

MODELOS DE REGRESIÓN BETA CON EFECTOS
MIXTOS:
ENFOQUE BAYESIANO

Por
Jorge Figueroa Zúñiga

SOMETIDA COMO REQUISITO PARA
OPTAR AL GRADO DE
DOCTOR EN ESTADÍSTICA
EN LA
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SANTIAGO, CHILE
NOVIEMBRE 2011

© Copyright by Jorge Figueroa Zúñiga, 2011

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTAMENTO DE
ESTADÍSTICA

Los firmantes certifican que han leído y recomendado a la Facultad de Ciencias Físicas y Matemáticas la aceptación de la Tesis titulada “**Modelos de Regresión Beta con Efectos Mixtos: Enfoque Bayesiano**” escrita por **Jorge Figueroa Zúñiga** como requerimiento para optar al grado de **Doctor en Estadística**.

Fecha: Noviembre 2011

Examinador Externo: _____
AAAAAAAAAAAAA

Orientador: _____
Reinaldo Arellano Valle

Comité Examinador: _____
DDDDDDDDDD

BBBBBBBBBBBB

CCCCCCCCCCCC

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

Fecha: **Noviembre 2011**

Autor: **Jorge Figueroa Zúñiga**

Título: **Modelos de Regresión Beta con Efectos Mixtos:
Enfoque Bayesiano**

Departamento: **Estadística**

Grado: **Ph.D.** Convocatoria: **Noviembre** Año: **2011**

Se le concede permiso a la Pontificia Universidad Católica de Chile para hacer circular y copiar, con propósitos no comerciales, este documento según lo requieran personas y/o instituciones.

Firma del Autor

EL AUTOR SE RESERVA LOS DERECHOS DE OTRAS PUBLICACIONES. LA TESIS NI EXTRACTOS EXTENSOS DE ELLA PUEDEN SER IMPRESOS O REPRODUCIDOS SIN EL PERMISO ESCRITO DEL AUTOR.

EL AUTOR CERTIFICA QUE EL PERMISO SE HA OBTENIDO PARA EL USO DE CUALQUIER MATERIAL REGISTRADO QUE APAREZCA EN ESTA TESIS (CON EXCEPCIÓN DE LOS EXTRACTOS QUE REQUIEREN SOLAMENTE EL RECONOCIMIENTO APROPIADO DE LOS ESTUDIANTES) Y QUE TODO USO ESTÉ RECONOCIDO CLARAMENTE.

A mis padres:

Isaac Figueroa Muñoz

Norma Zúñiga Hurtado

Y A mi hermano:

Edgardo Figueroa Zúñiga

Tabla de Contenidos

Tabla de Contenidos	v
Índice de Tablas	vii
Índice de Figuras	viii
Resumen	ix
Agradecimientos	xi
1. Introducción	1
1.1. Distribución beta y sus propiedades	1
1.2. Modelos de Regresión beta	7
2. Revisión bibliográfica	10
2.1. Modelos Jerárquicos	10
2.2. Criterios de Comparación	11
2.2.1. Verosimilitud predictiva	12
2.2.2. Criterios de información basados en el desvío	13
2.3. Diagnóstico de Convergencia	14
2.3.1. Inspección visual	14
2.3.2. Criterio de Geweke	15
2.3.3. Criterio de Gelman y Rubin	16
2.3.4. Criterio de Heidelberger y Welch	17
3. Regresión beta con efectos mixtos: Enfoque Bayesiano	18
3.1. Introducción	19
3.2. Propuesta del modelo y justificación de prioris	22
3.3. Modelo de regresión beta mixto con parámetro de precisión molesto	25

3.3.1.	Discusión de prioris	26
3.3.2.	Ajuste del modelo usando algoritmo de Monte Carlo vía cadenas de Markov	27
3.3.3.	Ejemplo con datos simulados	30
3.3.4.	Aplicación del modelo a datos reales	36
3.3.5.	Apéndice	41
3.3.6.	Códigos BUGS	46
3.4.	Modelo de regresión beta con submodelo de regresión asociado al parámetro de precisión	51
3.4.1.	Discusión de prioris	51
3.4.2.	Ajuste del modelo usando algoritmo de Monte Carlo vía cadenas de Markov	52
3.4.3.	Ejemplo con datos simulados	54
3.4.4.	Aplicación del modelo a datos reales	58
3.4.5.	Apéndice	62
3.4.6.	Códigos BUGS	65
3.5.	Resumen y conclusiones	67
3.6.	Discusión	67
4.	Conclusiones Generales y Trabajos Futuros	70
	Bibliografía	71

Índice de cuadros

3.1. <i>Análisis de sensibilidad para la especificación de la priori del parámetro de precisión</i>	32
3.2. <i>Medias a posteriori exactas. Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto simulado</i>	33
3.3. <i>Test de diagnóstico de Gelman's, Geweke's y Heidelberg & Welch's</i>	34
3.4. <i>Análisis de sensibilidad para la especificación de la priori del parámetro de precisión</i>	38
3.5. <i>Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto dado en 1.c.4</i>	39
3.6. <i>Propuesta de submodelos de precisión con un modelo de localización común</i>	55
3.7. <i>Análisis de sensibilidad para la especificación del submodelo de precisión</i>	56
3.8. <i>Medias a posteriori exactas. Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto simulado</i>	56
3.9. <i>Propuesta de submodelos de precisión bajo diferentes modelos de localización</i>	59
3.10. <i>Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto dado en 2.c</i>	60

Índice de figuras

1.1.	<i>Distribuciones Beta considerando diferentes valores para los parámetros (α, β)</i>	3
1.2.	<i>Distribuciones Beta de 4 parámetros considerando diferentes valores para los parámetros (α, β, a, b)</i>	4
1.3.	<i>Distribuciones $Pert(0, 0, 4, 1, \alpha)$ considerando diferentes valores para el parámetro de apuntamiento α</i>	6
1.4.	<i>Distribuciones $beta(\mu\phi, (1 - \mu)\phi)$ considerando diferentes valores para los parámetros μ y ϕ</i>	7
3.1.	<i>Autocorrelación - Datos simulados y parámetro ϕ común</i>	35
3.2.	<i>Historia - Datos simulados y parámetro ϕ común</i>	35
3.3.	<i>Autocorrelación - Datos de Prater y parámetro ϕ común</i>	40
3.4.	<i>Historia - Datos de Prater y parámetro ϕ común</i>	40
3.5.	<i>Densidad - Datos de Prater y parámetro ϕ común</i>	41
3.6.	<i>Autocorrelación - Datos simulados y estructura de regresión para ϕ</i>	57
3.7.	<i>Historia - Datos simulados y estructura de regresión para ϕ</i>	57
3.8.	<i>Autocorrelación - Datos de Prater y estructura de regresión para ϕ</i>	61
3.9.	<i>Historia - Datos de Prater y estructura de regresión para ϕ</i>	61
3.10.	<i>Densidad - Datos de Prater y estructura de regresión para ϕ</i>	62

Resumen

El presente trabajo propone un modelo donde la variable de respuesta es distribuida beta y se tienen variables explicativas mixtas, es decir, se consideran variables explicativas tanto fijas como aleatorias en el modelo de regresión. Este tipo de modelos con efectos mixtos son muy utilizados en la actualidad, principalmente en el área de la salud y su estudio ha sido restringido principalmente a repuestas con distribución normal o con distribuciones simétricas, asumiendo además que la variable mencionada tiene soporte sobre todos los reales. Sin embargo, lo anterior parece no ser aplicable cuando la respuesta tiene soporte limitado a un intervalo (a, b) con $-\infty < a < b < \infty$ (a, b escalares conocidos), por ejemplo, en modelamientos para tasas y proporciones en donde la variable de respuesta está acotada a un intervalo $(0, 1)$. Además es lógico pensar también que este tipo de modelos no ajuste bien los datos cuando estos presentan asimetrías

La ventaja sustancial en considerar un modelamiento beta se debe a la gran flexibilidad que ella presenta, lo que permite un buen ajuste a los datos disponibles en un gran número de casos, sean estos simétricos o no. Se podrá prever que un posible inconveniente de esta distribución es que toma solo repuestas en el intervalo $(0, 1)$, pero en la práctica podremos llevar cualquier respuesta con soporte limitado en \mathbb{R} al intervalo $(0, 1)$ por medio de una simple transformación [17].

Con el fin de establecer una estructura de regresión para la media de la variable de respuesta, será necesario considerar un modelo beta reparametrizado en función de la media y un parámetro de dispersión (Paolino Philip, 2001) [18]. El modelo propuesto es usado en situaciones donde la variable de interés es continua y restringida

al intervalo $(0,1)$ y se relaciona con otras variables por medio de una estructura de regresión. Los parámetros del modelo de regresión beta son interpretables en términos de la media de la variable de respuesta, donde esta será convenientemente ligada con la estructura de regresión mediante el link logit (Ferrari and Cribari-Neto, 2004) [7].

Para abordar nuestro modelo propuesto, se utilizará el enfoque Bayesiano, donde se discutirán las diversas alternativas propuestas en la literatura sobre la especificación de prioris convenientes, se propondrá una priori alternativa a lo expuesto en la literatura y se realizarán las estimaciones de los parámetros de interés junto con las respectivas pruebas de diagnóstico, comparando la calidad del ajuste que proporciona cada modelo planteado, para tal efecto, se efectuarán diversas simulaciones. Finalmente se compararán los enfoques bayesiano y clásico del modelo mixto propuesto usando prioris clásicas en la literatura, con lo que serán debatidas y analizadas las ventajas y desventajas que presenta cada uno de estos enfoques.

Agradecimientos

Capítulo 1

Introducción

En este capítulo revisaremos la distribución beta y las diferentes reparametrizaciones que se han realizado de la misma con el fin de resolver diversas problemáticas en las distintas áreas de estudio. Mostraremos además, cómo a partir de una reparametrización en particular (Paolino, P. (2001))[18] se han desarrollado diversos estudios relacionados con modelos de regresión y respuesta distribuida beta.

1.1. Distribución beta y sus propiedades

La distribución beta es muy flexible, en cuanto su densidad puede adoptar una gran variedad de formas y presenta propiedades deseables que han sido estudiadas en detalle (Gupta, A. K. and Nadarajah, S. (2004))[14].

La función de densidad de probabilidad de una variable aleatoria Y que sigue una distribución beta está dada por:

$$f(y|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad 0 < y < 1 \quad (1.1.1)$$

Donde:

$$E(Y) = \frac{\alpha}{\alpha + \beta} = \mu \quad ; \quad Var(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (1.1.2)$$

Con $\alpha, \beta > 0$ y $\Gamma(\cdot)$ denota la función gamma.

Diferentes elecciones para α y β conllevan a diferentes formas en la función de densidad, como se detalla a continuación:

- $\alpha = 1, \beta = 1$ es una distribución uniforme (0,1)
- $\alpha < 1, \beta < 1$ es una densidad con forma de U
- $\alpha < 1, \beta \geq 1$ o $\alpha = 1, \beta > 1$ es estrictamente decreciente (forma de "L")
- $\alpha = 1, \beta < 1$ o $\alpha > 1, \beta \leq 1$ es estrictamente creciente (forma de "J")
- $\alpha = 1, \beta > 2$ o $\alpha > 2, \beta = 1$ es estrictamente convexo
- $\alpha = 1, 1 < \beta < 2$ o $1 < \alpha < 2, \beta = 1$ es estrictamente concavo
- $\alpha = 1, \beta = 2$ o $\alpha = 2, \beta = 1$ es una línea recta
- $\alpha > 1, \beta > 1$ es una densidad unimodal (simétrica, asimétrica a la izquierda o asimétrica a la derecha).
- Si $\alpha = \beta$ entonces la función de densidad es simétrica en torno a $1/2$.

En la figura 1.1 se muestra un ejemplo gráfico del modelo beta, utilizando distintos valores en sus parámetros α y β .

Existen otras variantes del modelo beta original visto en (1.1.1) y que han sido definidas a fin de satisfacer diferentes objetivos como veremos a continuación.

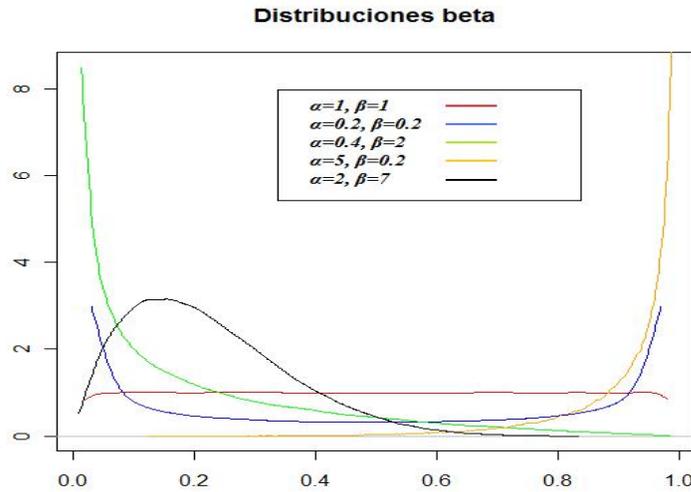


Figura 1.1: *Distribuciones Beta considerando diferentes valores para los parámetros (α, β) .*

Distribución Beta de 4 parámetros: La distribución beta de 4 parámetros es muy flexible en cuanto a forma y soporte, así que ha sido muy popular para adaptar conjuntos de datos donde se tiene una variable limitada a un intervalo (a, b) con $(a < b; a, b \in \mathbb{R})$. En tal caso, diremos que la variable aleatoria Y tiene distribución beta de 4 parámetros si su función de densidad está dada por:

$$f(y|\alpha, \beta, a, b) = \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{B(\alpha, \beta)(b-a)^{\alpha+\beta-1}} ; \alpha > 0, \beta > 0; a < y < b \quad (1.1.3)$$

Donde $B(\alpha, \beta)$ es la función beta. Denotaremos lo anterior como $Y|\alpha, \beta, a, b \sim \text{beta}(\alpha, \beta, a, b)$.

Además se tienen las siguientes propiedades:

$$E(Y) = a + \frac{\alpha}{\alpha + \beta}(b - a) \quad ; \quad V(Y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}(b - a)^2$$

$$\text{moda}(Y) = m = a + \frac{\alpha - 1}{\alpha + \beta - 2}(b - a) \quad \text{si } \alpha > 1, \beta > 1$$

(Nota: $m = a, b$ si $\alpha < 1, \beta < 1$; $m = a$ si $\alpha < 1, \beta \geq 1$ o si $\alpha = 1, \beta > 1$;
 $m = b$ si $\alpha \geq 1, \beta < 1$ o si $\alpha > 1, \beta = 1$; m adopta infinitas soluciones si $\alpha = 1, \beta = 1$).

En la figura 1.2 se muestra un ejemplo gráfico del modelo beta de 4 parámetros.

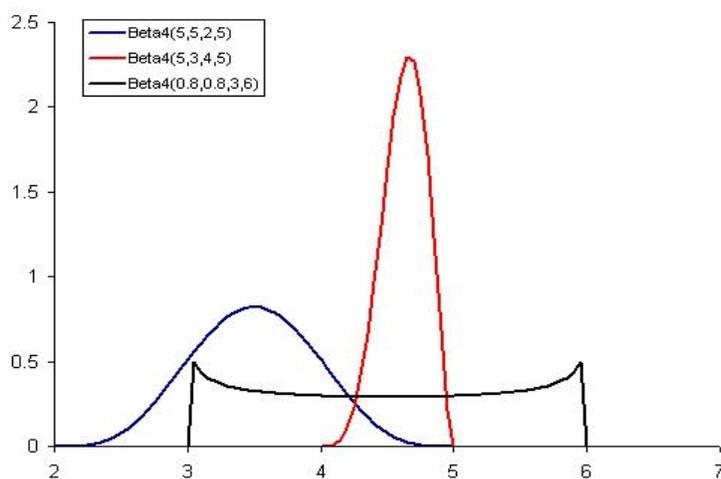


Figura 1.2: *Distribuciones Beta de 4 parámetros considerando diferentes valores para los parámetros (α, β, a, b) .*

Cabe mencionar que esta reparametrización es anloga a tomar $y^*(b - a) + a = y$ donde y^* es distribuida beta con función de densidad definida en (1.1.1).

Distribución PERT: La distribución PERT o también conocida como beta PERT se trata de una versión de la distribución beta y requiere los mismos tres parámetros que la distribución triangular, es decir, mínimo (a), máximo (b) y moda (m). Si Y tiene distribución PERT, su función de densidad de probabilidad quedará determinada por

la función:

$$f(y|v, w, a, b) = \frac{(y-a)^{v-1}(b-y)^{w-1}}{B(v, w)(b-a)^{v+w-1}} ; a < y < b \quad (1.1.4)$$

Donde $B(v, w)$ es la función beta y $v = 6\left[\frac{\mu-a}{b-a}\right]$, $w = 6\left[\frac{b-\mu}{b-a}\right]$ con $\mu = \frac{a+4m+b}{6}$ (Moda(Y)= m , E(Y)= μ). Denotaremos lo anterior como $Y|a, m, b \sim Pert(a, m, b)$.

Una versión mas generalizada del modelo PERT está dada al considerar el modelo dado en (1.1.4), pero tomando $\mu = \frac{a+\lambda m+b}{\lambda+2}$. Esto permite incorporar al modelo un parámetro de apuntamiento al tomar la siguiente reparametrización:

$$\lambda = \frac{2\alpha}{1-\alpha} \Leftrightarrow \alpha = \frac{\lambda}{\lambda+2} \text{ con } \mu = \alpha m + (1-\alpha)c, \text{ donde } 0 < \alpha < 1 \text{ con } c = \frac{a+b}{2}.$$

Tendremos entonces una distribución con 4 parámetros de la forma $Y|a, m, b, \lambda \sim Pert(a, m, b, \lambda)$ o bien $Y|a, m, b, \alpha \sim Pert(a, m, b, \alpha)$ donde $\alpha = \frac{\lambda}{\lambda+2}$ es considerado como parámetro de apuntamiento.

De esta forma, las estimaciones de los parámetros del modelo se hacen mas intuitivas, tal como se representa en la figura 1.3.

