
Incidencia de la factura electrónica en los impuestos por bienes y servicios en Latinoamérica: Un enfoque desde los modelos lineales generalizados mixtos

Electronic invoicing (e-invoicing) incidence for goods and services taxation in Latin America: A generalized linear mixed model approach

William Fernando Acero Ruge^a
waceror@dian.gov.co

Jaime Humberto Niño Peña^b
jninop@dian.gov.co

Resumen

La evasión tributaria es uno de los grandes problemas de los países en vía de desarrollo, afectando la equidad del sistema fiscal. Una de las posibles soluciones para evitar esta situación es hacer uso de la tecnología, específicamente de la factura electrónica en el ámbito de la digitalización de las obligaciones tributarias para aumentar el recaudo. En Latinoamérica países como Argentina, Brasil, Chile, Ecuador, México, Uruguay entre otros, han implementado de forma exitosa este sistema, cada uno con sus particularidades y han obtenido una incidencia positiva sobre los niveles de recaudo en el impuesto sobre la ventas de bienes y servicios. En este artículo proponemos un enfoque a través de los modelos lineales generalizados mixtos (GLMM) con el fin de determinar si existe una incidencia positiva de la implementación de la factura electrónica en el recaudo del impuesto de bienes y servicios en Latinoamérica. Los resultados indican que existe una incidencia positiva de la factura electrónica sobre el recaudo del impuesto de bienes y servicios, con una tendencia alcista a través de los años.

Palabras clave: Modelos lineales generalizados mixtos (GLMM), recaudo, impuesto, factura electrónica.

Abstract

Tax evasion is one of the major problems of developing countries, affecting the fairness of the tax system. One of the possible solutions to avoid this situation is to make use of technology, specifically electronic invoicing in the field of digitization of

^aDirección de impuestos y aduanas nacionales - DIAN

^bDirección de impuestos y aduanas nacionales - DIAN

tax obligations to increase revenue collection. In Latin America, countries such as Argentina, Brazil, Chile Ecuador, Mexico, Uruguay among others have successfully implemented this system, each with its own particularities and they have obtained a positive impact on collection levels, particularly in Value Added Tax. In this article we propose an approach through mixed generalized linear models (GLMM) in order to determine if electronic invoicing implementation yields a positive impact on tax collection for sales of goods and services across Latin America. The results indicate that there is a positive impact of e-invoicing over tax collection for sales of goods and services, with an upward trend over the years

Keywords: Generalized linear mixed models (GLMM), tax collection, electronic invoicing.

1. Introducción

La factura electrónica (FE) es la evolución de la factura tradicional que para efectos legales tiene la misma validez que el papel. Tributariamente es un soporte de transacciones de venta de bienes y/o servicios. En Latinoamérica los sistemas de FE coinciden en ciertas características aunque mantienen diferencias importantes, debido a las regulaciones locales en materia tributaria, de tal manera que no existen dos implementaciones de FE iguales, por ejemplo, difieren en el año en el que se empieza a implementar la FE y la manera en la que los contribuyentes se van haciendo partícipes en este sistema. Entre los puntos en común se encuentra que 1) La administración tributaria de cada país tiene todos los documentos (En el caso Colombiano es la Dirección de Impuestos y Aduanas Nacionales - DIAN), 2) Los Beneficios del contribuyente, en términos de la disminución de costos operativos como ahorro de papel y del espacio físico para el almacenamiento de los documentos y 3) La existencia de un formato estándar para todos los contribuyentes, entre otros.

Cinco países Latinoamericanos han llevado a cabo investigaciones para lograr determinar si en cada caso existe un efecto de la FE en el aumento de las diferentes variables de interés relacionadas con el recaudo. En general hasta el momento todos estos estudios han hecho uso de metodologías cuasi-experimentales (Rosenbaum & Rubin 1983). En Argentina, Artana & Templado (2018) utilizaron el método de diferencias en diferencias (DID) y de regresión discontinua (RD) sobre las *Ventas gravadas*, *Débito menos crédito fiscal*, *Recaudo*, encontrado efectos positivos en el incremento de estos. En Brasil, Naritomi (2019) encontró a través de DID que las *utilidades reportadas* aumentaron en 22% en 4 años. Para Ecuador, Ramírez et al. (2018) determinaron a través de DID que las *Ventas locales gravadas*, *Ventas no gravadas*, *Compras gravadas*, *Compras no gravadas* y *el Impuesto causado* que a nivel general el impacto de la implementación de FE favoreció positivamente el recaudo sobre estas variables. Castro et al. (2016) para México, a través de la RD, determinaron un impacto positivo sobre el *impuesto sobre la renta* y en Uruguay, Bérigolo et al. (2018) determinaron a través de DID un impacto positivo sobre el

monto de los pagos realizados por los contribuyentes.

En general, cada uno de estos países ha realizado estudios de forma separada para determinar el impacto de la FE en su respectivo caso. Debido a lo anterior, se hace necesario contar una metodología que determine de forma general si en Latinoamérica existe un efecto positivo de la implementación de la FE sobre el recaudo de impuestos incluso sobre aquellos países que a pesar de haber implementado este sistema aún no han realizado algún estudio similar. En este sentido y debido a que las condiciones socio-económicas de cada país son diferentes, en este estudio se tomará como fuente de información la publicada por la Organización para la Cooperación y el Desarrollo Económicos (OCDE).

En la sección 2, se describen los conceptos de los modelos lineales generalizados mixtos (GLMM) usados en este artículo, la sección 3 considera el análisis y descripción de la información utilizada, junto con el modelo planteado y los resultados asociados. Finalmente en la sección 4 contiene las conclusiones y trabajo futuro derivado de esta investigación.

2. Marco teórico

De forma general, autores como Rencher & Schaalje (2008) describen el modelo lineal clásico de la forma

$$y = X\beta + \epsilon \quad (1)$$

donde y es un vector de dimensión $n \times 1$ representando la variables respuesta, X es la matriz de covariables de dimensión $n \times p$, β un vector de efectos fijos de dimensión $p \times 1$ y ϵ un vector de errores independientes tal que $\epsilon_i \sim N(0, \sigma^2)$ y además $E(y_i) = \mu_i = x'_i\beta$, donde x'_i representa la i -ésima fila de la matrix X . Sin embargo, como no siempre la relación anterior es lineal, este modelo ha sido extendido a la familia exponencial (Nelder & Wedderburn 1972), donde la distribución de cada una de las variables independientes pertenece a la familia exponencial y posee las siguiente propiedades:

1. **Componente aleatoria:** La función de densidad de cada y_i depende de un parámetro θ y puede ser representada de la forma

$$f(y_i; \theta) = \exp[y_i b(\theta) + c(\theta) + d(y_i)] \quad (2)$$

2. **Componente sistemática:** Las covariables producen un conjunto de predictores lineales de la forma

$$\eta_i = x'_i\beta \quad (3)$$

3. **Función de enlace:** Una función $g(\cdot)$ monótona y diferenciable tal que $E(y_i) = \mu_i$ que describe la relación entre la componente aleatoria y la componente sistemática de la forma

$$g(\mu_i) = \eta_i = x_i' \beta \quad (4)$$

lo anterior constituye la base de los modelos lineales generalizados (GLM). En estos modelos el vector de parámetros β se considera fijo y las observaciones independientes entre sí. Sin embargo, existen situaciones donde estas observaciones suelen estar correlacionadas, por ejemplo con los datos de tipo panel o datos anidados. Es allí donde los modelos mixtos son útiles para modelar las correlaciones entre las observaciones.

