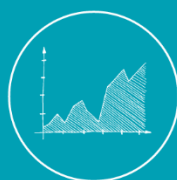
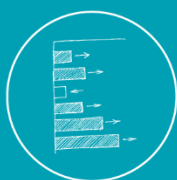




UNA APROXIMACIÓN A LA ESTIMACIÓN EN ÁREAS PEQUEÑAS DE LA DESNUTRICIÓN CRÓNICA INFANTIL EN ECUADOR

2022



Elaboración técnica:



Buenas cifras,
mejores vidas

Autoridades:**Roberto Castillo A.***Director Ejecutivo***Jorge García-Guerrero***Subdirector General***Lorena Moreno***Coordinadora General Técnica de Innovación en
Métricas y Análisis de la Información***Galo Egas G.***Director de Innovación en Métricas y Metodologías***Edición:**

Lorena Moreno

Leandro Chalela

Belén Escola

Autores:

Domingo Morales

Universidad Miguel Hernández de Elche, España

Diego Villacreses

*Instituto Nacional de Estadística y Censos, Ecuador***Citar como:**

INEC (2022). Una Aproximación a la Estimación en Áreas Pequeñas de la Desnutrición Crónica Infantil en Ecuador (Vol. 13, Ser. Cuadernos de Trabajo). Instituto Nacional de Estadística y Censos.

Los Cuadernos de Trabajo son documentos que presentan análisis, tanto metodológicos como temáticos, relacionados a fenómenos sociales, económicos y ambientales con el objetivo de promover la investigación e incentivar el debate.

Las interpretaciones y opiniones expresadas en este documento pertenecen a los autores y no reflejan el punto de vista oficial del Instituto Nacional de Estadística y Censos (INEC). El INEC ha realizado una revisión del documento, no obstante, no garantiza la exactitud de los datos que figuran en el documento.

Se extiende un especial agradecimiento a Vital Strategies, quienes a través de la *Iniciativa Datos para la Salud* financiada por Bloomberg Philanthropies, y en convenio con el INEC, hicieron posible la capacitación y acompañamiento académico impartida por Domingo Morales, para la elaboración del presente documento. Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad de los autores y pueden no coincidir con la posición de Vital Strategies.

Índice de contenidos

1.	Introducción.....	5
2.	Marco teórico	6
2.1	Estimación en áreas pequeñas basada en el modelo NER.....	6
2.1.1.	Modelo de regresión con errores anidados	8
2.1.2.	Distribución condicional de la no-muestra dada la muestra	10
2.1.3.	Predictores óptimos empíricos	10
2.2.	Selección de variables	11
3.	Consideraciones para la aplicación empírica	12
3.1.	Conformidad con las estimaciones directas de nivel superior	12
3.2.	Implicaciones de la Condición de Conformidad	13
3.3.	Estimación del error cuadrático medio	13
4.	Aplicación empírica.....	15
4.1.	Propuesta.....	15
4.2.	Construcción del Modelo.....	17
4.2.1.	Proceso de selección de variables	17
4.2.2.	Proceso de selección de modelo	18
5.	Resultados ECV-2014	20
5.1.	Selección de modelo	20
5.2.	Descriptivos relevantes	22
5.3.	Regresión	23
5.4.	Mapa provincial.....	25
6.	Resultados ENSANUT-2018.....	26
6.1.	Selección de modelo	26
6.2.	Descriptivos relevantes	28
6.3.	Regresión	28
6.4.	Mapa provincial.....	30
7.	Reflexiones finales.....	31
	Referencias	32
	Anexos.....	35

Índice de Gráficos

Gráfico No. 1: AIC y RMSE de Validación Cruzada de Modelos ECV-2014 ...	21
Gráfico No. 2: Total cantones clasificados por Coeficiente de Variación, por modelo. ECV-2014.....	22
Gráfico No. 3: Diferencia tasa DCI frente a promedio nacional por provincia. ECV-2014	23
Gráfico No. 4: EPPCs por cantón, mapa. ECV-2014.....	25

Gráfico No. 5: AIC y RMSE de Validación Cruzada de Modelos ENSANUT-2018 27

Gráfico No. 6: Total cantones clasificados por Coeficiente de Variación, por modelo. ENSANUT-2018.....27

Gráfico No. 7: Diferencia tasa DCI con relación a promedio nacional por provincia. ENSANUT-201828

Gráfico No. 8: EPPCs por cantón, mapa. ENSANUT-2018.....30

Gráfico No. 9: ECV-2014 vs ENSANUT-2018: estimaciones provinciales directas 35

Gráfico No. 10: Distribuciones de residuos por versión de modelo final. ECV-2014 35

Gráfico No. 11: Distribuciones de residuos por versión de modelo final. ENSANUT-201837

Índice de Tablas

Tabla No. 1: Resultados NER, versiones de modelo final. Parte I. ECV-2014.....24

Tabla No. 2: Resultados NER, versiones de modelo final. Parte I. ENSANUT-2018 29

1. Introducción

La Desnutrición Crónica Infantil (DCI) se calcula usando como referencia los estándares sugeridos por la Organización Mundial de la Salud - OMS. La tasa de DCI será el total de infantes (menores de 5 años o 60 meses) que presenten un z-score menor a -2 dividido para el total de infantes. El z-score es la distancia en desvíos estándares de la talla del infante con relación a la mediana propuesta por OMS para su edad en días o meses y sexo (WHO, 2006).

La DCI es una grave problemática que ha afectado constantemente a la población vulnerable de los países en desarrollo. Se ha demostrado que las consecuencias físicas de corto plazo de la DCI se perpetúan en el largo plazo a niveles tan profundos que la reversibilidad es difícil de lograr (Galler, et al., 2012). Por ejemplo, se ha encontrado que la presencia de DCI causa una reducción permanente de ingresos en la adultez (Dewey y Begum, 2011; Campbell, et al., 2012), afecta negativamente a varios aspectos de la salud (Galler, et al., 2012) y reduce considerablemente la capacidad cognitiva en la adultez (Ampaabeng y Ming, 2013; Schoenmaker, et al., 2015).

En Ecuador, al 2018, la DCI afecta a alrededor de uno de cada cuatro infantes. Entre el 2006 y el 2014 se observó una reducción de menos de 2 puntos porcentuales (p.p.) con una variación estadísticamente significativa; y, entre el 2014 y el 2018 se observó una variación no estadísticamente significativa. Afortunadamente, varios estudios concluyen que el costo-efectividad de una inversión en desarrollo de infantes es alto (Heckman, 2002, 2010a, 2010b; García, et al., 2019; Suárez, 2020); por lo cual, mantener dicha inversión en un nivel adecuado es uno de los retos más importantes para países en desarrollo.

Por ello, la medición y monitoreo de la DCI se convierte en una de las principales metas de desarrollo en países con altas prevalencias (Behrman y Hoddinott, 2005). Uno de los insumos más importantes para su reducción es la disposición de información confiable con la mayor desagregación territorial posible (Elbers, Lanjouw & Lanjouw, 2003; Molina, 2019). Así, las técnicas de Estimación en Áreas Pequeñas (Small Area Estimation - SAE) son necesarias.

Este documento presenta una aplicación de la metodología SAE a la estimación de la DCI en cantones de Ecuador. Específicamente, la propuesta consiste en usar predictores óptimos empíricos bajo el modelo de regresión con errores anidados (Nested Error Regression – NER). La sección 2 describe el marco teórico de SAE-NER, y las secciones 3 y 4 presenta una detallada aplicación de SAE-NER para la DCI en Ecuador. Las secciones 5 y 6 discute los resultados en base a la Encuesta de Condiciones de Vida 2014 – ECV y la Encuesta de Salud y Nutrición 2018 – ENSANUT, respectivamente; mientras que la sección 7 presenta las reflexiones finales.

2. Marco teórico

Esta sección establece un marco teórico dividido en tres subsecciones. Los dos primeros apartados contienen un resumen de la literatura actual requerida para nuestra aplicación. El tercer apartado presenta el desarrollo teórico necesario para implementar la metodología estadística propuesta.

2.1 Estimación en áreas pequeñas basada en el modelo NER

El modelo NER (Nested Error Regression) o modelo de regresión con errores anidados es un modelo mixto lineal (LMM - Linear Mixed Model). Los LMM establecen una relación lineal entre una variable objetivo y algunas variables explicativas. Al mismo tiempo, los LMM pueden manejar datos en los que las observaciones no son independientes. A diferencia de los modelos lineales de efectos fijos (LM - Linear Models), los LMM pueden modelar errores correlacionados. Los LMM son generalizaciones de los LM que permiten hacer un mejor análisis de una variable dependiente. Estos modelos pueden incorporar efectos aleatorios, efectos jerárquicos, medidas repetidas, correlaciones espaciales y temporales.

Los LMM tienen una amplia aplicabilidad en estimación en áreas pequeñas, donde la flexibilidad para combinar diferentes fuentes de información y explicar diferentes fuentes de errores es de gran ayuda. Los modelos mixtos aumentan la información efectiva utilizada en el proceso de estimación al vincular todas las observaciones de la muestra, y al mismo tiempo permitir distintos tipos de variabilidad entre áreas. De entre los libros que tratan de LMM destacan las obras de Searle, Casella y McCulloch (1992), Longford (1995), McCulloch y Searle (2001), Goldstein (2003), Demidenko (2004) y Jiang (2007).

Así, el modelo NER es un LMM con un efecto aleatorio que tiene D niveles asociados a los dominios o áreas pequeñas. Fueron Battese y otros (1988) quienes aplicaron el modelo NER por primera vez a problemas de SAE: predijeron las áreas plantadas de maíz y soja en 12 condados del centro-norte de Iowa. Desde entonces, el modelo NER es el modelo básico a nivel de unidad en SAE. Con respecto a las aplicaciones a datos reales, los predictores basados en modelos NER se han empleado ampliamente para estimar parámetros de dominio.

Basándose en el capítulo 7 de Morales, et al. (2021), a continuación, se describe el modelo NER, se proporciona la teoría básica para el cálculo de predictores óptimos empíricos de parámetros de dominio con estructura no necesariamente lineal, y se introduce un procedimiento bootstrap para la estimación de los errores cuadráticos medios.

Consideramos una variable cuantitativa de interés, y_{dj} , donde el subíndice d representa el dominio en el que se encuentra el individuo j . En el contexto SAE, el dominio puede referirse a cualquier subconjunto de interés de una población finita definido por variables geográficas, grupo de edad, género, grupo género-edad, situación migratoria, etc.

En lo referente a la estructura de la población, adoptamos la notación del capítulo 10 de Morales et al. (2021). Consideramos una población finita U de tamaño N . Sea D el número de dominios y sea N_d el número de unidades (individuos, hogares, empresas, etc.) en el dominio d . Por tanto, los subíndices tienen el recorrido $d = 1, \dots, D$ y $j = 1, \dots, N_d$. Suponemos que la población está particionada en dominios U_d , $d = 1, \dots, D$, de modo que $U = \bigcup_{d=1}^D U_d$, $U_{d_1} \cap U_{d_2} = \emptyset$ si $d_1 \neq d_2$ y $N = \sum_{d=1}^D N_d$. Una muestra de tamaño n extraída de la población, $s \subset U$, se descompone de manera similar en las submuestras s_d correspondientes a los dominios; es decir, $s = \bigcup_{d=1}^D s_d$ y $n = \sum_{d=1}^D n_d$.

$$\mathbf{y}_s = \underset{1 \leq d \leq D}{col}(\mathbf{y}_{ds}), \mathbf{y}_{ds} = \underset{1 \leq j \leq n_d}{col}(\mathbf{y}_{dj}), \quad (1)$$

$$\mathbf{y}_r = \underset{1 \leq d \leq D}{col}(\mathbf{y}_{dr}), \mathbf{y}_{dr} = \underset{n_d+1 \leq j \leq N_d}{col}(\mathbf{y}_{dj}), \quad (2)$$

$$\mathbf{y}_d = (\mathbf{y}'_{ds}, \mathbf{y}'_{dr})'. \quad (3)$$

Sea $h_d: R^{N_d} \mapsto R$ una función medible no necesariamente lineal. Supongamos que:

1. disponemos de \mathbf{y}_s , pero se desconoce \mathbf{y}_r ,
2. queremos obtener una estimación $\hat{h}_d = \hat{h}_d(\mathbf{y}_s)$ del parámetro $h_d = h_d(\mathbf{y}_d)$,
3. pretendemos que el error estándar de \hat{h}_d sea menor que algún umbral bajo el cual la estimación sea considerada útil o "representativa".

En tal caso, conviene usar procedimientos SAE para obtener una estimación adecuada de \hat{h}_d .

Cuando la variable objetivo sigue una distribución normal, se puede recurrir a una metodología SAE derivada de un modelo NER. En particular, se pueden usar los predictores óptimos empíricos (EBP - Empirical Best Predictors). Molina y Rao (2010) asignaron esta denominación ya que, al cumplirse ciertos supuestos, los EBP dan estimaciones óptimas (con error cuadrático medio mínimo dentro de clase de predictores aproximadamente insesgados) considerando la información disponible. Estos autores presentaron una metodología de cálculo para obtener el EBP para una amplia variedad de funciones h_d , entre ellas, la del tipo el indicador de pobreza FGT_0 definido por Foster, Greer y Thorbecke (1984).

Para la estimación de la DCI en áreas pequeñas, la variable de interés y_{dj} es el estándar de talla para la edad de niños entre 0 y menos de 60 meses propuesto por la OMS (2006) para un niño con k meses de vida cumplidos, $0 \leq k < 60$; es decir:

$$y_{dj} = \frac{z_{dj} - \bar{Z}^{(k)}}{S_Z^{(k)}}, \quad (4)$$

donde z_{dj} es la talla en centímetros del niño j del dominio d , y $\bar{Z}^{(k)}$ y $S_Z^{(k)}$ son valores de estandarización (media y desviación estándar en cm) para la edad de k meses. La DCI del dominio d es:

$$F_d = \frac{1}{N_d} \sum_{j=1}^{N_d} F_{dj}, \quad F_{dj} = I(y_{dj} < y_0), \quad d = 1, \dots, D, j = 1, \dots, N_d, \quad (5)$$

donde y_0 es un valor umbral, $I(y_{dj} < y_0) = 1$ si $y_{dj} < y_0$ y $I(y_{dj} < y_0) = 0$ en caso contrario. La DCI del dominio d , F_d , está definida usando la misma fórmula que el indicador de pobreza FGT_0). En efecto, la DCI cumple las siguientes características: (i) y_{dj} es una variable continua que representa el estándar de talla para la edad, (ii) $y_0 = -2$, (iii) F_{dj} una variable dicotómica donde "1" representará un infante con DCI [nótese que (iii) se deriva de (i) y (ii)]. De esta manera, se puede utilizar la metodología de Molina y Rao (2010) basada en modelos NER para la estimación en áreas pequeñas de funciones del tipo FGT_0 .

2.1.1. Modelo de regresión con errores anidados

Como se introdujo anteriormente, el modelo NER es un LMM con un único efecto aleatorio que tiene D niveles, formalmente se describe como:

$$\mathbf{y}_{dj} = \mathbf{x}_{dj}\boldsymbol{\beta} + u_d + e_{dj}, \quad d = 1, \dots, D, j = 1, \dots, N_d, \quad (6)$$

donde

- \mathbf{y}_{dj} es la variable objetivo, medida en la unidad j del dominio d .
- \mathbf{x}_{dj} es un vector fila de dimensión p , $\mathbf{x}_{dj} = (x_{1dj}, x_{2dj}, \dots, x_{pdj})$, que contiene los valores de las variables auxiliares. En el archivo de datos, los valores de las variables auxiliares se disponen en columnas.
- $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ es un vector columna que contiene los parámetros de la regresión,
- $u_d \sim N(0, \sigma_u^2)$ es el efecto aleatorio asociado al dominio, y
- $e_{dj} \sim N(0, \sigma_e^2)$ es el error del modelo.

El modelo NER asume que los u_d 's y los e_{dj} 's son todos independientes. Este modelo puede considerarse como un modelo con intercepto aleatorio. La estructura de covarianzas del modelo NER es:

$$\text{var}(\mathbf{y}_{dj}) = \sigma_u^2 + \sigma_e^2, \quad \text{cov}(\mathbf{y}_{dj_1}, \mathbf{y}_{dj_2}) = \sigma_u^2 \text{ si } j_1 \neq j_2, \quad (7)$$

$$\text{cov}(\mathbf{y}_{d_1j_1}, \mathbf{y}_{d_2j_2}) = 0 \text{ si } d_1 \neq d_2, \quad (8)$$

de modo que las observaciones son dependientes dentro de los dominios e independiente entre los dominios. Definimos los vectores $\mathbf{e}_d = \underset{1 \leq j \leq N_d}{\text{col}}(e_{dj})$ y $\mathbf{1}_a = \underset{1 \leq j \leq a}{\text{col}}(1)$, donde el operador $\underset{1 \leq j \leq a}{\text{col}}$ define un vector columna de dimensión a . Este modelo puede escribirse en la forma vectorial:

$$\mathbf{y}_d = \mathbf{X}_d\boldsymbol{\beta} + \mathbf{1}_{N_d}u_d + \mathbf{e}_d, \quad d = 1, \dots, D, \quad (9)$$

donde

- los vectores $\mathbf{e}_d \sim N(0_{N_d}, \sigma_e^2 \mathbf{I}_{N_d})$ son independientes, \mathbf{I}_a es la matriz identidad $a \times a$, $\mathbf{0}_a = \underset{1 \leq j \leq a}{\text{col}}(0)$, y
- los vectores $\mathbf{y}_d \sim N(\boldsymbol{\mu}_d, \mathbf{V}_d)$ son independientes, $\boldsymbol{\mu}_d = \mathbf{X}_d\boldsymbol{\beta}$, $\mathbf{V}_d = \sigma_u^2 \mathbf{1}_{N_d} \mathbf{1}_{N_d}' + \sigma_e^2 \mathbf{I}_{N_d}$ y $\mathbf{X}_d = \underset{1 \leq j \leq N_d}{\text{col}}(\mathbf{x}_{dj})$.

En la práctica, solo disponemos de la información de una muestra extraída de la población finita U , donde (i) \mathbf{y}_{ds} es un sub-vector de \mathbf{y}_d correspondiente a los elementos muestrales con tamaño n_d , (ii) \mathbf{y}_{dr} el sub-vector de \mathbf{y}_d correspondiente a los elementos no muestrales con tamaño $N_d - n_d$. Sin pérdida de generalidad, escribimos $\mathbf{y}_d = (\mathbf{y}_{ds}', \mathbf{y}_{dr}')'$. El vector \mathbf{y}_d tiene tamaño N_d , donde $N_d = n_d + (N_d - n_d)$. Asumimos que los sub-vectores muestrales \mathbf{y}_{ds} siguen los modelos marginales derivados del modelo NER poblacional; es decir,

$$\mathbf{y}_{ds} = \mathbf{X}_{ds}\boldsymbol{\beta} + \mathbf{1}_{n_d}u_d + \mathbf{e}_{ds}, \quad d = 1, \dots, D, s = 1, \dots, n_d, \quad (10)$$

donde

- los vectores $\mathbf{e}_{ds} \sim N(\mathbf{0}_{n_d}, \sigma_e^2 \mathbf{I}_{n_d})$ son independientes,
- los vectores $\mathbf{y}_{ds} \sim N(\boldsymbol{\mu}_{ds}, \mathbf{V}_{ds})$ son independientes, con vector de medias $\boldsymbol{\mu}_{ds} = \mathbf{X}_{ds}\boldsymbol{\beta} + \mathbf{1}_{n_d}u_d$ y matriz de covarianza $\mathbf{V}_{ds} = \sigma_u^2 \mathbf{1}_{n_d} \mathbf{1}_{n_d}' + \sigma_e^2 \mathbf{I}_{n_d}$.
- La matriz de variables independientes es $\mathbf{X}_d = \underset{1 \leq j \leq n_d}{\text{col}}(x_{dj})$ y contiene por columnas los valores fijos (no aleatorios) de p variables auxiliares.

Si $\sigma_e^2 > 0$ y $\sigma_u^2 > 0$ son conocidos, entonces podemos calcular el estimador insesgado lineal óptimo (BLUE - Best Linear Unbiased Estimator) de $\boldsymbol{\beta}$ y el predictor insesgado lineal óptimo (BLUP - Best Linear Unbiased Predictor) de u_d , $d = 1, \dots, D$; es decir,

$$\tilde{\boldsymbol{\beta}} = \left(\sum_{d=1}^D \sum_{s=1}^{n_d} \mathbf{X}_{ds}' \mathbf{V}_{ds}^{-1} \mathbf{X}_{ds} \right)^{-1} \sum_{d=1}^D \sum_{s=1}^{n_d} \mathbf{X}_{ds}' \mathbf{V}_{ds}^{-1} \mathbf{y}_{ds}, \quad \tilde{u}_d = \sigma_u^2 \mathbf{1}_{n_d}' \mathbf{V}_{ds}^{-1} (\mathbf{y}_{ds} - \mathbf{X}_{ds} \tilde{\boldsymbol{\beta}}), \quad (11)$$

donde

$$\mathbf{V}_{ds}^{-1} = \frac{1}{\sigma_e^2} \left(\mathbf{I}_{n_d} - \frac{\gamma_d}{n_d} \mathbf{1}_{n_d} \mathbf{1}_{n_d}' \right), \quad \gamma_d = \frac{n_d \sigma_u^2}{n_d \sigma_u^2 + \sigma_e^2}. \quad (12)$$

Al desconocerse los parámetros del modelo NER, se obtienen sus estimadores aplicando el método de la máxima verosimilitud (Maximun Likelihood - ML) o el método de la máxima verosimilitud residual (Restricted Maximun Likelihood - REML). Siguiendo a Searle y otros (1992), este trabajo utiliza estimadores REML para ajustar el modelo NER. Estos procedimientos están implementados en la función *lmer* de la librería *lme4* de R¹. Definimos el estimador $\hat{\mathbf{V}}_{ds} = \hat{\sigma}_u^2 \mathbf{1}_{n_d} \mathbf{1}_{n_d}' + \hat{\sigma}_e^2 \mathbf{I}_{n_d}$ y sustituimos los estimadores $\hat{\sigma}_e^2$ y $\hat{\sigma}_u^2$ en BLUP, para obtener el EBLUE (empirical BLUE) de $\boldsymbol{\beta}$ y el EBLUP (empirical BLUP) de u_d ; es decir,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{d=1}^D \sum_{s=1}^{n_d} \mathbf{X}_{ds}' \hat{\mathbf{V}}_{ds}^{-1} \mathbf{X}_{ds} \right)^{-1} \sum_{d=1}^D \sum_{s=1}^{n_d} \mathbf{X}_{ds}' \hat{\mathbf{V}}_{ds}^{-1} \mathbf{y}_{ds}, \quad \hat{u}_d = \hat{\sigma}_u^2 \mathbf{1}_{n_d}' \hat{\mathbf{V}}_{ds}^{-1} (\mathbf{y}_{ds} - \mathbf{X}_{ds} \hat{\boldsymbol{\beta}}). \quad (13)$$

Se reemplazan los estimadores $\hat{\sigma}_e^2$ y $\hat{\sigma}_u^2$ en la ecuación (12), para obtener:

$$\hat{\gamma}_d = \frac{n_d \hat{\sigma}_u^2}{n_d \hat{\sigma}_u^2 + \hat{\sigma}_e^2}. \quad (14)$$

¹ R es un lenguaje y ambiente para computación estadística y generación de gráficos estadísticos (<https://www.r-project.org/about.html>).

