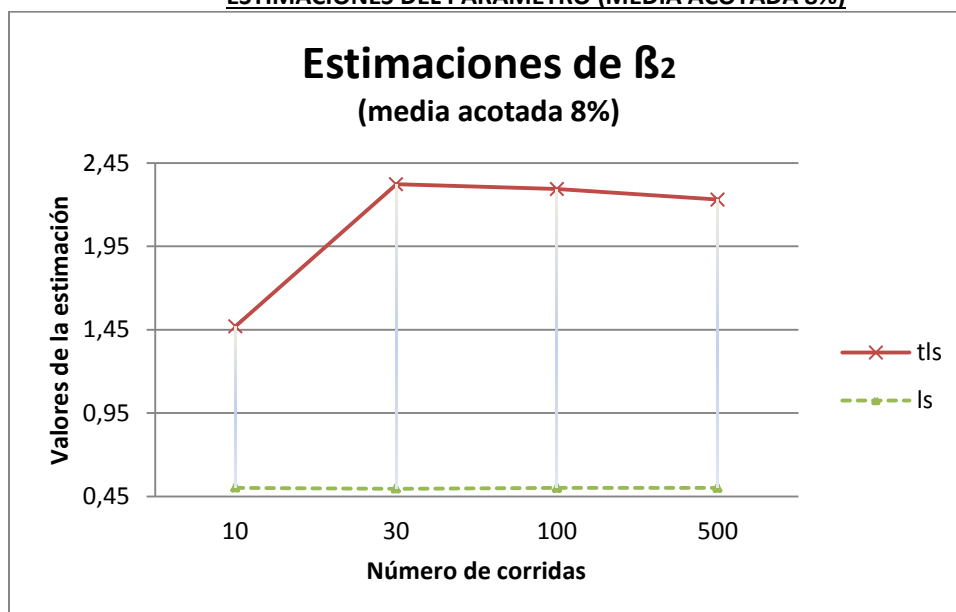
Elaboración: Autor.

FIGURA 21
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 5%)



Elaboración: Autor.

FIGURA 22
ESTIMACIONES DEL PARAMETRO (MEDIA ACOTADA 8%)



Elaboración: Autor.

TABLA 12
INTERVALO CONFIANZA PARAMETRO (β_2)

CASO 6: Nsimul = 10000				
	INTERVALO DE CONFIANZA MEDIA			
	T.L.S		L.S.	
	LIMITE INFERIOR	LIMITE SUPERIOR	LIMITE INFERIOR	LIMITE SUPERIOR
n = 10	0,29714	5,72261	0,48959	0,51494
n = 30	-11,99028	7,29869	0,49668	0,50880
n = 100	1,94428	3,11306	0,49975	0,50605
n = 500	2,23688	2,25907	0,49932	0,50208

Elaboración: Autor.

Para este último caso, como se aprecia en las figuras 20, 21 y 22 y en la tabla 12, los resultados en ambos casos presentan inconsistencias respecto al valor fijado de (β_2).

Como comentario final para cerrar la discusión de los resultados obtenidos en los casos 1, 2, 3 y 5 en donde se mostró que la regresión ortogonal ha dado mejores resultados al momento de obtener los parámetros de la regresión, lo cual es acorde a lo indicado por Casella (2002), Fuller (1987), Van Huffel (2004) y Davidov (2004) así como Maddala (1996) en su texto de Econometría.

CONCLUSIONES:

Uno de los principales resultados para un problema de dos variables es que en la regresión ortogonal no interesa cual variable sea considerada la regresora ya que se obtiene la misma recta de regresión, lo cual se mostró no ocurre en la regresión ordinaria.

Tal como se ha verificado en las tablas de resultados, los estimadores de regresión ortogonal resultaron ser más adecuados por su exactitud respecto a los valores que se obtuvieron por el método de regresión ordinaria.

A pesar que se utilizó pocos datos ($n = 10$), ya se tiene una buena aproximación a los resultados con los cuales se generaron los datos para la regresión. Aunque también en esta situación aparecen valores del error cuadrático medio mayores que cuando se utilizó los otros valores para n .

Si bien es cierto, al incrementar el valor de la varianza de la variable que representaría el ruido blanco (proceso estocástico de media nula) en las variables originales del modelo, ambos métodos se vieron afectados en su exactitud, el resultado de la regresión ortogonal sigue siendo de mejor rendimiento, aunque cuando se asignaron valores diferentes a la varianza de los elementos de error ambas regresiones resultaron inexactas en su estimación.