2.1. Modelos lineales generalizados mixtos (GLMM)

Bajo un modelo lineal simple, suponga que dado un vector de efectos aleatorios b , las observaciones y_1, \dots, y_n son (condicionalmente) independientes tal que

$$y_i | b_i \sim N(x_i' \beta + z_i' b_i, \sigma^2) \quad (5)$$

donde x_i y z_i son vectores conocidos, β y b son vectores desconocidos de coeficientes de regresión asociados a los efectos fijos y efectos aleatorios respectivamente, y σ^2 es una varianza desconocida. En este sentido existen dos elementos claves para construir un modelo lineal mixto gaussiano: (i) La independencia condicional y una distribución condicional y (ii) la distribución de los efectos aleatorios. Los anteriores elementos pueden ser usados para construir un GLMM (Jiang & Nguyen 2007). Estos modelos consideran que dado un vector de efectos aleatorios b y de observación y_1, \dots, y_n , independientes, la distribución de y_i dado b es un miembro de la familia exponencial con función de densidad de probabilidad (p.d.f) dada por

$$f(y_i | b) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c_i(y_i, \phi) \right\} \quad (6)$$

donde $b(\cdot)$, $a_i(\cdot)$ y $c_i(\cdot, \cdot)$ son funciones conocidas y ϕ es un parámetro de precisión que puede o no ser conocido. El parámetro canónico (ó parámetro natural, (Nelder & Wedderburn 1972)) θ_i , está asociado con la media condicional $\mu_i = E(y_i | b_i)$, que a su vez está asociado con un predictor lineal

$$\eta_i = x_i' \beta + z_i' b_i \quad (7)$$

donde se asume que $b_i \sim N(0, G)$, siendo G la matriz de varianzas y covarianza de los efectos aleatorios. η_i se relaciona con la variable respuesta a través de una función de enlace $g(\cdot)$ tal que

$$g(\mu_i) = \eta_i \quad (8)$$

Para los GLM, el proceso de estimación de parámetros se puede realizar median la construcción de la función de máxima verosimilitud. Esta estimación (de máxima verosimilitud) consta de tres pasos: describir la probabilidad de los datos, resolver las ecuaciones de *score* para encontrar una estimación y utilizar la información de Fisher para describir la varianza del estimador. En los modelos jerárquicos (como los GLMM), la distribución conjunta de los datos observados y no observados es fácil de describir, pero la distribución marginal de datos observados a menudo no lo es. Incluso en modelos completamente observados, encontrar las funciones de *score* puede resultar difícil y llevar mucho tiempo. Estas dificultades se multiplican cuando se intenta hallar la información de Fisher (observada). En este artículo se utiliza el enfoque propuesto por Kristensen et al. (2016) que combina la diferenciación automática con la aproximación de Laplace, y está diseñado de forma modular utilizando bibliotecas de software modernas y de alto rendimiento, además de haber demostrado reducción en los tiempos de procesamiento.

2.2. Modelo beta para datos longitudinales

La distribución beta es una distribución de probabilidad continua definida sobre el intervalo (0,1) con función de densidad

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi(1-\mu))} y^{\mu\phi-1} (1-y)^{\phi(1-\mu)-1} \quad (9)$$

donde ϕ es el parámetro de precisión y se tiene que $E(Y) = \mu$ y $V(Y) = \frac{\mu(1-\mu)}{1+\phi}$. La distribución beta es una distribución que pertenece a la familia exponencial y en los modelos de regresión beta el parámetro de media se expresa mediante una función de covariables, mientras que el parámetro ϕ es tratado como un ruido. Para mapear el predictor lineal en el espacio de valores observados en el intervalo (0,1) se puede hacer uso de la función de enlace logística, de tal forma que

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = x'_{it}\beta \quad (10)$$

En los análisis longitudinales o en el caso de que los sujetos se agrupen dentro de unidades de muestreo o entidades geográficas, las mediciones dentro de la misma persona o unidad suelen estar correlacionadas, violando el supuesto de observaciones condicionalmente independientes en los modelos de regresión (Fitzmaurice et al. 2011). Considere el caso de los diseño longitudinales donde $t = 1, \dots, T$ denota el tiempo en el que fue medida la i -ésima observación, con $i = 1, \dots, N$ y ν_t denota un efecto aleatorio para el sujeto i . En el modelo de regresión lineal normal, la inclusión de efectos aleatorios conduce un modelo lineal mixto (LMM) tal que

$$y_{it} = x'_{it}\beta + z'_{it}b_i + \epsilon_{it}; \quad b_i \sim N(0, G) \quad y \quad \epsilon_{it} \sim N(0, \sigma^2) \quad (11)$$

De manera similar al añadir los efectos aleatorios a la regresión beta en (10), se tiene que el GLMM beta está dado por:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = x'_{it}\beta + z'_{it}b_i + \epsilon_{it}; \quad b_i \sim N(0, G) \quad (12)$$

Notese que en la ecuación (12) la distribución de los errores no se especifica, por lo tanto no es de esperarse que los residuales sigan también una distribución normal, esto debido a que la variables respuesta y no es normal. Sin embargo, este problema se puede resolver mediante la creación de residuos fácilmente interpretables para modelos lineales generalizados (mixtos) que están estandarizados a valores entre 0 y 1, y que pueden interpretarse intuitivamente como residuos para el modelo lineal. Esto se logra mediante un enfoque basado en simulación que transforma los residuos a una escala uniforme (Dunn & Smyth 1996, Gelman & Hill 2006).

3. Aplicación

Los datos sobre los ingresos del sector público, y sobre los impuestos en particular, son insumos básicos para la mayoría de las descripciones económicas estructurales y los análisis económicos y se utilizan cada vez más en las comparaciones internacionales. En este sentido, la OCDE proporciona un marco conceptual que define qué ingresos del gobierno deben considerarse con impuestos y clasifica sus diferentes tipos. Las tablas¹ muestran entre otras cosas, datos de ingresos por tipo de impuesto en moneda nacional, como porcentaje del Producto Interno Bruto (PIB).