2.1.2. Distribución condicional de la no-muestra dada la muestra

Los sub-vectores no muestrales \mathbf{y}_{dr} son desconocidos por lo que deben ser predichos. Para este fin se utilizará la distribución condicional:

$$\mathbf{y}_{dr}|\mathbf{y}_{ds} \sim N(\boldsymbol{\mu}_{dr|s}, \mathbf{V}_{dr|s}), \quad (15)$$

donde

$$\begin{aligned} \boldsymbol{\mu}_{dr|s} &= \mathbf{X}_{dr}\boldsymbol{\beta} + \sigma_u^2 \mathbf{1}_{N_d-n_d} \mathbf{1}'_{n_d} \mathbf{V}_{ds}^{-1} (\mathbf{y}_{ds} - \mathbf{X}_{ds}\boldsymbol{\beta}), \\ \mathbf{V}_{dr|s} &= \sigma_u^2 (1 - \gamma_d) \mathbf{1}_{N_d-n_d} \mathbf{1}'_{n_d} + \sigma_e^2 \mathbf{I}_{N_d-n_d}. \end{aligned} \quad (16)$$

Como se desconocen los parámetros necesarios para calcular $\boldsymbol{\mu}_{dr|s}$, $\mathbf{V}_{dr|s}$, éstos se sustituyen por estimadores. Los estimadores del vector de medias y de la matriz de covarianzas condicionales son

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{dr|s} &= \mathbf{X}_{dr}\hat{\boldsymbol{\beta}} + \hat{\sigma}_u^2 \mathbf{1}_{N_d-n_d} \mathbf{1}'_{n_d} \hat{\mathbf{V}}_{ds}^{-1} (\mathbf{y}_{ds} - \mathbf{X}_{ds}\hat{\boldsymbol{\beta}}) = \mathbf{X}_{dr}\hat{\boldsymbol{\beta}} + \mathbf{1}_{N_d-n_d} \hat{u}_d, \\ \hat{\mathbf{v}}_{dr|s} &= \hat{\sigma}_u^2 (1 - \hat{\gamma}_d) \mathbf{1}_{N_d-n_d} \mathbf{1}'_{n_d} + \hat{\sigma}_e^2 \mathbf{I}_{N_d-n_d} \end{aligned} \quad (17)$$

Para toda unidad $j \in U_d - s_d$, se tiene que

$$\boldsymbol{\mu}_{dj|s} = \begin{cases} \mathbf{X}_{dj}\boldsymbol{\beta} + \gamma_d (\hat{\mathbf{Y}}_d - \hat{\mathbf{X}}_d\boldsymbol{\beta}) & \text{si } n_d \neq 0, \\ \mathbf{X}_{dj}\boldsymbol{\beta} & \text{si } n_d = 0, \end{cases} \quad \gamma_d = \frac{n_d \sigma_u^2}{n_d \sigma_u^2 + \sigma_e^2}. \quad (18)$$

donde $\hat{\mathbf{Y}}_d = n_d^{-1} \sum_{j=1}^{n_d} y_{dj}$ y $\hat{\mathbf{X}}_d = n_d^{-1} \sum_{j=1}^{n_d} x_{dj}$. Para toda unidad $j \in U_d - s_d$, la varianza condicional es

$$\mathbf{v}_{dj|s} = \begin{cases} \sigma_u^2 (1 - \gamma_d) + \sigma_e^2 & \text{si } n_d \neq 0, \\ \sigma_u^2 + \sigma_e^2 & \text{si } n_d = 0. \end{cases} \quad (19)$$

2.1.3. Predictores óptimos empíricos

Asumiendo que la variable objetivo y_{dj} sigue el modelo NER, este apartado presenta el EBP de la DCI del dominio d , F_d , definida en (20). Siguiendo a Molina y Rao (2010), el objetivo es predecir $F_d = F_d(\mathbf{y})$ usando la información muestral de \mathbf{y}_s . Para este fin, se busca encontrar un predictor \hat{F}_d insesgado (o aproximadamente insesgado) que minimice el error cuadrático medio (MSE - Mean Square Error)

$$MSE(\hat{F}_d) = E[(\hat{F}_d - F_d)^2], \quad (20)$$

donde E representa la esperanza matemática con respecto a la distribución conjunta del vector poblacional \mathbf{y} . El mejor predictor (BP - Best Predictor) es aquel predictor insesgado que minimiza el $MSE(\hat{F}_d)$, definido en la ecuación (20). El BP de F_d es

$$\hat{F}_d^B = E[F_d|\mathbf{y}_s]. \quad (21)$$

Nótese que este BP es insesgado y puede descomponerse para el vector muestral \mathbf{y}_s y no muestral \mathbf{y}_r . Tras aplicar la esperanza condicional presentada en (20), se tiene:

$$\hat{F}_d^B = \frac{1}{N_d} \left(\sum_{j \in s_d} F_{dj} + \sum_{j \in r_d} \hat{F}_{dj}^B \right), \quad (22)$$

donde \hat{F}_{dj}^B es el BP de F_{dj} , definido por:

$$\hat{F}_{dj}^B = E[F_{dj} | \mathbf{y}_s]. \quad (23)$$

La distribución condicional de y_r , dado y_s , depende del vector $\boldsymbol{\theta} = (\beta', \sigma_u^2, \sigma_e^2)'$ de parámetros desconocidos del modelo, que deben ser estimados; es decir,

$$E[F_{dj} | \mathbf{y}_s] = E[F_{dj} | \mathbf{y}_s; \boldsymbol{\theta}]. \quad (24)$$

Sea $\hat{\boldsymbol{\theta}} = (\hat{\beta}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'$ un estimador basado en los datos muestrales y_s . El EBP de F_{dj} es

$$\hat{F}_{dj}^{EB} = E[F_{dj} | \mathbf{y}_s; \hat{\boldsymbol{\theta}}]. \quad (25)$$

donde la esperanza condicional viene a ser igual a

$$\hat{F}_{dj}^{EB} = P(N(0,1) < \hat{\alpha}_{dj}) = \Phi(\hat{\alpha}_{dj}), \quad \hat{\alpha}_{dj} = \hat{\mathbf{v}}_{dj|s}^{-1/2} (y_0 - \hat{\mu}_{dj|s}), \quad (26)$$

Donde $\Phi(\cdot)$ es la función de distribución de una variable aleatoria normal estándar y $\hat{\mu}_{dj|s}$ y $\hat{\mathbf{v}}_{dj|s}$ se obtienen de la ecuación (18) y (19). El EBP del total de niños con DCI en el dominio d , $T_d = N_d F_d$, es $\hat{T}_d^{EB} = N_d \hat{F}_d^{EB}$.

Para el cálculo del EBP de F_d es necesario disponer de un archivo muestral S y un archivo censal C con las siguientes propiedades:

1. C contiene los valores \mathbf{x}_{dj} , $d = 1, \dots, D$, $j = 1, \dots, N_d$, de las variables auxiliares.
2. Los valores \mathbf{x}_{dj} , $d = 1, \dots, D$, $j \in s_d$, son los mismos en los archivos C y S .
3. C y S usan los mismos identificadores de unidad, de modo que cada dominio U_d , del archivo C , puede particionarse en su parte muestral s_d y en su parte no muestral r_d .

En algunas aplicaciones se tiene un archivo muestral S y un archivo censal C que verifican la propiedad 1 pero no cumplen las propiedades 2 y 3. Por tanto, no se puede calcular el EBP. Una alternativa es usar predictores proyectivos empíricos (EPP - empirical projective predictor). Así, para la proporción de desnutrición crónica infantil, los EPP de F_d y T_d son:

$$\hat{F}_d^{EP} = \frac{1}{N_d} \sum_{j \in U_d} \hat{F}_{dj}^{EP}, \quad \hat{T}_d^{EP} = N_d \hat{F}_d^{EP}. \quad (27)$$

2.2. Selección de variables

Partiendo del concepto del modelo real para una variable y en un proceso de regresión Gaussiana:

$$\mathbf{y} = \mathbf{X}^{**} \boldsymbol{\beta} + \mathbf{e} \quad (28)$$

Donde \mathbf{X}^{**} es una matriz con p variables. Para aproximarse al problema en el mundo real se tendrá la matriz \mathbf{Z} con k variables. Todas las variables de los dos vectores se encuentran correlacionadas:

$$\mathbf{y} = \mathbf{X}^{**}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e} \quad (29)$$

Un algoritmo oráculo permitirá separar \mathbf{X}^{**} de \mathbf{Z} con un alto nivel de éxito (Leng, Lin & Wahba, 2006), ya que, en el mundo real desconocemos las variables pertenecen a cada grupo.

Según Zou & Zhang (2009) Adaptative Elastic-Net es un algoritmo oráculo superior a LASSO y Adaptative LASSO. Lo cual significa que, Adaptative Elastic-Net será el mejor algoritmo para selección de variables para procesos de regresión Gaussiana. Regresión que, pese a no incluir efectos aleatorios, consideramos es la mejor opción para nuestra aplicación, ya que no hemos podido encontrar hasta el momento aplicaciones de Adaptative Elastic-NET para Regresión Lineal Mixta.

Finalmente, se debe añadir una consideración adicional, tanto LASSO como Ada-ENET presentan resultados sensibles frente a la selección de los grupos para Validación Cruzada. Por esta razón se incorpora la siguiente propuesta.

3. Consideraciones para la aplicación empírica

3.1. Conformidad con las estimaciones directas de nivel superior

Si los parámetros de dominio (por ejemplo, cantones) son de tipo media y están calculados a partir de un modelo NER, no se puede garantizar que el promedio de los estimadores de dominios coincida con la estimación directa de un nivel de agregación superior (por ejemplo, provincia). Sea F_p^{dir} un estimador directo de la proporción de población con DCI en el nivel superior de interés p de tamaño N_p . Si en ese nivel p hay D_p dominios, se tiene que $N_p = \sum_{d=1}^{D_p} N_d$.

Sean $\hat{F}_1^{EB}, \dots, \hat{F}_{D_p}^{EB}$ los EBP de F_1, \dots, F_{D_p} obtenidos a partir de un modelo NER. Sean $\hat{F}_1^{dir}, \dots, \hat{F}_{D_p}^{dir}$ los correspondientes estimadores directos. En general, no se cumplirá la propiedad de Conformidad. Es decir, ocurrirá qué:

$$\hat{F}_p^{dir} \neq \hat{F}_p^{EB} \quad \hat{F}_p^{EB} = \frac{1}{N_p} \sum_{d=1}^{D_p} N_d \hat{F}_d^{EB}, \quad \hat{F}_p^{dir} = \frac{1}{\sum_{d=1}^{D_p} \sum_{j=1}^{n_d} w_{dj}} \sum_{d=1}^{D_p} \sum_{j=1}^{n_d} w_{dj} F_{dj}, \quad (30)$$

donde los w_{dj} son los pesos finales del diseño muestral. Unos estimadores conformes son:

$$\hat{F}_d^{EBc} = \lambda_d \hat{F}_d^{EB}, \quad \text{donde} \quad \lambda_d = \hat{F}_p^{dir} / \hat{F}_p^{EB}. \quad (31)$$

Los estimadores conformes verifican la igualdad $\hat{F}_p^{dir} = \sum_{d=1}^D \hat{F}_d^{EBc}$. Para estimar la varianza de los estimadores conformes, se pueden usar la aproximación:

$$\widehat{\text{var}}(\hat{F}_d^{EBc}) \sim \lambda_d^2 \widehat{\text{var}}(\hat{F}_d^{EB}), \quad d = 1, \dots, D. \quad (32)$$

Sin embargo, es preferible realizar tales estimaciones por procedimientos de remuestreo.

Los estimadores conformes derivados de los EBP se calculan de forma similar a los obtenidos a partir de los EBP y se denotan \hat{F}_d^{epc} , $d = 1, \dots, D$.

3.2. Implicaciones de la Condición de Conformidad

Consideremos el requerimiento de SAE-NER: $\mathbf{y} = (\mathbf{y}'_s, \mathbf{y}'_r)'$, que significa que desde un vector poblacional \mathbf{y} , obtendremos una muestra \mathbf{y}_s , y podremos identificar la información no presente en la muestra \mathbf{y}_r . Lo cual es equivalente a afirmar que los dos vectores provienen de información equivalente.

Empíricamente para que este requerimiento se cumpla se necesita que la información muestral y poblacional se levanten en un mismo período de tiempo o que los determinantes \mathbf{X} de \mathbf{y} no se vieran afectados en el tiempo. Además, tenemos que el modelo NER:

$$\mathbf{y}_d = \mathbf{X}_d \boldsymbol{\beta} + \mathbf{1}_{N_d} u_d + \mathbf{e}_d, \quad d = 1, \dots, D, \quad (33)$$

no permite la inclusión de Factores de Expansión, lo cual podría afectar a nuestra estimación de $\boldsymbol{\beta}$ y $\hat{\mu}_{dr|s}$.

Dadas estas restricciones, que empíricamente son difíciles de cumplir se propone la Condición de Conformidad (CC), la cual ayudaría a mejorar los siguientes aspectos del proceso inferencial de SAE-NER:

- *Coherencia de estimaciones de provincia y cantón:* La estimación directa a nivel provincial en las encuestas se realizan considerando el diseño muestral y el uso de factores de expansión. La estimación para cantones se realiza a través del modelo NER, el cual no soporta actualmente la inclusión de factores de expansión. Esto podría afectar la coherencia.
- *Coherencia de estimaciones para los cantones:* Si tuviéramos estimaciones directas con suficiente precisión para compararlas con SAE-NER por cantón, CC reducirá la diferencia entre estas dos estimaciones.
- *Coherencia temporal:* ECV se levantó en el 2014 y ENSANUT entre el 2018 y el 2019. Por otro lado, CPV se dio en el 2010. Esta diferencia temporal podría afectar la coherencia de las estimaciones SAE-NER.

3.3. Estimación del error cuadrático medio

Esta sección supone que existe un archivo censal \mathcal{C} verificando la propiedad 1 pero no las propiedades 2 y 3. El vector de parámetros del modelo NER es:

$$\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_u^2, \sigma_e^2)' \quad (34)$$

Sean U_d , s_d y r_d los conjuntos asociados a la población, la muestra y la no muestra del dominio d . A continuación, se introduce un procedimiento de remuestreo Bootstrap para la estimación del error cuadrático medio (MSE - Mean Squared Error) del predictor EPPC.

González-Manteiga y otros (2018a) introdujeron un procedimiento bootstrap para estimar el error cuadrático medio del EBLUP que fue posteriormente adaptado al EBP por Molina y Rao (2010). Para calcular un estimador por bootstrap paramétrico de $MSE(\hat{P}_d^{eppc})$, donde \hat{F}_d^{eppc} es el estimador conforme derivado del EPP, el procedimiento introducido por los autores citados queda de la siguiente forma.

1. Ajustar el modelo a los datos muestrales y calcular $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}_u^2, \hat{\sigma}_e^2)'$.
2. Repetir B veces ($b = 1, \dots, B$):

- a. Población bootstrap: Generar $u_d^{*(b)} \sim N(0, \hat{\sigma}_u^2)$, $e_{dj}^{*(b)} \sim N(0, \hat{\sigma}_e^2)$ $d = 1, \dots, D$ $j = 1, \dots, N_d$. Generar la población bootstrap

$$y_{dj}^{*(b)} = x_{dj}\hat{\beta} + u_d^{*(b)} + e_{dj}^{*(b)}, \quad d = 1, \dots, D, j = 1, \dots, N_d,$$

- b. y calcular los DCI de la población bootstrap

$$F_d^{*(b)} = \frac{1}{N_d} \sum_{j \in U_d} I(y_{dj}^{*(b)} < z), \quad d = 1, \dots, D.$$

- c. Muestra bootstrap: la muestra bootstrap tiene las mismas unidades que la muestra de datos reales, i.e. $s_d^{*(b)} = s_d$, $b = 1, \dots, B$. Es decir, $s_d^{*(b)} = s_d$, $b = 1, \dots, B$. Como el modelo se define sobre la población, la fuente de aleatoriedad proviene de la generación de la población.
- d. Modelo bootstrap: Ajustar un modelo NER a la muestra bootstrap $(y_{dj}^{*(b)}, x_{dj})$, $d = 1, \dots, D$, $j \in s_d$. Calcular el estimador del vector de parámetros del modelo $\hat{\theta}^{*(b)} = (\hat{\beta}^{*(b)'}, \hat{\sigma}_u^{2*(b)}, \hat{\sigma}_e^{2*(b)})'$. Calcular los predictores de la proporción $P_d^{*(b)}$ y del total $T_d^{*(b)} = N_d P_d^{*(b)}$ de la población Bootstrap; es decir

$$\hat{P}_d^{cpp*(b)} = \frac{1}{N_d} \sum_{j \in U_d} \hat{p}_{dj}^{ebp*(b)}, \quad \hat{T}_d^{cpp*(b)} = N_d \hat{P}_d^{cpp*(b)}$$

- e. Donde

$$\begin{aligned} \hat{p}_{dj}^{ebp*(b)} &= P(N(0,1) < \hat{\alpha}_{dj}^{*(b)}) = \Phi(\hat{\alpha}_{dj}^{*(b)}), \quad \hat{\alpha}_{dj}^{*(b)} = \hat{v}_{dj|s}^{-\frac{1}{2}*(b)} (y_0 - \hat{\mu}_{dj|s}^{*(b)}), \\ \mu_{dj|s}^{*(b)} &= \begin{cases} x_{dj}\hat{\beta}^{*(b)} + \hat{\gamma}_d^{*(b)} (\bar{y}_{ds}^{*(b)} - \bar{x}_{ds}\hat{\beta}^{*(b)}) & \text{si } n_d \neq 0, \\ x_{dj}\hat{\beta}^{*(b)} & \text{si } n_d = 0, \end{cases} \quad \hat{\gamma}_d^{*(b)} = \frac{n_d \hat{\sigma}_u^{2*(b)}}{n_d \hat{\sigma}_u^{2*(b)} + \hat{\sigma}_e^{2*(b)}}, \\ \hat{v}_{dj|s}^{*(b)} &= \begin{cases} \hat{\sigma}_u^{2*(b)} (1 - \hat{\gamma}_d^{*(b)}) + \hat{\sigma}_e^{2*(b)} & \text{si } n_d \neq 0, \\ \hat{\sigma}_u^{2*(b)} + \hat{\sigma}_e^{2*(b)} & \text{si } n_d = 0, \end{cases} \quad \bar{x}_{ds} = \frac{1}{n_d} \sum_{j=1}^{n_d} x_{dj}, \quad \bar{y}_{ds}^{*(b)} = \frac{1}{n_d} \sum_{j=1}^{n_d} y_{dj}^{*(b)}. \end{aligned}$$

- f. Los predictores proyectivos conformes de $P_d^{*(b)}$ y $T_d^{*(b)}$ son

$$\hat{P}_d^{cepp*(b)} = \frac{\hat{T}_d^{cepp*(b)}}{N_d} \quad \text{y} \quad \hat{T}_d^{cepp*(b)} = \lambda_d^{*(b)} \hat{T}_d^{cpp*(b)},$$

- g. donde

$$\lambda_d^{*(b)} = \frac{\hat{T}_p^{dir*(b)}}{\hat{T}_p^{cpp*(b)}}, \quad \hat{T}_p^{dir*(b)} = \sum_{d=1}^{D_p} \sum_{j=1}^{n_d} w_{dj} y_{dj}^{*(b)}, \quad \hat{T}_p^{cpp*(b)} = \sum_{d=1}^{D_p} \hat{T}_d^{cpp*(b)}.$$

3. Salida final:

$$mse^*(\hat{P}_d^{cepp}) = \frac{1}{B} \sum_{b=1}^B (\hat{P}_d^{cepp*(b)} - P_d^{*(b)})^2.$$

4. Aplicación empírica

Para obtener las estimaciones deseadas basadas en nuestro marco teórico se elaboraron varias funciones en el lenguaje de programación R. Para las estimaciones de SAE-NER se proponen cuatro funciones: i) cálculo de los EPPs por dominio basadas en modelo NER (sin Conformidad); ii) bootstrap paramétrico para la estimación de los MSEs de los predictores EPP; iii) cálculo de los EPPs por dominio basadas en modelo NER con la Condición de Conformidad; iv) bootstrap paramétrico para cálculo de los MSEs de los predictores obtenidos en iii). Estas funciones tienen las siguientes ventajas:

- Uso de una cantidad k de regresores, limitada únicamente por las capacidades del computador y/o R. Los nombres de los regresores se especifican como un vector de caracteres, lo cual facilita considerablemente su manejo.
- Velocidad y eficiencia computacional: se utilizó el paquete de R, *data.table*, siguiendo prácticas de programación para reducir el uso de RAM y acelerar la velocidad de cómputo.
- Se proponen funciones SAE-NER con Condición de Conformidad.

Para la selección de variables se utilizó Adaptive Elastic-Net usando el paquete *glmnet* de R. Este paquete únicamente permite calcular n validaciones cruzadas al partir la base de datos en n pedazos. Para aumentar su estabilidad se realiza este proceso k veces de tal forma que en cada iteración se divida la base de datos en n pedazos diferentes, aumentando así la estabilidad de los resultados. El Anexo No. 3 proporciona una descripción detallada de estas funciones.

El Anexo No. 3 contiene los códigos que evidencian la creación de los parámetros y funciones para la estimación del SAE-NER.

