Distribución Beta para modelar proporciones en áreas pequeñas

Luz Karime Bernal Muñoz
Estadística

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
2017

Distribución Beta para modelar proporciones en áreas pequeñas

Luz Karime Bernal Muñoz
Estadística

Trabajo de grado presentado para optar al título de
Maestría en Ciencias Bioestadística

Director
Luz Mery Gonzáles García
Doctor en Estadística

Universidad Nacional de Colombia
Facultad de Ciencias
Departamento de Estadística
Bogotá, D.C.
2017

Índice general

Índice de tablas

Índice de figuras

Introducción

La necesidad de obtener información confiable y detallada sobre características de subpoblaciones para las que el diseño muestral no fue planeado no es reciente, su importancia radica en que contar con información con mayores niveles de desagregación orientarían la toma de decisiones e implementación de programas con mayor eficiencia ya sea en el sector público o privado. El problema de estimación en áreas pequeñas - SAE, por sus siglas en inglés (*Small Area Estimation*), consiste en que los estimadores directos, propuestos por Chochran (1977) en la teoría clásica de muestreo, carecen de tamaños de muestra que permitan producir resultados con la precisión adecuada. Una posible estrategia para afrontar el problema es considerar las subpoblaciones desde la planeación del diseño muestral, aumentando así el tamaño de muestra y a su vez los costos de la investigación. Estas implicaciones hacen intratable la estrategia desde el punto de vista práctico dando como resultado el desarrollo de diversos planteamientos teóricos de estimación, tratados en la actualidad, principalmente mediante el modelamiento estadístico.

El modelo Fay-Herriot es posiblemente el modelo más famoso de las técnicas SAE basada en modelos, Fay y Herriot (1979) estimaron el ingreso per cápita en áreas pequeñas de Estados Unidos a partir de un modelo lineal mixto. En los años siguientes se desarrollaron tanto aplicaciones como generalizaciones al modelo Fay-Herriot. Battese *et al.* (1988), utilizaron modelos de efectos mixtos para predecir la cantidad de maíz y soya en 12 condados de Iowa, Estados Unidos. Pfeffermann (1997) relacionó los modelos propuestos por MacGibbon y Tomberlin en 1989; Farrell, MacGibbon y Tomberlin en 1997. Por otra parte, Ghos y Rao (1994) y Rao (1999), motivados por la creciente demanda de estadísticas confiables en áreas pequeñas, presentaron una revisión de los métodos de estimación. Año seguido, Rao (2000) recopiló los artículos de simposios y talleres [1]. Esta monografía la presentó posteriormente en el Seminario Internacional de Estadística de EUSTAT.

En paralelo, desde la teoría de modelos lineales generalizados mixtos, Jiang y Lahiri (1998) incluyeron variables respuesta que siguen distribuciones Binomial y Poisson. Pfeffermann (2002) realizó una revisión crítica de los principales avances de los métodos de estimación en áreas pequeñas con especial énfasis en la predicción de las áreas sin información, así como de los nuevos modelos aplicados a medidas discretas y modelos

[1] National Institute on Drug Abuse, Priceton Conference (National Institute on Drug Abuse, 1979); International Symposium on Small Area Statistics, Ottawa, 1986; International Scientific Conference on Small Area Statistics and Survey Designs, Varsovia, 1992; International association of Survey Statisticians Satellite Conference on Small Area Estimation, Riga, 1999

de series temporales. Nandram y Erhardt (2004), usaron el algoritmo SIR (Sequential Importance Resampling) en lugar de simulación MCMC (Markov Chain Monte Carlo) para ajustar modelos de Poisson y regresión lógistica. Ranalli & Vicarelli (2010) aplicaron modelos lineales mixtos a nivel de unidad y modelos logísticos mixtos para variables respuesta binaria e información auxiliar completa.

Así mismo, con el creciente auge de métodos de estimación desde el enfoque Bayesiano, se encuentran aplicaciones como la desarrollada por Fadila *et al.* (2015) quienes estimaron la tasa de analfabetismo en Indonesia, mediante un modelo bayesiano jerárquico a nivel de unidad. En el mismo año, Satriya *et al.* adaptaron el modelo Fay-Herriot para estimar el ingreso per-capita en distritos de Indonesia, introduciendo a prioris para las varianzas muestrales estimadas. Sugasawa (2017) propone un estimador Bayesiano Empírico robusto para reducir la sensibilidad de los estimadores Fay-Herriot, ante la presencia de outliers. Boubeta M. (2017) desarrolló su tesis doctoral entorno al modelo de Poisson mixto bajo estimación clásica.

En este punto, se ha evidenciado que las técnicas de estimación en áreas pequeñas han tenido el principal desarrollo desde el enfoque clásico para parámetros que pueden ser modelados a través de distribuciones Normal, Binomial y Poisson, con algunas propuestas recientes de estimación desde el enfoque Bayesiano. No obstante, una práctica común en todas las investigaciones, es considerar indicadores asociados a tasas, cocientes o proporciones que tienen dominio en el intervalo (0,1) y que han sido tratados en SAE mediante modelos a nivel de unidad y regresión logística. Un caso más realista, se presenta cuando se dispone de información agregada por consideraciones de reserva estadística sobre las unidades, orientando la necesidad de implementar modelos a nivel de área, como sucede con el modelo Fay-Herriot pero que, dada la naturaleza de la variable respuesta, resulta inadecuado modelarlo como una variable normal por diversas razones. Pérez R.*et al.* (2006), señalan que el rango en el que está definida una proporción difiere del rango de la distribución normal luego, la función de esperanza condicional no es lineal, y por lo tanto no existe una relación lineal entre la media de la variable de interés y las covariables. Por otro lado, la distribución de probabilidad de proporciones es, en general, asimétrica y en tal caso las pruebas de hipótesis bajo este supuesto distribucional pueden ser erróneas. Intentar resolver estas deficiencias bajo transformaciones como la logit o la logarítmica inversa logran linealizar la relación entre la variable respuesta y los predictores, pero no logran estabilizar la varianza.

Una distribución que permite modelar variables aleatorias con dominio en el intervalo (0,1) es la distribución Beta, propuesta por Cepeda (2001) y Cepeda & Gamerman (2005) desde el enfoque Bayesiano, mientras que Ferrari y Cribari-Neto (2004) la abordan desde el enfoque de estimación clásica. Con la naciente línea de investigación, el interés por estructuras más complejas, como modelos no lineales doblemente generalizados son estudiados por Smithson y Verkuilen (2006), Simas *et al.* (2010) y Cepeda-Cuervo & Achcar (2010). Modelos con supuestos de correlación espacial son presentados por Cepeda y Garrido *et al.* (2011) y Cepeda & Nuñez-Anton (2011). Cepeda *et al.* (2014) tratan los modelos lineales generalizados con efectos aleatorios, Gutierrez (2014) trabaja los modelos no lineales generalizados con efectos aleatorios y Tejedor (2014) desarrolla el modelamiento conjunto de media y varianza, en lugar de la dispersión, en modelos mixtos con respuesta Beta.

El presente trabajo relaciona las propuestas de modelamiento beta para la estimación de proporciones en áreas pequeñas, no sólo considerando la relevancia que tienen estos parámetros en la línea de investigación de muestreo, sino los nuevos desarrollos en torno a estimación Bayesiana que vincula las dos temáticas. Por lo anterior, se considera adecuado dividir el marco teórico en dos capítulos el primero dedicado a las técnicas de SAE y el segundo, al modelamiento de variables que siguen una distribución beta. Específicamente, el capítulo 1, inicia con una breve discusión sobre los alcances y limitaciones de las diferentes técnicas encontradas en la literatura para el tratamiento de áreas pequeñas, para sustentar el uso de la estimación basada en modelos, posteriormente se hace énfasis en los modelos a nivel de área, así como de los respectivos métodos de estimación.

En relación al marco teórico de modelamiento en la distribución beta, en el capítulo 2, se presenta una breve discusión sobre la pertinencia metodológica, seguida de los conceptos propios de la técnica, así como de los métodos de estimación. En esta línea, se diferencia la estimación mediante el enfoque clásico (Máxima Verosimilitud Beta – MVB) y el enfoque bayesiano (Bayesiano Beta –BB). Adicionalmente, considerando que la estimación en áreas pequeñas basada en modelos es desarrollada a partir de modelos de efectos mixtos, para el caso de la estimación en la distribución beta, se presentan dos alternativas de tratamiento, uno, mediante estimación clásica de Modelos de Mixturas Finitas (Mixtura Finita Beta – MFB) y dos, mediante estimación bayesiana en el contexto de Modelos Mixtos Doblemente Generalizados (Bayesiano Mixto Beta –BMB). El capítulo finaliza con la descripción de los criterios de evaluación de los modelos.

Ahora bien, dado el soporte teórico, en los siguiente dos capítulos se aborda la implementación en datos reales con base en la información de la Encuesta Multiproposito de Bogota 2011. Para dar un panorama completo del problema de investigación, la naturaleza de las variables, los alcances y limitaciones de las mismas, se desarrolla el capítulo 3 con fin de contextualizar las necesidades de información que permitan focalizar adecuadamente las decisiones en materia de política pública, para un caso específico de interés como es, la desigualdad medida desde el enfoque de capacidades. El capítulo 4 está destinado a la presentación de los resultados, en el entendido que, obedecen al objetivo principal del documento, es decir, a la propuesta de estimación de proporciones en áreas pequeñas mediante la distribuición beta. El capítulo se encuentra desarrollado en cinco secciones, cuatro de ellas con base en los planateamientos teóricos descritos en el capítulo 2; específicamente, se realiza el modelamiento de las proporciones mediante cuatro alternativas de estimación diferenciadas por el enfoque clásico y bayesiano a través de Modelamiento beta clásico (MVB) y modelamiento beta bayesiano (BB), respectivamente. En relación al tratamiento de heterogeneidad debida a la presencia de grupos que pueden o no, estar asociados a las áreas mayores, se utiliza el modelamiento clásico en mixturas finitas beta (MFB) y modelamiento bayesiano mixto beta (BMB). Para la validación de supuestos se calculan los residuales ponderados por la matriz Hat en modelo MFB y BMB, en razón a que los paquetes usados para la estimación en cada caso, no proveían al momento de la implementación, tales resultados; por su parte para el modelo BB, a pesar de que el paquete usado para la estimación, suministra el cálculo de los residuales, se detectó un error en la programación que fué notificada a los autores, de tal manera que para el desarrollo de este documento, se calcularon de forma externa al paquete. Adicional a este aporte, se realizó el cálculo de bandas simuladas para la verificación del supuesto de normalidad en los residuales, ya que no hace parte de ninguno

de los paquetes y software utilizados al momento de la implementación. Para finalizar el capítulo, en la quinta sección, se resumen los hallazgos y se comparan los resultados mediante remuestreo bootstrap, sugerido en la literatura de SAE como una metodología para calcular los errores estándar asociados a los estimadores para cada área pequeña.

En el apéndice A, se hace un recorrido por los algoritmos de estimación, iniciando con Esperanza-Maximización (EM) usado para el modelamiento de mixturas finitas; para el modelamiento Bayesiano se presenta la teoría de estimación Monte Carlo vía Cadenas de Markov (MCMC), los algoritmos Metropolis–Hasting, Muestreo de Gibbs y su combinación, así mismo, se presentan los métodos de análisis de convergencia; el apéndice concluye con la descripción del método de remuestreo Bootstrap. El apéndice B está destinado a la presentación de resultados propios de la aplicación, entre los cuales están la metodología estadística y operativa de la fuente de información, así como los resultados de todos los modelos ajustados bajo las cuatro técnicas. El documento finaliza con las conclusiones, sugerencias sobre trabajos futuros y bibliografía.

CAPÍTULO 1

Estimación en áreas pequeñas

Según Rao (2003, pp.1), "Un área es considerada pequeña si la muestra para el dominio específico, no es lo suficientemente grande como para soportar estimaciones directas con precisión adecuada". El término es utilizado para referirse a un área geográfica o a un dominio pequeño, es decir, un grupo específico sobre el que se requiere estimar una característica poblacional, pero para el cual, el diseño muestral no fue planeado. El interés creciente por contar con información estadística confiable lo más exhausta posible, ha propiciado desarrollos teóricos diferenciados claramente en tres grandes técnicas de estimación: estimadores directos, indirectos y basados en modelos.

Los estimadores **directos** están propuestos en la teoría clásica de muestreo, caracterizada por evaluar las propiedades de sesgo y varianza de los estimadores basándose en el diseño muestral. En esta línea, existe la posibilidad de incorporar modelos dentro del proceso de estimación buscando aumentar la eficiencia de los estimadores a partir de información auxiliar relacionada con la variable de interés, no obstante, su inclusión no modifica el proceso de evaluación de propiedades, es decir, permanecen basadas en el diseño muestral y en tal sentido son estimadores directos conocidos en la literatura como estimadores asistidos por modelos. Debido a que los estimadores directos usan la información únicamente de las unidades seleccionadas en el dominio, la estimación en áreas pequeñas es restringida, puesto que presentan errores estándar inaceptablemente altos; incluso, puede no haber sido seleccionada ninguna unidad muestral en alguna de las áreas pequeñas. Adicionalmente, Ghosh y Rao (1994), señalan que los resultados obtenidos en las estrategias de estimación asistidas por modelos son más susceptibles a inconsistencias debido a errores de especificación del modelo conforme el tamaño del dominio se incrementa.

Los estimadores **indirectos** usan información auxiliar de registros administrativos o de las áreas mayores. En el primer caso se ubican las técnicas demográficas que dieron origen a la estimación de áreas pequeñas, denominadas por Purcell y Kish (1980) como "Técnicas con Consideraciones Sintomáticas" haciendo referencia al uso de datos obtenidos únicamente de registros administrativos actualizados como nacimientos, defunciones y migraciones, así como de información censal, sin involucrar muestreo. El inconveniente primordial de la técnica radica en que los censos de población generalmente están su-

jetos a problemas de omisiones, duplicidad, así como a la temporalidad de un censo a otro.

En el segundo caso, en donde se utiliza para cada área pequeña la información disponible del área mayor bajo el supuesto de homogeneidad, se encuentra el estimador *sintético*, propuesto por Gonzales (1973); el estimador sintético incrementa notablemente su precisión, sin embargo, presenta la desventaja de ser un estimador sesgado, situación que logra reducirse –sujeto al incremento de varianza– mediante el estimador *compuesto* propuesto por Nichol (1977) que corresponde a una suma ponderada entre un estimador directo y el estimador sintético, además del incremento en la varianza, una dificultad adicional se presenta en la selección de los pesos involucrados en el estimador.

Finalmente, gracias en parte a los avances tecnológicos, los métodos de estimación en áreas pequeñas tienen su más reciente desarrollo en las **técnicas basadas en modelos** que han recibido especial atención porque de acuerdo con Rao (2003, pp. 75), presentan las siguientes ventajas:

1. Puede realizarse diagnóstico del modelo, incluyendo análisis de residuales, selección de las variables auxiliares y detección de datos influyentes.
2. A diferencia de los estimadores sintéticos y los combinados, permiten obtener medidas estables de variabilidad para cada área pequeña.
3. Pueden ajustarse modelos lineales mixtos, modelos no lineales y modelos lineales genealizados con efectos aleatorios para cada área.
4. Se pueden incorporar complejas estructuras de dependencia espacial y temporal.

En este contexto, la efectividad de la estimación depende de la calidad de la información disponible que permita un buen ajuste del modelo. Para definir la estrategia de estimación basada en modelos es necesario determinar dos aspectos, el primero asociado a la disponibilidad de información y el segundo al método de estimación. En términos de disponibilidad de información, en áreas pequeñas se diferencian dos tipos de modelos: uno, cuando se puede acceder a la información de cada elemento seleccionado en la muestra y la predicción se hace a ese nivel, denominados **modelos a nivel de unidad** y dos, cuando la información se encuentra únicamente disponible en forma agregada y en tal caso se trata de **modelos a nivel de área**.

Referente al método de estimación, se cuenta con tres métodos: El primero, se basa en la teoría de modelos lineales mixtos bajo el supuesto de normalidad que da origen al Mejor Predictor Lineal Empírico Insesgado **EBLUP** por sus siglas en inglés (Empirical Best Linear Unbiased Prediction). Los siguientes dos métodos son aplicables para variables en las que no es adecuado asumir tal distribución y sustentan su construcción en las técnicas de modelamiento Bayesiano Empírico y Bayesiano Jerárquico dando lugar a los estimadores **EB** y **HB**, respectivamente.

Antes de abordar el marco conceptual de los métodos de estimación basada en modelos, se hace necesario precisar que, el objetivo de la estimación en áreas pequeñas es obtener estimadores puntuales y sus respectivos errores estándar para un parámetro poblacional

de interés, en el contexto tradicional de muestreo, parámetro que, no debe confundirse con los denominados parámetros del modelo cuya estimación no solo asisten la estimación puntual del estimador de interés en torno a SAE, sino que se involucran directamente en el cálculo de los respectivos errores estandándar.

Bajo este contexto, se presenta a continuación el marco conceptual de los modelos a nivel de área incorporando resultados bajo el supuesto de normalidad en los casos que se requiera, así como los métodos de estimación para variables que no siguen esta distribución, con fundamento en Rao (2003).

1.1. Modelos a nivel de área

La idea básica de los métodos de estimación en áreas pequeñas basada en modelos, plantea que el estimador directo $\widehat{\theta}_d$ para un área específica d, difiere del verdadero valor del parámetro de interés en dicha área θ_d, tan solo por el error muestral e_d y que, adicionalmente, el parámetro sigue un modelo poblacional. Formalmente, siguiendo a Rao (2003, pp. 75), considérese la población objetivo conformada por M áreas de las cuales m son seleccionadas y se requiere estimar los parámetros θ_d, $d = 1, \cdots, m$. Además, los estimadores directos $\widehat{\theta}_d$ no cuentan con la precisión deseada y su relación con el parámetro está dada por:

$$\widehat{\theta}_d = \theta_d + e_d, \tag{1.1}$$

en donde, $\widehat{\theta}_d$ es una medida agregada, es decir, tal que $\widehat{\theta}_d = h(\widehat{\overline{Y}}_d)$ con $h(\cdot)$ una función para la cual $\theta_d = h(\overline{Y}_d)$ y los errores e_d son independientes, centrados en cero y con varianza conocida ψ_d, $d = 1, \ldots, m$. Esta relación determina la primera etapa o componente del modelo que, mediante la especificación de una distribución de probabilidad f_d, puede expresarse como:

$$\widehat{\theta}_d | \theta_d, \psi_d \sim f_d(\theta_d, \psi_d). \tag{1.2}$$

La segunda etapa o componente está dada por la relación entre el parámetro de la d-ésima área, θ_d, y la información auxiliar al mismo nivel, a través de un modelo explícito:

$$\theta_d = \boldsymbol{x}_d^\intercal \boldsymbol{\beta} + \boldsymbol{z}_d^\intercal \boldsymbol{u}_d. \tag{1.3}$$

Con $\boldsymbol{\beta}$ el p–vector de parámetros de regresión $\beta_1, \ldots, \beta_p$, y, $\boldsymbol{u}_d$ independientes, con media cero y varianza común desconocida σ_u^2 y las covariables a nivel de área están representadas por $\boldsymbol{x}_d$ y $\boldsymbol{z}_d$. La segunda etapa formulada mediante distribución de probabilidad, g_d, es:

$$\theta_d | u_d \sim g_d(\boldsymbol{x}_d^\intercal \boldsymbol{\beta} + \boldsymbol{z}_d^\intercal \boldsymbol{u}_d, \sigma_u^2). \tag{1.4}$$

Combinando los componentes de las ecuaciones (1.1) y (1.3), se obtiene el modelo básico a nivel de área, que corresponde a un modelo lineal mixto:

$$\widehat{\theta}_d = \boldsymbol{x}_d^\intercal \boldsymbol{\beta} + \boldsymbol{z}_d^\intercal \boldsymbol{u}_d + e_d, \tag{1.5}$$

en donde $\boldsymbol{u}_d$ son los efectos aleatorios que modelan la variabilidad de las áreas pequeñas que no es recogida por las covariables, a través del parámetro σ_u^2. El modelo así definido involucra variables aleatorias basadas en el diseño, e_d, y variables aleatorias basadas en

el modelo $\boldsymbol{u}_d$, asumiendo que son independientes entre sí. Las características de estas variables aleatorias, de acuerdo con la notación tradicional de muestreo son:

$$\begin{aligned} E_p(e_d|\theta_d) &= 0 & V_p(e_d|\theta_d) &= \psi_d \\ E_m(\boldsymbol{u}_d) &= 0 & V_m(\boldsymbol{u}_d) &= \sigma_u^2 (>= 0), \end{aligned}$$

donde E_p y V_p denotan el valor esperado y la varianza basados en el diseño muestral $p(\cdot)$, por lo tanto, ψ_d, la varianza muestral, se asume conocida. Debido a que este supuesto restringe muchas aplicaciones, generalmente se debilita mediante el uso de la varianza muestral estimada $\widehat{\psi}_d$.

El modelo de la ecuación (1.5) puede generalizarse a partir de la notación matricial así:

$$\widehat{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}. \tag{1.6}$$

Siendo $\widehat{\theta}$ el vector de estimadores puntuales de las m x 1 áreas muestreadas, $\mathbf{X}$ y $\mathbf{Z}$ son matrices de rango completo conocidas de orden m x p y m x h, respectivamente, $\mathbf{u}$ y $\mathbf{e}$ son independientes con media $\mathbf{0}$ y matrices de covarianza $\mathbf{G}$ y $\mathbf{R}$ respectivamente.

Mediante las ecuaciones normales de Henderson (1963), el mejor predictor lineal insesgado **BLUP** para el vector de parámetros del modelo $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{u}})^\intercal$ está dado por:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\intercal \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\intercal \mathbf{V}^{-1} \widehat{\theta}$$

y,

$$\widehat{\boldsymbol{u}} = \mathbf{G}\mathbf{Z}^\intercal \mathbf{V}^{-1} (\widehat{\theta} - \mathbf{X}\widehat{\boldsymbol{\beta}}).$$

Donde, $\mathbf{V} = \mathbf{V}(\widehat{\theta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^\intercal + \mathbf{R}$. En la práctica las componentes de las matrices de covarianza son desconocidas y, por lo tanto, al ser reemplazadas por estimaciones dan lugar al denominado mejor predictor lineal empírico insesgado, **EBLUP**.

Un caso particular del modelo definido en la ecuación (1.5), es propuesto por Fay y Herriot (1979), conocido ampliamente en la literatura de estimación en áreas pequeñas como el modelo Fay–Herriot:

$$\widehat{\theta}_d = \boldsymbol{\beta}_0 + X_d \boldsymbol{\beta}_1 + u_d + e_d. \tag{1.7}$$

Adicionalmente asumen f_d y g_d normales dada la especificación distribucional de los errores e_d y u_d, quedando así completamente definido el modelo en etapas de acuerdo con las ecuaciones (1.2) y (1.4). Bajo estas consideraciones el estimador Fay–Herriot (**HF**) de áreas pequeñas es:

$$\widetilde{\theta}_d^{HF} = \gamma_d \widehat{\theta}_d + (1 - \gamma_d) \boldsymbol{x}_d^\intercal \widetilde{\boldsymbol{\beta}} \qquad \text{donde,} \qquad \gamma_d = \frac{\sigma_u^2}{(\phi_d + \sigma_u^2)}. \tag{1.8}$$

Para estimar $\boldsymbol{\beta}$, u_d y σ_u^2, los autores utilizan el enfoque clásico de estimación vía máxima verosimilitud restringida – REML. Adicionalmente, estiman la varianza muestral ϕ_d a partir de la información muestral como $\widehat{\phi}_d = \dfrac{s_d^2}{n_d}$, $d = 1, \ldots, m$.

La expresión del estimador Fay–Herriot corresponde a una suma ponderada de un estimador directo y un estimador sintético, es decir, tiene la estructura del estimador compuesto mencionado dentro de las técnicas de estimación indirecta, con la condición de facilitar la especificación de los pesos, dados los valores estimados de las respectivas componentes.

El planteamiento de estimación a partir de etapas, es decir, la especificación de la relación del estimador muestral y parámetro de interés, seguida de la especificación del modelamiento para el parámetro, ecuaciones (1.2) y (1.4) respectivamente, facilita la introducción a los métodos de estimación bayesiana en SAE, cuando los supuestos de normalidad no son sustentables, o incluso, como lo proponen You y Chapman (2006), para modificar el modelo Fay–Herriot con el propósito de modelar la varianza muestral ϕ para corregir los inconvenientes que presenta el uso de la varianza muestral estimada $\widehat{\phi}_d$. Trabajos posteriores discuten diferentes a prioris para estimación basada en modelos con enfoque bayesiano, Sugasawa, *et. al*, (2015).

1.2. Método de estimación Empírica de Bayes

En la sección anterior se describió de forma general las implicaciones que conlleva la disponibilidad de información para el planteamiento del modelo, la extensión matricial a diversas estructuras de la matriz de covarianza, así como la identificación de las componentes asociadas al modelo Fay–Herriot, que basado en el supuesto de normalidad permite obtener de manera explícita el estimador EBLUP y que, adicionalmente, por las restricciones sobre el conocimiento de las varianzas muestrales, hace uso de la información de la muestra para su estimación, conduciendo así al estimador EBLUP. Aunque el supuesto de normalidad no es requerido para la estimación puntual, si lo es para calcular las medidas de precisión de los estimadores y la consecuente inferencia estadística.

Cuando las variables dependientes no son continuas, no es adecuado utilizar el estimador EBLUP, por ejemplo, si la información se encuentra a nivel de unidad y la característica de interés corresponde a ausencia/presencia de una condición o incluso, si se dispone de información a nivel de área en donde la presencia/ausencia de una condición corresponde a conteos, los modelos lineales mixtos no son recomendados. Los casos mencionados son ampliamente abordados en la literatura de SAE haciendo uso de la teoría de Modelos Lineales Generalizados– MLG usando modelos Binomial para información a nivel de unidad y de Poisson para información a nivel de área, así como del uso de los métodos de estimación EB y HB. Para el caso específico y más realista de contar con información a nivel de área, otro tipo de parámetros de interés pueden ser continuos pero restringidos a un intervalo específico, tal como sucede con las proporciones, objeto primordial de este documento, que pueden ser modeladas a partir de regresión Beta que tiene la particularidad de no pertenecer a la familia exponencial, requisito indispensable para modelarse mediante MLG. En esta sección se abordan los conceptos básicos del método de estimación Empírica de Bayes y en la siguiente, se trata el método de estimación jerárquico de Bayes, asumiendo la pertenencia de la variable aleatoria a la familia exponencial, en el capítulo 2, se analiza la teoría del modelamiento de respuestas Beta.

Rao (2003, pp. 179) resume el procedimiento para el desarrollo del método Empírico Bayesiano (EB) en los siguientes pasos:

1. Obtener la densidad a posteriori $f(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \boldsymbol{\lambda})$ de los parámetros de interés $\boldsymbol{\theta}$ de las áreas pequeñas, dados los datos $\widehat{\boldsymbol{\theta}}$; usando la densidad condicional $f(\widehat{\boldsymbol{\theta}}|\boldsymbol{\theta}, \boldsymbol{\lambda_1})$ y la densidad $f(\boldsymbol{\theta}|\boldsymbol{\lambda_2})$, donde $\boldsymbol{\lambda}$ denota el vector de parámetros del modelo.
2. Estimar los parámetros del modelo $\boldsymbol{\lambda}$, a partir de la densidad marginal de $\widehat{\boldsymbol{\theta}}$, $f(\widehat{\boldsymbol{\theta}}|\boldsymbol{\lambda})$.
3. Usar la densidad a posteriori estimada, $f(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\lambda}})$ para hacer inferencias acerca de $\boldsymbol{\theta}$, donde $\widetilde{\boldsymbol{\lambda}}$ es el estimador de $\boldsymbol{\lambda}$.

Dentro del contexto de este documento, en la sección 1.1 se vinculó para cada área $d = 1, \ldots, m$, el estimador directo $\widehat{\theta}_d$ y el parámetro θ_d con la media estimada $\widehat{\overline{Y}}_d$ y la media poblacional $\overline{Y}_d$ de la variable de interés, respectivamente, a través de una función $h(\cdot)$. Sin pérdida de generalidad, para introducir la notación del enfoque bayesiano en áreas pequeñas, considérese $\overline{Y}_d = \mu_d$ y por lo tanto $\theta_d = h(\mu_d)$, con lo cual la ecuación (1.2) para MLG se expresa como:

$$\widehat{\theta}_d = h(\mu_d) + e_d.$$

En este caso, $h(\cdot)$ corresponde a la función de enlace, que linealiza la relación del valor esperado a posteriori de θ_d con las covariables y los efectos aleatorios. Además, los errores muestrales e_d son tales que, $E_p(e_d|\theta_d) = 0$ y $Var_p(e_d|\theta_d) \approx \left[h^{'}(\theta_d)\right]^2 V_p(\widehat{\theta}_d)$.

Para llevar a cabo una estimación empírica de Bayes, se asume que el vector de estimadores directos $\widehat{\theta}_d$ sigue una distribución $f_d(\widehat{\theta}_d|\theta_d)$, así mismo, que los parámetros θ_d siguen independientemente una distribución $g_d(\theta_d, \phi)$ donde ϕ es un vector de parámetros desconocido. Entonces, la distribución a posteriori de θ_d para $d = 1, \ldots, m$ está dada por:

$$\pi(\theta_d|\widehat{\theta}_d; \phi) = \frac{f_d(\widehat{\theta}_d|\theta_d) g_d(\theta_d, \phi)}{\int f_d(\widehat{\theta}_d|\theta_d) g_d(\theta_d, \phi) d\theta_d}.$$

El estimador de Bayes $\widetilde{\theta}_d$ de θ_d bajo función de pérdida del error cuadrático medio condicional de θ_d dado $\widehat{\theta}_d$ es:

$$\widetilde{\theta}_d \equiv E\left[\theta_d|\widehat{\theta}_d; \phi\right] = \frac{\int \theta_d f_d(\widehat{\theta}_d|\theta_d) g_d(\theta_d, \phi) d\theta_d}{\int f_d(\widehat{\theta}_d|\theta_d) g_d(\theta_d, \phi) d\theta_d}.$$

Debido a que el estimador de Bayes $\widehat{\theta}_d$ depende del parámetro ϕ, desconocido en el modelo, éste puede ser estimado a partir de la distribución marginal de los estimadores directos $\widehat{\theta}_d$ de acuerdo a:

$$L(\phi) = \prod_{d=1}^{m} \int f_d(\widehat{\theta}_d|\theta_d) g_d(\theta_d, \phi) d\theta_d.$$

Con la función de distribución marginal de los estimadores directos queda completamente definida la función de verosimilitud a maximizar, así, el estimador Empírico Bayesiano

EB obtenido al estimar ϕ mediante el enfoque clásico, está dado por:

$$\widetilde{\theta}_d^{EB} = E\left[\theta_d|\widehat{\theta}_d;\widehat{\phi}\right].$$

Bajo los supuestos distribucionales para la estructura del modelo Fay–Herriot (1.7), el estimador Bayesiano es:

$$\widetilde{\theta}_d^{B}(\boldsymbol{\beta},\sigma_u^2) = \gamma_d\widehat{\theta}_d + (1-\gamma_d)\boldsymbol{x}_d^{\intercal}\widehat{\boldsymbol{\beta}}. \tag{1.9}$$

Al reemplazar en $\widetilde{\theta}_d^{B}$ con los estimadores de los parámetros del modelo mediante REML, se obtiene el estimador Empírico Bayesiano-EB, $\widetilde{\theta}_d^{EB}$, que coincide con el estimador EBLUP de la ecuación (1.8), esto es, $\widetilde{\theta}_d^{HF} = \widetilde{\theta}_d^{EB}$.

1.3. Método de estimación Jerárquica de Bayes

Como se ha mencionado, un aspecto importante en el planteamiento de los modelos a nivel de área para SAE, es el supuesto de varianzas muestrales conocidas para los errores e_d que asocian el estimador directo y el parámetro en la d–ésima área. Al ser un supuesto poco factible en la práctica, se acude a su estimación clásica o mediante procedimientos de remuestreo como bootstrap (Pérez A., 2008). En el caso de estimación Jerárquica de Bayes este problema no existe, puesto que al ser un método completamente bayesiano, es posible a partir de la especificación de distribuciones a priori y el uso del teorema de Bayes, determinar las distribuciones a posteriori y como consecuencia medidas de localización y escala de los parámetros del modelo. Más aún, cuando las medidas no tienen una expresión analítica, la teoría provee algoritmos como el muestreador de Gibbs o el Metropolis–Hastings, para aproximarlas.

De esta manera, en el método de estimación Jerárquica de Bayes, tanto los parámetros θ_d del área, como los parámetros del modelo $\boldsymbol{\phi = (\beta, \sigma_u^2)}$ se consideran aleatorios y son estimados desde el enfoque Bayesiano especificando distribuciones a priori para realizar inferencia con base en la distribución marginal a posteriori. En particular θ_d es estimada por la media posteriori $E(\theta_d|\widehat{\theta}_d)$ y la variabilidad es estimada por la varianza posteriori $V(\theta_d|\widehat{\theta}_d)$.

Considerando los supuestos distribucionales del modelo Fay–Herriot de la ecuación (1.7), en el planteamiento completamente bayesiano, se tiene, tal como en el método Empírico de Bayes:

i. $\widehat{\theta}_d|\theta_d,\beta,\sigma_u^2 \sim N(\theta_d,\psi_d)$,

ii. $\theta_d|\beta,\sigma_u^2 \sim N(\mathbf{x}_i^{\intercal}\boldsymbol{\beta},\sigma_u^2)$.

Sin embargo, en este caso adicionalmente, se requiere la especificación de las distribuciones a priori para (β, σ_u^2) consideradas variables aleatorias y pueden ser tales que:

iii. $f(\beta) \propto 1$ si σ_u^2 es conocida o,
$f(\beta,\sigma_u^2) = f(\beta)f(\sigma_u^2) \propto f(\sigma_u^2)$,

donde $i = 1, \ldots, m$ y $f(\sigma_u^2)$ es la apriori sobre σ_u^2.

Rao (2003, pp.237), muestra que para el caso de σ_u^2 conocida y apriori de $f(\beta)$ no informativa o uniforme, el estimador $\widetilde{\theta}_d^{HB}$ es equivalente al estimador BLUP del modelo Fay–Farriot (1.8), esto es, $\widetilde{\theta}_d^{HB} = E(\theta_d|\widehat{\theta}_d, \sigma_u^2) = \widetilde{\theta}_d^{HF}$.

1.4. Error Cuadrático Medio

La precisión de los estimadores de áreas pequeñas basados en modelos, es medida a través del error cuadrático medio predicho (MSPE, por sus siglas en inglés), debido a que es el criterio de minización para la estimación de los **BLUP**. Para este caso, Rao (2003, p.98) demuestra que el error cuadrático medio del estimador puede descomponerse en dos partes, una, debida a la variabilidad del estimador y otra, debida a la variabilidad de los coeficientes. De acuerdo con la expresión del modelo lineal mixto de la ecuación (1.6), el estimador BLUP, $t(\delta, \widehat{\theta})$, es obtenido a partir de la combinación lineal $\mu = \mathbf{L}\beta + \mathbf{Mu}$ esto es:

$$\begin{aligned} t(\delta, \widehat{\theta}) &= \mathbf{L}\widehat{\beta} + \mathbf{M}\widehat{\mathbf{u}} \\ &= t^*(\delta, \widehat{\theta}, \beta) + \mathbf{d}^\intercal(\widehat{\beta} - \beta), \end{aligned}$$

donde δ son parámetros de las matrices de varianzas y covarianzas $\mathbf{R}$ y $\mathbf{G}$ de los errores $\mathbf{e}$ y $\mathbf{u}$ respectivamente, cuyo uso en el proceso de estimación se describe en la ecuación (1.1). Por su parte, $t^*(\delta, \widehat{\theta}, \beta)$ es el BLUP asumiendo que β es conocido, así:

$$t^*(\delta, \widehat{\theta}, \beta) = \mathbf{l}^\intercal\beta + \mathbf{b}^\intercal(\widehat{\theta} - \mathbf{X}\beta),$$

con,

$$\mathbf{b}^\intercal = m^\intercal\mathbf{G}\mathbf{Z}^\intercal\mathbf{V}^{-\mathbf{1}},$$

y,

$$\mathbf{d}^\intercal = \mathbf{l}^\intercal - \mathbf{b}^\intercal X.$$

Con lo anterior, el error cuadrático medio del estimador **BLUP** está dado por:

$$ECM[t(\delta, \widehat{\theta})] = ECM[t^*(\delta, \widehat{\theta}, \beta)] + Var[\mathbf{d}^\intercal(\widehat{\beta} - \beta)] = g_1(\delta) + g_2(\delta),$$

para el cual,

$$g_1(\delta) = m^\intercal\mathbf{G} - \mathbf{G}\mathbf{Z}^\intercal\mathbf{V}^{-\mathbf{1}}\mathbf{Z}\mathbf{G}m,$$

y,

$$g_2(\delta) = \mathbf{d}^\intercal\mathbf{X}^\intercal\mathbf{V}^{-\mathbf{1}}\mathbf{X}^{-1}\mathbf{d}.$$

De esta manera, en términos de precisión a cada área pequeña, se asocia el error estándar $EE(\widehat{\theta})$ dado por la raiz cuadrada del error cuadrático medio, $EE(\widehat{\theta}) = \sqrt{ECM(\widehat{\theta})}$. Adicionalmente, con base en ésta medida de precisión se construyen los intervalos de confianza para cada estimador de área pequeña, no obstante, se acuerdo con Chatterjee, Lahiri y Li (2006, p. 1.222), "Las probabilidades de cobertura de este tipo de intervalos pueden converger al valor nominal de $(1 - \alpha)$, pero los intervalos pueden ser ineficientes debido a que presentan problemas de subcobertura o sobrecobertura, dependiendo del valor par-

ticular estimado, MSPE. El error de cobertura de este tipo de intervalos podría exceder el order $O(n^{-1})$, indicado que no es suficientemente preciso para la estimación en áreas pequeñas dados los tamaños muestrales involucrados" . Por su parte, en la metodología Bayesiana sobre la que se basan los estimadores **EB** y **JB**, es posible generar intervalos de credibilidad de tal manera que la probabilidad a posteriori de que el intervalo contenga el parámetro sea $(1 - \alpha)$, sin embargo, un problema adicional surge cuando el parámetro de interés es una proporción, como en el caso de esta aplicación, para el cual "independientemente de los métodos empleados para obtenerlos, la información en general genera comportamientos inesperados" (Cepeda–Cuervo E.*et. al*(2008)).

Bajo estas consideraciones, Chatterjee, Lahiri y Li (2006, p. 1.222) proponen abordar el problema de aproximación a la distribución del predictor mediante Bootstrap (Ver apéndice (A.3)), principalmente en el contexto del estimador EBLUP.

CAPÍTULO 2

Modelamiento de proporciones mediante la distribución Beta

A lo largo del capítulo 1 se ha mencionado que la literatura de SAE basada en modelos, trata ampliamente las técnicas para variables que siguen una distribución normal ya sea en modelos a nivel de unidad o de área, utilizando los métodos EBLUP, EB o HB. Otros desarrollos en MLG permiten el uso de la Binomial cuando se cuenta con información a nivel de unidad o Poisson cuando está a nivel de área, bajo los métodos de estimación EB y HB. Sin embargo, considérese el caso de contar con información a nivel de área y el interés está en modelar proporciones, cocientes o índices, que son indicadores altamente utilizados en encuestas por muestreo, para los cuales la precisión a nivel de área, no permite tomar decisiones. Es este caso, para implementar las técnicas de SAE basada en modelos, podría recurrirse a una práctica antigua de la literatura de modelamiento, que consistía en transformar la variable y asumir normalidad, lo cual no es conceptualmente correcto, porque, por un lado, de acuerdo con Pérez R.*et al.* (2006), las transformaciones no controlan la sobre dispersión natural de este tipo de variables y adicionalmente es necesario modificar valores límites para evitar nulos o indefinidos; en cuanto a la normalidad, el rango de las predicciones puede diferir del soporte (0,1) en donde están definidas las proporciones y las pruebas de hipótesis podrían ser erróneas dado que su distribución de probabilidad es, en general, asimétrica.

Cepeda (2001) propone modelar la media y la dispersión para variables que siguen una distribución Beta desde el enfoque Bayesiano, que permitiría la aplicación de los métodos EB y HB de SAE. En cuanto al enfoque clásico, sobre el que se basa el EBLUP, Ferrari y Cribari-Neto (2004) proponen el modelamiento Beta bajo el supuesto de homocedasticidad. Desde los dos enfoques se han desarrollado extensiones teóricas con el propósito de permitir el análisis de información con estructuras complejas. En este capítulo, se presenta la distribución beta y sus características para posteriormente, detallar la teoría de su modelamiento estadístico.

2.1. Distribución Beta como función de su media y dispersión

De acuerdo con la notación desarrollada sobre SAE, se dice que el estimador directo $\widehat{\theta}_d$, con $d = 1, \cdots, m$ y m áreas muestreadas, sigue una distribución beta con parámetros α_d y λ_d si su función de densidad está dada por:

$$f(\widehat{\theta}_d; \alpha; \lambda) = \frac{\Gamma(\alpha_d + \lambda_d)}{\Gamma(\alpha_d)\Gamma(\lambda_d)} \widehat{\theta}_d^{\alpha_d - 1}(1 - \widehat{\theta}_d)^{\lambda_d - 1}, 0 < \widehat{\theta}_d < 1, \tag{2.1}$$

donde $\alpha_d > 0$, $\lambda_d > 0$ y $\Gamma(\cdot)$ es la función Gamma $\Gamma(z) = \int_0^\infty t^{z-1}e^{-z}dz$. La media y la varianza de $\widehat{\theta}_d$ son respectivamente:

$$E(\widehat{\theta}_d) = \frac{\alpha_d}{(\alpha_d + \lambda_d)}, \qquad V(\widehat{\theta}_d) = \frac{\alpha_d \lambda_d}{(\alpha_d + \lambda_d)^2(\alpha_d + \lambda_d + 1)}.$$

La distribución de la ecuación (2.1) fue reparametrizada por Jorgensen (1997) y, tanto Cepeda (2001) como Ferrari y Cribari-Neto (2004) lo implementaron con el propósito de adecuar el modelamiento en la forma tradicional al valor esperado de la variable $\widehat{\theta}_d$. La reparametrización consiste en transformar la distribución en función de la media haciendo $E(\widehat{\theta}_d) = \mu_d = \frac{\alpha_d}{(\alpha_d + \lambda_d)}$ y $\phi_d = \alpha_d + \lambda_d$, de tal manera que $\widehat{\theta}_d | \mu_d, \phi_d \sim beta(\mu_d\phi_d, (1-\mu_d)\phi_d)$ y por lo tanto la distribución se reescribe como:

$$f(\widehat{\theta}_d; \mu_d, \phi_d) = \frac{\Gamma(\phi_d)}{\Gamma(\mu_d\phi_d)\Gamma((1-\mu_d)\phi_d)} \widehat{\theta}_d^{\mu_d\phi_d - 1}(1 - \widehat{\theta}_d)^{(1-\mu_d)\phi_d - 1}, \quad 0 < \widehat{\theta}_d < 1. \tag{2.2}$$

Las medidas de centralidad y escala generadas son:

$$E(\widehat{\theta}_d) = \mu_d, \qquad V(\widehat{\theta}_d) = \frac{V(\mu_d)}{(1 + \phi_d)} = \frac{\mu_d(1 - \mu_d)}{(1 + \phi_d)}. \tag{2.3}$$

El parámetro ϕ_d se conoce como parámetro de precisión porque conforme éste crece, para un μ_d fijo, la varianza de la variable $\widehat{\theta}_d$ disminuye.