Si la facturación electrónica mejora la gestión tributaria y reduce la evasión fiscal, esto debería verse reflejado por ejemplo, en el impuesto recaudado por bienes y servicios (IBS). Por esta razón se decide tomar el IBS como variable respuesta expresada como porcentaje del PIB (de ahora en adelante se hará referencia esta variable como IBS). Se tienen en cuenta solo países latinoamericanos (debido a que tiene condiciones socio-económicas más similares) y se excluye del análisis Venezuela y Puerto Rico. El primero debido a que es una economía que tiene bastante problemas como la hiperinflación y por lo tanto puede generar ruido no deseado en los diferentes análisis. El segundo, se excluye debido a que no se cuenta con información reportada en el portal de información de la OCDE. Para cada país, se investigó el año en el cual empezó la facturación electrónica con el fin de poder relacionarlo con la variable de interés y así poder establecer si existe o no una relación entre la implementación de la FE y y el IBS. La ventana de tiempo utilizada para el análisis es desde el año 1990 hasta el 2019.

¹La información fue obtenida del portal de estadísticas de la OCDE <https://stats.oecd.org/>

La base de datos utilizada esta conformada por 538 registros que identifican el año, el país, el IBS (como porcentaje del PIB) y una variable dicotómica que toma el valor de uno (1) si el país en determinado año empezó a implementar FE y de ahí en adelante toma el mismo valor. En otro caso toma el valor de cero (0). En la figura 1 se puede observar que países como Brasil y Chile son los primeros en empezar la implementación de FE (línea vertical) desde el año 2003, seguido de otros países como México (2005), Costa Rica y Guatemala (2007), entre otros. De la misma manera se observa que la tendencia general al recaudo del IBS es positiva después de la implementación de la FE, lo cual lleva a pensar que existe alguna incidencia positiva de la FE sobre el recaudo. Sin embargo, es necesario aclarar que posterior a estos periodos de implementación pueden haber surgido diferentes políticas públicas que también hallan incidido en el recaudo del IBS y que resulta difícil conocer la minucia de cada país a lo largo de esta ventana de observación. Por esta razón es necesario incluir un efecto dentro del modelo que contemple estas variaciones a través del tiempo.

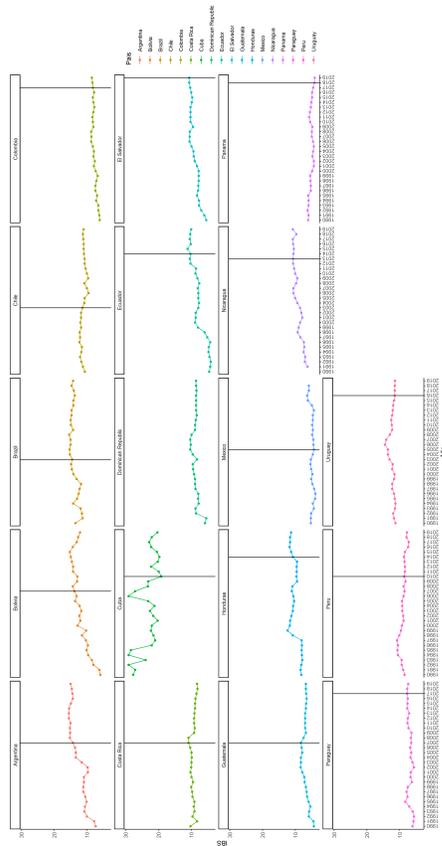


Figura 1: Comportamiento del IBS (%) a través de los años en los países de latinoamerica vs el año de implementación de FE (línea vertical).

Por otra parte, la figura 2 muestra que cada uno de los países presenta un promedio del IBS entre los años 1990-2019 diferente, destacando que Cuba es aquel con mayor valor medio del IBS frente al resto de países. Sin embargo, los intervalos de confianza (barras verticales) para la media de cada país muestran no ser muy distantes lo cual nos lleva sospechar que cada país debido a sus dinámicas económicas recauda a un ritmo constante a través del tiempo. En la figura 3 se observa en general que el IBS a través del tiempo ha tenido un comportamiento con tendencia alcista, es decir, cada vez se genera mayor recaudo con uno que otro periodo de bajas y en términos de variabilidad se logra ver que esta es alta para cada año pero homogénea entre estos. Siendo así y con base a estos análisis exploratorios los GLMM permiten a capturar este tipo de correlación en el tiempo y a diferenciar los efectos propios de cada país. Con el fin de evitar que los resultados y los modelos utilizados en este artículo se vean afectados por el comportamiento tan diferente que tiene Cuba frente a los demás países, este no se tendrá en cuenta de ahora en adelante.

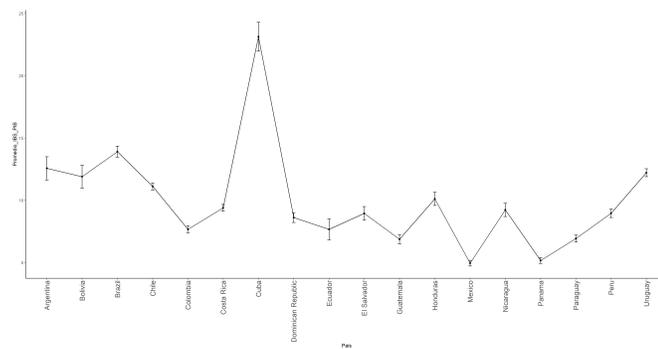


Figura 2: Promedio e intervalo de confianza del IBS (%) por país

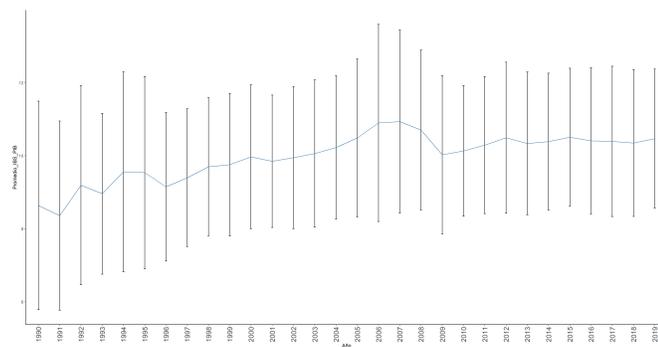


Figura 3: Promedio e intervalo de confianza del IBS (%) por periodo

3.1. Modelo inicial (Modelo 1)

Dado que es de interés conocer el efecto marginal promedio de cada país del IBS con la variable dicotómica de FE, se incluyen estas como efectos fijos dentro del modelo y por otra parte la variable identificadora del año se toma como efecto aleatorio. En este sentido, la forma funcional del GLMM se puede especificar como:

$$\begin{cases} y_{it} \sim \text{Beta}(\mu_{it}\phi_i, \phi_i(1 - \mu_{it})) \\ g(\mu_{it}) = \eta_{1,it} = x'_{it}\beta + z'_{it}b_i + \epsilon_{it} = x'_{it}\beta + \nu_t + \epsilon_{it} \end{cases} \quad (13)$$