4.1. Propuesta

Para la presente investigación se utilizaron tres fuentes de información: 1) Encuesta de Condiciones de Vida 2014 (ECV-2014), 2) Encuesta Nacional de Salud y Nutrición 2018 (ENSANUT-2018), 3) Censo de Población y Vivienda 2010 (CPV-2010). Considerando que utilizamos dos encuestas diferentes (ECV-2014 y ENSANUT-2018) para obtener y_{ds} y sus variables predictoras X_{ds} , tendremos dos resultados de SAE-NER: el primero para el 2014 y el segundo para el 2018. Las dos encuestas, individualmente, sirven para realizar las estimaciones mencionadas en 2.1.1, las cuales al mezclarse con el CPV-2010 presentan los predictores EPPC a nivel del cantón de interés. A continuación, se detallan algunos aspectos importantes de dichas fuentes.

La *Encuesta de Condiciones de Vida*² (ECV) es una encuesta multipropósito que permite medir diferentes aspectos y dimensiones del bienestar y calidad de vida de los hogares: consumo y pobreza de consumo de los hogares, acceso de servicios públicos, salud, vivienda y educación, entre otras características sociodemográficas. La ECV, sexta ronda, se levantó mediante visita de personal del Instituto Nacional de Estadística y Censos del Ecuador (INEC) a los hogares

² Como se puede observar en el Documento Metodológico de ECV-2014 presentado en la página web oficial de INEC: <https://bit.ly/3BMQExF>

seleccionados. Se recogió información completa de los miembros de 28.970 hogares entre noviembre 2013 y octubre 2014. Su tamaño muestral fue calculado con el fin de obtener errores relativos del 10% al 95% de confianza en: 1) tasa de desempleo, 2) proporción de la población en estado de pobreza, 3) proporción de la población en estado de extrema pobreza. Si bien la ECV no considera a la DCI como una variable de diseño, vemos que sus estimaciones por provincia presentan un Coeficiente de Variación (CV) menor al 30%, por lo que se consideran adecuados para su uso como base para la Condición de Conformidad. Finalmente, con base a la limpieza de la base de datos realizada nos quedamos con 11.289 observaciones válidas para el cálculo de la DCI a nivel de provincia (menores de 5 años).

La *Encuesta de Salud y Nutrición*³ (ENSANUT) es una encuesta especializada en salud y nutrición. Además, capta características sociodemográficas de los miembros del hogar, características de la vivienda y acceso a servicios públicos. Entre otras, la ENSANUT tiene las siguientes variables de interés: 1) Desnutrición Crónica en menores de 5 años (DCI); 2) sobrepeso y obesidad en niños entre 5 y 11 años; 3) proporción de mujeres en edad fértil (MEF) de 12 a 49 años. A nivel provincial se espera un error relativo del 15% al 95% de confianza, por lo que esta muestra será más grande que la ECV-2014. La muestra total de menores con información válida para el cálculo de la DCI será de 20.285 observaciones.

El *Censo de Población y Vivienda* (CPV) 2010⁴ es un recuento de la población y de las viviendas para generar información sobre la estructura, crecimiento y distribución de la población y de sus características económicas, sociales y demográficas. También permite conocer medidas demográficas y características de las viviendas en las que habita la población ecuatoriana, con el fin de evaluar las condiciones de vida y servicios inherentes. El CPV 2010, como base de datos presenta 14.483.499 observaciones, de estos se considera 1.308.910 que son aquellos menores de cinco años de edad donde se enfoca el objeto de este estudio.

Previo a la estimación de SAE-NER se identifica las variables comparables entre el censo y las encuestas. Este trabajo utiliza la preparación de datos entre CPV y ECV realizada por Molina y otros (2015) para estimar la pobreza y desigualdad por consumo mediante la metodología SAE. En este estudio, los autores realizan el proceso de armonización de datos⁵ y la comparación de las variables, basado en criterios de similitud en la pregunta, respuesta y universo de encuestados. Finalmente, generan variables referidas a características demográficas, material de vivienda, servicios, educación y ocupación a nivel de todos los miembros del hogar, para el jefe y cónyuge.

De manera similar, se procede con la preparación de los datos entre la ENSANUT y CPV. Es importante mencionar que la armonización de datos para la ENSANUT se realiza con referencia a la DPA 2014, debido a que no se establece de manera clara la correspondencia del CPV con la cartografía 2018.

³ Como se puede observar en el Documento Metodológico de ENSANUT-2018 presentado en la página web oficial de INEC en: <https://bit.ly/3k7eGx0>

⁴ Como se puede observar en la página web oficial de INEC, sección '¿Qué es el Censo de Población y Vivienda?' en: <https://bit.ly/3ENDe6a>

⁵ Dado que el CPV y ECV se encuentran codificadas en base al Clasificador Geográfico Estadístico- DPA 2010, se debe reflejar la DPA en base al 2014.

Así, del CPV-2010 se obtendrán las variables \mathbf{X}_{dr} , condicional a que cada variable corresponda a la misma medición de muestral en \mathbf{X}_{ds} .

Por esta razón, debemos recalcar que, dada esta restricción, potencialmente se excluirán determinantes importantes de la DCI, como información de salud del niño y de la madre. Pero, dadas las restricciones de información, en el proceso de selección de variables y modelos se buscará un modelo que minimice el error cuadrático medio de los estimadores.

Con base en estas condiciones se obtienen alrededor de 400 variables tanto para ECV-2014 como para ENSANUT-2018. Esta cantidad de variables se debe a que cada variable categórica con k categorías es transformada a $(k-1)$ variables binarias. En general, las variables que aparecen tanto en las dos encuestas como en el censo pueden ser clasificadas en los siguientes tipos:

- Demográfico: edad, sexo, relación con jefe de hogar, idioma que habla, provincia y área de residencia.
- Social: discapacidad, autoidentificación étnica, educación.
- Laboral: se encuentra trabajando (sí o no), rama de actividad, ocupación.
- Vivienda: información de techo, paredes y piso; acceso a agua y servicios básicos; pobreza por Necesidades Básicas Insatisfechas-NBI.

En el ajuste de los modelos NER a los datos, también se consideran interacciones de la información demográfica, social y laboral para el jefe de hogar y para el cónyuge. Si el hogar no reporta cónyuge estas variables toman el valor 0.

4.2. Construcción del Modelo

4.2.1. Proceso de selección de variables

Dada la cantidad de potenciales variables independientes para NER (~400) nos encontramos frente a un *problema de dimensionalidad alta*, o sea, se tienen tantas variables que es difícil lograr generar un subconjunto de variables que minimicen alguna medida de eficiencia; y, de incluirse todas las variables, se obtendría un modelo sobre-parametrizado poco útil (Hastie, Tibshirani & Friedman, 2009). Frente a este problema se utilizan algoritmos de selección de variables y, dadas las reflexiones en (2.2), se aplica el Ada-ENET.

Se debe considerar que entre Econometría Tradicional y Machine Learning -ML existe una diferencia al momento de minimizar el MSE. En la primera generalmente se busca la incorporación de un modelo que permita comprender relaciones causales de una variable x_j con relación a la variable dependiente y , lo cual representaría que su objetivo principal sería la reducción del sesgo en la fórmula de MSE, o al menos la relacionada a la variable x_j (inferencia en la jerga de ML). Por otro lado, en ML se considera que su principal fin es la predicción: no importa el sesgo o varianza de la variable x_j siempre y cuando el sesgo y varianza del modelo en conjunto sea el mínimo posible (considérese que es común el caso en el cual el sesgo que introduce una variable se compensa con el sesgo que introduce otra, al tener signos diferentes). Ada-ENET tiende a reducir las dos componentes del MSE, pero su enfoque es principalmente la predicción (James, Witten, Hastie, & Tibshirani, Capítulo 2, 2013) (Friedman, Hastie, & Tibshirani, Sección 3.3, 2009).

Si bien existen herramientas ML específicas para selección de variables para *inferencia*, estas se encuentran fuera del alcance de este documento. Ada-ENET tiene un *hyperparámetro* α , mientras más alto es este valor Ada-ENET será menos

restrictivo con la selección de variables y viceversa (Friedman, Hastie & Tibshirani, Sección 2, 2010). Dado que desconocemos si nuestra información se adapta mejor estos valores extremos de alfa o sus intermedios planteamos la utilización de $\alpha = \{0,1; 0,3; 0,5; 0,7; 0,9\}$.

En ideal, el modelo con el menor MSE de validación cruzada será el mejor, a menos que presente residuos muy distantes de la distribución normal⁶ o rompa otros supuestos fundamentales para SAE-NER.

Pero, sabemos desde la sección anterior que no es posible realizar VC y además tomar en cuenta los efectos aleatorios u_d . Por ello, se busca alguna alternativa que presente la mejor aproximación posible por lo que se propone el uso de Ada-ENET.

Adicionalmente, al romperse el supuesto de $\mathbf{y} = (\mathbf{y}_s, \mathbf{y}_r)$, probablemente se pierda el evidente supuesto de $\mathbf{X} = (\mathbf{X}_s, \mathbf{X}_r)$, para mitigar este problema, se propone crear subconjuntos grandes de variables auxiliares:

- *Todas*: Todas las variables disponibles;
- *Todas-sd=1*: Todas las variables disponibles con variables continuas estandarizadas;
- *Simil 5%*: Variables cuya estimación puntual tenga una diferencia absoluta máxima del 5% entre CPV y encuesta
- *Simil 10%*: Diferencia absoluta máxima del 10%
- *Simil 15%*: Diferencia absoluta máxima de 15%.

Ya que tendremos 5 subconjuntos de potenciales variables independientes, en cada uno de ellos se estimará Ada-NET con cada alfa anteriormente definido. Por lo que tendremos 25 modelos para la selección del mejor modelo.

4.2.2. Proceso de selección de modelo

La selección de un modelo predictivo es un proceso mediante el cual, frente a varios posibles candidatos, se minimiza una función de pérdida o se optimiza algún otro criterio: error cuadrático medio-MSE, error absoluto medio, AIC, BIC, entre otros. El modelo que optimice dicha función será seleccionado (Zucchini, 2000). Este proceso ha sido históricamente diferente entre econometría y estadística. Los estadísticos generalmente buscaban realizar algún tipo de Validación Cruzada (VC)⁷ para evaluar la calidad de sus predicciones. Por otro lado, en econometría se argumentaba que el tamaño de muestra que usualmente usaban era demasiado pequeño para realizar esta validación.

Por otro lado, la VC es generalmente usada en ML pues se dispone de información considerablemente grande por lo que partir las bases de datos no representa un problema desde el punto de vista del tamaño de los pedazos resultantes. Para realizar VC en modelos NER o lineales mixtos se requiere que cada dominio presente un tamaño de muestra grande tal que sea posible partir el dominio d en bases de entrenamiento y comprobación: *training* y *test* (Braun, Held & Sabanés, 2014). Este

⁶ Se plantea como restricción Normalidad en Residuos porque es un supuesto de SAE-NER.

⁷ VC se refiere al proceso de cálculo de funciones de pérdida en información futura (información *test*) con relación a la información sobre la cual se realizan las estimaciones de modelo (información *train*). Cuando la información futura no se encuentra disponible, se puede partir la información disponible para realizar un proceso parecido al anteriormente mencionado. Para más información refiérase a la sección 7 de Friedman, Hastie, Tibshirani &, Friedman (2009).

no es el caso ni de la ECV-2014 ni de la ENSANUT-2018 que presentan 155 y 112 dominios con menos de 50 observaciones.

Además, Vaida & Blanchard (2005) sugieren el uso de AIC sin validación cruzada para la selección de modelo. Pero, su propuesta considera otros supuestos restrictivos: se requiere que $\mathbf{y} = (\mathbf{y}_s, \mathbf{y}_r)$, lo que requiere que tanto la muestra como la no-muestra extraída de CPV-2010 provengan del mismo vector poblacional. Por ejemplo, se puede simular una distribución poblacional de tamaño N , de la cual se pueden sacar dos subconjuntos: $\mathbf{y}_{s,t2}$ y \mathbf{y}_r . Y, asumiendo que $\mathbf{y}_{s,t2} = \mathbf{y}_s * \mathbf{s}$, con $\mathbf{y}_s * \mathbf{s} = (y_{s1}s_1, y_{s2}s_2, y_{s3}s_3, \dots, y_{sn}s_n)$, donde el vector \mathbf{s} representa los cambios que ha tenido \mathbf{y}_s en el tiempo. Si la distribución de \mathbf{s} fuese aleatoria, se podría utilizar la metodología de Vaida & Blanchard (2005). Por otra parte, si \mathbf{s} es no aleatorio, el supuesto de $E[(\mathbf{y}_s, \mathbf{y}_r)|\mathbf{X}_s] \neq E[(\mathbf{y}_{s,t2}, \mathbf{y}_r)|\mathbf{X}_s]$ se rompería y por lo que se generarían estimaciones SAE con desviaciones con relación al MSE ideal, y por tanto AIC, con efectos sobre sesgo y varianza desconocidos.

Dado que el vector CPV-2010 (\mathbf{y}_r), y el vector muestral (\mathbf{y}_s) presentan diferencias temporales de 4 y 8 años, es intuitivo pensar que el vector \mathbf{s} no es aleatorio, por lo que la metodología de Vaida & Blanchard (2005) tendrá efectividad limitada.

Frente a estas restricciones, con el fin de mitigar en la medida de lo posible estas limitaciones, planteamos usar AIC dentro de la muestra; y, minimizar varianza con Condición de Conformidad en estimaciones fuera de muestra ($\hat{\mathbf{y}}_r$). Adicionalmente, tras encontrar el mejor modelo de los 25 modelos planteados en la sección anterior, se proponen cuatro modificaciones adicionales al mismo. Por facilidad de notación, el mejor modelo original se denomina Modelo v1 (versión 1). Los modelos modificados se obtienen después de realizar las siguientes acciones:

- Modelo v2 (versión 2). El modelo v2 se obtiene eliminando las variables auxiliares con coeficientes betas no significativos, siempre y cuando su extracción no afecte a los signos del resto de coeficientes. Hay que comprobar que el p-valor de todas las variables explicativas de v2 sea menor igual al 10%.
- Modelo v3 (versión 3). El modelo v3 se obtiene eliminando las variables auxiliares con coeficientes betas no significativos siempre y cuando su extracción no afecte a los signos del resto de coeficientes. Hay que comprobar que el p-valor de todas las variables explicativas de v2 sea menor igual al 5%.
- Modelo v4 (versión 4). El modelo v4 se obtiene eliminando las variables auxiliares con coeficientes contra-intuitivos. Un coeficiente es contra-intuitivo si su signo no se corresponde con el que debería tener según el estado actual del conocimiento.
- Modelo v5 (versión 5). Se realiza v2 y v4 simultáneamente.
- Modelo v6 (versión 6). Se realiza v3 y v4 simultáneamente.

De estos 30 modelos se seleccionará finalmente el modelo para los cálculos SAE-NER.

Estas propuestas se realizan con objeto de incidir en los dos sumandos que intervienen en la fórmula del MSE ; es decir, $MSE = sesgo^2 + varianza$. Las modificaciones v2 y v3 tienen como fin reducir la parte de $varianza$ de la anterior ecuación, v4 la parte de $sesgo$ y v5-v6 las dos simultáneamente. Si nos encontráramos en un escenario en el que la metodología propuesta por Vaida & Blanchard (2005) fuese suficiente, no sería necesario realizar estos modelos adicionales.

Empíricamente podemos saber que nos encontramos en dicho escenario si la variabilidad del AIC (o MSE de validación cruzada) por modelo tuvieran una alta correlación con la distribución del error estándar de los EPPC. Caso contrario, sabremos que nos encontramos en un escenario donde $E[y_{s,t2}|s]$ es no aleatorio, por lo cual requeriremos criterios adicionales. Para este escenario, proponemos utilizar como tercer criterio la distribución del error estándar de los EPPC en censo ya que: 1) es una predicción fuera de muestra por lo que evitará sobreparametrización, 2) contiene la Condición de Conformidad, por lo cual estos errores estándares aumentarán con el sesgo entre estimaciones EPP frente a estimaciones directas.

5. Resultados ECV-2014

5.1. Selección de modelo

Se presentan los AIC relativos de cada modelo al mínimo AIC de nuestros 30 modelos, mediante $AIC_{rel,i} = \left(\frac{AIC_i}{\min(\{AIC_1, \dots, AIC_{29}\})} - 1 \right) 100\%$. El AIC relativo del modelo i se interpreta del siguiente modo: el modelo i tiene un AIC que es $x\%$ más alto que el AIC mínimo. Se sigue el mismo proceso para el RMSE⁸, calculado a partir de valores observados y predichos en un proceso de validación cruzada, y se tendrá la misma interpretación. Este RMSE se calcula mediante Validación Cruzada con $n = 10$ y 100 iteraciones⁹, durante el proceso de selección de Ada-ENET.

Se puede observar en el Gráfico No. 1 que tanto para el AIC como para el RMSE los modelos presentan una variabilidad muy pequeña (al excluir el grupo de modelos que parten de todas las variables). La distancia promedio respecto al AIC mínimo es de 0,388% y 20 de los 29 modelos presentan un AIC relativo menor a 0.5%.

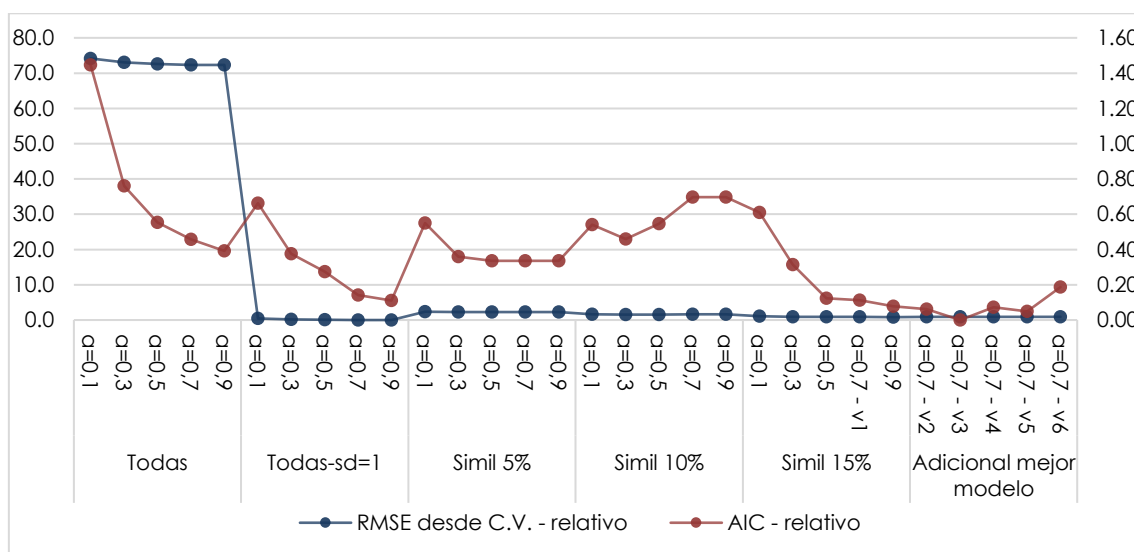
Por otra parte, el RMSE promedio es 13.11% al incluir todos los modelos, al excluir los modelos que utilizan todas las variables sin estandarización este valor llega a 1.16%, ya que este grupo es un atípico con relación a RMSE pero no a AIC.

Al observar los resultados de RMSE (excluyendo el grupo de atípicos), resulta interesante notar que solo ciertos modelos se alejan del resto. Por ejemplo, al comparar los modelos obtenidos por: 1) Ada-ENET desde el total de variables potenciales cuyas continuas fueron estandarizadas ($sd = 1$); 2) Ada-ENET desde el total de variables potenciales con una diferencia máxima del 15% (estimación puntual encuesta vs. censo); únicamente 1 modelo tiene una diferencia relativa mayor al 1% (1,08% de hecho). Se puede ver un fenómeno similar en el resto de los grupos de modelos con magnitudes un poco más grandes (entre 1% al 2%).

⁸ RMSE: Raíz cuadrada de MSE. El MSE se calcula como es usual: $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$. Durante Validación cruzada de k -foldings, se estima el modelo en la información partida $k-1$, se predice en k y se calcula MSE, este proceso se realiza para todos los k y se promedia. Cuanto se itera sobre los k -foldings el proceso se repite las n veces requeridas y se promedia sus MSE, sobre ese se resultado se obtiene el RMSE a usarse.

⁹ Siguiendo la recomendación de Wong & Yeh (2020), se realizaron simulaciones con diferentes valores de iteraciones, encontrando que en general se tiene estabilidad de resultados para hyper-parámetro λ con el valor seleccionado.

Gráfico No. 1: AIC y RMSE de Validación Cruzada de Modelos ECV-2014



Fuente: ECV-2014

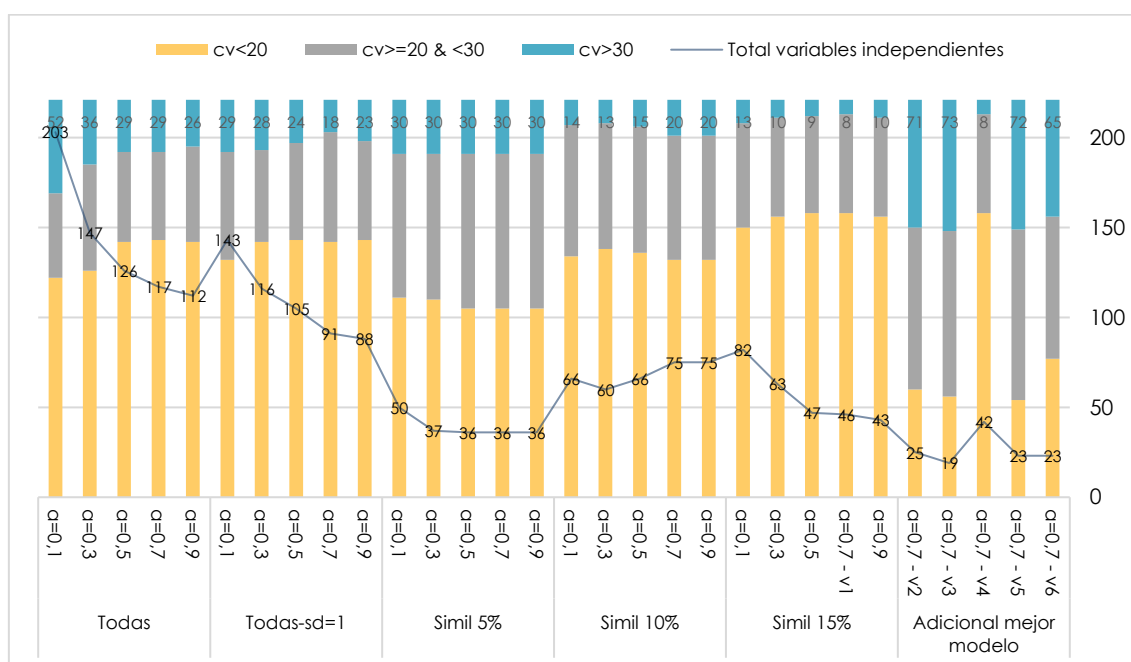
Elaboración: Autores

Usualmente en esta etapa se selecciona el modelo sin observar resultados de predicción, pero considerando que, entre todos los modelos, 29 de 30 tienen una diferencia relativa menor a 0,8% en AIC (el modelo 1 tiene una diferencia 1,5%) diferencia; que 20 de los 30 modelos presentan una diferencia relativa menor al 1,6% en RMSE, es necesario recurrir a criterios complementarios. Para este caso parece coherente utilizar una varianza estimada fuera de muestra con corrección de sesgo mediante la Condición de Conformidad dados los criterios presentados en la subsección de *Construcción modelo*. Ya que no nos interesa toda la distribución muestral de EPPC, sino, solo los coeficientes de variación mayores a 30%, minimizaremos ese criterio.