Distribución beta-regression: En el contexto de análisis de regresión con variable de respuesta distribuida beta, es común reparametrizar la distribución beta original dada en (1.1.1). Esta reparametrización se realiza en términos de su media, lo que permite establecer un enlace de esta variable (variable de respuesta) con la estructura de regresión que contiene las variables explicativas del modelo. De esta forma, la función de densidad asociada queda determinada por:

$$f(y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, 0 < y < 1 \quad (1.1.5)$$

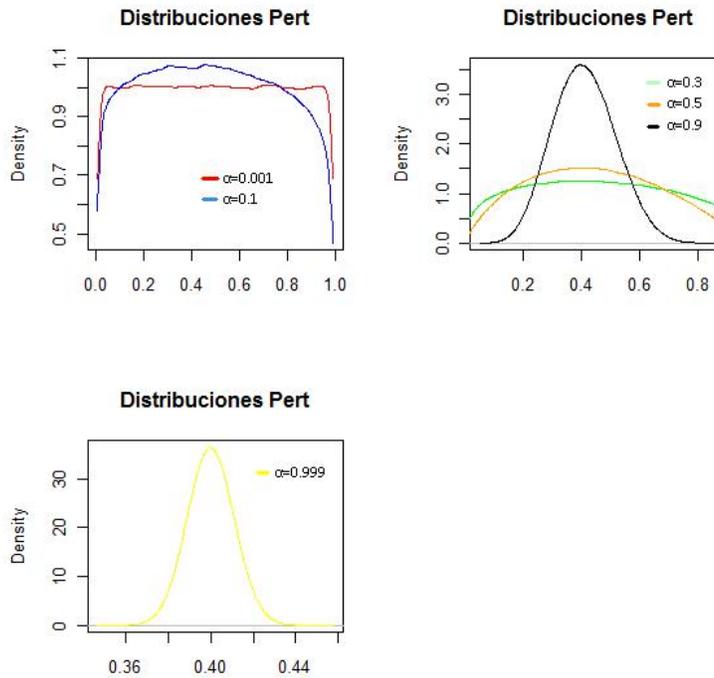


Figura 1.3: *Distribuciones Pert(0,0,4,1, α) considerando diferentes valores para el parámetro de apuntamiento α .*

Donde:

$$E(y) = \frac{\alpha}{\alpha + \beta} = \mu$$

y

$$Var(y) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{\mu(1 - \mu)}{1 + \phi} = \frac{V(\mu)}{1 + \phi}$$

con $\phi = \alpha + \beta$.

De esta forma $0 < \mu < 1$ y $\phi > 0$, luego la forma de la densidad dada en (1.1.5) cambia dependiendo del valor de estos dos parámetros. Particularmente se

tendrá simetría cuando $\mu = 1/2$ y asimetría cuando $\mu \neq 1/2$. Notamos además que para un μ fijo, el parámetro ϕ controla la dispersión del modelo, de forma que a medida que aumenta ϕ la varianza en el modelo disminuye, luego diremos que ϕ es un parámetro de precisión.

Denotaremos la distribución de esta variable aleatoria por $Y|\mu, \phi \sim \text{beta}(\mu\phi, (1-\mu)\phi)$.

En la figura 1.4 se representa gráficamente el modelo expuesto anteriormente.

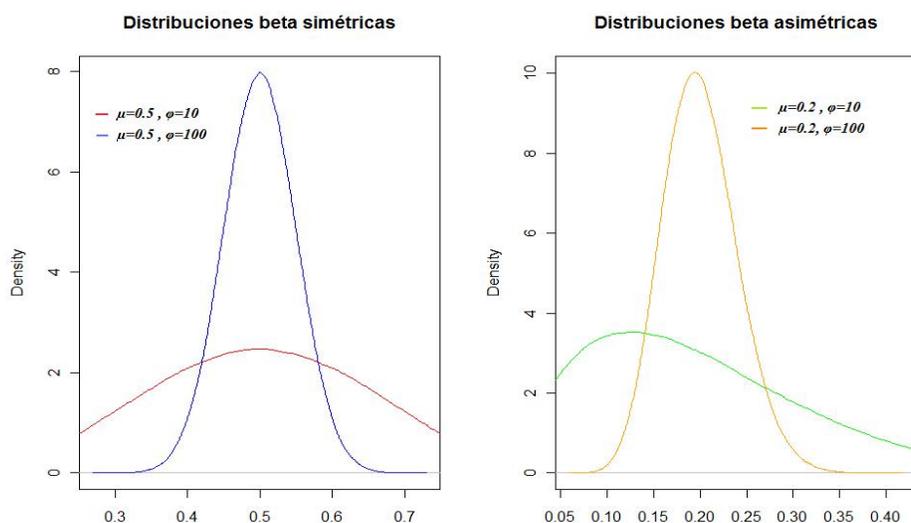


Figura 1.4: *Distribuciones $\text{beta}(\mu\phi, (1-\mu)\phi)$ considerando diferentes valores para los parámetros μ y ϕ .*

1.2. Modelos de Regresión beta

La familia beta es una elección natural para modelar datos continuos restringidos a un intervalo $(0,1)$, esto debido a la flexibilidad que presenta en cuanto puede adoptar una gran variedad de formas en su densidad. Los modelos de regresión con variable de respuesta beta han sido recientemente estudiados, comenzando con la

reparametrización efectuada por Paolino (2001) [18] y principalmente con el estudio hecho por Ferrari y Cribari-Neto (2004) [7] quienes estudiaron el modelo de regresión en el contexto clásico utilizando la reparametrización dada en (1.1.5) donde fueron discutidas sus principales características y hechas las pruebas de diagnóstico correspondientes. El modelo de regresión propuesto en este estudio es:

$$G(\mu) = X\beta \tag{1.2.1}$$

Donde $X_{n \times k}$ es la matriz de variables explicativas, $\beta \in \mathbb{R}^k$ y $G(\cdot)$ es una función que liga la respuesta media μ con la estructura de regresión del modelo y donde fue considerada, para tal efecto, la función link logit.

Luego un estudio hecho por Smithson y Verkuilen (2006) [23] replicó el estudio hecho por Ferrari y Cribari-Neto [7], pero asignando un submodelo de regresión para el parámetro de dispersión, de la forma:

$$\phi_i = \exp(-w_i^t \delta) \tag{1.2.2}$$

Donde $w_{i(k^* \times 1)}$ y $\delta_{(k^* \times 1)}$ ($k^* \leq k$).

Lo anterior permite que, en muchos casos, el ajuste realizado en modelos lineales generalizados sea de mejor calidad que cuando no se considera este submodelo para el parámetro de dispersión, esto queda debidamente justificado por Smyth (1988) [24].

Un estudio posterior, hecho por Branscum, Johnson y Thurmond (2007) [3], retomó el modelo propuesto por Ferrari y Cribari-Neto [7] pero desde la perspectiva Bayesiana [6], asignando distribuciones normales a los coeficientes de regresión y distribución gamma para el parámetro de dispersión ϕ . En este estudio se estimaron

los parámetros de interés y se propuso seleccionar la función link que liga la variable dependiente con la estructura de regresión usando el método de Factor de Bayes. Además se propuso un modelo de regresión beta semiparamétrico, de la forma:

$$\text{logit}(\mu) = \beta_0 + \beta_1 x + \dots + \beta_p x^p + \sum_{j=1}^J b_j ((x - k_j)_+)^p \quad (1.2.3)$$

donde los k_j son nudos fijos y $(x)_+ = xI_{(0,\infty)}(x)$, luego se usaron funciones spline penalizadas para implementar la regresión beta Bayesiana semiparamétrica.

Finalmente Zimprich (2010) [27] realiza un estudio para el modelo de regresión beta con efectos mixtos de la forma:

$$\ln\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = x_{ij}^t \beta + z_{ij}^t b_i \quad (1.2.4)$$

Con $x_{ij(p \times 1)}$, $\beta_{(p \times 1)}$ y $z_{ij(q \times 1)}$ vector de variables predictoras para los efectos aleatorios b_i donde $b_i \sim N_q(0, D)$ donde D es la matriz de varianzas-covarianzas de los efectos aleatorios. En este trabajo se consideró un modelo análogo al de Smithson y Verkuilen (2006) [23], es decir, se consideró además un submodelo de regresión para el parámetro de dispersión. Luego, se desarrollaron las estimaciones de los parámetros de interés, considerando y comparando diferentes modelos de localización y submodelos de dispersión.

Capítulo 2

Revisión bibliográfica

En este capítulo realizaremos una revisión bibliográfica sobre temas que serán abordados y aplicados en los capítulos siguientes. En la sección inicial serán descritos los modelos jerárquicos y sus aplicaciones. En la sección 2.2 son descritos algunos criterios de comparación de modelos que serán citados y utilizados a lo largo del presente trabajo, en las secciones de aplicación con datos simulados. En la última sección se presentan los criterios mas usados para evaluar la convergencia en los algoritmos MCMC utilizados para simular muestras de distribuciones a posteriori cuando su forma es desconocida.

2.1. Modelos Jerárquicos

Un modelo jerárquico toma en cuenta una estructura jerárquica en la población a ser investigada, donde las unidades mas bajas en la población deben estar organizadas por jerarquía en niveles superiores. Por ejemplo, podemos considerar escolares agrupados en clases, que a su vez están organizadas por escuelas y éstas por municipios. Podemos entonces describir los resultados de un alumno como una suma de efectos de alumno, de la clase, de la escuela y del municipio al cual pertenece. Éstos efectos

pueden ser considerados permutables y con una distribución descrita por un componente de varianza. También se pueden tener coeficientes de regresión en algunos o en todos los niveles.

Un modelo jerárquico normal que ha sido muy utilizado tiene la forma

$$y_{ij} \sim N(\mu + \nu_j, \sigma^2), i = 1, \dots, m, j = 1, \dots, n_i$$

$$\nu_j \sim N(0, \sigma_\nu^2).$$

El modelo anterior, podría asumir distribuciones para y_{ij} diferentes a la normal, como por ejemplo, distribuciones pertenecientes a la familia exponencial y en particular, como es el caso en el presente trabajo, una distribución beta cuando la respuesta hace referencia a una tasa o una proporción. Un punto importante en la estimación de los parámetros en los modelos jerárquicos hace referencia a las varianzas que muchas veces son asumidas conocidas por la dificultad que presentan al ser estimadas. Gelman (2006) [11] propone distribuciones a priori para los parámetros de varianza en los modelos jerárquicos.

Los modelos jerárquicos son aplicados frecuentemente en la modelación de datos de muestras con diseño complejo, que comúnmente involucran estratificación y conglomeración en varias etapas para investigar poblaciones con estructuras jerárquicas. Tal estructura es reflejada en el modelo a través de efectos aleatorios y componentes de varianza que permiten diferenciar los niveles jerárquicos.

2.2. Criterios de Comparación

En el proceso de modelación, frecuentemente ocurren situaciones en que es necesario decidir sobre la inclusión o exclusión de parámetros o bien sobre la forma de la

distribución que adoptan las observaciones. En tales situaciones, surge la necesidad de efectuar una comparación entre los modelos alternativos a través de la utilización de criterios que permitan evaluar la calidad del ajuste proporcionado por el modelo. En la presente sección, describiremos los criterios de selección de modelos utilizados en este trabajo.

2.2.1. Verosimilitud predictiva

Este criterio compara las capacidades predictivas de diferentes modelos a través de la verosimilitud predictiva dada por

$$p(y_r|y, M) = \int p(y_r, \Psi|y, M)d\Psi \quad (2.2.1)$$

Donde Ψ es el conjunto de todos los parámetros del modelo M , $y = (y_1, \dots, y_n^t)$ es el vector de observaciones e $y_r = (y_1, \dots, y_h^t)$ es un vector de datos no observados o que no fueron utilizados en la obtención de la distribución a posteriori de Ψ . Luego, el modelo seleccionado será aquel que presente un mayor valor en su verosimilitud predictiva. Del desarrollo de (2.2.1) se obtiene

$$\begin{aligned} p(y_r|y, M) &= \int p(y_r, \Psi|y, M)d\Psi \\ &= \int p(y_r|\Psi, y, M)p(\Psi|y, M)d\Psi \\ &= E_{\Psi|y, M}[p(y_r|\Psi, y, M)] \end{aligned} \quad (2.2.2)$$

Ahora, si tenemos una aproximación de la distribución a posteriori de Ψ formada por $(\Psi^{(1)}, \dots, \Psi^{(L)})$, entonces una aproximación para (2.2.2) estaría dada por

$$\hat{E}_{\Psi|y, M}[p(y_r|\Psi, y, M)] = \frac{1}{L} \sum_{l=1}^L \prod_{i=1}^h p(y_{ir}|\Psi^{(l)}, y, M)$$

Cabe señalar que en el caso multivariado, la aproximación tiene la misma forma que y_{ir} sustituida por el vector $y_{ir} = (y_{i1}, \dots, y_{ik})^t$ donde k es el número de variables observadas y con dependencia entre sí.

2.2.2. Criterios de información basados en el desvío

Sea $L(y|\Psi_k, M_k)$ una función de verosimilitud para el modelo M_k , donde Ψ_k reúne todos los parámetros del modelo. Se define la función de desvío como $D(\Psi_k) = -2\log L(y|\Psi_k, M_k)$. Los criterios DIC (Deviance Information Criterion), definido en Spiegelhalter et al. (2002) [25] y los criterios AIC (Akaike Information Criterion) y BIC (Bayesian Information Criterion) son dados por

$$AIC(M_k) = D(E[\Psi_k|y, M_k]) + 2d_k;$$

$$BIC(M_k) = D(E[\Psi_k|y, M_k]) + \log(n)d_k;$$

$$DIC(M_k) = 2E[D(\Psi_k)|y, M_k] - D(E[\Psi_k|y, M_k])$$

Donde d_k representa el número de parámetros del modelo M_k . Luego, cuanto menor es el valor de los criterios, mejor es el ajuste del modelo. El criterio DIC también puede ser escrito como

$$DIC(M_k) = E[D(\Psi_k)|y, M_k] + p_D(M_k)$$

Donde $p_D(M_k) = E[D(\Psi_k)|y, M_k] - D[E(\Psi_k|y, M_k)]$ representa el número de parámetros efectivos en el modelo.

Suponga que $\{\Psi_k^{(1)}, \dots, \Psi_k^{(L)}\}$ corresponde a una muestra de distribución a posteriori, entonces las siguientes aproximaciones de Monte Carlo pueden ser aplicadas (Silva y Lopes (2008)) [22]

$$E[D(\Psi_k)|y, M_k] \approx L^{-1} \sum_{l=1}^L D(\Psi_k^{(l)}) \quad y \quad E[\Psi_k|y, M_k] \approx L^{-1} \sum_{l=1}^L \Psi_k^{(l)} \quad (2.2.3)$$

Cabe mencionar que los criterios definidos en esta sección pueden presentar valores negativos. No obstante el valor de p_D debe ser siempre positivo. En caso contrario, pueden existir indicios de una mala especificación del modelo. Las ecuaciones de los tres criterios valen directamente para el caso multivariado, pues dependen de funciones de verosimilitud que incorporan una estructura de dependencia entre las variables observadas.

2.3. Diagnóstico de Convergencia

En este trabajo se utiliza inferencia Bayesiana para la obtención de las estimaciones de los parámetros de los modelos propuestos. En la mayoría de los ejemplos, las densidades a posteriori no poseen forma conocida y se recurre a métodos de simulación indirecta para muestrear los parámetros de los modelos. Los métodos de simulación aplicados en este trabajo fueron el muestreo de Gibbs y el algoritmo de Metropolis-Hastings, que se basan en algoritmos de Monte Carlo vía Cadenas de Markov (MCMC). Sus descripciones pueden ser vistas en Gamerman (2006) [9]. Cuando existe la convergencia en la cadena, se alcanza entonces la distribución objetivo (posteriori), después de una fase de calentamiento. Como describiremos a continuación, existen métodos para monitorear o diagnosticar la existencia de convergencia.

2.3.1. Inspección visual

Una inspección visual de los gráficos de las cadenas es una forma simple de monitorear la convergencia. La observación de las trayectorias de diferentes cadenas partiendo de valores iniciales distintos permite que se verifique si existe una mezcla de las cadenas a medida que aumenta el número de iteraciones, indicando convergencia

en distribución. A partir de los gráficos de autocorrelación, se verifica si las muestras generadas pueden ser consideradas independientes de los valores iniciales. Los histogramas posibilitan el análisis de la forma de la densidad a posteriori, identificando bimodalidad, por ejemplo.

2.3.2. Criterio de Geweke

Geweke (1992) [13] sugiere un criterio para evaluar la convergencia en base a la comparación de medias, en intervalos diferentes, después de una fase de calentamiento del algoritmo. Si la convergencia fué alcanzada, los comportamientos en esos intervalos debe ser semejante.

Considere $\psi = t(\theta)$ una función real de un parámetro de interés θ , para el cual fué obtenida una muestra a partir del algoritmo MCMC. Sea m el número de iteraciones correspondiente a la fase de calentamiento y $m + n$ el total de iteraciones. Se construyen las medias $\bar{\psi}_a$ y $\bar{\psi}_b$ en base a los grupos de iteraciones $n_a < n$ y $n_b < n$. Luego,

$$z_G = \frac{\bar{\psi}_a - \bar{\psi}_b}{\sqrt{Var(\bar{\psi}_a) + Var(\bar{\psi}_b)}} \rightarrow N(0, 1) \quad (2.3.1)$$

en distribución, bajo la hipótesis de convergencia.

De esta forma, valores grandes de z_G indican falta de convergencia. Sin embargo, valores pequeños de esta estadística no significan que existe convergencia. Luego, la decisión debe ser tomada conjuntamente con otros criterios y con la inspección visual. Para implementar el criterio de Geweke, basta con efectuar sólo una cadena larga del algoritmo MCMC.

2.3.3. Criterio de Gelman y Rubin

Gelman y Rubin (1992) [10] proponen un criterio que consiste en la comparación de las varianzas entre y dentro de las cadenas. Para desarrollar el criterio, son necesarias $I \geq 2$ cadenas, considerando que I no debe ser muy grande para evitar desperdicios de tiempo computacional. Sean θ el parámetro de interés; θ_i^j el j -ésimo valor muestreado en la i -ésima cadena, para $i = 1, \dots, I$ y $j = 1, \dots, J$; $\bar{\theta}_i$ la media de las observaciones de la cadena i ; $\bar{\theta}$ la media global. Entonces, las varianzas entre cadenas \hat{B} y dentro de las cadenas \hat{W} son dadas por

$$\hat{B} = \frac{J}{I-1} \sum_{i=1}^I (\bar{\theta}_i - \bar{\theta})^2$$

y

$$\hat{W} = \frac{1}{I(J-1)} \sum_{i=1}^I \sum_{j=1}^J (\theta_i^j - \bar{\theta}_i)^2$$

Todos los IJ valores de θ fueron muestreados de su distribución a posteriori y la varianza de θ puede ser estimada por \hat{B} , \hat{W} y por la media ponderada de $\hat{\sigma}^2 = (1 - 1/J)\hat{W} + (1/J)\hat{B}$.