Donde:

- y_{it} es el recaudo de impuestos por bienes y servicios como porcentaje del PIB del país i en el año t .
- x_{it} indica si el país i en el año t implementa o no factura electrónica.
- β es el vector de coeficientes asociados al vector x_{it} .
- ν_t es el efecto aleatorio del año t y se asume que $\nu_t \sim N(0, G)$.
- ϵ_{it} es el error del modelo para el país i en el año t .
- ϕ_i es conocido como el parámetro de precisión.
- $g(\cdot)$ es la función de enlace logit.

Partiendo del modelo general dado en la ecuación (13), se generan todos los posibles modelos a partir de las distintas combinaciones de las variables explicativas, en la tabla 1 se encuentran estas combinaciones ordenadas de menor a mayor según el criterio de Akaike corregido (AICc), además se encuentra su peso relativo de evidencia (*Akaike weight*) que nos indica la probabilidad de que un modelo sea el mejor modelo aproximado en relación con nuestros datos. En este sentido y siguiendo el criterio de parsimonia, el modelo que presenta un mejor ajuste es el modelo que no incluye la variable de factura electrónica (FE). Sin embargo, la mejora que se obtiene al eliminar esta variable del modelo no es importante, por lo cual el modelo que se utilizará para explicar el IBS es el modelo presentado en la ecuación (13).

Tabla 1: Modelos de todas las posibles combinaciones de variables explicativas

Id	cond((Int))	disp((Int))	cond(FE)	cond(Pais)	df	logLik	AICc	delta	weight
4	-1.96	+	+	+	20	1532.9	-3024.0	0.0	0.63
3	-1.95	+		+	19	1531.3	-3023.0	1.1	0.37
2	-2.34	+	+		4	1112.7	-2217.4	806.7	0.00
1	-2.29	+			3	1098.1	-2190.1	833.9	0.00

En la tabla 2, se puede observar el cálculo del R^2 propuesto por Nakagawa & Schielzeth (2013), así como el coeficiente de correlación intraclase (ICC) y la raíz

del error cuadrático medio (RMSE) del modelo seleccionado. De allí se observa que la inclusión de efectos aleatorios en el modelo brinda un mejor ajuste en el modelo ya que su $R^2_{condicional}$ es mayor que el $R^2_{marginal}$. De la misma manera el ICC puede ser visto como el porcentaje de varianza explicada por la inclusión de los efectos aleatorios (Nakagawa et al. 2017), que para este caso es del 33 %, lo cual indica que contemplar que utilizar efectos es relevante en el modelo. Por último, el RMSE del modelo es muy bajo, lo cual es también un buen indicador de ajuste del modelo.

Tabla 2: Estadísticas de ajuste del modelo

$R^2_{condicional}$	$R^2_{marginal}$	ICC	RMSE
0.88	0.82	0.33	0.01

Las respectivas estimaciones de los efectos fijos del modelo y estadísticos de significancia se encuentran en la tabla 3. De acuerdo con esto se puede ver que el hecho de implementar factura electrónica hace que el IBS aumente en 4 % en promedio y aunque no resuelve ser significativo (a un nivel de significancia del 5 %) hasta el momento es un buen indicador de que estos cambios en el sistema transaccional de los países generan un movimiento positivo en el recaudo de este tipo de impuestos. Respecto a los coeficientes de los países se observa que el único país que tiene un IBS mayor a respecto a Argentina es Brasil, el resto de los países posee menores niveles de recaudo de dicho impuesto.

Tabla 3: Efectos fijos del modelo

Coefficiente	Estimación	Error Std.	Estadístico z	$\Pr(> z)$
(Intercept)	-1.96	0.03	-69.76	0.00
FE1	0.04	0.02	1.81	0.07
PaisBolivia	-0.06	0.03	-2.03	0.04
PaisBrazil	0.12	0.03	3.90	0.00
PaisChile	-0.14	0.03	-4.45	0.00
PaisColombia	-0.52	0.04	-14.73	0.00
PaisCosta Rica	-0.32	0.03	-9.74	0.00
PaisDominican Republic	-0.40	0.04	-11.30	0.00
PaisEcuador	-0.56	0.04	-15.76	0.00
PaisEl Salvador	-0.36	0.03	-10.40	0.00
PaisGuatemala	-0.66	0.04	-18.35	0.00
PaisHonduras	-0.23	0.03	-6.99	0.00
PaisMexico	-1.01	0.04	-25.06	0.00
PaisNicaragua	-0.34	0.03	-10.02	0.00
PaisPanama	-0.95	0.04	-23.77	0.00
PaisParaguay	-0.64	0.04	-17.42	0.00
PaisPeru	-0.37	0.03	-11.19	0.00
PaisUruguay	-0.01	0.03	-0.42	0.68

En la figura 4, se puede ver gráficamente el exponencial de los coeficientes tanto de

los efectos fijo como de los efectos aleatorios, los puntos marcados en azul indican un impacto positivo en el recaudo del IBS tanto en los efectos fijo como en los efectos aleatorios. Sobre esto últimos es interesante ver que antes del año 2000 aprox. los países en general tenían un comportamiento a la baja sobre el IBS y va aumentando a medida que pasa el tiempo.