El Gráfico No. 2 muestra los 30 modelos en orden ascendente por el total de cantones con un CV>30%. Nuestro mejor modelo presenta ocho (8) cantones con un CV alto, o sea, solo ocho (8) cantones tienen una presión demasiado baja como para considerarse como un dato confiable, por lo tanto el modelo seleccionado para SAE-NER en ECV será $\alpha = 0,7 - v4^{10}$.

¹⁰ Nótese que en los gráficos se tiene $\alpha=0,1$, $\alpha=0,3$..., lo cual de hecho representa $\alpha = 0,1$, $\alpha = 0,3$,...

Gráfico No. 2: Total cantones clasificados por Coeficiente de Variación, por modelo. ECV-2014



Fuente: ECV-2014

Elaboración: Autores

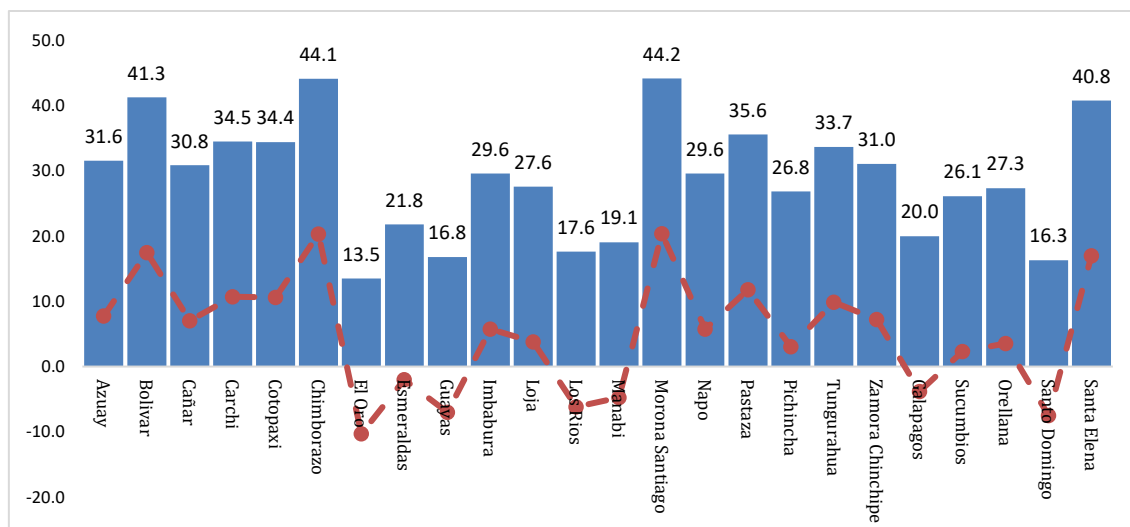
Podemos ver que intra-grupos los diferentes modelos Ada-ENET presentan resultados muy parecidos respecto de la varianza del EPPCs, lo cual significa que la matriz de correlaciones de $E[y_s|X_s]$ presenta en general correlaciones muy bajas. Esta conclusión se puede deducir ya que pese a cambiar el hiper-parámetro alfa de Ada-ENET (el cual al aumentar aumenta la probabilidad de excluir variables con alta correlación) el conjunto de variables seleccionadas se mantiene casi invariante.

Intuitivamente se esperaría que, al restringir el posible grupo de variables independientes a las variables más similares entre censo y encuesta (por ejemplo, 5% vs 15%), la cantidad de cantones con CVs bajos aumente. Esto se debe a que se podría reducir el sesgo descrito en la Sección 3.2 como $y_{s,t2}$. Pero, al reducir la cantidad de variables incluidas, posiblemente la varianza del estimador del MSE aumentaría. Sería interesante realizar una descomposición del MSE para comprobar formalmente las observaciones realizadas, pero, ese tipo de análisis se encuentra fuera del alcance de nuestra investigación.

5.2. Descriptivos relevantes

En el Gráfico No. 3 se observa la diferencia de la estimación puntual de cada provincia con relación al promedio nacional (23,89%). Podemos ver que la provincia con mayor DCI a nivel nacional es Chimborazo en la Sierra con un 44,1%, seguida por Morona Santiago 44% (Oriente), Santa Elena 40,6% (Costa) y Bolívar con 40,5%. Por otro lado, únicamente las provincias de la Costa presentan tasas de DCI menores al promedio nacional, con el Oro (16,6%) y Guayas (17,6%) con las menores tasas de DCI a nivel nacional. Este orden se mantiene igual en ENSANUT-2018. Resulta interesante observar que Santa Elena es la única provincia de la Costa con DCI mayor al promedio nacional (23.89%); y, a la vez, presenta la mayor tasa a nivel nacional.

Gráfico No. 3: Diferencia tasa DCI frente a promedio nacional por provincia. ECV-2014



Fuente: ECV-2014

Elaboración: Autores

5.3. Regresión

Extendiendo lo mencionado en la Subsección 3.2 sobre el proceso de selección de variables, los coeficientes de regresión estimados deben interpretarse con cuidado. Ada-ENET (y su familia) minimizan el MSE. Durante esa minimización, la reducción de la varianza tiene el mismo peso que la minimización del sesgo. Por lo cual, pueden existir coeficientes con diferentes niveles de sesgo. Resulta contra-intuitivo pensar que un modelo con coeficientes no interpretables (con un signo que no se corresponde a lo esperable según la teoría socio-económica) pueda realizar predicciones adecuadas de cualquier variable de interés. Sin embargo, en el análisis estadístico de datos es frecuente encontrar ocasiones en las que sea necesario llegar a un punto de equilibrio entre la interpretabilidad y la significatividad de los coeficientes de regresión estimados (James, Witten, Hastie, & Tibshirani, Capítulo 2, 2013).

Tómese en cuenta que en nuestras regresiones la variable dependiente es la talla para la edad esperada según OMS o Z-Score. Por ello si se quiere relacionar los coeficientes presentados con la DCI se debe cambiar el signo del coeficiente. Por ejemplo, en las Tablas No. 1, podemos ver que en las cuatro especificaciones propuestas¹¹ se tiene un signo negativo para la edad en meses, que se interpreta como una relación negativa entre el aumento de la edad (en el promedio de su distribución) y el Z-Score (manteniendo el resto de las variables constantes). Por otro lado, la misma variable se interpreta como una relación positiva, aumenta la probabilidad de DCI.

¹¹ Nótese que el orden (1), (2), ..., es equivalente a v_1 , v_2 , ...

Tabla No. 1: Resultados NER, versiones de modelo final. Parte I. ECV-2014.

	Talla para la edad (OMS)					
	[1]	[2]	[3]	[4]	[5]	[6]
Edad en meses	-0,058***	-0,058***	-0,059***	-0,058***	-0,058***	-0,059***
Edad en meses^2	0,882***	0,885***	0,893***	0,881***	0,885***	0,889***
sexo	-0,119***	-0,118***	-0,118***	-0,119***	-0,118***	-0,116***
(1==Azuay)	-0,147			-0,148		
(1==Los Ríos)		0,068			0,060	
(1==Morona Santiago)	-0,264**			-0,268***		
(1==Pastaza)	-0,184			-0,177		
(1==Pichincha)	-0,307***			-0,303**		
(1==Tungurahua)	-0,402***			-0,405***		
(1==Zamora Chinchipe)	-0,341***			-0,360***		
(1==Bolívar)	-0,299**			-0,305*		
(1==Orellana)	0,072					
(1==Santa Elena)	-0,617***			-0,624***		
(1==Carchi)	-0,283*			-0,282*		
(1==Cotopaxi)	-0,134			-0,133		
(1==Chimborazo)	-0,238*			-0,242*		
(1==El Oro)	0,106			0,103		
Región Sierra	-0,243***	-0,320***	-0,354***	-0,260***	-0,329***	-0,253***
Total en el hogar secundaria completa	0,131***	0,135***	0,134***	0,129***	0,134***	0,159***
Pago de vivienda: Por servicios	0,145*	0,149*	0,156*			0,111
Total carencias en hogar(NBI)^0,5	-0,123***	-0,127***	-0,142***	-0,117***	-0,127***	
Total niños (<+12 años)^0,5	-0,340***	-0,381***	-0,388***	-0,340***	-0,381***	-0,410***
Agua que toman: beben tal como llega al hogar	-0,051	-0,063		-0,015	-0,022	-0,099**
Agua que toman: La hierven	-0,048	-0,058			-0,079*	
Niños (<=12) en el hogar 1-3	0,065			0,060		
Servicio agua: Por tubería fuera de la vivienda	-0,074**	-0,087**	-0,095***	-0,082**	-0,089***	-0,120***
Servicio higiénico: Letrina	-0,181**	-0,188**	-0,197**	-0,184**	-0,193**	-0,249***
Servicio higiénico: Alcantarillado y pozo séptico	0,049			0,047		
Material piso: Pisos que no son de tierra	0,093*	0,095*		0,098*	0,098*	0,186***
Tiene electricidad por red pública	-0,086					
Infante asiste a guardería pública/ONG	-0,053*	-0,050		-0,058*	-0,052*	-0,042
(1==etnia indígena)	-0,170**	-0,186***	-0,197***	-0,167**	-0,187***	-0,226***
(1==etnia indígena, montubio, o mestizo)	-0,150**	-0,154***	-0,158***	-0,148**	-0,154***	-0,153***
(1==22+ años)	0,091			0,088		
(1== Primaria incompleta)	-0,077*	-0,070*	-0,082*	-0,077*	-0,069*	-0,087**
(1==Terciaria o superior o postgrado)	0,107**	0,116**	0,118**	0,109**	0,118**	0,160***
(1== habla lengua indígena)	-0,193***	-0,199***	-0,211***	-0,190***	-0,204***	-0,223***
(1==habla lengua extranjera)	0,117			0,124		
(1==nació en otro país)	0,164			0,166		
(1== tiene trabajo)	-0,154**	-0,156**	-0,149**	-0,149**	-0,153**	-0,165***
(1==trabaja en rama comercio)	0,059	0,059		0,058	0,059	
(1==no sabe leer y escribir)	-0,102*	-0,106*		-0,101*	-0,102*	-0,113*
(1== Edad entre 15-24 años)	-0,153***	-0,161***	-0,165***	-0,152***	-0,161***	-0,197***
(1==años de educación)	0,013***	0,013***	0,014***	0,013***	0,013***	0,016***
(1==educación postgrado)	0,165*	0,167*	0,155*	0,166*	0,167*	0,177*
(1==es estudiante)	0,233		0,236			
constante	0,677***	0,793***	0,878***	0,565***	0,755***	0,536***
N	11.289	11.289	11.289	11.289	11.289	11.289
AIC	36.287	36.269	36.246	36.272	36.263	36.314
BIC	36.646	36.474	36.407	36.602	36.454	36.505

*** p < 0,001; ** p < 0,01; * p < 0,05,

Nota: Modelos con alfa=0,7 y simil del 15%.

Fuente: ECV-2014

Elaboración: Autores

Cabe recalcar que los coeficientes del modelo NER se interpretan igual que los coeficientes de una regresión lineal simple que incluya las variables indicadoras de cada cantón. Resulta importante observar que entre los cuatro modelos se observan signos y valores de coeficientes betas muy similares, lo que se puede interpretar como una fuerte robustez de los modelos. La regresión (4), en la cual eliminamos las variables con coeficientes contra-intuitivos, solo tiene cuatro variables menos que el modelo original, reiterando la robustez de este. Adicionalmente, es importante analizar la información disponible sobre el diagnóstico de calidad del modelo. Tómese en cuenta que no existe una implementación de pruebas de heterocedasticidad o presencia de relación lineal entre y y X . Por lo cual nuestras pruebas de calidad del modelo se limitan al análisis

de la distribución de los residuos, que como se puede ver en el Anexo 1.2 son adecuados.

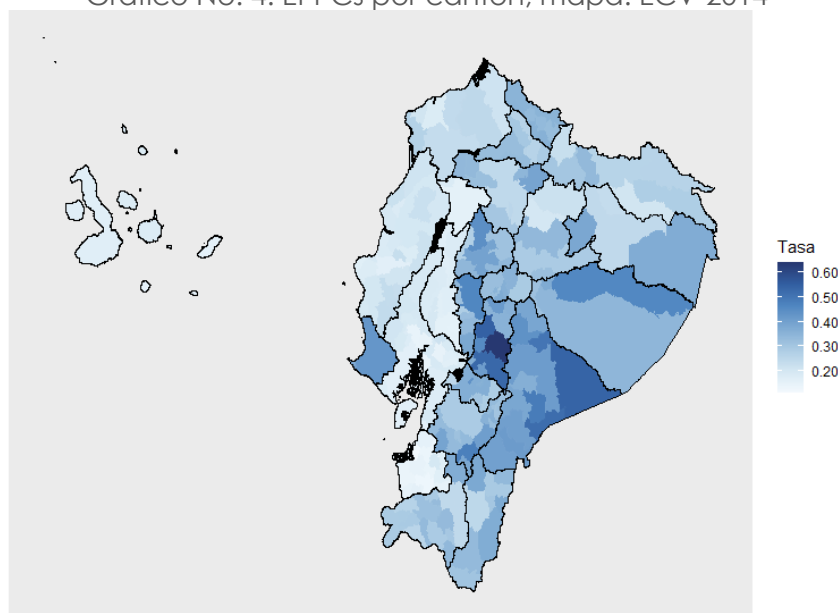
Se puede observar que el modelo seleccionado mediante Ada-ENET eligió al menos una variable para cada grupo descrito en la Sección 3.1: demográfica, social, laboral y vivienda. Lo cual se podría interpretar como una señal de la complejidad del fenómeno estudiado. No sería posible comprender la DCI sin alguna de las dimensiones mencionadas. Probablemente se requerirá mucha más información para comprender las razones de la DCI en Ecuador. Por ejemplo, las variables sociales relacionadas a auto-identificación a la etnia indígena aumentan la probabilidad de DCI y las variables relacionadas a baja educación presentan la misma relación (ver detalle de regresión con errores estándar Anexo No.4).

En general, las variables que representen baja calidad de la vivienda tienen una relación negativa con el Z-Score y la información de mercado laboral, claramente relacionada con ingresos, presenta una relación positiva con el Z-Score. Finalmente, al igual que se observó en la información descriptiva de la DCI por provincias, nuestra regresión presenta una clara relación negativa del Z-Score con las provincias de la Sierra y positiva con la mayoría de las provincias de la Costa.

5.4. Mapa provincial

Los resultados ilustrados en el Gráfico No. 4 de EPPCs cantonales son similares a los resultados provinciales: las tasas de DCI más altas se encuentran en la Sierra central y dos provincias de la Amazonía (Morona Santiago y Pastaza).

Gráfico No. 4: EPPCs por cantón, mapa. ECV-2014



Fuente: ECV-2014
Elaboración: Autores

En varios estudios¹² se encuentra que la Costa en general presenta muy bajos niveles de DCI, pero podemos observar que Santa Elena es un atípico en este caso. Se sugiere realizar estudios adicionales para comprender este fenómeno. Además, es necesario señalar que el uso e interpretación de la información de Santa Elena se debería realizar con precaución.

Por otra parte, se encuentra que los cantones menos afectados por DCI se mantienen en la región Costa y Galápagos. Los cantones en la región Costa presentan poca variabilidad de DCI, mientras que los cantones de la Sierra y Amazonía presentan mayor variabilidad. Por ejemplo, de los 21 cantones en Manabí, 18 presentan DCI menor a 26%. Se puede calcular que los EPPC de los 21 cantones presentan un desvío estándar de 3,1. Guayas, El Oro y Los Ríos tienen valores alrededor de ~1,8 en este estadístico. Por otro lado, la Sierra Central y varias provincias de la Amazonía presentan dispersiones mucho más grandes entre sus cantones, por ejemplo, el desvío estándar: Pastaza (6,4), Chimborazo (5,3), Orellana (5,2), Cotopaxi (4,9) y Bolívar (4,9).

6. Resultados ENSANUT-2018

6.1. Selección de modelo

Al igual que en ECV-2014, se encuentra muy poca variabilidad tanto en AIC como en RMSE entre los 30 modelos (excluyendo los atípicos en RMSE al usar todas las potenciales variables), pero ligeramente menor. En promedio, el AIC relativo de todos los modelos 0,2% y el RMSE al incluir todos los modelos 23,5% y 0,56% al excluir los modelos que usan todas las variables potenciales que resulta aquí, al igual que en ECV, un grupo atípico para RMSE únicamente.

El grupo con menor AIC relativo se encuentra en el conjunto de variables con una diferencia porcentual del 15%, igual que en ECV. La mayor diferencia frente a AIC menor es 0,61% en el modelo derivado de Ada-ENET con $\alpha=0,1$ partiendo del subgrupo de variables con una estimación puntual con una diferencia máxima del 10% entre encuesta y censo (ver Gráfico No. 5)

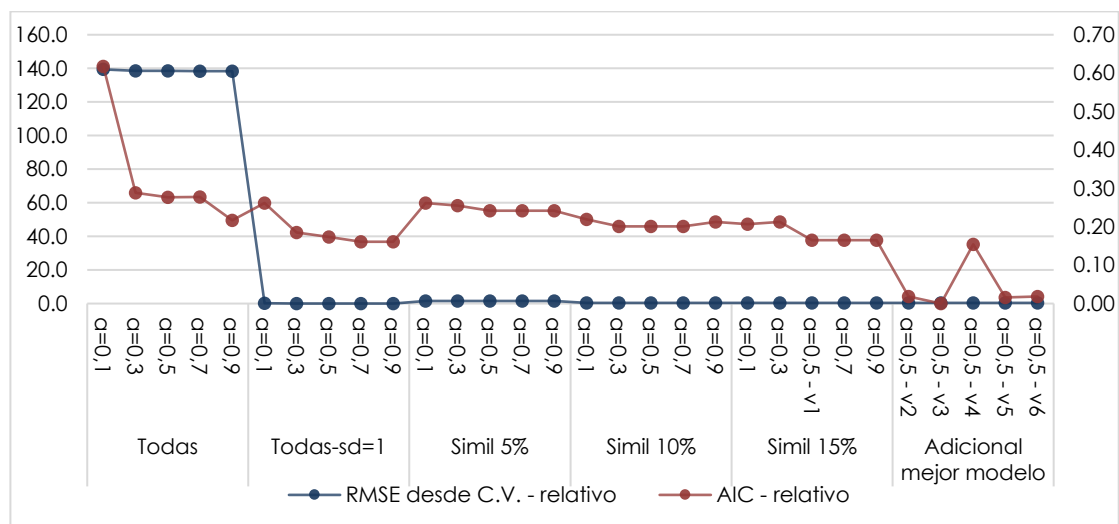
Estos resultados no nos permiten tener un modelo claramente superior al resto, razón por la cual nos encontramos en el mismo escenario que en ECV, por lo tanto, recurriremos al mismo criterio complementario: total de cantones con $CV>30\%$.

En el Gráfico No. 6 encontramos el mismo orden de grupos que en ECV: i) potenciales variables con: i) total variables, ii) total variables con variables continuas estandarizadas, iii) similitud del 5%, iv) 10%, v) 15%, v) modelos adicionales del mejor modelo. Vemos que en el grupo de los mejores modelos se pierden entre 65 a 97 cantones, una caída drástica frente a ECV-2014 donde se perdían alrededor de 8 cantones (al excluir modelos variantes del mejor), y el mejor modelo logró generar una pérdida únicamente de 7 cantones. Esta diferencia se puede deber principalmente a la distancia entre censo y encuesta, que en este caso es el doble que ECV (2010 a 2014 vs. 2010 a 2018). Esto sucede pese a que la muestra de niños entre 0 a 5 años en ENSANUT es el doble (~20 mil) que ECV (~11 mil). No se debe olvidar el efecto adicional que genera el no tomar en cuenta el diseño muestral.

¹² Amplias recopilaciones de resultados se pueden encontrar en: Secretaría Técnica Ecuador Crece Sin Desnutrición Infantil (2021), Larrea (2006) & Vaca (2018)

En este caso la selección de modelo es más complicada que en ECV. Las variantes al mejor modelo empeoran levemente (62 a 65 para v4) sus resultados en unos casos y en otros drásticamente (62 a alrededor de 100 para el resto de versiones). Al existir esta diferencia del comportamiento de los modelos alternativos al mejor entre ECV y ENSANUT se requiere profundizar las razones a este fenómeno, pero el mismo se encuentra fuera del alcance del presente documento. Se escoge el modelo v4 para buscar mantener algo de coherencia con los resultados de ECV (solo se pierde tres cantones con respecto a v1).