Nótese que para el caso de proporciones no es posible asumir que $\mu_d = \theta_d$, tal como se asumió en el capítulo 1, en donde $E(\widehat{\theta}) = \theta$ como lo establece la ecuación (1.2). Esto se explica porque en teoría de muestreo, los parámetros tratados como proporciones son formulados como razón de totales $R_d = \frac{t_{y_d}}{t_{z_d}}$, donde t_{y_d} corresponde al total de una variable de estudio y, y t_{z_d} corresponde al total de una variable z. En el caso específico de estudio de este documento, t_{y_d} indica el total de elementos que cumplen con la característica de interés y t_{z_d} el total de elementos del dominio, éste último, en la práctica es un tamaño desconocido y no constante, haciendo necesaria la estimación tanto de numerador como del denominador. El sesgo relativo de un estimador de razón es despreciable conforme el tamaño de muestra crece, situación que no necesariamente podría aplicar a la teoría de áreas pequeñas, por lo que futuros trabajos podrían analizar el efecto del sesgo de diseño para la estimación de proporciones en áreas pequeñas basada en modelos.

2.2. Modelamiento de variables que siguen una distribución beta

Desde el enfoque de clásico, Ferrari y Cribari-Neto (2004) propusieron modelar la media de la distribución Beta bajo el supuesto de varianza constante, mediante un modelo lineal generalizado para el parámetro de localización:

$$h_1(\mu_d) = \eta_d = x_d^\intercal \beta,$$

con $h_1(\cdot)$ la función logit, asumiendo ϕ constante y conocida. No obstante, Kosmidis y Firth (2010), demostraron que la inferencia vía Máxima Verosimilitud para el modelo beta propuesto conduce a conclusiones desacertadas, pues tiende a subestimar los errores estándar de los estimadores debido al sesgo del método de estimación. Al respecto, Simas *et al.* (2010) además de obtener expresiones analíticas para la corrección del sesgo de los estimadores máxima verosimilitud, desarrollaron el modelamiento en la distribución beta desde el enfoque clásico, considerando que el parámetro de dispersión no es constante y que por el contrario puede ser modelado de manera similar al parámetro de localicación, mediante la especificación del modelo:

$$h_2(\phi_d) = \xi_d = z_d^\intercal \gamma,$$

donde, β y γ son los coeficientes de regresión de los predictores lineales η_d y ξ_d, estimados vía máxima verosimilitud.

La teoría de Modelos Lineales Generalizados (MLG) permite modelar variables que no siguen una distribución normal, con la condición de que tal variable pertenezca a la familia exponencial uniparamétrica para garantizar entre otras cosas, la existencia de estadísticos suficientes y completos. En el caso de la distribución beta, al no pertenecer a la familia exponencial uniparamétrica, el uso de la teoría de MLG es restringido; No obstante, Cepeda (2001) y Cepeda & Gamerman (2005) a partir de la reparametrización de la distribución beta en función de su media (ecuación (2.2)) y la teoría de la familia exponencial biparamétrica, realizaron una extensión hacia los Modelos Doblemente Generalizados (MDG) desde el enfoque Bayesiano, que adicionalmente facilita el modelamiento del parámetro de dispersión lo cual permite modelar variabilidad adicional que no es capturada a través de la media y que, en el contexto de SAE basada en modelos implica capturar información para la estimación de los errores estándar.

Se dice que una variable aleatoria y_i $(i = 1, \ldots, n)$ sigue una distribución que pertenece a la familia exponencial biparamétrica (Gelfand & Dalal (1990), Dey, Gelfand & Peng (1997)), si su función de densidad se escribe como:

$$f(y|\zeta, \tau) = b(y) exp\{\zeta y + \tau T(y) - \rho(\zeta, \tau)\}, \tag{2.4}$$

donde $b(\cdot)$, $T(\cdot)$ y $\rho(\cdot)$ son funciones reales conocidas y, ζ y τ los parámetros asociados con la media y la varianza de y.

Así, para modelar la media y la dispersión de forma simultánea, se trabaja la estimación de los parámetros de forma intercalada de acuerdo con dos modelos lineales generalizados,

uno que modela la media, μ, a partir de un conjunto de covariables con coeficientes β, y la escala, ϕ, modelada con las mismas u otras covariables y coeficientes γ. Smyth (1989) desarrolló la propuesta de la siguiente manera:

1. Se consideran valores iniciales $\beta^{(0)}$ y $\gamma^{(0)}$ a partir de los cuales se generan muestras bajo distribuciones pertenecientes a la familia exponencial uniparamétrica.

2. Para actualizar el valor del parámetro β, se asume γ fija y se estima mediante la teoría de MLG.

3. Se estima el valor del parámetro γ a partir de una variable instrumental obtenida por la aproximación lineal:

$$\varphi_i = h_2(\phi_i) + \frac{d(h_2(\phi_i))}{d\phi_i}(r_i^D - \phi_i),$$

   donde $h_2(\cdot)$ es la función de enlace para la escala y los residuos $r_i^D = sg(y_i - \widehat{\mu}_i)\sqrt{d_i}$ con d_i es la contribución de cada dato al modelo evaluada con la deviance.
   Así, a partir del valor fijo para β, la actualización del nuevo valor de γ es obtenido mediante MLG de acuerdo a las variables instrumentales ζ_d construidas y sus respectivas covariables.

4. Se repiten los pasos 2 a 3 hasta lograr convergencia.

De lo anterior, se considera que los estimadores directos $\widehat{\theta}_d$, $d = 1, \ldots, m$ corresponden a la realización de variables aleatorias independientes e idénticamente distribuidas de acuerdo a la densidad de la ecuación (2.4) y que puede reparametrizarse como la expresión de la ecuación (2.2), de tal manera que para su modelamiento, se establecen los siguientes MLG para los parámetros:

$$\begin{aligned} h_1(\mu_d) &= \eta_d = x_d^\intercal \beta \\ h_2(\phi_d) &= \xi_d = z_d^\intercal \gamma, \end{aligned} \tag{2.5}$$

Las componentes sistemáticas η_d y ξ_d relacionan las covariables con la media y la dispersión mediante $h_1(\cdot)$ y $h_2(\cdot)$, las funciones de enlace, monótonas, continuas y doblemente diferenciables, generalmente $h_1(\cdot) = logit(\cdot)$ y $h_2(\cdot) = log(\cdot)$, respectivamente. Los coeficientes, β y γ son estimados desde el enfoque clásico, vía máxima verosimilitud restringida y con enfoque bayesiano a través del algoritmo Metropolis–Hastings.

2.3. SAE mediante mixturas finitas Beta

Los efectos aleatorios en SAE, permiten predecir diferentes valores de medias para cada área pequeña y asociar medidas de precisión pues incorporan las variaciones debidas a efectos de las áreas mayores. En el caso de modelos beta, la incorporación de dichos efectos aleatorios es trabajada desde el enfoque completamente bayesiano, algunas propuestas fueron desarrolladas por Cepeda *et al.* (2014), Gutierrez (2014) y Tejedor (2014). Así mismo, un modelo general para trabajar efectos mixtos de acuerdo con Figueroa–Zuñiga *et al.* (2013), podría expresarse como:

$$\begin{aligned} logit(\mu_d) &= \eta_d = x_d^\intercal \beta + z_d^\intercal u \\ log(\phi_d) &= \xi_d = w_d^\intercal \delta + l_d^\intercal \gamma. \end{aligned} \tag{2.6}$$

Sin embargo, un método alternativo para capturar la heterogeneidad debida a observaciones agrupadas, es mediante modelamiento de mixturas finitas, presentando dos ventajas: una, permitir el agrupamiento de observaciones para situaciones en las que variables observadas no están directamente influenciadas por las áreas mayores y dos, para el caso de SAE la ampliación de tamaños de grupos que mejore la predicción no solo por el incremento del tamaño sino por la construcción de grupos más homogéneos, esta estrategia fue tratada por Torkashvand *et al.* (2017) en donde propone agrupar áreas pequeñas con base en la distancia euclidiana para mejorar el error cuadrático medio del EBLUP en modelos lineales mixtos a nivel de área. Para el caso de distribución beta, el agrupamiento puede darse de forma natural, si se asume que se desconoce la pertenencia al área mayor y se aborda la estimación mediante modelamiento de mixturas finitas.

El modelamiento de mixturas finitas (MF) es usado como una estrategia para modelar datos con presencia de heterogeneidad que puede deberse a que los datos provienen de dos o más subpoblaciones denominadas componentes. Pearson (1984), da origen a la teoría de MF utilizando el método de momentos para mixturas de dos componentes con varianzas iguales. En esta misma línea, trabajos posteriores fueron implementados por Charlier (1906), Charlier y Wicksell (1924), Cohen (1967), y Tan y Chang (1972). La estimación Máxima Verosimilitud para mixturas fue desarrollado por Rao (1948) y Hasselblad (1966, 1969). Desde el enfoque Bayesiano, Garrido L. (2010) en sus tesis Doctoral, plantea la estimación de parámetros en mixturas finitas de distribuciones pertenecientes a la familia exponencial biparamétrica, específicamente sobre la Normal, Exponencial, Gamma y Weibull, mediante muestreador de Gibbs y el algoritmo Metropolis–Hastings. En 2012, Garrido L. & Cepeda–Cuervo E., utilizan métodos de Monte Carlo vía Cadenas de Markov para modelar la media y la dispersión en el caso de mixturas de distribuciones Normal y Gamma. Finalmente, Garrido L. & Cepeda–Cuervo E. (2014), proponen métodos de estimación de los parámetros de media y dispersión, para la heterocedasticidad en modelos con mixtura de las distribuciones Weibull y Normal.

En las siguientes secciones se aborda la teoría relacionada con mixturas finitas para la distribución beta considerando el modelo dado en la ecuación (2.5) y en la siguiente, la metodología bayesiana con las especificaciones de las aprioris y desarrollos propios de la técnica considerando el modelo mixto de la ecuación (2.6) y se plantea para trabajos futuros el uso de mixturas para estimación de proporciones desde el enfoque bayesiano en áreas pequeñas.

Para la implementación en áreas pequeñas, se asume que las M áreas se encuentran agrupadas en K componentes no necesariamente dadas por las áreas mayores, tal como se asumió entorno a modelos mixtos en el capítulo 1. Por lo tanto, la pertenencia de los estimadores directos a cada componente se desconoce y dada la naturaleza del estimador directo, pueden modelarse mediante distribución beta con el fin de estimar los parámetros de media y dispersión de cada componente.

De acuerdo con Grüe *et al.* (2012), el modelo mixturas finitas para distribuciones beta está dado por:

$$g(\widehat{\theta}; x, z, c, \lambda) = \sum_{k=1}^{K} \pi(k; c, \alpha) f(\widehat{\theta}; h_1^{-1}(x^{\intercal}\beta_k), h_2^{-1}(z^{\intercal}\gamma_k)), \tag{2.7}$$

donde $g(\cdot;\cdot)$ es la densidad de mixtura finita de K componentes, $f(\widehat{\theta};\mu,\phi)$ se denominan las densidades de los componentes y corresponden para el caso, a la distribución beta usando la reparametrización de media y dispersión de la ecuación (2.2) quienes adicionalmente, son modelados mediante los modelos lineales generalizados en la ecuación (2.5). Por su parte $\pi(k;\cdot)$ se denominan pesos de la mixtura y representan la probabilidad de que la realización $\widehat{\theta}_d$, haya sido generada por las K densidades distintas, por lo tanto, tiene las características de estar acotadas en el intervalo abierto $[0,1]$ y su suma sobre todas las componentes equivale a la unidad.

Para determinar los pesos asociados a los componentes, supóngase el problema genérico de máxima verosimilitud, para el cual se cuenta con una muestra aleatoria de la variable $\widehat{\theta}$, con función de densidad $p(\widehat{\theta}|\lambda)$ en donde λ son los parámetros a estimar de acuerdo con las realizaciones $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_d)$, la función de verosimilitud de la mixtura está dada por:

$$L(\Lambda|\widehat{\theta}) = \prod_{d=1}^{m} \sum_{k=1}^{K} \pi_k f_K(\widehat{\theta}_d|\lambda_k),$$

Por lo tanto, la función de $log - verosimilitud$ es:

$$l(\Lambda|\widehat{\theta}) = \log L(\Lambda|\widehat{\theta}) = \log \prod_{d=1}^{m} \left\{ \sum_{k=1}^{K} \pi_k f_K(\widehat{\theta}_d|\lambda_k) \right\} = \sum_{d=1}^{m} \log \left\{ \sum_{k=1}^{K} \pi_k f_K(\widehat{\theta}_d|\lambda_k) \right\}, \quad (2.8)$$

en donde el procedimiento de maximización puede no tener solución analítica debido a la presencia de la suma de logaritmos, por lo que en la práctica se utilizan algoritmos basados en Newton–Rapson, Fisher Scoring o Esperanza – Maximización (Gómez, A. (2014)).

2.3.1. Estimación clásica del modelo en mixturas finitas

Una metodología de estimación vía máxima verosimilitud para distribuciones de mixturas finitas es el algoritmo Esperanza-Maximización (EM), propuesto por Dempster *et al.* (1977), que permite obtener máximos de la función de verosimilitud cuando no es posible hacerlo de forma analítica, propuesto en principio para tratar datos faltantes. El uso fue luego extendido a situaciones en las que los parámetros de la distribución dependen de variables no observables. En mixturas finitas puede abordarse la técnica debido a que la pertenencia a cada clúster es desconocida a priori y, por lo tanto, aunque no obedezca a una situación de datos faltantes si lo es de datos incompletos.

La técnica introduce una variable indicadora para identificar la pertenencia de cada observación a un clúster, siendo considerada como una variable latente definida por $W = (W_1, W_2 \ldots, W_m)$ cuya realización se nota $w = (w_1, w_2 \ldots, w_m)$. El arreglo de datos completos puede representarse por la dupla $(\widehat{\theta}, W)$ en la que cada realización de $\widehat{\theta}_d$ se encuentra asociada una realización w_d. Así, la variable aleatoria w_k es dicotómica k-dimensional cuyo d-ésimo elemento indica la pertenencia de la observación $\widehat{\theta}_d$ a la k-ésima componente de la mixtura, $d = 1, 2 \ldots, m.$; $k = 1, 2 \ldots, K.$, luego W_k sigue una distribución multinomial de una única realización sobre las k componentes con probabilidad π tal que $W \sim Multi(1, \pi)$ con $\pi = (\pi_1, \ldots, \pi_k)$ tales que $\sum \pi_k = 1$ y por lo tanto, la

distribución está dada por:

$$p(W_k) = p(W_k = w_k) = \binom{1}{w_{d1} w_{d2} \dots w_{dk}} \prod_{k=1}^{K} \pi_k^{w_{dk}}.$$

De esta manera, la función de densidad de los datos completos dada por $f(\widehat{\theta}_k, W_k)$ puede calcularse a partir de la regla de Bayes mediante $f(\widehat{\theta}_k|W_k)p(W_k)$ así:

$$f(\widehat{\theta}_k|W_k)p(W_k) = \left\{ \prod_{k=1}^{K} f_k(\widehat{\theta}_d|\lambda_k)^{w_{dk}} \right\} \left\{ \prod_{k=1}^{K} \pi_k^{w_{dk}} \right\} = \prod_{k=1}^{K} \left[\pi_k f_k(\widehat{\theta}_d|\lambda_k) \right]^{w_{dk}},$$

con lo cual, la función de verosimilitud está dada por:

$$L(\Lambda|\widehat{\theta}, w) = \prod_{d=1}^{m} \prod_{k=1}^{K} \left[\pi_k f_k(\widehat{\theta}_d|\lambda_k) \right]^{w_{dk}}. \tag{2.9}$$

La función así obtenida considera los pesos de la mixtura sin presentar los inconvenientes de maximización de la log verosimilitud, tal como sucedía en la ecuación (2.3). Considerado que uno de los objetivos principales del modelamiento mediante mixturas finitas es obtener la partición adecuada de las subpoblaciones que generan la heterogeneidad en las observaciones, la probabilidad de pertenencia de cada observación al clúster se actualiza una vez se realizan las estimaciones de los parámetros del modelo mediante el teorema de bayes, convirtiéndose así, en probabilidades a posteriori.

El algoritmo EM (Ver Anexo A.1) es implementado por Grüe *et al.* (2012), en el paquete betareg del Software R – Project, para modelamiento de mixturas en distribuciones Beta, además incluyó procedimientos de corrección de sesgo de los estimadores MV y regresión particionada. En esta implementación, para los pesos $\pi(k; c, \alpha)$ de la ecuación (2.7) los autores definen una función que permite inicializarlos a partir de la inclusión de variables c, que puedan influenciar la construcción de los clúster:

$$\pi(k; c, \alpha) = \frac{\exp\{c^\intercal \alpha_k\}}{\sum_{u=1}^{K} \exp\{c^\intercal \alpha_u\}},$$

que corresponde a un modelo logístico multinomial, que relaciona las probabilidades de pertenencia a cada grupo con variables independientes c, denominadas concomitantes. Debido a que éste planteamiento corresponde a una metodología de estimación clásica - MV, la estimación obedecería en el contexto de áreas pequeñas al método EBLUP.

2.4. SAE mediante Estimación Bayesiana Beta

Dentro del contexto de SAE, el planteamiento bayesiano no requiere el supuesto de las convergencias asintóticas de los estimadores máxima verosimilitud para realizar inferencia, por lo tanto, no requiere aproximaciones de grandes muestras, convirtiéndolo en una estrategia adecuada para áreas pequeñas. Para la distribución Beta en el contexto de distribuciones pertenecientes a la familia exponencial biparámetrica, Cepeda-Cuervo (2001) y Cepeda *et al.* (2005), proponen una estrategia basada en algoritmos MCMC

para estimar de forma conjunta la media μ_d y dispersión ϕ_d, para la cual es necesario especificar distribuciones a priori para todos los parámetros del modelo de la ecuación (2.5).

Considerando que $L(\Theta|\widehat{\theta}_d)$ es la verosimilitud de $\Theta = (\beta, \gamma)$ y $p(\Theta)$ es su distribución a priori conjunta, la distribución a posteriori de Θ es tal que $\pi(\Theta|\widehat{\theta}_d) \propto L(\Theta)p(\Theta)$. Bajo el supuesto de normalidad sobre la distribución a priori de los parámetros, la distribución a posteriori $\pi(\Theta)$ no es tratable analíticamente, de tal manera que la propuesta de estimación se basa en un algoritmo que alterna de forma iterativa la estimación de los parámetros, muestreando valores de β y γ originados por las distribuciones condicionales a posteriori $\pi(\beta|\gamma, \widehat{\theta}_d)$ y $\pi(\gamma|\beta, \widehat{\theta}_d)$ respectivamente, basadas en kernels de transición normales, pues estas distribuciones condicionales a posteriori tampoco tienen una expresión analítica que facilite el cálculo de sus valores esperados.

Para la construcción de los kernels de transición, los autores definen variables de trabajo que permitan aproximarse a las funciones de enlace $h_1(\cdot)$ y $h_2(\cdot)$ mediante linealización de Taylor de primer orden y variables aleatorias t_1 y t_2 insesgadas para μ y ϕ respectivamente. Así, de acuerdo con el modelo para la media definido en la ecuación (2.5), la variable de trabajo está dada por:

$$\widetilde{\theta}_d = x_d^\intercal \beta^{(c)} + \frac{\widehat{\theta}_d - \mu_d^{(c)}}{\mu_d^{(c)}(1 - \mu_d^{(c)})},$$

donde $\beta^{(c)}$ y $\mu_d^{(c)}$ son los valores actualizados de β y μ_d, respectivamente y $d = 1, \ldots, m$.

De la reparametrización de la distribución beta en función de la media, se sabe que $V(\mu) = \mu(1-\mu)$ (ver ecuación 2.3), por lo tanto, el segundo factor de la suma en la variable aleatoria $\widetilde{\theta}_d$ puede considerarse una medida de ajuste de error dada por el residuo de la estimación y la varianza del parámetro modelado en la actualización (c).

En relación a las propiedades de centralidad y dispersión de la variable de trabajo $\widetilde{\theta}_d$ se tiene que, $E(\widetilde{\theta}_d) = x_d^\intercal \beta^{(c)}$ y $V(\widetilde{\theta}_d) = \left(\frac{dh_1(\mu_d)}{\mu_d}\right)^2 Var(\widehat{\theta}_d)$ de tal manera que al asumir una función de probabilidad, puede inducirse la generación de cadenas de Markov, que mediante algoritmos de selección Metropolis–Hastings o Gibbs logren estado estacionario.

Bajo el supuesto de normalidad sobre la distribución a priori condicional $\beta|\gamma \sim N(b, B)$ y la variable de trabajo $\widetilde{\theta}_d \sim N(x_d^\intercal \beta, V(\widetilde{\theta}_d))$ el kernel de transición q_1 está dado por la distribución a posteriori de β, $q_1(\beta|\beta^{(c)}, \gamma^{(c)}) \sim N(b^*, B^*)$ donde,

$$b^* = B^*(B^{-1}b + X^\intercal \Sigma^{-1} \widetilde{\theta}) \tag{2.10}$$

$$B^* = (B^{-1} + X^\intercal \Sigma^{-1} X)^{-1}, \tag{2.11}$$

y Σ es la matriz diagonal con valores $V(\widetilde{\theta}_d) = \widetilde{\sigma}_d$.

Considerando la distribución condicional a posteriori $\pi(\gamma|\beta)$ para los coeficientes γ que modelan la dispersión ϕ como se presenta en la ecuación (2.5), su tratamiento no es viable de forma analítica tal como en el caso de la condicional a posteriori completa para β, por lo tanto, según la metodología propuesta por los autores, es necesario que $E(t_2) = \phi$ luego,

con base en la ecuación (2.1):

$$t_{2d} = \frac{(\alpha_d + \lambda_d)^2}{\alpha_d}\widehat{\theta}_d = \frac{\phi_d^2}{\mu_d\phi_d}\widehat{\theta}_d = \frac{\phi_d}{\mu_d}\widehat{\theta}_d.$$

Nuevamente mediante linealización de Taylor al rededor del valor actualizado $\gamma^{(c)}$ de ϕ_d, la variable de trabajo se encuentra dada por:

$$\breve{\theta}_d = z_d^\intercal\gamma^{(c)} + \frac{\widehat{\theta}_d}{\mu_d^{(c)}} - 1,$$

luego, la variable de trabajo $\breve{\theta}_d$ puede entenderse como la predicción de ϕ_d corregida por el error de estimación de la media $\widehat{\theta}_d - \mu_d^{(c)}$, con respecto al valor esperado $E(\widehat{\theta})$ en la actualización (c). Para la generación de cadenas de Markov, el valor esperado y la varianza de la variable de trabajo para γ, son respectivamente, $E(\breve{\theta}_d) = z_d^\intercal\gamma$ y $V(\breve{\theta}_d) = \left(\frac{dh_2(\phi_d)}{\phi_d}\right)^2 Var(t_{2d})$.

Bajo los supuestos distribucionales, $\gamma|\beta \sim N(g, G)$ y $\breve{\theta}_d \sim N(z_d^\intercal\gamma, V(\breve{\theta}_d))$ el kernel de transición de la distribución condicional a posteriori de γ es $q_2(\gamma|\gamma^{(c)}, \beta^{(c)}) \sim N(g^*, G^*)$ donde:

$$g^* = G^*(G^{-1}g + Z^\intercal\Phi^{-1}\breve{\theta}) \tag{2.12}$$

$$G^* = (G^{-1} + Z^\intercal\Phi^{-1}Z)^{-1}, \tag{2.13}$$

y Φ es la matriz diagonal de entradas $\widetilde{\sigma}_d$.

2.4.1. Estimación bayesiana del modelo Beta

El proceso de estimación se basa en la extracción de valores generados por los kernels de transición q_1 y q_2 para β y γ respectivamente, mediante el algoritmo Metropolis–Hastings de MCMC, bajo el siguiente esquema de actualización:

1. Definir valores (β^0, γ^0) para inicializar la cadena en la iteración $j = 1$
2. Generar un nuevo valor ψ de la densidad propuesta $q_1(\beta^{j-1}, \cdot)$
3. Calcular la probabilidad de aceptación $\alpha(\beta^{j-1}, \psi)$, en caso de aceptar el cambio de estado, hacer $\beta^j = \psi$, en otro caso $\beta^j = \beta^{j-1}$ (Ver apéndice A.2.1)
4. Generar un nuevo valor ψ de la densidad propuesta $q_2(\gamma^{j-1}, \cdot)$
5. Calcular la probabilidad de aceptación $\alpha(\gamma^{j-1}, \psi)$, en caso de aceptar el cambio de estado, hacer $\gamma^j = \psi$, en otro caso $\gamma^j = \gamma^{j-1}$
6. Generar una nueva iteración para el paso $j + 1$
7. Repetir los pasos hasta alcanzar convergencia.

2.5. SAE mediante Estimación Bayesiana Beta con efectos mixtos

En la sección 2.3.1 se presentó una manera alterna de estimar los parámetros de media y dispersión de cada una de las K componentes de la mixtura finita vía máxima verosimilitud, que considera implícitamente un comportamiento debido a subgrupos poblacionales. Por su parte, el tratamiento de modelos que formulan explícitamente los efectos aleatorios, obedece a los modelos mixtos doblemente generalizados que, en distribuciones beta se desarrolla principalmente desde el enfoque Bayesiano.

De acuerdo con Figueroa–Zuñiga *et. al*(2013), considérese $\widehat{\theta}_1, \ldots, \widehat{\theta}_L$ vectores aleatorios continuos con realizaciones $\widehat{\theta}_l = (\widehat{\theta}_{1l}, \widehat{\theta}_{2l}, \ldots, \widehat{\theta}_{dl}, \ldots, \widehat{\theta}_{n_l l})^\intercal$, es decir cada área mayor l, tiene realizaciones que varían de $d = 1, \ldots, n_l$ en donde $\sum_{l=1}^{L} n_l = m$ siendo m el total de áreas pequeñas. Cada realización $\widehat{\theta}_{dl}$ siguen una distribución Beta que puede reparametrizarse en función de su media μ_d y su parámetro de dispersión ϕ_d como en la ecuación (2.2). Adicionalmente la esperanza condicional dados los efectos aleatorios, $h_1(E(\widehat{\theta}_d|u_d))$, pueden modelarse a partir del componente sistemático η_d como:

$$logit(\mu_d) = \eta_d = X_d\beta + Z_d u_d, \tag{2.14}$$

donde, $\mu_d = (\mu_1, \ldots, \mu_m)^\intercal$ es el vector de medias, X_d es la matriz de diseño de orden n_l x p de p variables conocidas, con vector de coeficientes $\beta = (\beta_1, \ldots, \beta_p)$ fijos. Los efectos aleatorios son modelados mediante la matriz de diseño Z_d de tamaño n_l x q con vector de coeficientes aleatorios u_d. En cuanto al parámetro de dispersión ϕ se plantea el modelo general:

$$log(\phi_d) = \xi_d = w_d\gamma + b_d\delta_d, \tag{2.15}$$

donde w_d es el vector de diseño de tamaño p', correspondiente a los efectos fijos γ que puede o no ser del contenido y tamaño de los efectos fijos que modelan la media; análogamente, b_d corresponde al vector de diseño de tamaño q' que modela los efectos aleatorios δ_d.

Para poder comparar los resultados obtenidos mediante los modelos de mixturas finitas y de efectos mixtos, una estrategia sería considerar $L = K$, sin embargo, las probabilidades a posteriori π_k son resultado de la función de verosimilitud, el método de estimación y actualización de valores conjuntos, por lo cual un trabajo futuro podría consistir en la estimación de áreas pequeñas mediante mixturas finitas con enfoque bayesiano.

2.5.1. Estimación bayesiana del modelo Beta con efectos mixtos

Se asume $\widehat{\theta}_d|\eta, \beta, u_d, \xi, \delta, b_d \sim beta(\mu_d\phi_d, (1-\mu_d)\phi_d)$, $d = 1, \ldots, m.$, tal que μ_d y ϕ_d pueden ser modelados mediante las relaciones de las ecuaciones (2.14) y (2.15) respectivamente. Para el desarrollo del método **HB** de SAE descrito en la sección 1.3, es necesario definir las distribuciones a priori para todos los parámetros, considerados en éste enfoque como variables aleatorias, para que, mediante el teorema de Bayes se obtengan distribuciones conjuntas condicionales a posteriori que faciliten el cálculo de medidas de localización y escala que se postulan como estimadores puntuales y de precisión de cada área pequeña. Comúnmente se asume distribuciones normales apriori para los efectos aleatorios y los

efectos fijos, sin embargo, Figueroa–Zuñiga *et. al*(2013) señalan que resulta conveniente utilizar distribuciones con colas más pesadas como la t-Student multivariada, para controlar posibles inconvenientes ante la presencia de outliers, adicionalmente, para muestras pequeñas la elección de las apriori influyen en mayor medida las inferencias debido al gran número de parámetros a estimar. Bajo esta línea, y considerando que conforme los grados de libertad de una t-Student crecen, su forma se aproxima más a una distribución Normal multivariada, se asumen las siguientes distribuciones a priori.

i. $u_d|gl_{1u}, \mu_u, \Sigma_u \sim t_q(gl_{1u}, \mu_u, \Sigma_u)$,

ii. $gl_{u1} \sim \epsilon(a)$,

iii. $\Sigma_u|\phi, c \sim IW_q(\psi, c)$.

Es decir, bajo el modelo de la ecuación (2.14), los efectos aleatorios u_d, $d = 1, \ldots, m.$, siguen una distribución t-Student tal que, $u_d|gl_{u1}, \mu_u, \Sigma_u \sim t_q(gl_{u1}, \mu_u, \Sigma_u)$, donde los grados de libertad, gl_{u1}, son modelados a partir de una distribución exponencial con media $1/a$; los efectos aleatorios u_d, están centrados en cero, $\mu_u = 0$, y Σ_u que corresponde a una matriz de parámetros de escala en la distribución t–Student, simétrica definida positiva de orden q x q y está distribuida Wishart Inversa, $\Sigma_u|\phi, c \sim IW_q(\psi, c)$.

En relación a la dispersión ϕ, Figueroa–Zuñiga *et. al* (2013), consideran dos casos, uno, asumiendo que es constante, y en tal caso el modelo de la ecuación (2.15) puede escribirse como:

$$log(\phi_d) = \xi_d = \gamma. \tag{2.16}$$

De tal manera que para estimar el parámetro, introducen una distribución a priori no informativa que puede definirse mediante una Gamma Inversa o una distribución menos informativa como la uniforme. De lo anterior, faltaría realizar la especificación de las a priori para los coeficientes que modelan la media así:

iv. $\beta|gl_{2\beta}, \mu_\beta, \Sigma_\beta \sim t_p(gl_{2\beta}, \mu_\beta, \Sigma_\beta)$,

v. $\phi \sim GI(t_1, t_2)$,

en donde $gl_{2\beta}$ son los grados de libertad de los coeficientes en la distribución.

En el segundo caso, los autores consideran que la dispersión es también modelada mediante efectos mixtos tal como en la ecuación (2.15) y por lo tanto, el literal *v.* de las a prioris antes específicadas se extiende para modelar las a prioris para los parámetros adicionales:

v. $\delta_d|gl_\delta, \Sigma_\delta \sim t_{q'}(gl_\delta, 0, \Sigma_\delta)$,

vi. $gl_\delta \sim \epsilon(e)$,

vii. $\Sigma_b|\phi*, s \sim IW_q(\psi*, s)$,

viii. $\gamma|gl_\gamma, \mu_\gamma, \Sigma_\gamma \sim t_p(gl_\gamma, \mu_\gamma, \Sigma_\gamma)$,

donde w_d es el vector de diseño de tamaño p' correspondiente a los efectos fijos γ que puede o no, ser del contenido y tamaño de los efectos fijos que modelan la media; análogamente,

b_d corresponde al vector de diseño de tamaño q' que modela los efectos aleatorios δ_d; así mismo $gl.$ corresponden a los grados de libertad de los respectivos coeficientes que para el caso de los efectos aleatorios es estocástica, mientras que para los efectos fijos no lo es.

Bajo estas estructuras, Figueroa–Zuñiga *et. al* (2013) demuestra que las componentes sistemáticas que modelan las medias de las áreas pequeñas η_d son independientes y tiene función de densidad tal que $f(\eta_d|\beta, \Sigma_u, gl_u) \propto f(u_d|\beta, \Sigma_u, gl_u)$, de forma similar sucede con las componentes sistemáticas ξ, condicionalmente sobre γ, Σ_b y gl_b son independientes y además su función de densidad condicional es proporcional a la densidad condicional de los efectos aleatorios b_d: $f(\xi_d|\gamma, \Sigma_b, gl_b) \propto f(b_d|\gamma, \Sigma_b, gl_b)$. Con lo que finalmente, bajo el supuesto de independencia de todas las variables aleatorias, la densidad a posteriori conjunta, condicionada a los estimadores directos para el modelo con dispersión constante, está dada por:

$$f(\beta, \Sigma_u, gl_u, \phi, \eta|\widehat{\theta}) \propto \left[\prod_{d=1}^{m}\prod_{l=1}^{n_l} f(\widehat{\theta}_{dl}|\eta_{dl}, \phi)\right]\left[\prod_{d=1}^{m} f(\eta_d|\beta, \Sigma_u, gl_u)\right] f(\beta)f(\Sigma_u)f(gl_u)f(\phi). \tag{2.17}$$

Mientras que, la distribución a posteriori conjunta de los parámetros, condicionada a los estimadores directos, cuando el parámetro de dispersión se modela con efectos aleatorios, está dada por:

$$\begin{aligned} f(\beta, \Sigma_u, gl_u, \gamma, \Sigma_b, gl_b, \eta, \xi|\widehat{\theta}) \propto & \left[\prod_{d=1}^{m}\prod_{l=1}^{n_l} f(\widehat{\theta}_{dl}|\eta_{dl}, \xi_{dl})\right]\left[\prod_{d=1}^{m} f(\eta_d|\beta, \Sigma_u, gl_u)\right] \\ & \times \left[\prod_{d=1}^{m} f(\xi_d|\gamma, \Sigma_b, gl_b)\right] f(\beta)f(\Sigma_u)f(gl_u)f(\gamma)f(\Sigma_b)f(gl_b) \end{aligned} \tag{2.18}$$

Debido a que a partir de las densidades $f(\beta, \Sigma_u, gl_u, \phi, \eta|\widehat{\theta})$ y $f(\beta, \Sigma_u, gl_u, \gamma, \Sigma_b, gl_b, \eta, \xi|\widehat{\theta})$ no es posible obtener de forma analítica los valores esperados y dispersiones de interés, se opta por generar una secuencia de muestras aleatorias que siguen la densidad objetivo, sobre las que se evalúan sus propiedades de centralidad y dispersión; esto es posible a partir de los algoritmos basados en Monte Carlo vía Cadenas de Markov (MCMC), como el Metrópolis – Hastings y el Muestreador de Gibbs (Ver anexo A.2), siendo el último, el utilizado por los autores.

2.6. Criterios de selección y evaluación de modelos

Para validar la predicción de los modelos, es necesario en principio seleccionar entre un grupo de alternativas, el más adecuado y una vez elegido, evaluar su desempeño. En relación a la selección, las decisiones son soportadas a través de mecanismos basados en la comparación de las verosimilitudes definidos en Spiegelhalter *et al.* (2002) . Una selección adicional en mixturas finitas, corresponde al número de componentes K que mejor construyen los grupos, pero en todo caso, puede evaluarse a partir de los mismos

criterios de selección, sometiendo a prueba modelos que varían la cantidad de grupos. En el caso de evaluación de desempeño se considera la capacidad predictiva de los modelos y la tradicional evaluación de residuales.

2.6.1. Criterios de selección tipo desvío

Sea $L(y|\Theta_p, M_a)$ la función de verosimilitud de un modelo alternativo M_a, donde Θ_p contiene todos los p' parámetros incluidos en el modelo para m observaciones, la función de desvío $D(\Theta_p) = -2\log L(y|\Theta_p, M_a)$ es una generalización del análisis de varianza, que permite construir medidas de discrepancia en MLG, bajo la cual se definen los criterios de información de Akaike (AIC), Bayesiano (BIC) y de la Devianza (DIC), como sigue:

$$\begin{aligned} AIC(M_a) &= D\left[E(\Theta_p|y, M_a)\right] + 2p'. \\ BIC(M_a) &= D\left[E(\Theta_p|y, M_a)\right] + \log(m)p'. \\ DIC(M_a) &= 2E\left[D(\Theta_p|y, M_a)\right] - D\left[E(\Theta_p|y, M_a)\right]. \end{aligned}$$

Para la decisión, se opta por el modelo M entre los a modelos alternativos si éste tiene el menor valor de acuerdo con los criterios de información expuestos.

2.6.2. Análisis de residuales

Para concluir, se realiza el análisis de residuales, en este sentido, para regresión Beta Espinheira *et al.*(2008) y Ferrari *et al.*(2011) plantean los tradicionales residuales estandarizados definidos como:

$$re_d^p = \frac{\widehat{\theta}_d - \widehat{\mu}_d}{\sqrt{(\widehat{V}(\widehat{\theta}_d))}},$$

en donde, de acuerdo con la ecuación (2.3), $\widehat{V}(\widehat{\theta}_d) = \frac{\widehat{\mu}_d(1-\widehat{\mu}_d)}{\widehat{\phi}_d+1}$, denominados residuales de Pearson, una desventaja de estos residuales es que en general su distribución para datos no Normales es asimétrica y por lo tanto no es de esperar que pruebas paramétricas o visuales indiquen lo contrario.

De esta manera los residuales propuestos por los autores para el caso de modelamiento beta, son los residuales estandarizados ponderados datos por:

$$re_d^w = \frac{\widehat{\theta}_d^* - \widetilde{\mu}_d^*}{\sqrt{v_d(1-l_{dd})}},$$

en los que:

$$\begin{aligned} \widehat{\theta}_d^* &= \log\left[\frac{\widehat{\theta}_d}{1-\widehat{\theta}_d}\right], \\ \widetilde{\mu}_d^* &= \psi(\mu_d\phi_d) - \psi((1-\mu_d)\phi_d), y, \\ v_d &= \psi'(\mu_d\phi_d) + \psi'((1-\mu_d)\phi_d), \end{aligned}$$

con $\psi(\cdot)$ la función digamma, es decir, para la función Gamma $\Gamma(z)$ se tiene que $\psi(z) = \frac{d\log\Gamma(z)}{dz} = \frac{\Gamma'(z)}{\Gamma(z)}$, además, l_{dd} es la d-ésima componente de la matriz Hat

ponderada $H = W^{1/2}X(X^{\intercal}WX)X^{\intercal}W^{1/2}$ siendo W la matriz diagonal con elementos $w_d = \phi_d v_d[1/\{h_1'(\mu_d)\}]$. Espinheira *et al.* (2008) muestran que los errores así definidos, tienen propiedades de convergencia a la distribución normal y adicionalmente con el factor de ponderación l_{dd} es posible identificar datos influyentes. El cálculo de los errores Pearson y Ponderados son considerados dentro de los paquetes Betareg y Bayesianbetareg del Software R–Project, mas no para Betamix, por lo cual se construyen para este trabajo (Ver anexo B.3.3).

CAPÍTULO 3

Planteamiento del problema y descripción de la información

En este capítulo se especifica el parámetro que se busca estimar y su importancia en el contexto de políticas públicas, contexto mismo que permite plantear qué indicadores podrían describir adecuadamente el comportamiento del parámetro de interés y que, dentro del ajuste del modelo tienen el rol de variables independientes. Adicionalmente, se presenta la descripción básica relacionada con la metodología estadística de la encuesta, para definir lo que para este trabajo se considera área pequeña, destacando la relevancia y tamaño de las áreas mayores en terminos de la Nación. Por último, el análisis descriptivo tiene tres alcances: El primero, sobre el comportamiento territorial que sugiere la viabilidad de usar estrategias de agrupamiento de áreas pequeñas propias de la métodología de mixturas finitas, así como las límitaciones para incluir estructuras de correlación espacial o temporal en el modelo. En segundo lugar, se describe el comportamiento de las áreas pequeñas mediante un análisis descriptivo, en función del cumplimiento de los supuestos sobre el soporte de la variable respuesta en la distribución beta y correlaciones entre las variables; así mismo, se retoma la metodología estadística, para evidenciar que se tiene cobertura completa sobre la población objetivo y que no hay presencia de áreas pequeñas no muestreadas, luego el uso del modelo tiene como propósito la estimación de los errores estándar del parámetro, pero no la estimación del mismo para áreas faltantes. En tercer lugar, y para concluir, se presenta la composición de las variables para poder abordar la aplicación que se desarrolla en el capitulo 4.

3.1. Justificación del modelo y parámetro de interés

La mayoría de los indicadores de encuestas socio–económicas o demográficas y de salud, definen su comportamiento en el intervalo $(0, 1)$, sin embargo, un tema latente en este tipo de investigaciones y que se ha convertido en un desafío constante y creciente asumido por académicos, gobiernos y organizaciones, está dado por la necesidad de comprender la problemática de la pobreza y de la desigualdad que afronta la población.

En el país, desde hace varios años se cuenta con índices de pobreza, calidad de vida y de desarrollo humano, tales como el Índice de Necesidades Básicas Insatisfechas (NBI)

o el Índice de Condiciones de Vida (ICV) o los construidos bajo enfoque estrictamente monetario, como son los métodos de Línea de pobreza, indigencia o el Coeficiente de GINI, este último, para evaluar la desigualdad en la distribución del ingreso. Los platemientos de fondo de éstos indicadores, se basan en una concepción utilitarista en la cual se vincula directamente el ingreso con la capacidad de adquisición (SDP, 2011). No obstante, Amartya Sen (1990) platea un nuevo concepto que se resume en la frase "La principal riqueza de una nación es su gente" para denotar que la riqueza potencial o el bienestar, depende de las condiciones del individuo y no únicamente de los factores de ingreso y/o adquisición de bienes o servicios. Más adelante, afirma también que existe "una correspondencia estrecha entre la pobreza vista como escasez del ingreso y la pobreza vista como la falta de capacidad para satisfacer algunas necesidades elementales y esenciales" (Sen, 1997); constituyendo así el concepto de **capacidades**.

Desde el enfoque de capacidades se desarrollan nuevas metodologías de medición de la pobreza, entre las más recientes, se destaca el Índice de Pobreza Multidimensional (IPM), que busca describir las múltiples carencias que sufre una población, involucrando variables de potencial humano. Este indicador es propuesto y desarrollado por Oxford Poverty & Human Development Initiative, OPHI en 2008; En 2010, fue adaptado para Colombia por el Departamento Nacional de Planeación, DNP y para el Distrito Capital fue adaptado en el año 2011 por la Secretaría Distrital de Planeación, SDP.