Cuando las cadenas no convergen, las trayectorias de los parámetros siguen siendo influenciadas por los valores iniciales. Hasta que la convergencia sea alcanzada, σ^2 será superestimada por $\hat{\sigma}^2$ y subestimada por \hat{W} . Siguiendo este razonamiento, Brooks y Gelman (1998) [4] propusieron un análisis gráfico que tiene como objetivo verificar si: las varianzas $\hat{\sigma}^2$ y \hat{W} se estabilizan como función de J ; el valor del PSRF (potential scale reduction factor) $\hat{R} = \hat{V}/\hat{W}$, con $\hat{V} = \hat{\sigma}^2 + \hat{B}/(IJ)$ se aproxima a 1 (un valor aceptable indicado por el criterio "Rule of thumb" o "regla del pulgar" es de un $\hat{R} < 1,2$). Así se verifica la convergencia cuando estas condiciones son satisfechas.

2.3.4. Criterio de Heidelberger y Welch

El test de convergencia de Heidelberger y Welch [15] [16] usa el estadístico Cramer-von-Mises para testear la hipótesis nula que indica que la cadena de Markov proviene de una distribución estacionaria. El test se aplica sucesivamente, en primer lugar a toda la cadena, luego si la hipótesis nula es rechazada se descarta el primer 10 % de la cadena, si se sigue rechazando la hipótesis nula se descarta el 20 %, luego el 30 %, 40 %, etc... hasta que la hipótesis nula es aceptada o hasta que se ha descartado un 50 % de la cadena. Este último resultado constituye falla del test de estacionaridad e indica que es necesario correr una cadena MCMC más larga. Si el test de estacionaridad es pasado, software como R-Project (paquete lattice; librería CODA) [20] reportan el número de iteraciones que son necesarias mantener y el número de iteraciones que se pueden descartar.

Cumplida la etapa anterior, se usa la porción de la cadena que ha pasado el test de estacionaridad para ejecutar el test half-width que calcula el ancho medio del intervalo de credibilidad al $(1 - \alpha) \%$ para la media. Si la razón entre el half-width y la media es menor que un valor ϵ , entonces la cadena pasa el test half-width. De otro modo, se concluye que la cadena considerada no es lo suficientemente larga como para estimar la media con suficiente precisión.

Capítulo 3

Regresión beta con efectos mixtos: Enfoque Bayesiano

Típicamente los modelos lineales generalizados han sido tratados asumiendo Normalidad, homocedasticidad y considerando un soporte de la variable de respuesta sobre todos los reales, aún cuando en un buen número de casos se tengan variables naturalmente acotadas por un mínimo y un máximo, como por ejemplo altura, peso, etc. La propuesta del presente capítulo es modelar respuestas que tienen soporte sobre un intervalo acotado (a, b) con $a < b$ y $a, b \in \mathbb{R}$, las cuales pueden presentar heterocedasticidad y algún tipo de asimetría. Para tal efecto, se propone una variable de respuesta modelada por medio de la distribución beta de la forma $y|\mu, \phi \sim \text{beta}(\mu\phi, (1 - \mu)\phi)$ como se ha descrito en (1.1.5).

Abordaremos en este capítulo un modelo de Regresión beta bajo el enfoque Bayesiano, pero considerando además un modelo extendido en donde se dan cabida a efectos fijos y aleatorios.

3.1. Introducción

Comúnmente es necesario analizar datos que se encuentran relacionados con otras variables y en donde se pueden distinguir dos grupos de variables explicativas, uno con variables que fueron medidas de forma exacta o sin errores y otro grupo con variables que fueron medidas con cierto sesgo. Además si pensamos que un experimento puede ser replicado mas de una vez en cada individuo en diferentes instantes, entonces podremos esperar que exista alguna dependencia entre estas mediciones por individuo. Este tipo de situaciones comúnmente son resueltas aplicando un modelo de regresión lineal con efectos mixtos. Este tipo de modelo se basa habitualmente en algunas presunciones como el tener distribución Normal de la variable de respuesta o bien que ésta posea alguna distribución simétrica con soporte en \mathbb{R} . Luego, no serán apropiados, por ejemplo, para modelar tasas y proporciones en donde la respuesta está restringida al intervalo $(0,1)$, pues por una parte, las estimaciones podrían exceder estos límites y por otra parte porque en la práctica, este tipo de datos suelen presentarse asimétricos. Nuestro objetivo es proponer un modelo de regresión con efectos mixtos, adecuado para situaciones en donde la variable dependiente o de respuesta (y) sea una medición continua, acotada a un intervalo (a,b) con $a < b$ ($a, b \in \mathbb{R}$) y distribuya de forma asimétrica. De esta forma, proponemos un modelo de la familia beta, que es conocida por ser muy flexible, puesto que su densidad puede adoptar diferentes formas (simétricas o asimétricas) dependiendo de los dos parámetros indexados en la distribución.

En ocasiones, se tiene además valiosa información relacionada con el estudio en cuestión, ya sea proveniente de estudios anteriores o bien que es aportada por expertos en el tema. Luego, será conveniente adicionar de alguna manera esta información a

priori en nuestro modelo. Esto es posible utilizando el enfoque Bayesiano, el cual presenta como ventajas el que se puede incluir información a priori sobre los parámetros, no tiene dependencia con los tamaños muestrales, lo que podría disminuir eventuales costos de muestreo y por otra parte su implementación computacional no presenta grandes dificultades. El análisis Bayesiano nos permite pensar en distribuciones a priori sobre los parámetros, una vez observado el proceso de muestreo.

Consideraremos entonces en este capítulo un enfoque Bayesiano para modelar la regresión beta con efectos mixtos. Este modelo especifica la media poblacional transformada mediante una función link, como una combinación lineal de covariables.

Un enfoque clásico del modelo de regresión beta fue estudiado en detalle por Ferrari & Cribari-Neto (2004) [7], donde se usaron las propiedades asintóticas de los estimadores máximo verosímiles para realizar inferencia. Por el contrario, las inferencias realizadas bajo la perspectiva Bayesiana, no requieren de estas aproximaciones basadas en grandes muestras. Esto es de gran beneficio en un sin número de casos prácticos en donde se dispone de un conjunto de datos con pocas observaciones. Sin embargo, cuando se disponen de pocas observaciones, las prioris en el análisis Bayesiano tendrán mayor influencia en la inferencia que cuando se dispone de un mayor número de observaciones. Este enfoque Bayesiano del modelo de regresión beta fue estudiado por Branscum, Johnson & Thurmond (2007) [3], que complementa de alguna forma el estudio hecho por Ferrari & Cribari-Neto (2004) [7], pero que no incluye efectos del tipo aleatorios en las variables independientes o explicativas.

Tal como se ha expuesto, buscamos modelar una variable dependiente continua y limitada entre dos extremos conocidos. Como es sabido, el soporte de la distribución beta está restringido únicamente al intervalo $(0,1)$, luego será necesario transformar

los datos que se escapen de estas bandas mediante una simple transformación de la forma $y^* = (y - a)/(b - a)$ donde b y a son el máximo y mínimo valor posible respectivamente que puede tomar la variable de respuesta en el experimento. Además será necesario comprimir aún mas este rango para evitar los ceros y unos tomando $y^{**} = [y^*(N - 1) + 1/2]/N$ donde N es el tamaño muestral (McMillan and Creelman, 2005, pp. 8-9) [17].

En concordancia con lo anterior, para realizar el análisis Bayesiano implementaremos aproximaciones del tipo MCMC, puesto que los cálculos directos para efectuar la estimación de los parámetros son difíciles de realizar en este tipo de modelos. Un caso especial de aproximaciones MCMC que requiere solo de la especificación de las distribuciones condicionales posterioris completas para cada parámetro es el Gibbs sampler, que en nuestro caso no resulta difícil de implementar.

En lo que sigue del presente trabajo, obtendremos las respectivas distribuciones posterioris condicionales para implementar la aproximación Gibbs sampler, posteriormente se realizarán las estimaciones de los parámetros de interés y las pruebas de diagnóstico correspondientes. Probaremos los modelos propuestos realizando diversas simulaciones, con lo que podremos vislumbrar cuan buenas son las estimaciones conseguidas respecto al valor de los parámetros reales.

3.2. Propuesta del modelo y justificación de pri- oris

La distribución beta es una elección natural para modelar datos continuos restringidos a un intervalo (0,1). Además, cuando se quiere realizar un análisis de regresión, es común considerar la reparametrización dada en (1.1.5), pues lo que se busca es modelar la media de la variable de respuesta y en ese sentido resulta conveniente dejarla expresada como un parámetro del modelo.

De acuerdo a lo expuesto hasta ahora, en el presente trabajo se propone un modelo de Regresión con la siguiente estructura:

$$G(E(Y_i|b_i)) = X_i\beta + Z_ib_i = \eta_i \quad (3.2.1)$$

Donde $G(\cdot)$ es una función que liga la media de la variable de respuesta con la estructura de regresión del modelo y donde Y_1, Y_2, \dots, Y_n son vectores aleatorios independientes con $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})^t$; $b_i = (b_{i1}, b_{i2}, \dots, b_{iq})^t$ $i=1, 2, \dots, m$.

Comúnmente se considera $b_i|\psi \sim N_q(0, \psi)$, supuesto que será discutido más adelante. Además tenemos que $Y_{ij}|b_i, \beta, \phi \stackrel{ind}{\sim} \text{beta}(\mu_{ij}\phi, (1 - \mu_{ij})\phi)$ $i=1, 2, \dots, m$; $j=1, 2, \dots, n_i$. Es decir, $Y_{ij}|b_i, \beta, \phi \stackrel{dist}{=} Y_{ij}|\mu_{ij}, \phi$ donde su función de densidad queda definida como:

$$f(y_{ij}|\mu_{ij}, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_{ij}\phi)\Gamma((1 - \mu_{ij})\phi)} y_{ij}^{\mu_{ij}\phi-1} (1 - y_{ij})^{(1-\mu_{ij})\phi-1} , 0 < y_{ij} < 1$$

La interpretación de los regresores y parámetros de regresión en (3.2.1) es el siguiente:

1. Y_i es el $(n_i \times 1)$ vector de repuesta para las observaciones en el i-simo grupo.

2. X_i es el $(n_i \times p)$ modelo matricial para los efectos fijos para observaciones en el i -simo grupo.
3. β es el $(p \times 1)$ vector aleatorio de coeficientes de efectos fijos.
4. Z_i es el $(n_i \times q)$ modelo matricial para los efectos aleatorios para observaciones en el i -simo grupo.
5. b_i es el $(q \times 1)$ vector de coeficientes de efectos aleatorios para el grupo i -simo.
6. ψ es la $(q \times q)$ matriz de covarianzas aleatoria para los efectos aleatorios.
7. μ_i es el vector de valores esperados del vector Y_i .
8. ϕ es una medida de precisión de Y_{ij} .

Existen diferentes alternativas para explicitar la función link $G(\cdot)$, siendo las siguientes funciones las elecciones más comunes:

$$G(w) = \begin{cases} \log \frac{w}{1-w}, & \text{logística} \\ \Phi^{-1}(w), & \text{inversa de la función de distribución acumulada} \\ & \text{de la Normal estándar} \\ \log(-\log(1-w)), & \text{complemento log-log .} \end{cases}$$

En el planteamiento de un modelo de regresión beta, es común utilizar la función link logit, esto por la facilidad que presenta en cuanto a su implementación y además por la interpretabilidad que se obtiene sobre los parámetros estimados. Por ejemplo, la expresión $\exp(\beta_k)$ representa la razón de chances cuando se aumenta la variable x_k en una unidad y se mantienen las demás variables constantes (Ferrari y Cribari-Neto (2004)) [7].

Reescribiendo la ecuación (3.2.1) se expone a continuación el modelo en estudio usando la función link logit:

$$\begin{aligned} \eta_{ij} = G(\mu_{ij}) &= \log \left\{ \frac{\mu_{ij}}{1 - \mu_{ij}} \right\} = x_{ij}^t \beta + z_{ij}^t b_i & (3.2.2) \\ \Leftrightarrow \mu_i &= \frac{\exp(x_{ij}^t \beta + z_{ij}^t b_i)}{1 + \exp(x_{ij}^t \beta + z_{ij}^t b_i)} = \frac{\exp(\eta_{ij})}{1 + \exp(\eta_{ij})} \end{aligned}$$

Con $x_{ij} = (x_{ij1}, \dots, x_{ijp})^t$ y $z_{ij} = (z_{ij1}, \dots, z_{ijq})^t$.

En modelos de regresión con efectos mixtos, el vector de efectos aleatorios b_i asume típicamente una distribución Normal multivariada, sin embargo, dicha distribución podría no ser apropiada en situaciones donde las mediciones de las variables explicativas medidas con error, presenten outliers. En tales casos, resulta más apropiado considerar una distribución que posea colas mas pesadas. En este sentido, una distribución t-student multivariada es un mejor candidato para modelar b_i , es decir, consideramos $b_i \sim t_q(\nu_b, \mu_b, \Sigma_b)$ con $\nu_b > 0$ grados de libertad, $\mu_b = (0, \dots, 0)^t \in \mathbb{R}^q$ y parámetro de escala Σ_b (matriz $q \times q$ simétrica y definida positiva). No obstante lo anterior, podemos notar que para una valor grande de ν_b , la distribución t-student multivariada es aproximadamente equivalente a la distribución Normal multivariada.

En el modelo de regresión beta mixto propuesto antes, hemos considerado el parámetro de precisión ϕ como un parámetro común o definido también como un parámetro molesto (Ferrari and Cribari-Neto (2004))[7]. Sin embargo, para una formulación mas general de este modelo, podemos considerar diferentes parámetros de precisión ϕ_{ij} asociados a cada respuesta y_{ij} $i = 1, \dots, m$, $j = 1, \dots, n_i$. Bajo esta situación, es usual asumir un modelo lineal mixto para el logaritmo de los parámetros

de precisión ϕ_{ij} (Smithson and Verkuilen (2006)) [23], de la forma

$$\ln(\phi_{ij}) = w_{ij}^t \delta + h_{ij}^t d_i, \quad (3.2.3)$$

donde $w_{ij} = (w_{ij1}, \dots, w_{ijp^*})^t$ es el vector de diseño correspondiente al $p^* \times 1$ vector δ de efectos fijos y $h_{ij} = (h_{ij1}, \dots, h_{ijq^*})^t$ es el vector de diseño correspondiente al $q^* \times 1$ vector d_i de efectos aleatorios. Note que, las matrices de diseño $W_i = (w_{i1}, \dots, w_{in_i})^t$ y $H_i = (h_{i1}, \dots, h_{in_i})^t$ puede, aunque esto no es un requisito, contener las mismas variables predictoras que las matrices $X_i = (x_{i1}, \dots, x_{in_i})^t$ y $Z_i = (z_{i1}, \dots, z_{in_i})^t$, respectivamente. Mas aún, en general consideraremos para W_i algunas columnas de X_i con $q^* \leq q$ y para H_i algunas columnas de Z_i con $p^* \leq p$.

A continuación, analizaremos el modelo de regresión beta con efectos mixtos, diferenciando entre modelos con parámetro de precisión molesto y modelos con parámetro de precisión asociado a un submodelo de regresión con efectos mixtos.

3.3. Modelo de regresión beta mixto con parámetro de precisión molesto

En esta sección consideraremos el parámetro de precisión ϕ como molesto, es decir, sin una estructura de regresión asociada, por lo que nos abocaremos a discutir qué priori es la mas apropiada para modelar este parámetro, basándonos tanto en propuestas formuladas en la literatura como en una propuesta propia. Además, se discutirán las prioris que resultan ser más adecuadas para la especificación del modelo propuesto y luego se realizará el ajuste del mismo usando un software adecuado.

3.3.1. Discusión de prioris

Para realizar una especificación completa del modelo Bayesiano es necesario elicitar distribuciones a priori para todos los parámetros desconocidos en el modelo. Los efectos fijos son comúnmente modelados usando una distribución Normal multivariada, es decir, se asume $\beta | \mu_\beta, \Sigma_\beta \sim N_p(\mu_\beta, \Sigma_\beta)$. Usualmente, se desconoce algún tipo de información sobre el parámetro β , luego se suelen especificar prioris vagas para dicho parámetro tomando, por ejemplo, valores altos en la matriz de covarianzas Σ_β . Sin embargo, el impacto que tiene la elección de este parámetro de escala bajo el modelo Normal, no puede ser negado. Luego, es razonable utilizar una estrategia alternativa, por lo que consideramos apropiado modelar el efecto fijo mediante una distribución t-multivariada con grados de libertad pequeños, pues a través de esta elección se consigue una distribución de colas mas pesadas y en este sentido, una elección de una priori mas vaga para β , con lo anterior, consideramos $\beta | \nu_\beta, \mu_\beta, \Sigma_\beta \sim t_p(\nu_\beta, \mu_\beta, \Sigma_\beta)$. Si asumimos que el vector de efectos aleatorios b_i sigue una distribución t-multivariada, es decir, $b_i | \nu_b, \mu_b, \Sigma_b \sim t_q(\nu_b, \mu_b, \Sigma_b)$, donde $\mu_b = 0 \in \mathbb{R}^q$, entonces la distribución a priori para los grados de libertad la podemos asumir discreta como en Albert and Chib (1993) ([1]) y Besag et al. (1995) ([2]), o bien continua como en Geweke (1993) [12]. Para efectos de este estudio, seleccionaremos esta última alternativa, mas específicamente consideraremos una priori exponencial para los grados de libertad. La priori para el parámetro de escala matricial Σ_b de los efectos aleatorios, es escogida principalmente por simplicidad computacional, luego consideramos la distribución Wishart inversa como priori para este parámetro, tal como ha sido usada en Fong et al. (2010) [8], es decir, $\Sigma_b | \psi, c \sim IW(\psi, c)$. Una priori alternativa para Σ_b es la distribución Wishart contraída (Philip and Carl (2000)) [19].

Tal como se ha expuesto en el título de esta sección, consideraremos un modelo mixto beta con un parámetro de precisión común ϕ (Ferrari and Cribari-Neto (2004)). Típicamente, en el contexto Bayesiano, se selecciona una priori gamma inversa para este parámetro, pero siguiendo las recomendaciones dadas por Gelman (2006) [11], la priori uniforme elevada al cuadrado es un mejor candidato para este parámetro, es decir, $\phi = U^2$ donde $U \sim Unif(0, a)$ (típicamente se considera $a = 50$) [11]. Siguiendo esta analogía y considerando que las distribuciones $Unif(0, 1)$ y $beta(1, 1)$ son equivalentes, hemos propuesto una priori para el parámetro ϕ de la forma $\phi = (a * B)^2$ donde $B \sim beta(1 + \epsilon, 1 + \epsilon)$ con $\epsilon \downarrow 0$, con lo que se consigue una priori mas flexible y que contiene a la priori propuesta por Gelman (2006) [11] y que puede ser además mas vaga que ésta, dependiendo del valor que se le asigne a ϵ .