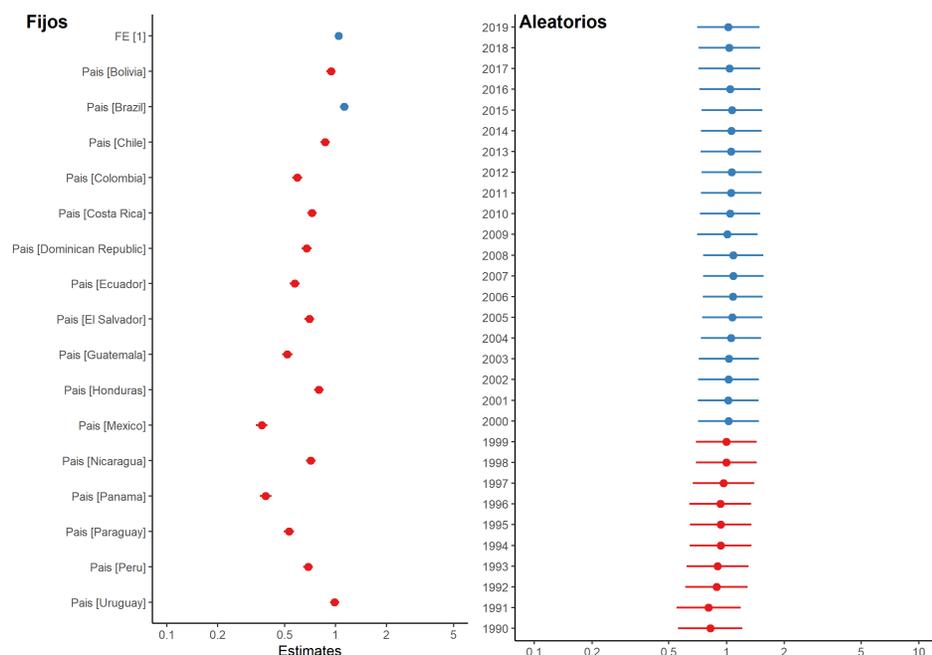


Figura 4: Exponencial de los coeficientes de los efectos fijos y aleatorios

En la tabla 4 se pueden observar las medias marginales estimadas de cada país al implementar FE o no, así como sus errores estándar y los límites inferiores y superiores a un nivel de confianza del 95 %. Para el caso de Colombia se observa que haber implementado FE el IBS aumenta en promedio un 0.3 % frente a no hacerlo. En general se espera que en promedio el IBS se encuentre entre 7.4 % y 8.5 % a un nivel de significancia del 95 %. Respecto a los demás países se puede ver que Argentina es aquel con mayor IBS estimado, mientras que México y Panamá son aquellos con menor IBS.

Tabla 4: Media marginales estimadas según FE por cada país junto con su intervalo de confianza al 95 %

FE	País	IBS	SE	LI	LS
0	Argentina	0.123	0.003	0.117	0.129
1	Argentina	0.127	0.003	0.121	0.134

Continuación en la siguiente página

Tabla 4 – continuación de la página anterior

FE	País	IBS	SE	LI	LS
0	Bolivia	0.116	0.003	0.111	0.122
1	Bolivia	0.120	0.003	0.114	0.127
0	Brazil	0.136	0.003	0.130	0.143
1	Brazil	0.141	0.003	0.135	0.148
0	Chile	0.108	0.003	0.103	0.114
1	Chile	0.112	0.003	0.107	0.118
0	Colombia	0.077	0.002	0.073	0.081
1	Colombia	0.080	0.003	0.074	0.085
0	Costa Rica	0.092	0.003	0.087	0.097
1	Costa Rica	0.096	0.003	0.090	0.101
0	Dominican Republic	0.086	0.002	0.082	0.091
1	Dominican Republic	0.089	0.003	0.084	0.096
0	Ecuador	0.074	0.002	0.070	0.079
1	Ecuador	0.077	0.003	0.072	0.082
0	El Salvador	0.089	0.002	0.085	0.094
1	El Salvador	0.093	0.003	0.087	0.099
0	Guatemala	0.068	0.002	0.064	0.072
1	Guatemala	0.070	0.002	0.066	0.075
0	Honduras	0.100	0.003	0.095	0.105
1	Honduras	0.104	0.003	0.098	0.110
0	Mexico	0.049	0.002	0.045	0.052
1	Mexico	0.051	0.002	0.047	0.054
0	Nicaragua	0.091	0.002	0.086	0.096
1	Nicaragua	0.094	0.003	0.089	0.100
0	Panama	0.051	0.002	0.048	0.055
1	Panama	0.053	0.002	0.049	0.058
0	Paraguay	0.069	0.002	0.065	0.073
1	Paraguay	0.072	0.003	0.067	0.077
0	Peru	0.088	0.002	0.083	0.093
1	Peru	0.091	0.003	0.086	0.097
0	Uruguay	0.122	0.003	0.116	0.127
1	Uruguay	0.126	0.004	0.119	0.133

3.1.1. Diagnóstico

El primer supuesto a probar es la normalidad de los efectos aleatorios, en la figura 5 se puede ver el diagrama cuantil-cuantil y mediante la prueba de Kolmogorov-Smirnov (Conover 1999) con un p-valor de 0.067 a un nivel de significancia del 5 % se puede afirmar que los residuales siguen aproximadamente distribución normal. Sin embargo, este p-valor está muy cerca al nivel de significancia, por lo cual no consideramos que efectivamente se puede afirmar con certeza que si son normales.

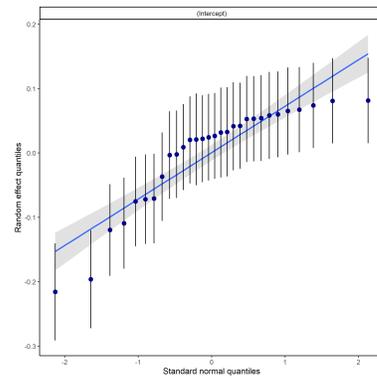


Figura 5: Diagrama cuantil-cuantil para los efectos aleatorios del modelo

Por otra parte, en este modelo la distribución de la respuesta no es normal, por lo cual no es de esperarse que los residuales totales del modelo comporten normalmente. La figura 6 resume múltiples cosas, la primera es que se puede observar que los residuales no siguen una distribución uniforme acorde a la metodología planteada. También se lleva a cabo un test de dispersión no paramétrico de varianza basado en el ratio de la varianza de los residuales observados contra los residuales simulados y se evidencia que los residuales tienen varianza constante. Por último, en la parte derecha del gráfico se encuentran las predicciones del modelo *vs* los residuales del mismo, allí es de esperarse que no muestren ningún patrón de tendencia y/o agrupamiento ya que de ser así se tendrían errores de especificación del modelo. En la parte derecha del gráfico se evidencia lo anteriormente dicho, se lleva a cabo una regresión cuantílica en los cuantiles 1, 2 y 3; en la cual se compara la locación observada *vs* la esperada (0.5) y se marca en rojo cuando la regresión ajustada se desvía de la esperada, por lo que se concluye que los residuales presentan una tendencia sobre los cuantiles 1 y 2.

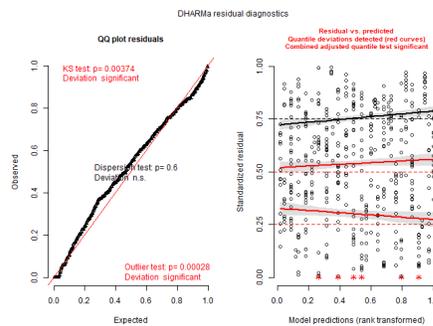


Figura 6: Diagrama cuantil-cuantil de los residuales para contraste de la uniformidad (lado izquierdo) y diagrama de residuales contra los valores predichos (lado derecho)

Teniendo en cuenta este diagnóstico, se observa que el supuesto de uniformidad en los residuales no se cumple, además supone existir algún tipo de sobre-dispersión, razón por la cual se procede a probar un nuevo modelo con covariables explicativas para el parámetro de dispersión.