En general, los indicadores de pobreza para Bogotá, al momento de la encuesta, mostraron un comportamiento positivo ya que entre 2007 y 2011 el NBI tuvo una reducción de aproximadamente 2 puntos porcentuales y el ICV en 2011 tuvo una ganancia de 1.4 puntos sobre la calificación calculada para 2007; la línea de pobreza y de indigencia, carecen de comparación debido al cambio metodológico de la línea de pobreza – (SDP, 2011). Caso contrario sucede con la desigualdad, por ejemplo Colombia, de acuerdo con el coeficiente de concentración del ingreso GINI se encuentra dentro de los cinco países con mayores niveles de desigualdad en la distribución del ingreso, con un valor a 2010 de 0.560, siendo superado tan solo por Panamá, Bolivia, Haití y Sudáfrica. Por su parte, el Distrito Capital no es ajeno al comportamiento nacional ya que en 2011 presenta un coeficiente de 0.542 que, si bien es inferior a la concentración del ingreso nacional, éste ha crecido en relación a 2007 cuando el indicador se ubicó en 0,511.

Bajo la consideración de que para alcanzar la igualdad, condiciones como los ingresos o la educación de los padres, el sexo, la etnia, el lugar de procedencia o migración, etc., no debería afectar, según el Banco Mundial, el éxito personal, pues son circunstancias sobre las cuales no se tiene decisión o control, se desprende el Índice de Oportunidades Humanas (IOH), que busca evaluar la desigualdad, midiendo qué tanto impactan las circunstancias personales exógenas, la probabilidad de que un niño acceda a servicios necesarios tales como educación oportuna, agua potable o conexión a servicio de energía y que le permitan "desarrollarse integralmente en un ambiente menos inequitativo que el de sus padres, para contar en la edad adulta con mejores perspectivas de inserción en el mercado laboral e independencia económica, por la vía de la generación de ingresos" (Velez *et al.* (2011)), eje central del concepto de **igualdad de oportunidades**.

Por lo anterior, el IOH está enfocado a las personas de 17 años y menos, lo que se justifica en primer lugar, porque según James Heckman, premio Nobel en economía (2000), "es muy probable que la inversión en la formación de capital humano en una fase temprana del ciclo de vida sea más eficiente que la mitigación de la desventaja a edades más avanzadas". En segundo lugar, porque dada su vulnerabilidad económica, las oportunidades de los niños es de interés político, en Colombia por ejemplo, el riesgo de vivir en un hogar en pobreza extrema para la población infantil es el doble que la de los adultos mayores (Nuñez J., 2009).

La metodología estadística para calcular el IOH, se basa en regresión logística para lo que se requiere información a nivel de unidad que, como se ha indicado, generalmente por razones de reserva estadística no es de fácil acceso; luego considerando que se dispone de información a nivel de área, se propone abordar el concepto de desigualdad con el enfoque de capacidades, a partir de las oportunidades humanas y de las dimensiones del IPM, ya que éste indicador incluye dentro de las dimensiones de cálculo, uno asociado específicamente a la población de 17 años y menos, denominado precisamente, niñez y juventud, las restantes dimensiones, permitirían analizar cómo las condiciones de los hogares, afectan las asociadas a dicha población, permitiendo identificar atráves de la estimación en áreas pequeñas, focos de población en vulnerabilidad y desigualdad, así como también factores de riesgo asociados.

En resumen, el interés es estimar el porcentaje de hogares con carencias para la población de 17 años y menos, de acuerdo con las condiciones de los hogares donde residen con base en la distribución beta, cuyo modelamiento es planteado siguiendo el concepto de igualdad de oportunidades. Esta propuesta es, desde el punto de vista académico, novedoso, puesto que no se encontró en la revisión bibliográfica de SAE basada en modelos, aplicaciones o desarrollos para respuesta beta, incluso para Colombia dentro de la misma revisión, no se encontraron aplicaciones para ningún tipo de variable dependiente (Normal, Binomial, Poisson).

3.2. Justificación del tratamiento de áreas pequeñas para el Distrito Capital

La información para la estimación de proporciones en áreas pequeñas mediante distribución Beta, es tomada de la Encuesta Multipropósito de Bogotá – 2011, que evaluó las condiciones socio–económicas de la población civil Bogotana, residente en el área urbana de la capital en 2011; ésta se diseña como una encuesta por muestreo probabilístico estratificado, de conglomerados, con entrevista cara a cara e informante directo, en donde el criterio de estratificación corresponde a las localidades que conforman el territorio urbano de la capital. Los detalles metodológicos y operativos básicos de la encuesta se describen en apéndice (B.1), no obstante, mayor detalle y sus resultados se encuentran disponibles en la página Web de la Secretaría Distrital de Planeación, http://www.sdp.gov.co/portal/page/portal/PortalSDP.

De acuerdo con el parámetro de interés del presente trabajo, en la tabla 3.1 se muestra la estimación del porcentaje de hogares con carencias en las condiciones de la niñez y juventud según localidad, en donde se observa que las localidades con menor población joven y

de niñez con carencias son Chapinero (16,54 %), La Candelaria (18,18 %) y Teusaquillo (19,11 %); mientras que la situación de carencia para los niños y jovenes de las localidades Bosa, Rafael Uribe Uribe, San Cristóbal, Usme y Ciudad Bolívar, superan el (36 %) de los hogares.

Localidad	$\widehat{\theta}\,\%$
Usaquén	27,18
Chapinero	16,54
Santa Fe	30,15
San Cristóbal	37,59
Usme	40,44
Tunjuelito	32,87
Bosa	36,51
Kennedy	30,52
Fontibón	27,76
Engativá	28,00
Suba	33,26
Barrios Unidos	24,36
Teusaquillo	19,11
Los Mártires	25,45
Antonio Nariño	29,70
Puente Aranda	25,67
La Candelaria	18,18
Rafael Uribe Uribe	36,62
Ciudad Bolívar	42,06
Total Bogotá	31,63

TABLA 3.1. Porcentaje de hogares con carencias en condiciones de la niñez y la juventud por localidad.

Fuente: SDP, EMB 2011.

La problemática se acentúa si se analiza en relación al tamaño poblacional al que hace referencia la estimación, ya que el Distrito Capital, además de ser la ciudad de mayor población del país, está conformada territorial o geográficamente, como se indicó, por las denominadas localidades, cuya composición puede compararse a estructuras de ciudad considerando que trece de ellas tenían a la fecha de la encuesta (2011), más de 200 mil habitantes según proyecciones de población DANE 2005-2015. Para dimensionar las estructuras, por ejemplo, localidades como Suba y Kennedy se sitúan en el cuarto y quinto lugar de territorios con más de un millón de habitantes, por debajo de Medellín, Cali y Barranquilla; Por su parte, localidades como Engativá y Ciudad Bolívar tienen más habitantes que Cúcuta; Bosa, más que Soledad, Ibagué y Bucaramanga; y la localidad de Usaquén supera en tamaño poblacional a ciudades capitales como Pereira, Santa Marta, Villavicencio, Pasto, Montería y Valledupar.

Además del hecho de los amplios tamaños poblacionales, la dinámica de la capital sugiere la existencia de alta diversidad al interior de las localidades, sospecha que no puede verificarse mediante estimadores a niveles de desagregación de mayor detalle, con la precisión adecuada por las restricciones sobre los tamaños de muestra, convirtiéndolo así

en un típico problema de áreas pequeñas.

Dentro de la estructura geográfica de Bogotá se encuentra una unidad territorial intermedia entre localidad y barrios, denominadas Unidades de Planeamiento Zonal – UPZ, que como su nombre lo indica, son territorios o áreas geográficas menores, que le permiten a la administración Distrital planificar el desarrollo de la ciudad a este nivel. Lo anterior implica que, por el alcance geográfico y administrativo, las UPZ's son una unidad de análisis importante para realizar estimación de indicadores que facilite una mejor inversión de recursos mediante la identificación de necesidades focalizadas. Por lo tanto, para el objeto de este documento, las áreas pequeñas corresponden a las UPZ cuyos resultados se obtienen mediante SAE basada en el modelo Beta, adecuado para modelar la proporción de hogares con carencias en condiciones de la niñez y la juventud.

3.3. Análisis descriptivo de la información

Como se mencionó en el preámbulo del capítulo, el análisis descriptivo se desarrolla en función de las variables que enmarcan la propuesta de medición de desigualdades desde el enfoque de capacidades para la población joven e infante. El análisis tiene tres fines relevantes para el planteamiento de los modelos, el primero en relación a la pertinencia de agrupar UPZ's no necesariamente mediante las localidades que pueden considerarse las áreas mayores naturalmente constituidas por la estructura territorial, así como exponer las restricciones técnicas para incorporar correlación espacial en el entendido que se trata de áreas pequeñas geográficas. En segundo lugar, definir el alcance del modelamiento en términos de áreas pequeñas muestreadas ya que al estar todas las UPZ's en la muestra, los modelos se usan con fines de estimar la precisión de la estimación, así como analizar las correlaciones presentes en las variables del modelo propuesto; En tercer lugar, describir la composición de las variables incluidas en el modelo, para los posteriores análisis de resultados.

3.3.1. Descripción del comportamiento de las localidades como áreas mayores naturales

Para iniciar el análisis por estructura geográfica, es importante mencionar, que las variables de interés corresponden a los indicadores que representan las dimensiones del IPM y no al índice agregado. Sin embargo, sólo se encuentran disponibles en las publicaciones de la SPD mediante mapas, los resultados del índice agregado. En todo caso, el análisis del IPM, permite además de evidenciar cambios en el tiempo y en la distribución geográfica, generar el supuesto de la presencia grupos de UPZ's no necesariamente dados por las localidades. Para una mejor comprensión de los mapas, en todos los casos se incluye un croquis de apoyo para realizar el análisis. La localidad de Sumapaz (20), ubicada en el extremo sur de la ciudad, no se encuentra dentro de la población objetivo pues está catalogada como una UPZ completamente rural, razón por la cual no está mapeada su información.

El Índice de Pobreza Multidimensional (IPM), combina la incidencia y la intensidad de la pobreza, es decir, por un lado, considera el porcentaje de hogares pobres porque carecen de al menos el 30 % de las 15 condiciones ponderadas que son evaluadas en Colombia,

corregido por el promedio de condiciones que carecen de forma simultánea los hogares pobres; esto da a significar que no es lo mismo un hogar que es pobre por carecer de 3 condiciones que serlo por carecer de 10, por ejemplo (Bernal, L.K. *et al.* (2011)).

De la figura 3.1, se puede evidenciar un comportamiento diferencial que no necesariamente se encuentra dado por la localidad sino por grupos de localidades con un patrón geográfico y temporal. En 2011, se puede ver como la zona sur de Bogotá tiene los indicadores más desalentadores y es separado del Norte (mejores IPM's), por un grupo de localidades geográficamente ubicado en la zona central de la capital. En el ámbito temporal, se observa que con el tiempo aunque las condiciones de los habitantes de Bogotá, han tendido a mejorar, los grupos de localidades con mayor pobreza aún lo son: Ciudad Bolívar, Bosa, Tunjuelito, Rafael Uribe Uribe, Los Mártires y San Cristóbal.

Sin embargo, al analizar el porcentaje de hogares pobres, figura 3.2, así como la intensidad de la pobreza, figura 3.3, la estructura de agrupamiento geográfico no se conserva, ni tienen un comportamiento claro en el tiempo, por ejemplo, en el caso del porcentaje de hogares pobres a 2011, Suba presenta un comportamiento similar al de la zona centro. En cuanto a la intensidad de la pobreza, se observa que Usaquén, aunque no tiene un porcentaje alto de pobres, los que lo son, carecen de forma simultánea de tantas carencias promedio como los de la zona sur más críticos.

Estos aspectos son claves en el modelamiento porque sustentan el hecho de no utilizar estructuras de covarianzas dada la posición en el espacio o tiempo. Para analizar estructuras temporales o geográficas por UPZ's se tienen restricciones principalmente debidas a disponibilidad de información. En el caso de mediciones en el tiempo, solamente la encuesta multipropósito 2011 incluyó dentro de las variables observadas el CHIP catastral que identifica de manera única la vivienda, permitiendo así asociar las respectivas UPZ's. Finalmente, a pesar de contar con la identificación de UPZ en el 2011, se carece de la georreferenciación de manzanas catastro y su homologación a manzanas DANE.

En resumen, en principio puede suponerse que la localidad al ser un área mayor, es la variable asociada a los efectos aleatorios, no obstante, se observa viable considerar agrupaciones como lo sugiere Torkashvand E.,*et al.* (2017) quienes implementan métodos de clasificación entorno a SAE como mecanismo para mejorar la estimación; o mediante modelamiento de mixturas finitas, como se plantea abordar en este documento, asumiendo que se desconoce la pertenencia de cada área pequeña a las componentes de la mixtura, siguiendo la propuesta de Grüe B.*et al.* (2012) para la distribución beta.

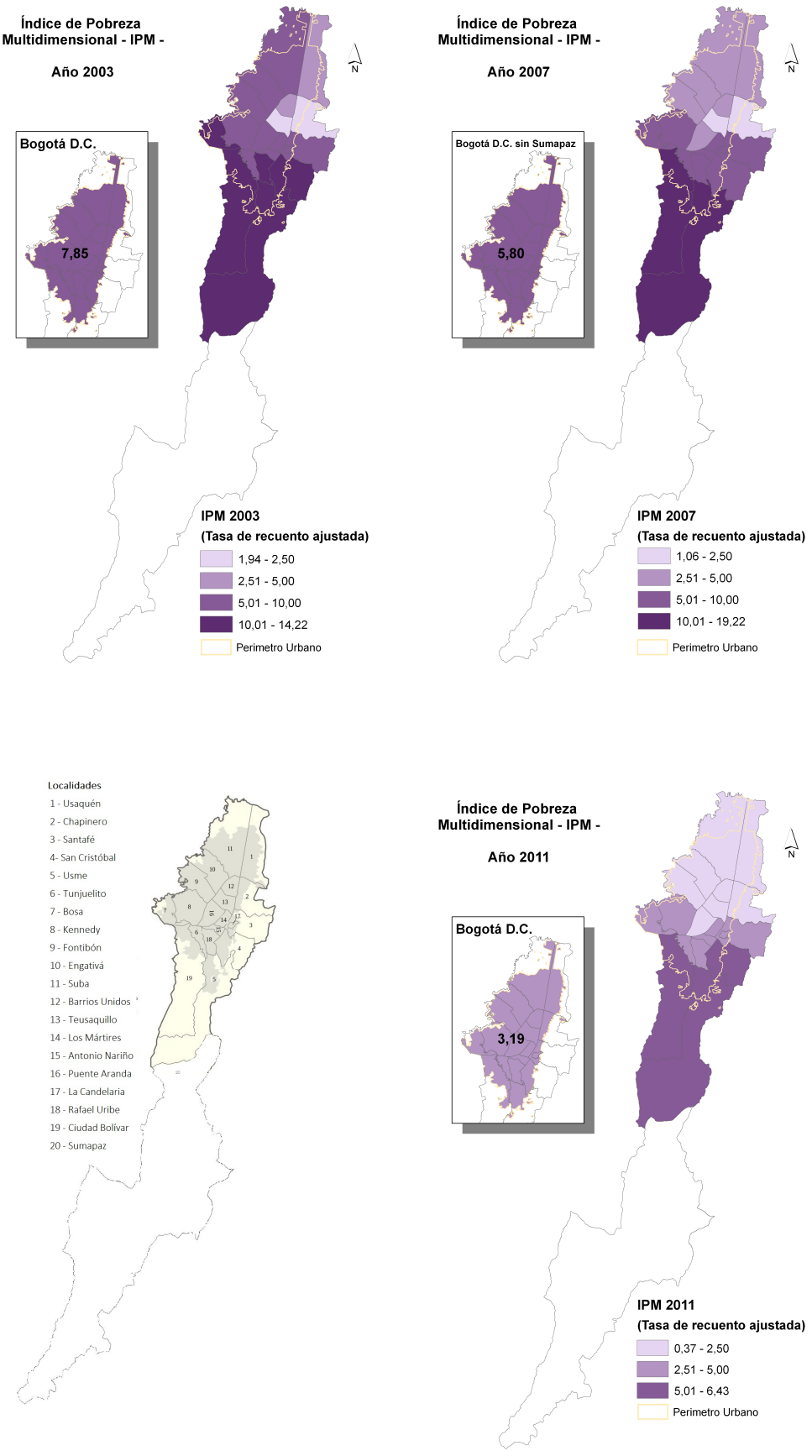


FIGURA 3.1. Índice de Pobreza Multidimensional. Bogotá.

Fuente: SDP, ECV (2003–2007), EMB 2011.

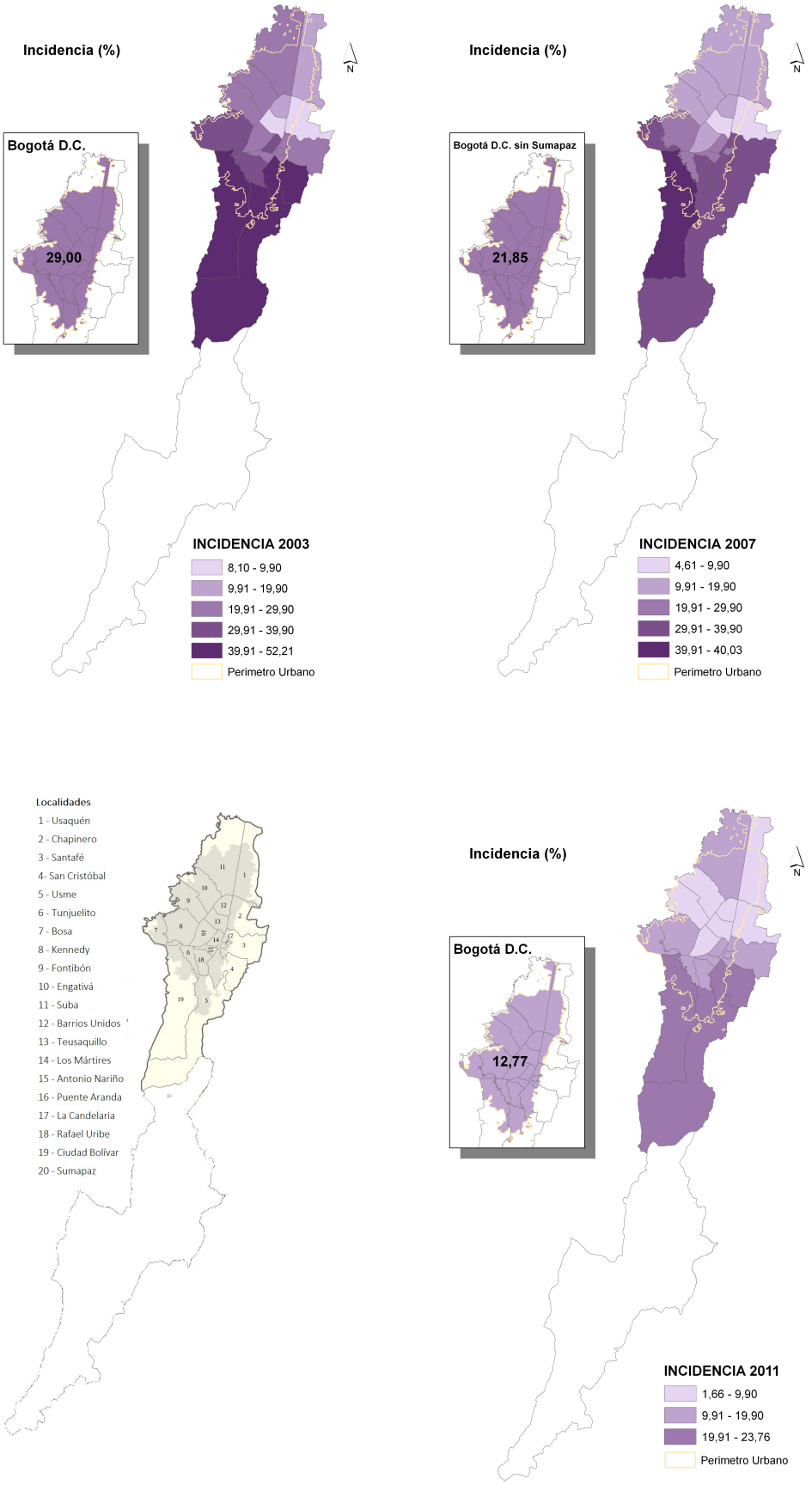


FIGURA 3.2. Incidencia de Hogares Pobres. Bogotá.

Fuente: SDP, ECV (2003–2007), EMB 2011.

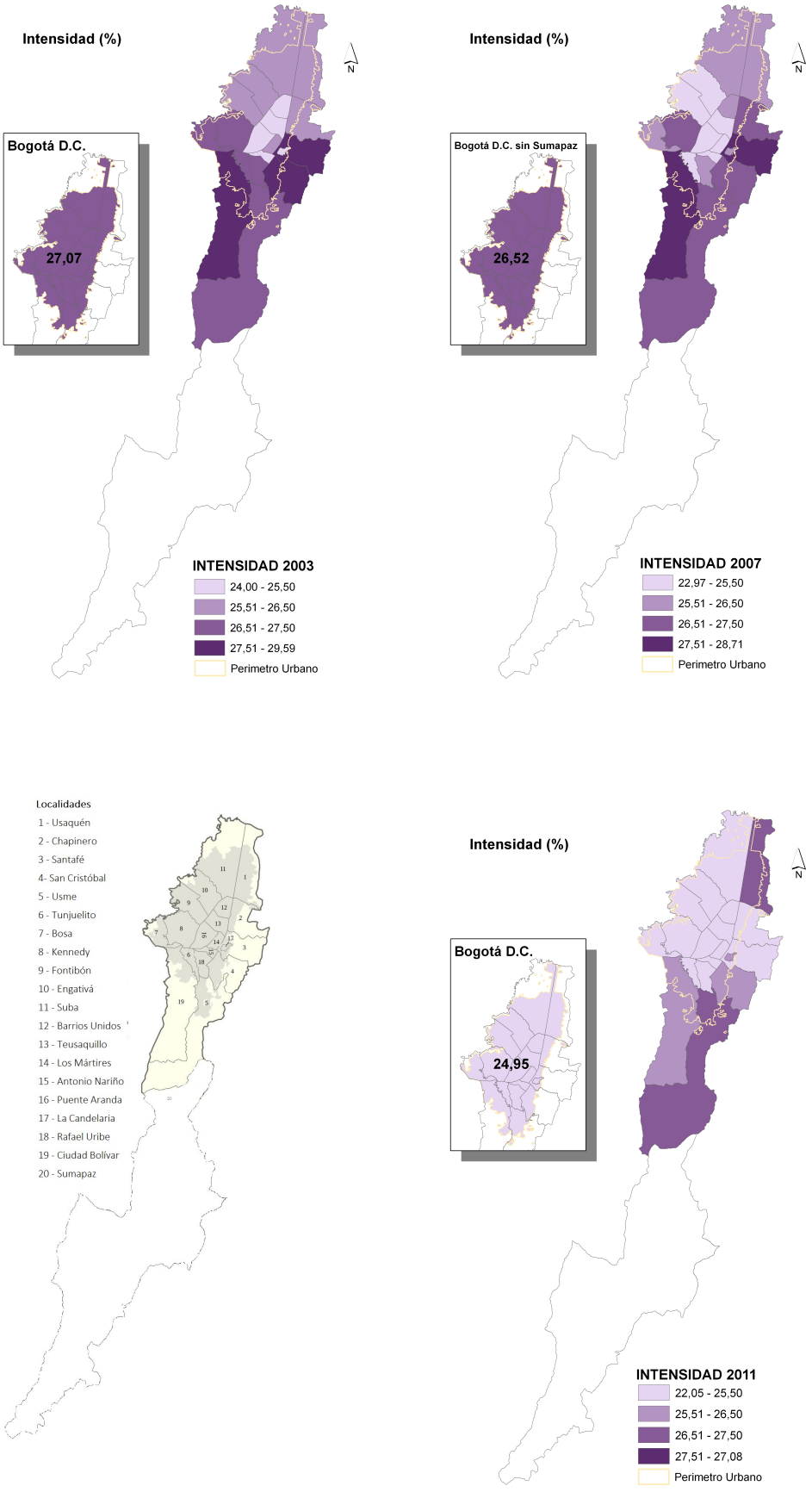


FIGURA 3.3. Intensidad de la pobreza. Bogotá.

Fuente: SDP, ECV (2003–2007), EMB 2011.

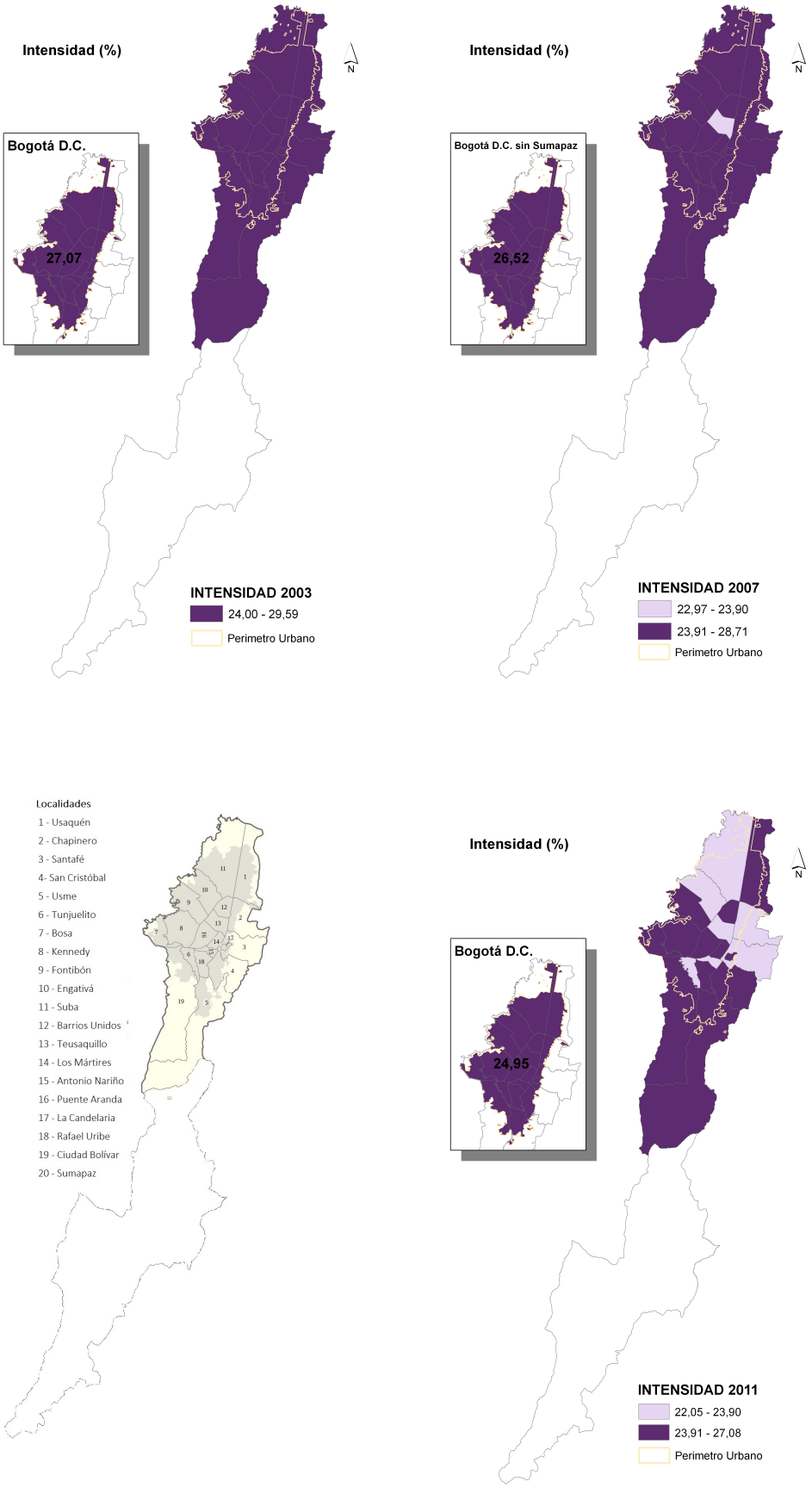


FIGURA 3.4. Intensidad de la pobreza con amplitud de rangos. Bogotá.

Fuente: SDP, ECV (2003–2007), EMB 2011.

3.3.2. Descripción de la información de las UPZ's como áreas pequeñas de interés

A partir de la descripción del comportamiento del IPM por localidades, consideradas las áreas mayores naturales, se describe brevemente, el comportamiento de las UPZ's que constituyen las áreas pequeñas de interés. En la tabla B.1 se verifica que los porcentajes de niños y jóvenes con carencias en los hogares bogotanos por UPZ's, son positivos y no nulos, cumpliendo con las condiciones de soporte de variables en la distribución beta. Así mismo, el histograma la figura B.1, presenta una leve asimetría negativa que, como se indicó en el marco teórico, es un comportamiento tradicional en este tipo de variables y por lo tanto intentar modelarlo asumiendo normalidad puede generar conclusiones erróneas.

Por otro lado, al analizar el boxplot, que no incluye las localidades con tamaños mínimos de muestra como La Candelaria (1 UPZ), Antonio Nariño, Los Mártires y Tunjuelito (2 UPZ's cada una), en la figura B.2, se observa que las estructuras de las UPZ's al interior de las áreas mayores constituidas territorialmente (localidades) reflejan heterocedasticidad, adicionalmente, se presentan datos atípicos, lo cual podría apoyar la estrategia de agrupar las áreas pequeñas en clases diferentes a los territorios, mediante mixturas finitas o una clasificación previa para el análisis bayesiano. En cuanto a relaciones de las variables, en la matriz de dispersión (figura B.3), se aprecia que las condiciones de carencias educativas del hogar, tienen relación lineal positiva con las condiciones de carencia de la niñez y juventud, así como con las carencias laborales y de salud. Las condiciones de la vivienda parecen no tener relación con las demás variables, sin embargo, presenta algunos datos con comportamiento atípico. Esto sugiere que, posiblemente en el modelo la inclusión de todas las variables redunde información, la matriz de correlaciones lineales se presenta en la tabla B.2.

Finalmente, sobre coberturas de información se tiene que Bogotá a la fecha de la encuesta contaba con 120 UPZ's, sin embargo, la muestra está conformada por 104 en razón a que 8 están en zona rural, incluso la localidad de Sumapaz con una única UPZ, no hace parte del objeto de estudio, dirigido únicamente a las zonas urbanas. Por otro lado, el marco muestral estuvo conformado por predios en zonas con predominio residencial, lo que motivó a que las otras 8 UPZ's no hayan sido seleccionadas: Paseo los libertadores (Usaquén), Parque Entrenubes (Usme), Aeropuerto El Dorado (Fontibón), Jardín Botánico (Engativá), La Academia y Guaymaral (Suba), El Mochuelo y Monteblando (Ciudad Bolívar), quienes tienen uso de suelo predominantemente dotacional. Esto implica que no hay áreas pequeñas faltantes o no muestreadas, para las que se quisiera realizar predicción; adicionalmente, la localidad de La Candelaria está conformada por una única UPZ, esto quiere decir que ella en si misma tiene un estimador directo con precisión adecuada, por lo tanto, se excluye para el análisis de áreas pequeñas. En la tabla B.3 se encuentra la distribución del número de las áreas pequeñas por localidad, así como los detalles de inclusión/exclusión en la muestra.

3.3.3. El parámetro y la composición de las variables propuestas para el modelo

De acuerdo con la justificación conceptual del modelo para la aplicación, se observa un cambio entre los mapas de la figura 3.3 y los mapas de la figura 3.4, debida solamente a la amplitud de los rangos, sin embargo, ampliar los rangos en el último caso, permite

sugerir que, en 2003 los hogares pobres eran "igualmente pobres", mientras que a 2011 un grupo de localidades logró que su población tuviera menos número de carencias en promedio, siendo un proxy de desigualdad basada en capacidades que, como se mencionó, es el enfoque que busca atender el IOH.

Las áreas pequeñas están constituidas por las Unidades de Planeamiento Zonal (UPZ) y el propósito es obtener estimadores puntuales de un parámetro de interés, para cada una de ellas con precisión adecuada; como ya se ha indicado, la heterogeneidad debida a agrupaciones de UPZ's puede ser tratada mediante el modelamiento en mixturas finitas o considerando las áreas mayores (localidades) como efectos aleatorios, bajo modelos de efectos mixtos. Incluso, considerando que el modelamiento de variables que siguen una distribución beta permiten modelar además de la media, la dispersión, se plantean los modelos de estimación clásica beta y beta bayesiano (libres de estructuras de grupos), para analizar si el modelamiento de la dispersión, es una condición suficiente, por lo menos para este caso, que permita capturar completamente la heterogeneidad presente en los datos.

Con base en lo anterior, la variable de interés en éste caso es el porcentaje de hogares con carencias en las condiciones de la niñez y la juventud ***(Ninez:N)*** que se estimará mediante las técnicas SAE basada en modelos. Dada la naturaleza del parámetro a estimar, se modela de acuerdo con la distribución beta en función de las condiciones educativas ***(Educacion:E)***, de salud ***(Salud:S)***, de trabajo ***(Trabajo:T)*** y de vivienda ***(Vivienda:V)*** del hogar, siguiendo el marco del IOH, que los plantea como factores potenciales de desigualdad en la población.

- **Porcentaje de hogares con privación en condiciones de la niñez y la juventud:** Dentro de la población infantil y hasta los 17 años, se evalúa la pobreza en ésta dimensión de acuerdo a cuatro condiciones sobre las que se espera la población no se encuentre en privación o con carencia: Asistencia escolar, cursar el grado para la edad, acceder a salud y nutrición adecuada en la primera infancia, así como, no encontrarse realizando actividades catalogadas como trabajo infantil. De esta manera, la variable se calcula como un cociente entre la cantidad de hogares con al menos una carencia en las condiciones evaluadas y el número de hogares con población de referencia.

  Las condiciones evaluadas son:

  1. Logro educativo: El hogar se encuentra en privación de la condición de logro educativo, si tiene al menos un niño entre 6 y 16 años que no asista a una institución educativa.
  2. Sin rezago escolar: Existe carencia de esta condición en el hogar, si al menos uno de sus miembros con edad entre 7 y 17 años, tiene menos años aprobados de acuerdo a la norma nacional para su edad.
  3. Atención integral para la primera infancia: Hogares con al menos un niño en primera infancia (0 a 5 años), que no acceda de manera simultánea a servicios de cuidado integral, a saber: salud, nutrición adecuada y educación inicial, se consideran con carencia AIPI.
  4. Libre de trabajo infantil: Un hogar se considera en privación o carencia, si existe por lo menos un miembro en condición de trabajo infantil, es decir, los niños

que realicen oficios del hogar por más de 15 horas a la semana, niños hasta los 14 años que trabajen (ocupados) y niños de 15 a 17 años que realicen trabajo no ligero.

- **Porcentaje de hogares con privación en condiciones de Educación del hogar:** Es el cociente entre la cantidad de hogares con carencias en al menos una de las dos condiciones evaluadas en la dimensión y el número de hogares, las condiciones son:

  1. Logro educativo: Un hogar es considerado en carencia si tiene menos de 9 años promedio de educación, este umbral se define de acuerdo con norma nacional, teniendo en cuenta que hasta el grado noveno se adquieren las competencias mínimas para insertarse en el mercado laboral y se calcula para la población de 15 años y más.
  2. Alfabetismo: Si al menos una persona de 15 años y más no sabe leer ni escribir, el hogar es considerado en carencia de la condición.

- **Porcentaje de hogares con privación en condiciones de trabajo:**Es el cociente entre la cantidad de hogares con carencias en al menos una de las dos condiciones de la dimensión y el número de hogares; específicamente, las condiciones evaluadas en la dimensión son:

  1. Sin desempleo de larga duración: Un hogar es considerado en carencia de la condición, si al menos una persona desempleada, lo ha estado por más de 12 meses, o si el hogar carece de población económicamente activa.
  2. Empleo formal: Un hogar se considera en carencia de empleo formal si al menos un ocupado se encuentra sin afiliación a pensiones, considerada como una proxy de formalidad. Adicionalmente, si el hogar tiene población en situación de desempleo o no tiene población económicamente activa, es considerado en carencia de la condición. Para evitar duplicidad de conteos se excluyen los desempleados de larga duración, así como los menores de 18 incluidos en trabajo infantil.

- **Porcentaje de hogares con privación en condiciones de salud:** Es el cociente entre la cantidad de hogares con carencias en al menos una de las dos condiciones de salud evaluadas y el número de hogares, las condiciones son:

  1. Aseguramiento: Si alguno de los miembros del hogar mayores de 5 años, no cuenta con servicio de afiliación a salud, el hogar es considerado en carencia de ésta condición.
  2. Acceso a servicios dada una necesidad: Si algún miembro del hogar en los últimos 30 días enfrentó una enfermedad, accidente, problema odontológico o alguna otra dificultad de salud que no haya implicado hospitalización y que para tratarla no acudió a un médico general, especialista, odontólogo, terapista o institución de salud, el hogar se encuentra en carencia de ésta condición.

- **Porcentaje de hogares con privación en condiciones de vivienda:** Es el cociente entre la cantidad de hogares con carencias en al menos una de las cinco condiciones de vivienda y el número de hogares, las condiciones son:

  1. Acceso a fuente de aguas mejoradas: El hogar carece de ésta condición si no cuenta con servicio de acueducto.

2. Eliminación adecuada de excretas: Si el hogar no cuenta con servicio de alcantarillado, presenta carencia en la condición.
3. Pisos adecuados: Si el hogar tiene pisos en tierra, se encuentra en carencia de pisos adecuados.
4. Paredes exteriores adecuadas: Se consideran que el hogar tiene carencia de paredes exteriores adecuadas, si al encontrarse en zona urbana los materiales de las paredes exteriores son en madera burda, tabla, tablón, guadua, otro vegetal, zinc, tela, cartón, desechos o están sin paredes.
5. Sin hacinamiento crítico: Un hogar urbano se considera en privación si hay en el hogar tres o más personas por cuarto.

CAPÍTULO 4

Estimadores de proporciones para áreas pequeñas basados en el modelo Beta. Aplicación a datos de condiciones de la niñez y juventud

En este capítulo se aborda la estimación de proporciones para áreas de los modelos presentados en el marco teórico del capítulo 2. En primera instancia, se desarrollan los relacionados con el enfoque clásico, puesto que éste enfoque da origen al método de estimación **EBLUP** de SAE. En la sección 4.1 se inicia con el modelamiento en mixturas finitas de acuerdo con la teoría presentada en la sección 2.3, bajo la consideración que, SAE incorpora las estruturas de grupos a partir de modelos mixtos y aunque son tratamientos distintos, las mixturas finitas como se ha indicado podría ser una estrategia para la reducción de los errores asociados al modelo y de acuerdo con el análisis descriptivo, dicha agrupación toma sentido dada la estructura de la información. Posteriormente, se implementa el modelo beta clásico tratado de forma general en la sección 2.2, caso para el cual los datos se asumen libres de estructuras agrupadas con el objeto de analizar si, mediante el modelamiento del parámetro de dispersión es posible capturar toda la variabilidad contenida en los datos, los resultados relativos a este modelamiento se presentan en la sección 4.2.

Ahora bien, para obtener estimadores **HB** de la teoría de SAE, en la sección 4.3 se implementa el modelo beta bayesiano que no incorpora estructura de grupos, siguiendo la teoría de la sección 2.4, con el propósito de darle continuidad a los resultados del mismo planteamiento (libre de grupos), del enfoque clásico. La inclusión de efectos aleatorios desde el enfoque bayesiano se presenta en la sección 4.4, cuya extensión obedece a la implementación de varios escenarios de modelamiento; en primer lugar, dados por la naturaleza de los efectos aleatorios, en donde se evalúa los debidos a las localidades y los debidos a agrupaciones de UPZ's, con base en un análisis de clasificación, siguiendo a Torkashvand et al. (2017) quienes utilizan distancias euclidianas en el contexto del EBLUP, para mejorar la estimación del error cuadrático medio. En segundo lugar, los escenarios se construyen de acuerdo a la incorporación de efectos mixtos solo para la media, asumiendo parámetro de dispersión constante; y los demás escenarios corresponden a variantes para el modelamiento del parámetro de dispersión.

En todos los casos se validan los supuestos sobre los residuales y en estimación bayesiana se realizan los diagnósticos de convergencia. Es de resaltar que se implementan para la aplicación los cálculos de los errores ponderados por la matriz Hat, descritos en la sección 2.6.2, debido a que los paquetes utilizados para el caso de modelos de mixtura finita y modelos mixtos bayesianos no incluían resultados para el respectivo análisis al momento de la implementación. Por otro lado, el paquete usado para el modelo beta bayesiano, aunque incorpora el cálculo dentro del proceso de estimación, presentaba un leve error de programación, por lo que se calcularon también de forma externa al paquete. Adicionalmente, las gráficas de bandas simuladas para la validación del supuesto de normalidad no estaban disponibles en los paquetes utilizados por lo que también su desarrollo representa aportes para el presente trabajo.

El capítulo concluye en la sección 4.5, con la comparación de los modelos con base en la metodología de SAE para la estimación de los errores estándar de los estimadores de cada área, presentada en la sección 1.4, como resultado se selecciona el modelo que presenta los menores errores estándar; adicionalmente, se presenta una expresión analítica para el cálculo de las medidas de precisión, teniendo en cuenta que, dadas las características de la distribución beta en la familia exponencial biparamétrica, es posible realizar la estimación de forma directa sin requerir técnicas de remuestreo.

4.1. Estimador de Proporciones en Mixturas Finitas Beta. EMFB

Para ajustar el modelo del parámetro de proporciones mediante mixturas finitas, se considera que se desconoce la composición de los grupos de áreas pequeñas como se desarrolló en el capítulo 2, específicamente en la sección 2.3. Se estiman diferentes modelos para que, de acuerdo con los criterios de selección, escoger el que mejor se adecúa a la estructura de los datos. El modelo general planteado corresponde al modelamiento de la media μ_d y dispersión ϕ_d con todas las variables explicativas. Así:

$$logit(\mu_{dk}) = \beta_{0k} + \beta_{1k}Educacion_d + \beta_{2k}Trabajo_d + \beta_{3k}Salud_d + \beta_{4k}Vivienda_d,$$
$$log(\phi_{dk}) = \gamma_{0k} + \gamma_{1k}Educacion_d + \gamma_{2k}Trabajo_d + \gamma_{3k}Salud_d + \gamma_{4k}Vivienda_d,$$

donde $k = 1, \ldots, K$ que a su vez varía entre todos los tamaños de componentes posibles $K = 1, \ldots, 19$ con el fin de evidenciar si la partición puede estar dada por las 19 localidades. El procedimiento de estimación EM para mixturas finitas es implementado a través del paquete betareg del Software R – Project, con la función *betamix*, para el cual bajo el modelo general no se logra convergencia del algoritmo incluso para componentes de tamaño 2.