3.3.2. Ajuste del modelo usando algoritmo de Monte Carlo vía cadenas de Markov

Sea $y^t = (y_1^t, \dots, y_m^t)$ y $\eta^t = (\eta_1^t, \dots, \eta_m^t)$. Donde condicionalmente en β, Σ_b y ν_b los η_i son independientes con $f(\eta_i|\beta, \Sigma_b, \nu_b) \propto f(b_i|\beta, \Sigma_b, \nu_b)$, $i = 1, \dots, m$. Para obtener las estimaciones de los parámetros desconocidos del modelo, es necesario muestrear de la distribución a posteriori conjunta dada a continuación

$$\begin{aligned} f(\beta, \Sigma_b, \nu_b, \phi, \eta|y) &\propto f(y|\eta, \phi)f(\eta|\beta, \Sigma_b, \nu_b)f(\Sigma_b)f(\nu_b)f(\beta)f(\phi) \\ &\propto \left[\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right] \left[\prod_{i=1}^m f(\eta_i|\beta, \Sigma_b, \nu_b) \right] f(\Sigma_b)f(\nu_b)f(\phi)f(\beta) \end{aligned} \quad (3.3.1)$$

Para la especificación de (3.3.1) consideramos que $\Sigma_b \perp\!\!\!\perp \nu_b \perp\!\!\!\perp \phi \perp\!\!\!\perp \beta$.

Podemos entonces utilizar el muestreo de Gibbs para generar una muestra Monte

Carlo de la posteriori conjunta $f(\beta, \Sigma_b, \nu_b, \phi, \eta|y)$. El muestreo de Gibbs en este contexto involucra un muestreo iterativo desde las distribuciones condicionales completas $f(\Sigma_b|\nu_b, \beta, \phi, \eta, y)$, $f(\nu_b|\Sigma_b, \beta, \phi, \eta, y)$, $f(\beta|\Sigma_b, \nu_b, \phi, \eta, y)$, $f(\phi|\beta, \Sigma_b, \nu_b, \eta, y)$ y $f(\eta_i|\eta_k, \beta, \Sigma_b, \nu_b, \phi, y_i)$, $i, k = 1, \dots, m$, $i \neq k$, las cuales pueden ser descompuestas como sigue

$$f(\Sigma_b|\nu_b, \beta, \phi, \eta, y) \propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\Sigma_b) \quad (3.3.2)$$

$$f(\nu_b|\Sigma_b, \beta, \phi, \eta, y) \propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\nu_b) \quad (3.3.3)$$

$$f(\beta|\Sigma_b, \nu_b, \phi, \eta, y) \propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\beta) \quad (3.3.4)$$

$$f(\phi|\beta, \Sigma_b, \nu_b, \eta, y) \propto \left(\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(\phi) \quad (3.3.5)$$

$$f(\eta_i|\eta_k, \beta, \Sigma_b, \nu_b, \phi, y_i) \propto \left(\prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(b_i|\beta, \Sigma_b, \nu_b) \quad (i \neq k; i, k = 1, \dots, m) \quad (3.3.6)$$

Demostración: Antes que nada, demostraremos que $f(\eta_i|\beta, \Sigma_b, \nu_b) \propto f(b_i|\beta, \Sigma_b, \nu_b)$.

Notemos que $f(\eta|\beta, \Sigma_b, \nu_b) = \prod_{i=1}^m f(\eta_i|\beta, \Sigma_b, \nu_b)$

donde $g(b_i) = \eta_i = X_i\beta + Z_i b_i \Leftrightarrow b_i = (Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i\beta)$.

luego, $f(\eta_i|\beta, \Sigma_b, \nu_b) = f(b_i|\beta, \Sigma_b, \nu_b) \left| \frac{\partial g^{-1}(\eta_i)}{\partial \eta_i} \right| \propto f(b_i|\beta, \Sigma_b, \nu_b)$

De esta forma se obtiene en (3.3.2) que

$$\begin{aligned} f(\Sigma_b|\nu_b, \beta, \phi, \eta, y) &\propto f(\eta|\beta, \Sigma_b, \nu_b) f(\Sigma_b) \propto f(b|\beta, \Sigma_b, \nu_b) f(\Sigma_b) \\ &\propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\Sigma_b) \end{aligned}$$

De forma análoga en (3.3.3) y (3.3.4) se tiene

$$\begin{aligned}
 f(\nu_b|\Sigma_b, \beta, \phi, \eta, y) &\propto f(\eta|\beta, \Sigma_b, \nu_b)f(\nu_b) \propto f(b|\beta, \Sigma_b, \nu_b)f(\nu_b) \\
 &\propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\nu_b) \\
 f(\beta|\Sigma_b, \nu_b, \phi, \eta, y) &\propto f(\eta|\beta, \Sigma_b, \nu_b)f(\beta) \propto f(b|\beta, \Sigma_b, \nu_b)f(\beta) \\
 &\propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\beta)
 \end{aligned}$$

La ecuación (3.3.5) se obtiene de forma directa

$$f(\phi|\beta, \Sigma_b, \nu_b, \eta, y) \propto f(y|\eta, \phi)f(\phi) \propto \left(\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(\phi)$$

Y la ecuación (3.3.6) se obtiene fácilmente como

$$\begin{aligned}
 f(\eta_i|\eta_k, \beta, \Sigma_b, \nu_b, \phi, y_i) &\propto f(y_i|\eta_i, \phi)f(\eta_i|\beta, \Sigma_b, \nu_b) \propto f(y_i|\eta_i, \phi)f(b_i|\beta, \Sigma_b, \nu_b) \\
 &\propto \left(\prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(b_i|\beta, \Sigma_b, \nu_b) \quad (i \neq k; i, k = 1, \dots, m)
 \end{aligned}$$

Cabe señalar que si a los efectos aleatorios b_i se les asigna una distribución Normal multivariada y se mantienen las prioris descritas en la sección (3.3.1), entonces se consiguen prioris conjugadas para los parámetros β y Σ_b pero no así para los parámetros ϕ y $\eta_i|\beta, \Sigma_b$ ($i = 1, \dots, m$). Sin embargo, el presente trabajo se ha desarrollado bajo el supuesto de que los efectos aleatorios poseen distribución t-multivariada, con lo cual no se consiguen prioris conjugadas para ninguno de los parámetros desconocidos del modelo (ver apéndice).

No obstante lo anterior, el modelo puede ser implementado sin mayor dificultad en el software WinBUGS. Inferencias a posteriori en $\beta, \Sigma_b, \phi, \nu_b$ y, principalmente en las respuestas esperadas $(\mu_{ij}; i = 1, \dots, m, j = 1, \dots, n_i)$ son fácilmente obtenidas en el software mencionado. Los test de hipótesis respecto a los coeficientes de regresión

y las respuestas esperadas son también fáciles de obtener.

3.3.3. Ejemplo con datos simulados

Para ilustrar la metodología propuesta, consideraremos el siguiente modelo lineal mixto con datos simulados:

$$\eta_{ij} = \beta_1 + x_{ij2}\beta_2 + x_{ij3}\beta_3 + b_{i1} + z_{ij2}b_{i2},$$

$i = 1, \dots, 100$; $j = 1, \dots, 5$, donde $\eta_{ij} = \ln\{\mu_{ij}/(1 - \mu_{ij})\}$, $z_{ij1} = x_{ij1} = 1$, $z_{ij2} = x_{ij2}$. El valor de las covariables fue generado desde una distribución uniforme definida sobre el intervalo unitario. Se han especificado los vectores de efectos fijos y aleatorios por $\beta = (\beta_1, \beta_2, \beta_3)^\top$ y $b_i = (b_{i1}, b_{i2})^\top$ respectivamente. Adicionalmente, para realizar la simulación, también asumimos que:

$$y_{ij}|b_i, \phi, \beta \sim \text{beta}(\mu_{ij}\phi, (1 - \mu_{ij})\phi),$$

$$\phi = 49,$$

$$b_i|\nu_b, \Sigma_b \sim t_2(\nu_b, 0, \Sigma_b),$$

$$\nu_b = 1000,$$

$$\Sigma_b = \begin{pmatrix} 1,09 & -0,36 \\ -0,36 & 0,13 \end{pmatrix},$$

$$\beta = (-2, 1, 2),$$

Estos valores asignados para cada uno de los parámetros mencionados, servirán como referencia para evaluar cuán buenas son las estimaciones de los parámetros luego de realizar el ajuste de los modelos propuestos.

Para realizar el ajuste del modelo y de acuerdo con lo expuesto en la sección (3.3.1), se utilizarán para β , ν_b y Σ_b las siguientes prioris:

$$\begin{aligned} \beta &\sim t_3(\nu_\beta, \mu_\beta, \Sigma_\beta) \\ \nu_\beta &= 20; \quad \mu_\beta = (0, 0, 0); \quad \Sigma_\beta = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\ b &\sim t_2(\nu_b, 0, \Sigma_b) \\ \nu_b &\sim \exp(0, 001); \quad \Sigma_b \sim IW(\psi, c); \quad \psi = \begin{pmatrix} 5 & 15 \\ 15 & 45 \end{pmatrix}; \quad c = 5. \end{aligned}$$

Una distribución gamma inversa del tipo $\phi \sim IG(\epsilon, \epsilon)$ con valores pequeños para ϵ ($\epsilon = 0,5, 0,1, 0,01, 0,001, \dots$) sería una elección natural para la distribución a priori del parámetro de precisión ϕ . Tal como mencionamos anteriormente Gelman (2006) [11] sugiere que la distribución a priori $\phi = U^2$ con $U \sim U(0, a)$ (con valores grandes para a ; $a = 50$ por ejemplo) es menos informativa que una priori gamma inversa. Continuando con nuestra propuesta (ver sección 3.3.1), adoptaremos una priori alternativa, es decir, asumiremos que $\phi = (a * B)^2$ con $B \sim \beta(1 + \epsilon, 1 + \epsilon)$.

El cuadro 3.1 reporta el "deviance information criterion" (DIC) para los modelos ajustados con diferentes distribuciones a priori para ϕ . En este cuadro podemos observar que las diferentes prioris propuestas, conducen a un DIC similar. Sin embargo el modelo 1d muestra un ajuste ligeramente mejor que los otros modelos propuestos.

Cuadro 3.1: *Análisis de sensibilidad para la especificación de la priori del parámetro de precisión*

<i>Model</i>	<i>Prior for ϕ</i>	<i>DIC</i>
model 1a	$\phi \sim IG(0,01, 0,01)$	-1245,02
model 1b	$\phi = U^2, U \sim U(0, 50)$	-1245,14
model 1c	$\phi = (50 * B_1)^2, B_1 \sim beta(1,1, 1,1)$	-1245,15
model 1d	$\phi = (50 * B_2)^2, B_2 \sim beta(1,5, 1,5)$	-1245,41

En el cuadro 3.2 reportamos la estimación de los parámetros del modelo 1d. Estos resultados muestran que los parámetros estimados mediante la metodología Bayesiana propuesta aquí, son similares a los reales valores de los parámetros del modelo.

Para realizar el ajuste del modelo, consideramos 100,000 iteraciones Monte Carlo y los resultados son presentados considerando las últimas 90,000 iteraciones. Adicionalmente, se han realizado los test de diagnóstico pertinentes, con lo que se han observado conductas deseables de las cadenas (ver cuadro 3.3, Figuras 3.1 y 3.2, y los comentarios hechos mas abajo). También se ha realizado un análisis de sensibilidad con respecto a la especificación de las prioris de los coeficientes de regresión y de la matriz de dispersión de los efectos aleatorios. En cada caso, las inferencias realizadas a posteriori no presentan alteraciones apreciables en comparación con los resultados presentados en el cuadro 3.2.

Cuadro 3.2: *Medias a posteriori exactas. Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto simulado*

Parameter	Posterior Inference				
	<i>True</i>	<i>Mean</i>	<i>MCError</i>	<i>Median</i>	<i>95 %CI</i>
β_1	-2	-2,065	0.00208	-2,065	(-2,281,-1,856)
β_2	1	1.037	0.00145	1.035	(0.9036,1.171)
β_3	2	1.949	0.001263	1.949	(1.824,2.065)
ϕ	49	44.16	0.07068	44.09	(38.34,50.61)
$\Sigma_{b_{11}}$	1.09	0.9814	0.003648	0.9636	(0.7086,1.33)
$\Sigma_{b_{12}}$	-0,36	-0,2574	0.001453	-0,2515	(-0,4078,-0,1306)
$\Sigma_{b_{22}}$	0.13	0.08028	5.979E-4	0.07683	(0.03528,0.1455)

La versión multivariada del diagnóstico de convergencia de Gelman y Rubin's propuesta por Brooks & Gelman (1998) [4] indica que la cadena es convergente, luego que el "multivariate proportional scale reduction factor" (mprf) es igual a 1.01 (mprf < 1,2). Además, al analizar la convergencia de la cadena para cada parámetro, podemos concluir que la convergencia es alcanzada en cada caso. Esta última conclusión es corroborada por diferentes test de convergencia como son el diagnóstico de convergencia de Gelman y Rubin's (Gelman & Rubin, 1992)[10], diagnóstico de Geweke's (Geweke, 1992)[13] y el diagnóstico de Heidelberg y Welch's (Heidelberger & Welch (1981) y Heidelberger & Welch (1983)[15] y [16]), los cuales fueron obtenidos usando las librerías `lattice` y `coda` (Plummer et al., 2006)[20] del software `R-Project`, que puede ser obtenido de forma libre desde la página <http://www.r-project.org/>. Para obtener el diagnóstico de convergencia de Gelman y Rubin, fueron inicializadas dos cadenas con distintos puntos de inicio y se realizaron 100,000 iteraciones Monte Carlo,

de las cuales han sido consideradas solo las últimas 90,000 iteraciones. Los resultados de los diagnósticos realizados son presentados en el cuadro 3.3. Adicionalmente, las figuras 3.1 y 3.2 sugieren que la cadena para cada parámetro es no correlacionada y estacionaria respectivamente. Estos resultados son esenciales para lograr una estimación adecuada de los parámetros del modelo.

Cuadro 3.3: *Test de diagnóstico de Gelman's, Geweke's y Heidelberg & Welch's*

Test	Test statistic						
	β_1	β_2	β_3	ϕ	$\Sigma_{b_{11}}$	$\Sigma_{b_{12}}$	$\Sigma_{b_{22}}$
Gelman							
Point Est.	1.00	1.00	1.00	1.00	1.00	1.00	1.01
Geweke							
Chain 1	-0,36638	0.01835	-0,41155	-0,27470	-0,82727	-0,17245	-1,09279
Chain 2	0.5569	0.4910	-1,2369	1.2388	-0,9661	0.7677	-0,3432
Heidel							
Chain 1							
Stationarity Test	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1
P-value	0.990	0.865	0.865	0.679	0.915	0.996	0.264
Halfwidth Test	passed	passed	passed	passed	passed	passed	passed
Mean	-2,0698	1.0394	1.9482	44.1187	0.9813	-0,2595	0.0815
Halfwidth	0.00625	0.00393	0.00366	0.17634	0.00940	0.00421	0.00168
Chain 2							
Stationarity Test	passed	passed	passed	passed	passed	passed	passed
Start iteration	1	1	1	1	1	1	1
P-value	0.9326	0.8428	0.8287	0.0624	0.6013	0.8085	0.4262
Halfwidth Test	passed	passed	passed	passed	passed	passed	passed
Mean	-2,061	1.034	1.949	44.195	0.981	-0.255	0.079
Halfwidth	0.00633	0.00460	0.00389	0.22512	0.00982	0.00416	0.00169

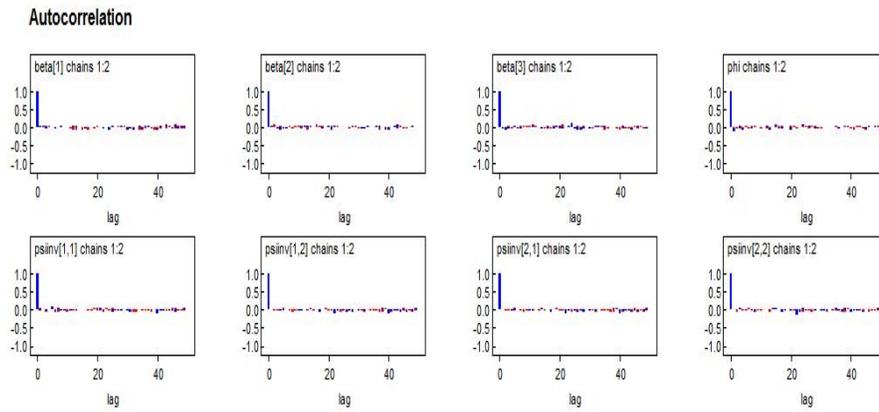


Figura 3.1: Autocorrelación - Datos simulados y parámetro ϕ común

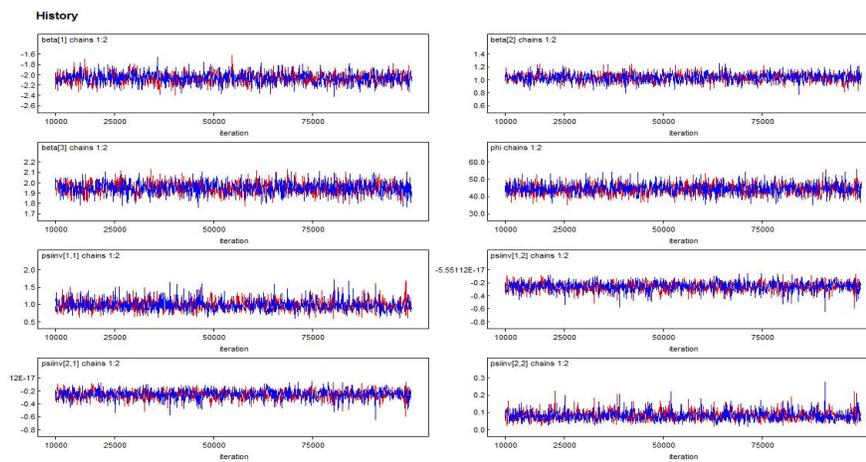


Figura 3.2: Historia - Datos simulados y parámetro ϕ común

3.3.4. Aplicación del modelo a datos reales

Este es un ejemplo famoso y muy estudiado en diversos artículos y hace referencia a los datos de petróleo de Nilon H. Prater (Prater, 1956). Prater, recolectó información del rendimiento de la gasolina de petróleo crudo en las diversas etapas conocidas como "puntos finales" del proceso de refinación. A cada petróleo crudo se le han hecho tres tipos de mediciones: una para la gravedad específica y dos mediciones diferentes para la presión de vapor.

Para clasificar los datos se pone de manifiesto que, aunque existen 32 observaciones, solo han sido medidos 10 crudos diferentes, donde no se tiene la misma cantidad de mediciones por crudo, por lo que se tienen datos desbalanceados.

Ajustaremos entonces un modelo de regresión beta con efectos mixtos como se ha propuesto en Venables (2000) [26], donde se enfoca el problema desde el punto de vista clásico y bajo una visión mas simple que la propuesta en este trabajo.