3.2. Modelo con covariables explicativas para el parámetro ϕ (Modelo 2)

Retomando lo mencionado en la sección 2, ϕ es conocido como el parámetro de precisión que puede ser modelado con función de enlace $h(\cdot)$ de tal manera que $h(\phi_i) = \eta_{2,i} = w_i'\delta$, donde w_i es similar al vector x_{it} pero con las covariables a utilizar para el modelo de dispersión y δ el vector de coeficientes asociado a la matriz W (Hardin et al. 2007). Se escogen como variables para la matriz W las variables de Año y País ya que son aquellas donde se espera que sean suficientes para explicar tanto las diferencias de variabilidad del IBS entre países como las diferencias entre los años. En este sentido la forma funcional del modelo está dada por:

$$\begin{cases} y_{it} \sim \text{Beta}(\mu_{it}\phi_i, \phi_i(1 - \mu_{it})) \\ g(\mu_{it}) = \eta_{1,it} = x'_{it}\beta + z'_{it}b_i + \epsilon_{it} = x'_{it}\beta + \nu_t + \epsilon_{it} \\ h(\phi_i) = \eta_{2,i} = w_i'\delta \end{cases} \quad (14)$$

Donde $h(\cdot)$ al igual que $g(\cdot)$ se escoge como la función de enlace logit. La tabla 5 muestra todos los posibles modelos que se pueden ajustar ordenados de menor a mayor por el AICc, de allí se puede observar que el modelo más conservador en cuanto a interpretabilidad y ajuste, es el modelo con Id 16.

Tabla 5: Modelos de todas las posibles combinaciones de variables explicativas bajo el modelado del parámetro de precisión

Id	cond((Int))	disp((Int))	cond(FE)	cond(Pais)	disp(Año)	disp(Pais)	logLik	AICc	weight
15	-1.79	3.58		+	+	+	1671.8	-3196.8	0.71
16	-1.81	3.67	+	+	+	+	1672.2	-3195.0	0.29
11	-1.94	5.72		+		+	1605.2	-3135.1	0.00
12	-1.94	5.72	+	+		+	1605.2	-3132.8	0.00
8	-1.83	4.36	+	+	+		1615.2	-3121.6	0.00
7	-1.80	4.44		+	+		1605.7	-3105.2	0.00
4	-1.96		+	+			1532.9	-3024.0	0.00
3	-1.95			+			1531.3	-3023.0	0.00
10	-2.33	4.46	+			+	1237.0	-2432.2	0.00
9	-2.29	4.23				+	1230.9	-2422.2	0.00
14	-2.30	3.13	+		+	+	1246.4	-2384.2	0.00
13	-2.28	2.97			+	+	1244.1	-2381.9	0.00
2	-2.34		+				1112.7	-2217.4	0.00
1	-2.29						1098.1	-2190.1	0.00
6	-2.34	4.37	+		+		1115.5	-2160.3	0.00
5	-2.29	4.21			+		1100.8	-2133.2	0.00

En cuanto a los criterios presentados en la tabla 2, para el nuevo modelo no se incluye las estadísticas $R^2_{condicional}$, $R^2_{marginal}$ e ICC, ya que no incluyen el efecto

del modelo de dispersión. Sin embargo, este modelo sigue siendo mejor en términos del AICc.

Tabla 6: Estadísticas de ajuste bajo el modelado del parámetro ϕ

<u>RMSE</u>
<u>0.01</u>

En la tabla 7 se puede observar que bajo el nuevo modelo el hecho de implementar factura electrónica hace que el IBS aumente en promedio un 2%, aunque este aumento sigue sin ser significativo (a un nivel de significancia del 5%). De igual forma se mantiene la interpretación inicial, de que esto es un buen indicador en cuanto a un movimiento positivo en el recaudo del IBS.

Tabla 7: Efectos fijos bajo incluyendo el modelado del parámetro ϕ

Coefficiente	Estimación	Error Std.	Estadístico z	Pr($> z $)
(Intercept)	-1.81	0.03	-56.53	0.00
FE1	0.02	0.02	0.92	0.36
PaisBolivia	-0.10	0.03	-3.27	0.00
PaisBrazil	-0.00	0.02	-0.14	0.89
PaisChile	-0.37	0.03	-11.54	0.00
PaisColombia	-0.65	0.03	-23.12	0.00
PaisCosta Rica	-0.52	0.03	-16.21	0.00
PaisDominican Republic	-0.56	0.03	-16.31	0.00
PaisEcuador	-0.56	0.04	-12.72	0.00
PaisEl Salvador	-0.43	0.03	-14.94	0.00
PaisGuatemala	-0.81	0.03	-27.37	0.00
PaisHonduras	-0.41	0.04	-11.23	0.00
PaisMexico	-1.20	0.03	-36.67	0.00
PaisNicaragua	-0.42	0.03	-14.57	0.00
PaisPanama	-1.12	0.03	-34.88	0.00
PaisParaguay	-0.79	0.03	-25.23	0.00
PaisPeru	-0.60	0.03	-19.19	0.00
PaisUruguay	-0.18	0.03	-5.91	0.00

La figura 7 muestra el gráfico de efectos fijos y efectos aleatorios la cual mantiene en términos generales un estructura similar a la figura 4. En la tabla 8 se puede observar para el caso de Colombia que haber implementado FE el IBS aumenta en promedio un 0.1% (0.12% al tener en cuenta toda la precisión decimal) frente a no hacerlo. En general se espera que en promedio el IBS se encuentre entre 7.7% y 8.3% a un nivel de significancia del 95%. De la misma manera se puede interpretar para los demás países.