En la tabla B.4, se relacionan los modelos ajustados a partir de las variables en distintos escenarios de inclusión y exclusión, indicando en qué casos no se alcanzó convergencia, y en caso de alcanzarla sus criterios para selección BIC y AIC. El modelo que mejor comportamiento tiene para estos datos es el M_{14}, expresado así;

Modelo MFB (Mixtura Finita Beta):

$$\begin{aligned} logit(\mu_{dk}) &= \beta_{0k} + \beta_{1k} Educacion_d, \\ log(\phi_{dk}) &= \gamma_{0k} + \gamma_{1k} Educacion_d. \end{aligned} \tag{4.1}$$

Contrastando con los resultados del análisis descriptivo, la educación tiene una asociación lineal positiva con las demás variables y por lo tanto puede recoger la información de las restantes. La mejor partición encontrada mediante la estimación Esperanza Maximización para clases latentes bajo el modelo de la ecuación (4.1) está dada por tres componentes $k = 1, 2, 3$ con los siguiente coeficientes estimados y errores estándar:

	Componente 1		Componente 2		Componente 3	
μ_d	$\widehat{\beta}_{i1}$	Error Est.	$\widehat{\beta}_{i2}$	Error Est.	$\widehat{\beta}_{i3}$	Error Est.
Intercepto	-2.49	0.102	-1.55	0.075	-0.664	0.030
Educacion	2.69	0.158	2.18	0.146	1.051	0.050
ϕ_d	$\widehat{\gamma}_{i1}$		$\widehat{\gamma}_{i2}$		$\widehat{\gamma}_{i3}$	
Intercepto	12.075	1.020	3.00	0.364	2.682	0.934
Educacion	-27.88	1.822	5.84	0.953	16.60	3.069

Nota: Todos los coeficientes son significativos a cualquier nivel.

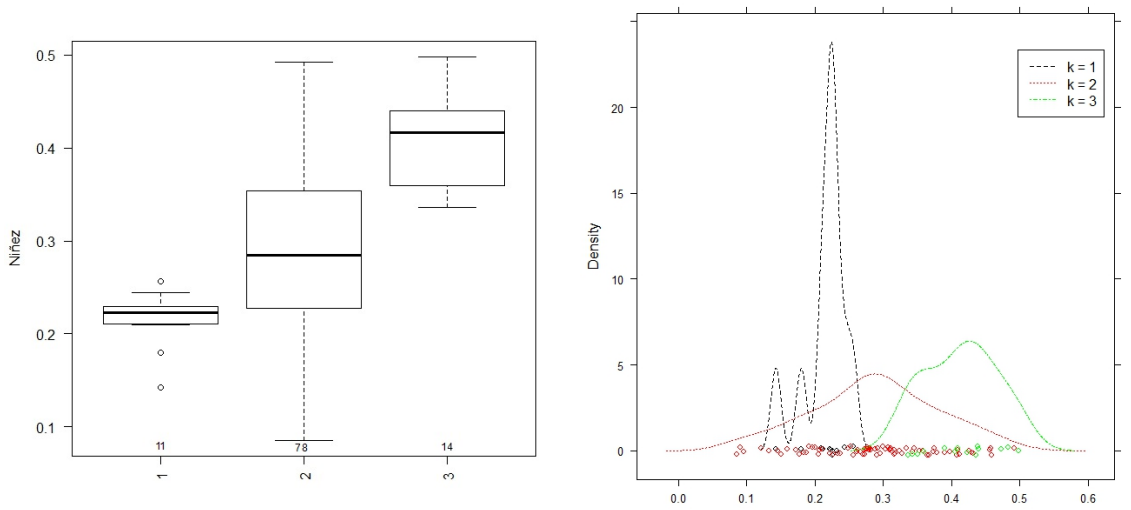
La convergencia es alcanzada después de 88 iteraciones

Tamaños de componentes: 11/78/14 – Criterios: BIC =-239.317; AIC=-276.203

TABLA 4.1. Resumen Modelo Mixturas Finitas Beta (MFB).

La estructura de los grupos de UPZ's de acuerdo con las 3 componentes a posteriori para cada localidad se presenta en la tabla B.5, encontrándose que, localidades como Usaquén, Kennedy, Fontibón, Suba y Teusaquillo tienen UPZ's distribuidas en las tres grupos, lo que evidencia fuente de heterocedasticidad adicional al interior de las mismas. Por otro lado, el boxplot de la figura (4.1a) muestra la segunda componente como la que tienen mayor dispersión pero es aproximadamente simétrica, mientras que la componente tres con menor dispersión, es asimétrica a la izquierda. Por su parte, la componente uno, agrupa UPZ's con bajos porcentajes de niños y jóvenes en carencias y porcentajes de carencias en educación de hasta el 45 %, presentando tres datos atípicos, que a su vez influyen sobre la gráfica de densidades (Ver figura (4.1b)), correspondientes a las localidades de Parque el Salitre y Kennedy Central, con porcentaje de niños y jóvenes en carencia de 14.3 % y 17.9 % respectivamente. La UPZ Usaquén de la misma localidad, presenta el dato atípico por encima del grupo (25.6 %).

En la figura 4.2 se presentan las rectas ajustadas para las medias de cada componente (k=1,2,3.), encontrándose que la primera componente es influenciada por UPZ's con carencias educativas del hogar aproximadamente altas (30 - 50 %) pero con carencias en las condiciones de los niños y jóvenes bajas, lo que desde el punto de vista del IOH es un comportamiento deseado, que se esperaría refleje a futuro movilidad social.



(A) Boxplot. (B) Densidades.

FIGURA 4.1. Niñez y juventud por componentes de la mixtura del modelo MFB.

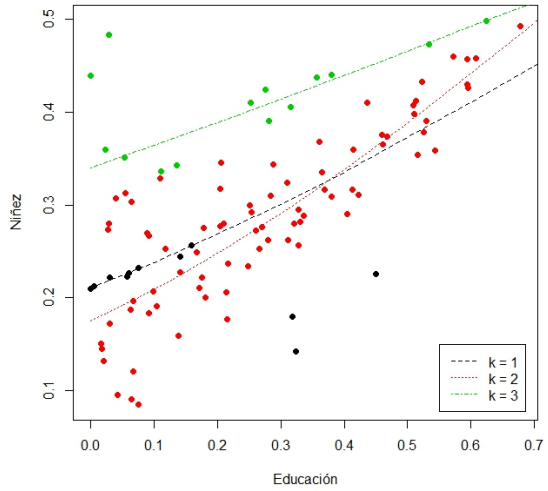


FIGURA 4.2. Ecuaciones ajustadas para las medias, según las componentes de la mixtura en el modelo MFB.

Para validar el modelo, se realiza el cálculo de los residuales Pearson re_d^p (ver sección 2.6.2), en la figura 4.3 que el autocorrelograma evidencia independencia y la gráfica de cuantiles sugiere que la distribución empírica de los errores puede aproximarse a la Normal. Específicamente, el test de *Ljung Box* sobre independencia presenta un p–valor de 0.43 y el test de *Shapiro Wilk* para normalidad de 0.40. En cuanto a la gráfica de residuales versus medias ajustadas, al carecer de patrón puede interpretarse que no es necesaria más información que modele la dispersión.

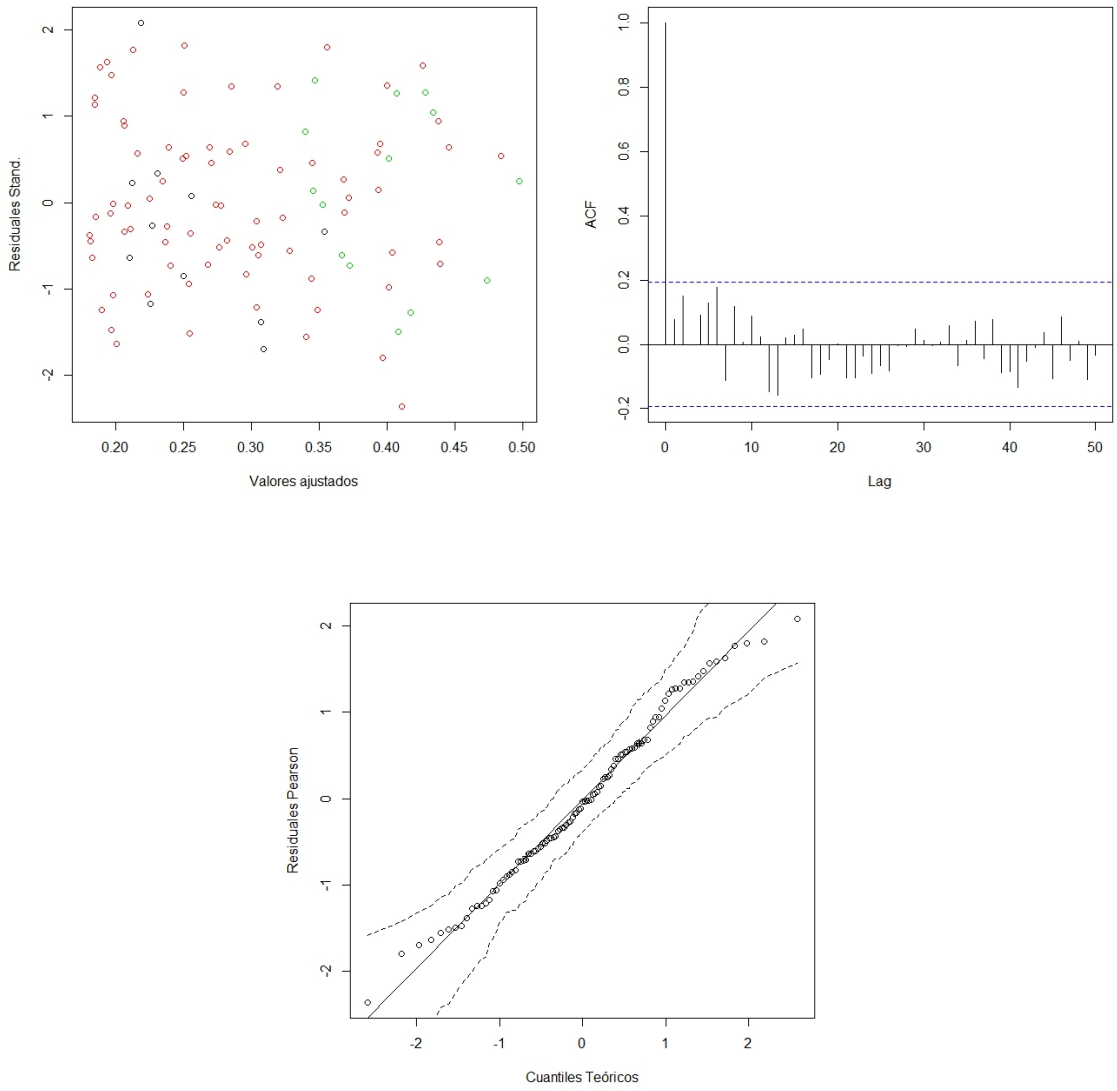


FIGURA 4.3. Gráficas de residuales de Pearson sobre el modelo MFB.

Finalmente, considerando los residuales ponderados re_d^w sugeridos para validación de supuestos en regresión beta, se encuentran resultados análogos a los residuales Pearson, con p–valores de 0.42 y 0.63 en los test *Ljung Box* y *Shapiro Wilk* de independencia y normalidad respectivamente. La figura 4.4 presenta las pruebas gráficas de supuestos del modelo.

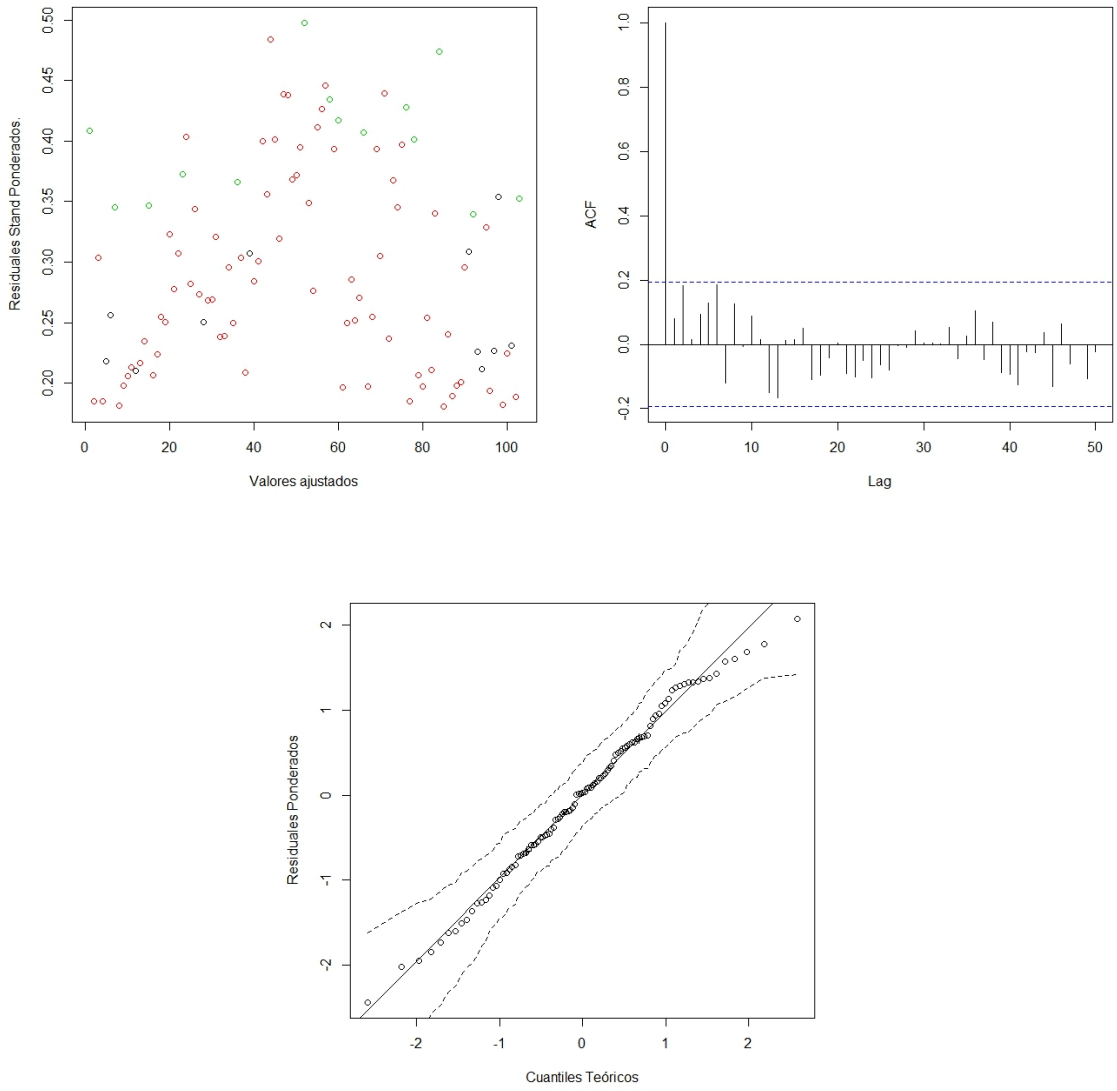


FIGURA 4.4. Gráficas de residuales ponderados sobre el modelo MFB.

En el apéndice B.3.3, se presenta el código diseñado en R–Project para el cálculo de residuales en razón a que el paquete *betamix* implementa el procedimiento de estimación más no de validación de supuestos.

4.2. Estimador de Proporciones Máxima Verosimilitud Beta. EMVB

Como se refirió en la introducción del capítulo, se implementa el modelo beta clásico definido en la sección 2.2, considerando que, el planteamiento del modelamiento conjunto de media y dispersión en la distribución beta, permite definir completamente el comportamiento del parámetro de interés, en este sentido, el propósito es analizar la capacidad del modelo para recopilar toda la información de variabilidad que no es modelada a través

de la media, sin incorporar de grupos o componentes como en el caso de mixturas finitas.

Se considera así, el modelo que ajusta para cada UPZ los mismos interceptos y pendientes dada la variable educación:

Modelo MVB (Máxima Verosimilitud Beta):

$$\begin{aligned} logit(\mu_d) &= \beta_0 + \beta_1 Educacion_d, \\ log(\phi_d) &= \gamma_0 + \gamma_1 Educacion_d, \end{aligned} \tag{4.2}$$

μ_d	$\widehat{\beta}_i$	Error Est.
Intercepto	-1.38	0.070
Educacion	1.89	0.159
ϕ_d	$\widehat{\gamma}_i$	
Intercepto	2.80	0.236
Educacion	3.72	0.754

Número de iteraciones 16 (BFGS) + 1 (Fisher scoring)

Criterios: BIC =-248.442 ; AIC=-258.981

TABLA 4.2. Resumen Modelo Máxima Verosimilitud Beta (MVB).

La recta ajustada para el modelo de la ecuación 4.2 se presenta en la figura 4.5, en la que se observa, como se esperaba, que el ajuste genera errores más amplios que los generados sobre el modelo de la ecuación (4.1).

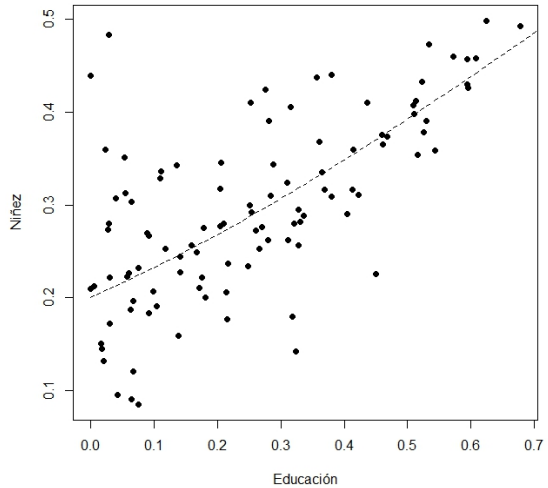


FIGURA 4.5. Ecuación ajustada para la media del modelo MVB.

No obstante, se cumplen sus supuestos ya que la normalidad probada mediante el test de *Shapiro Wilk* para los errores ponderados re_d^w presenta un p–valor de 0.127, por su parte, la independencia probada con el test *Ljung Box* tiene un p–valor de 0.508.

Finalmente, al comparar los criterios de información resumidos en la tabla (4.3) el modelo MFB puede preferirse si se considera que el BIC se encuentra penalizado para la cantidad de coeficientes estimados.

Modelo	AIC	BIC	Desvío
Mixturas Finitas Beta (MFB)	-276.203	-239.317	-294.93
Máxima Verosimilitud Beta (MVB)	-258.981	-248.442	-266.98

TABLA 4.3. Criterios de información para comparación de modelos MFB y MVB.

4.3. Estimador de Proporciones Completamente Bayesiano Beta. ECBB

Mediante los modelos planteados en las secciones anteriores, se realiza estimación vía máxima verosimilitud, cuyo análogo en SAE corresponde al método EBLUP, tratado en el capítulo 1. Mediante el enfoque completamente bayesiano es posible obtener los estimadores HB tratados en la sección 1.3, esta distinción es necesaria para evidenciar que no se tratan de estimadores obtenidos con ajuste de modelos Bayesianos Empíricos. En esta sección se ajusta el modelo beta bayesiano que al igual que el modelo beta clásico no considera existencia de grupos, y se implementa con el mismo propósito de evaluar la capacidad de recoger la variabilidad que no es modelada por el parámetro de localización.

De acuerdo con la propuesta de estimación de Cepeda–Cuervo (2001) y Cepeda *et al.* (2005) para la estimación conjunta de media μ_d y dispersión ϕ_d para variables dependientes en la distribución beta en el marco de modelos pertenecientes a la familia exponencial biparamétrica, se plantea el modelo:

Modelo CBBE (Completamente Bayesiano Beta Equivalente)

$$\begin{aligned} logit(\mu_d) &= \beta_0 + \beta_1 Educacion_d, \\ log(\phi_d) &= \gamma_0 + \gamma_1 Educacion_d. \end{aligned} \tag{4.3}$$

Se denomina ***Equivalente*** ya que coincide en forma funcional con el modelo máxima verosimilitud beta (MVB) de la ecuación (4.2), sin embargo desde el planteamiento bayesiano se asumen que los coeficientes de regresión también son variables aleatorias y por lo tanto siguen una distribución de probabilidad. Para el caso, las a priori son tales que $\beta \sim N(b^*, B^*)$ y $\gamma \sim N(g^*, G^*)$, donde b^*, B^*, g^* y G^* están definidas de acuerdo a las ecuaciones (2.10) y (2.12).

La estimación se implementa a través del paquete Bayesianbetareg del Software R – Project, con los siguientes resultados:

μ_d	$\widehat{\beta}_i$	Error Est.	Inf IC (95 %)	Sup IC (95 %)
Intercepto	-0.26	0.082	-0.417	-0.096
Educacion	2.59	0.268	2.068	3.122
ϕ_d	$\widehat{\gamma}_i$			
Intercepto	3.06	0.333	2.383	3.704
Educacion	-0.24	0.945	-2.081	1.619

Tamaño de la cadena 100.000, Burn in 0.2, Salto 30

Criterios: BIC =225.3774 ; AIC=220.1079; Desvío=216.1079

TABLA 4.4. Resumen Modelo Completamente Bayesiano Beta Equivalente (CBBE).

En la figura 4.6 se presentan los gráficos de trayectorias de las muestras generadas a partir de las distribuciones a posteriori de cada parámetro, observándose un buen comportamiento en el sentido que explora de forma alternada los estados de la cadena y los periodos de transición pequeños permiten suponer convergencia.

De acuerdo, con las pruebas de convergencia, ver apéndice (A.2.4.2), los estadísticos Geweke sobre la diferencia de medias ergódicas no son mayores que el percentil 95 de una distribución Normal, con lo cual, si bien no indica convergencia tampoco es posible concluir que no la alcance, sin embargo, dado el comportamiento gráfico de las trayectorias, así como los p-valores de la prueba Heidelberger – Welch se concluye que se alcanza convergencia dado que no se rechaza la hipótesis de estacionariedad y las longitudes medias de los intervalos de credibilidad (Halfwidth) son pequeños.

μ_d	Geweke	Heidelberger y Welch	Halfwidth
Intercepto	0.8504	0.692	0.00321
Educacion	-0.9928	0.653	0.01018
ϕ_d			
Intercepto	0.4048	0.745	0.0232
Educacion	-0.6100	0.654	0.0623

Fracciones de cadenas Geweke: 0.1- 0.5, Burn in 0.2, Salto 30

Prueba Rifter – Lewis: $q = 0.025$, $r = 0.005$, $s = 0.95$ para al menos 3.746 muestras

TABLA 4.5. Criterios de convergencia de las cadenas en el Modelo CBBE.

La verificación de supuestos de los errores ponderados por la matriz Hat re_d^w indican que es posible asumir normalidad ya que mediante la prueba *Shapiro Wilk* se obtiene un p–valor de 0.321. Caso contrario ocurre con la independencia, debido a que el p-valor asociado a la prueba *Ljung Box* es tan solo de 0.0306. Bajo esta consideración y teniendo en cuenta que, como se observa en la tabla 4.4 la variable educación no resulta significativa para modelar la dispersión en el enfoque completamente bayesiano, no es posible tener un modelo símil al obtenido mediante estimación clásica, esto tiene implicaciones por ejemplo, para la implementación directa del estimador Bayesiano Empírico.

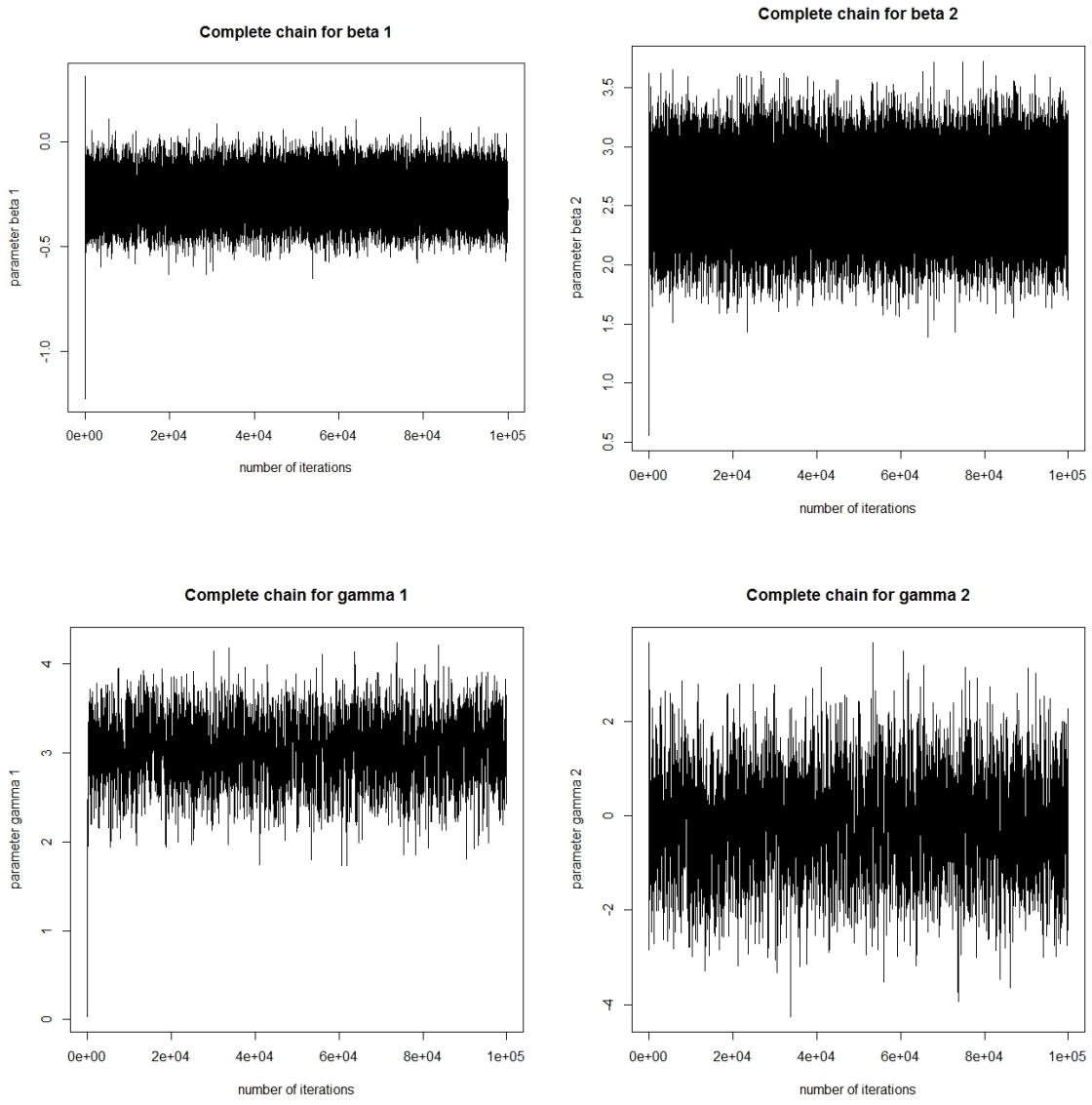


FIGURA 4.6. Gráfico de trayectorias del Modelo CBBE.

En las tablas B.6 y B.7 se presentan algunos de los modelos estimados teniendo en cuenta que la inclusión– exclusión de variables afectaban o bien la convergencia, la significancia o los supuestos. De acuerdo a lo anterior, los modelos M_{12} y M_{14} presentan convergencia y cumplen los supuestos de normalidad e independencia en los errores. Los resultados se describen a continuación, considerando en principio el modelo M_{14} ya que, de acuerdo con los criterios de selección, tienen mejor desempeño:

Modelo CBB (Completamente Bayesiano Beta)

$$\begin{aligned} logit(\mu_d) &= \beta_1 Educacion_d + \beta_2 Salud_d, \\ log(\phi_d) &= \gamma_0. \end{aligned} \tag{4.4}$$

En donde, los coeficientes y sus medidas de error estándar e intervalos de credibilidad se especifican en la tabla 4.6, las trayectorias en la figura 4.7 y las respectivas pruebas de convergencia en la tabla 4.7.

μ_d	$\widehat{\beta}_i$	Error Est.	Inf IC (95 %)	Sup IC (95 %)
Educacion	1.163	0.173	0.829	1.511
Salud	3.873	0.881	2.210	5.580
ϕ_d	$\widehat{\gamma}_i$			
Intercepto	2.321	0.130	2.054	2.567

Tamaño de la cadena 100.000, Burn in 0.2, Salto 30

Criterios: BIC =91.9668 ; AIC=86.6974; Desvío=82.697

TABLA 4.6. Resumen Modelo Completamente Bayesiano Beta (CBB).

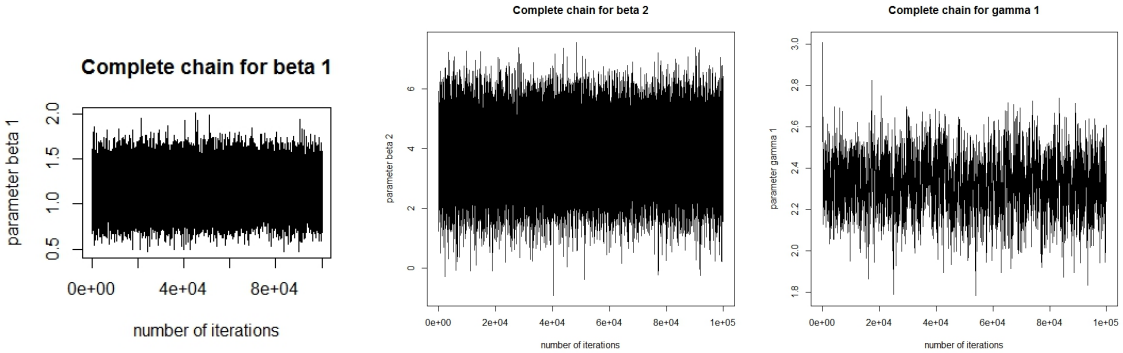


FIGURA 4.7. Gráfico de trayectorias del Modelo CBB.

μ_d	Geweke	Heidelberger y Welch	Halfwidth
Educacion	-0.0121	0.973	0.00658
Salud	0.81688	0.766	0.03344
ϕ_d			
Intercepto	-0.3047	0.539	0.011

Fracciones de cadenas Geweke: 0.1- 0.5, Burn in 0.2, Salto 30

Prueba Rifter – Lewis: q = 0.025, r = 0.005, s = 0.95 para al menos 3.746 muestras

TABLA 4.7. Criterios de convergencia de las cadenas en el Modelo CBB.

Como se ha indicado, éste modelo presenta criterios de selección mejores que cualquier otro modelo considerado en la metodología completamente bayesiana y además los p–valores de las pruebas de normalidad e independencia *Shapiro Wilk* y *Ljung Box* para los errores ponderados re_d^w, respectivamente son 0.058 y 0.198. Sin embargo, en la figura 4.8 la gráfica de dispersión de los errores ponderados versus los valores ajustados de la media, además de tener dos datos extremos, los puntos restantes tienen un comportamiento de cono invertido, indicando que el modelo no ha recogido completamente la dispersión de los datos, en otras palabras hay presencia de heterogeneidad; por su parte el gráfico de bandas simuladas para chequear la normalidad contradice el resultado de la prueba analítica de *Shapiro Wilk*, situación que puede deberse a que dado el p–valor, la hipótesis de normalidad

podría ser rechazada a un nivel de significacia del 10 %, pero además que el resultado de la prueba puede estar siendo afectado por la heterocedasticidad presente.

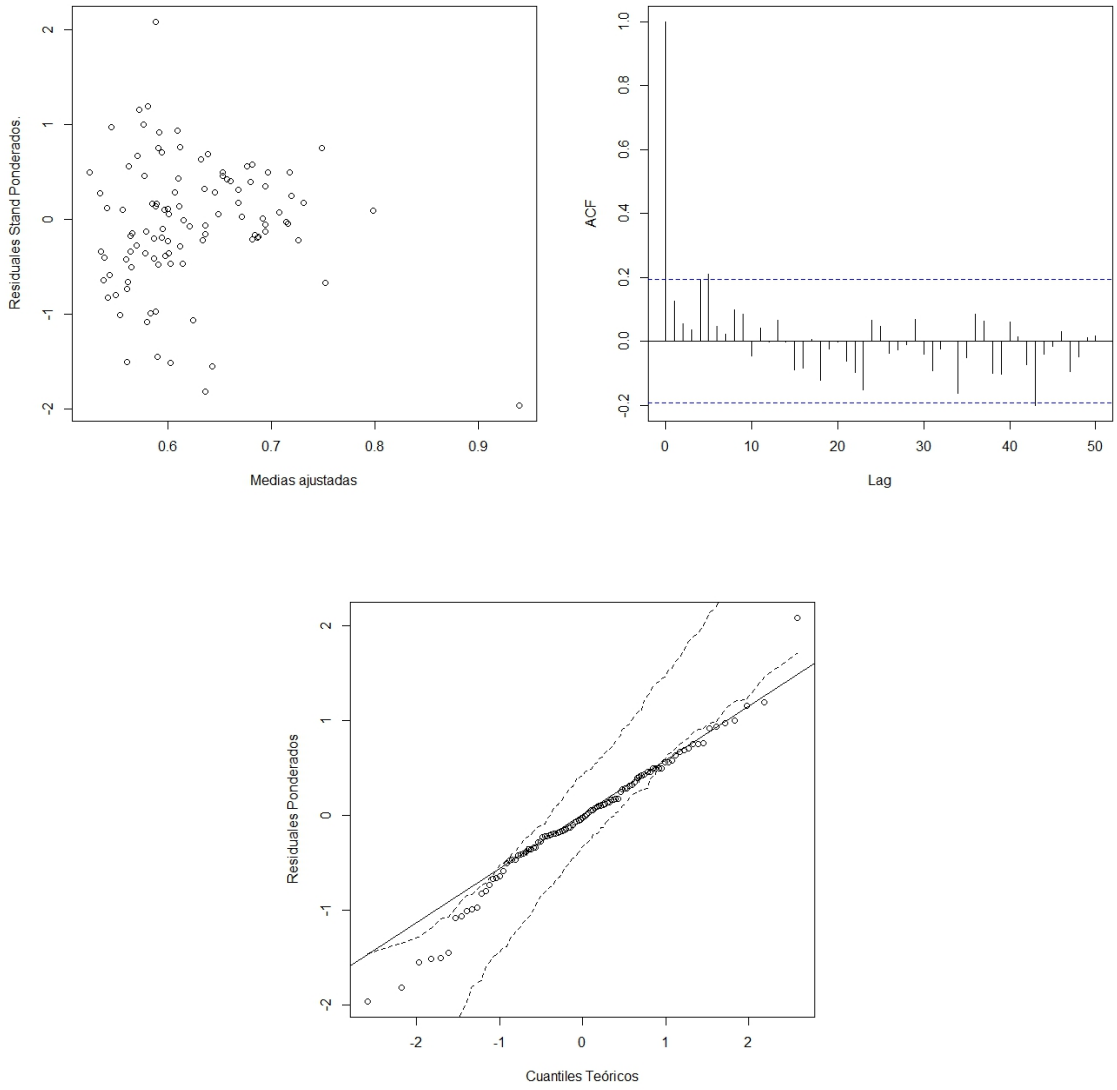


FIGURA 4.8. Gráficas de residuales ponderados sobre el modelo CBB.

Por su parte el modelo M_{12}, también cumple con las condiciones de convergencia, significancia y supuestos de normalidad e independencia en los errores, con p–valores de 0.118 y 0.509, respectivamente, así mismo, la gráfica de dispersión de los errores frente a la media ajustada presenta un comportamiento que permite suponer que la varianza presente en los datos ha sido completamente capturada por el modelo, indicando homocedastidad en los errores, el gráfico de autocorrelación incluye dentro de los umbrales de aleatoriedad todos los rezagos y el gráfico de bandas simuladas, es acorde con la prueba de *Shapiro Wilk*.

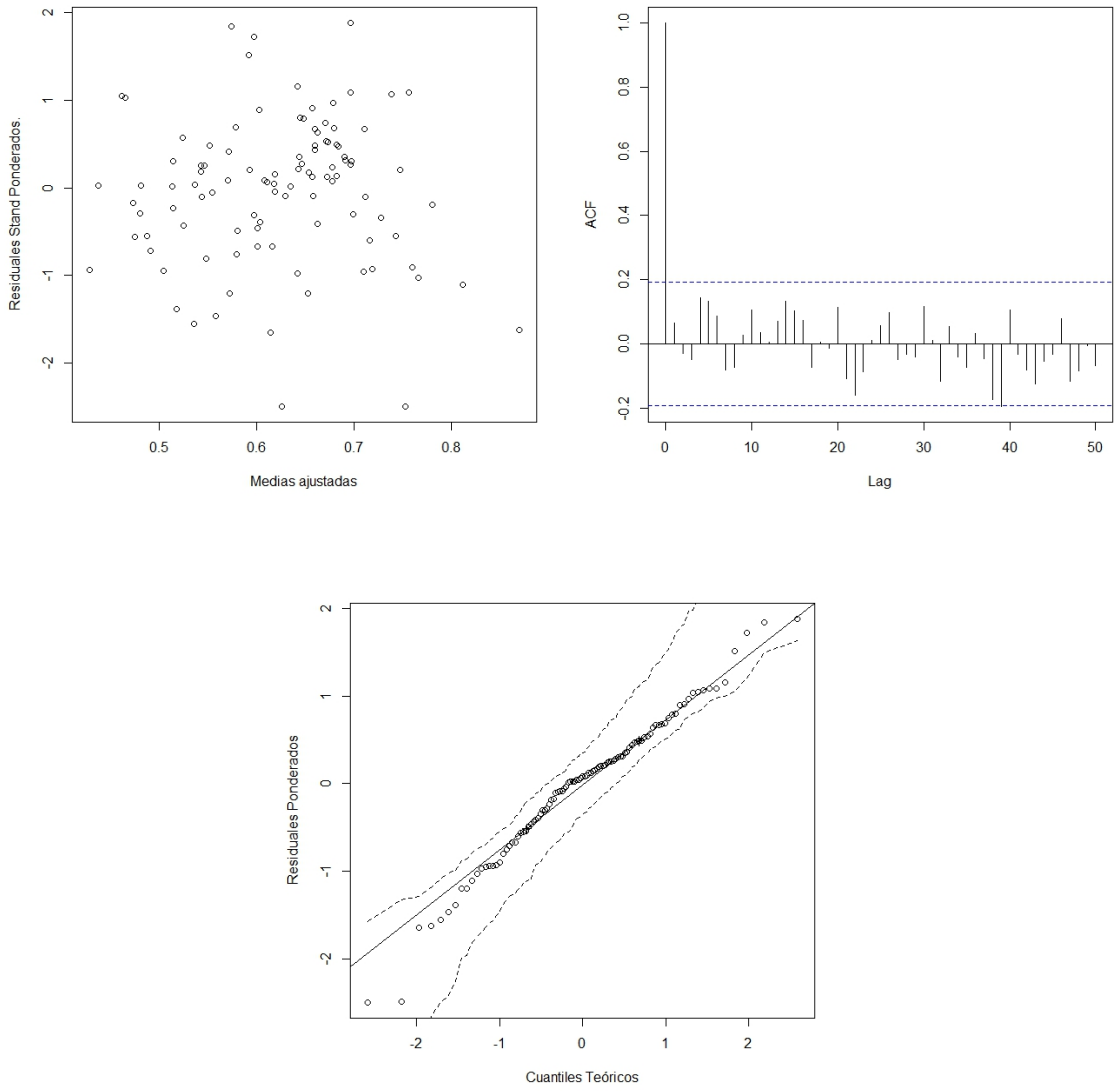


FIGURA 4.9. Gráficas de residuales ponderados sobre el modelo CBBa.

En consecuencia, el modelo M_{12} tiene resultados más confiables y por lo tanto bajo la sigla BBa sus estimaciones e información de convergencia se presenta a continuación:

Modelo CBBa (Completamente Bayesiano Beta Alterno)

$$\begin{aligned} logit(\mu_d) &= \beta_0 + \beta_1 Educacion_d + \beta_2 Trabajo_d, \\ log(\phi_d) &= \gamma_0. \end{aligned} \tag{4.5}$$

Los coeficientes estimados, errores estándar e intervalos de credibilidad se presentan en la tabla 4.8, las trayectorias y pruebas de convergencia en la figura 4.10 y la tabla 4.15, respectivamente.

μ_d	$\widehat{\beta}_i$	Error Est.	Inf IC (95 %)	Sup IC (95 %)
Intercepto	-0.489	0.087	-0.667	-0.321
Educacion	0.959	0.314	0.364	1.588
Trabajo	4.444	0.428	3.605	5.304
ϕ_d	$\widehat{\gamma}_i$			
Intercepto	2.987	0.139	2.705	3.251

Tamaño de la cadena 100.000, Burn in 0.2, Salto 30

Criterios: BIC =177.142 ; AIC=169.2378; Desvío=163.2378

TABLA 4.8. Resumen Modelo Completamente Bayesiano Beta Alterno (CBBa).

μ_d	Geweke	Heidelberger y Welch	Halfwidth
Intercepto	1.6998	0.107	0.00301
Educacion	-0.597	0.642	0.01072
Trabajo	-1.721	0.104	0.01483
ϕ_d			
Intercepto	1.062	0.713	0.00967

Fracciones de cadenas Geweke: 0.1- 0.5, Burn in 0.2, Salto 30

Prueba Rifter – Lewis: $q = 0.025$, $r = 0.005$, $s = 0.95$ para al menos 3.746 muestras

TABLA 4.9. Criterios de convergencia de las cadenas en el Modelo CBBa.

Los modelos beta bajo el enfoque bayesiano que mejor se ajustaron a los datos, corresponden a modelos de dispersión constante, cuya estructura no afecta la estimación de la precisión de cada área pequeña en SAE, debido a que la varianza de la variable aleatoria depende de la estimación de μ_d, ver ecuación (2.3).

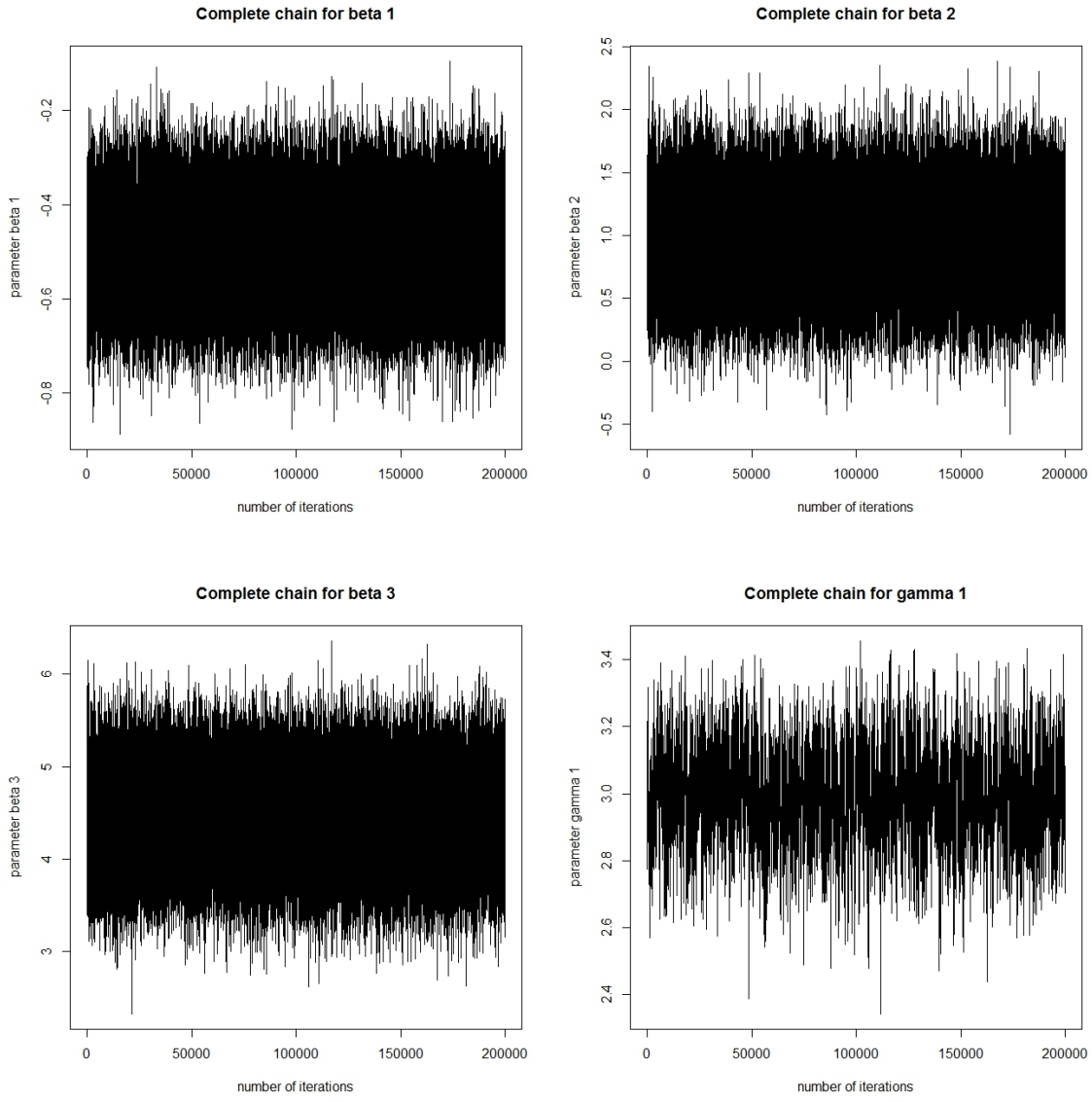


FIGURA 4.10. Gráfico de trayectorias del Modelo CBBa.