Las variables a considerar en el modelo se definen a continuación:

y_{ij} : Rendimiento como un porcentaje del crudo para el i -ésimo petróleo crudo del cual ha sido tomada la j -ésima muestra.

EP_{ij} : Volatilidad deseada de la gasolina (El punto final. Varía dentro de la muestra) para el i -ésimo petróleo crudo del cual ha sido tomada la j -ésima muestra

Donde $i = 1, \dots, 10$ y $j = 1, \dots, n_i$ con $n_1 = 4, n_2 = 3, n_3 = 3, n_4 = 4, n_5 = 3, n_6 = 3, n_7 = 4, n_8 = 3, n_9 = 2, n_{10} = 3$.

Para abordar el problema con los datos de Prater, llevamos la variable de respuesta

al intervalo $(0, 1)$ usando la transformación:

$$y_{ij}^* = \frac{y_{ij}}{100}$$

Luego, se plantea el modelo lineal mixto como:

$$\ln \left\{ \frac{\mu_{ij}}{1 - \mu_{ij}} \right\} = \eta_{ij} = x_{ij}^\top \beta + z_{ij}^\top b_i,$$

donde $\mu_{ij} = E(y_{ij}^* | b_i)$, $x_{ij} = (1, EP_{ij})^\top$, $z_{ij} = (1, EP_{ij})^\top$, $\beta = (\beta_1, \beta_2)$ y $b_i = (b_{i1}, b_{i2})$

Además, para realizar el ajuste del modelo, consideramos los siguientes supuestos:

$$y_{ij} | b_i, \phi, \beta \sim \text{beta}(\mu_{ij}\phi, (1 - \mu_{ij})\phi),$$

$$\beta \sim t_2(\nu_\beta, \mu_\beta, \Sigma_\beta)$$

$$\nu_\beta = 10; \quad \mu_\beta = (0, 0); \quad \Sigma_\beta = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

$$b \sim t_2(\nu_b, 0, \Sigma_b)$$

$$\nu_b \sim \text{exp}(0, 001); \quad \Sigma_b \sim IW(\psi, c); \quad \psi = \begin{pmatrix} 1000 & 0 \\ 0 & 1000 \end{pmatrix}; \quad c = 4$$

En el cuadro 3.4 se entrega el ajuste del modelo considerando algunas variantes en la propuesta del modelo de localización y en la priori utilizada para modelar el parámetro de precisión. Aquí podemos observar que las diferentes prioris propuestas ajustan similarmente bien el modelo de acuerdo al criterio DIC, sin embargo, el mejor ajuste se logra cuando se considera un modelo de localización que contiene solo el intercepto aleatorio. Particularmente, se puede observar que la priori beta propuesta, logra un buen ajuste del modelo.

Cuadro 3.4: *Análisis de sensibilidad para la especificación de la priori del parámetro de precisión*

<i>Model</i>	<i>LocationModel</i>	<i>Prior for ϕ</i>	<i>DIC</i>
Model 1.a.1	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}(\beta_2 + b_{i2})$	$\phi \sim IG(0,01, 0,01)$	-123,091
Model 1.a.2	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}(\beta_2 + b_{i2})$	$\phi = \mathbf{U}^2, \mathbf{U} \sim U(0, 50)$	-122,764
Model 1.a.3	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}(\beta_2 + b_{i2})$	$\phi = (50 * B1)^2, B1 \sim beta(1,1, 1,1)$	-122,825
Model 1.a.4	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}(\beta_2 + b_{i2})$	$\phi = (50 * B2)^2, B2 \sim beta(1,5, 1,5)$	-122,944
Model 1.b.1	$\eta_{ij} = \beta_1 + EP_{ij}(\beta_2 + b_{i2})$	$\phi \sim IG(0,01, 0,01)$	-123,050
Model 1.b.2	$\eta_{ij} = \beta_1 + EP_{ij}(\beta_2 + b_{i2})$	$\phi = \mathbf{U}^2, \mathbf{U} \sim U(0, 50)$	-122,664
Model 1.b.3	$\eta_{ij} = \beta_1 + EP_{ij}(\beta_2 + b_{i2})$	$\phi = (50 * B1)^2, B1 \sim beta(1,1, 1,1)$	-122,793
Model 1.b.4	$\eta_{ij} = \beta_1 + EP_{ij}(\beta_2 + b_{i2})$	$\phi = (50 * B2)^2, B2 \sim beta(1,5, 1,5)$	-122,865
Model 1.c.1	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}\beta_2$	$\phi \sim IG(0,01, 0,01)$	-142,474
Model 1.c.2	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}\beta_2$	$\phi = \mathbf{U}^2, \mathbf{U} \sim U(0, 50)$	-142,555
Model 1.c.3	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}\beta_2$	$\phi = (50 * B1)^2, B1 \sim beta(1,1, 1,1)$	-142,567
Model 1.c.4	$\eta_{ij} = (\beta_1 + b_{i1}) + EP_{ij}\beta_2$	$\phi = (50 * B2)^2, B2 \sim beta(1,5, 1,5)$	-142,595

En el cuadro 3.5 se presenta la estimación de los parámetros del modelo 1.c.4, el cual presentó un mejor ajuste en comparación con los otros modelos propuestos. Para realizar el ajuste del modelo, consideramos 200,000 iteraciones Monte Carlo y los resultados son presentados considerando las últimas 190,000 iteraciones. Adicionalmente, se han realizado los test de diagnóstico pertinentes, con lo que se han observado conductas deseables de las cadenas (ver figuras 3.3, 3.4 y 3.5).

Cuadro 3.5: *Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto dado en 1.c.4*

Parameter	Posterior Inference			
	<i>Mean</i>	<i>MCError</i>	<i>Median</i>	<i>95 %CI</i>
$\beta_1 + b_{11}$	-4.366	0.005409	-4.371	(-4.751,-3.971)
$\beta_1 + b_{21}$	-4.776	0.004731	-4.779	(-5.144,-4.392)
$\beta_1 + b_{31}$	-4.542	0.004726	-4.543	(-4.904,-4.176)
$\beta_1 + b_{41}$	-5.013	0.005462	-5.019	(-5.416,-4.585)
$\beta_1 + b_{51}$	-4.943	0.005473	-4.947	(-5.359,-4.516)
$\beta_1 + b_{61}$	-5.032	0.005661	-5.04	(-5.421,-4.607)
$\beta_1 + b_{71}$	-5.513	0.00576	-5.514	(-5.921,-5.098)
$\beta_1 + b_{81}$	-5.554	0.005908	-5.558	(-5.988,-5.108)
$\beta_1 + b_{91}$	-5.646	0.006673	-5.657	(-6.105,-5.172)
$\beta_1 + b_{10\ 1}$	-6.022	0.005887	-6.029	(-6.496,-5.525)
β_2	0.01071	1.485E-5	0.01074	(0.009606,0.01175)

La versión multivariada del diagnóstico de convergencia de Gelman y Rubin's propuesta por Brooks & Gelman (1998) [4] indica que la cadena es convergente, luego que el "multivariate proportional scale reduction factor" (mprf) es igual a 1.09 (mprf < 1,2). Adicionalmente, las figuras 3.3 y 3.4 sugieren que la cadena para cada parámetro es no correlacionada y estacionaria respectivamente, la figura 3.5 nos muestra que la función de densidad a posteriori de cada parámetro no presenta bimodalidad, lo cual es un buen indicio puesto que la presencia de esta suele venir acompañada con problemas en la convergencia de la cadena. Estos resultados mencionados son esenciales para lograr una estimación adecuada de los parámetros del modelo, por lo que las estimaciones descritas en el cuadro 3.5 son confiables.

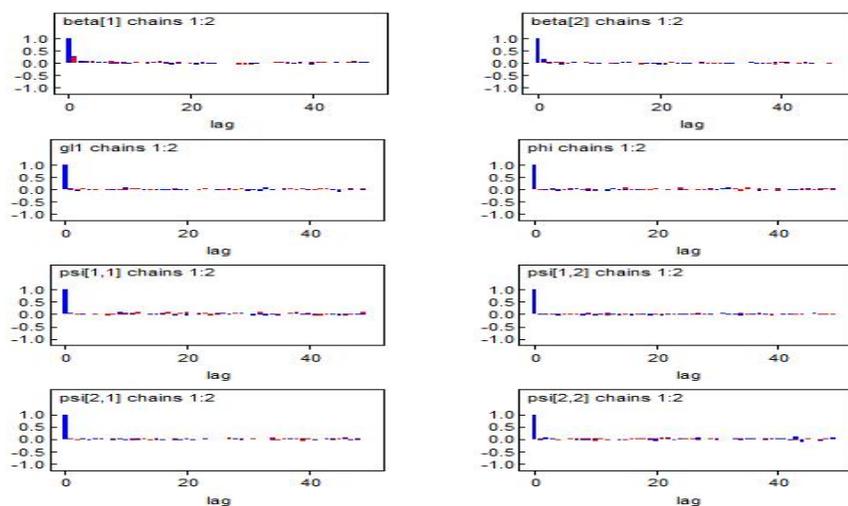


Figura 3.3: *Autocorrelación - Datos de Prater y parámetro ϕ común*

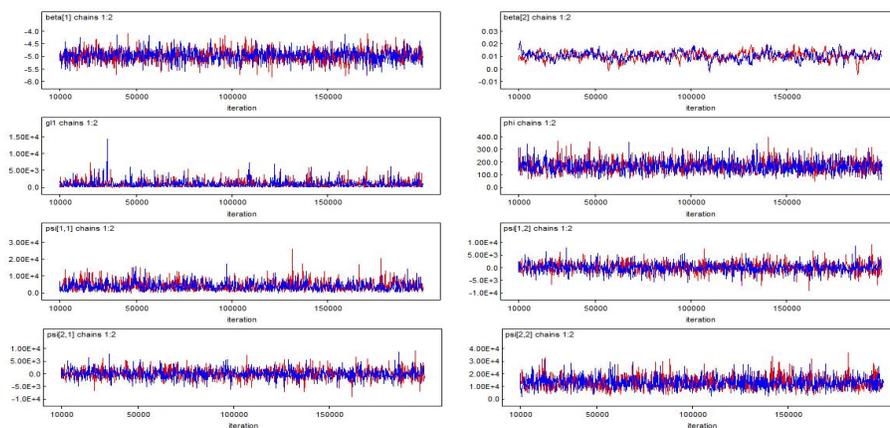


Figura 3.4: *Historia - Datos de Prater y parámetro ϕ común*

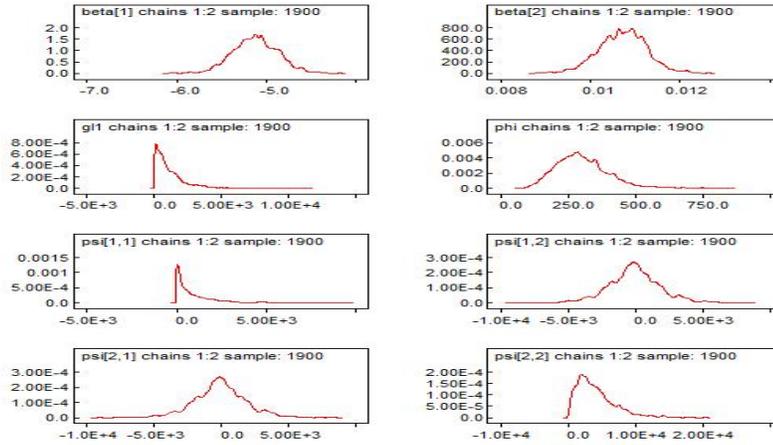


Figura 3.5: Densidad - Datos de Prater y parámetro ϕ común

3.3.5. Apéndice

A.1. Distribución a posteriori bajo el modelo propuesto

Note que de la ecuación (3.3.1) se obtiene

$$\begin{aligned}
 f(\beta, \Sigma_b, \nu_b, \phi, \eta|y) &\propto \left[\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right] \left[\prod_{i=1}^m f(\eta_i|\beta, \Sigma_b, \nu_b) \right] f(\Sigma_b) f(\nu_b) f(\phi) f(\beta) \\
 &\propto \left[\frac{(\Gamma(\phi))^n}{\prod_{i=1}^m \prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})}\phi\right) \right]} \prod_{i=1}^m \prod_{j=1}^{n_i} \left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi^{-1} (1-y_{ij})^{\frac{1}{1+\exp(\eta_{ij})}\phi^{-1}} \right) \right] \\
 &\times \left[\frac{\Gamma(\frac{\nu_b+q}{2})}{\Gamma(\frac{\nu_b}{2})(\nu_b\pi)^{\frac{q}{2}}} \right]^m |\Sigma_b|^{-m/2} \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \\
 &\times \frac{|\psi|^{\frac{c}{2}} |\Sigma_b|^{-\frac{(c+q+1)}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr}(\psi \Sigma_b^{-1}) \right\}}{2^{\frac{cq}{2}} \Gamma_q \left(\frac{c}{2} \right)} \\
 &\times \lambda_0 e^{-\lambda_0 \nu_b} \times \frac{1}{2a} \phi^{-1/2} \frac{\Gamma(2+2\epsilon)}{\Gamma^2(1+\epsilon)} \left(\frac{\phi^{1/2}}{a} \right)^\epsilon \left(1 - \frac{\phi^{1/2}}{a} \right)^\epsilon \\
 &\times \left[\frac{\Gamma(\frac{\nu_\beta+p}{2})}{\Gamma(\frac{\nu_\beta}{2})(\nu_\beta\pi)^{\frac{p}{2}}} \right] |\Sigma_\beta|^{-1/2} \left[1 + \frac{1}{\nu_\beta} ((\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta)) \right]^{\frac{\nu_\beta+p}{2}}
 \end{aligned}$$

Donde $b_i = (Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta)$, $n = \sum_{i=1}^m n_i$ y $\Gamma_c(\bullet)$ es la función de densidad Gamma c-variada.

A.2. Condicionales completas bajo el modelo propuesto

De las ecuaciones (3.3.2) a la (3.3.6) se obtienen las siguientes condicionales completas:

$$\begin{aligned} f(\Sigma_b | \nu_b, \beta, \phi, \eta, y) &\propto \left(\prod_{i=1}^m f(b_i | \beta, \Sigma_b, \nu_b) \right) f(\Sigma_b) \\ &\propto |\Sigma_b|^{-m/2} \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \times |\Sigma_b|^{-\frac{(c+q+1)}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr}(\psi \Sigma_b^{-1}) \right\} \end{aligned}$$

$$\begin{aligned} f(\nu_b | \Sigma_b, \beta, \phi, \eta, y) &\propto \left(\prod_{i=1}^m f(b_i | \beta, \Sigma_b, \nu_b) \right) f(\nu_b) \\ &\propto \left[\frac{\Gamma(\frac{\nu_b+q}{2})}{\Gamma(\frac{\nu_b}{2})(\nu_b \pi)^{\frac{q}{2}}} \right]^m \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \times e^{-\lambda_0 \nu_b} \end{aligned}$$

$$\begin{aligned} f(\beta | \Sigma_b, \nu_b, \phi, \eta, y) &\propto \left(\prod_{i=1}^m f(b_i | \beta, \Sigma_b, \nu_b) \right) f(\beta) \\ &\propto \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \times \left[1 + \frac{1}{\nu_\beta} ((\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta)) \right]^{\frac{\nu_\beta+p}{2}} \end{aligned}$$

$$\begin{aligned} f(\phi | \beta, \Sigma_b, \nu_b, \eta, y) &\propto \left(\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \eta_{ij}, \phi) \right) f(\phi) \\ &\propto \left[\frac{(\Gamma(\phi))^n}{\prod_{i=1}^m \prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \phi\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})} \phi\right) \right]} \prod_{i=1}^m \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \phi - 1} (1 - y_{ij})^{\frac{1}{1+\exp(\eta_{ij})} \phi - 1} \right) \right] \\ &\times \phi^{-1/2} \left(\frac{\phi^{1/2}}{a} \right)^\epsilon \left(1 - \frac{\phi^{1/2}}{a} \right)^\epsilon \end{aligned}$$

$$\begin{aligned}
f(\eta_i|\eta_k, \beta, \Sigma_b, \nu_b, \phi, y_i) &\propto \left(\prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(b_i|\beta, \Sigma_b, \nu_b) \quad (i \neq k; i, k = 1, \dots, m) \\
&\propto \left[\frac{1}{\prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})}\phi\right) \right]} \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi-1} (1-y_{ij})^{\frac{1}{1+\exp(\eta_{ij})}\phi-1} \right) \right] \\
&\times \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}}
\end{aligned}$$

Se observa entonces que las condicionales completas bajo estudio no son conjugadas, ni tienen una forma cerrada.

B.1. Distribución a posteriori bajo el modelo con prioris tradicionales

Tal como se comentó en esta sección, tradicionalmente el modelo con efectos mixtos es propuesto considerando efectos aleatorios Normales multivariados, con priori gamma inversa para el parámetro ϕ , priori wishart inversa para Σ_b y priori Normal multivariada para β . Bajo este escenario, se consiguen algunas condicionales completas conjugadas, aunque la posteriori bajo este escenario tampoco tiene una forma cerrada como veremos a continuación.

Considerando que $\beta \perp\!\!\!\perp \Sigma_b \perp\!\!\!\perp \phi$ se tiene

$$\begin{aligned}
f(\beta, \Sigma_b, \phi, \eta|y) &\propto \left[\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right] \left[\prod_{i=1}^m f(\eta_i|\beta, \Sigma_b) \right] f(\Sigma_b) f(\phi) f(\beta) \\
&\propto \left[\frac{(\Gamma(\phi))^n}{\prod_{i=1}^m \prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})}\phi\right) \right]} \prod_{i=1}^m \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi-1} (1-y_{ij})^{\frac{1}{1+\exp(\eta_{ij})}\phi-1} \right) \right] \\
&\times \left[(2\pi)^{-\frac{n}{2}} \left(\prod_{i=1}^m |Z_i \Sigma_b Z_i^t|^{-\frac{1}{2}} \right) \text{Exp} \left\{ -\frac{1}{2} \sum_{i=1}^m (\eta_i - X_i \beta)^t (Z_i \Sigma_b Z_i^t)^{-1} (\eta_i - X_i \beta) \right\} \right] \\
&\times \frac{|\psi|^{\frac{c}{2}} |\Sigma_b|^{-\frac{(c+q+1)}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr}(\psi \Sigma_b^{-1}) \right\}}{2^{\frac{cq}{2}} \Gamma_q\left(\frac{c}{2}\right)} \\
&\times \frac{b_o^{a_0}}{\Gamma(a_0)} \phi^{-(a_0+1)} \text{Exp} \left(-\frac{b_0}{\phi} \right) \times \left[(2\pi)^{-\frac{n}{2}} |\Sigma_\beta|^{-\frac{1}{2}} \text{Exp} \left\{ -\frac{1}{2} (\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\} \right]
\end{aligned}$$

Con $\eta_i|\beta, \Sigma_b \sim N_{n_i}(X_i\beta, Z_i\Sigma_b Z_i^t)$, $n = \sum_{i=1}^m n_i$ y $\Gamma_c(\bullet)$ función de densidad Gamma c-variada.