Se pudo también verificar que los sesgos que ocurren en la regresión ordinaria y que son el principal motivo para no usar dicho método se cumplieron tal y cual se obtuvo en la derivación teórica.

Así también los intervalos de confianza que se calcularon para ambos procedimientos, muestran el mejor comportamiento del método de regresión ortogonal a los que se han obtenido por el método de regresión ordinaria.

Por otro lado, hay algunos aspectos teóricos que aun necesitan revisarse y se augura hacerlo en otros trabajos, con lo cual se sugiere nuevos análisis especialmente en los casos en que no hubo mayor exactitud en los resultados.

Otros aspectos también a tomar en consideración, son las aplicaciones prácticas que se derivan de este trabajo. Es muy común la aplicación de la regresión ordinaria en diversos aspectos y campos de la ciencias sociales y aun en las ciencias duras, sin tomar en consideración los supuestos que deben cumplirse y respetarse, especialmente en las ciencias económicas donde se aplica indiscriminadamente la regresión ordinaria, la cual como se ha demostrado no obtiene los mejores resultados cuando el supuesto de la no aleatoriedad de las variables regresoras no se cumple.

Finalmente, en muchas aplicaciones de regresión, los datos no son obtenidos a partir de experimentos diseñados, sino que son observacionales, y no se ha fijado ninguna variable regresora. Si el interés es predecir una variable en términos de otras, la regresión ordinaria es aceptable. Pero si lo que interesa es la relación en sí, y ninguna de las variables involucradas es fijada de antemano, es conveniente utilizar regresión ortogonal.

REFERENCIAS BIBLIOGRAFICAS

Al-sharadqah, A., Chernov, N., & Huang, Q. (2011). *Errors-In-Variables regression and the problem of moments*. Recuperado el 2012 de Julio, de Brazilian Journal of Probability and Statistics: <http://www.math.uab.edu/~chernov/cl>

Azarang, M. R., & García, E. (1996). *Simulación y análisis de Modelos Estocásticos*. México: McGraw-Hill.

Baker, K. (2005). *Singular Value Decomposition Tutorial*. Recuperado el Agosto de 2012, de Ohio State University: http://www.ling.ohio-state.edu/~kbaker/pubs/Singular_Value_Decomposition_Tutorial.pdf

Blyth, T. S., & Robertson, E. F. (2002). *Further Linear Algebra*. Londres: Springer.

Boggs, P. T., & Rogers, J. E. (1990). *Orthogonal Distance Regression*. Recuperado el Julio de 2012, de Center for Computing and Applied Mathematics: http://docs.scipy.org/doc/external/odr_ams.pdf

Casella, G., & Berger, R. L. (2002). *Statistical Inference*. Thomson Learning.

Davidov, O. (2004). *Estimating the slope in measurement error models—a different perspective*. Recuperado el 2012 de Julio, de Statistics & Probability Letters: www.elsevier.com/locate/stapro

De Groen, P. (1996). *An Introduction to Total Least Squares*. Recuperado el Julio de 2012, de Vrije Universiteit Brussel, Department of Mathematics.: arXiv:math/9805076v1

Di Ciccio, T., & Tibshirani, R. (1991). *Department of Statistics, University of Toronto*. Recuperado el 10 de enero de 2014, de <http://www.utstat.toronto.edu/wordpress/WSFiles/technicalreports/9107.pdf>

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*. New York: Chapman & Hall/CRC.

Eaton, M. L. (2007). *Multivariate Statistics: A vector space approach*. Ohio: Institute of Mathematical Statistics.

Freund, J. E., Miller, I., & Miller, M. (2000). *Estadísticas matemática con aplicaciones*. México: Pearson Educación.

Fuller, W. A. (1987). *Measurement error models*. New York: JOHN WILEY & SONS.

Gillard, J. W. (2006). *An Historical Overview of Linear Regression with Errors with variables*. Recuperado el Julio de 2012, de School of Mathematics, Cardiff University: http://cardiff.ac.uk/math/resources/Gillard_Tech_Report.pdf

Gleser, L. J. (1981). *Estimation in a multivariate "errors in variables" regression models: large sample results*. Recuperado el 2012 de Mayo, de Annals of Statistics: www.jstor.org

Golub, G. H., & Van Loan, C. F. (1980). *An analysis of the total least squares problem*. Recuperado el Mayo de 2012, de Society for Industrial and Applied Mathematics: <http://www.cs.cornell.edu/cv/ResearchPDF/Analysis.total.least.squares.prob.pdf>

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.