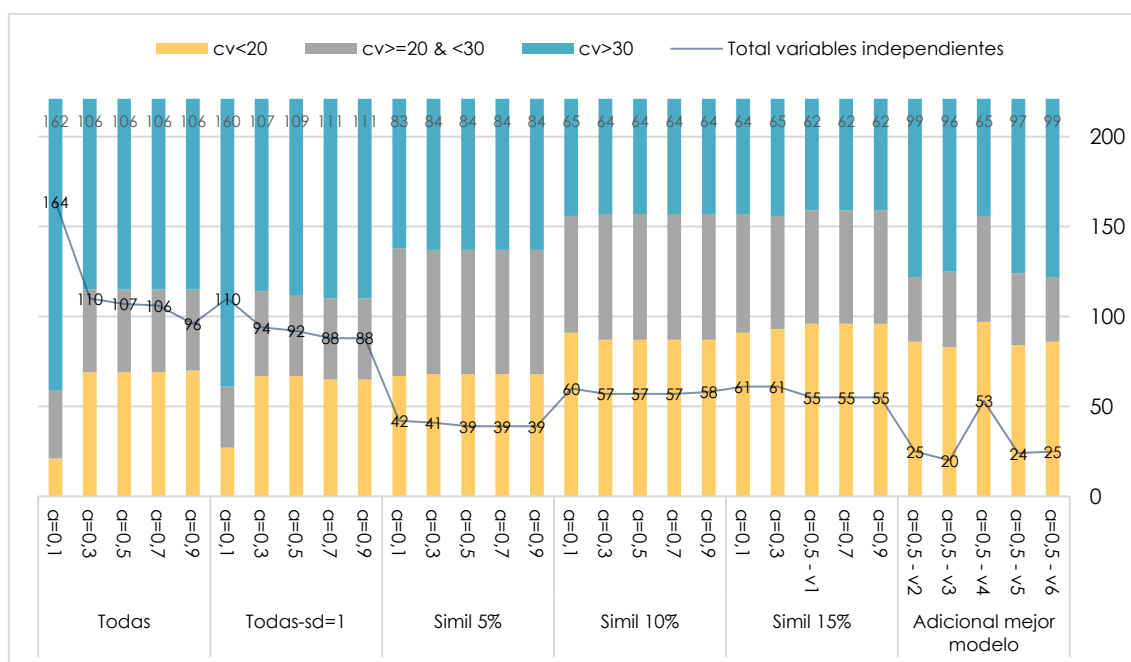
Gráfico No. 5: AIC y RMSE de Validación Cruzada de Modelos ENSANUT-2018



Fuente: ENSANUT-2018

Elaboración: Autores

Gráfico No. 6: Total cantones clasificados por Coeficiente de Variación, por modelo. ENSANUT-2018



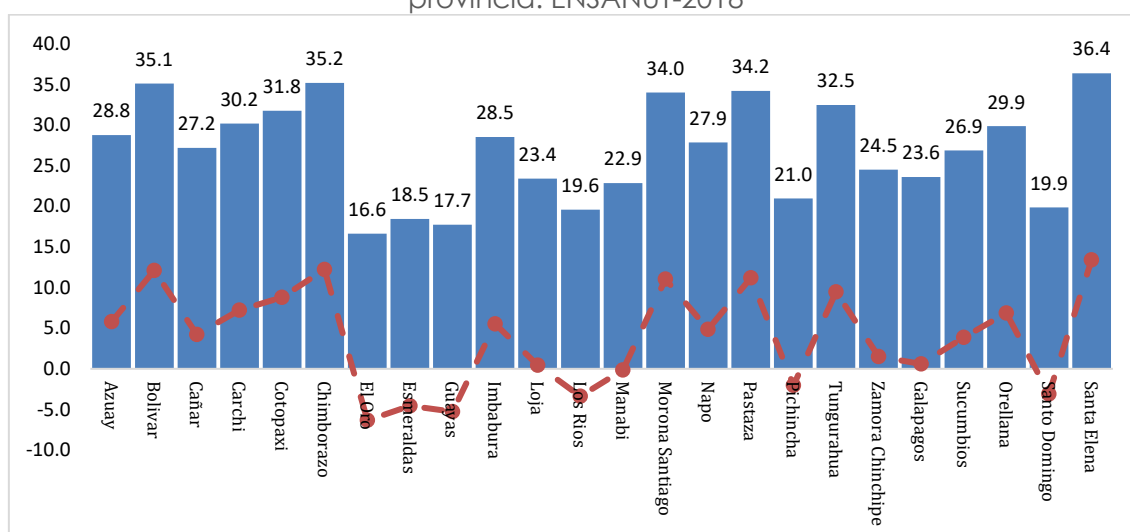
Fuente: ENSANUT-2018

Elaboración: Autores

6.2. Descriptivos relevantes

En el Gráfico No. 7 se observa la diferencia entre la estimación puntual de cada provincia y el promedio nacional (23,0%). Así, los resultados sugieren que la provincia con mayor DCI a nivel nacional es Santa Elena (Costa) con 36,4%, seguida de Chimborazo 35,2% (Sierra) y Bolívar (Sierra) 35,1%, Pastaza (Oriente) 34,2% y Morona Santiago 34% (Oriente). Al igual que en ECV-2014, solo las provincias de la Costa presentan tasas de DCI menores al promedio nacional siendo el Oro (16,6%) y Guayas (17,6%) las provincias con menores tasas de DCI a nivel nacional. En general, los resultados confirman a Santa Elena como la única provincia de la Costa con DCI mayor al promedio nacional y con la mayor tasa a nivel nacional (ver Gráfico No. 7).

Gráfico No. 7: Diferencia tasa DCI con relación a promedio nacional por provincia. ENSANUT-2018



Fuente: ENSANUT-2018

Elaboración: Autores

6.3. Regresión

Para ENSANUT-2018 se presentan las mismas variaciones al mejor modelo observando patrones muy similares que en ECV-2014. Los seis (6) modelos estimados presentan los mismos signos en los coeficientes que comparten entre sí (ver detalle de regresión con errores estándar Anexo No.5). En general, los coeficientes presentan diferencias muy pequeñas entre modelos, curiosamente el signo de los efectos de edad y edad al cuadrado se mantienen con el mismo signo entre las dos encuestas (ver Tablas No. 1-4). También se puede observar que el rango de variables seleccionadas por Ada-ENET recoge todos los grupos de variables del censo, al igual que en ECV-2014.

Pese a que el modelo (3) tiene casi 35 variables menos que el modelo base (1) la diferencia máxima que podemos observar entre el máximo AIC y el mínimo AIC (BIC) de los cuatro modelos es 0,17% (0,48%). Esta es una robustez de resultados poco común en la práctica econométrica.

Tabla No. 2: Resultados NER, versiones de modelo final. Parte I. ENSANUT-2018

	Talla para la edad (OMS)					
	[1]	[2]	[3]	[4]	[5]	[6]
Edad en meses	-0,042 ***	-0,042 ***	-0,042 ***	-0,042 ***	-0,042 ***	-0,042 ***
edad en meses^2	0,641 ***	0,640 ***	0,641 ***	0,641 ***	0,640 ***	0,640 ***
sexo	-0,138 ***	-0,138 ***	-0,139 ***	-0,138 ***	-0,138 ***	-0,138 ***
(1==Azuay)	-0,033			-0,030		
(1==Loja)	0,174			0,172		
(1==Los Ríos)	0,204 *			0,203 *		
(1==Manabí)	0,005			0,003		
(1==Pichincha)	0,125			0,122		
(1==Tungurahua)	-0,241			-0,243		
(1==Orellana)	0,164	0,188				0,188
(1==Santo Domingo)	0,278 *			0,274		
(1==Santa Elena)	-0,537 ***	-0,563 ***	-0,554 ***	-0,536 ***	-0,562 ***	-0,563 ***
(1==Cañar)	-0,090			-0,095		
(1==Carchi)	-0,195	-0,238 *	-0,252 *	-0,201	-0,246 *	-0,238 *
(1==Chimborazo)	-0,209 *	-0,244 *	-0,253 *	-0,210 *	-0,253 *	-0,244 *
(1==Esmeraldas)	-0,083			-0,087		
Región Costa	0,220 ***	0,276 ***	0,278 ***	0,202 **	0,266 ***	0,276 ***
Región Sierra	-0,068			-0,086		
Ciudad Quito	-0,171			-0,171		
(1== indígena)	-0,125 *	-0,174 ***	-0,181 ***	-0,124 *	-0,172 ***	-0,174 ***
(1== negro/a)	0,131			0,133		
(1== indígena, montubio, o mestizo)	-0,235 ***	-0,258 ***	-0,257 ***	-0,236 ***	-0,259 ***	-0,258 ***
(1==tiene 25-64 años)	0,052	0,073 *		0,051	0,074 *	0,073 *
(1== participa en el mercado laboral)	0,063			0,064		
(1== trabaja en rama actividad, agricultura, ganadería y pesca)	-0,046			-0,046		
(1== trabaja en rama actividades profesionales)	-0,100			-0,101		
(1== trabaja en rama artísticas)	0,194 **			0,196 **		
(1== trabaja en rama suministro de electricidad)	0,154			0,153		
(1==Logaritmo de edad)	0,048			0,048		
Agua que toman: La hierven	-0,030					
Dependencia demográfica	-0,014			-0,014		
Total de hijos vivos de mujeres 12-49 en el hogar	-0,027 ***	-0,024 **	-0,026 ***	-0,027 ***	-0,024 **	-0,024 **
Total hijos nacidos siguen vivos	0,102	0,100		0,103	0,101	0,100
El hogar tiene acceso a teléfono convencional	0,064 *	0,076 **	0,087 **	0,063 *	0,076 **	0,076 **
Pago de vivienda: Propia y la está pagando	0,059			0,060		
Pago de vivienda: Arrendada	-0,055	-0,054		-0,055	-0,054	-0,054
(1== etnia indígena)	-0,058			-0,057		
(1== tiene 15-24 años)	-0,040			-0,040		
(1==tiene 22+ años)	0,054	0,066 *	0,080 **	0,055	0,066 *	0,066 *
(1==trabaja en ocupación empleado oficina)	0,045			0,045		
(1== trabaja en ocu, operadores de instalación)	-0,089			-0,089		
(1== no participa en el mercado laboral)	-0,045	-0,056 *	-0,060 **	-0,046	-0,056 *	-0,056 *
(1==trabaja en rama actividad actividades)	0,129			0,130		
(1== trabaja en rama actividad de los hogares)	0,115			0,112		
Total cuartos exclusivos para dormir ^0,5	0,083 **	0,096 ***	0,100 ***	0,084 **	0,096 ***	0,096 ***
Total niños (<=12 años)^0,5	-0,240 ***	-0,260 ***	-0,246 ***	-0,238 ***	-0,260 ***	-0,260 ***
Fuente agua: Agua por tubería red pública o pozo	0,039	0,059		0,040	0,059	0,059
Servicio higiénico: Conectado a pozo séptico	-0,025			-0,026		
Índice de calidad (más mejor)	0,052 ***	0,057 ***	0,059 ***	0,052 ***	0,056 ***	0,057 ***
Material paredes: Madera	-0,095 *	-0,108 **	-0,107 **	-0,095 *	-0,105 **	-0,108 **
Estado paredes: Regular	-0,023			-0,024		
Material piso: Duela, parquet, tablón o piso flotante	0,078			0,078		
Material piso: Ladrillo o cemento	-0,070 **	-0,087 ***	-0,090 ***	-0,070 **	-0,088 ***	-0,087 ***
Material piso: Pisos que no son de tierra	0,163 **	0,166 **	0,165 **	0,164 **	0,170 **	0,166 **
Tipo vivienda: Casa o Villa	-0,069 *	-0,069 *	-0,063 *	-0,070 *	-0,069 *	-0,069 *
constante	-0,683 ***	-0,562 ***	-0,409 ***	-0,676 ***	-0,554 ***	-0,562 ***
N	20.287	20.287	20.287	20.287	20.287	20.287
AIC	74.149	74.040	74.026	74.140	74.038	74.040
BIC	74.608	74.262	74.208	74.584	74.252	74.262

*** p < 0,001; ** p < 0,01; * p < 0,05

Nota: Modelos con alfa=0,5 y similitud del 15%.

Fuente: ENSANUT-2018

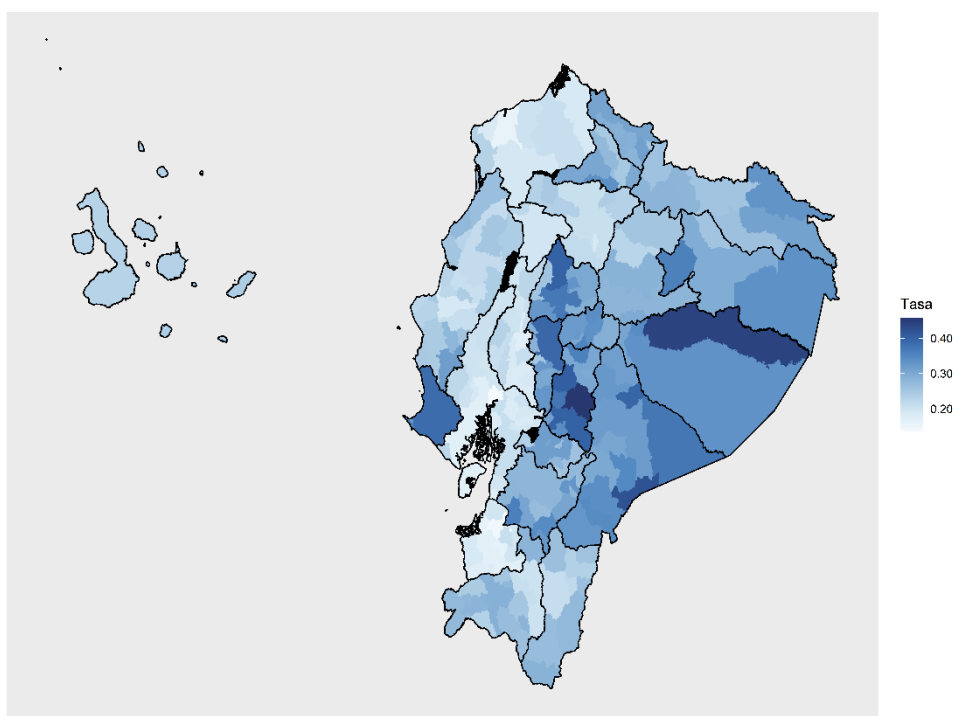
Elaboración: Autores

Para ENSANUT-2018 se presentan las mismas variaciones al mejor modelo observando patrones muy similares que en ECV-2014. Los seis (6) modelos estimados presentan los mismos signos en los coeficientes que comparten entre sí. En general, los coeficientes presentan diferencias muy pequeñas entre modelos, curiosamente el signo de los efectos de edad y edad al cuadrado se mantienen con el mismo signo entre las dos encuestas (ver Tablas No. 1-4). También se puede observar que el rango de variables seleccionadas por Ada-ENET recoge todos los grupos de variables del censo, al igual que en ECV-2014.

6.4. Mapa provincial

En el Gráfico No. 8 se aprecia un resumen general de los resultados a nivel cantonal del mejor SAE-NER para DCI en ENSANUT-2018, exactamente el modelo que parte de variables potenciales con similitud del 15% utilizando un Ada-ENET con $\alpha = 0,5$ y sin modificaciones adicionales. Este mapa resulta muy similar al de ECV-2014. Santa Elena resulta un atípico en la Costa. En general esta región se mantiene con niveles bajos de DCI y poca variabilidad entre cantones. Por otro lado, la Sierra Central presenta los cantones con mayores tasas de DCI y a la vez presenta mayor variabilidad en la misma. Otro caso interesante es la Amazonía, tanto en Pastaza como en Morona se tienen mayoritariamente cantones azules medianamente claros y dos cantones azules marcadamente oscuros (Gráfico No. 8).

Gráfico No. 8: EPPCs por cantón, mapa. ENSANUT-2018



Fuente: ENSANUT-2018
Elaboración: Autores

7. Reflexiones finales

Dado que la metodología de cálculo de DCI es teóricamente idéntica a un FGT_0 , o tasa de pobreza, la metodología de Molina & Rao (2010) se puede aplicar sin modificación para obtener estimaciones más precisas a nivel cantonal. Las propuestas adicionales sobre selección de variables, modelo y Conformidad no dependen de la estimación de DCI o algún FGT_0 , sino, son generales para la metodología.

La propuesta del Condición de Conformidad (CC) resulta ser una corrección necesaria que genera propiedades deseables e ineludibles de un proceso de estimación SAE, pero aumenta considerablemente los errores estándares. Estudios adicionales se requerirán para comprender a mayor profundidad las propiedades y efectos en la selección de modelo y variables de la CC.

Para obtener las estimaciones presentadas hemos desarrollado códigos en R, que, resultan novedosos en la literatura SAE-NER y SAE-NER con Conformidad de alta eficiencia computacional; Ada-ENET con n iteraciones sobre la validación cruzada original. Se presentan estos códigos en el Anexo 3 para el uso de algún lector interesado.

Los resultados obtenidos muestran que, para ECV-2014, 158 cantones resultan tener un CV menor a 20% y solamente 8 cantones poco precisos resultan tener un CV mayor al 30%. Para ENSANUT-2018, se pierde precisión: se obtienen 97 cantones con un CV menor al 20% y 65 poco precisos con CV mayor al 30%. En las dos encuestas se mantiene tasas de DCI relativamente altas en Santa Elena y en la Sierra Central. Igualmente, desde ambas encuestas se estima bajos niveles de DCI en la Costa.

Los resultados finales con desagregación cantonal, sin llegar a constituir cifras oficiales, se espera sean una contribución útil a la toma de decisiones basadas en evidencia.

Referencias

- Ampaabeng, S. & Ming, C. (2013). The long-term cognitive consequences of early childhood malnutrition: The case of famine in Ghana. *Journal of Health Economics*. <https://doi.org/10.1016/j.jhealeco.2013.08.001>
- Braun, J.; Held, L.; Sabanés B. (2014). Choice of generalized linear mixed models using predictive crossvalidation. *Computational Statistics Data Analysis*, <https://doi.org/10.1016/j.csda.2014.02.008>
- Battese, G.E, Harter, R.M., Fuller, W.A. (1988). An error-component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.
- Behrman, J.R. and Hoddinott, J. (2005). Programme evaluation with unobserved heterogeneity and selective implementation: The Mexican PROGRESA Impact on Child Nutrition. <https://doi.org/10.1111/j.1468-0084.2005.00131.x>
- Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*. <https://doi.org/10.1126/science.1248429>
- Demidenko E. (2004). *Mixed models, theory and applications*. John Wiley, New York.
- Dewey, K. & Begum, K. (2011). Long-term consequences of stunting in early life. *Maternal and Child Nutrition*. <https://doi.org/10.1111/j.1740-8709.2011.00349.x>
- Elbers, C., Lanjouw, J. O., & Lanjouw, P., (2003). Micro-Level Estimation of Poverty and Inequality, *Econometrica*, Vol. 71, No. 1, 355–364. <https://www.jstor.org/stable/3082050>
- Foster, J., Greer, J. and Thorbecke, E. (1984). A class of decomposable poverty measures, *Econometrica*, 52, 761-766.
- Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer. 12th review 2017. https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12_toc.pdf
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22. <https://doi.org/10.18637/jss.v033.i01>
- Galler, J. R., Bryce, C., Waber, D. P., Zichlin, M. L., Fitzmaurice, G. M., & Eaglesfield, D. (2012). Socioeconomic outcomes in adults malnourished in the first year of life: a 40-year study. *Pediatrics*. <https://doi.org/10.1542/peds.2012-0073>
- García, J., Heckman, J., Leaf, D. & Prados, M. J. (2019). *Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program*. National Bureau of Economic Research - Working Paper Series. <https://doi.org/10.3386/w23479>
- Goldstein, H. (2003). *Multilevel Statistical Models*. Arnold, London.
- González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., Santamaría, L. (2008a). Bootstrap mean squared error of small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78, 443-462.
- Hobza, T., Morales, D. (2016). Empirical best prediction under unit-level logit mixed models. *Journal of official statistics*, 32, 3, 661-69.

- Heckman, J. J., N. Hohmann, J. Smith, & M. Khoo (2000). Substitution and dropout bias in social experiments: A study of an influential social experiment. *Quarterly Journal of Economics*.
- Heckman, J. J., Moon, S. H., R. Pinto, P., Savelyev, & Yavitz, A. Q. (2010a, July). Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program. *Quantitative Economics*
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. Q. (2010b, February). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*
- James, G., Witten, D., Hastie, T. & Tibshirani, T. (2013). *Introduction to Statistical Learning with applications in R*. Springer.
https://web.stanford.edu/~hastie/ISLR2/ISLRv2_website.pdf
- Jiang, J. & Rao, J. (2020) Robust Small Area Estimation: An Overview. *Annual Review of Statistics and Its Application*, <https://doi.org/10.1146/annurev-statistics-031219-041212>
- Jiang, J., P. Lahiri. (2001). Empirical best prediction for small area inference with binary data. *Annals of the Institute of Statistical Mathematics*, 53, 217-243.
- Jiang, J. (2003). Empirical best prediction for small-area inference based on generalized linear mixed models. *Journal of statistical planning and inference*, 111, 117-127.
- Jiang, J (2007). *Linear and generalized linear mixed models and their applications*. Springer.
- Larrea, C. (2006) *Desnutrición, Etnicidad y Pobreza en el Ecuador y el Área Andina*. Universidad Andina Simón Bolívar, Sede Ecuador.
<https://repositorio.uasb.edu.ec/bitstream/10644/856/1/LARREAC-CON0008-DESNUTRICION.pdf>
- Leng, C., Lin, Y., & Wahba, G. (2006). A note on the Lasso and Related procedures in model selection. *Statistica Sinica*, 16(4), 1273-1284.
<http://www.jstor.org/stable/24307787>
- Longford, N.T. (1995). *Random coefficient models*. Clarendon Press, London.
- McCulloch, C.E., Searle, S.R. (2001). *Generalized, Linear and Mixed Models*. John Wiley, New York.
- Molina, A., Cabrera, E., Moreno, L., Sharman, M.A., y Cueva F. (2015) Mapa de Pobreza y Desigualdad por consumo Ecuador 2014. Instituto Nacional de Estadística y Censos y Banco Mundial (INEC-BM). Quito-Ecuador.
- Molina, I. (2019) Desagregación de datos en encuestas de hogares: Metodologías de estimación en áreas pequeñas. *Series Estudios Estadísticos*, No 97. CEPAL.
https://repositorio.cepal.org/bitstream/handle/11362/44214/5/S1900419_es.pdf
- Molina, I. & Rao, J.N.K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38, 369-385.
- Morales, D., Esteban, M., Pérez, A., Hobza, T. (2021). *A course on small area estimation and mixed models*. Springer
- Roberts S. & Nowak G. (2014) Stabilizing the lasso against cross-validation variability. *Computational Statistics & Data Analysis*
- Searle, S.R., Casella, G., McCulloch, C.E. (1992). *Variance Components*. John Wiley, New York.