4.4. Estimador de Proporciones Bayesiano Mixto Beta. EBMB

De acuerdo con el planteamiento de SAE basada en modelos, el procedimiento de estimación establece modelos mixtos que resultan de considerar los errores asociados al diseño muestral, ver ecuación 1.1 y los asociados al modelo superpoblacional que sigue el parámetro de interés, de acuerdo con la ecuación (1.2), el rol de estos errores en el modelo consiste en generar estimadores de varianzas basadas en modelos que conduzcan finalmente a la estimación de las medidas de precisión en áreas pequeñas; en este sentido, debido al desconocimiento de la varianza de diseño, se requiere el uso de sus estimadores muestrales, lo que incorpora sesgos en la estimación; con base en ello, se encuentran en la literatura planteamientos que hacen uso del enfoque baye-

siano definiendo distribuciones a priori para modelar la varianza (You y Chapman (2006)).

Desde el punto de vista de modelos beta en la familia exponencial biparamétrica, estos inconvenientes no se presentan ya que se establecen modelos directos para los parámetros de media y dispersión, tal como se desarrolló en la sección previa 4.3. No obstante, se observa que al ajustar una única ecuación para la media, los residuales pueden ser de tamaño considerable en comparación con los obtenidos cuando se ajustan diferentes medias ante la existencia por ejemplo, de grupos latentes, tal como se desarrolló en la sección 4.1 con mixturas finitas, que incorporan los grupos a partir de estimación EM. Con lo anterior, una estrategia para modelar diferentes medias dentro del enfoque bayesiano es la inclusión de efectos aleatorios, teniendo en cuenta que desde la distribución beta biparamétrica, no se incorporan para estimar varianzas como en el caso tradicional, sino buscando mejorar las medidas de precisión de las áreas pequeñas.

En esta sección se consideran los siguientes modelos de efectos mixtos en la distribución beta biparamétrica, con base en la teoría presentada en la sección 2.5:

Modelo BMBDc (Bayesiano Mixto Beta con Dispersión constante)

$$\begin{aligned} logit(\mu_d) &= X_d\beta + Z_d u_d, \\ log(\phi_d) &= \gamma, \end{aligned} \tag{4.6}$$

Modelo BMBDm (Bayesiano Mixto Beta con modelamiento del parámetro Dispersión)

$$\begin{aligned} logit(\mu_d) &= X_d\beta + Z_d u_d, \\ log(\phi_d) &= W_d\gamma \end{aligned} \tag{4.7}$$

Modelo BMBDmm (Bayesiano Mixto Beta con Dispersión mixta)

$$\begin{aligned} logit(\mu_d) &= X_d\beta + Z_d u_d, \\ log(\phi_d) &= W_d\gamma + B_d\delta_d, \end{aligned} \tag{4.8}$$

El modelo de la ecuación (4.6) tiene la misma forma funcional que los desarrollados por Figueroa–Zuñiga *et al.* (2013), presentados en la sección 2.5 (Ver ecuaciones (2.14) y (2.16)).

Por su parte, los modelos de las ecuaciones (4.7) y (4.8) corresponden a un caso particular y a una extensión del modelo para la dispersión propuesto por los autores en la ecuación (2.15), respectivamente. En cuanto a la especificación de las distribuciones a priori se tiene en cada caso:

i. $u_d|\mu_u, \Sigma_u \sim N(\mu_u, \Sigma_u)$,

ii. $\Sigma_u|\psi \sim G(\psi)$,

iii. $\beta|\mu_\beta, \Sigma_\beta \sim N(\mu_\beta, \Sigma_\beta)$,

iv. $\phi \sim GI(t_1, t_2)$, No informativa.

Para el modelo de dispersión constante en la ecuación (4.6) y, para las dispersiones modeladas de acuerdo con la ecuación (4.7) se intercambia el literal *iv.* por:

iv. $\gamma|\mu_\gamma, \Sigma_\gamma \sim N(\mu_\gamma, \Sigma_\gamma)$,

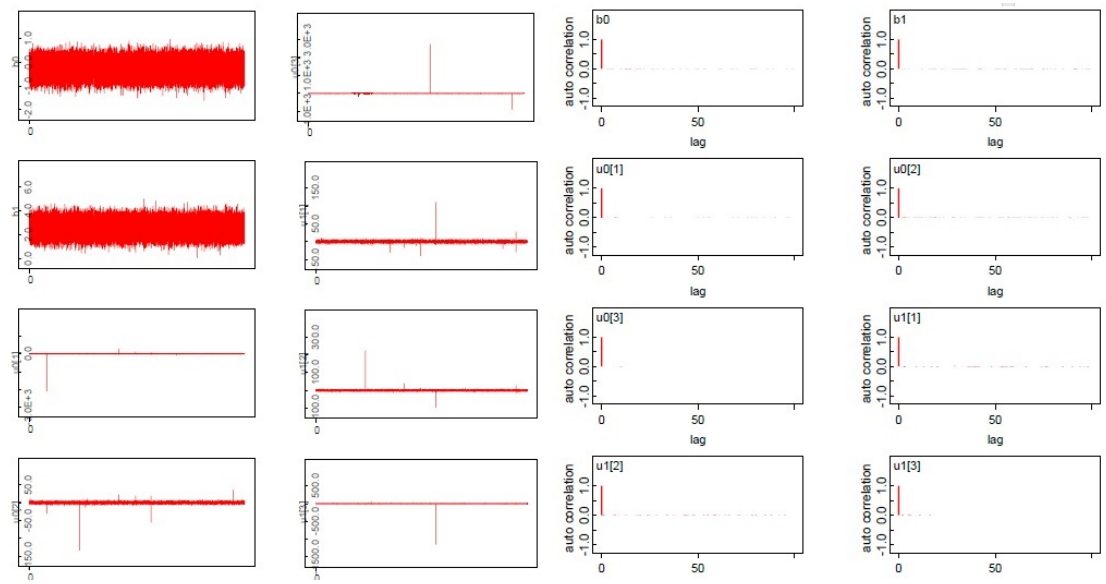
Para concluir la etapa de especificación de las a priori, en el caso de las dispersiones que siguen el modelo mixto de la ecuación (4.8) se intercambia *iv.* y se adicionan *v.* y *vi.*

iv. $b_d|\mu_b, \Sigma_b \sim N(\mu_b, \Sigma_b)$,

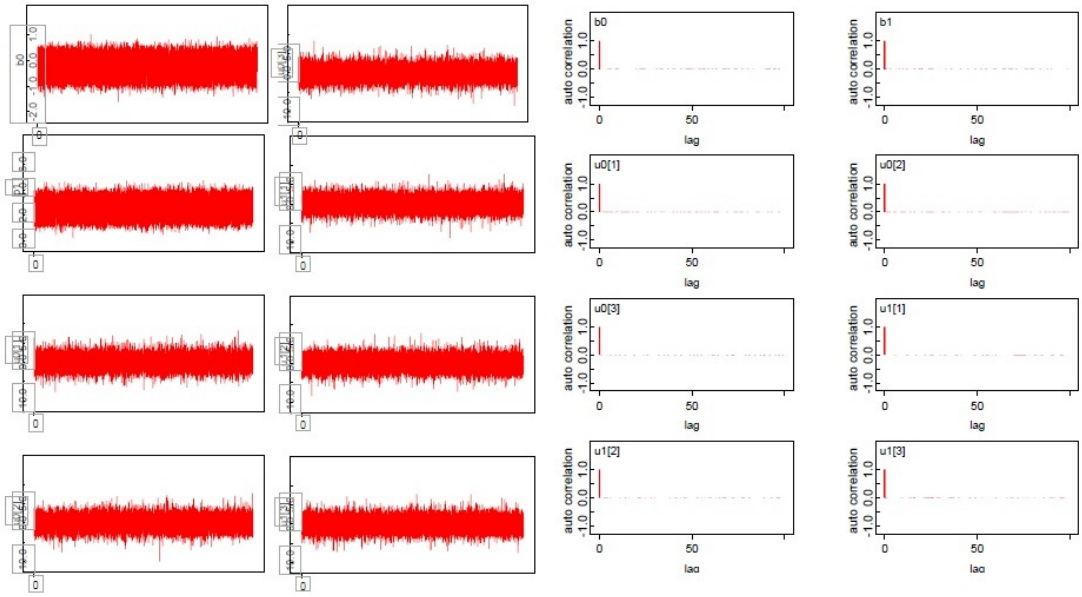
v. $\Sigma_b|\psi* \sim G(\psi*)$,

vi. $\gamma|\mu_\gamma, \Sigma_\gamma \sim N(\mu_\gamma, \Sigma_\gamma)$,

Tal y como se han especificado las apriori, las distribuciones propuestas difieren de las usadas por Figueroa–Zuñiga *et. al* (2013), ya que durante el proceso de ajuste de los modelos, se observó que la inclusión de mayor número de parámetros con muestras pequeñas, generaban cambios importantes y distorsiones ante leves modificaciones en los grados de libertad. Ejemplo de ello se presenta en la figura 4.11, que ilustran las trayectorias de las cadenas y las de autocorrelación para un ensayo sobre los datos de la implementación para el cual se consideró la agrupación dada por las componentes de mixturas finitas. Razón por la cual se opta por realizar la aplicación sobre las a priori ya mencionadas, basadas en distribuciones normales para coeficientes de las componentes sistemáticas de medias y dispersiones.



(A) A priori t–Student



(B) A prioris Normales

FIGURA 4.11. Trayectorias y autocorrelación modificando las distribuciones a priori

Finalmente, los modelos son implementados en **JAGS** por sus siglas en inglés (**Just Another Gibbs Sampler**) que, como su nombre lo indica se basa en el algoritmo muestreo de Gibbs, ver apéndice A.2.2 para la simulación de MCMC en modelos Bayesianos Jerárquicos. Aunque el lenguaje de programación no difiere sustancialmente del software WinBUGS, entre otras cosas, la operabilidad es más intuitiva.

4.4.1. Modelo de efecto aleatorio natural

Una vez especificadas las formas funcionales de los modelos de efectos mixtos que serán considerados, así como las distribuciones a priori, falta la definición del efecto aleatorio

y las variables a incluir. Para este caso, se ha denominado efecto aleatorio "natural" porque corresponde a las áreas mayores que contienen a las áreas pequeñas; es decir, la primera propuesta es considerar que el efecto aleatorio es la localidad con d niveles $d = 1, \ldots, 18$. Por otro lado, se considera únicamente como variable independiente a las condiciones de educación del hogar ya que resultó la variable que mejor modela la media y dispersión de las condiciones de la niñez y la juventud, bajo las diversas metodologías implementadas, recapitulando: de la sección 4.1 el modelo de mixtura finita beta (**MFB**); de la sección 4.2 el modelo máxima verosimilitud beta (**MVB**) y de la sección 4.3 el modelo completamente bayesiano beta alterno (**CBBa**).

Bajo el modelo **BMBDc** de la ecuación (4.6) los resultados de la estimación de cada coeficiente, se presenta en la tabla B.8 en donde, se observa que la variable Educación no resulta significativa en la mayoría de las localidades, caso contrario sucede con los interceptos, en su mayoría significativos, pero se presentan problemas cuando el tamaño de muestra es escaso, situación en la cual ningún coeficiente es significativamente diferente de cero. Desde la figura B.4 a la B.13, se muestran las trayectorias de las cadenas para los interceptos y las pendientes, así como las gráficas de autocorrelación. En el último caso se evidencia además, que la localidad 15 (Antonio Nariño), no tiene una trayetoria regular y existe correlación serial para diferentes rezagos. Los estadísticos de *Geweke* para intercepto y pendiente en este grupo son (2.65,26.593) indicando que no alcanza la convergencia y aún cuando pasan las pruebas de estacionariedad de Heidelberger y Welch (p–valor:0.3219, 0.3221), no así las de las longitudes medias de intervalos de credibilidad (Halfwidth: 5.482,21.36). Los problemas de autocorrelación podrían solucionarse mediante la generación de cadenas más largas y periodos de burn in mayores, sin embargo, la ausencia de significancia impide generar ecuaciones diferenciadas por localidad, situación que no se mejora con la especificación de modelos para el parámetro de dispersión.

Como conclusión de considerar los efectos aleatorios naturales, es posible que la combinación: alto número de efectos aleatorios con presencia de tamaños mínimos de muestra y desbalanceados, tal como sucede con los datos de ésta aplicación, impidan la convergencia de las cadenas a distribuciones estacionarias ya que en los casos de tamaños mínimos pueden presentarse inconvenientes para lograr cambios de estado en cada iteración, tal y como ocurre con la localidad de Antonio Nariño. En todo caso, validar esta hipótesis mediante un análisis de factibilidad en el que se consideren diversos universos y tamaños de muestra, se encuentra fuera del alcance de la aplicación, quedando como opción de trabajos futuros.

Con lo anterior, se propone el diseño de clusters para analizar estructuras de grupos latentes, tal como se aborda desde la metodología de mixturas finitas, solo que, para el caso, dichos grupos no aplican dado que su construcción obedece a supuestos distribucionales y metodos de estimación que difieren del enfoque bayesiano, lo que posiblemente tendría incidencia en los resultados.

4.4.2. Análisis de clasificación para el diseño de clases latentes

En consideración al desempeño del modelo de mixturas finitas, así como el análisis de las áreas mayores (localidades) descrito en la sección 3.3.1, un planteamiento alternativo se basa en la diseño de grupos que, por un lado permita el modelamiento de diversas

ecuaciones para la media y la dispersión en aras de reducir los residuales cuando las condiciones del efecto aleatorio "natural" no permiten la convergencia; por otro lado, es posible que el incremento del tamaño de muestra facilite los cambios de estado en la cadena hasta alcanzar estacionariedad. La estrategia es además soportada, a través del trabajo de Torkashvand *et al.* (2017) quienes en el marco de SAE basada en modelos lineales a nivel de área, construyen grupos con base en distancias euclidianas con el propósito de reducir del error cuadrático medio del EBLUP.

Mediante el análisis de componentes principales, previo a la construción de clusters, se incluyen todas las variables definidas dentro de la propuesta del IPM y que han sido tratadas a lo largo del documento. Dado lo anterior, se encuentra que el primer plano factorial concentra el 80 % de la variabilidad total de la información, principalmente explicada por el primer eje que concentra 68.3 %, resultado congruente con el análisis descriptivo que indicaba altas correlaciones entre las variables incluidas en el modelo. Por su parte, la variable que más contribuye en la conformación del primer factor es, condiciones educativas del hogar, de hecho en la figura B.14 se observa que ésta prácticamente define el eje, situación análoga a los resultados de modelos implementados. El segundo eje, que concentra el 12.2 % de la variabilidad de los datos, está principalmente descrito por las condiciones de la niñez y la juventud así como por las condiciones de trabajo del hogar.

En relación a la representación de los individuos en el plano factorial, de acuerdo con la figura B.15 se observa que en el cuadrante 1, se encuentran las UPZ's con población que puede denominarse *vulnerable* pues está caracterizada por presentar los índices más altos de carencias en condiciones educativas del hogar y en condiciones de la niñez y juventud; pero además, se encuentran tipificadas por presentar condiciones de vivienda con altos porcentajes de carencia, comportamiento diferencial del resto de cuadrantes del plano, máxime cuando desde el contexto de ciudad, las coberturas de servicios públicos son un tema que se considera "superado", por lo menos en las áreas urbanas. Este hecho es el resultado de un dato extremo dentro de la localidad de Usme, correspondiente a la UPZ Ciudad Usme, cuyo comportamiento se explica por tener estructuras similares a zonas rurales si se analiza a partir de su densidad poblacional, ya que dicha UPZ concentra el 30 % de la extensión hectaria de la localidad, según cifras de la Secretaría Distrital de Planeación para 2011 y tamaño poblacional de 14.852 habitantes, de acuerdo con la misma fuente.

El cuarto cuadrante, se asocia a la población de hogares con problemas educativos y que tienen gran afectación en sus condiciones de trabajo, ya que presentan mayores tasas de desempleo y desempleo de larga duración, es decir presentan una *pobreza estructural*. Las UPZ's del tercer cuadrante pueden denominarse *en pobreza coyuntural* puesto que sus mayores carencias la presentan en condiciones de trabajo pero con bajos porcentajes en las condiciones educativas, de vivienda y de condiciones de la niñez y la juventud. Finalmente, en el segundo cuadrante, se encuentran las UPZ's que teniendo bajas carencias educativas, presentan simultáneamente altas carencias en condiciones de la niñez y la juventud, así como en las condiciones de trabajo.

Para terminar, en la figura B.16 se observan los clústers de UPZ's con tamaños de 4 a 6 grupos, no obstante, en todos los escenarios Ciudad Usme por sí misma corresponde a un cluster que, como se ha visto, tiene implicación en la convergencia de la cadena. Por

ejemplo, para la información clasificada en 6 grupos, bajo el modelo de la ecuación (4.6) se observa en la figura B.18 que las cadenas asociadas a los coeficientes del sexto grupo (Ciudad Usme), no son estacionarias.

Finalmente, debido al efecto que tiene la variable condiciones de la vivienda sobre la proyección de los individuos en el plano factorial, su significancia en los modelos implementados, así como su impacto dentro del contexto temático de zonas urbanas, se realiza un nuevo análisis factorial considerando únicamente las variables asociadas a educación y trabajo, con el sustento de ser las dos variables que más contribuyen en la construcción de los ejes del ACP anterior. Sin entrar en análisis detallado de los resultados en la figura B.17 se observa que el porcentaje de varianza recogida por los ejes del plano factorial alcanza el 100 % básicamente debido a la variable educación (88.8 %).

4.4.3. Modelo de efecto aleatorio por clases latentes

En el *primer* escenario implementado se considera como efecto aleatorio, el dado por las clases latentes del ACP y se modelan los parámetros de interés en función de las carencias en condiciones educativas del hogar, sin embargo, los resultados obtenidos indican que para el *primer cluster* en ningún modelo (dispersión constante, modelada o mixta) tiene intercepto o pendiente significativa. De esta manera, se desarrolla un *segundo* escenario de modelamiento que consiste en modificar la variable independiente que explica el comportamiento de la niñez y la juventud en el cluster, por las condiciones de trabajo del hogar encontrando los mismos resultados; situación que limita el ajuste de un modelo para el grupo mencionado, por lo cual, al analizar la relación de las variables del ACP con la dependiente, se encuentra que el comportamiento de las condiciones de la niñez y la juventud en el cluster 1, se relaciona con el cociente entre las condiciones educativas del hogar y las condiciones de trabajo, como se muestra en la figura 4.12, es decir, se observa que las carencias de la niñez crecen o decrecen conforme la relación Educación/Trabajo lo hace.

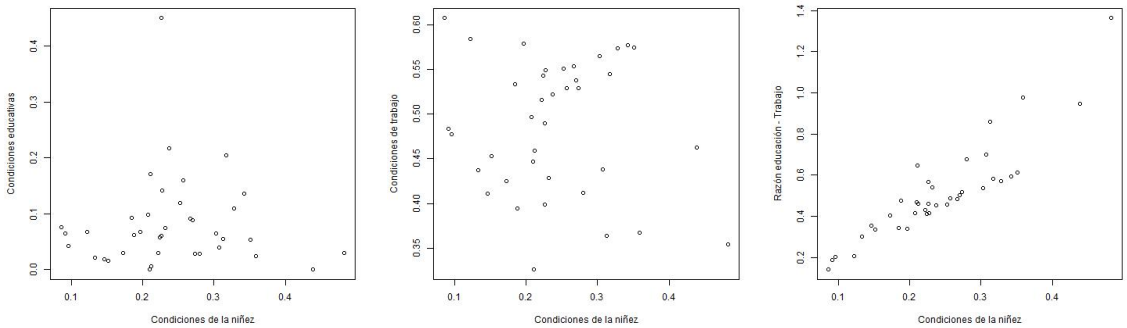


FIGURA 4.12. Gráficas de dispesión variable dependiente vs independientes. Cluster 1

Bajo este contexto, se realiza el modelamiento incorporando este cociente como variable explicativa para el cluster mencionado. Adicionalmente, en relación a las distribuciones a priori, se evidenció que la inclusión de varianzas estocásticas definidas en los literales (*ii*) para los efectos aleatorios de los tres modelos puestos a prueba, afectan

las autcorrelaciones de las cadenas como lo ilustra por ejemplo, la figura B.19 para el modelo **BMBDc** en donde además, el criterio de selección DIC es (-246.4), mayor que el obtenido para el mismo modelo asumiendo varianzas no estocásticas. Finalmente, las cadenas requirieron de un alto número de iteraciones, en comparación con los modelos implementados en otras metodologías pasando de 100.000 a 1.000.000, principalmente sobre leves rezagos de orden 1 existentes en las cadenas asociadas al tercer grupo.

Los resultados de los modelos basados en los tres escenarios planteados en las ecuaciones (4.6), (4.7) y (4.8) se presentan a continuación, para que, una vez descritos los resultados se defina el mejor modelo mediante los criterios de selección, sobre el cual se verifican los supuestos de los residuales.

4.4.3.1. Modelo Bayesiano Mixto Beta con Dispersión constante

Para el modelo de dispersión ϕ constante **BMBDc**, se tiene $\widehat{\phi} = 41.777$ y la estimación de los coeficientes del modelo para las medias son:

	Cluster 1		Cluster 2		Cluster3	
μ_d	$\widehat{\beta}_{i1}$	Error Est.	$\widehat{\beta}_{i2}$	Error Est.	$\widehat{\beta}_{i3}$	Error Est.
Intercepto	-1.269	0.062	-1.469	0.211	-1.742	0.434
Educacion			2.126	0.765	2.591	0.832
Cociente (E/T)	0.115	0.028				

Nota: Todos los coeficientes son significativos a cualquier nivel.

Tamaños de los clusters: 38/40/25 – Criterio DIC: -252.6

TABLA 4.10. Resumen modelo Bayesiano Mixto Beta con dispersión constante (BMBDc).

4.4.3.2. Modelo Bayesiano Mixto Beta con modelamiento del parámetro Dispersión

Para el modelo de dispersión ϕ_d modelada **BMBDm**, se asume que la relación entre el logaritmo del parámetro sigue una relación lineal simple con relación a las condiciones educativas del hogar, es decir:

$$log(\phi_d) = \gamma_0 + Educacion\gamma_1$$

En donde,

ϕ_d	$\widehat{\gamma}_i$	Error Est.
Intercepto	3.001	0.227
Educacion	3.468	0.766

Por su parte, la estimación de los parámetros asociados a las medias se presentan en la tabla 4.11:

	Cluster 1		Cluster 2		Cluster3	
μ_d	$\widehat{\beta}_{i1}$	Error Est.	$\widehat{\beta}_{i2}$	Error Est.	$\widehat{\beta}_{i3}$	Error Est.
Intercepto	-1.240	0.078	-1.479	0.213	-1.754	0.272
Educacion			2.161	0.730	2.602	0.480
Cociente (E/T)	0.115	0.040				

Nota: Todos los coeficientes son significativos a cualquier nivel.

Tamaños de los clusters: 38/40/25 – Criterio DIC: -270.9

TABLA 4.11. Resumen Modelo Bayesiano Mixto Beta con dispersión modelada (BMBDm).

4.4.3.3. Modelo Bayesiano Mixto Beta con Dispersión mixta

En cuanto a la inclusión de efectos aleatorios, sobre el modelo de dispersión ϕ_d mixta **BMBDmm**, se realizan tres modelos, uno que incorpora los mismos clusters y variables tanto para media como para dispersión **BMBDmm1**, el segundo surge a partir del hecho de que, como se observa en la tabla 4.12, la mayoría de los coeficientes que modelan la dispersión resultan no significativos, por lo cual, para el escenario 2, se asume que la estructura de los clusters se conserva, pero se modela la dispersión sólo en función de la variable educación como se observa en la tabla 4.13 y el modelo será notado **BMBDmm2**. El tercer modelo (4.14) modifica también los grupos en razón a que el modelo **BMBDmm2** descrito anteriormente, muestra algunos coeficientes significativos, por lo tanto se unen los no significativos en un solo grupo, bajo el modelo notado **BMBDmm3**.

Modelo Bayesiano Mixto Beta, Dispersión Mixta 1: (BMBDmm1)

	Cluster 1		Cluster 2		Cluster3	
μ_d	$\widehat{\beta}_{i1}$	Error Est.	$\widehat{\beta}_{i2}$	Error Est.	$\widehat{\beta}_{i3}$	Error Est.
Intercepto	-1.708	0.362	-1.470	0.201	-1.775	0.190
Educacion			2.138	0.751	2.640	0.355
Cociente (E/T)	1.063	0.643				
ϕ_d	$\widehat{\gamma}_{i1}$	Error Est.	$\widehat{\gamma}_{i2}$	Error Est.	$\widehat{\gamma}_{i3}$	Error Est.
Intercepto	3.993	0.875	4.176	0.819	4.218**	2.383
Educacion			-1.866**	2.945	2.433**	4.609
Cociente (E/T)	0.018**	0.829				

***: Coeficientes No significativos. Tamaños de los clusters: 38/40/25 – Criterio DIC: -261.5*

TABLA 4.12. Resumen Modelo Bayesiano Mixto Beta con dispersión mixta (BMBDmm1).

Modelo Bayesiano Mixto Beta con Dispersión mixta 2: (BMBDmm2)

	Cluster 1		Cluster 2		Cluster3	
μ_d	$\widehat{\beta}_{i1}$	Error Est.	$\widehat{\beta}_{i2}$	Error Est.	$\widehat{\beta}_{i3}$	Error Est.
Intercepto	-1.251	0.062	-1.476	0.204	-1.752	0.191
Educacion			2.165	0.748	2.601	0.353
Cociente (E/T)	0.116	0.044				
ϕ_d	$\widehat{\gamma}_{i1}$	Error Est.	$\widehat{\gamma}_{i2}$	Error Est.	$\widehat{\gamma}_{i3}$	Error Est.
Intercepto	2.735	0.341	4.150	0.842	3.605**	2.583
Educacion	8.259	3.479	-1.733**	3.015	3.611**	4.968

***: Coeficientes No significativos. Tamaños de los clusters: 38/40/25 – Criterio DIC: -277.0*

TABLA 4.13. Resumen Modelo Bayesiano Mixto Beta con dispersión mixta solo a partir de educación (BMBDmm2).

Modelo Bayesiano Mixto Beta con Dispersión mixta 3: (BMBDmm3)

ϕ_d	$\widehat{\gamma}_{i1}$	Error Est.	$\widehat{\gamma}_{i2}$	Error Est.
Intercepto	4.001	0.444	5.481	0.295
Educacion	9.500	3.520	-2.669	0.565

	Cluster 1		Cluster 2		Cluster3	
μ_d	$\widehat{\beta}_{i1}$	Error Est.	$\widehat{\beta}_{i2}$	Error Est.	$\widehat{\beta}_{i3}$	Error Est.
Intercepto	-1.261	0.054	-1.449	0.216	-1.785	0.186
Educacion			2.059	0.795	2.659	0.356
Cociente (E/T)	0.112	0.035				

Criterio DIC: -281.0

TABLA 4.14. Resumen Modelo Bayesiano Mixto Beto con dispersión mixta y grupos colapsados (BMBDmm3).

Bajo el criterio DIC, el mejor modelo dentro del grupo de modelos beta mixtos para media y dispersión es el **BMBDmm3**, con valor de -281, sin embargo, éste modelo presenta inconvenientes sobre los supuestos de los errores, toda vez que el test de *Shapiro Wilk* para normalidad presenta un p–valor de 0.0075 resultado que se confirma con la gráfica de bandas simuladas de la figura 4.13 y aunque cumple con los supuestos de independencia (p– valor de 0.11 para el test *Ljung-Box*) no puede ser considerado como un buen modelo, si además, en la gráfica de dispersión de las medias ajustadas frente a los residuales ponderados se observa una marcada forma de cono, indicando que la varianza no ha sido correctamente modelada.

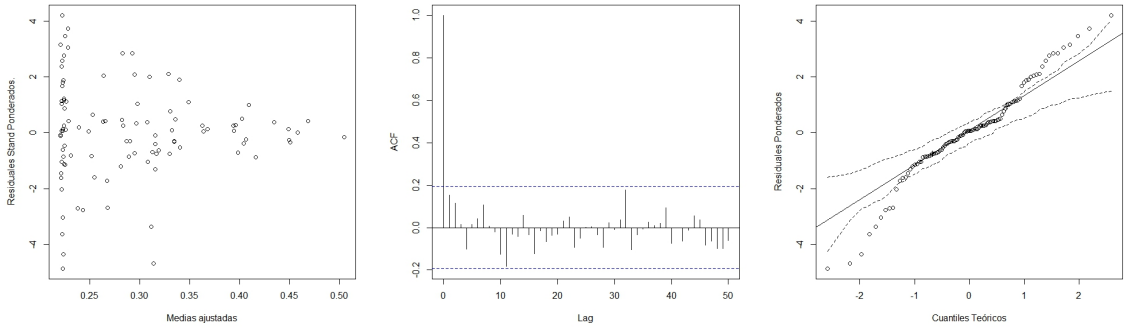


FIGURA 4.13. Gráficas de residuales ponderados sobre el modelo BMBDmm3.

Con los anteriores escenarios, se concluye que el modelo de dispersión de componente sistemática lineal simple, puede modelar adecuadamente la dispersión, en este caso, la verificación de los supuestos de los errores presentan para el test de *Ljung Box* un p–valor de 0.2658 así mismo, el gráfico de bandas simuladas presentan un comportamiento de normalidad, como se observa en la figura 4.14.

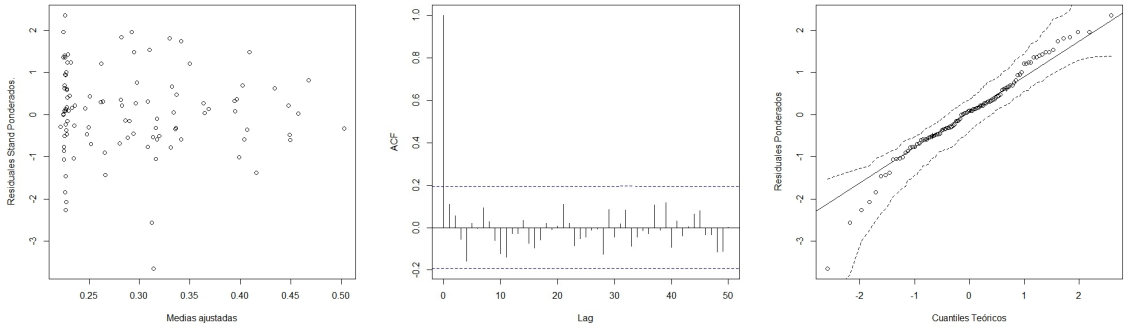


FIGURA 4.14. Gráficas de residuales ponderados sobre el modelo BMBDm.

Las gráficas de trayectorias de las cadenas, así como de las autocorrelaciones se presentan en las figuras 4.15 y 4.16, respectivamente, para la inspección visual de convergencia.

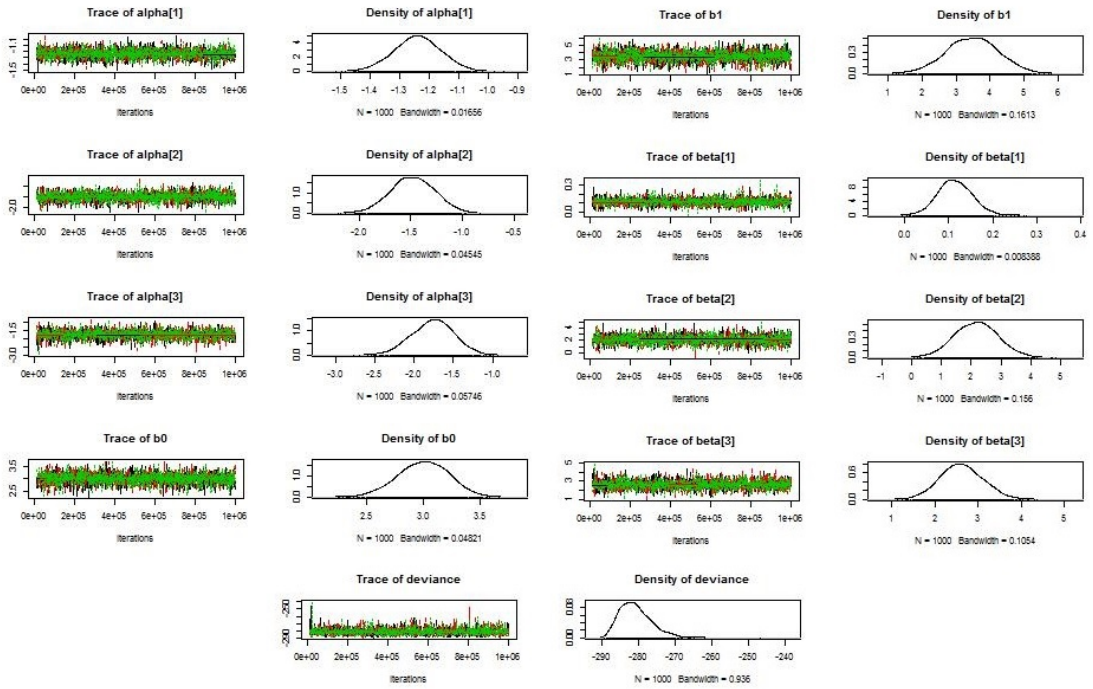


FIGURA 4.15. Trayectoria de las cadenas del modelo BMBDm.

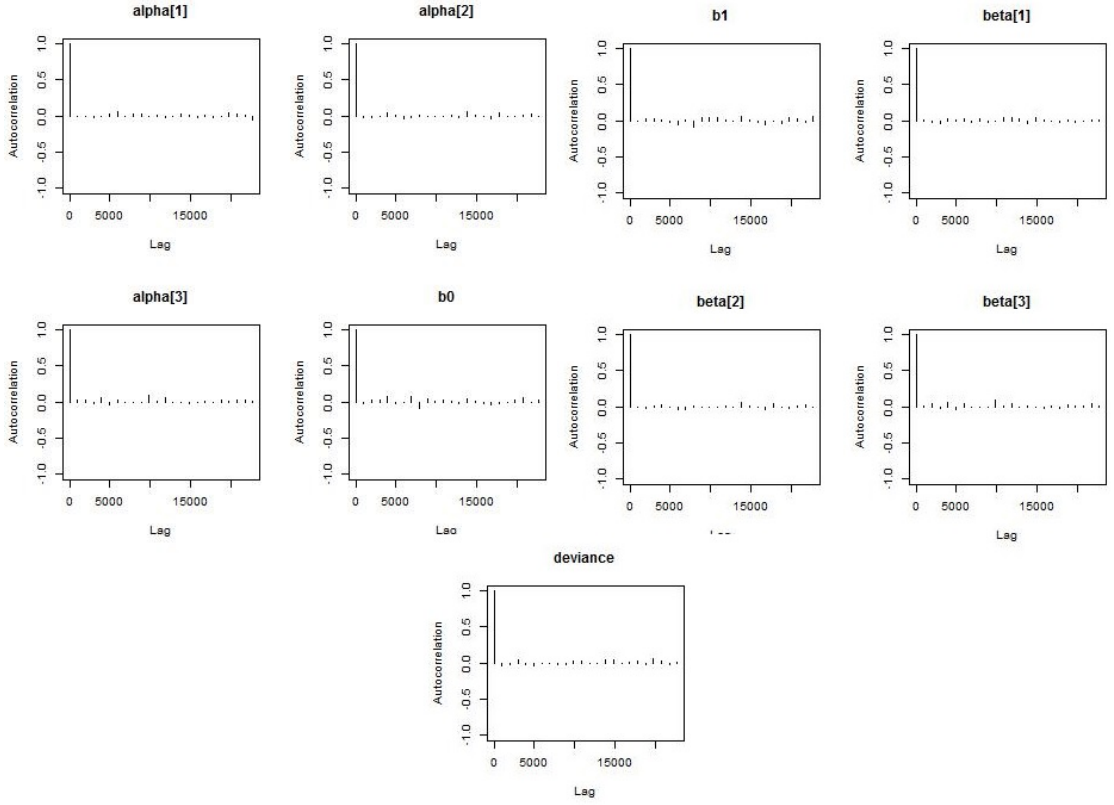


FIGURA 4.16. Autocorrelaciones de las cadenas del modelo BMBDm.

Para concluir el análisis del modelo, las pruebas analíticas confirman la convergencia para todas las cadenas.

μ_d	Cluster	Geweke	Heidelberger y Welch	Halfwidth
Intercepto	1	0.42240	0.9869	0.00377
	2	-0.45084	0.0717	0.01295
	3	0.55572	0.7567	0.02649
Pendiente	1	1.21640	0.8615	0.00172
	2	0.52303	0.0973	0.04925
	3	0.03576	0.6252	0.05176
ϕ				
Intercepto		-0.12459	0.0648	0.36690
Pendiente		-0.07431	0.1238	0.36690

TABLA 4.15. Criterios de convergencia de las cadenas del Modelo BMBDm.

4.5. Análisis comparativo de estimadores de proporciones para áreas pequeñas basados en el modelo Beta

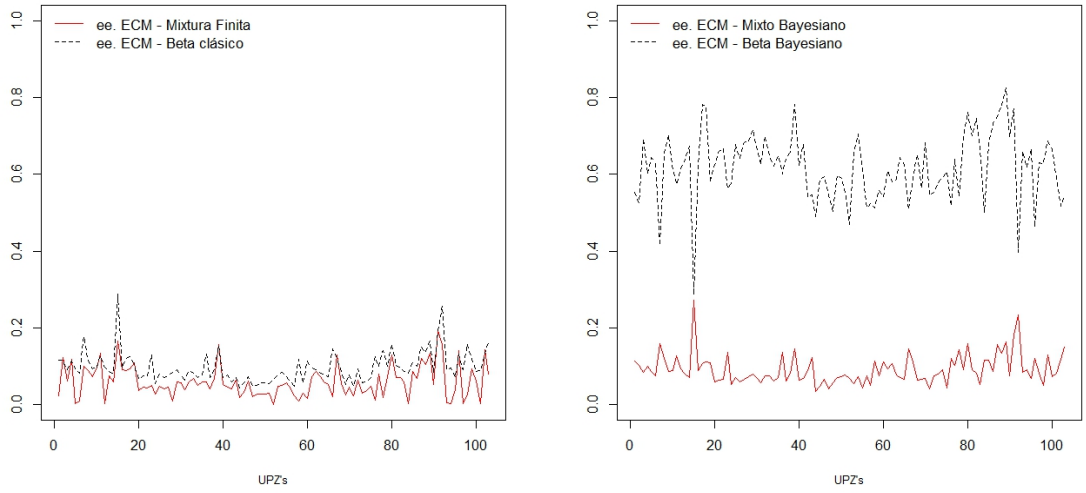
A lo largo del capítulo se abordó la estimación de variables que siguen una distribución beta biparamétrica desde el enfoque clásico, mediante el tratamiento de mixturas finitas, así como de estimación clásica beta; posteriormente se abordó la estimación bayesiana sin incluir efectos aleatorios considerando que la distribución beta permite realizar de forma directa, la estimación del parámetro de dispersión, y por lo tanto la posterior varianza del estimador; los efectos aleatorios incorporados en la parte final de la aplicación se introducen como una estrategia para mejorar el ajuste del modelo, caso para el cual, el efecto natural no resulta significativo, por lo que se consideraron grupos dados por clasificación vía ACP y se ajustan modelos bajo diferentes escenarios de inclusión de los efectos.

Bajo cada metodología, se obtuvo "el mejor" modelo ajustado, la respectiva verificación de supuestos y, para los casos de estimación bayesiana, se incluyó su análisis de convergencia. En la práctica, sólo un modelo es seleccionado con el objeto de que en términos generales, además de cumplir con las cualidades de ajuste, sea posible la publicación de cifras con la precisión adecuada para cada área pequeña. Como se presentó en la sección 1.4, la medida utilizada en SAE para tal fin, es el error cuadrático medio, sin embargo, debido, entre otras cosas, a la ausencia de expresiones analíticas para su estimación, Chatterjee, Lahiri y Li (2006) proponen el uso de técnicas de remuestreo como bootstrap para su cálculo, ver apéndice A.3. Desde el enfoque Bayesiano, la medida puede ser estimada a partir de las cadenas generadas dentro del procedimiento de estimación, sin embargo, de acuerdo con Herrador M.*et. al* (2009), los resultados de cada modelo dependen de las especificaciones y los supuestos asociados, lo que impide su comparación directa.

De lo anterior, en esta sección se presentan los resultados de estimación de las medidas precisión mediante la técnica de bootstrap, que permite generar criterios de decisión aislando el efecto de la metodología de estimación; los estimadores sujetos al análisis de precisión son: Estimador en Mixturas Finitas Beta (**EMFB**), Estimador Máxima Verosimilitud Beta (**EMVB**), Estimador Completamente Bayesiano Beta (**ECBB**) y

Estimador Bayesiano Mixto Beta (**EBMB**), obtenidos mediante el ajuste de los modelos respectivos: mixtura finita beta (**MFB**) desarrollado en la sección 4.1; modelo clásico beta (**MVB**) de la sección 4.2; modelo bayesiano beta alterno (**BBa**) de la sección 4.3 y finalmente, de la sección previa 4.4, el modelo bayesiano mixto beta con modelamiento del parámetro dispersión lineal simple (**BMBDm**). Es importante notar que, en tres de los cuatro modelos, la componente sistemática asociada al parámetro de dispersión es lineal para las condiciones de educación del hogar. Mientras que para el modelo bayesiano beta alterno (**BBa**) el mejor modelo correspondió a un ajuste de dispersión constante.