B.2. Condicionales completas bajo el modelo con prioris tradicionales

En el modelo con prioris tradicionales, las condicionales completas son

$$\begin{aligned}
f(\Sigma_b|\beta, \phi, \eta, y) &\propto \left(\prod_{i=1}^m f(\eta_i|\beta, \Sigma_b) \right) f(\Sigma_b) \\
&\propto \left[\left(\prod_{i=1}^m |Z_i \Sigma_b Z_i^t|^{-\frac{1}{2}} \right) \text{Exp} \left\{ -\frac{1}{2} \sum_{i=1}^m (\eta_i - X_i \beta)^t (Z_i \Sigma_b Z_i^t)^{-1} (\eta_i - X_i \beta) \right\} \right] \\
&\times |\Sigma_b|^{-\frac{(c+q+1)}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr}(\psi \Sigma_b^{-1}) \right\} \\
&\propto \left[|\Sigma_b|^{-\frac{m}{2}} \text{Exp} \left\{ -\frac{1}{2} \sum_{i=1}^m \text{tr} \left((Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta) \right)^t \Sigma_b^{-1} \left((Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta) \right) \right\} \right] \\
&\times |\Sigma_b|^{-\frac{(c+q+1)}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr}(\psi \Sigma_b^{-1}) \right\} \\
&\propto |\Sigma_b|^{-\frac{m+c+q+1}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr} \left[\left(\psi + \sum_{i=1}^m \left((Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta) \right) \left((Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta) \right)^t \right) \Sigma_b^{-1} \right] \right\}
\end{aligned}$$

De esta forma:

$$\Sigma_b | \beta, \phi, \eta, y \sim IW \left(\left(\sum_{i=1}^m (Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta) (\eta_i - X_i \beta)^t Z_i (Z_i^t Z_i)^{-1} \right) + \psi, m + c \right)$$

$$\begin{aligned} f(\beta | \Sigma_b, \phi, \eta, y) &\propto \left(\prod_{i=1}^m f(\eta_i | \beta, \Sigma_b) \right) f(\beta) \\ &\propto \left[\text{Exp} \left\{ -\frac{1}{2} \sum_{i=1}^m ((Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta))^t \Sigma_b^{-1} ((Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta)) \right\} \right] \\ &\times \text{Exp} \left\{ -\frac{1}{2} (\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\} \end{aligned}$$

$$\text{Sea } A_i = (Z_i^t Z_i)^{-1} Z_i^t, \quad W_i = A_i X_i \quad \text{y} \quad H = \left[\left(\sum_{i=1}^m W_i^t \Sigma_b^{-1} W_i \right) + \Sigma_\beta^{-1} \right]^{-1}$$

Entonces

$$\begin{aligned} f(\beta | \Sigma_b, \phi, \eta, y) &\propto \left[\text{Exp} \left\{ -\frac{1}{2} \sum_{i=1}^m (\eta_i^t - \beta^t X_i^t) A_i^t \Sigma_b^{-1} A_i (\eta_i - X_i \beta) \right\} \right] \\ &\times \text{Exp} \left\{ -\frac{1}{2} (\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta) \right\} \\ &\propto \left[\text{Exp} \left\{ -\frac{1}{2} \sum_{i=1}^m \beta^t W_i^t \Sigma_b^{-1} W_i \beta - 2\beta^t W_i^t \Sigma_b^{-1} A_i \eta_i \right\} \right] \times \text{Exp} \left\{ -\frac{1}{2} (\beta^t \Sigma_\beta^{-1} \beta - 2\beta^t \Sigma_\beta^{-1} \mu_\beta) \right\} \\ &\propto \text{Exp} \left\{ -\frac{1}{2} \left(\beta^t \left[\left(\sum_{i=1}^m W_i^t \Sigma_b^{-1} W_i \right) + \Sigma_\beta^{-1} \right] \beta - 2\beta^t \left[\left(\sum_{i=1}^m W_i^t \Sigma_b^{-1} A_i \eta_i \right) + \Sigma_\beta^{-1} \mu_\beta \right] \right) \right\} \end{aligned}$$

Luego

$$\beta | \Sigma_b, \phi, \eta, y \sim N \left(H \left[\left(\sum_{i=1}^m W_i^t \Sigma_b^{-1} A_i \eta_i \right) + \Sigma_\beta^{-1} \mu_\beta \right], H \right)$$

$$\begin{aligned}
f(\phi|\beta, \Sigma_b, \eta, y) &\propto \left(\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(\phi) \\
&\propto \left[\frac{(\Gamma(\phi))^n}{\prod_{i=1}^m \prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})}\phi\right) \right]} \prod_{i=1}^m \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi-1} (1-y_{ij})^{\frac{1}{1+\exp(\eta_{ij})}\phi-1} \right) \right] \\
&\times \phi^{-(a_0+1)} \text{Exp}\left(-\frac{b_0}{\phi}\right)
\end{aligned}$$

$$\begin{aligned}
f(\eta_i|\eta_k, \beta, \Sigma_b, \phi, y_i) &\propto \left(\prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \phi) \right) f(b_i|\beta, \Sigma_b) \quad (i \neq k; i, k = 1, \dots, m) \\
&\propto \left[\frac{1}{\prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})}\phi\right) \right]} \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi-1} (1-y_{ij})^{\frac{1}{1+\exp(\eta_{ij})}\phi-1} \right) \right] \\
&\times \left[\text{Exp}\left\{ -\frac{1}{2}(\eta_i - X_i\beta)^t A_i^t \Sigma_b^{-1} A_i (\eta_i - X_i\beta) \right\} \right]
\end{aligned}$$

Concluimos entonces que bajo el modelo con efectos mixtos, que considera distribución normal multivariada para los efectos aleatorios y prioris tradicionales, se consiguen formas cerradas para las condicionales completas de Σ_b y β , pero no para las condicionales completas de ϕ y η_i .

3.3.6. Códigos BUGS

C.1. Códigos BUGS para la aplicación del modelo beta mixto usando los datos simulados (datos balanceados)

Esta sección presenta variadas piezas del código BUGS usadas para ajustar el modelo de regresión beta mixto considerando la aplicación con datos simulados.

Priori gamma inversa para ϕ

```

model
{
  for( i in 1 : m ) {
    for( j in 1 : n ) {
      Y[i , j] ~ dbeta(a1[i,j] ,a2[i,j])
      a1[i,j] <- mu[i , j]*phi
      a2[i,j] <- (1-mu[i , j])*phi
      logit(mu[i , j]) <- inprod(x[i, j, ], beta[ ])+inprod(z[i, j, ], b[i,1,])
    }
    b[i,1,1:q] ~ dmt(cevec [ ] ,psi[ , ],g11)
  }
  g11~dexp(a0)
  beta[1:p] ~ dmt(alpha[ ] , V1[ , ],g12)
  V1[1:p ,1:p] <- inverse(V[ , ])
  psi[1:q,1:q] ~ dwish(R0[ , ], c0)
  psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
  phiinv ~ dgamma(a00,a00)
  phi<-1/phiinv
}

```

Priori uniforme para ϕ

```

model
{
  for( i in 1 : m ) {
    for( j in 1 : n ) {
      Y[i , j] ~ dbeta(a1[i,j] ,a2[i,j])
      a1[i,j] <- mu[i , j]*phi
      a2[i,j] <- (1-mu[i , j])*phi
      logit(mu[i , j]) <- inprod(x[i, j, ], beta[ ])+inprod(z[i, j, ], b[i,1,])
    }
    b[i,1,1:q] ~ dmt(cevec [ ] ,psi[ , ],g11)
  }
  g11~dexp(a0)
  beta[1:p] ~ dmt(alpha[ ] , V1[ , ],g12)
  V1[1:p ,1:p] <- inverse(V[ , ])
  psi[1:q,1:q] ~ dwish(R0[ , ], c0)
  psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])

  phir ~ dunif(a00,b01)
  phi<- phir*phir
}

```

Priori beta para ϕ

```

model
{
  for( i in 1 : m ) {
    for( j in 1 : n ) {
      Y[i , j] ~ dbeta(a1[i,j] ,a2[i,j])
      a1[i,j] <- mu[i , j]*phi
      a2[i,j] <- (1-mu[i , j])*phi
      logit(mu[i , j]) <- inprod(x[i, j, ], beta[ ])+inprod(z[i, j, ], b[i,1,])
    }
    b[i,1,1:q ] ~ dmt(cerovec [ ] ,psi[ , ],g11)
  }
  g11~dexp(a0)
  beta[1:p] ~ dmt(alpha[ ] , V1[ , ],g12)
  V1[1:p ,1:p] <- inverse(V[ , ])
  psi[1:q,1:q] ~ dwish(R0[ , ], c0)
  psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
  phiinicial ~ dbeta(a00,b0)
  phi<-phiinicial*b11
}

```

C.2. Códigos BUGS para la aplicación del modelo beta mixto usando los datos de Prater (datos desbalanceados)

Esta sección presenta variadas piezas del código BUGS usadas para ajustar el modelo de regresión beta mixto considerando la aplicación con los datos de petróleo de Prater.

Priori gamma inversa para ϕ

```

model
{
  for( j in 1 : m ) {
    for( i in (N[j]+1) : N[j+1] ) {
      Y[i] ~ dbeta(a1[i] ,a2[i])
      a1[i] <- mu[i]*phi
      a2[i] <- (1-mu[i])*phi
      logit(mu[i]) <- inprod(x[i, ], beta[ ])+inprod(z[i, ], b[j,1,])
    }
    b[j,1,1:q ] ~ dmt(cerovec [ ] ,psi[ , ],g11)
  }
  g11~dexp(a0)
}

```

```

beta[1:p] ~ dmt(alpha[ ], V1[ ],gl2)
V1[1:p,1:p] <- inverse(V[ ])
psi[1:q,1:q] ~ dwish(R0[ ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
phiinv ~ dgamma(a00,a00)
phi<-1/phiinv
for( j in 1 : m ) {
  Inter[j]<-beta[1]+b[j,1,1]
  EP[j]<-beta[2]+b[j,1,2]
}
}

```

Priori uniforme para ϕ

```

model
{
for( j in 1 : m ) {
for( i in (N[j]+1) : N[j+1] ) {
Y[i] ~ dbeta(a1[i] ,a2[i])
a1[i] <- mu[i]*phi
a2[i] <- (1-mu[i])*phi
logit(mu[i]) <- inprod(x[i, ], beta[ ])+inprod(z[i, ], b[j,1,])
}
b[j,1,1:q] ~ dmt(cerovec [ ],psi[ ],gl1)
}
gl1~dexp(a0)
beta[1:p] ~ dmt(alpha[ ], V1[ ],gl2)
V1[1:p,1:p] <- inverse(V[ ])
psi[1:q,1:q] ~ dwish(R0[ ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
phir ~ dunif(a00,b01)
phi<- phir*phir
for( j in 1 : m ) {
  Inter[j]<-beta[1]+b[j,1,1]
  EP[j]<-beta[2]+b[j,1,2]
}
}
}

```

Priori beta para ϕ

```

model
{
for( j in 1 : m ) {
for( i in (N[j]+1) : N[j+1] ) {
Y[i] ~ dbeta(a1[i] ,a2[i])

```

```

a1[i] <- mu[i]*phi
a2[i] <- (1-mu[i])*phi
logit(mu[i]) <- inprod(x[i, ], beta[ ])+inprod(z[i, ], b[j,1,])
}
b[j,1,1:q] ~ dmt(cevec [ ],psi[ , ],g11)
}
g11~dexp(a0)
beta[1:p] ~ dmt(alpha[ ], V1[ , ],g12)
V1[1:p ,1:p] <- inverse(V[ , ])
psi[1:q,1:q] ~ dwish(R0[ , ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
phiinicial ~ dbeta(a00,b0)
phi<-(b11*phiinicial)*(b11*phiinicial)
for( j in 1 : m ) {
  Inter[j]<-beta[1]+b[j,1,1]
  EP[j]<-beta[2]+b[j,1,2]
}
}

```

3.4. Modelo de regresión beta con submodelo de regresión asociado al parámetro de precisión

En esta sección consideraremos que cada variable de respuesta y_{ij} tiene asociado un parámetro de precisión ϕ_{ij} , al cual se le asocia una estructura de regresión, por lo que nos abocaremos a discutir qué priori es la mas apropiada para modelar los parámetros de regresión de este submodelo de precisión. Se discutirán las prioris utilizadas en la especificación del modelo de localización y luego se realizará el ajuste del modelo usando un software adecuado.

3.4.1. Discusión de prioris

Para una formulación mas general del modelo planteado, consideraremos diferentes parámetros de precisión, a los cuales los denotaremos por ϕ_{ij} , asociados a cada variable de respuesta y_{ij} . Asumiremos entonces un modelo lineal con efectos mixtos para el logaritmo de ϕ_{ij} como se plantea a continuación:

$$\ln(\phi_{ij}) = \tau_{ij} = w_{ij}^\top \delta + h_{ij}^\top d_i, \quad (3.4.1)$$

donde $w_{ij} = (w_{ij1}, \dots, w_{ijp^*})^\top$ es el vector de diseño correspondiente al $p^* \times 1$ vector de efectos fijos δ y $h_{ij} = (h_{ij1}, \dots, h_{ijq^*})^\top$ es el vector de diseño correspondiente al $q^* \times 1$ vector de efectos aleatorios d_i . Note que, las matrices de diseño $W_i = (w_{i1}, \dots, w_{in_i})^\top$ y $H_i = (h_{i1}, \dots, h_{in_i})^\top$ pueden, aunque no es un requisito, contener las mismas variables predictoras que las matrices $X_i = (x_{i1}, \dots, x_{in_i})^\top$ y $Z_i = (z_{i1}, \dots, z_{in_i})^\top$, respectivamente. En otras palabras, comúnmente la matriz W_i queda determinada por algunas columnas de X_i con $q^* \leq q$ y la matriz H_i queda determinada por algunas columnas de Z_i con $p^* \leq p$.

Al modelo descrito en (3.4.1), le llamaremos submodelo de precisión, el cual tiene la misma estructura que el modelo de localización dado en (3.2.2).

Por consiguiente, se especificarán las prioris para los parámetros δ y d_i 's de la misma forma en que fueron especificadas en el modelo de localización para los parámetros β y b_i 's respectivamente (ver sección 3.3.1).

3.4.2. Ajuste del modelo usando algoritmo de Monte Carlo vía cadenas de Markov

Sea $y^t = (y_1^t, \dots, y_m^t)$, $\eta^t = (\eta_1^t, \dots, \eta_m^t)$ y $\tau^t = (\tau_1^t, \dots, \tau_m^t)$. Donde condicionalmente en β , Σ_b y ν_b los η_i son independientes con $f(\eta_i|\beta, \Sigma_b, \nu_b) \propto f(b_i|\beta, \Sigma_b, \nu_b)$ y condicionalmente en δ , Σ_d y ν_d los τ_i son independientes con $f(\tau_i|\delta, \Sigma_d, \nu_d) \propto f(d_i|\delta, \Sigma_d, \nu_d)$ $i = 1, \dots, m$. Para obtener las estimaciones de los parámetros desconocidos del modelo, es necesario muestrear de la distribución a posteriori conjunta dada a continuación:

$$f(\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, \eta, \tau|y) \propto \left[\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \tau_{ij}) \right] \left[\prod_{i=1}^m f(\eta_i|\beta, \Sigma_b, \nu_b) \right] \times \left[\prod_{i=1}^m f(\tau_i|\delta, \Sigma_d, \nu_d) \right] f(\beta)f(\Sigma_b)f(\nu_b)f(\delta)f(\Sigma_d)f(\nu_d). \quad (3.4.2)$$

Para la especificación de (3.4.2) consideramos que $\beta \perp\!\!\!\perp \Sigma_b \perp\!\!\!\perp \nu_b \perp\!\!\!\perp \delta \perp\!\!\!\perp \Sigma_d \perp\!\!\!\perp \nu_d$.

Podemos entonces utilizar el muestreo de Gibbs para generar una muestra Monte Carlo de la posteriori conjunta $f(\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, \eta, \tau|y)$. El muestreo de Gibbs en este contexto, involucra un muestreo iterativo desde las distribuciones condicionales completas $f(\Sigma_b|\nu_b, \beta, \delta, \Sigma_d, \nu_d, \eta, \tau, y)$, $f(\nu_b|\Sigma_b, \beta, \delta, \Sigma_d, \nu_d, \eta, \tau, y)$, $f(\beta|\Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, \eta, \tau, y)$, $f(\delta|\beta, \Sigma_b, \nu_b, \Sigma_d, \nu_d, \eta, \tau, y)$, $f(\Sigma_d|\beta, \Sigma_b, \nu_b, \delta, \nu_d, \eta, \tau, y)$, $f(\nu_d|\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \eta, \tau, y)$, $f(\eta_i|\eta_k, \tau_k, \beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, y_i)$, $i, k = 1, \dots, m, i \neq k$,

$f(\tau_i|\tau_k, \eta_k, \beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, y_i)$, $i, k = 1, \dots, m$, $i \neq k$ las cuales pueden ser descom-
puestas como sigue

$$f(\Sigma_b|\nu_b, \beta, \delta, \Sigma_d, \nu_d, \eta, \tau, y) \propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\Sigma_b) \quad (3.4.3)$$

$$f(\nu_b|\Sigma_b, \beta, \delta, \Sigma_d, \nu_d, \eta, \tau, y) \propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\nu_b) \quad (3.4.4)$$

$$f(\beta|\Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, \eta, \tau, y) \propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\beta) \quad (3.4.5)$$

$$f(\delta|\beta, \Sigma_b, \nu_b, \Sigma_d, \nu_d, \eta, \tau, y) \propto \left(\prod_{i=1}^m f(d_i|\delta, \Sigma_d, \nu_d) \right) f(\delta) \quad (3.4.6)$$

$$f(\Sigma_d|\beta, \Sigma_b, \nu_b, \delta, \nu_d, \eta, \tau, y) \propto \left(\prod_{i=1}^m f(d_i|\delta, \Sigma_d, \nu_d) \right) f(\Sigma_d) \quad (3.4.7)$$

$$f(\nu_d|\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \eta, \tau, y) \propto \left(\prod_{i=1}^m f(d_i|\delta, \Sigma_d, \nu_d) \right) f(\nu_d) \quad (3.4.8)$$

$$f(\eta_i|\eta_k, \tau_k, \beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, y_i) \propto \left(\prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \tau_{ij}) \right) f(b_i|\beta, \Sigma_b, \nu_b) \quad (3.4.9)$$

($i \neq k; i, k = 1, \dots, m$)

$$f(\tau_i|\tau_k, \eta_k, \beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, y_i) \propto \left(\prod_{j=1}^{n_i} f(y_{ij}|\eta_{ij}, \tau_{ij}) \right) f(d_i|\delta, \Sigma_d, \nu_d) \quad (3.4.10)$$

($i \neq k; i, k = 1, \dots, m$)

De acuerdo a las ecuaciones expuestas en (3.4.3 - 3.4.10) podemos concluir que no se consiguen prioris conjugadas para ninguno de los parámetros desconocidos del modelo (ver apéndice).