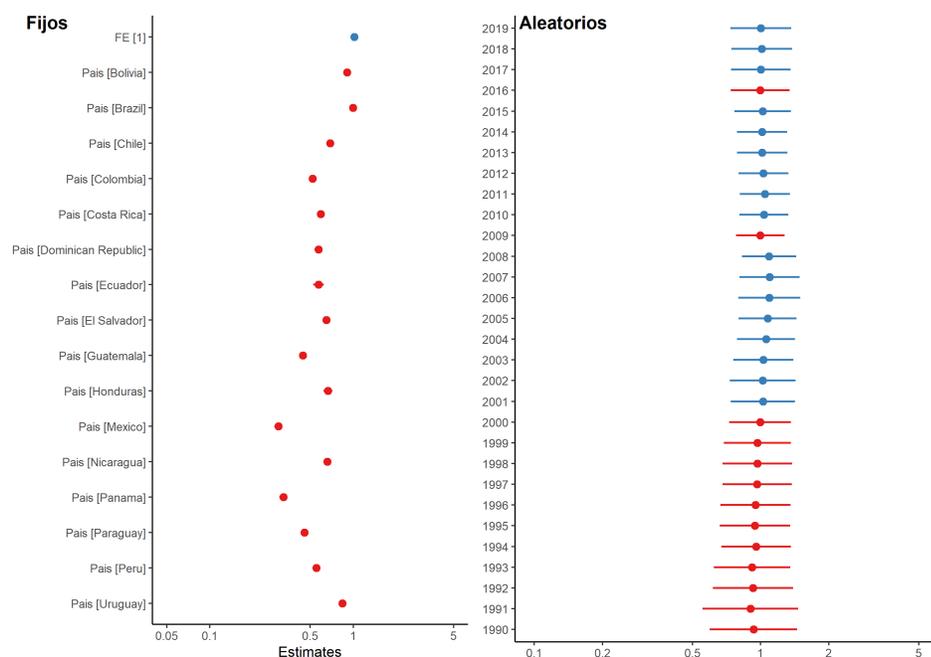


Figura 7: Exponencial de los coeficientes de los efectos fijos y aleatorios del modelo 2

Tabla 8: Media marginales estimadas según FE por cada país junto con su intervalo de confianza al 95 %

FE	País	IBS	SE	LI	LS
0	Argentina	0.141	0.004	0.134	0.149
1	Argentina	0.143	0.003	0.136	0.150
0	Bolivia	0.129	0.003	0.123	0.136
1	Bolivia	0.131	0.003	0.125	0.137
0	Brazil	0.141	0.003	0.135	0.146
1	Brazil	0.143	0.002	0.139	0.146
0	Chile	0.102	0.003	0.097	0.107
1	Chile	0.103	0.002	0.099	0.108
0	Colombia	0.079	0.001	0.077	0.081
1	Colombia	0.080	0.002	0.077	0.083
0	Costa Rica	0.089	0.002	0.085	0.093
1	Costa Rica	0.090	0.002	0.087	0.093
0	Dominican Republic	0.086	0.001	0.083	0.089
1	Dominican Republic	0.087	0.002	0.083	0.091
0	Ecuador	0.086	0.003	0.080	0.093
1	Ecuador	0.087	0.004	0.080	0.095

Continuación en la siguiente página

Tabla 8 – continuación de la página anterior

FE	País	IBS	SE	LI	LS
0	El Salvador	0.096	0.002	0.093	0.100
1	El Salvador	0.098	0.002	0.093	0.102
0	Guatemala	0.068	0.001	0.065	0.071
1	Guatemala	0.069	0.001	0.067	0.071
0	Honduras	0.098	0.002	0.094	0.103
1	Honduras	0.100	0.003	0.095	0.105
0	Mexico	0.047	0.001	0.045	0.050
1	Mexico	0.048	0.001	0.046	0.050
0	Nicaragua	0.098	0.002	0.093	0.102
1	Nicaragua	0.099	0.002	0.094	0.104
0	Panama	0.051	0.001	0.048	0.054
1	Panama	0.052	0.002	0.049	0.055
0	Paraguay	0.070	0.002	0.066	0.073
1	Paraguay	0.071	0.002	0.067	0.075
0	Peru	0.083	0.002	0.080	0.087
1	Peru	0.084	0.002	0.081	0.088
0	Uruguay	0.121	0.002	0.118	0.124
1	Uruguay	0.123	0.002	0.118	0.127

3.2.1. Diagnóstico

En cuanto al supuesto de normalidad de los efectos aleatorios, en la figura 8 se puede ver el diagrama cuantil-cuantil y mediante la prueba de Kolmogorov-Smirnov con un p-valor de 0.72 a un nivel de significancia del 5% se puede afirmar que los residuales siguen aproximadamente distribución normal, lo cual en contraste con el modelo 1 es mucho mejor.

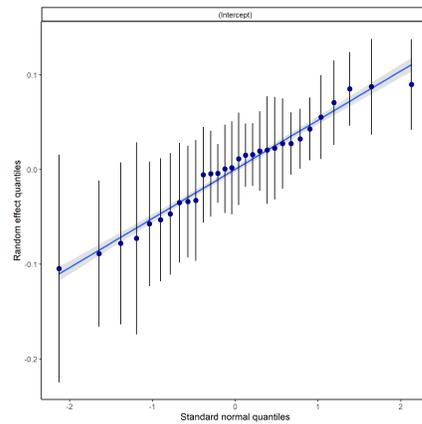


Figura 8: Diagrama cuantil-cuantil para los efectos aleatorios del modelo 2

Por otra parte, como se mencionó anteriormente en este modelo la distribución de la respuesta no es normal, por lo cual no es de esperarse que los residuales totales del modelo comporten normalmente. La figura 9 que a diferencia del modelo 1, en el modelo 2 los residuales siguen una distribución uniforme y que en cuanto a los residuales *vs* las predicciones del modelo, estos muestran un comportamiento mucho más cerca de lo esperado que el modelo 1.

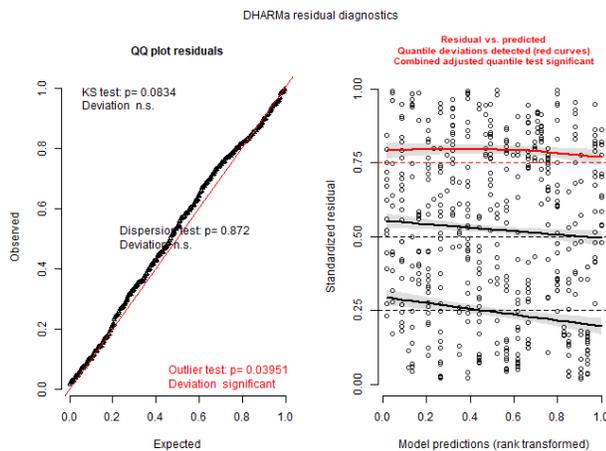


Figura 9: Diagrama cuantil-cuantil de los residuales para contraste de la uniformidad (lado izquierdo) y diagrama de residuales contra los valores predichos (lado derecho) para el modelo 2

3.2.2. Predicción

Una vez ajustado y validado el modelo, se escoge como modelo final el modelo presentado en la ecuación 14 (Modelo 2). La figura 10 muestra las predicción del IBS por país y año comparado con el dato real. Allí se observa que por ejemplo en el caso colombiano las predicciones del modelo son muy cercanas a lo real.