Secretaría Técnica Ecuador Crece Sin Desnutrición Infantil (2021). *Plan Estratégico Intersectorial para la Prevención y Reducción de la Desnutrición Crónica Infantil*. <https://www.infancia.gob.ec/wp-content/uploads/2021/09/Plan-Intersectorial.pdf>

Suárez, J. (2020). *El flagelo de la desnutrición infantil en América Latina*. Visiones – Banco de Desarrollo de América Latina. <https://www.caf.com/es/conocimiento/visiones/2020/03/el-flagelo-de-la-desnutricion-infantil-en-america-latina/>

Schoenmaker C., Juffer F., van Ijzendoorn, M., Linda van den Dries, Linting, M., van der Voort, A. & Bakermans-Kranenburg, M. (2015). Cognitive and health-related outcomes after exposure to early malnutrition: The Leiden longitudinal study of international adoptees. *Children and Youth Services Review*. <https://doi.org/10.1016/j.chilcyouth.2014.12.010>

Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, 92(2), 351-370. <https://doi.org/10.1093/biomet/92.2.351>

Vaca, M. (2018). Cuando las políticas fallan. Desafíos en la reducción de la desnutrición crónica infantil en el Ecuador. *Revista del CLAD Reforma y Democracia*, (71), 155-188.

Wong T. & Yeh P. (2020). Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering* <https://doi.org/10.1109/TKDE.2019.2912815>

World Health Organization. (2006). WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization. <http://doi.org/10.1037/e569412006-008>

Zou, H., & Zhang, H. H. (2009). On The Adaptive Elastic-Net with a Diverging Number of Parameters. *Annals of statistics*, <https://doi.org/10.1214/08-AOS625>

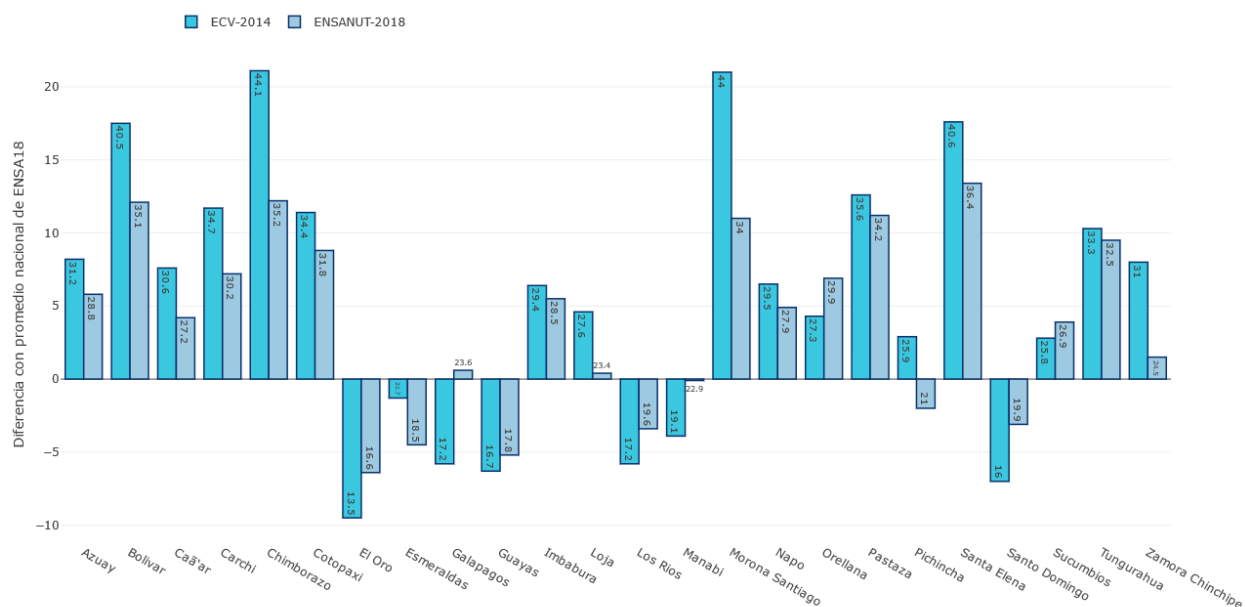
Zucchini, W. (2000) An Introduction to Model Selection. *Journal of Mathematical Psychology*, <https://doi.org/10.1006/jmps.1999.1276>

Anexos

Anexo No. 1: resultados complementarios

Anexo No. 1.1: Estimaciones provinciales DCI, ECV-2014 vs ENSANUT-2018

Gráfico No. 9: ECV-2014 vs ENSANUT-2018: estimaciones provinciales directas

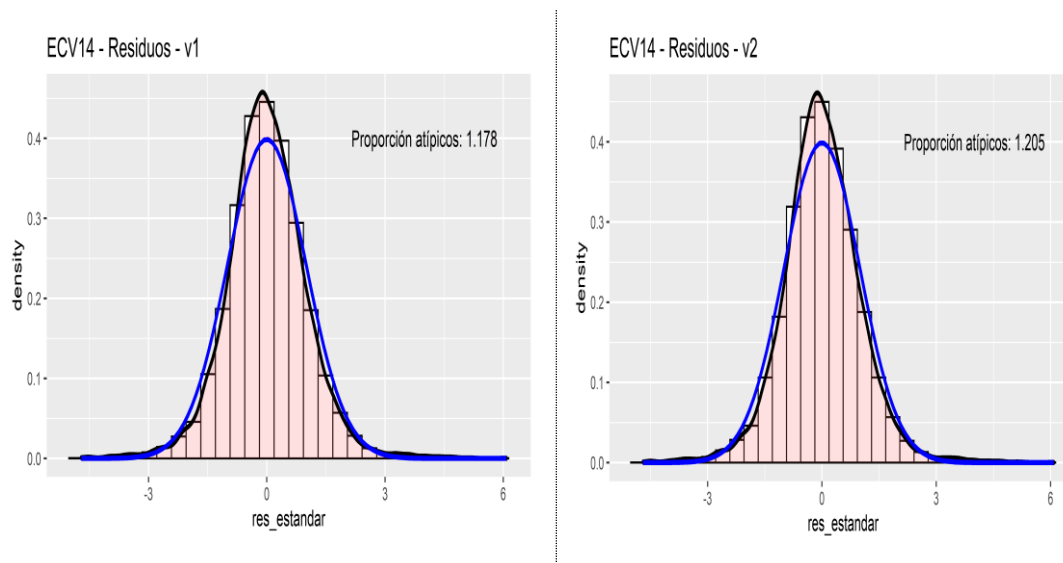


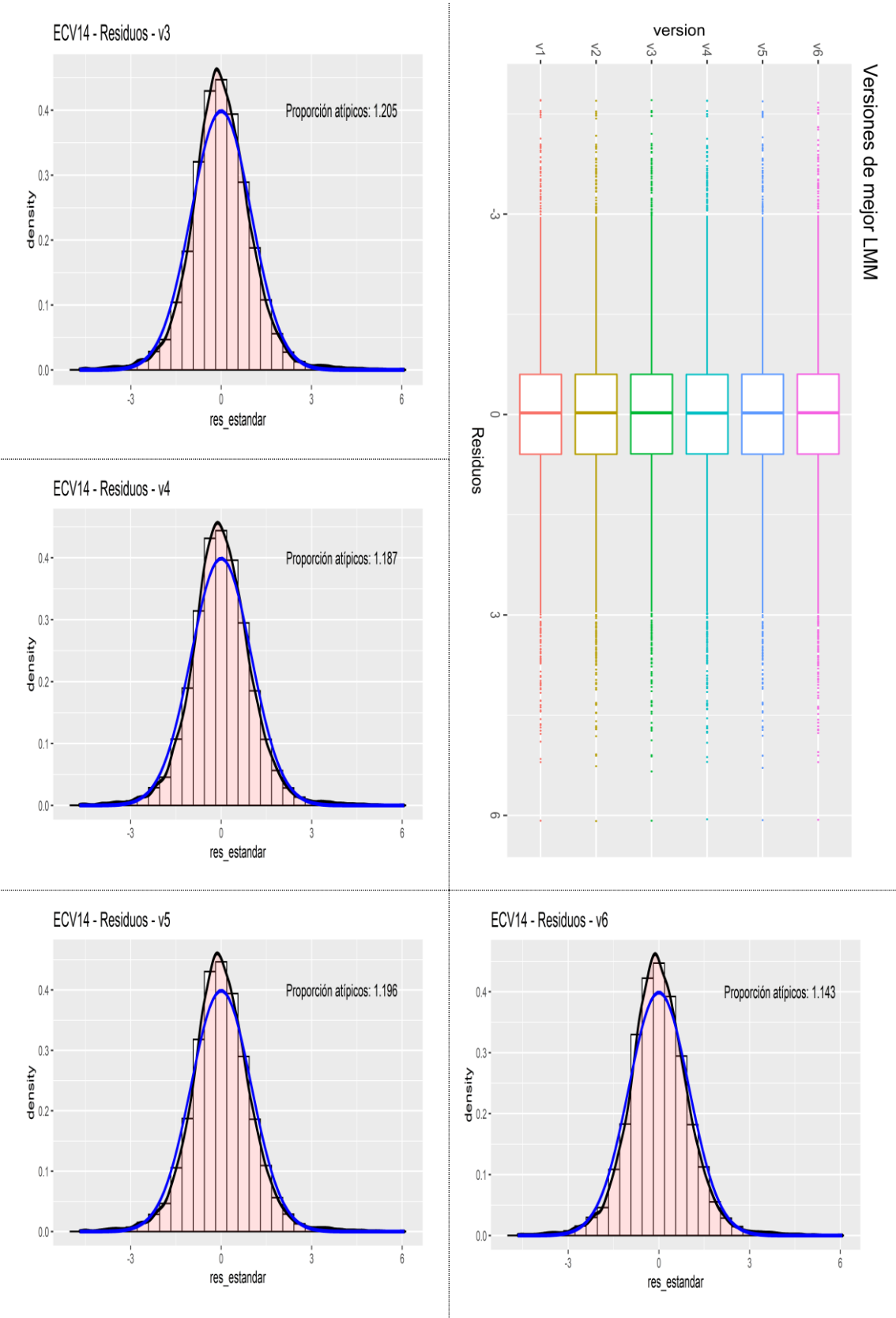
Fuente: ECV-2014 y ENSANUT-2018

Elaboración: Autores

Anexo No. 1.2: Distribución de residuos LMM por versión de modelo final, ECV-2014.

Gráfico No. 10: Distribuciones de residuos por versión de modelo final. ECV-2014

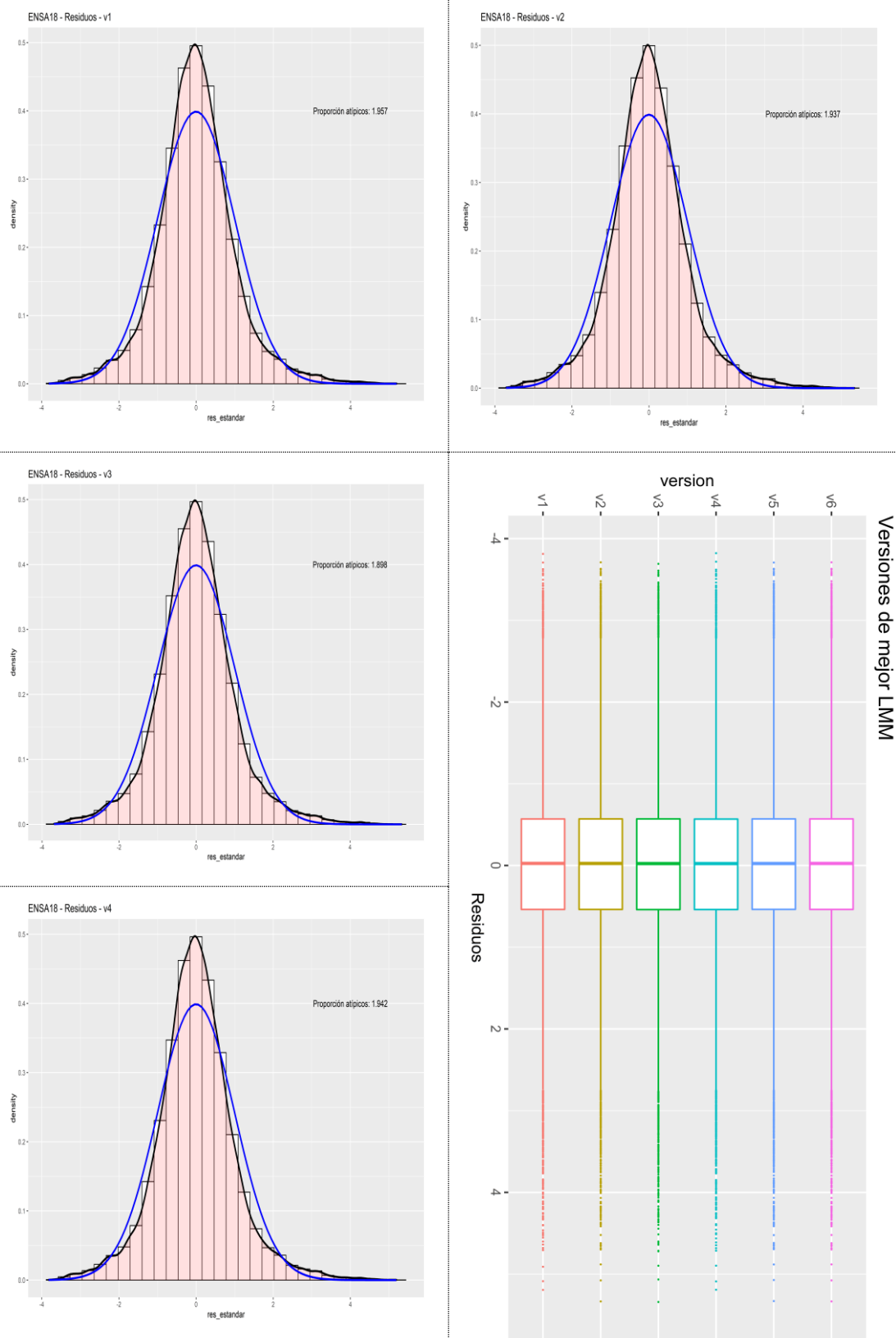


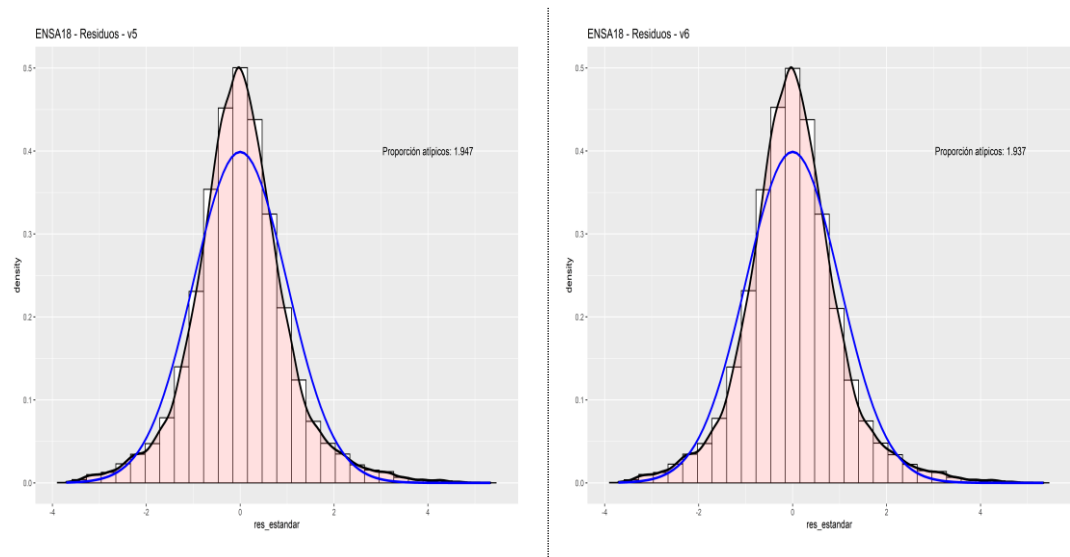


Fuente: ECV-2014
Elaboración: Autores

Anexo No. 1.3: Distribución de residuos LMM por versión de modelo final, ENSANUT-2018.

Gráfico No. 11: Distribuciones de residuos por versión de modelo final. ENSANUT-2018





Fuente: ECV-2014
Elaboración: Autores

Anexo No. 2: resultados EPPC por cantón, ECV-2014 & ENSANUT-2018

Anexo No. 2.1: resultados EPPC por cantón, ECV-2014

Provincia	Cantón	id_can	EPPC	ee_CC	cv_CC	n_muestral
Azuay	Cuenca	0101	29,0%	2,9%	9,97	508
Azuay	Girón	0102	33,7%	4,2%	12,41	19
Azuay	Gualaceo	0103	35,9%	4,2%	11,79	38
Azuay	Nabon	0104	47,8%	5,0%	10,53	23
Azuay	Paute	0105	33,4%	4,4%	13,24	11
Azuay	Pucara	0106	39,3%	4,3%	10,84	10
Azuay	San Fernando	0107	31,7%	4,6%	14,68	6
Azuay	Santa Isabel	0108	33,6%	4,2%	12,44	16
Azuay	Sigsig	0109	40,7%	4,5%	10,96	29
Azuay	Oña	0110	38,6%	5,3%	13,80	0
Azuay	Chordeleg	0111	32,5%	4,2%	13,05	18
Azuay	El Pan	0112	30,8%	5,2%	16,75	3
Azuay	Sevilla De Oro	0113	31,5%	4,4%	14,00	3
Azuay	Guachapala	0114	31,4%	4,6%	14,76	2
Azuay	Camilo Ponce Enríquez	0115	33,3%	4,1%	12,44	27
Bolívar	Guaranda	0201	46,4%	5,3%	11,40	208
Bolívar	Chillanes	0202	38,3%	5,1%	13,18	26
Bolívar	Chimbo	0203	32,6%	4,8%	14,60	25
Bolívar	Echeandia	0204	32,0%	4,3%	13,45	11
Bolívar	San Miguel	0205	33,9%	4,3%	12,68	41
Bolívar	Caluma	0206	28,1%	4,9%	17,43	22
Bolívar	Las Naves	0207	34,4%	4,9%	14,37	3
Cañar	Azogues	0301	28,6%	2,3%	8,06	88
Cañar	Biblian	0302	30,8%	3,4%	11,09	43
Cañar	Cañar	0303	36,7%	2,9%	7,81	110
Cañar	La Troncal	0304	24,9%	2,8%	11,28	76
Cañar	El Tambo	0305	36,0%	3,7%	10,41	36
Cañar	Deleg	0306	30,0%	4,6%	15,37	1
Cañar	Suscal	0307	43,5%	4,9%	11,32	0
Carchi	Tulcán	0401	35,3%	2,8%	7,82	142
Carchi	Bolívar	0402	36,1%	3,8%	10,45	10
Carchi	Espejo	0403	33,6%	3,7%	11,06	22
Carchi	Mira	0404	35,5%	3,5%	9,86	16
Carchi	Montufar	0405	32,9%	3,2%	9,74	52
Carchi	San Pedro De Huaca	0406	33,0%	4,0%	12,06	7
Cotopaxi	Latacunga	0501	31,4%	3,1%	9,86	215
Cotopaxi	La Mana	0502	27,5%	4,1%	15,02	55
Cotopaxi	Pangua	0503	34,4%	4,5%	13,04	17
Cotopaxi	Pujili	0504	39,7%	4,4%	11,07	76
Cotopaxi	Salcedo	0505	34,1%	3,7%	10,92	51
Cotopaxi	Saquisilí	0506	39,6%	3,8%	9,53	38
Cotopaxi	Sigchos	0507	44,3%	5,0%	11,19	19
Chimborazo	Riobamba	0601	35,7%	3,2%	8,92	151
Chimborazo	Alausi	0602	53,1%	5,4%	10,16	36
Chimborazo	Colta	0603	54,7%	4,6%	8,37	29
Chimborazo	Chambo	0604	35,7%	4,9%	13,70	6
Chimborazo	Chunchi	0605	41,2%	5,0%	12,02	20
Chimborazo	Guamote	0606	64,0%	5,8%	9,02	56
Chimborazo	Guano	0607	41,7%	4,4%	10,48	36
Chimborazo	Pallatanga	0608	45,3%	4,5%	9,82	11
Chimborazo	Penipe	0609	36,1%	4,6%	12,84	2
Chimborazo	Cumanda	0610	37,0%	4,4%	11,90	3
El Oro	Machala	0701	11,2%	3,4%	30,25	381
El Oro	Arenillas	0702	14,0%	4,4%	31,23	34
El Oro	Atahualpa	0703	14,2%	4,2%	29,84	6
El Oro	Balsas	0704	15,0%	4,1%	27,71	7
El Oro	Chilla	0705	20,0%	6,3%	31,50	2
El Oro	El Guabo	0706	15,4%	4,7%	30,25	69
El Oro	Huaquillas	0707	17,1%	4,4%	25,69	22
El Oro	Marcabelli	0708	15,8%	4,2%	26,31	9
El Oro	Pasaje	0709	13,8%	3,9%	28,27	64
El Oro	Piñas	0710	13,1%	3,4%	26,13	15
El Oro	Portovelo	0711	15,0%	3,8%	25,63	12
El Oro	Santa Rosa	0712	14,9%	3,8%	25,71	51
El Oro	Zaruma	0713	16,7%	5,1%	30,25	17