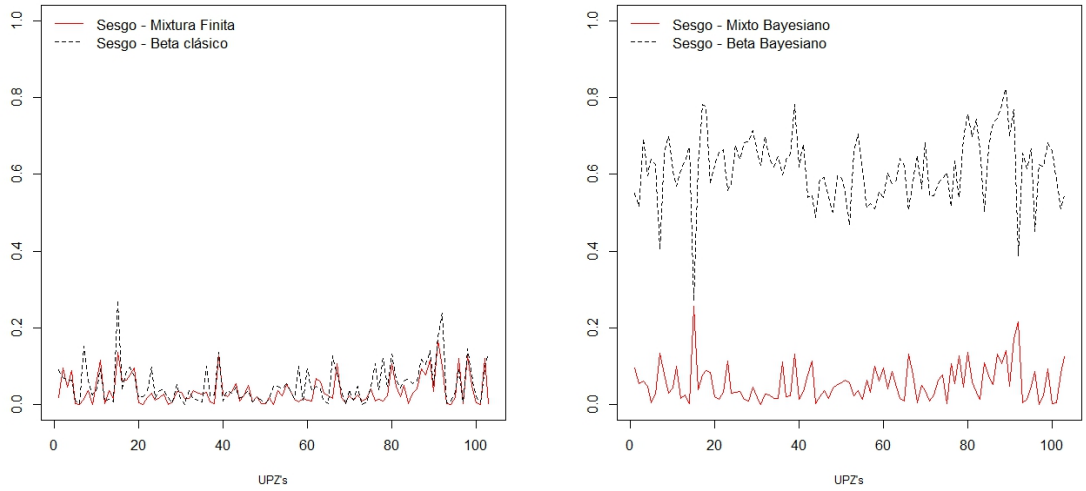
Los resultados del error estándar para $\widehat{\theta}_d$, calculado como la raíz de los errores cuadrados medios, con base en 10.000 muestras bootstrap se presentan en la figura 4.17, en donde se observa que el peor desempeño se presenta en el modelo de dispersión constante.



(A) Estimación Clásica (B) Estimación Bayesiana

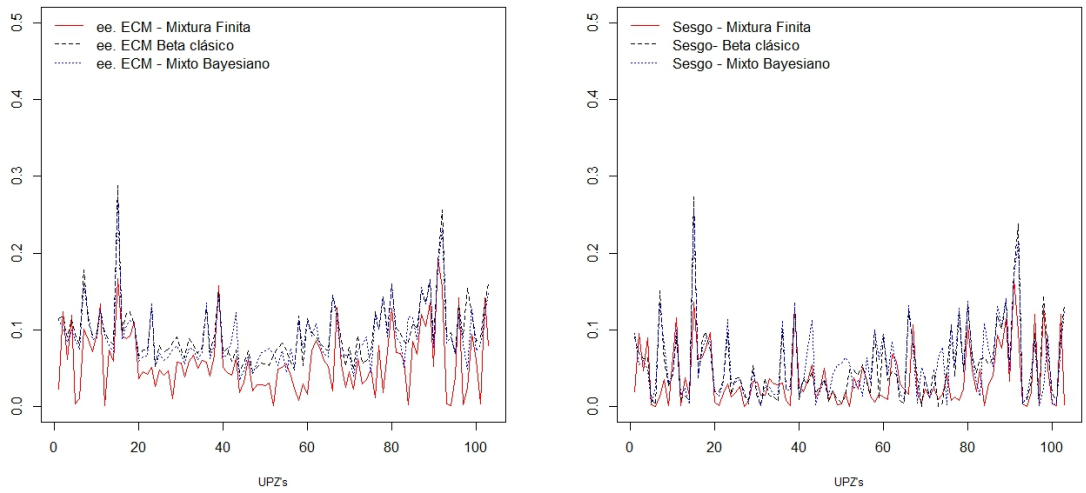
FIGURA 4.17. Error estándar para el porcentaje de hogares con carencias en condiciones de la niñez y juventud.

Adicionalmente, al analizar los sesgos, se observa en la figura 4.18, que el comportamiento del error estándar del modelo beta bayesiano, está fuertemente influenciado por la amplitud de los residuales. Por su parte, aunque la inclusión de grupos por mixturas o clusters pareciera no representar una ganancia significativa frente al modelo beta clásico, en términos del modelamiento de la media que permita producir estimaciones para áreas no muestreadas, si se observan grandes desviaciones para algunas UPZ's pero además, tiene menor desempeño en relación a los errores estándar, frente a los competidores que incorporan grupos. En todo caso, presenta un buen desempeño, lo que sugiere que, por lo menos para estos datos, el modelamiento de la dispersión logra capturar gran parte de la variabilidad contenida los mismos.



(A) Estimación Clásica (B) Estimación Bayesiana

FIGURA 4.18. Sesgos bootstrap de las predicciones de $\widehat{\mu}_d$.



(A) Errores estándar (B) Sesgo

FIGURA 4.19. Error estándar y sesgos para modelos con dispersión modelada.

Excluyendo del análisis el modelo de dispersión constante, se observa en la figura 4.19 que los modelos presentan resultados muy similares, pero que podría preferirse el modelo de mixturas finitas, ya que presenta menores errores estándar y menores diferencias entre el estimador directo y la predicción de la media.

Para concluir, una ventaja del modelamiento para un estimador directo que sigue una distribución beta, es que al pertenercer a la familia exponencial biparamétrica y utilizar modelamiento de los dos parámetros, es posible estimar el error cuadrático medio

directamente de la información modelada.

Debido a que $ECM[\widehat{\theta}_d] = V(\widehat{\theta}_d) + B^2(\widehat{\theta}_d)$, (donde B nota el sesgo del estimador) y de acuerdo con las ecuaciones (2.3), la expresión analítica del ECM para un estimador directo en la familia exponencial biparámetrica beta, está dada por:

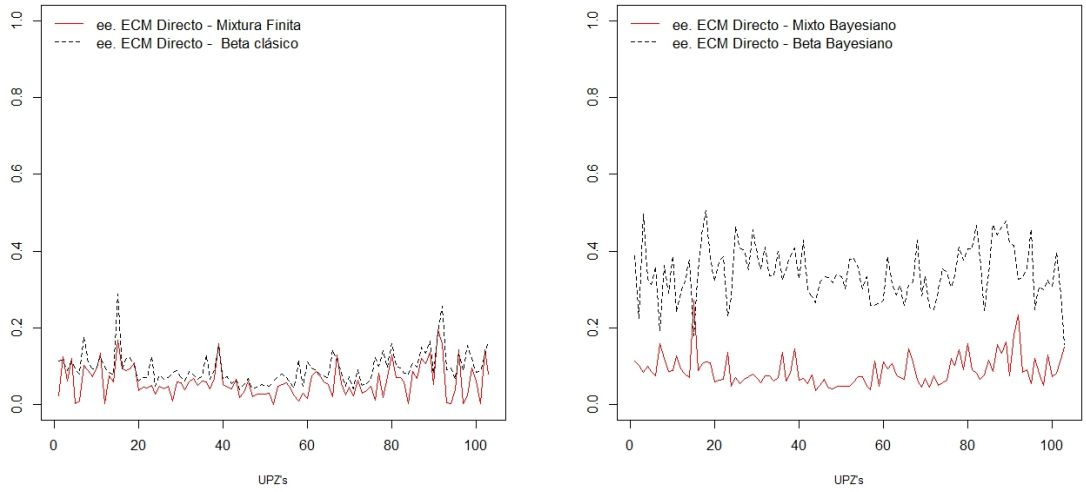
$$ECM(\widehat{\theta}_d) = \frac{\mu_d(1-\mu_d)}{(1+\phi_d)} + (\widehat{\theta}_d - \mu_d)^2.$$

Con estimador:

$$\widehat{ECM}(\widehat{\theta}_d) = \frac{\widehat{\mu}_d(1-\widehat{\mu}_d)}{(1+\widehat{\phi}_d)} + (\widehat{\theta}_d - \widehat{\mu}_d)^2. \tag{4.9}$$

De esta manera, una vez ajustado el modelo es posible reemplazar de forma directa todas las componentes para obtener el error estándar del estimador directo, haciendo $\widehat{EE_{SAE}}(\widehat{\theta}) = \sqrt{\widehat{ECM}(\widehat{\theta}_d)}$.

En la figura 4.20, se observan las estimaciones directas del error estándar para el porcentaje de hogares con carencias en condiciones de niñez y juventud.



(A) Estimación Clásica (B) Estimación Bayesiana

FIGURA 4.20. Error estándar para el porcentaje de hogares con carencias en condiciones de la niñez y jueventud.

Comparando los resultados del estimador plug–in del error cuadrático medio, con los obtenidos vía bootstrap en figura 4.17, se observa que, aunque es viable utilizar de forma directa los estimadores del error cuadrático medio, ante la falta de ajuste se tiende a subestimar los errores estándar como sucede con el modelo de dispersión constante **BBa**.

Finalmente, con los hallazgos dados por la comparación de los modelos, se opta por el modelo de mixturas finitas **MFB**, pues además de generar resultados con alta precisión; en relación a sus competidores, presenta los menores residuales, siendo esto útil para la predicción de estimadores en áreas pequeñas no muestreadas (así no sea el caso de la aplicación). Los estimadores puntuales, así como los errores estándar asociados se presentan en la tabla 4.17, encontrando que de las 103 UPZ's 46 presentan errores estándar inferiores al 5 %, 41 entre el 5 % y 10 % , mientras que 16 deben ser tratadas con precausión.

Localidad	UPZ	$\widehat{\theta}\,\%$	$\widehat{EE_{SAE}(\widehat{\theta})}\,\%$
Antonio Nariño	Ciudad Jardín	27,29	4,63
Antonio Nariño	Restrepo	29,97	5,69
Barrios Unidos	Los Andes	25,25	7,47
Barrios Unidos	Doce de Octubre	24,92	5,99
Barrios Unidos	Los Alcázares	20,00	6,87
Barrios Unidos	Parque Salitre	14,26	19,42
Bosa	Apogeo	27,98	4,46
Bosa	Bosa Occidental	37,53	2,95
Bosa	Bosa Central	35,94	3,52
Bosa	El Porvenir	35,37	4,82
Bosa	Tintal Sur	43,67	1,08
Chapinero	El Refugio	17,21	8,04
Chapinero	San Isidro Patios	40,95	1,77
Chapinero	Pardo Rubio	18,37	7,25
Chapinero	Chicó Lago	15,10	8,69
Chapinero	Chapinero	9,57	12,10
Ciudad Bólivar	Arborizadora	25,26	5,10
Ciudad Bólivar	San Francisco	35,85	5,59
Ciudad Bólivar	Lucero	45,93	4,02
Ciudad Bólivar	El Tesoro	45,79	2,32
Ciudad Bólivar	Ismael Perdomo	44,00	0,79
Ciudad Bólivar	Jerusalén	40,75	2,92
Engativá	Las Ferias	17,68	9,25
Engativá	Minuto de Dios	27,61	4,54
Engativá	Boyacá Real	28,87	4,23
Engativá	Santa Cecilia	34,26	5,07
Engativá	Bolivia	18,72	7,42
Engativá	Garcés Navas	31,70	8,64
Engativá	Engativá	34,33	7,33
Engativá	Álamos	35,08	7,98
Fontibón	Fontibón	28,01	5,99
Fontibón	Fontibón San Pablo	29,19	5,23
Fontibón	Zona Franca	42,38	2,09
Fontibón	Ciudad Salitre Occidental	22,62	0,26
Fontibón	Granjas de Techo	13,21	9,50
Fontibón	Modelia	23,18	0,29
Fontibón	Capellanía	30,74	14,26

TABLA 4.16. Porcentaje de hogares con carencias en condiciones de la niñez y la juventud y errores estándar (MFB)- Parte 1.

Localidad	UPZ	$\widehat{\theta}\,\%$	$\widehat{EE_{SAE}(\widehat{\theta})}\,\%$
Kennedy	Américas	33,62	5,84
Kennedy	Carvajal	29,52	4,00
Kennedy	Castilla	20,70	6,80
Kennedy	Kennedy Central	17,97	15,85
Kennedy	Timiza	30,96	5,13
Kennedy	Tintal Norte	30,31	12,96
Kennedy	Calandaima	23,68	5,44
Kennedy	Corabastos	39,73	2,51
Kennedy	Gran Britalia	28,13	4,51
Kennedy	Patio Bonito	42,57	2,29
Kennedy	Las Margaritas	21,03	6,28
Kennedy	Bavaria	22,75	6,17
Los Mártires	Santa Isabel	23,39	5,82
Los Mártires	La Sabana	26,19	5,26
Puente Aranda	Ciudad Montes	22,23	5,88
Puente Aranda	Muzú	27,50	6,74
Puente Aranda	San Rafael	27,67	6,00
Puente Aranda	Zona Industrial	30,93	3,89
Puente Aranda	Puente Aranda	22,57	2,36
Rafael Uribe Uribe	San José	24,39	0,99
Rafael Uribe Uribe	Quiroga	33,47	3,87
Rafael Uribe Uribe	Marco Fidel Suárez	37,85	3,20
Rafael Uribe Uribe	Marruecos	36,78	6,14
Rafael Uribe Uribe	Diana Turbay	42,96	2,08
San Cristóbal	San Blas	39,01	2,62
San Cristóbal	Sosiego	26,25	4,80
San Cristóbal	20 de Julio	31,63	4,13
San Cristóbal	La Gloria	43,22	4,12
San Cristóbal	Los Libertadores	40,95	6,24
Santa Fe	Sagrado Corazón	9,06	12,88
Santa Fe	La Macarena	20,58	7,02
Santa Fe	Las Nieves	19,07	6,99
Santa Fe	Las Cruces	29,03	5,85
Santa Fe	Lourdes	47,26	0,22
Suba	San José de Bavaria	19,67	7,29
Suba	Britalia	26,98	9,51
Suba	El Prado	32,83	13,37
Suba	La Alhambra	20,96	0,15
Suba	Casa Blanca Suba	48,27	16,74
Suba	Niza	26,71	9,23
Suba	La Floresta	15,90	8,92
Suba	Suba	34,58	10,98
Suba	El Rincón	31,66	3,59
Suba	Tibabuyes	40,53	1,54

TABLA 4.17. Porcentaje de hogares con carencias en condiciones de la niñez y la juventud y errores estándar (MFB)- Parte 2.

Localidad	UPZ	$\widehat{\theta}\,\%$	$\widehat{EE_{SAE}(\widehat{\theta})}\,\%$
Teusaquillo	Galerías	12,11	10,55
Teusaquillo	Teusaquillo	8,54	13,53
Teusaquillo	Parque Simón Bolívar - CAN	43,88	15,85
Teusaquillo	La Esmeralda	22,33	0,38
Teusaquillo	Quinta Paredes	21,21	0,11
Teusaquillo	Ciudad Salitre Oriental	31,31	14,18
Tunjuelito	Venecia	32,35	5,03
Tunjuelito	Tunjuelito	31,07	4,83
Usaquén	Verbenal	38,99	2,26
Usaquén	La Uribe	27,98	12,42
Usaquén	San Cristóbal Norte	25,64	6,07
Usaquén	Toberín	27,32	11,95
Usaquén	Los Cedros	22,15	0,30
Usaquén	Usaquén	25,66	0,97
Usaquén	Country Club	35,91	10,29
Usaquén	Santa Bárbara	14,55	8,88
Usme	La Flora	49,22	1,83
Usme	Danubio	45,63	2,79
Usme	Gran Yomasa	36,52	2,80
Usme	Comuneros	37,36	2,76
Usme	Alfonso López	41,16	3,04
Usme	Ciudad Usme	49,79	0,07

TABLA 4.18. Porcentaje de hogares con carencias en condiciones de la niñez y la juventud y errores estándar (MFB). Parte Final

CAPÍTULO 5

Conclusiones y Trabajos Futuros

5.1. Conclusiones

Las conclusiones se exponen en forma de afirmación con base en los resultados para el caso específico de la información disponible, las generalizaciones requieren de procedimientos sobre poblaciones controladas, vía simulación, que permitan responder a cada uno de los items que se sugieren como hipótesis.

1. Se implementó el modelo beta para la estimación de parámetros con dominio en el intervalo (0,1) dentro del contexto de estimación de áreas pequeñas basada en modelos. La implementación es novedosa, puesto que dentro de la revisión bibliográfica realizada, no se encontraron desarrollos sobre este tema particular.

2. Una ventaja de la distribución beta en la familia exponencial biparamétrica, es que no requiere la inclusión de efectos aleatorios para estimar la varianza del estimador directo, debido a que el modelamiento conjunto de los parámetros de la distribución, provee toda la información requerida para la predicción en áreas no muestreadas y estimación de errores estándar basados en el error cuadrático medio.

3. Se implementó la metodología de mixturas finitas para SAE, como una estrategia alterna que difiere del modelamiento de efectos mixtos, sobre los que tradicionalmente se basa la técnica de estimación en áreas pequeñas; éste tema no había sido abordo hasta el momento de la implementación, en la literatura de SAE de acuerdo con la revisión bibliográfica.

4. El paquete de R–Project, betareg y la función betamix para la implementación de mixturas finitas, no disponía al momento de la aplicación, de procedimientos para validar los supuestos del modelo, por lo anterior, se implementó en este trabajo el cálculo de residuales Pearson y residuales ponderados por las componentes de la matriz Hat, para los respectivos fines del documento.

5. Se identifican inconvenientes sobre el cálculo de los errores ponderados por la matriz Hat en el paquete Bayesianbetareg, de R–Project, por lo tanto, para la implementación se calcularon de forma externa al paquete.

6. La inclusión de efectos aleatorios en el contexto de SAE para estimadores directos que siguen una distribución beta biparamétrica, así como la generación de grupos por mixturas finitas, se utilizan con el propósito de mejorar la estimación, presentando su mayor utilidad, en los casos de áreas pequeñas no muestreadas.

7. El modelamiento del parámetro de dispersión en la distribución beta, permitió obtener errores estándar para cada área pequeña con alta precisión. En contraste, el desempeño del modelo bayesiano con dispersión constante, se vió considerablemente afectado en términos de errores estándar, limitando su uso en el contexto de áreas pequeñas para esta aplicación.

8. A partir del plantemiento de la distribución beta biparamétrica, es posible obtener estimadores del error estándar SAE sin recurrir a procedimientos de remuestreo, no obstante, malos ajustes pueden sobreestimar o subestimar los errores estándar de las áreas pequeñas. En este sentido, se recomienda la técnica de remuestreo, con fines de supervisión.

9. El mejor modelo para el caso aplicado, resultó ser el modelo de mixtura finita beta **MFB**, pues estima con menor sesgo, los errores estándar son los más pequeños entre todas las metodologías implementadas y, bajo el procedimiento de bootstrap se supervisa la consistencia de sus resultados.

10. Acorde con el análisis descriptivo, las localidades no resultan ser un factor diferenciador entre las UPZ's, es decir, mediante mixturas finitas las componentes agruparon UPZ's de diversos territorios. Así mismo, la inclusión de la localidades como efecto aleatorio en el enfoque bayesiano, tampoco generó significacia estadística, lo anterior, limita la incorporación de estructuras de autocorrelación espacial, para la implementación de modelos más sofisticados en áras de encontrar mejores ajustes.

11. En todos los modelos, la variable condiciones educativas del hogar, se presenta como un factor determinante para el comportamiento de las condiciones de la niñez y la juventud. Esto tiene dos implicaciones, desde el punto de vista técnico, la estimación basada en modelos permite producir resultados apoyados en información auxiliar que facilite la producción de estimadores con mayor precisión o incluso, en función de estimar indicadores para áreas pequeñas no seleccionadas. Desde el punto de vista práctico, la identificación de factores de riesgo asociados a una problemática permitirían tomar decisiones más focalizadas. En el caso de esta aplicación, podría generarse movilidad social si se focalizan los esfuerzos en mitigar de las carencias en las condiciones de educación de la población.

12. La metodología desarrollada para este caso aplicado puede extenderse a indicadores de otras fuentes y objetivos, aplicados a bioestadística, por ejemplo, para temas de salud pública, epidemiología, nutrición, salud ambiental, poblaciones genéticas, etc.

5.2. Trabajos futuros

1. En la teoría de muestreo, el sesgo relativo de un estimador de razón es despreciable conforme el tamaño de muestra crece, situación que no necesariamente podría aplicar a la teoría de áreas pequeñas, por lo que en futuros trabajos se podría analizar el efecto del sesgo de diseño, para la estimación de proporciones en áreas pequeñas basada en modelos.

2. El modelamiento para mixturas finitas beta implementado con el software disponible en R–Project, se basa en componentes comunes para los parámetros de media y dispersión, un trabajo futuro, podría considerar componentes diferentes para cada parámetro o, componentes exclusivas para la media y modelamiento de la dispersión agregada o, modelamiento de media sin componentes y modelo de dispersión con componentes.

3. El modelamiento desarrollado en este trabajo para mixturas finitas, no hace uso de variables concomitantes, por lo que sería interesante implementar desarrollos con base en probabilidades de pertenencia a cada componente, dadas por un modelo logístico multinomial.

4. La metodología de mixturas finitas beta implementada, se basa en el algoritmo Esperanza Maximización, que se desarrolla mediante teoría clásica de estimación, un tema de investigación es la implementación de inferencia bayesiana en mixturas beta.

5. Considerar los efectos aleatorios naturales (localidades) dentro del modelo, arrojó resultados estadísticamente no significantes. Diversos escenarios exploratorios fueron realizados para esta implementación, a saber, se consideraron sólo localidades con mayor número de UPZ's, se consideraron sólo localidades con menor número de UPZ's, se ampliaron los tamaños de muestra en cada localidad a partir del método de remuestreo bootstrap, en todos los casos, buscando identificar los motivos de no convegencia o significancia; sin embargo, los resultados no son del alcance de este estudio de aplicación, pero sugiere la necesidad de realizar estudios de factibilidad, para examinar bajo qué condiciones puede llegarse a incluirse los efectos aleatorios naturales, cuando vienen estructurados por altos niveles del efecto (18 localidades), tamaños mínimos de muestra (localidades con 2 UPZ's), así como desbalanceados, tal como sucede con los datos de ésta aplicación. Determinar condiciones de factibilidad, permitiría por ejemplo, adicionar estructuras de correlación espacial en los errores.

6. Evaluar técnicas y estrategias de reducción del error cuadrático medio estimado. En el trabajo se utilizó mixturas finitas y efectos aleatorios dados por ACP, pero de los resultados se observa que el impacto principalmente se obtiene con fines de pronóstico mas no para la mejorar la precisión y los resultados pueden deberse a la estructura de la información más que a las técnicas.

APÉNDICE A

Algoritmos de estimación

A.1. Algoritmo EM

El algoritmo EM propuesto por Dempster *et al.* (1977), consta de dos etapas en cada iteración j, la primera sobre el cálculo de esperanza y la segunda para el proceso de maximización, lo cual motiva el nombre del algoritmo.

Paso E: Esperanza

1. Elegir un valor inicial para $\lambda^{(j=0)}$

2. Teniendo en cuenta que W es no observable, no es posible maximizar directamente la verosimilitud de la ecuación (2.9), por lo cual se maximiza su valor esperado, condicionado a los datos observados $\widehat{\theta}$ y el valor inicial $\lambda^{(j=0)}$, denominada como la "Función Q":

$$\begin{aligned} Q(\lambda|\lambda^{(j)}) =& E\left[\log L(\Lambda|\widehat{\theta}, w) \mid \widehat{\theta}, \lambda^{(j)}\right] \\ =& \sum_{d=1}^{m}\sum_{k=1}^{K} E\left[w_{dk}\log\pi_k + \log f_k(\widehat{\theta}_d|\lambda_k) \mid \widehat{\theta}, \lambda^{(j)}\right] \\ =& \sum_{d=1}^{m}\sum_{k=1}^{K} E\left[w_{dk} \mid \widehat{\theta}, \lambda^{(j)}\right]\left[\log\pi_k + \log f_k(\widehat{\theta}_d|\lambda_k)\right] \\ =& \sum_{d=1}^{m}\sum_{k=1}^{K} \widehat{\tau}_{dk}^{j}\left[\log\pi_k + \log f_k(\widehat{\theta}_d|\lambda_k)\right], \end{aligned}$$

donde $\widehat{\tau}_{dk}^{j}$ representan la máxima probabilidad, entre el conjunto de probabilidades a posteriori de que el valor observardo $\widehat{\theta}_d$ pertenezca a alguna de las componentes $k = 1, \ldots, K$, en la i–ésima iteración.

Paso M: Maximización

3. Elegir un nuevo valor para $\lambda^{(j+1)}$, seleccionando el que maximiza la función Q

$$\lambda^{(j+1)} = \arg\{Max_{\lambda} Q(\lambda|\lambda^{(j)})\}$$

4. Hacer $j = j + 1$

5. Volver a 2

6. Repetir los pasos hasta alcanzar convergencia, esto es, cuando la diferencia relativa de la verosimilitud $L(\Lambda|\widehat{\theta} = \widehat{\theta}_d)$ en la iteración $j + 1$ y la j, sea menor que ϵ.

A.2. Monte Carlo vía Cadenas de Markov y Diagnóstico de Convergencia

Siguiendo a Jiménez (2015), una cadena de Markov es una secuencia de variables $Y_0, Y_1, Y_2 \ldots$, tal que:

$$Pr(Y_{t+1} \in A|Y_0, Y_1, Y_2, \ldots, Y_t) = Pr(Y_{t+1} \in A|Y_t),$$

lo cual indica que la probabilidad condicional de que la variable aleatoria se encuentre en el estado A, en el periodo $t + 1$, depende únicamente del valor del estado en el periodo t y es independiente del pasado. Esta propiedad es importante porque garantiza que conforme la variable aleatoria cambie de estado, la cadena irá olvidando su valor inicial, es decir, irá alcanzando convergencia hacia una distribución que no depende ni del paso t, ni del valor inicial; esta distribución es conocida como **estacionaria o invariante**, de tal manera que al alcanzarla, se estará muestreando en la densidad objetivo. Estas cadenas, tienen propiedades muy importantes que las conecta con integración de Monte Carlo, debido a que en general son resumidas mediante la función $\overline{f}_n = \frac{\sum_t Y_t}{n}$, denominada media ergódica, cuyo comportamiento asintótico dá origen al **teorema ergódico**, que señala que si el valor esperado de la cadena existe, entonces su media ergódica converge en media cuadrática a su esperanza matemática: $P\left[\overline{f}_n \to E\left[f(Y)\right]\right] = 1$.

En el contexto Bayesiano, la construcción de cadenas de Markov busca que la distribución estacionaria sea la distribución a posteriori de los parámetros. De forma genérica, sea $(Y_1, Y_2, \ldots, Y_n)'$ una muestra aleatoria de una función de densidad $f_{Y_i}(y; \Theta)$ donde $\Theta = (\theta_1, \theta_2, \ldots, \theta_s)$ con distribución a priori $p(\Theta)$, entonces la distribución a posteriori, $\pi(\Theta)$, está dada por:

$$\pi(\Theta) \propto \prod_i f_{y_i}(y; \Theta) p(\Theta)$$

Para el caso $\pi(\Theta)$ es desconocida, de manera que para conocer su comportamiento se extraen muestras aleatorias $\widetilde{\Theta}_1, \widetilde{\Theta}_2, \ldots, \widetilde{\Theta}_r$, a partir de algoritmos que generan cadenas de Markov:

A.2.1. Algoritmo Metropolis–Hastings

Se considera una densidad $q(\cdot|\theta_t)$ conocida, denominada de transición (Metropolis *et al.*, (1953) ; Hastings (1970)), tal que $\frac{\pi(\theta)}{q(\theta|y,\theta)}$ es aproximadamente constante, el algoritmo Metropolis–Hastings se basa en mecanismos de aceptación y rechazo para la generación de las observaciones de $\pi(\theta)$, de la siguiente manera:

1. Se asume un valor inicial $\theta^{(0)}$, en el paso incial $j = 1$,
2. Se propone un nuevo valor $\theta^{(p*)}$ a partir de la densidad de transición $q(\theta^{(p*)})$,
3. Se define la probabilidad de aceptación y rechazo, sobre la cual se decide el cambio de un estado en $t-1$ a otro en t, como:

$$\alpha(\theta^{(j-1)}, \theta^{(p*)}) = \frac{\pi(\theta^{(p*)})q(\theta^{(j-1)}, \theta^{(p*)})}{\pi(\theta^{(j-1)})q(\theta^{(j-1)})}$$

4. Comparar la probabilidad de aceptación, $\alpha(\theta^{(j-1)}, \theta^{(p*)})$, frente a $u \sim U[0,1]$: Si $\alpha(\theta^{(j-1)}, \theta^{(p*)}) > u$, se acepta el cambio de estado y por lo tanto, se toma $\theta^{(j)} = \theta^{(p*)}$. En caso de rechazar el valor propuesto, se itera volviendo a generar un valor propuesto $\theta^{(p*)}$ mediante $q(\theta^{(p*)})$ del paso 2.
5. El proceso se repite reiniciando $j = j + 1$, hasta lograr convergencia.

El algoritmo garantiza la convergencia a la distribución estacionaria, independientemente de la forma de la distribución $q(\cdot|\cdot)$, sin embargo, la distribución propuesta debe tener la misma dimensión que la objetivo y debe tener la capacidad de generar valores que se acepten, en otro caso, la cadena puede durar largos periodos de tiempo en el mismo estado, de tal manera que la selección adecuada de la propuesta puede convertirse en un problema (Jiménez, 2015). Una vez se obtenga un valor que pertenezca a la distribución estacionaria, los siguientes también lo haran.

A.2.2. Algoritmo Muestreo de Gibbs

Para los s–parámetros de $\Theta = (\theta_1, \theta_2, \ldots, \theta_s)$ es posible establecer las funciones de densidad condicionadas completas, que determinan la a posteriori conjunta $\pi(\Theta)$ de la siguiente manera, (Gamerman y Freitas, 2006):

1. Se asume un vector de valores iniciales $\Theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_s^0)$ en el paso $j = 1$.
2. Se obtienen los valores actualizados al paso j, $\Theta^j = (\theta_1^j, \theta_2^j, \ldots, \theta_s^j)$ a partir de las distribuciones condicionales completas:

$$\begin{aligned}
\theta_1^j &\sim \pi(\theta_1|\theta_2^{j-1}, \ldots, \theta_s^{j-1}) \\
\theta_2^j &\sim \pi(\theta_2|\theta_1^{j-1}, \ldots, \theta_s^{j-1}) \\
&\vdots \\
\theta_s^j &\sim \pi(\theta_s|\theta_1^{j-1}, \ldots, \theta_{s-1}^{j-1})
\end{aligned}$$

3. Actualizar $j = j + 1$, hasta lograr convergencia.

A.2.3. Algoritmo Metropolis–Hastings con Gibbs

Como se espera, este algoritmo combina los algoritmos presentados en los apartados A.2.1 y A.2.2. Por un lado, el algoritmo de Metropolis–Hastings requiere de una especificación adecuada de la distribución propuesta $q(\cdot|\cdot)$ y por el otro el Muestreo de Gibbs el conocimiento de las distribuciones condicionales completas. De tal manera que la combinación de los algoritmos opera de la siguiente manera:

1. Se asume un vector de valores iniciales $\Theta^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_s^0)$ en el paso $j = 1$.
2. Se obtienen los valores actualizados al paso j, $\Theta^j = (\theta_1^j, \theta_2^j, \ldots, \theta_s^j)$ a partir de las distribuiciones propuestas $q_i(\theta_i|\cdot)$ para cada uno de los parámetros de interés y $i = 1, \ldots, s$:

$$\begin{aligned} \theta_1^{(p*)_j} &\sim q_1(\theta_1^{j-1}|\cdot) \\ \theta_2^{(p*)_j} &\sim q_2(\theta_2^{j-1}|\cdot) \\ &\vdots \\ \theta_s^{(p*)_j} &\sim q_s(\theta_s^{j-1}|\cdot) \end{aligned}$$

3. Actualizar $\theta_i^{j-1} = \theta_i^{(p*)_j}$ siempre y cuando se acepte el nuevo valor propuesto de acuerdo con el criterio de probabilidad de aceptación $\alpha(\theta_i^{(j-1)}, \theta_i^{(p*)_j})$.
4. Actualizar el contador $j = j + 1$ hasta lograr convergencia.

A.2.4. Diagnóstico de convergencia

Frente al tratamiento operativo de la generación de las Cadenas de Markov, generalmente se recurre a descartar un número alto de las primeras observaciones para eliminar el efecto de los valores iniciales, etapa denominada de calentamiento o burn-in; adicionalmente, a pesar de que la convergencia se garantiza teóricamente cuando los pasos o iteraciones tienden a infinito, en la práctica se establecen mecanismos de parada, principalmente por limitaciones computacionales, entonces, para validar que se ha logrado muestrear valores de la distribución a posteriori y por lo tanto las inferencias son viables, es necesario validar la convergencia de las cadenas mediante criterios como (Jiménez, 2015):

A.2.4.1. Métodos gráficos

Gráficos de trayectorias: Para verificar visualmente, que la cadena está explorando adecuadamente el espacio de estados, se grafican los valores θ_t versus t, que indicaran estacionariedad si no se producen desviaciones del trayecto. Otra alternativa es generar varias cadenas a partir de distintos puntos de arranque, permitiendo verificar que las cadenas logran mezclar adecuadamente los valores muestreados a medida que aumenta el número de observaciones.

Gráficos de autocorrelación: Permiten verificar que la cadena efectivamente solo depende del periodo anterior y que logró ser independiente de todos los estados de periodos anteriores.

Histogramas a posteriori: Permiten analizar la estructura de la función de densidad a posteriori de los parámetros de interés.

A.2.4.2. Pruebas de convergencia

Criterio de Geweke: Geweke (1992) sugiere la comparación de las medias ergódicas generadas a partir de datos particionados, correspondientes a una fracción de los primeros valores (después del burn-in) y otra de los valores finales de la cadena. El supuesto es que, si la diferencia del estadístico de prueba es grande comparada con una Normal, no se logró la convergencia. Sin embargo, estadístico de prueba cercano a 0 no garantiza la convergencia y por lo tanto es necesario apoyar la decisión con otros criterios y mediante la inspección visual.

Criterio de Heidelberger y Welch: El criterio de Heidelberger y Welch (Heidelberger, P. y Welch, P. (1983)), utiliza dos perspetivas para revisar la convergencia de las Cadenas de Markov, una desde el concepto de estacionariedad y otra evaluando la precisión de la estimación. La *prueba de estacionariedad*, contrasta la hipótesis nula de que las muestras tomadas provienen de una distribución estacionaria, bajo el estadístico de Cramér Von Mises que es aplicado de forma sucesiva, sobre el 100 % de la cadena, luego sobre el 90 %, el 80 % hasta el 50 %, en cada caso verificando que no se rechace la hipótesis nula. En caso de rechazarla la prueba en cada una de las fracciones de la cadena considerada y hasta el 50 %, significa que la cadena no logró convergencia y que el número de iteraciones es escaso para lograrlo.

Por su parte la prueba *half–width* es realizada sobre el porcentaje de la cadena que logró probar la convergencia a partir del test de estacionariedad y consiste en comparar el ancho medio del intervalo de credibilidad al $(1-\alpha)\,\%$ para la media, con el estimador puntual, generando así una medida de precisión de la estimación, que debería ser mínimo para indicar que efectivamente el número de iteraciones fue suficiente para lograr convergencia; de otro modo, se necesita más pasos para estimar la media con suficiente precisión.

A.3. Método Bootstrap

El método de muestreo bootstrap planteado por Efron B. (1979), hace parte de métodos Monte Carlo no paramétricos utilizados con el propósito de estimar características poblacionales mediante procedimientos de remuestreo (Gil N. (2014)).. Considérese el parámetro de interés θ y un estimador $\widehat{\theta}$, entonces la estimación bootstrap de la distribución de $\widehat{\theta}$ se obtiene bajo el siguiente procedimiento:

1. Generar B muestras bootstrap $x^{*(b)} = \{x_1^*, \ldots, x_n^*\}$ mediante muestreo con reemplazamiento a partir de la muestra observada $x^{*(b)} = \{x_1, \ldots, x_n\}$.

2. Calcular $\widehat{\theta}^{(b)}$ que corresponde a la b–ésima réplica a partir de la muestra boostrap correspondiente.

3. La estimación bootstrap de la función de probabilidad empírica $F_{\widehat{\theta}}(\cdot)$ es obtenida mediante las B réplicas de $\widehat{\theta}^{(b)}$, $\{\widehat{\theta}^{(1)}, \ldots, \widehat{\theta}^{(B)}\}$, por lo tanto es posible obtener medidas de centralidad y dispersión, de dicha distribución.

Debido a que el método bootstrap presenta convergencia a la verdadera distribución del parámetro, los estadísticos asociados a la media convergen al verdadero valor del parámetro.

APÉNDICE B

Anexos de la aplicación

B.1. Metodología estadística y operativa de la encuesta EMB 2011

La Encuesta Multiproposito de Bogotá fue realizada por primera vez en 2011 por la Secretaría Distrital de Planeación - SDP, con apoyo metodológico y operativo del Departamento Administrativo Nacional de Estadística (DANE), responsable de las Encuestas de Calidad de Vida del país, todos los detalles metodológicos y resultados pueden ser consultados en el sitio Web de la entidad, http://www.sdp.gov.co/portal/page/portal/PortalSDP. Se extrae en este apartado, una breve descripción del componente metodológico estadístico de la encuesta:

Tipo de operación estadística: La Encuesta Multipropósito Bogotá (EMB, 2011) es una encuesta por muestreo probabilístico dirigida a hogares con entrevista cara a cara e informante directo.

Universo: El universo para la EMB está compuesto por los hogares particulares y la población civil no institucional existente en el año 2011 en la parte urbana del distrito capital.

Población objetivo: La población objetivo está compuesta por los hogares particulares y población civil no institucional existente en el año 2011 en la parte urbana del distrito capital, se excluyen:

- Hogares ubicados en las zonas rurales de la ciudad en Usaquén, Chapinero, Santafé, San Cristóbal, Usme, Suba, Ciudad Bolívar y Sumapaz.
- Cárceles o centros de rehabilitación penitenciarios, orfanatos o albergues infantiles, hogares geriátricos o asilos de ancianos, conventos, seminarios o monasterios, internados de estudio, cuarteles guarniciones o estaciones de policía, campamentos de trabajo, albergues para desplazados y reinsertados, centros de rehabilitación no penitenciarios, ni unidades económicas o agropecuarias.

Cobertura y desagregación geográfica: Cabecera urbana de Bogotá (19 localidades) desagregada por localidad y estrato socio-económico.

Periodo de recolección: El periodo de recolección de la información comprende los días 7 de Febrero hasta 7 de abril de 2011.

Método de recolección: Dispositivos Móviles de Captura (DMC), por medio de un formulario para recolección de información, implementado en SysSurvey específicamente para la Encuesta Multipropósito para Bogotá, EMB 2011, en el cual se incorporan automáticamente las normas de validación y consistencia estipuladas para la misma. Esta información es recolectada por los encuestadores, con el informante directo.

Marco Muestral: Está conformado por el inventario cartográfico y el listado de viviendas y hogares a nivel de manzana, obtenidos de la información del Censo Nacional de Población y Vivienda de 2005 para la ciudad de Bogotá. Este marco se encuentra asociado al Código Homologado para Información Predial (CHIP) que identifica los predios a partir de la Base de Datos Catastral con corte a primero de marzo de 2010. El marco cuenta con 1.307.562 registros de predios urbanos, con algún uso habitacional, ubicados en 36.383 manzanas de las 19 localidades urbanas de Bogotá D.C.

Diseño Muestral: Probabilístico - estratificado por localidad, de conglomerados de segmentos conformados por grupos de tamaño promedio, 8 viviendas. Todas las vivienda, hogares y personas que conforman el segmento seleccionado, son observados.

Tamaño Muestral: Fueron seleccionados 1.860 segmentos, que dieron lugar a una muestra de viviendas de tamaño 15.832, 16.508 hogares y 54.614 personas.

Precisión: Medida en términos del error de muestreo, aproximadamente 5 % para indicadores con una prevalencia del 10 % a nivel de localidad.

Unidad de muestreo: Es el segmento, conformado por un conjunto de predios ubicados dentro de la misma manzana o manzanas cercanas. Todos los predios, viviendas, hogares y personas que conforman cada segmento corresponden a las unidades análisis.

Unidad de observación: Son los hogares y las personas que los conforman, al igual que las viviendas que habitan ubicadas dentro de un determinado predio. A cada predio se le asocian todas las viviendas, hogares y personas del cual están constituidos.

B.2. Estadísticas descriptivas gráficas de la estructura de las variables del modelo

	E	N	T	S	V
Min.	0.00000	0.08541	0.3260	0.0000	0.00000
1st Qu.	0.09095	0.22510	0.5415	0.1168	0.00000
Median	0.24951	0.28951	0.6431	0.1653	0.02385
Mean	0.25658	0.29596	0.6244	0.1634	0.04436
3rd Qu.	0.38044	0.36583	0.7325	0.2085	0.06769
Max.	0.67840	0.49795	0.8600	0.4298	0.56138

Tabla B.1. Medidas de localización de las variables incluidas en el IPM.

Porcentaje de hogares con carencias en las condiciones evaluadas en el IPM
E:Educación – T:Trabajo – S:Salud – V:Vivienda.

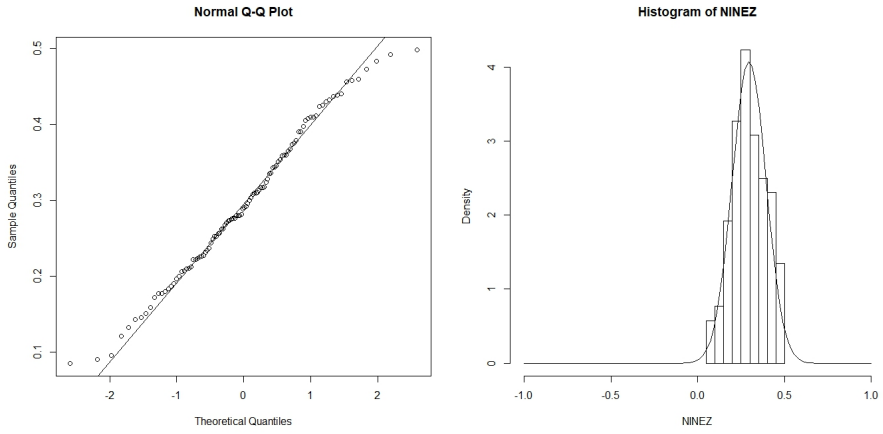


Figura B.1. Gráfica QQ – Histográma. Porcentaje de hogares con carencias en condiciones de la niñez y la juventud.