No obstante lo anterior, el modelo puede ser implementado sin mayor dificultad en el software WinBUGS. Inferencias a posteriori en $\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d$ y, principalmente en las respuestas esperadas ($\mu_{ij}; i = 1, \dots, m, j = 1, \dots, n_i$) son fácilmente obtenidas en el software mencionado. Los test de hipótesis respecto a los coeficientes

de regresión y las respuestas esperadas son también fáciles de obtener.

3.4.3. Ejemplo con datos simulados

Para ilustrar la metodología propuesta, consideraremos el modelo de localización dado por

$$\eta_{ij} = \beta_1 + x_{ij2}\beta_2 + x_{ij3}\beta_3 + b_{i1} + z_{ij2}b_{i2},$$

$i = 1, \dots, 100$; $j = 1, \dots, 5$, donde $\eta_{ij} = \ln\{\mu_{ij}/(1 - \mu_{ij})\}$, $z_{ij1} = x_{ij1} = 1$, $z_{ij2} = x_{ij2}$. y los datos provenientes de la simulación realizada en (3.3.3).

Para realizar el ajuste del modelo y de acuerdo con lo expuesto en (3.4.1) se utilizarán las siguientes prioris para el modelo en cuestión:

$$\beta \sim t_3(\nu_\beta, \mu_\beta, \Sigma_\beta)$$

$$\nu_\beta = 20; \quad \mu_\beta = (0, 0, 0); \quad \Sigma_\beta = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

$$\delta \sim t_3(\nu_\delta, \mu_\delta, \Sigma_\delta)$$

$$\nu_\delta = 20; \quad \mu_\delta = (0, 0, 0); \quad \Sigma_\delta = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

$$b \sim t_2(\nu_b, 0, \Sigma_b)$$

$$\nu_b \sim \exp(0, 001); \quad \Sigma_b \sim IW(\psi, c); \quad \psi = \begin{pmatrix} 5 & 15 \\ 15 & 45 \end{pmatrix}; \quad c = 5.$$

$$d \sim t_2(\nu_d, 0, \Sigma_d)$$

$$\nu_d \sim \exp(0,001); \quad \Sigma_d \sim IW(\psi_d, c_d); \quad \psi_d = \begin{pmatrix} 5 & 15 \\ 15 & 45 \end{pmatrix}; \quad c_d = 5.$$

En el cuadro 3.6 se presentan 5 opciones diferentes para explicitar la estructura de regresión del parámetro de precisión y se considera un modelo de localización común a cada submodelo de precisión. El cuadro 3.7 reporta la "deviance information criterion" para los cinco modelos ajustados, con lo que observamos que el submodelo de precisión que no incluye efectos aleatorios logra un mejor ajuste de nuestros datos.

En este caso, los modelos 2a, 2c y 2d son similarmente buenos. Sin embargo, el modelo 2a muestra un ajuste levemente mejor que las otras propuestas.

Cuadro 3.6: *Propuesta de submodelos de precisión con un modelo de localización común*

<i>Model</i>	<i>Location Submodel</i>	<i>Precision Submodel</i> ($\ln(\phi_{ij})$)
Model 2a	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij2} + x_{ij3}\beta_3$	δ_1
Model 2b	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij2} + x_{ij3}\beta_3$	$\delta_1 + d_{i1}$
Model 2c	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij2} + x_{ij3}\beta_3$	$\delta_1 + \delta_3x_{ij3}$
Model 2d	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij2} + x_{ij3}\beta_3$	$\delta_1 + \delta_2x_{ij2} + \delta_3x_{ij3}$
Model 2e	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})x_{ij2} + x_{ij3}\beta_3$	$(\delta_1 + d_{i1}) + (\delta_2 + d_{i2})x_{ij2} + \delta_3x_{ij3}$

En el cuadro 3.8 se reportan los parámetros estimados del modelo 2a. Estos resultados muestran que los parámetros estimados mediante la metodología Bayesiana propuesta aquí, son similares a los reales valores de los parámetros del modelo.

Cuadro 3.7: *Análisis de sensibilidad para la especificación del submodelo de precisión*

<i>Model</i>	<i>DIC</i>
model 2a	-1244.850
model 2b	-1221.180
model 2c	-1242.940
model 2d	-1241.530
model 2e	-1223.740

Para realizar el ajuste del modelo, consideramos 100,000 iteraciones Monte Carlo y los resultados son presentados considerando las últimas 90,000 iteraciones. Adicionalmente, se han realizado los test de diagnóstico pertinentes, con lo que se han observado conductas deseables de las cadenas (ver figuras 3.6 y 3.7, y los comentarios hechos mas abajo).

Cuadro 3.8: *Medias a posteriori exactas. Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95 % para el modelo de regresión beta mixto simulado*

Parameter	Posterior Estimation				
	<i>True</i>	<i>Mean</i>	<i>MCError</i>	<i>Median</i>	<i>95 %CI</i>
β_1	-2	-2.063	0.003866	-2.064	(-2.284,-1.836)
β_2	1	1.032	0.002024	1.033	(0.8902,1.165)
β_3	2	1.947	0.001293	1.947	(1.828,2.064)
$\Sigma_{b_{11}}$	1.09	0.9755	0.003577	0.9612	(0.7043,1.339)
$\Sigma_{b_{12}}$	-0.36	-0.2577	0.001783	-0.2536	(-0.4206,-0.1265)
$\Sigma_{b_{22}}$	0.13	0.08049	7.268E-4	0.07752	(0.03626,0.148)

Para el modelo 2a, la versión multivariada del diagnóstico de convergencia de Gelman y Rubin's, indica con un "multivariate proportional scale reduction factor" (mprf) igual a 1.00, que la cadena es convergente ($\text{mprf} < 1.2$). Las Figuras 3.6 y 3.7 sugieren, respectivamente, que la cadena para cada parámetro es no correlacionada y estacionaria. Luego, la estimación especificada es adecuada.

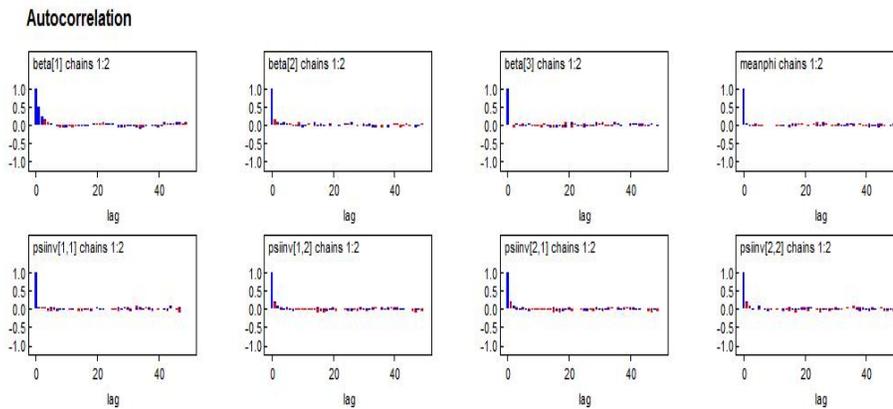


Figura 3.6: Autocorrelación - Datos simulados y estructura de regresión para ϕ

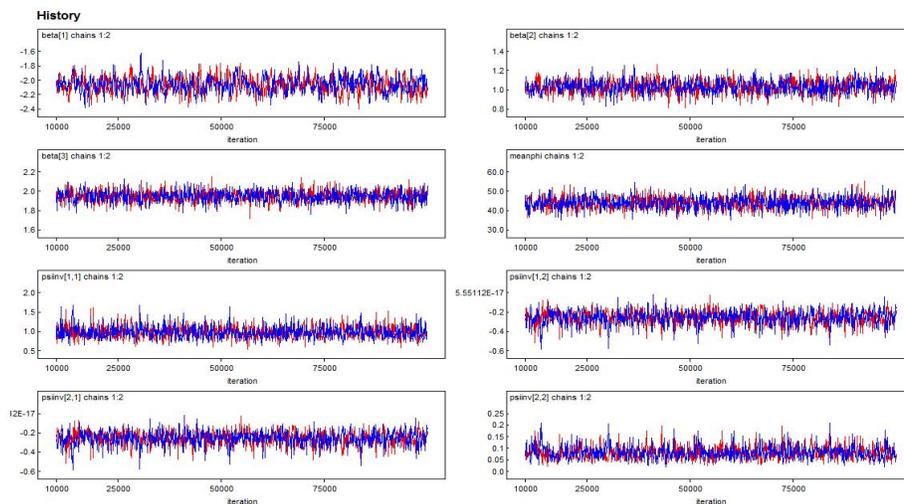


Figura 3.7: Historia - Datos simulados y estructura de regresión para ϕ

3.4.4. Aplicación del modelo a datos reales

Consideremos el ejemplo de los datos de petróleo de Prater descritos en la sección 3.3.4, donde el modelo de localización queda descrito por:

$$\ln \left\{ \frac{\mu_{ij}}{1 - \mu_{ij}} \right\} = \eta_{ij} = x_{ij}^\top \beta + z_{ij}^\top b_i,$$

donde $\mu_{ij} = E(y_{ij}^* | b_i)$, $x_{ij} = (1, EP_{ij})^\top$, $z_{ij} = (1, EP_{ij})^\top$, $\beta = (\beta_1, \beta_2)$ y $b_i = (b_{i1}, b_{i2})$

Además, para realizar el ajuste del modelo, consideramos los siguientes supuestos:

$$y_{ij} | b_i, \phi, \beta \sim \text{beta}(\mu_{ij}\phi, (1 - \mu_{ij})\phi),$$

$$\beta \sim t_2(\nu_\beta, \mu_\beta, \Sigma_\beta)$$

$$\nu_\beta = 10; \quad \mu_\beta = (0, 0); \quad \Sigma_\beta = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

$$\delta \sim t_2(\nu_\delta, \mu_\delta, \Sigma_\delta)$$

$$\nu_\delta = 10; \quad \mu_\delta = (0, 0); \quad \Sigma_\delta = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

$$b \sim t_2(\nu_b, 0, \Sigma_b)$$

$$\nu_b \sim \text{exp}(0, 001); \quad \Sigma_b \sim IW(\psi, c); \quad \psi = \begin{pmatrix} 1000 & 0 \\ 0 & 1000 \end{pmatrix}; \quad c = 4$$

$$d \sim t_2(\nu_d, 0, \Sigma_d)$$

$$\nu_d \sim \text{exp}(0, 001); \quad \Sigma_d \sim IW(\psi_d, c_d); \quad \psi_d = \begin{pmatrix} 1000 & 0 \\ 0 & 1000 \end{pmatrix}; \quad c_d = 4$$

En el cuadro 3.9 se entrega el ajuste del modelo considerando algunas variantes en la estructura del modelo de localización y en el submodelo de precisión. Aquí podemos observar que de acuerdo al criterio DIC, el modelo que mejor ajusta los datos es el

2.c que contiene solo el intercepto aleatorio tanto en el modelo de localización como en el submodelo de precisión.

Cuadro 3.9: *Propuesta de submodelos de precisión bajo diferentes modelos de localización*

<i>Model</i>	<i>Location Submodel</i>	<i>Precision Submodel</i> ($\ln(\phi_{ij})$)	<i>DIC</i>
Model 2.a.1	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})EP_{ij}$	$(\delta_1 + d_{i1}) + (\delta_2 + d_{i2})EP_{ij}$	-142,614
Model 2.a.2	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})EP_{ij}$	$(\delta_1 + d_{i1}) + \delta_2 EP_{ij}$	-131,504
Model 2.a.3	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})EP_{ij}$	$\delta_1 + (\delta_2 + d_{i2})EP_{ij}$	-142,767
Model 2.a.4	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})EP_{ij}$	$\delta_1 + \delta_2 EP_{ij}$	-131,389
Model 2.a.5	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})EP_{ij}$	δ_1	-121,505
Model 2.b	$\beta_1 + (\beta_2 + b_{i2})EP_{ij}$	$\delta_1 + (\delta_2 + d_{i2})EP_{ij}$	-142,488
Model 2.c	$(\beta_1 + b_{i1}) + \beta_2 EP_{ij}$	$(\delta_1 + d_{i1}) + \delta_2 EP_{ij}$	-145,858

En el cuadro 3.10 reportamos la estimación de los parámetros del modelo 2c, el cual presentó un mejor ajuste en comparación con los otros modelos propuestos. Para realizar el ajuste del modelo, consideramos 200,000 iteraciones Monte Carlo y los resultados son presentados considerando las últimas 190,000 iteraciones. Adicionalmente, se han realizado los test de diagnóstico pertinentes, con lo que se han observado conductas deseables de las cadenas (ver figuras 3.8, 3.9 y 3.10).

La versión multivariada del diagnóstico de convergencia de Gelman y Rubin's propuesta por Brooks & Gelman (1998) [4] indica que la cadena es convergente, luego que el "multivariate proportional scale reduction factor" (mprf) es igual a 1.08 (mprf < 1,2). Adicionalmente, las figuras 3.8 y 3.9 sugieren que la cadena para cada parámetro es no correlacionada y estacionaria respectivamente, la figura 3.10 nos muestra que la función de densidad a posteriori de cada parámetro no presenta bimodalidad, lo cual es un buen indicio puesto que la presencia de esta suele venir

Cuadro 3.10: *Medias y medianas a posteriori estimadas. Intervalos de credibilidad (CI) al 95% para el modelo de regresión beta mixto dado en 2.c*

Parameter	Posterior Inference in Location Submodel			
	<i>Mean</i>	<i>MCError</i>	<i>Median</i>	<i>95%CI</i>
$\beta_1 + b_{11}$	-4.232	0.01036	-4.233	(-4.703,-3.751)
$\beta_1 + b_{21}$	-4.574	0.009174	-4.579	(-5.032,-4.098)
$\beta_1 + b_{31}$	-4.323	0.00947	-4.331	(-4.78,-3.831)
$\beta_1 + b_{41}$	-4.801	0.01092	-4.8	(-5.28,-4.287)
$\beta_1 + b_{51}$	-4.696	0.01061	-4.695	(-5.197,-4.199)
$\beta_1 + b_{61}$	-4.816	0.01069	-4.816	(-5.286,-4.301)
$\beta_1 + b_{71}$	-5.214	0.01118	-5.212	(-5.731,-4.673)
$\beta_1 + b_{81}$	-5.259	0.01225	-5.258	(-5.806,-4.688)
$\beta_1 + b_{91}$	-5.45	0.01161	-5.447	(-5.969,-4.887)
$\beta_1 + b_{10\ 1}$	-5.799	0.01186	-5.8	(-6.332,-5.222)
β_2	0.01013	2.793E-5	0.01012	(0.008851,0.01139)

acompañada con problemas en la convergencia de la cadena. Estos resultados mencionados son esenciales para lograr una estimación adecuada de los parámetros del modelo, por lo que las estimaciones descritas en el cuadro 3.10 son confiables.

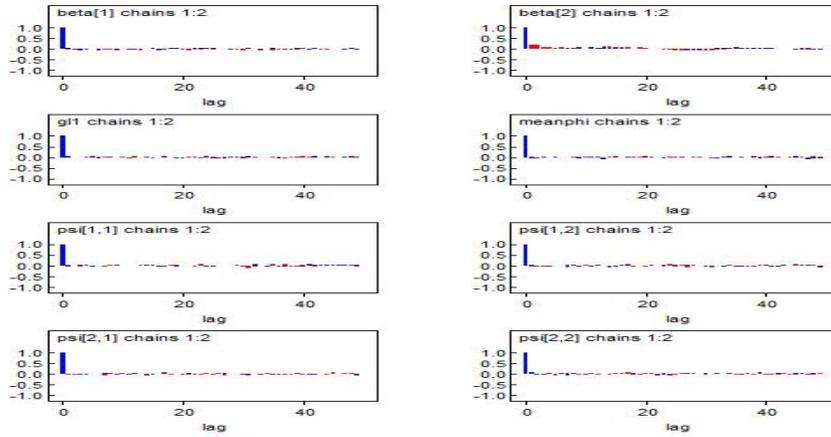


Figura 3.8: Autocorrelación - Datos de Prater y estructura de regresión para ϕ

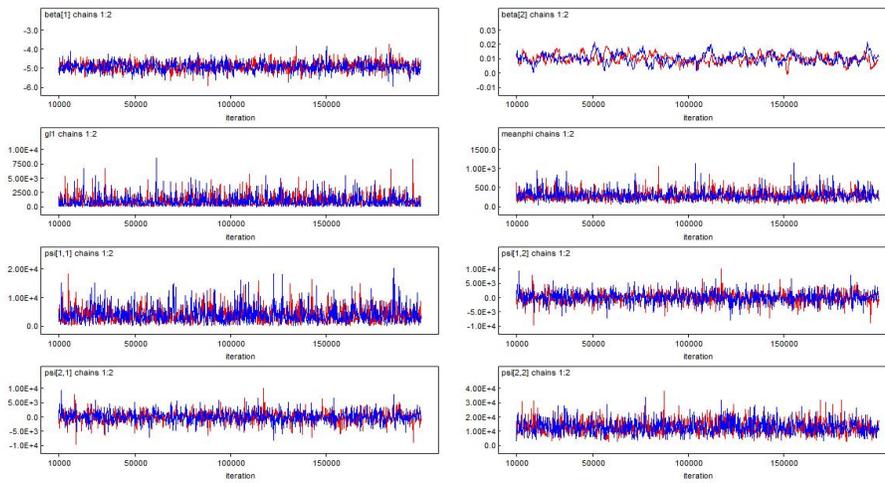


Figura 3.9: Historia - Datos de Prater y estructura de regresión para ϕ

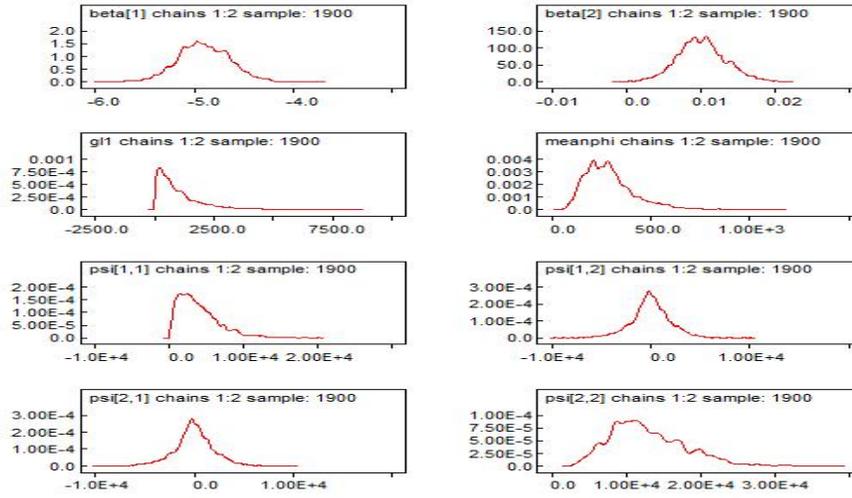


Figure 3.10: *Densidad - Datos de Prater y estructura de regresión para ϕ*

3.4.5. Apéndice

A.1. Distribución a posteriori

Note que de la ecuación (3.4.2) se obtiene

$$f(\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, \eta, \tau | y) \propto \left[\prod_{i=1}^m \prod_{j=1}^{n_i} f(y_{ij} | \eta_{ij}, \tau_{ij}) \right] \left[\prod_{i=1}^m f(\eta_i | \beta, \Sigma_b, \nu_b) \right] \times \\ \left[\prod_{i=1}^m f(\tau_i | \delta, \Sigma_d, \nu_d) \right] \times f(\Sigma_b) f(\nu_b) f(\delta) f(\Sigma_d) f(\nu_d) f(\beta)$$