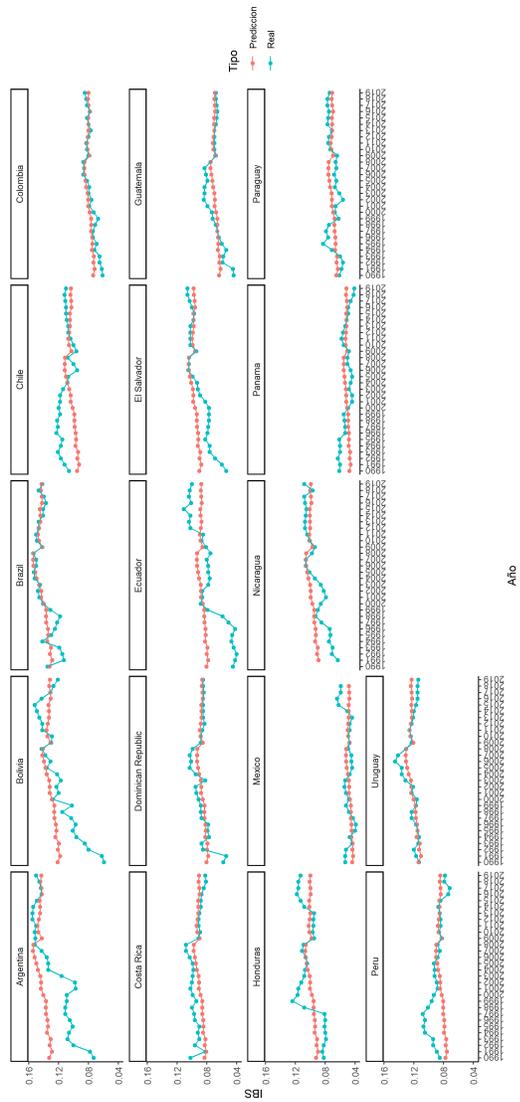


Figura 10: Predicción del IBS (%) a través de los años por cada país

4. Conclusiones y trabajos futuros

En este trabajo se ha presentado la aplicación de un modelo lineal generalizado mixto (GLMM). Esta técnica proporciona el tratamiento de datos que se presumen tiene alguna relación ya sea en tiempo o en estructura. Se modeló el IBS expresado como el porcentaje del PIB de cada país frente a la implementación de FE con una ventana de observación desde 1990 hasta el 2019. Al respecto se plantearon dos (2) modelos, uno que solamente modelaba la media y otro que modelaba la media y la varianza. Los resultados mostraron que el segundo modelo posee un mejor ajuste y logra cumplir los supuestos estadísticos de dicha metodología, razón por la cual este es el modelo predilecto para poder determinar la incidencia de la FE sobre el IBS.

En general se observa que bajo el modelo 2, aunque el efecto de FE no resultó ser estadísticamente significativo, si se evidencia que la incidencia de esta implementación en general a nivel Latinoamérica es positiva. En Colombia, al seguir implementando factura electrónica, el recaudo de impuestos en bienes y servicios como porcentaje del PIB se espera que sea de 8 % en promedio, moviéndose entre 7,7 % y 8,4 % con un nivel de confianza del 95 %. En términos generales se evidencia un comportamiento constante a lo largo de los años sobre el IBS, por lo cual no es de esperarse grandes cambios a través del tiempo. y en algunos casos como por ejemplo Colombia y Paraguay, la ventana de observación a partir de la cual se implementa FE es aún muy corta, por lo cual, es muy pronto para que el modelo capture la incidencia de la implementación de FE en estos países.

Para trabajos futuros es importante considerar otras covariables además de FE que se encuentren medidas de forma estándar en cada uno de los distintos países, con el fin de aislar un poco más el efecto que pueda tener la implementación de FE sobre el IBS, además de contar con más periodos de observación para países que son relativamente “nuevos” en esta implementación.

Recibido:

Aceptado:

Referencias

- Agresti, A. (2015), *Foundations of linear and generalized linear models*, John Wiley & Sons.
- Artana, D. & Templado, I. (2018), ‘Análisis del impacto de la factura electrónica en la argentina’, *Inter-American Development Bank Discussion Paper* (562).
- Bérgolo, M., Ceni, R. & Sauval, M. (2018), Factura electrónica y cumplimiento tributario: Evidencia a partir de un enfoque cuasi-experimental, Technical report, Inter-American Development Bank.
- Castro, H. J., Carrillo, A. Z., Cortés, S. B., Aragón, G. A., Diez, M. E. & Salgado, M. M. S. (2016), ‘Impacto en la evasión por la introducción de la factura electrónica’.
- Conover, W. J. (1999), *Practical nonparametric statistics*, Vol. 350, John Wiley & Sons.
- Dunn, P. K. & Smyth, G. K. (1996), ‘Randomized quantile residuals’, *Journal of Computational and Graphical Statistics* **5**(3), 236–244.
- Fitzmaurice, G. M., Laird, N. M. & Ware, J. (2011), ‘Wiley series in probability and statistics. applied longitudinal analysis . hoboken’.
- Gelman, A. & Hill, J. (2006), *Data analysis using regression and multilevel/hierarchical models*, Cambridge university press.
- Hardin, J. W., Hardin, J. W., Hilbe, J. M. & Hilbe, J. (2007), *Generalized linear models and extensions*, Stata press.
- Jiang, J. & Nguyen, T. (2007), *Linear and generalized linear mixed models and their applications*, Vol. 1, Springer.
- Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. M. (2016), ‘Tmb: Automatic differentiation and laplace approximation’, *Journal of Statistical Software* **70**(5).
*<http://dx.doi.org/10.18637/jss.v070.i05>
- Nakagawa, S., Johnson, P. C. & Schielzeth, H. (2017), ‘The coefficient of determination r^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded’, *Journal of the Royal Society Interface* **14**(134), 20170213.
- Nakagawa, S. & Schielzeth, H. (2013), ‘A general and simple method for obtaining r^2 from generalized linear mixed-effects models’, *Methods in ecology and evolution* **4**(2), 133–142.
- Naritomi, J. (2019), ‘Consumers as tax auditors’, *American Economic Review* **109**(9), 3031–72.

- Nelder, J. A. & Wedderburn, R. W. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society: Series A (General)* **135**(3), 370–384.
- Ramírez, J., Oliva, N. & Andino, M. (2018), Facturación electrónica en Ecuador: Evaluación de impacto en el cumplimiento tributario, Technical report, Inter-American Development Bank.
- Rencher, A. C. & Schaalje, G. B. (2008), *Linear models in statistics*, John Wiley & Sons.
- Rosenbaum, P. R. & Rubin, D. B. (1983), 'The central role of the propensity score in observational studies for causal effects', *Biometrika* **70**(1), 41–55.