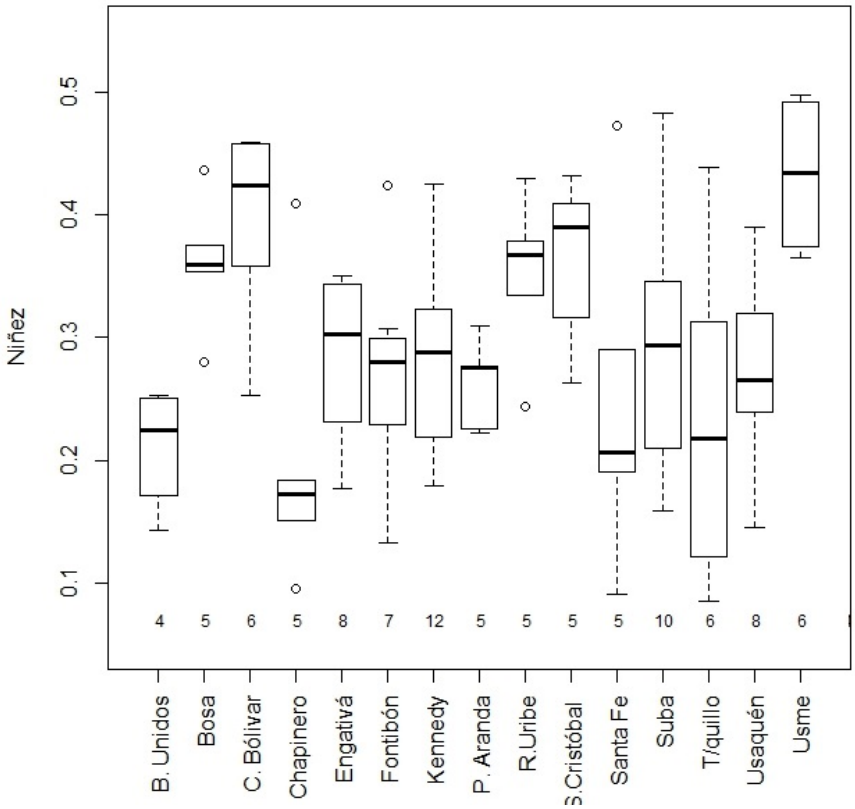


FIGURA B.2. Boxplot por localidades. Porcentaje de niños y jóvenes con carencias.

No incluye localidades con número de UPZ mínimo:La Candelaria (1 UPZ),)
Antonio Nariño, Los Mártires y Tunjuelito (2 UPZ's cada una)

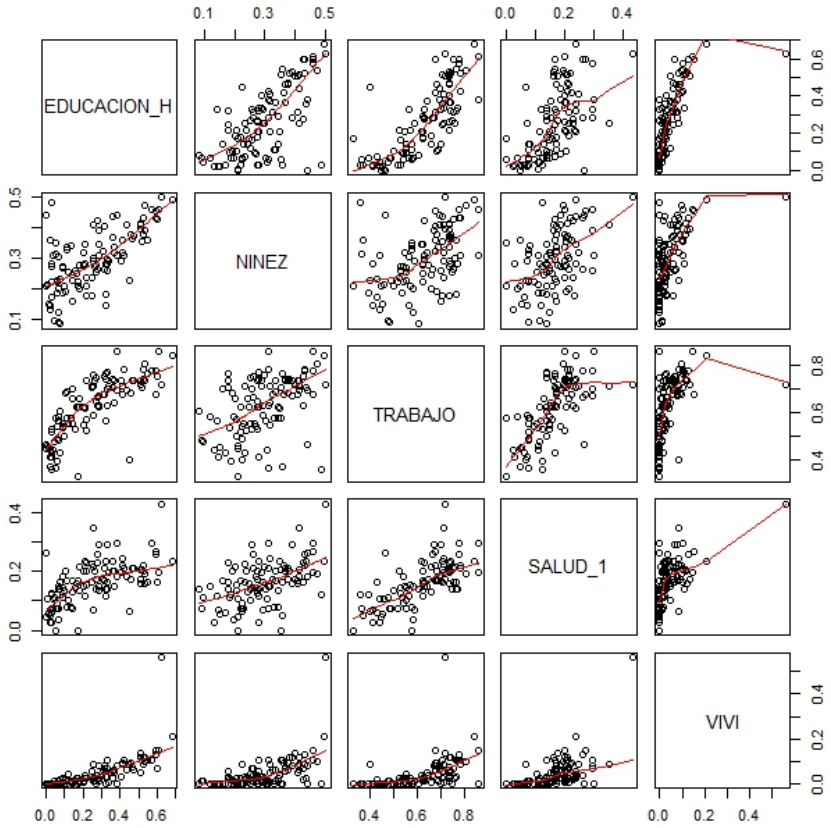


FIGURA B.3. Matriz de dispersión entre variables del IPM.

Porcentaje de hogares con carencias en las condiciones evaluadas en el IPM

	E	N	T	S	V
E	1.00	0.66	0.77	0.60	0.70
N	0.66	1.00	0.46	0.51	0.55
T	0.77	0.46	1.00	0.68	0.47
S	0.60	0.51	0.68	1.00	0.59
V	0.70	0.55	0.47	0.59	1.00

TABLA B.2. Matriz de correlación lineal entre variables del IPM.

Porcentaje de hogares con carencias en las condiciones evaluadas en el IPM
E:Educación – T:Trabajo – S:Salud – V:Vivienda.

Localidad	Total	Urbanas	Muestreadas	No muestreadas
Usaquén	10	9	8	1
Chapinero	6	5	5	0
Santa Fe	6	5	5	0
San Cristóbal	6	5	5	0
Usme	8	7	6	1
Tunjuelito	2	2	2	0
Bosa	5	5	5	0
Kennedy	12	12	12	0
Fontibón	8	8	7	1
Engativá	9	9	8	1
Suba	13	12	10	2
Barrios Unidos	4	4	4	0
Teusaquillo	6	6	6	0
Los Mártires	2	2	2	0
Antonio Nariño	2	2	2	0
Puente Aranda	5	5	5	0
La Candelaria	1	1	1	0
Rafael Uribe Uribe	5	5	5	0
Ciudad Bolívar	9	8	6	2
Sumapaz	1	0	na	na

TABLA B.3. Cobertura de muestral de las áreas pequeñas.

Número de UPZ's del territorio, muestreadas y no muestreadas

Las UPZ's no muestreadas obedecen a UPZ's no pertenecientes al marco muestral pues son de uso dotacional.

B.3. Resultados de modelos Mixturas Finitas

B.3.1. Modelos en Mixturas Finitas considerados

Cada línea de la tabla B.4 corresponde a un modelo implementado, en cada caso se indica en asterisco (*), qué variables fueron consideradas, así como la descripción de si presentó o no convergencia y, en caso de alcanzar la convergencia se presentan las respectivas medidas de los criterios de selección. Por ejemplo el modelo M1 corresponde a la implementación del modelo:

$$\begin{aligned} logit(\mu_{dk}) &= \beta_{0k} + \beta_{1k} Educacion_d + \beta_{2k} Trabajo_d + \beta_{3k} Salud_d + \beta_{4k} Vivienda_d, \\ log(\phi_{dk}) &= \gamma_{0k}. \end{aligned}$$

En todos los casos se considera el intercepto tanto para el modelo de la media como el de dispersión.

	Media μ				Dispersión ϕ				Resultados		
Variables	E	T	S	V	E	T	S	V	Converge	BIC	AIC
M1	*	*	*	*						-219,665	-253,917
M2					*	*	*	*		-146,888	-181,139
M3	*	*	*	*	*					-225,771	-265,292
M4	*	*	*	*	*	*			No		
M5	*	*	*	*	*		*		No		
M6	*	*	*	*	*			*	No		
M7	*	*	*	*		*				-217,041	-256,562
M8	*	*	*	*		*	*		No		
M9	*	*	*	*		*		*	No		
M10	*	*	*	*			*		No		
M11	*	*	*	*			*	*	No		
M12	*	*	*	*				*	No		
M13	*									-230,263	-248,706
M14	*				*					-239,317	-276,203
M15	*							*	No		
M16	*					*				-232,467	-256,180
M17	*						*		No		
M18		*								-194,862	-213,305
M19		*			*					-195,335	-219,048
M20		*			*		*			-202,752	-231,734
M21	*	*			*				No		
M22	*	*				*			No		
M23	*		*		*					-233,848	-262,831
M24	*		*			*			No		
M25	*			*	*					-232,938	-261,920
M26	*			*		*				-226,837	-255,819
M27	*	*	*		*				No		

TABLA B.4. Resultados de los modelos alternos de Mixtura Finitas para selección.

E:Educación – T:Trabajo – S:Salud – V:Vivienda.

B.3.2. Código R para el cálculo de residuales re_d^p y re_d^w para mixturas finitas Beta

```
set.seed(1234567)

r4<-betamix(formula = ninez ~ educacion|educacion, data = datos,
            k = 2:4,  nstart = 10, link = "logit", link.phi = "log")

clusters_r4<-cbind(datos,clusters(r4))
clusters_r4<- data.frame(clusters_r4)
clusters_r4$ordenini<- 1:nrow(clusters_r4)
clusters_r4<- clusters_r4 %>%
arrange(clusters.r4.)

cf   <- rbind(coef(r4, model = "mean", component = 1:3))
miu  <- numeric()
miuG <- numeric()

 for(ii in 1:3){
   miu <- c(miu, plogis(cf[ii, 1] + cf[ii, 2] * subset(clusters_r4, clusters.r4. =
       = ii)$educacion))
 }

cf <- rbind(coef(r4, model = "precision", component = 1:3))
phi<- numeric()
 for(ii in 1:3){
   phi <- c(phi, exp(cf[ii, 1] + cf[ii, 2] * subset(clusters_r4, clusters.r4. ==
       ii)$educacion))
  }

#PARA LA CONSTRUCCION DE LOS ERRORES PEARSON
clusters_r4$pred  <- miu
clusters_r4$prec  <- phi
clusters_r4$Vpred <-(miu*(1-miu))/(phi+1)
clusters_r4$Sdpred<-sqrt(clusters_r4$Vpred)
clusters_r4$resSt <-(clusters_r4$ninez - clusters_r4$pred)/clusters_r4$Sdpred

#PARA LA CONSTRUCCION DE LOS ERRORES PONDERADOS
Y.s   <- log(clusters_r4$ninez/(1-clusters_r4$ninez))
mu.s  <- digamma(clusters_r4$pred*clusters_r4$prec)-digamma((1-clusters_r4$pred)*
    clusters_r4$prec)
var.s <- trigamma(clusters_r4$pred*clusters_r4$prec)+trigamma((1-clusters_r4$pred)*
    clusters_r4$prec)
W <- diag((as.numeric(clusters_r4$prec*var.s*(clusters_r4$pred*(1-clusters_r4$pred)
    )^2)),nrow=length(Y.s))
H <- (W^0.5) %*%clusters_r4$educacion %*%solve(t(clusters_r4$educacion) %*%W%*%
    clusters_r4$educacion) %*%t(clusters_r4$educacion) %*%(W^0.5)
swr2<-(Y.s-mu.s)/sqrt(var.s*(1-diag(H)))
clusters_r4$swr2<-swr2
```

B.3.3. Estructura de UPZ's por componentes y localidades

Localidad	Componente 1	Componente 2	Componente 3
Usaquén	2	4	2
Chapinero		4	1
Santa Fe		4	1
San Cristóbal		5	
Usme		5	1
Tunjuelito		2	
Bosa		4	1
Kennedy	1	10	1
Fontibón	2	4	1
Engativá		6	2
Suba	1	7	2
Barrios Unidos	1	3	
Teusaquillo	2	3	1
Los Martires		2	
Antonio Nariño		2	
Puente Aranda	1	4	
Rafael Uribe Uribe	1	4	
Ciudad Bólivar		5	1

Tabla B.5. Distribución de UPZ por localidad, según componentes de Mixtura Finita.

B.4. Resultados de algunos modelos Beta Bayesianos

Adicional a los resultados de las tablas siguientes, se realizaron pruebas incluyendo la clasificación obtenida mediante mixturas finitas para analizar el comportamiento por grupos y su incidencia en el modelo, algunos resultados mostraron que la interacción Educación y el componente 1 era significativa para el modelo, no obstante Educación como variable de efecto fijo no lo era. Los criteríos, convergencias y significancias en general no orientaron la inclusión/exlcusión de variables por lo que no se describen en las tablas resumen.

Los resultados se distribuyen en dos tablas, en la primera, B.6, se describen las variables incluidas en cada modelo, así como la evaluación de significancia mediante los intervalos de credibilidad y, en la segunda tabla, B.7, se indica el resultado de los diagnósticos de convergencia (En caso de no alcanzar convergencia en alguna cadena, se indica en qué variable no se alcanzó) así como los criterios de selección y el cumplimento de los supuestos de los errores. Así por ejemplo, el modelo M1 corresponde a la siguiente forma funcional:

$$\begin{aligned} logit(\mu_d) &= \beta_0 + \beta_1 Educacion_d, \\ log(\phi_d) &= \gamma_0 + \gamma_1 Educacion_d. \end{aligned}$$

Para el cual en el modelo para la dispersión la variable Educación no resulta significativa. Adicionalmente, con la tabla de diagnóstico B.7 se relaciona que, el modelo M1, presenta

convergencia en las cadenas, el Desvío es de 216.1, el criterio AIC de 220.1 y BIC de 225.4, además que cumple con el supuesto de normalidad pero no el de independecia en los errores.

	Media μ					Dispersión ϕ				
Variables	I	E	T	S	V	I	E	T	S	V
M1	*	*				*	* (ns)			
M2	*	*				*				
M3	*	*	*	*	*	*	*	*	*	* (ns)
M4	*	*	*	*	*	*	*	*	*	
M5	*	*	*	*	*	*	* (ns)	* (ns)		
M6	*	*	*	*	*	*	* (ns)	* (ns)		
M7	*	*	*	*	*	*	* (ns)			
M8	*	*	*	*	*	*		* (ns)		
M9	*	*	*	*	*	*			* (ns)	
M10	*	*	*	*	*	*				
M11	*	*	*		*	*				
M12	*	*	*			*				
M13	*(ns)	*		*		*				
M14		*		*		*				

Tabla B.6. Variables incluidas en algunos modelos bayesianos implementados para selección.

I: Intercepto – E:Educación – T:Trabajo – S:Salud – V:Vivienda.
(ns): Variable no significativa de acuerdo a los intervalos de credibilidad del 95 %.

Variables	Convergencia	Deviance	AIC	BIC	Normalidad	Independencia
M1	Si	216,1	220,1	225,4	Si	No
M2	Si	209,8	213,8	219,1	Si	No
M3	No	1144,3	1154,3	1167,5		
M4	No	1057,7	1067,7	1080,9		
M5	No en ϕ	330,3	340,3	353,5		
M6	No en ϕ	330,3	340,3	353,5		
M7	No en ϕ	323,2	333,2	346,4		
M8	No en ϕ	338,9	348,9	362,0		
M9	No en ϕ	357,2	367,2	380,4		
M10	No S	332,6	342,6	355,7	Si	Si
M11	Si	222,7	230,7	241,2	Si	Si
M12	Si	163,2	169,2	177,1	Si	Si
M13	No S	90,9	96,9	104,8	Si	Si
M14	Si	83,6	87,6	92,8	Si	Si

Tabla B.7. Convegencia y criterios de selección de algunos modelos bayesianos implementados.

B.5. Resultados estimación bayesiana Beta mixta

B.5.1. Modelo de efecto aleatorio natural (Localidades)

Parámetros	mu.vect	sd.vect	2.5 %	25 %	50 %	75 %	97.5 %	N
alpha[1]	-1.115	0.169	-1.461	-1.227	-1.111	-1.000	-0.797	8*
alpha[2]	-2.015	0.274	-2.588	-2.184	-1.996	-1.826	-1.514	5*
alpha[3]	-2.181	0.344	-2.888	-2.398	-2.166	-1.946	-1.560	5*
alpha[4]	-1.799	0.766	-3.382	-2.279	-1.787	-1.267	-0.380	5*
alpha[5]	-1.709	0.892	-3.480	-2.320	-1.695	-1.102	0.030	6
alpha[6]	-0.579	1.579	-3.672	-1.608	-0.608	0.438	2.615	2
alpha[7]	-0.838	0.858	-2.468	-1.422	-0.847	-0.277	0.884	5
alpha[8]	-1.302	0.201	-1.705	-1.434	-1.295	-1.167	-0.924	12*
alpha[9]	-1.436	0.236	-1.913	-1.590	-1.430	-1.275	-0.997	7*
alpha[10]	-0.998	0.269	-1.539	-1.177	-0.995	-0.813	-0.481	8*
alpha[11]	-1.041	0.171	-1.386	-1.154	-1.037	-0.927	-0.707	10*
alpha[12]	-0.577	0.550	-1.621	-0.928	-0.585	-0.214	0.503	4
alpha[13]	-0.611	0.235	-1.098	-0.759	-0.610	-0.454	-0.161	6*
alpha[14]	-1.866	2.496	-6.938	-3.518	-1.701	-0.177	2.758	2
alpha[15]	1.750	11.217	-20.908	-5.548	2.156	8.813	24.509	2
alpha[16]	-1.020	0.399	-1.836	-1.281	-1.022	-0.744	-0.235	5*
alpha[17]	-1.296	0.419	-2.143	-1.570	-1.278	-1.014	-0.515	5*
alpha[18]	-1.363	0.557	-2.520	-1.724	-1.354	-0.984	-0.303	6*
beta[1]	1.038	1.007	-0.979	0.381	1.029	1.708	3.035	8
beta[2]	6.403	1.894	2.666	5.122	6.381	7.669	10.154	5*
beta[3]	3.699	0.947	1.897	3.052	3.688	4.309	5.642	5*
beta[4]	2.796	1.698	-0.368	1.627	2.737	3.877	6.242	5
beta[5]	2.591	1.574	-0.460	1.544	2.595	3.665	5.612	6
beta[6]	-0.580	4.277	-9.453	-3.319	-0.490	2.251	7.787	2
beta[7]	0.648	2.053	-3.474	-0.686	0.677	2.025	4.524	5
beta[8]	1.393	0.627	0.181	0.965	1.403	1.816	2.610	12*
beta[9]	3.102	1.328	0.431	2.233	3.101	4.000	5.649	7*
beta[10]	0.348	1.229	-2.002	-0.454	0.328	1.188	2.762	8
beta[11]	1.102	0.926	-0.699	0.485	1.108	1.714	2.892	10
beta[12]	-3.898	2.796	-9.825	-5.648	-3.731	-2.020	1.083	4
beta[13]	-15.700	4.774	-25.110	-18.857	-15.640	-12.551	-6.170	6*
beta[14]	2.574	8.822	-14.282	-3.373	2.119	8.442	20.034	2
beta[15]	-10.499	43.826	-99.578	-38.267	-12.312	18.153	78.021	2
beta[16]	-0.055	1.334	-2.804	-0.912	-0.031	0.822	2.484	5
beta[17]	1.698	0.952	-0.071	1.058	1.665	2.328	3.580	5
beta[18]	1.946	1.113	-0.195	1.176	1.933	2.700	4.174	6
phi	43.232	7.249	30.543	38.019	42.777	47.849	58.833	*
deviance	-260.719	10.627	-279.147	-268.294	-261.496	-253.958	-237.361	

*Tamaño de la cadena 100.000, Burn in 0.2, *Variables significativas*

Tabla B.8. Resumen Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc).

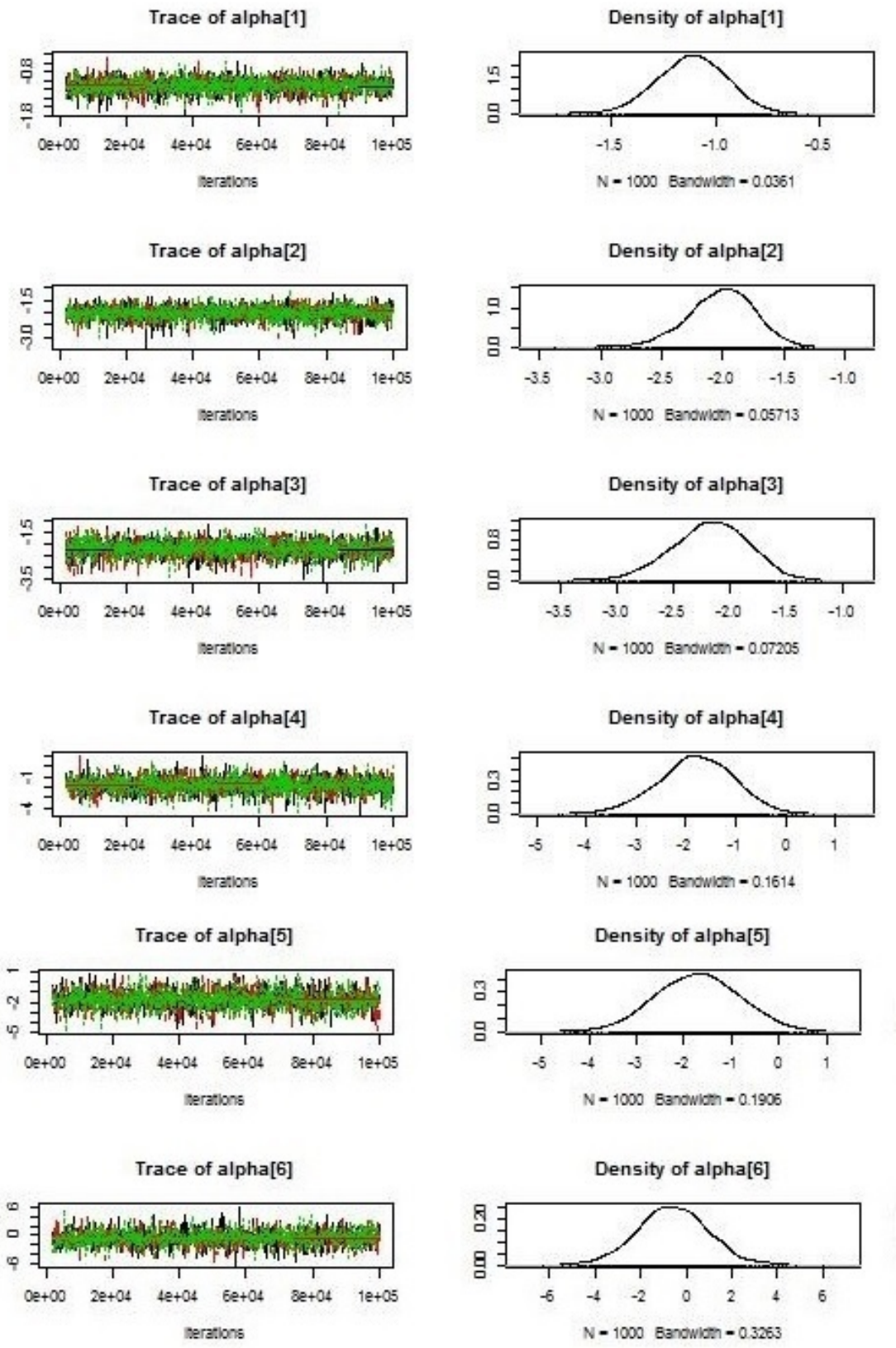


FIGURA B.4. Trayectorias de los interceptos del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 1.

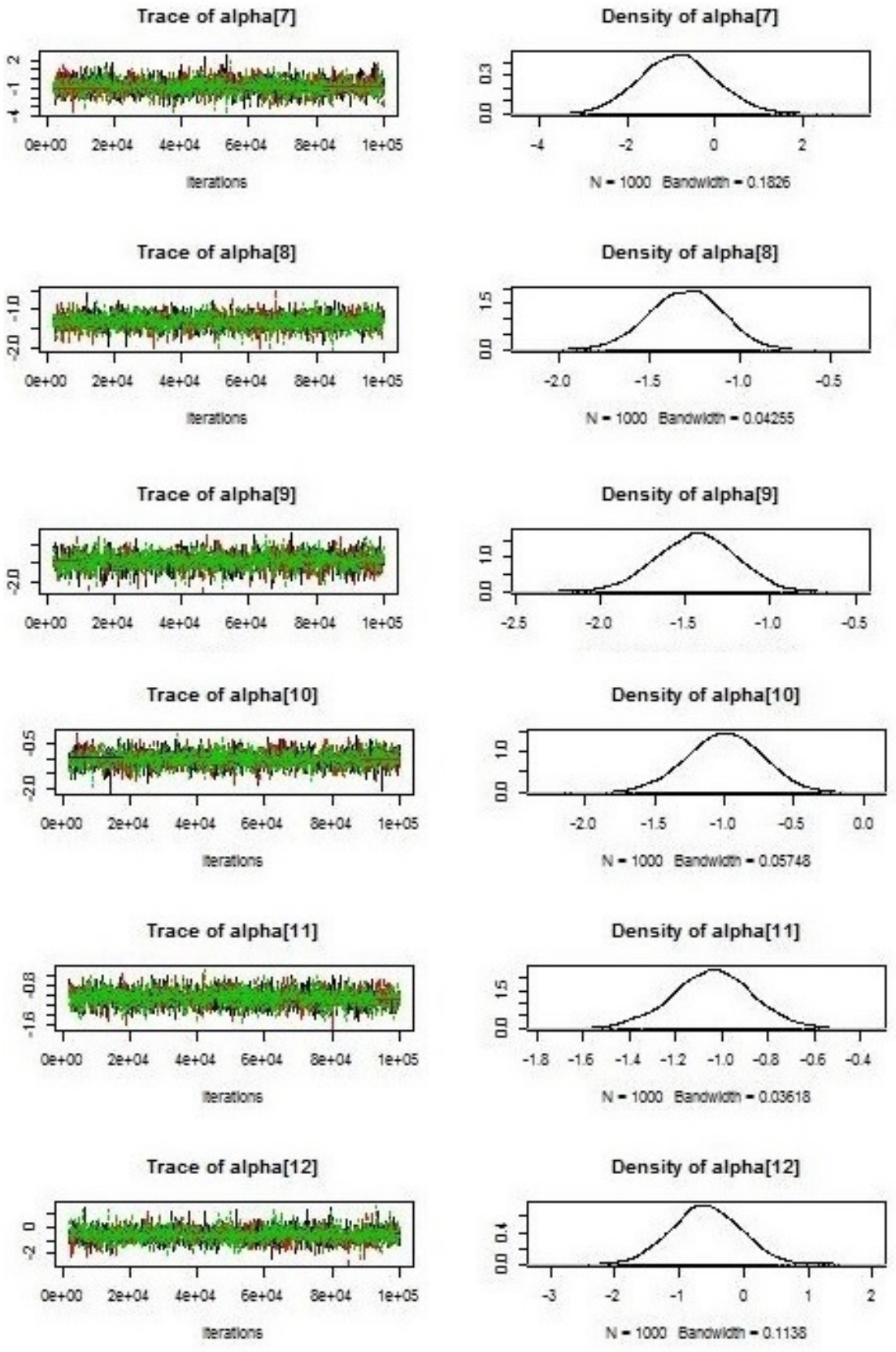


FIGURA B.5. Trayectorias de los interceptos del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 2.

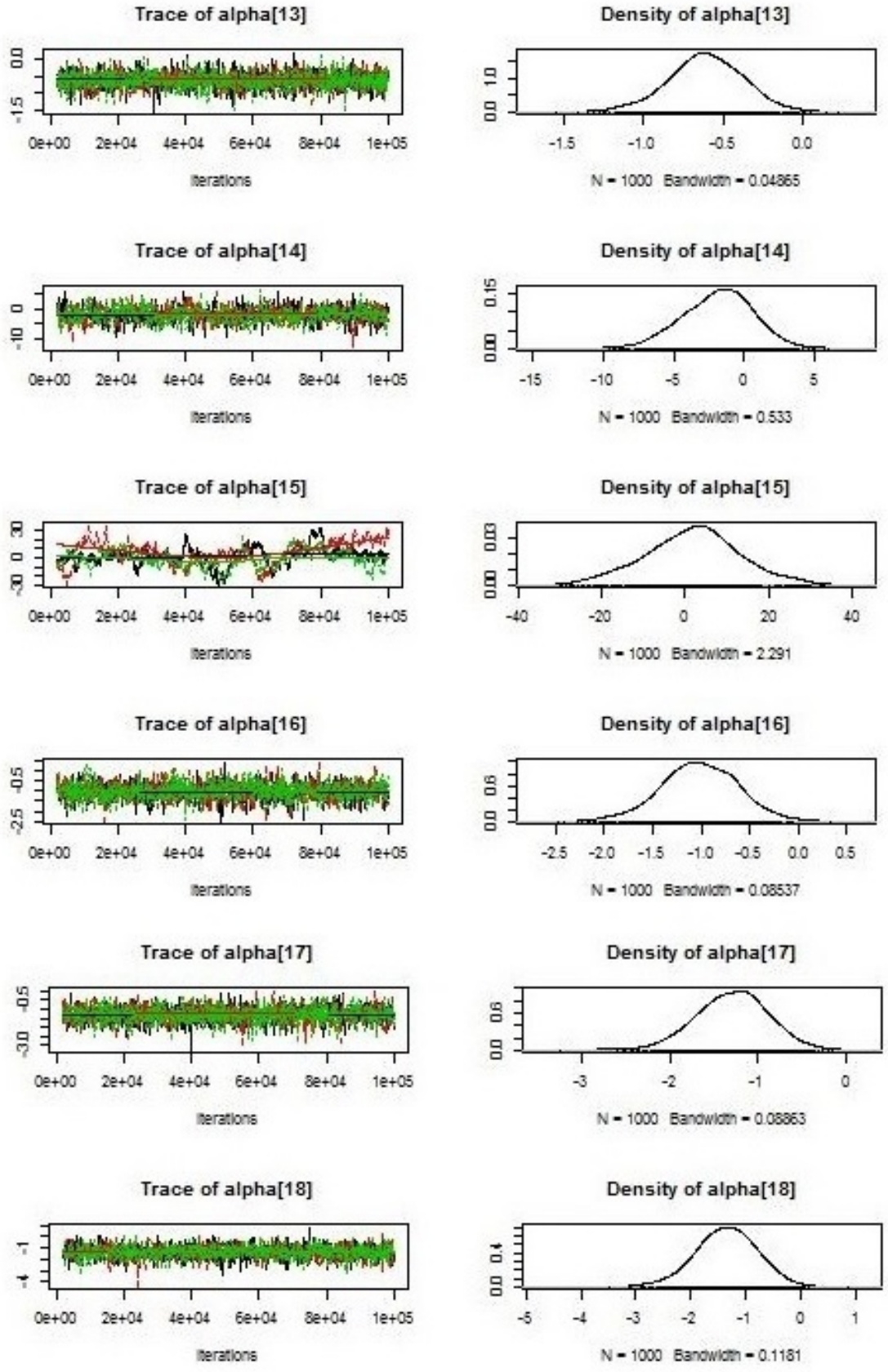


FIGURA B.6. Trayectorias de los interceptos del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc)3.

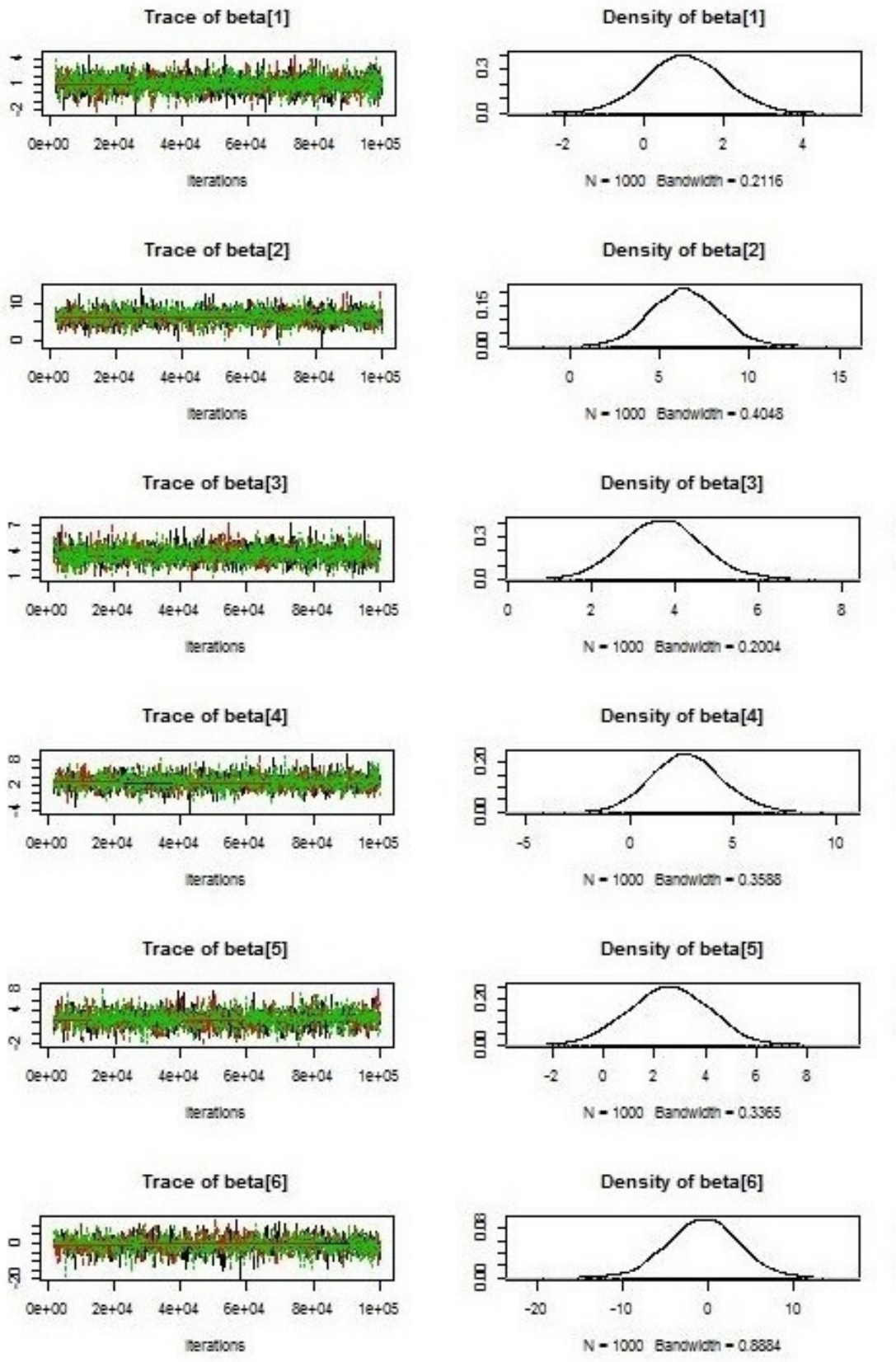


FIGURA B.7. Trayectorias de las pendientes del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 1.

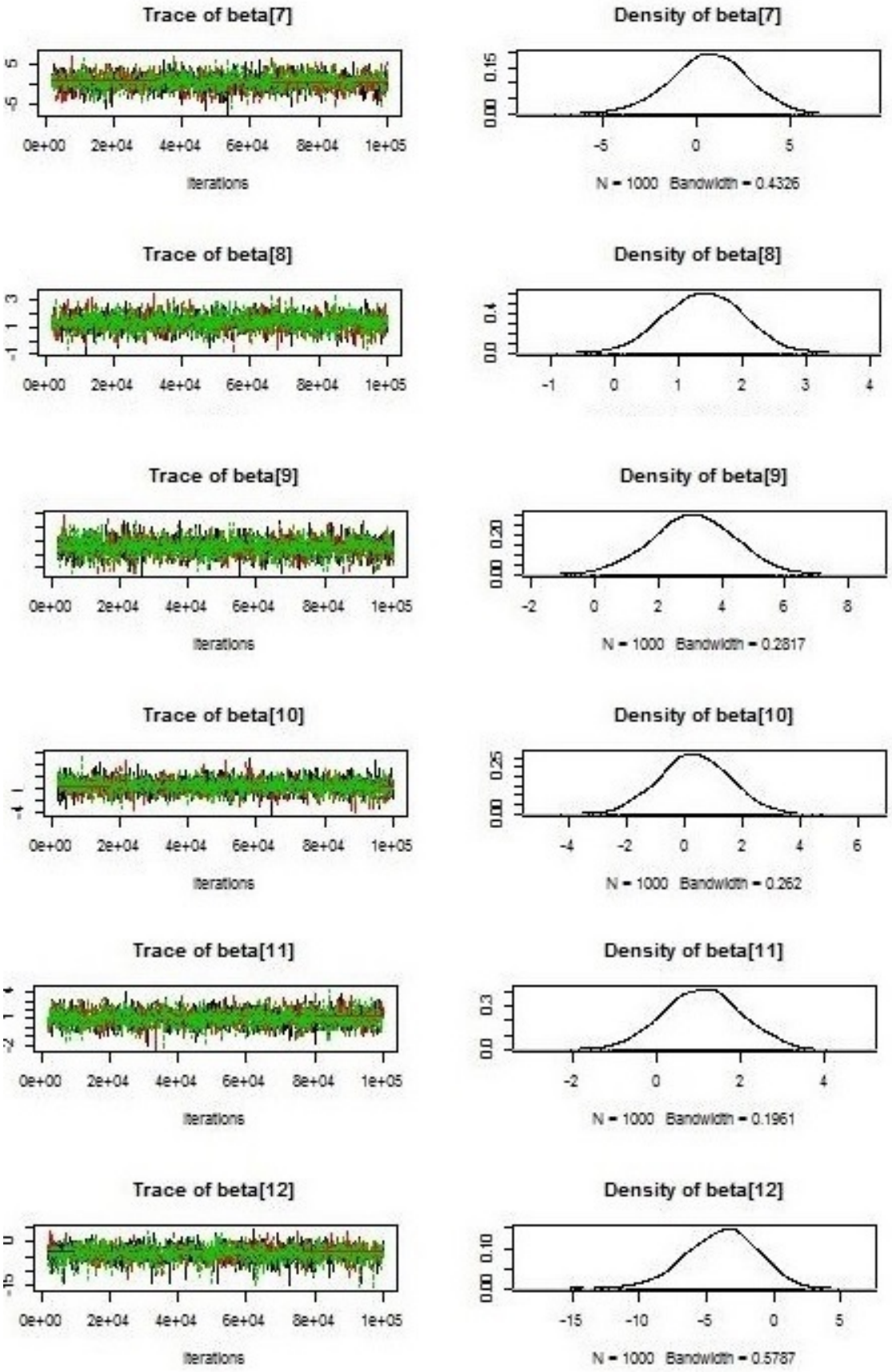


FIGURA B.8. Trayectorias de las pendientes del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 2.

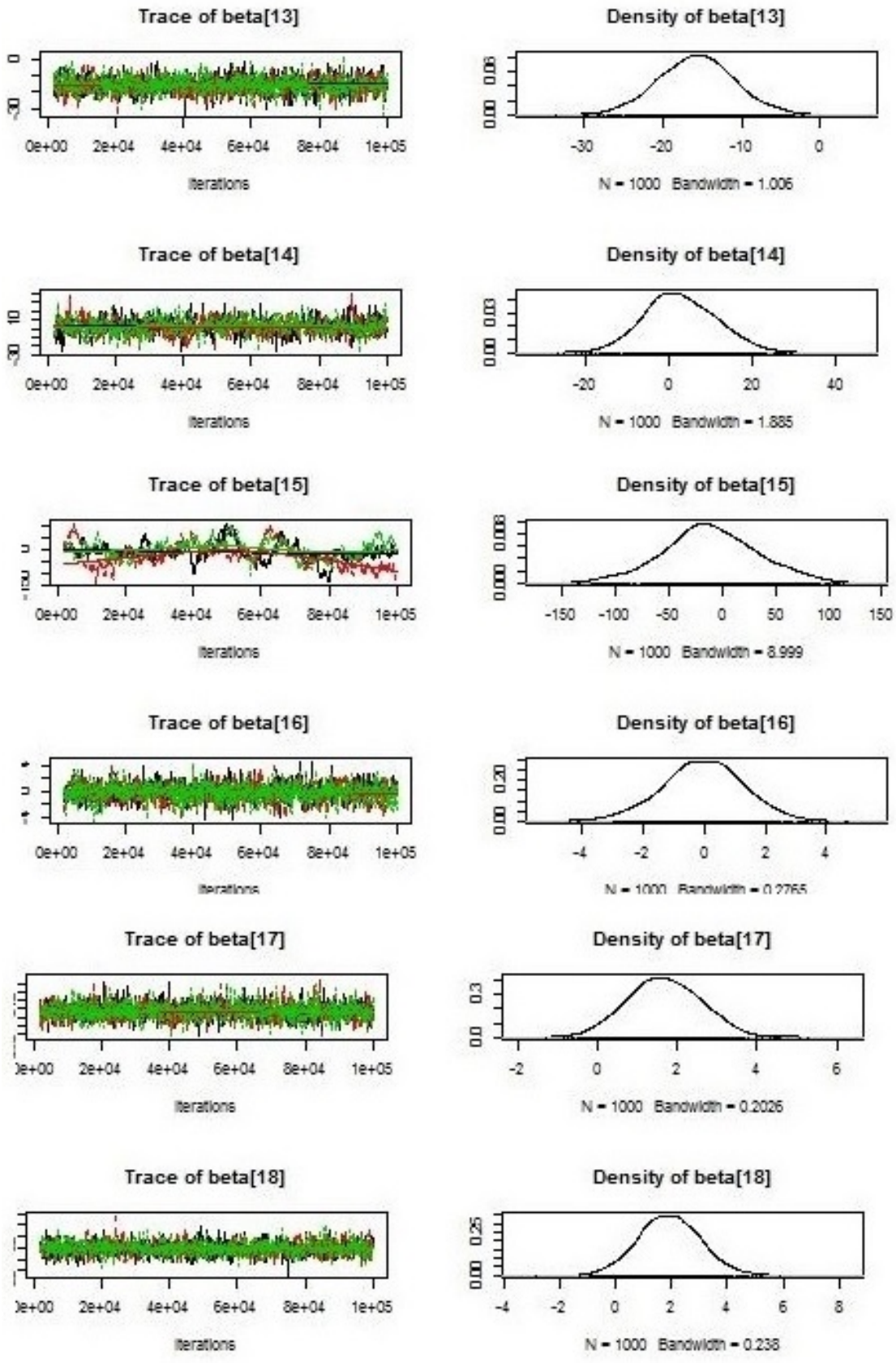


FIGURA B.9. Trayectorias de las pendientes del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 3.