El Oro	Las Lajas	0714	15,1%	4,8%	31,68	6
Esmeraldas	Esmeraldas	0801	18,2%	2,0%	10,82	192
Esmeraldas	Eloy Alfaro	0802	25,0%	2,8%	11,18	56
Esmeraldas	Muisne	0803	27,0%	2,7%	9,88	35
Esmeraldas	Quininde	0804	23,6%	2,6%	11,25	119
Esmeraldas	San Lorenzo	0805	21,4%	2,5%	11,80	39
Esmeraldas	Atacames	0806	20,9%	2,7%	12,82	31
Esmeraldas	Rioverde	0807	24,3%	2,7%	10,94	26
Guayas	Guayaquil	0901	16,4%	2,6%	16,11	647
Guayas	Alfredo Baquerizo Moreno	0902	21,8%	4,5%	20,80	39
Guayas	Balao	0903	19,4%	4,4%	22,64	8
Guayas	Balzar	0904	19,5%	4,5%	23,02	14
Guayas	Colimes	0905	20,3%	4,7%	22,90	16
Guayas	Daule	0906	14,7%	3,5%	23,60	68
Guayas	Duran	0907	14,0%	3,2%	22,60	72
Guayas	Empalme	0908	18,9%	3,4%	18,25	46
Guayas	El Triunfo	0909	19,2%	4,2%	21,70	11
Guayas	Milagro	0910	14,9%	3,7%	24,78	64
Guayas	Naranjal	0911	17,8%	3,5%	19,42	57
Guayas	Naranjito	0912	17,0%	3,9%	22,80	15
Guayas	Palestina	0913	19,8%	3,4%	17,37	12
Guayas	Pedro Carbo	0914	20,8%	4,2%	20,39	19
Guayas	Samborondon	0916	14,9%	3,1%	20,48	8
Guayas	Santa Lucia	0918	19,1%	4,4%	23,13	10
Guayas	Salitre	0919	19,8%	4,1%	20,72	27
Guayas	San Jacinto De Yaguachi	0920	19,4%	3,9%	20,34	41
Guayas	Playas	0921	17,6%	3,7%	20,84	20
Guayas	Simón Bolívar	0922	19,2%	3,8%	19,99	28
Guayas	Cnel. Marcelino Maridueña	0923	16,5%	3,9%	23,60	4
Guayas	Lomas De Sargentillo	0924	18,3%	3,9%	21,21	6
Guayas	Nobol	0925	17,9%	3,9%	22,01	4
Guayas	Gnral. Antonio Elizalde	0927	17,5%	3,6%	20,55	1
Guayas	Isidro Ayora	0928	19,6%	4,1%	20,74	0
Imbabura	Ibarra	1001	26,5%	2,2%	8,33	141
Imbabura	Antonio Ante	1002	26,7%	3,2%	11,97	40
Imbabura	Cotacachi	1003	32,7%	3,6%	10,90	46
Imbabura	Otavalo	1004	33,3%	2,8%	8,29	115
Imbabura	Pimampiro	1005	28,9%	3,7%	12,79	14
Imbabura	San Miguel De Urcuqui	1006	29,9%	3,8%	12,59	7
Loja	Loja	1101	24,5%	2,1%	8,64	141
Loja	Calvas	1102	28,9%	3,3%	11,44	24
Loja	Catamayo	1103	24,6%	3,1%	12,75	16
Loja	Celica	1104	29,4%	3,7%	12,75	11
Loja	Chaguarpamba	1105	29,2%	3,7%	12,79	10
Loja	Espindola	1106	34,1%	3,5%	10,35	19
Loja	Gonzanama	1107	33,4%	4,3%	12,99	6
Loja	Macara	1108	25,9%	3,5%	13,45	10
Loja	Paltas	1109	32,1%	3,6%	11,29	27
Loja	Puyango	1110	26,6%	3,3%	12,35	16
Loja	Saraguro	1111	37,0%	3,6%	9,85	32
Loja	Sozoranga	1112	33,5%	3,9%	11,66	6
Loja	Zapotillo	1113	29,0%	3,8%	13,08	18
Loja	Pindal	1114	29,1%	3,5%	12,13	14
Loja	Quilanga	1115	30,0%	3,9%	12,85	0
Loja	Olmedo	1116	33,2%	3,6%	10,77	5
Los Ríos	Babahoyo	1201	15,8%	4,0%	25,29	98
Los Ríos	Baba	1202	19,8%	4,8%	24,45	35
Los Ríos	Montalvo	1203	15,3%	3,8%	24,61	13
Los Ríos	Puebloviejo	1204	16,6%	4,3%	26,01	29
Los Ríos	Quevedo	1205	16,9%	3,5%	20,84	87
Los Ríos	Urdaneta	1206	17,0%	4,6%	26,76	8
Los Ríos	Ventanas	1207	17,1%	3,9%	22,73	34
Los Ríos	Vinces	1208	18,6%	4,2%	22,60	58
Los Ríos	Palenque	1209	18,9%	5,0%	26,59	22
Los Ríos	Buena Fe	1210	17,0%	4,2%	24,71	21
Los Ríos	Valencia	1211	18,3%	4,7%	25,44	39
Los Ríos	Mocache	1212	18,3%	4,4%	24,19	19
Los Ríos	Quinsaloma	1213	17,7%	4,6%	26,20	20
Manabí	Portoviejo	1301	16,3%	3,1%	19,17	55
Manabí	Bolívar	1302	18,2%	4,3%	23,66	13
Manabí	Chone	1303	20,1%	3,4%	17,12	46

Manabí	El Carmen	1304	20,1%	3,7%	18,41	31
Manabí	Flavio Alfaro	1305	22,5%	4,6%	20,36	13
Manabí	Jipijapa	1306	20,4%	4,4%	21,68	18
Manabí	Junín	1307	23,2%	4,5%	19,27	12
Manabí	Manta	1308	16,7%	2,8%	16,99	132
Manabí	Montecristi	1309	18,1%	4,0%	22,20	7
Manabí	Paján	1310	24,1%	4,4%	18,11	16
Manabí	Pichincha	1311	22,1%	4,7%	21,40	17
Manabí	Rocafuerte	1312	18,5%	4,2%	22,45	8
Manabí	Santa Ana	1313	20,9%	4,6%	22,15	20
Manabí	Sucre	1314	18,3%	3,8%	20,59	13
Manabí	Tosagua	1315	19,5%	4,6%	23,46	8
Manabí	24 De Mayo	1316	23,8%	4,6%	19,43	14
Manabí	Pedernales	1317	21,2%	4,9%	22,97	26
Manabí	Olmedo	1318	23,3%	4,8%	20,60	6
Manabí	Puerto Lopez	1319	19,6%	3,5%	18,05	0
Manabí	Jama	1320	22,3%	4,2%	18,73	7
Manabí	Jaramijo	1321	19,3%	4,1%	21,37	0
Manabí	San Vicente	1322	19,8%	4,6%	23,22	0
Morona Santiago	Morona	1401	41,2%	4,4%	10,68	87
Morona Santiago	Gualaquiza	1402	40,2%	4,4%	10,94	55
Morona Santiago	Limón Indanza	1403	40,4%	4,7%	11,67	23
Morona Santiago	Palora	1404	38,7%	5,4%	13,95	11
Morona Santiago	Santiago	1405	38,6%	4,2%	10,95	27
Morona Santiago	Sucua	1406	36,4%	4,6%	12,63	57
Morona Santiago	Huamboya	1407	50,2%	5,6%	11,13	23
Morona Santiago	San Juan Bosco	1408	41,4%	5,3%	12,84	8
Morona Santiago	Taisha	1409	54,1%	5,5%	10,10	182
Morona Santiago	Logroño	1410	48,7%	5,3%	10,86	12
Morona Santiago	Pablo Sexto	1411	43,7%	6,2%	14,09	0
Morona Santiago	Tiwintza	1412	52,0%	5,4%	10,34	27
Napo	Tena	1501	29,0%	3,5%	12,17	298
Napo	Archidona	1503	33,9%	3,9%	11,42	97
Napo	El Chaco	1504	22,0%	3,8%	17,39	13
Napo	Quijos	1507	20,1%	3,4%	16,91	19
Napo	Carlos Julio Arosemena Tola	1509	32,2%	5,2%	16,02	4
Pastaza	Pastaza	1601	34,6%	3,1%	9,07	309
Pastaza	Mera	1602	31,6%	3,6%	11,31	52
Pastaza	Santa Clara	1603	37,6%	4,7%	12,48	12
Pastaza	Arajuno	1604	46,1%	5,1%	11,18	94
Pichincha	Quito	1701	24,7%	4,2%	17,15	669
Pichincha	Cayambe	1702	39,2%	5,9%	14,97	79
Pichincha	Mejía	1703	30,2%	5,2%	17,14	42
Pichincha	Pedro Moncayo	1704	34,7%	6,1%	17,58	29
Pichincha	Rumiñahui	1705	25,5%	4,4%	17,17	35
Pichincha	San Miguel De Los Bancos	1707	32,3%	6,2%	19,25	12
Pichincha	Pedro Vicente Maldonado	1708	33,7%	6,4%	18,92	9
Pichincha	Puerto Quito	1709	34,0%	6,2%	18,22	10
Tungurahua	Ambato	1801	32,4%	4,7%	14,38	240
Tungurahua	Baños De Agua Santa	1802	29,1%	4,7%	16,14	6
Tungurahua	Cevallos	1803	30,4%	5,1%	16,65	10
Tungurahua	Mocha	1804	32,7%	5,0%	15,28	2
Tungurahua	Patate	1805	35,6%	5,5%	15,35	6
Tungurahua	Quero	1806	37,5%	6,3%	16,86	12
Tungurahua	San Pedro De Pelileo	1807	36,5%	5,4%	14,87	35
Tungurahua	Santiago De Pillaro	1808	35,6%	5,3%	14,99	21
Tungurahua	Tisaleo	1809	35,4%	5,9%	16,60	11
Zamora Chinchipe	Zamora	1901	25,0%	3,6%	14,42	67
Zamora Chinchipe	Chinchipe	1902	31,0%	5,6%	17,97	32
Zamora Chinchipe	Nangaritza	1903	37,2%	5,9%	16,01	14
Zamora Chinchipe	Yacuambi	1904	38,1%	6,0%	15,71	21
Zamora Chinchipe	Yantzaza	1905	29,3%	5,3%	18,04	73
Zamora Chinchipe	El Pangui	1906	33,9%	5,0%	14,70	28
Zamora Chinchipe	Centinela Del Condor	1907	30,4%	4,8%	15,86	24
Zamora Chinchipe	Palanda	1908	33,6%	5,7%	16,97	41
Zamora Chinchipe	Paquisha	1909	37,9%	5,9%	15,66	14
Galápagos	San Cristóbal	2001	15,6%	2,7%	17,55	55
Galápagos	Isabela	2002	17,0%	3,9%	23,14	17
Galápagos	Santa Cruz	2003	18,1%	2,4%	13,29	90
Sucumbíos	Lago Agrio	2101	26,3%	2,5%	9,38	192
Sucumbíos	Gonzalo Pizarro	2102	30,4%	3,7%	12,02	36

Sucumbíos	Putumayo	2103	26,5%	3,3%	12,38	26
Sucumbíos	Shushufindi	2104	20,9%	2,9%	13,91	81
Sucumbíos	Sucumbíos	2105	22,9%	3,8%	16,40	6
Sucumbíos	Cáscales	2106	33,8%	4,0%	11,89	49
Sucumbíos	Cuyabeno	2107	28,2%	4,3%	15,27	2
Orellana	Orellana	2201	24,6%	2,8%	11,32	284
Orellana	Aguarico	2202	37,3%	4,3%	11,57	14
Orellana	La Joya De Los Sachas	2203	23,6%	2,7%	11,45	96
Orellana	Loreto	2204	38,4%	3,4%	8,84	71
Santo Domingo De Los Tsáchilas	Santo Domingo	2301	15,7%	10,0%	63,30	325
Santo Domingo De Los Tsáchilas	La Concordia	2302	18,4%	10,1%	54,59	13
Santa Elena	Santa Elena	2401	42,7%	3,1%	7,29	224
Santa Elena	La Libertad	2402	37,7%	3,7%	9,84	111
Santa Elena	Salinas	2403	40,2%	3,4%	8,52	79

Anexo No. 2.1: Resultados EPPC por cantón, ENSA-2014

Provincia	Cantón	id_can	EPPC	ee_CC	cv_CC	n_muestral
Azuay	Cuenca	0101	28,0%	2,7%	9,66	436
Azuay	Girón	0102	26,9%	3,8%	14,12	17
Azuay	Gualaceo	0103	32,8%	3,8%	11,66	18
Azuay	Nabon	0104	34,4%	4,7%	13,65	6
Azuay	Paute	0105	27,6%	4,2%	15,39	6
Azuay	Pucara	0106	36,0%	4,0%	11,12	7
Azuay	San Fernando	0107	27,2%	4,2%	15,42	0
Azuay	Santa Isabel	0108	30,7%	4,1%	13,24	0
Azuay	Sigsig	0109	30,6%	4,0%	13,20	13
Azuay	Oña	0110	32,7%	4,4%	13,55	0
Azuay	Chordeleg	0111	27,9%	3,7%	13,26	0
Azuay	El Pan	0112	26,9%	5,0%	18,53	0
Azuay	Sevilla De Oro	0113	27,6%	3,8%	13,67	0
Azuay	Guachapala	0114	27,8%	4,0%	14,46	0
Azuay	Camilo Ponce Enríquez	0115	28,1%	3,8%	13,60	30
Bolívar	Guaranda	0201	39,3%	2,6%	6,66	204
Bolívar	Chillanes	0202	32,8%	3,5%	10,55	45
Bolívar	Chimbo	0203	32,6%	3,4%	10,34	78
Bolívar	Echeandia	0204	28,5%	4,2%	14,85	15
Bolívar	San Miguel	0205	30,2%	3,5%	11,66	72
Bolívar	Caluma	0206	25,2%	3,6%	14,19	38
Bolívar	Las Naves	0207	29,3%	3,8%	12,96	22
Cañar	Azogues	0301	26,2%	5,3%	20,33	159
Cañar	Biblian	0302	27,8%	6,2%	22,14	58
Cañar	Cañar	0303	31,7%	6,7%	21,11	130
Cañar	La Troncal	0304	23,8%	5,8%	24,48	166
Cañar	El Tambo	0305	26,7%	6,4%	23,89	16
Cañar	Deleg	0306	24,8%	6,4%	25,92	36
Cañar	Suscal	0307	30,8%	7,9%	25,73	15
Carchi	Tulcán	0401	30,8%	4,3%	14,00	254
Carchi	Bolívar	0402	28,3%	4,9%	17,16	98
Carchi	Espejo	0403	28,9%	4,5%	15,39	63
Carchi	Mira	0404	29,2%	4,5%	15,49	73
Carchi	Montufar	0405	31,3%	4,7%	14,98	91
Carchi	San Pedro De Huaca	0406	25,7%	5,0%	19,60	29
Cotopaxi	Latacunga	0501	30,2%	2,5%	8,39	300
Cotopaxi	La Mana	0502	25,7%	3,3%	13,02	50
Cotopaxi	Pangua	0503	32,7%	4,2%	12,80	4
Cotopaxi	Pujili	0504	37,0%	3,1%	8,34	125
Cotopaxi	Salcedo	0505	29,0%	3,0%	10,35	127
Cotopaxi	Saquisilí	0506	33,2%	3,3%	10,07	46
Cotopaxi	Sigchos	0507	40,0%	4,2%	10,41	26
Chimborazo	Riobamba	0601	30,3%	3,8%	12,38	337
Chimborazo	Alausi	0602	40,1%	4,8%	12,05	78

Chimborazo	Colta	0603	40,5%	5,0%	12,42	68
Chimborazo	Chambo	0604	29,9%	4,5%	15,03	18
Chimborazo	Chunchi	0605	34,1%	4,3%	12,68	14
Chimborazo	Guamote	0606	45,8%	5,2%	11,31	87
Chimborazo	Guano	0607	35,9%	4,6%	12,69	69
Chimborazo	Pallatanga	0608	37,3%	5,1%	13,74	0
Chimborazo	Penipe	0609	30,4%	5,0%	16,50	16
Chimborazo	Cumanda	0610	31,4%	4,3%	13,71	7
El Oro	Machala	0701	15,9%	8,7%	54,41	619
El Oro	Arenillas	0702	18,7%	10,0%	53,45	103
El Oro	Atahualpa	0703	16,2%	9,1%	56,18	18
El Oro	Balsas	0704	18,0%	9,6%	53,10	19
El Oro	Chilla	0705	20,9%	12,5%	59,90	0
El Oro	El Guabo	0706	19,4%	10,1%	52,12	144
El Oro	Huaquillas	0707	17,1%	9,3%	54,70	98
El Oro	Marcabeli	0708	17,2%	8,9%	51,89	9
El Oro	Pasaje	0709	14,2%	9,2%	64,82	204
El Oro	Piñas	0710	16,8%	8,5%	50,71	53
El Oro	Portovelo	0711	17,0%	9,5%	55,94	29
El Oro	Santa Rosa	0712	16,8%	8,9%	53,25	166
El Oro	Zaruma	0713	19,6%	9,9%	50,44	42
El Oro	Las Lajas	0714	17,5%	10,1%	57,88	17
Esmeraldas	Esmeraldas	0801	16,0%	9,7%	60,43	503
Esmeraldas	Eloy Alfaro	0802	21,0%	11,8%	56,23	0
Esmeraldas	Muisne	0803	23,2%	11,8%	51,01	30
Esmeraldas	Quinde	0804	19,2%	11,3%	59,11	268
Esmeraldas	San Lorenzo	0805	18,7%	10,5%	56,10	0
Esmeraldas	Atacames	0806	18,4%	10,6%	57,37	144
Esmeraldas	Rioverde	0807	20,3%	11,5%	56,91	47
Guayas	Guayaquil	0901	17,0%	8,3%	48,84	974
Guayas	Alfredo Baquerizo Moreno	0902	18,6%	10,1%	54,34	31
Guayas	Balao	0903	19,6%	10,0%	51,20	0
Guayas	Balzar	0904	20,5%	10,4%	50,80	18
Guayas	Colimes	0905	20,0%	10,5%	52,47	23
Guayas	Daule	0906	17,3%	9,1%	52,32	77
Guayas	Duran	0907	17,7%	8,2%	46,50	67
Guayas	Empalme	0908	19,8%	9,7%	48,78	36
Guayas	El Triunfo	0909	19,3%	9,5%	48,95	9
Guayas	Milagro	0910	17,8%	9,5%	53,27	71
Guayas	Naranjal	0911	21,0%	9,2%	43,98	32
Guayas	Naranjito	0912	18,9%	9,2%	48,62	4
Guayas	Palestina	0913	19,5%	9,1%	46,39	2
Guayas	Pedro Carbo	0914	22,1%	9,8%	44,31	41
Guayas	Samborondon	0916	13,7%	8,3%	60,60	43
Guayas	Santa Lucia	0918	20,0%	10,1%	50,70	21
Guayas	Salitre	0919	19,7%	10,1%	51,05	35
Guayas	San Jacinto De Yaguachi	0920	21,0%	9,7%	46,38	34
Guayas	Playas	0921	22,3%	9,0%	40,31	31
Guayas	Simón Bolívar	0922	20,3%	9,7%	47,63	7
Guayas	Cnel. Marcelino Maridueña	0923	18,2%	9,3%	51,05	12
Guayas	Lomas De Sargentillo	0924	18,9%	9,6%	50,64	9
Guayas	Nobol	0925	19,0%	9,3%	48,69	0
Guayas	Gnral. Antonio Elizalde	0927	19,1%	9,2%	48,50	0
Guayas	Isidro Ayora	0928	20,1%	9,9%	49,23	0
Imbabura	Ibarra	1001	24,9%	2,5%	10,16	279
Imbabura	Antonio Ante	1002	27,0%	3,5%	12,84	52
Imbabura	Cotacachi	1003	30,4%	3,8%	12,65	39
Imbabura	Otavalo	1004	33,8%	3,3%	9,89	186
Imbabura	Pimampiro	1005	30,0%	3,6%	11,84	42
Imbabura	San Miguel De Urcuqui	1006	27,5%	3,8%	13,77	16
Loja	Loja	1101	20,2%	3,4%	16,92	332

Loja	Calvas	1102	26,9%	4,5%	16,73	15
Loja	Catamayo	1103	21,0%	4,4%	20,74	36
Loja	Celica	1104	28,3%	5,2%	18,27	41
Loja	Chaguarpamba	1105	25,3%	5,0%	19,64	20
Loja	Espindola	1106	26,4%	4,8%	18,15	26
Loja	Gonzanama	1107	25,4%	5,1%	20,19	39
Loja	Macara	1108	23,5%	4,5%	19,31	46
Loja	Paltas	1109	28,8%	4,7%	16,44	37
Loja	Puyango	1110	24,6%	5,0%	20,23	24
Loja	Saraguro	1111	28,3%	4,6%	16,25	44
Loja	Sozoranga	1112	28,9%	6,0%	20,77	0
Loja	Zapotillo	1113	27,5%	4,4%	16,07	29
Loja	Pindal	1114	27,9%	4,7%	16,70	20
Loja	Quilanga	1115	28,4%	5,4%	19,07	13
Loja	Olmedo	1116	27,6%	5,6%	20,22	4
Los Ríos	Babahoyo	1201	18,0%	4,6%	25,54	310
Los Ríos	Baba	1202	21,9%	5,6%	25,62	113
Los Ríos	Montalvo	1203	19,0%	4,8%	25,12	46
Los Ríos	Puebloviejo	1204	18,5%	5,1%	27,64	41
Los Ríos	Quevedo	1205	19,4%	4,4%	22,80	309
Los Ríos	Urdaneta	1206	21,0%	5,4%	25,68	35
Los Ríos	Ventanas	1207	17,2%	5,1%	29,81	140
Los Ríos	Vinces	1208	21,0%	5,1%	24,32	144
Los Ríos	Palenque	1209	23,3%	5,6%	24,06	61
Los Ríos	Buena Fe	1210	19,7%	4,9%	24,73	111
Los Ríos	Valencia	1211	22,5%	5,3%	23,51	111
Los Ríos	Mocache	1212	20,0%	5,1%	25,75	63
Los Ríos	Quinsaloma	1213	20,6%	5,3%	25,61	26
Manabí	Portoviejo	1301	19,0%	5,0%	26,07	283
Manabí	Bolívar	1302	20,9%	6,1%	28,99	85
Manabí	Chone	1303	21,5%	5,6%	26,15	159
Manabí	El Carmen	1304	24,3%	5,7%	23,35	51
Manabí	Flavio Alfaro	1305	25,4%	7,0%	27,45	0
Manabí	Jipijapa	1306	24,6%	5,8%	23,78	65
Manabí	Junin	1307	24,2%	6,1%	25,03	11
Manabí	Manta	1308	22,3%	5,1%	22,95	148
Manabí	Montecristi	1309	23,1%	5,6%	24,35	30
Manabí	Pajan	1310	31,7%	6,3%	20,01	56
Manabí	Pichincha	1311	24,2%	6,9%	28,61	22
Manabí	Rocafuerte	1312	22,2%	6,1%	27,57	15
Manabí	Santa Ana	1313	20,9%	6,2%	29,63	73
Manabí	Sucre	1314	22,9%	5,3%	23,29	59
Manabí	Tosagua	1315	21,0%	6,3%	29,87	43
Manabí	24 De Mayo	1316	27,9%	6,6%	23,81	21
Manabí	Pedernales	1317	26,6%	6,5%	24,54	146
Manabí	Olmedo	1318	25,7%	7,1%	27,53	0
Manabí	Puerto Lopez	1319	24,8%	5,6%	22,61	32
Manabí	Jama	1320	27,5%	6,5%	23,68	11
Manabí	Jaramijo	1321	25,1%	6,1%	24,26	22
Manabí	San Vicente	1322	25,5%	6,2%	24,32	22
Morona Santiago	Morona	1401	31,9%	5,1%	16,10	179
Morona Santiago	Gualaquiza	1402	32,8%	5,8%	17,56	66
Morona Santiago	Limón Indanza	1403	34,0%	6,4%	18,71	54
Morona Santiago	Palora	1404	31,5%	5,6%	17,70	11
Morona Santiago	Santiago	1405	30,1%	5,9%	19,51	19
Morona Santiago	Sucua	1406	30,4%	5,0%	16,32	58
Morona Santiago	Huamboya	1407	39,5%	6,1%	15,56	67
Morona Santiago	San Juan Bosco	1408	34,4%	5,8%	16,84	13
Morona Santiago	Taisha	1409	37,2%	6,6%	17,80	61
Morona Santiago	Logroño	1410	34,7%	6,2%	17,81	48
Morona Santiago	Pablo Sexto	1411	32,6%	6,2%	18,92	13