Maddala, G. S. (1996). *Introducción a la Econometría*. México: Prentice-Hall Hispanoamericana, S.A.

Markovsky, I., & Van Huffel, S. (2007). *Overview of total least squares methods*. Recuperado el Julio de 2012, de School of Electronics and Computer Science, University of Southampton.: http://eprints.soton.ac.uk/263855/1/tls_overview.pdf

Martin, S. B. (1998). *An alternative method of least squares linear regression*. Recuperado el Junio de 2012, de Mc Murry University: www.mcm.edu/mathdept/sarah.pdf

Montoya, J. A. (2008). *La verosimilitud perfil en la Inferencia Estadística*. Guanajuato: CIMAT.

Nievergelt, Y. (1994). *Total least squares: State-of-the-art regression in numerical analysis*. Recuperado el Mayo de 2012, de Society for industrial and applied mathematics: <http://people.duke.edu/~hpgavin/ce200/nievergelt94.pdf>

Noble, B., & Daniel, J. W. (1989). *Algebra Lineal Aplicada*. México: Prentice-Hall Hispanoamericana, S.A.

Otamendi, J. (2006). Las etapas en la gestación del método de Montecarlo. En A.H.E.P.E, *Historia de la Probabilidad y la Estadística (III)* (págs. 117 - 123). Madrid: Publicaciones Delta.

Peña, D. (2002). *Análisis de datos multivariantes*. Madrid: McGraw Hill.

Petras, I., & Podlubny, I. (2010). *Least Squares or Least Circles? A comparison of classical regression and orthogonal regression*. Recuperado el Junio de 2012, de American Statistical Association: www.amstat.org/membership/index.cfm

Petras, I., Bednárová, D., & Podlubny, I. (2008). *Description of behavior of national economies in state space*. Recuperado el 2012 de Julio, de Acta Montanistica Slovaca: <http://actamont.tuke.sk/pdf/2008/n1/27petras.pdf>

Schaffrin, B., & Wieser, A. (2010). *Total least-squares adjustment of condition equations*. Recuperado el Mayo de 2012, de Springer Link: <http://link.springer.com/article/10.1007/s11200-011-0032-3#page-1>

Van Huffel, S. (2004). *Total Least Squares and Errors-in-Variables Modeling: Bridging the gap between Statistics, Computational Mathematics and Engineering*. Recuperado el Mayo de 2012, de K.U.Leuven, Dept. of Electrical Engineering: www.esat.kuleuven.ac.be/sista.

Van Huffel, S., & Zha, H. (1993). *The total least squares problem*. Recuperado el Junio de 2012, de Handbook of Statistics: citeseer.uark.edu:8080/citeseerx/showciting

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Londres: Oxford.

ANEXO:

CODIGO EN R PARA LA SIMULACION DE LOS RESULTADOS DE REGRESION ORDINARIA Y ORTOGONAL

```
tls=function(X,y)
{
  Xc=scale(X,scale=FALSE)
  yc=scale(y,scale=FALSE)
  a=cbind(Xc,yc)
  p=ncol(X)
  d=svd(a)$d[p+1]
  b=solve(crossprod(Xc)-diag(rep(d^2,p)),crossprod(Xc,yc))
  b0=mean(y)-sum(colMeans(X)*b)
  c(b0,b)
}
b=c(0,1,1)
corrida2=function(n)
{
  eta1=rnorm(n)
  eta2=rnorm(n)
  x1=eta1+rnorm(n)
  x2=eta2+rnorm(n)
  y=b[1]+b[2]*eta1+b[3]*eta2+rnorm(n)
  X=cbind(x1,x2)
  btls=tls(X,y)
  bls=coef(lm(y~X))
  res=c(btls,bls)
  names(res)=c(paste("tls.b",0:2,sep=""),paste("ls.b",0:2,sep=""))
}
```

```

    res
  }
  Nsimul=10000
  n=500
  simul=sapply(rep(n,Nsimul),corrida2)
  rowMeans(simul)
  mean(simul["tls.b2",],trim=.05)
  mean(simul["ls.b2",],trim=.05)
  sum(abs(simul["tls.b2",]-1)>3)
  sum(abs(simul["ls.b2",]-1)>3)
  rowMeans((simul-b)^2)
  apply((simul-b)^2,1,mean,trim=.05)

```