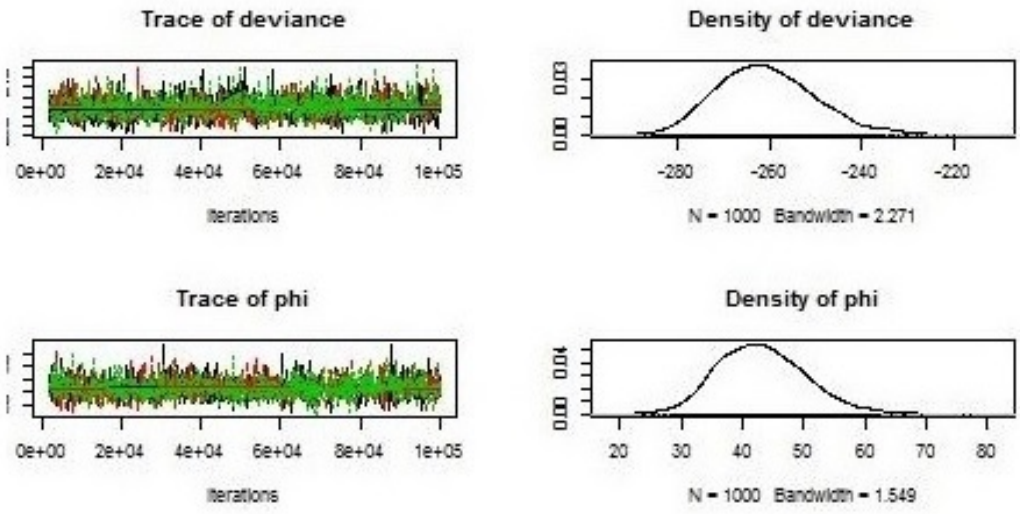
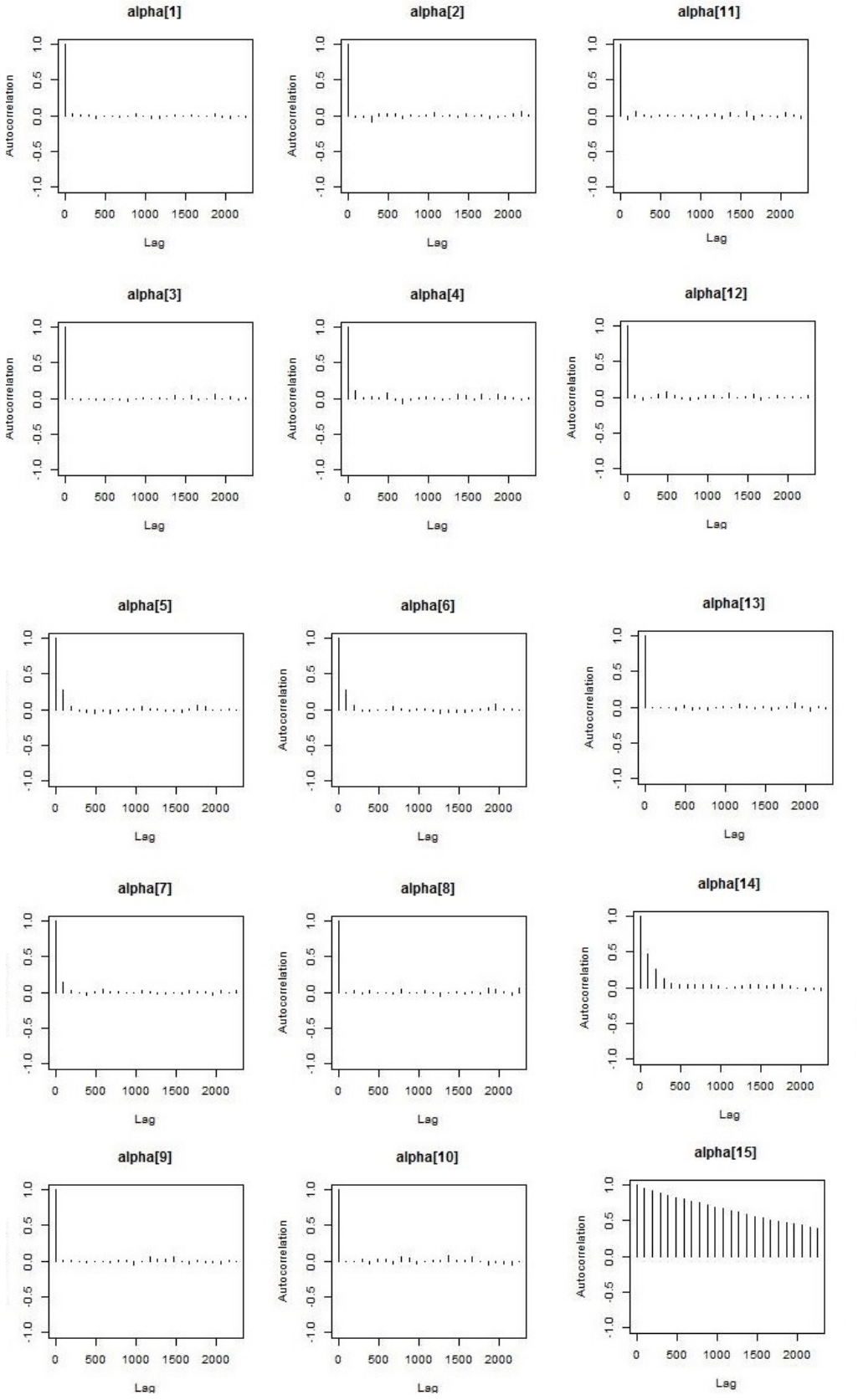


FIGURA B.10. Trayectorias de las pendientes del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 4.



alpha[15]: Intercepto para la localidad Antonio Nariño

FIGURA B.11. Autocorrelaciones del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 1.

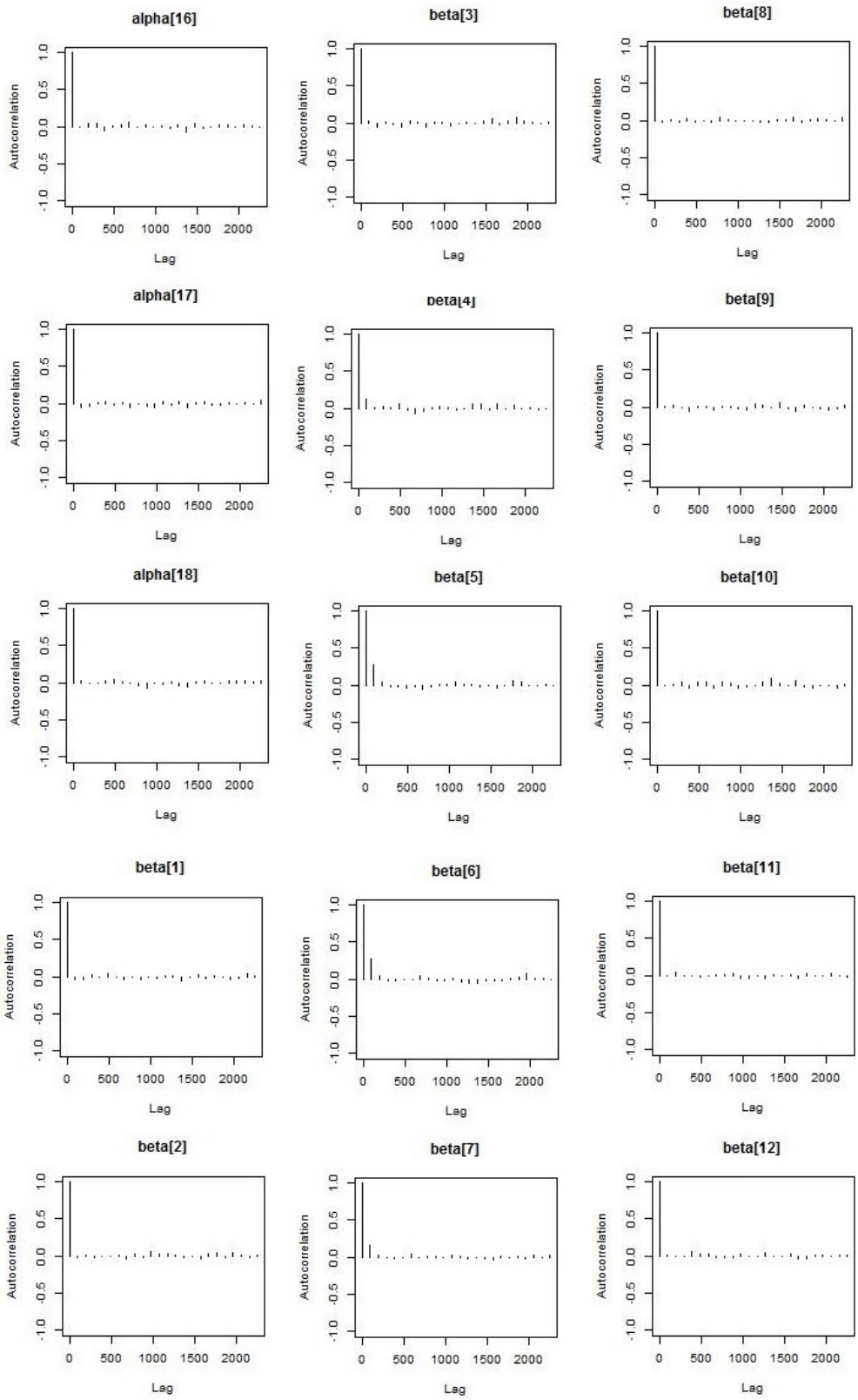
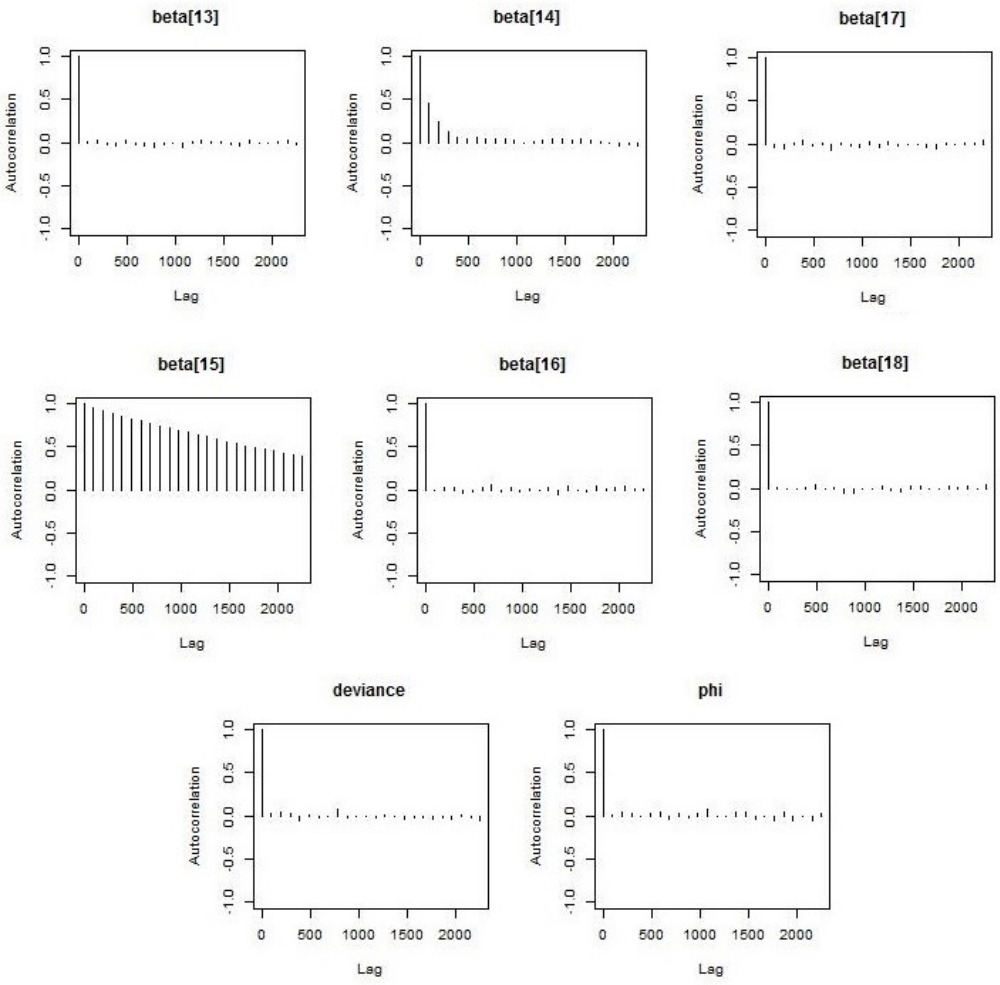


FIGURA B.12. Autocorrelaciones del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 2.



beta[15]: Pendiente para la localidad Antonio Nariño

FIGURA B.13. Autocorrelaciones del Modelo Bayesiano Mixto Beta con dispersión constante, efecto natural (BMBDc) 3.

B.6. Resultados de Análisis de clasificación para la conformación de clusters

B.6.1. ACP con todas las variables del IPM

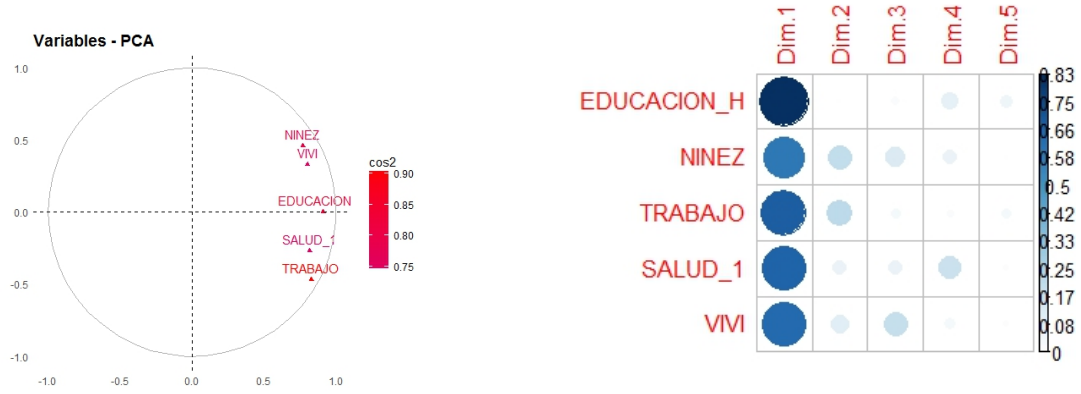


FIGURA B.14. Círculo de correlaciones del ACP y correlación de las variables con los factores.

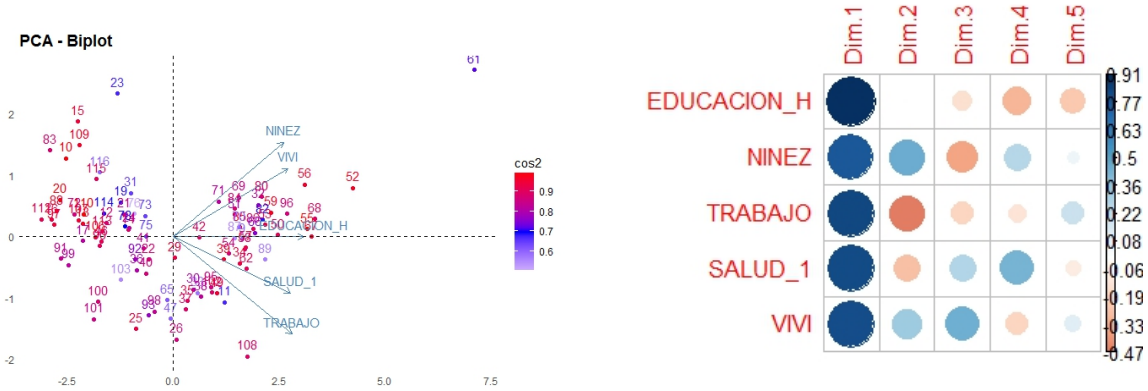


FIGURA B.15. Individuos en el plano factorial del ACP y correlación entre variables.

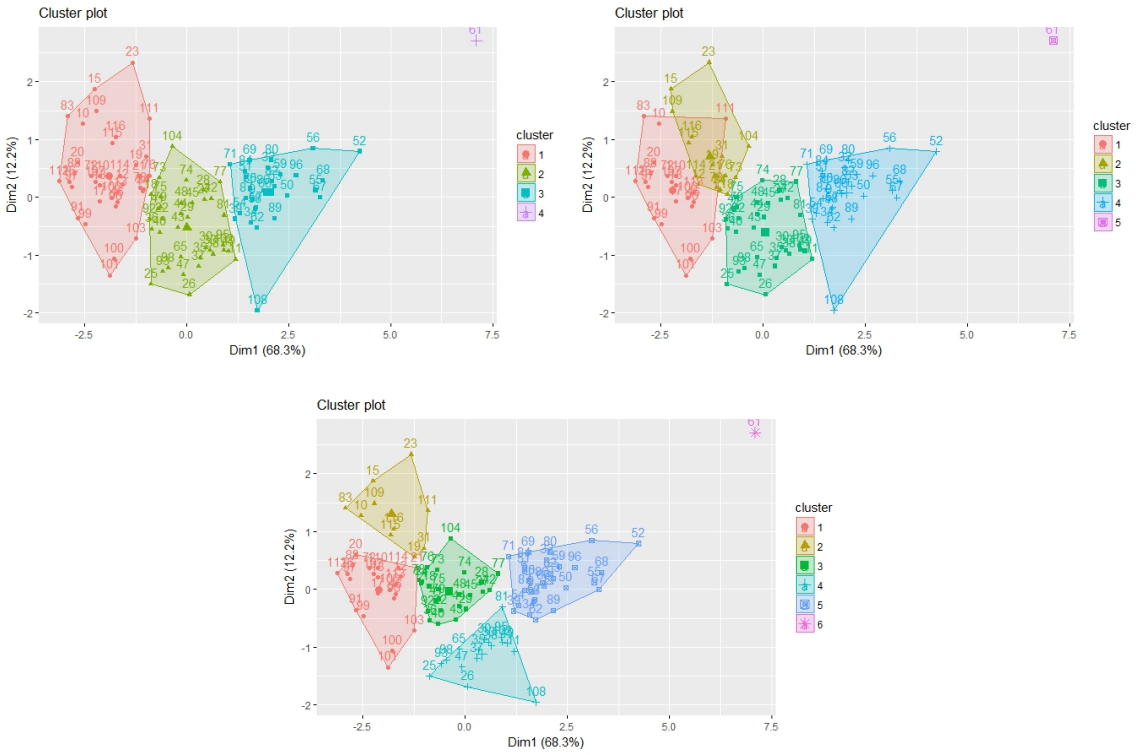


FIGURA B.16. Representación gráfica de los clusters tamaños 4,5 y 6

B.6.2. ACP con Educación y Trabajo

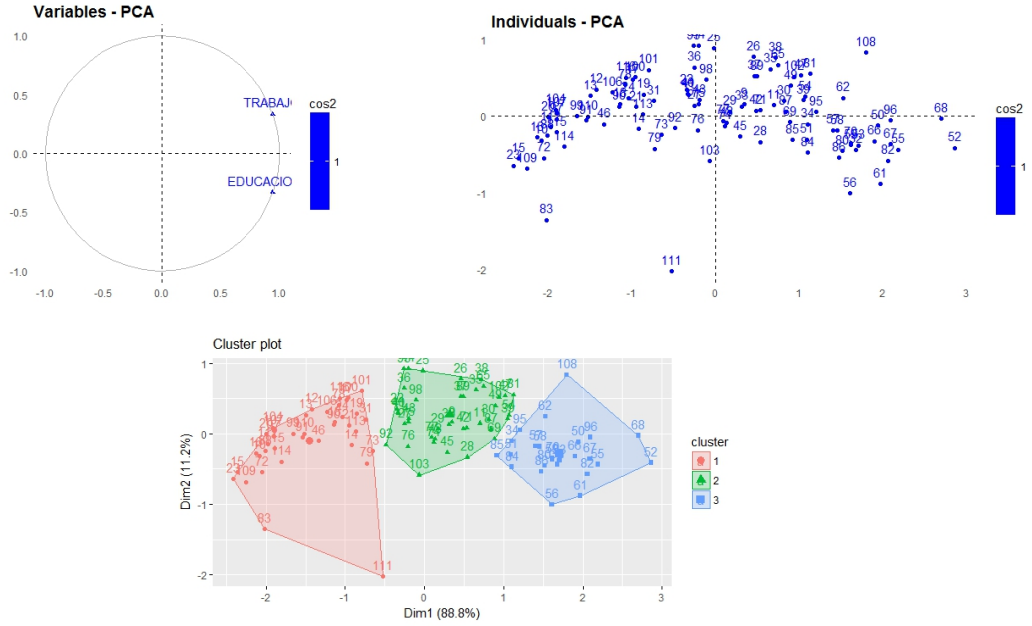


FIGURA B.17. Gráficas ACP para educación y trabajo en la construcción de efectos latentes

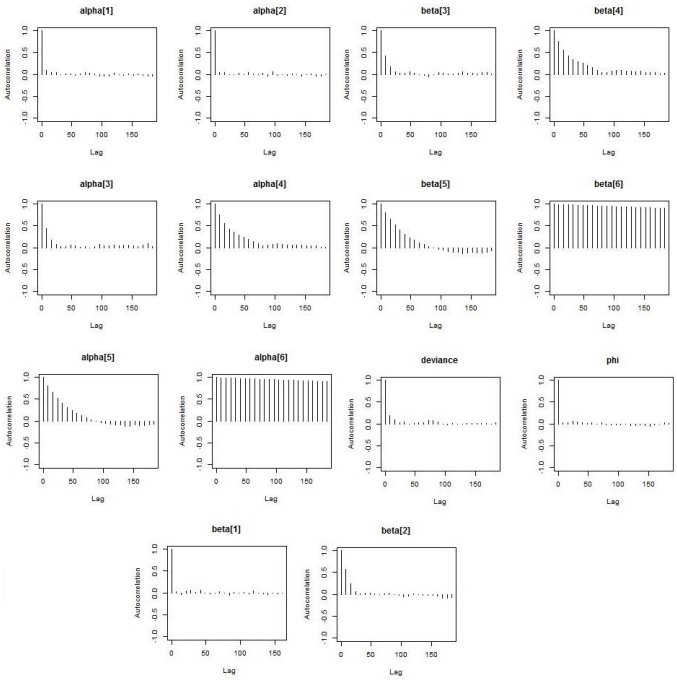


FIGURA B.18. Gráfico de autocorrelaciones del Modelo BMB efecto aleatorio agrupado.

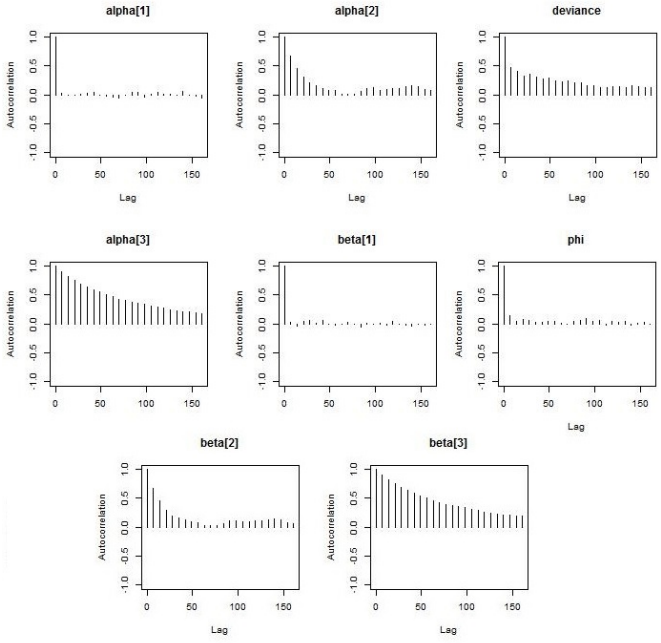


FIGURA B.19. Gráfico de autocorrelaciones del Modelo BMBDc con varianzas estocásticas en los efectos aleatorios.

B.7. Evaluación de convergencia en modelos Mixtos

B.7.1. Diagnósticos de convergencia de los modelos BMBDmm

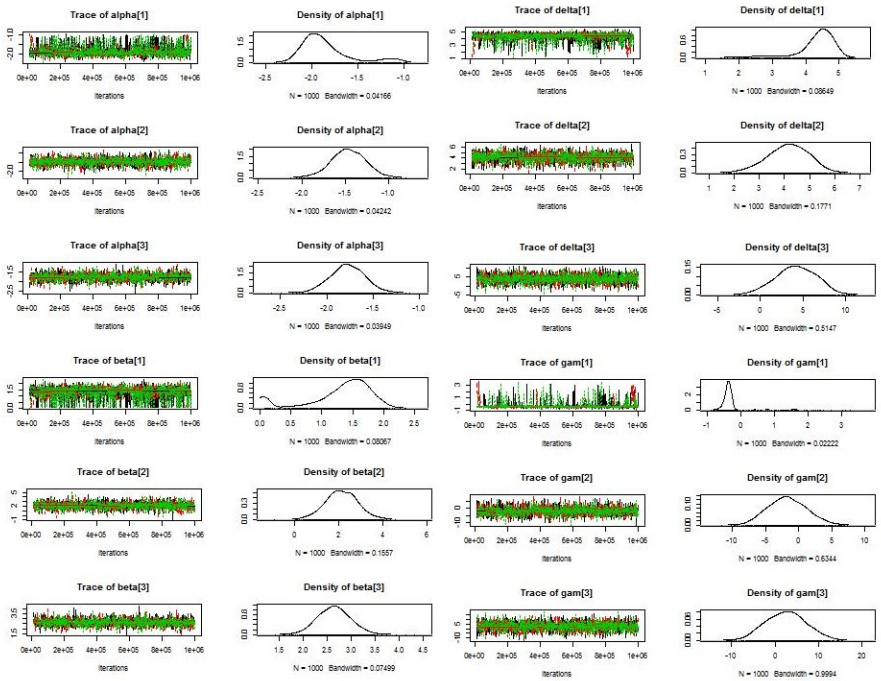


FIGURA B.20. Trayectoria de las cadenas del modelo BMBDmm.

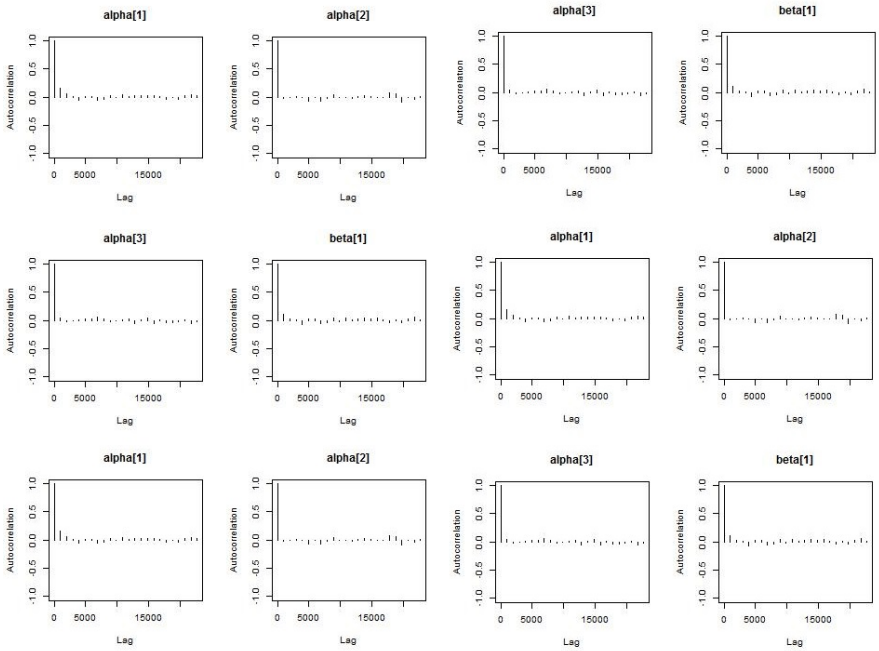


FIGURA B.21. Autocorrelaciones de las cadenas del modelo BMBDmm.

μ_d	Cluster	Geweke	Heidelberger y Welch	Estimador	Halfwidth
Intercepto	1	-1.4454	0.437	-1.857	0.0204
	2	0.7160	0.762	-1.461	0.0125
	3	-1.1135	0.517	-1.774	0.0127
Pendiente	1	1.2735	0.472	1.330	0.0364
	2	-0.8351	0.619	2.110	0.0465
	3	1.1750	0.479	2.639	0.0235
ϕ					
Intercepto	1	0.7025	0.934	4.324	0.0496
	2	0.8407	0.959	4.198	0.0502
	3	-0.3762	0.260	4.191	0.1492
Pendiente	1	-1.7261	0.658	-0.241	0.0458
	2	-0.2612	0.581	-1.937	0.1793
	3	0.5205	0.170	2.462	0.2894

TABLA B.9. Criterios de convergencia de las cadenas del Modelo BMBDmm3.

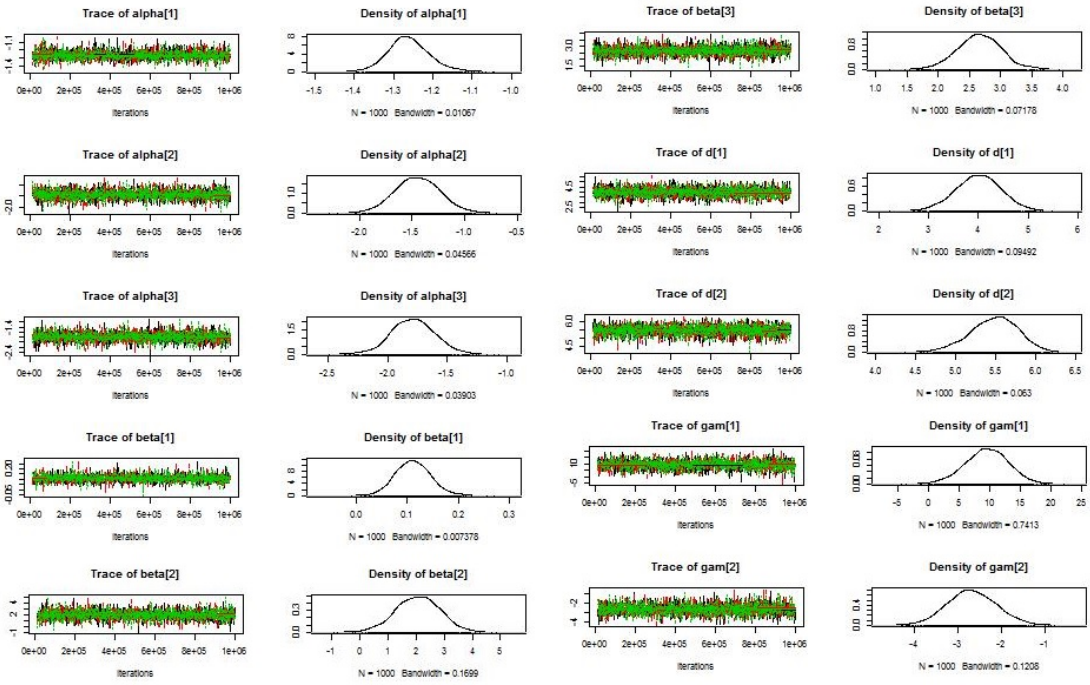


FIGURA B.22. Trayectoria de las cadenas del modelo BMBDmm3.

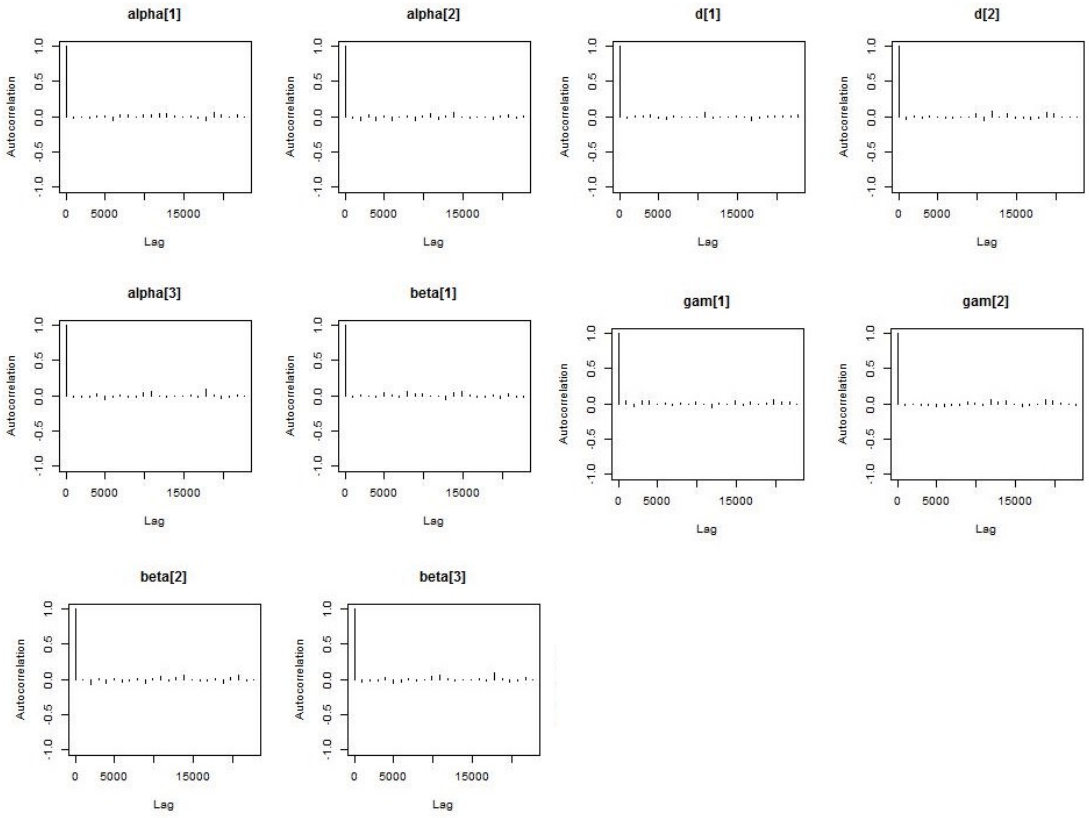


FIGURA B.23. Autocorrelaciones de las cadenas del modelo BMBDmm3.

μ_d	Cluster	Geweke	Heidelberger y Welch	Halfwidth
Interceptos	1	0.7740	0.896	0.00329
	2	0.3383	0.635	0.01161
	3	1.3073	0.538	0.01150
Pendientes	1	0.4009	0.188	0.00224
	2	-0.3865	0.539	0.04369
	3	-1.6470	0.429	0.02198
ϕ				
Intercepto	1	-1.6470	0.853	0.02745
	2-3	0.8136	0.519	0.01828
Pendiente	1	0.4278	0.346	0.22498
	2-3	-0.5363	0.393	0.03529

TABLA B.10. Criterios de convergencia de las cadenas del Modelo BMBDmm3.

Bibliografía

[1] Alkire, S. and Foster, J., *Counting and Multidimensional Poverty Measurement*, Tech. Report 7, OPHI Working Paper Series, (1995).

[2] Angulo, R., Díaz, Y., and Pardo, R., *Índice de Pobreza Multidimensional para Colombia (IPM-Colombia) 1997–2010*, Tech. Report 382, Archivos de economía, Departamento Nacional de Planeación, (2011).

[3] Balgobin, N. and Erhardt, E., *Fitting Bayesian Two-Stage Generalized Linear Models Using Random Samples via the SIR Algorithm*, Sankhy$\bar{a}$: The Indian Journal of Statistics **Vol; 66** (2004), no. 4, p. 733–755.

[4] Battese, G.E., Harter, R.M., and Fuller, W.A., *An error component model for prediction of county crop areas using survey and satellite data*, J. Amer. Statist. Ass **vol; 83** (1988), p. 28–36.

[5] Berg, E.J. and Fuller, W.A., *Small Area Prediction of Proportions Whit Aplications to the Canadian Labour Force Survey*, Journal of Survey Statistic and Methodology **vol; 2** (2014), p. 227–256.

[6] Bernal M., L.K., Quiroga, D., and Niño, A., *Índice de Pobreza Multidimensional para Bogotá 2003–2007*, Tech. Report 29, Bogotá Ciudad de Estadísticas, (2011).

[7] Boubeta, M., *Poisson mixed models: applications to small area data*, Abril (2017).

[8] Cepeda Cuervo, E., Aguilar, W., Cervantes, V., Corrales, M., Díaz, I., and Rodríguez, D., *Intervalos de confianza e intervalos de credibilidad para una proporción*, Revista Colombiana de Estadística **vol; 31** (2008), p. 211–228.

[9] Cepeda, E. , *Beta regression models: Joint mean and variance modelling*, http://www.bdigital.unal.edu.co/6207/1/varianceBetaRegresion.pdf, (2012).

[10] Cepeda, E., *Modelagem da variabilidad em modelos lineares generalizados*, Ph.D. thesis, Universidade Federal Do Rio De Janeiro, (2001).

[11] Cepeda, E. and Gamerman, D., *Bayesian methodology for modelling parameters in the two parameter exponential family*, Estadistica **vol; 57** (2005), p. 93–105.

[12] Cepeda, E. and Garrido, L., *Bayesian Beta Regression Models Joint Mean and Precision Modelling*, http://www.bdigital.unal.edu.co/5947/1/BayesianBetaRegresion.pdf, (2011).

[13] Cepeda, E., Migon, H., Garrido , L., and Achcar, J., *Generalized linear models with random effects in the two–parameter exponential family*, Journal of Statistical Computation and Simulation **vol; 84 (3)** (2014), p. 513–525.

[14] Chib, S. and Greenberg, E., *Understanding the Metropolis-Hastings Algorithm*, The American Statistician **vol; 49** (1995), no. 4, p. 327–335.

[15] Chochran, W.G., *Sampling techniques*, Wiley 3ra ed., New York, (1977).

[16] Dey, D.K., Gelfand, A.E., and Peng, F., *Overdispersed Generalized Linear Models*, Journal of Statistical Planning and Inference **vol: 64** (1997), p. 93–107.

[17] Domingez, R., *Estimación en áreas pequeñas: el ingreso medio mensual por comarca en los hogares gallegos*, Ph.D. thesis, Universidad de Santiago de Compostela, (2009).

[18] Erciulescu, A.L., *Small area prediction based on unit level models when the covariate mean is measured with error*, Ph.D. thesis, Iowa State University, (2015).

[19] Fadila, R., Rumiati, A., and Iriawan, N., *Pendugaan angka melek huruf di kabupaten bangkalan menggunakan small area estimation dengan pendekatan hierarchical bayes*, **vol:3** (2015), no. 2.

[20] Fay, R.E. and Herriot, R.A, *Estimates of income for small places: An application of james- stein procedures to census data*, Journal of the American Statistical Association **Vol: 74** (1979), p. 269–277.

[21] Ferrari, S. and Cribari-Neto, F., *Beta regression for modelling rates and proportions*, J. Appl Stat **vol: 31** (2004), no. 7, p. 799–815.

[22] Figueroa–Zúñiga, J., Abellano–Valle, R., and Ferrari, S., *Mixed Beta Regression: A Bayesian Perspective*, Computational Statistics and Data Analysis **vol: 61** (2013), p. 137–147.

[23] Garrido, L., *A Generalization of Bayesian Estimation in Finite Mixture of Distributions*, Ph.D. thesis, Universidad Nacional de Colombia, (2010).

[24] Garrido, L. and Cepeda–Cuervo, E., *Heteroscedastic Weibull–Normal Mixture Models: A Bayesian Approach*, Communications in Statistics–Simulation and Computation **vol: 43** (2014), p. 249–265.

[25] Garrido, L. and Cepeda–Cuervo, E., *Mixture of Distributions in the Biparametric Exponential Family: A Bayesian Approach*, Communications in Statistics–Simulation and Computation **vol: 41** (2012), p. 355–375.

[26] Ghosh, M. and Rao, J.N.K., *Small Area Estimation: An Appraisal*, Statistical Science **Vol: 9** (1994), no. 1, p. 55–93.

[27] Gil, S.N., *Bootstrap en poblaciones finitas*, Ph.D. thesis, Universidad de Granada, (2014).

[28] Goméz, A., *Modelos de mixturas finitas para la caracterización y mejora de la redes de monitorización de la calidad del aire*, (2014).

[29] Gonzales, M.E., *Use and evaluation of synthetic estimators*, Proceedings of the Social Statistics Section (1973), p. 33–36.

[30] Grüe, B., Kosmidis, I., and Zeileis, A., *Extended beta regression in r: Shaken, stirred, mixed, and partitioned*, Journal of Statistical Software **vol: 48** (2012).

[31] Gutierrez, A., *The Use of Working Variables in the Bayesian Modeling of Mean and Dispersion Parameters in Generalized Nonlinear Models with Random Eects*, http://www.tandfonline.com, (2014).

[32] Herrador, M., Morales, D., Esteban, M.C., Santamaria, L., Maruenda, Y., Peréz, A., and Molina, I., *Estimadores de áreas pequeñas basados en modelos para la encuesta de población activa*, Estadística Española **Vol: 51** (2009).

[33] Jiménez, J., *Métodos Monte Carlo basados en Cadenas de Markov*, (2015).

[34] Larsen, M.D., *Estimation of small area proportions using covariates and survey data*, Journal of Statistical Planning and Inference **vol: 112** (2003), p. 89–98.

[35] Lohr, L. and Prasad, N.G.N., *Small Area Estimation with Auxiliary Survey Data*, The Canadian Journal of Statistics **Vol: 31** (2003), no. 4, p. 383–369.

[36] Marker, D.A. , *Organization of small area estimators using a generalized linear regression framewor*, Journal of Official Statistic **15** (1999), 1–24.

[37] Nichol, S., *A regression approach to small area estimation*, Unpublished manuscript, Australian Bureaun of Statistics. Canberra Australia (1977).

[38] Nuñez, J., *Incidencia del gasto público social en la distribución del ingreso, la pobreza y la indigencia*, Tech. Report No.359, Archivos de economía, Departamento Nacional de Planeación, (2009).

[39] Pfeffermann, D., *Small Area Estimation: New Developments and Directions*, International Statistical Review **Vol: 70** (2002), no. 1, p. 125–143.

[40] Prasad, N.G.N. and Rao, J.N.K., *On robust estimation using a simple random effects model*, Survey Methodology **Vol: 25** (1999), p. 67–72.

[41] Pérez, A., *Estimación en áreas pequeñas bajo modelos lineales mixtos con dos factores aleatorios anidados*, Ph.D. thesis, Universidad Miguel Hernández de Elch, (2008).

[42] Pérez, R., Burgos, L., and Salinas, A., *Modelos de regresión para variables expresadas como una proporción continua*, Salud Pública Mex **Vol: 48** (2006), p. 395–404.

[43] Purcell, N.J. and Kish, L., *Postcensal estimates for local areas (or domains)*, Internat. Statist **vol: 48** (1980), p. 3–18.

[44] Rao, J. N. K., *Statistical methodology for indirect estimations is small area*, EUSTAT **Vol: 39** (2000).

[45] Rao, J.N. K., *Some recent advances in model-based small area estimation*, Survey Methodology, Statistics Canada **Vol: 22** (1999), no. 2, Catalogue 12–001, p. 175–186.

[46] Rao, J.N.K., *Small area estimation*, John Wiley and Son, (2003).

[47] Satriya A.M., Iriawan, N., and Brodjol, S.U., *Small area estimation pengeluaran per kapita di kabupaten bangkalan dengan metode hierarchical bayes*, Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam **vol: 3** (2015), no. 2.

[48] Secretaría Distrital de Planeación, SDP., *Principales resultados de la primera encuesta multipropósito de Bogota*, Tech. Report No.32, Bogotá Ciudad de Estadísticas, (2011).

[49] Simas, A.B., Barreto–Souza, W., and Rocha, A.V., *Improved Estimator for a General Class of Beta Regression Models*, Computational Statistics and Data Analysis **Vol: 54 (2)** (2010), p. 348–366.

[50] Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Linde, A., *Bayesian measures of model complexity and fit*, Journal of the Royal Statistical Society **series B 64** (2002), p. 583–639.

[51] Sugasawa S., *Robust Empirical Bayes Small Area Estimation with Density Power Divergence*, http://arxiv.org/abs/1702.06635v1, (2017).

[52] Sugasawa, S., Tamae, H., and Kubokawa, T., *Bayesian Estimators for Small Area Models Shrinking Both Means and Variances*, http://sindominio.net/ash, (2015).

[53] Tejedor, F.H. , *Modelamiento conjunto de media y varianza en modelos mixtos con respuesta Beta: perspectiva Bayesiana*, (2014).

[54] Torkashvand, E., Jarafi Jozani, M., and Torabi, M., *Clustering in small area estimation with area level linear mixed models*, http://arxiv.org/abs/1507.05179v2, (2017).

[55] Velez, Carlos E., Azevedo, J., and Posso, C., *Oportunidades para los niños colombianos: cuánto avanzamos en esta década*, Tech. Report No.673, Borradores de Economía, Banco Mundial – Banco de la República, (2011).

[56] You, Y. and Chapman, B., *Small area estimation using area level models and estimated sampling variances*, Survey Methodology **vol; 32** (2006), p. 97–103.