Morona Santiago	Tiwintza	1412	42,1%	7,2%	16,98	45
Napo	Tena	1501	28,1%	10,5%	37,45	508
Napo	Archidona	1503	28,5%	11,0%	38,57	214
Napo	El Chaco	1504	25,5%	8,7%	34,21	45
Napo	Quijos	1507	22,8%	8,2%	36,18	31
Napo	Carlos Julio Arosemena Tola	1509	29,8%	11,6%	38,99	20
Pastaza	Pastaza	1601	33,4%	3,9%	11,73	532
Pastaza	Mera	1602	30,6%	3,7%	12,19	80
Pastaza	Santa Clara	1603	34,2%	5,2%	15,19	32
Pastaza	Arajuno	1604	44,1%	4,9%	11,19	75
Pichincha	Quito	1701	20,8%	4,5%	21,71	706
Pichincha	Cayambe	1702	23,4%	5,5%	23,36	20
Pichincha	Mejía	1703	21,6%	5,0%	23,21	12
Pichincha	Pedro Moncayo	1704	24,1%	5,1%	21,26	9
Pichincha	Rumiñahui	1705	18,8%	4,4%	23,42	14
Pichincha	San Miguel De Los Bancos	1707	23,8%	5,9%	24,73	14
Pichincha	Pedro Vicente Maldonado	1708	24,8%	5,8%	23,46	5
Pichincha	Puerto Quito	1709	23,4%	5,7%	24,57	14
Tungurahua	Ambato	1801	32,1%	3,0%	9,37	447
Tungurahua	Baños De Agua Santa	1802	29,1%	3,9%	13,50	20
Tungurahua	Cevallos	1803	30,7%	4,0%	13,07	13
Tungurahua	Mocha	1804	30,4%	4,1%	13,59	8
Tungurahua	Patate	1805	33,6%	3,9%	11,64	6
Tungurahua	Quero	1806	35,4%	4,2%	12,00	13
Tungurahua	San Pedro De Pelileo	1807	33,7%	3,3%	9,90	38
Tungurahua	Santiago De Pillaro	1808	34,0%	4,1%	11,98	9
Tungurahua	Tisaleo	1809	32,7%	4,0%	12,21	19
Zamora Chinchipe	Zamora	1901	21,1%	8,5%	40,48	172
Zamora Chinchipe	Chinchipe	1902	28,6%	10,1%	35,35	58
Zamora Chinchipe	Nangaritza	1903	27,5%	11,2%	40,94	32
Zamora Chinchipe	Yacuambi	1904	26,5%	11,4%	43,03	36
Zamora Chinchipe	Yantzaza	1905	23,6%	9,9%	41,84	137
Zamora Chinchipe	El Pangui	1906	25,8%	10,1%	39,28	53
Zamora Chinchipe	Centinela Del Condor	1907	21,8%	10,1%	46,20	86
Zamora Chinchipe	Palanda	1908	26,9%	10,6%	39,51	56
Zamora Chinchipe	Paquisha	1909	26,5%	11,0%	41,49	6
Galápagos	San Cristóbal	2001	24,2%	3,8%	15,60	137
Galápagos	Isabela	2002	23,0%	5,2%	22,69	8
Galápagos	Santa Cruz	2003	23,4%	3,6%	15,31	250
Sucumbíos	Lago Agrio	2101	25,6%	7,6%	29,73	417
Sucumbíos	Gonzalo Pizarro	2102	28,3%	8,6%	30,50	35
Sucumbíos	Putumayo	2103	33,1%	8,7%	26,25	43
Sucumbíos	Shushufindi	2104	26,3%	7,7%	29,28	240
Sucumbíos	Sucumbíos	2105	26,1%	7,6%	29,05	30
Sucumbíos	Cascales	2106	28,1%	8,6%	30,69	46
Sucumbíos	Cuyabeno	2107	31,1%	8,5%	27,30	54
Orellana	Orellana	2201	29,5%	3,7%	12,60	533
Orellana	Aguarico	2202	33,6%	4,8%	14,32	25
Orellana	La Joya De Los Sachas	2203	25,7%	3,4%	13,33	177
Orellana	Loreto	2204	35,9%	4,1%	11,50	148
Santo Domingo De Los Tsáchilas	Santo Domingo	2301	19,6%	4,6%	23,35	1320
Santo Domingo De Los Tsáchilas	La Concordia	2302	22,2%	5,2%	23,30	165
Santa Elena	Santa Elena	2401	38,6%	5,2%	13,55	303
Santa Elena	La Libertad	2402	32,6%	5,4%	16,61	207
Santa Elena	Salinas	2403	36,9%	5,1%	13,79	114

Anexo No. 3: Códigos para ejemplificación de uso códigos SAE-NER

Links de acceso en la página institucional correspondiente a la publicación:

- SaeNer-Anexo 3.1 Selección de Variables:
[https://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Libros/cuadernos_trabajo/Anexo 3.1 Selección de Variables.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Libros/cuadernos_trabajo/Anexo_3.1_Selección_de_Variables.pdf)
- SaeNer-Anexo 3.2 Estimación SAE-NER:
[https://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Libros/cuadernos_trabajo/Anexo 3.2 Estimación SAE-NER.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Libros/cuadernos_trabajo/Anexo_3.2_Estimación_SAE-NER.pdf)
- SaeNer-Anexo 3.3 Obtención de Versiones adicionales:
[https://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Libros/cuadernos_trabajo/Anexo 3.3 Obtención de Versiones adicionales.pdf](https://www.ecuadorencifras.gob.ec/documentos/web-inec/Bibliotecas/Libros/cuadernos_trabajo/Anexo_3.3_Obtención_de_Versiones_adicionales.pdf)

Anexo No. 4: Resultados NER, versiones de modelo final con errores estándar. ECV-2014.

	Talla para la edad (OMS)					
	[1]	[2]	[3]	[4]	[5]	[6]
Edad en meses	-0,058*** (0,003)	-0,058*** (0,003)	-0,059*** (0,003)	-0,058*** (0,003)	-0,058*** (0,003)	-0,059*** (0,003)
Edad en meses^2	0,882*** (0,046)	0,885*** (0,046)	0,893*** (0,045)	0,881*** (0,046)	0,885*** (0,046)	0,889*** (0,046)
sexo	-0,119*** (0,023)	-0,118*** (0,023)	-0,118*** (0,023)	-0,119*** (0,023)	-0,118*** (0,023)	-0,116*** (0,023)
(1==Azuay)	-0,147 (0,088)			-0,148 (0,088)		
(1==Los Ríos)	0,068 (0,073)			0,060 (0,073)		
(1==Morona Santiago)	-0,264** (0,080)			-0,268*** (0,079)		
(1==Pastaza)	-0,184 (0,100)			-0,177 (0,099)		
(1==Pichincha)	-0,307*** (0,089)			-0,303** (0,090)		
(1==Tungurahua)	-0,402*** (0,105)			-0,405*** (0,106)		
(1==Zamora Chinchipe)	-0,341*** (0,087)			-0,360*** (0,087)		
(1==Bolívar)	-0,299** (0,102)			-0,305** (0,103)		
(1==Orellana)	0,072 (0,096)					
(1==Santa Elena)	-0,617*** (0,098)			-0,624*** (0,099)		
(1==Carchi)	-0,283* (0,110)			-0,282* (0,111)		
(1==Cotopaxi)	-0,134 (0,089)			-0,133 (0,090)		
(1==Chimborazo)	-0,238* (0,094)			-0,242* (0,095)		
(1==El Oro)	0,106 (0,076)			0,103 (0,076)		
Región Sierra	-0,243*** (0,056)	-0,320*** (0,045)	-0,354*** (0,044)	-0,260*** (0,056)	-0,329*** (0,045)	-0,253*** (0,043)
Total en el hogar secundaria completa	0,131*** (0,029)	0,135*** (0,029)	0,134*** (0,029)	0,129*** (0,029)	0,134*** (0,029)	0,159*** (0,029)
Pago de vivienda: Por servicios	0,145* (0,074)	0,149* (0,074)	0,156* (0,074)			0,111 (0,074)
Total carencias en hogar(NBI)^0,5	-0,123*** (0,019)	-0,127*** (0,017)	-0,142*** (0,016)	-0,117*** (0,019)	-0,127*** (0,017)	
Total niños (<+12 años)^0,5	-0,340*** (0,043)	-0,381*** (0,030)	-0,388*** (0,030)	-0,340*** (0,043)	-0,381*** (0,030)	-0,410*** (0,030)

Agua que toman: La beben tal como llega al hogar	-0,051 (0,035)	-0,063 (0,035)		-0,015 (0,026)	-0,022 (0,026)	-0,099** (0,035)
Agua que toman: La hierven	-0,048 (0,033)	-0,058 (0,033)			-0,079* (0,033)	
Niños (<=12) en el hogar 1-3	0,065 (0,045)			0,060 (0,045)		
Servicio agua: Por tubería fuera de la vivienda	-0,074** (0,027)	-0,087** (0,027)	-0,095*** (0,027)	-0,082** (0,027)	-0,089*** (0,027)	-0,120*** (0,027)
Servicio higiénico: Letrina	-0,181** (0,069)	-0,188** (0,067)	-0,197** (0,067)	-0,184** (0,069)	-0,193** (0,067)	-0,249*** (0,067)
Servicio higiénico: Alcantarillado y pozo séptico	0,049 (0,035)			0,047 (0,035)		
Material piso: Pisos que no son de tierra	0,093* (0,047)	0,095* (0,048)		0,098* (0,047)	0,098* (0,048)	0,186*** (0,046)
Tiene electricidad por red pública	-0,086 (0,058)					
Infante asiste a guardería pública/ONG	-0,053* (0,027)	-0,050 (0,026)		-0,058* (0,026)	-0,052* (0,026)	-0,042 (0,026)
(1==etnia indígena)	-0,170** (0,056)	-0,186*** (0,056)	-0,197*** (0,056)	-0,167** (0,056)	-0,187*** (0,056)	-0,226*** (0,056)
(1==etnia indígena, montubio, o mestizo)	-0,150** (0,046)	-0,154*** (0,046)	-0,158*** (0,046)	-0,148** (0,046)	-0,154*** (0,046)	-0,153*** (0,046)
(1==22+ años)	0,091 (0,066)			0,088 (0,066)		
(1== Primaria incompleta)	-0,077* (0,033)	-0,070* (0,033)	-0,082* (0,033)	-0,077* (0,033)	-0,069* (0,033)	-0,087** (0,033)
(1==Terciaria o superior o postgrado)	0,107** (0,040)	0,116** (0,040)	0,118** (0,040)	0,109** (0,040)	0,118** (0,040)	0,160*** (0,040)
(1== habla lengua indígena)	-0,193*** (0,058)	-0,199*** (0,057)	-0,211*** (0,057)	-0,190*** (0,057)	-0,204*** (0,057)	-0,223*** (0,057)
(1==habla lengua extranjera)	0,117 (0,108)			0,124 (0,108)		
(1==nació en otro país)	0,164 (0,107)			0,166 (0,107)		
(1== tiene trabajo)	-0,154** (0,049)	-0,156** (0,049)	-0,149** (0,049)	-0,149** (0,049)	-0,153** (0,049)	-0,165*** (0,049)
(1==trabaja en rama comercio)	0,059 (0,034)	0,059 (0,034)		0,058 (0,034)	0,059 (0,034)	
(1==no sabe leer y escribir)	-0,102* (0,051)	-0,106* (0,051)		-0,101* (0,051)	-0,102* (0,051)	-0,113* (0,051)
(1== Edad entre 15-24 años)	-0,153*** (0,032)	-0,161*** (0,032)	-0,165*** (0,032)	-0,152*** (0,032)	-0,161*** (0,032)	-0,197*** (0,032)
(1==años de educación)	0,013*** (0,003)	0,013*** (0,003)	0,014*** (0,003)	0,013*** (0,003)	0,013*** (0,003)	0,016*** (0,003)
(1==educación postgrado)	0,165* (0,070)	0,167* (0,070)	0,155* (0,070)	0,166* (0,070)	0,167* (0,070)	0,177* (0,070)
(1==es estudiante)	0,233 (0,333)		0,236 (0,333)			
constante	0,677*** (0,162)	0,793*** (0,110)	0,878*** (0,093)	0,565*** (0,151)	0,755*** (0,109)	0,536*** (0,104)
N	11.289	11.289	11.289	11.289	11.289	11.289
AIC	36.287	36.269	36.246	36.272	36.263	36.314
BIC	36.646	36.474	36.407	36.602	36.454	36.505

*** p < 0,001; ** p < 0,01; * p < 0,05

Anexo No. 4: Resultados NER, versiones de modelo final con errores estándar. ENSANUT-2018.

	Talla para la edad (OMS)					
	[1]	[2]	[3]	[4]	[5]	[6]
Edad en meses	-0,042 *** (0,003)	-0,042 *** (0,003)	-0,042 *** (0,003)	-0,042 *** (0,003)	-0,042 *** (0,003)	-0,042 *** (0,003)
edad en meses^2	0,641 *** (0,043)	0,640 *** (0,043)	0,641 *** (0,043)	0,641 *** (0,043)	0,640 *** (0,043)	0,640 *** (0,043)
sexo	-0,138 ***	-0,138 ***	-0,139 ***	-0,138 ***	-0,138 ***	-0,138 ***

	(0,021)	(0,021)	(0,021)	(0,021)	(0,021)	(0,021)
(1==Azuay)	-0,033			-0,030		
	(0,130)			(0,131)		
(1==Loja)	0,174			0,172		
	(0,095)			(0,096)		
(1==Los Ríos)	0,204 *			0,203 *		
	(0,077)			(0,078)		
(1==Manabí)	0,005			0,003		
	(0,075)			(0,076)		
(1==Pichincha)	0,125			0,122		
	(0,136)			(0,137)		
(1==Tungurahua)	-0,241			-0,243		
	(0,123)			(0,124)		
(1==Orellana)	0,164	0,188				0,188
	(0,107)	(0,115)				(0,115)
(1==Santo Domingo)	0,278 *			0,274		
	(0,133)			(0,135)		
(1==Santa Elena)	-0,537 ***	-0,563 ***	-0,554 ***	-0,536 ***	-0,562 ***	-0,563 ***
	(0,117)	(0,124)	(0,127)	(0,119)	(0,126)	(0,124)
(1==Cañar)	-0,090			-0,095		
	(0,107)			(0,107)		
(1==Carchi)	-0,195	-0,238 *	-0,252 *	-0,201	-0,246 *	-0,238 *
	(0,107)	(0,103)	(0,105)	(0,107)	(0,104)	(0,103)
(1==Chimborazo)	-0,209 *	-0,244 *	-0,253 *	-0,210 *	-0,253 *	-0,244 *
	(0,103)	(0,099)	(0,100)	(0,104)	(0,099)	(0,099)
(1==Esmeraldas)	-0,083			-0,087		
	(0,105)			(0,105)		
Región Costa	0,220 ***	0,276 ***	0,278 ***	0,202 **	0,266 ***	0,276 ***
	(0,063)	(0,044)	(0,044)	(0,062)	(0,044)	(0,044)
Región Sierra	-0,068			-0,086		
	(0,068)			(0,067)		
Ciudad Quito	-0,171			-0,171		
	(0,117)			(0,117)		
(1== indígena)	-0,125 *	-0,174 ***	-0,181 ***	-0,124 *	-0,172 ***	-0,174 ***
	(0,056)	(0,038)	(0,038)	(0,056)	(0,038)	(0,038)
(1== negro/a)	0,131			0,133		
	(0,101)			(0,101)		
(1== indígena, montubio, o mestizo)	-0,235 ***	-0,258 ***	-0,257 ***	-0,236 ***	-0,259 ***	-0,258 ***
	(0,053)	(0,047)	(0,047)	(0,053)	(0,047)	(0,047)
(1==tiene 25-64 años)	0,052	0,073 *		0,051	0,074 *	0,073 *
	(0,033)	(0,031)		(0,033)	(0,031)	(0,031)
(1== participa en el mercado laboral	0,063			0,064		
	(0,042)			(0,042)		
(1== trabaja en rama actividad	-0,046			-0,046		
agricultura, ganadería y pesca)	(0,029)			(0,029)		
(1== trabaja en rama actividades	-0,100			-0,101		
profesionales)	(0,105)			(0,105)		
(1== trabaja en rama artísticas)	0,194 **			0,196 **		
	(0,072)			(0,072)		
(1== trabaja en rama suministro de	0,154			0,153		
electricidad)	(0,116)			(0,116)		
(1==Logaritmo de edad)	0,048			0,048		
	(0,042)			(0,042)		
Agua que toman: La hierven	-0,030					
	(0,022)					
Dependencia demográfica	-0,014			-0,014		
	(0,021)			(0,021)		
Total de hijos vivos de mujeres 12-49 en el	-0,027 ***	-0,024 **	-0,026 ***	-0,027 ***	-0,024 **	-0,024 **
hogar	(0,008)	(0,007)	(0,007)	(0,008)	(0,007)	(0,007)
Total hijos nacidos siguen vivos	0,102	0,100		0,103	0,101	0,100
	(0,054)	(0,054)		(0,054)	(0,054)	(0,054)
El hogar tiene acceso a teléfono	0,064 *	0,076 **	0,087 **	0,063 *	0,076 **	0,076 **
convencional	(0,030)	(0,029)	(0,028)	(0,030)	(0,029)	(0,029)
Pago de vivienda: Propia y la está	0,059			0,060		
pagando	(0,043)			(0,043)		
Pago de vivienda: Arrendada	-0,055	-0,054		-0,055	-0,054	-0,054
	(0,029)	(0,028)		(0,029)	(0,028)	(0,028)
(1== etnia indígena)	-0,058			-0,057		
	(0,058)			(0,058)		
(1== tiene 15-24 años)	-0,040			-0,040		
	(0,036)			(0,036)		

(1==tiene 22+ años)	0,054 (0,029)	0,066 * (0,027)	0,080 ** (0,026)	0,055 (0,029)	0,066 * (0,027)	0,066 * (0,027)
(1==trabaja en ocupación empleado oficina)	0,045 (0,084)			0,045 (0,084)		
(1== trabaja en ocupación operadores de instalaciones y maquinas)	-0,089 (0,126)			-0,089 (0,126)		
(1== no participa en el mercado laboral)	-0,045 (0,025)	-0,056 * (0,023)	-0,060 ** (0,023)	-0,046 (0,025)	-0,056 * (0,023)	-0,056 * (0,023)
(1==trabaja en rama actividad actividades)	0,129 (0,142)			0,130 (0,142)		
(1== trabaja en rama actividad de los hogares)	0,115 (0,090)			0,112 (0,090)		
Total cuartos exclusivos para dormir ^0,5	0,083 ** (0,029)	0,096 *** (0,028)	0,100 *** (0,027)	0,084 ** (0,029)	0,096 *** (0,028)	0,096 *** (0,028)
Total niños (<=12 años)^0,5	-0,240 *** (0,047)	-0,260 *** (0,038)	-0,246 *** (0,037)	-0,238 *** (0,047)	-0,260 *** (0,038)	-0,260 *** (0,038)
Fuente agua: Agua por tubería red pública o pozo	0,039 (0,034)	0,059 (0,033)		0,040 (0,034)	0,059 (0,033)	0,059 (0,033)
Servicio higiénico: Conectado a pozo séptico	-0,025 (0,026)			-0,026 (0,026)		
Índice de calidad (más mejor)	0,052 *** (0,010)	0,057 *** (0,010)	0,059 *** (0,009)	0,052 *** (0,010)	0,056 *** (0,010)	0,057 *** (0,010)
Material paredes: Madera	-0,095 * (0,041)	-0,108 ** (0,040)	-0,107 ** (0,040)	-0,095 * (0,041)	-0,105 ** (0,040)	-0,108 ** (0,040)
Estado paredes: Regular	-0,023 (0,023)			-0,024 (0,023)		
Material piso: Duela, parquet, tablón o piso flotante	0,078 (0,045)			0,078 (0,045)		
Material piso: Ladrillo o cemento	-0,070 ** (0,025)	-0,087 *** (0,024)	-0,090 *** (0,024)	-0,070 ** (0,025)	-0,088 *** (0,024)	-0,087 *** (0,024)
Material piso: Pisos que no son de tierra	0,163 ** (0,062)	0,166 ** (0,061)	0,165 ** (0,061)	0,164 ** (0,062)	0,170 ** (0,061)	0,166 ** (0,061)
Tipo vivienda: Casa o Villa	-0,069 * (0,028)	-0,069 * (0,027)	-0,063 * (0,026)	-0,070 * (0,028)	-0,069 * (0,027)	-0,069 * (0,027)
constante	-0,683 *** (0,201)	-0,562 *** (0,115)	-0,409 *** (0,103)	-0,676 *** (0,201)	-0,554 *** (0,115)	-0,562 *** (0,115)
N	20.287	20.287	20.287	20287	20287	20287
AIC	74.149	74.040	74.026	74.140	74.038	74.040
BIC	74.608	74.262	74.208	74.584	74.252	74.262

*** p < 0,001; ** p < 0,01; * p < 0,05



@ecuadorencifras



@ecuadorencifras



@InecEcuador



t.me/equadorencifras



INEC/Ecuador



INECEcuador

Administración Central (Quito)
Juan Larrea N15-36 y José Riofrio,
Teléfonos: (02) 2544 326 - 2544 561 Fax: (02) 2509 836
Código postal: 170410
correo-e: inec@inec.gob.ec

www.ecuadorencifras.gob.ec