$$\begin{aligned}
& \propto \left[\frac{(\Gamma(\phi_{ij}))^n}{\prod_{i=1}^m \prod_{j=1}^{n_i} \left[\Gamma\left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi_{ij}\right) \Gamma\left(\frac{1}{1+\exp(\eta_{ij})}\phi_{ij}\right) \right]} \right] \prod_{i=1}^m \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}\phi_{ij}-1} (1-y_{ij})^{\frac{1}{1+\exp(\eta_{ij})}\phi_{ij}-1} \right) \\
& \times \left[\frac{\Gamma(\frac{\nu_b+q}{2})}{\Gamma(\frac{\nu_b}{2})(\nu_b\pi)^{\frac{q}{2}}} \right]^m |\Sigma_b|^{-m/2} \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \\
& \times \left[\frac{\Gamma(\frac{\nu_d+q^*}{2})}{\Gamma(\frac{\nu_d}{2})(\nu_d\pi)^{\frac{q^*}{2}}} \right]^m |\Sigma_d|^{-m/2} \prod_{i=1}^m \left[1 + \frac{1}{\nu_d} (d_i^t \Sigma_d^{-1} d_i) \right]^{\frac{\nu_d+q^*}{2}} \\
& \times \frac{|\psi|^{c/2} |\Sigma_b|^{-\frac{(c+q+1)}{2}} \text{Exp}\left\{-\frac{1}{2}\text{tr}(\psi \Sigma_b^{-1})\right\}}{2^{\frac{cq}{2}} \Gamma_q\left(\frac{c}{2}\right)} \\
& \times \lambda_0 e^{-\lambda_0 \nu_b} \times \left[\frac{\Gamma(\frac{\nu_\delta+p^*}{2})}{\Gamma(\frac{\nu_\delta}{2})(\nu_\delta\pi)^{\frac{p^*}{2}}} \right] |\Sigma_\delta|^{-1/2} \left[1 + \frac{1}{\nu_\delta} ((\delta - \mu_\delta)^t \Sigma_\delta^{-1} (\delta - \mu_\delta)) \right]^{\frac{\nu_\delta+p^*}{2}} \\
& \times \frac{|\psi_d|^{c/2} |\Sigma_d|^{-\frac{(c_d+q^*+1)}{2}} \text{Exp}\left\{-\frac{1}{2}\text{tr}(\psi_d \Sigma_d^{-1})\right\}}{2^{\frac{c_d q^*}{2}} \Gamma_{q^*}\left(\frac{c}{2}\right)} \\
& \times \lambda_{d0} e^{-\lambda_{d0} \nu_d} \times \left[\frac{\Gamma(\frac{\nu_\beta+p}{2})}{\Gamma(\frac{\nu_\beta}{2})(\nu_\beta\pi)^{\frac{p}{2}}} \right] |\Sigma_\beta|^{-1/2} \left[1 + \frac{1}{\nu_\beta} ((\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta)) \right]^{\frac{\nu_\beta+p}{2}}
\end{aligned}$$

Donde $\ln(\phi_{ij}) = \tau_{ij} = w_{ij}^t \delta + h_{ij}^t d_i$, $b_i = (Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta)$, $d_i = (H_i^t H_i)^{-1} H_i^t (\tau_i - W_i \delta)$, $n = \sum_{i=1}^m n_i$ y $\Gamma_c(\bullet)$ es la función de densidad Gamma c-variada.

A.2. Condicionales completas

De las ecuaciones (3.4.3) a la (3.4.10) se obtienen las siguientes condicionales completas

$$\begin{aligned}
f(\Sigma_b | \nu_b, \beta, \delta, \Sigma_d, \nu_d, \eta, \tau, y) & \propto \left(\prod_{i=1}^m f(b_i | \beta, \Sigma_b, \nu_b) \right) f(\Sigma_b) \\
& \propto \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \times |\Sigma_b|^{-\frac{(m+c+q+1)}{2}} \text{Exp}\left\{-\frac{1}{2}\text{tr}(\psi \Sigma_b^{-1})\right\}
\end{aligned}$$

$$\begin{aligned}
f(\nu_b|\Sigma_b, \beta, \delta, \Sigma_d, \nu_d, \eta, \tau, y) &\propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\nu_b) \\
&\propto \left[\frac{\Gamma(\frac{\nu_b+q}{2})}{\Gamma(\frac{\nu_b}{2})(\nu_b\pi)^{\frac{q}{2}}} \right]^m \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \times e^{-\lambda_0 \nu_b}
\end{aligned}$$

$$\begin{aligned}
f(\beta|\Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, \eta, \tau, y) &\propto \left(\prod_{i=1}^m f(b_i|\beta, \Sigma_b, \nu_b) \right) f(\beta) \\
&\propto \prod_{i=1}^m \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b+q}{2}} \times \left[1 + \frac{1}{\nu_\beta} ((\beta - \mu_\beta)^t \Sigma_\beta^{-1} (\beta - \mu_\beta)) \right]^{\frac{\nu_\beta+p}{2}}
\end{aligned}$$

$$\begin{aligned}
f(\delta|\beta, \Sigma_b, \nu_b, \Sigma_d, \nu_d, \eta, \tau, y) &\propto \left(\prod_{i=1}^m f(d_i|\delta, \Sigma_d, \nu_d) \right) f(\delta) \\
&\propto \prod_{i=1}^m \left[1 + \frac{1}{\nu_d} (d_i^t \Sigma_d^{-1} d_i) \right]^{\frac{\nu_d+q^*}{2}} \times \left[1 + \frac{1}{\nu_\delta} ((\delta - \mu_\delta)^t \Sigma_\delta^{-1} (\delta - \mu_\delta)) \right]^{\frac{\nu_\delta+p^*}{2}}
\end{aligned}$$

$$\begin{aligned}
f(\Sigma_d|\beta, \Sigma_b, \nu_b, \delta, \nu_d, \eta, \tau, y) &\propto \left(\prod_{i=1}^m f(d_i|\delta, \Sigma_d, \nu_d) \right) f(\Sigma_d) \\
&\propto \prod_{i=1}^m \left[1 + \frac{1}{\nu_d} (d_i^t \Sigma_d^{-1} d_i) \right]^{\frac{\nu_d+q^*}{2}} \times |\Sigma_d|^{-\frac{(m+c_d+q^*+1)}{2}} \text{Exp} \left\{ -\frac{1}{2} \text{tr}(\psi_d \Sigma_d^{-1}) \right\}
\end{aligned}$$

$$\begin{aligned}
f(\nu_d|\beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \eta, \tau, y) &\propto \left(\prod_{i=1}^m f(d_i|\delta, \Sigma_d, \nu_d) \right) f(\nu_d) \\
&\propto \left[\frac{\Gamma(\frac{\nu_d+q^*}{2})}{\Gamma(\frac{\nu_d}{2})(\nu_d\pi)^{\frac{q^*}{2}}} \right]^m \prod_{i=1}^m \left[1 + \frac{1}{\nu_d} (d_i^t \Sigma_d^{-1} d_i) \right]^{\frac{\nu_d+q^*}{2}} \times e^{-\lambda_{d0} \nu_d}
\end{aligned}$$

$$\begin{aligned}
f(\eta_i | \eta_k, \tau_k, \beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, y_i) &\propto \left(\prod_{j=1}^{n_i} f(y_{ij} | \eta_{ij}, \tau_{ij}) \right) f(b_i | \beta, \Sigma_b, \nu_b) \\
&\propto \left[\frac{1}{\prod_{j=1}^{n_i} \left[\Gamma \left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \phi_{ij} \right) \Gamma \left(\frac{1}{1+\exp(\eta_{ij})} \phi_{ij} \right) \right]} \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \phi_{ij} - 1} (1 - y_{ij})^{\frac{1}{1+\exp(\eta_{ij})} \phi_{ij} - 1} \right) \right] \\
&\times \left[1 + \frac{1}{\nu_b} (b_i^t \Sigma_b^{-1} b_i) \right]^{\frac{\nu_b + q}{2}} \\
&(i \neq k; i, k = 1, \dots, m)
\end{aligned}$$

$$\begin{aligned}
f(\tau_i | \tau_k, \eta_k, \beta, \Sigma_b, \nu_b, \delta, \Sigma_d, \nu_d, y_i) &\propto \left(\prod_{j=1}^{n_i} f(y_{ij} | \eta_{ij}, \tau_{ij}) \right) f(d_i | \delta, \Sigma_d, \nu_d) \\
&\propto \left[\frac{1}{\prod_{j=1}^{n_i} \left[\Gamma \left(\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \phi_{ij} \right) \Gamma \left(\frac{1}{1+\exp(\eta_{ij})} \phi_{ij} \right) \right]} \prod_{j=1}^{n_i} \left(y_{ij}^{\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} \phi_{ij} - 1} (1 - y_{ij})^{\frac{1}{1+\exp(\eta_{ij})} \phi_{ij} - 1} \right) \right] \\
&\times \left[1 + \frac{1}{\nu_d} (d_i^t \Sigma_d^{-1} d_i) \right]^{\frac{\nu_d + q^*}{2}} \\
&(i \neq k; i, k = 1, \dots, m)
\end{aligned}$$

Con $\ln(\phi_{ij}) = \tau_{ij} = w_{ij}^t \delta + h_{ij}^t d_i$, $b_i = (Z_i^t Z_i)^{-1} Z_i^t (\eta_i - X_i \beta)$, $d_i = (H_i^t H_i)^{-1} H_i^t (\tau_i - W_i \delta)$, $n = \sum_{i=1}^m n_i$ y $\Gamma_c(\bullet)$ es la función de densidad Gamma c-variada.

Se observa entonces que las condicionales completas bajo estudio no son conju-
gadas, ni tienen una forma cerrada.

3.4.6. Códigos BUGS

B.1. Códigos BUGS para la aplicación del modelo beta mixto usando los datos simulados (datos balanceados)

Esta sección presenta una pieza del código BUGS usada para ajustar el modelo de regresión beta mixto considerando la aplicación con datos simulados (datos balanceados).

Submodelo para el parámetro ϕ

```

model
{
  for( i in 1 : m ) {
    for( j in 1 : n ) {
      Y[i , j] ~ dbeta(a1[i,j] ,a2[i,j])
      a1[i,j] <- mu[i , j]*phi[i,j]
      a2[i,j] <- (1-mu[i , j])*phi[i,j]
      log(phi[i,j])<-inprod(x[i, j, ], delta[ ])+inprod(z[i, j, ], gama[i,1,])
      logit(mu[i , j]) <- inprod(x[i, j, ], beta[ ])+inprod(z[i, j, ], b[i,1,])
    }
    b[i,1,1:q ] ~ dmt(cerovec [ ] ,psi[ , ],g11)
    gama[i,1,1:q ] ~ dmt(cerovec [ ] ,psi[ , ],g11)
  }
  gl1~dexp(a0)
  beta[1:p] ~ dmt(alpha[ ] , V1[ , ],g12)
  delta[1:p] ~ dmt(alpha[ ] , V1[ , ],g12)
  V1[1:p ,1:p] <- inverse(V[ , ])
  psi[1:q,1:q] ~ dwish(R0[ , ], c0)
  psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
  meanphi<-mean(phi[,])
}

```

B.2. Códigos BUGS para la aplicación del modelo beta mixto usando los datos de Prater (datos desbalanceados)

Esta sección presenta una pieza del código BUGS usada para ajustar el modelo de regresión beta mixto considerando la aplicación con los datos de petróleo de Prater (datos desbalanceados).

Submodelo para el parámetro ϕ

```

model
{
  for( j in 1 : m ) {
    for( i in (N[j]+1) : N[j+1] ) {
      Y[i] ~ dbeta(a1[i] ,a2[i])
      a1[i] <- mu[i]*phi[i]
    }
  }
}

```

```

a2[i] <- (1-mu[i])*phi[i]
log(phi[i])<-inprod(x[i, ], delta[ ])+inprod(z[i, ], gama[j,1,])
logit(mu[i]) <- inprod(x[i, ], beta[ ])+inprod(z[i, ], b[j,1,])
}
b[j,1,1:q] ~ dmt(cevec [ ],psi[ , ],g11)
gama[j,1,1:q] ~ dmt(cevec [ ],psi[ , ],g11)
}
g11~dexp(a0)
beta[1:p] ~ dmt(alpha[ ], V1[ , ],g12)
delta[1:p] ~ dmt(alpha[ ], V1[ , ],g12)
V1[1:p ,1:p] <- inverse(V[ , ])
psi[1:q,1:q] ~ dwish(R0[ , ], c0)
psiinv[1:q,1:q]<-inverse(psi[1:q,1:q])
meanphi<-mean(phi[])
for( j in 1 : m ) {
  Inter[j]<-beta[1]+b[j,1,1]
  EP[j]<-beta[2]+b[j,1,2]
}
}

```

3.5. Resumen y conclusiones

3.6. Discusión

The beta regression model has gained increasing popularity from the reparameterization of the beta density in terms of its mean and a precision parameter [18], but mainly after the work of [7], in which this reparameterization was applied to a beta regression model with constant precision. A complementary approach proposed by [23] considers also the modeling the precision parameter. A Bayesian beta regression model was studied by [3]. In this paper, we extended these ideas for the mixed beta regression model under a Bayesian approach.

The present paper consider a Bayesian approach for the analysis of mixed beta regression based on two different criteria. First, the specification of different priors for the precision parameter ϕ is studied. Second, this parameter is modeled with a its

own regression structure. In the first case, we considered a uniform to squared prior for ϕ , i.e., a prior of the type $\phi = \mathbf{U}^2$, with $\mathbf{U} \sim U(0, a)$, and where it is common to consider a initial value $a = 50$ ([11]). We also considered an alternative to this prior, namely, a prior distribution of the type $\phi = (a * B)^2$ with $B \sim \text{beta}(1 + \epsilon, 1 + \epsilon)$ and ($\epsilon = 0,001, 0,01, 0,1, 0,5, \dots$), which delivered good results in terms of the model fit and the performance of the diagnostics tests. In the second case, we considered the same structure for $(\ln(\phi_{ij}))$ that for $\text{logit}(\mu_{ij})$, i.e, a mixed model with a multivariate t -prior for the fixed and random effects, under which good results were obtained in terms of the model fit and the performance of the diagnostics tests. It is worth mentioning that in this context it is necessary perform a model selection for ϕ_{ij} , including more or fewer fixed and random effects, since it is not clear in advance, the criteria for resolving the selection of which model is more efficient [**Cuando se plantea el submodelo de precisión, es necesario plantear y testear las diferentes opciones en su estructura (como en la tabla 4), no existiendo un criterio claro a priori para decidir que estructura es la mas adecuada**].

A classic version of this problem was raised by [27], where mixed beta regression models were estimated using the SAS procedure NLMIXED ([21]), employing adaptive Gaussian quadrature. This approach achieves good results, but the implementation of the mixed beta regression model for random effects that are non-normally distributed is very challenging. In this sense, own approach is more flexible, because one can be easily implement it when the distribution of the random effects is normal, Student- t , skew normal or other distributions, by using simple and accessible software such as WinBUGS. Other advantage of this approach is the easy implementation for the imputation of missing data [5], a common situation in practice and for which a

classic approach is much more complicated.

Bibliografía

- [1] James Albert and Siddhartha Chib, *Bayesian Analysis of Binary and Polychotomous Response Data*, Journal of the American Statistical Association **88** (1993), 669–679.
- [2] J. Besag, P. Green, D. Higdon, and K. Mengersen, *Bayesian computation and stochastic systems*, Statistical Science **10** (1995), 3–66.
- [3] Branscum, Johnson, and Thurmond, *Bayesian Beta Regression: Applications to Household expenditure data and genetic distance between foot-and-mouth disease viruses*, Australian and New Zealand Journal of Statistics **49** (2007), 287–301.
- [4] S.P. Brooks and A. Gelman, *General methods for monitoring convergence of iterative simulations*, Journal of Computational and Graphical Statistics **7** (1998), 434–455.
- [5] Gretchen Carrigan, Adrian Barnett, Annette Dobson, and Gita Mishra, *Compensating for missing data from longitudinal studies using winbugs*, Journal of Statistical Software **19** (2007), issue 7.
- [6] Dey, Ghosh, and Mallick, *Generalized linear models: A bayesian perspective*, first ed., Marcel Dekker, New York, 2000.
- [7] Silvia Ferrari and Francisco Cribari-Neto, *Beta Regression For Modelling Rates And Proportions*, Journal of Applied Statistics **31** (2004), 799–815.

- [8] Youyi Fong, Hvard Rue, and Jon Wakefield, *Bayesian inference for generalized linear mixed models*, *Biostatistics* **11** (2010), 397 – 412.
- [9] Dani Gamerman, *Markov chain monte carlo: Stochastic simulation for bayesian inference*, Chapman and Hall, 2006.
- [10] Gelman and Rubin, *Inference from iterative simulation using multiple sequences*, *Statistical Science* **7** (1992), 457–511.
- [11] A. Gelman, *Prior distributions for variance parameters in hierarchical models*, *Bayesian Analysis* **1** (2006), 515–533.
- [12] J. Geweke, *Bayesian treatment of the independent Student-t linear model*, *Journal of Applied Econometrics* **S8** (1993), 19–40.
- [13] John Geweke, *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*, In *Bayesian Statistics 4* (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith) (1992), 169–193.
- [14] A. K. Gupta and S. Nadarajah, *Handbook of Beta Distribution and Its Applications.*, Marcel Dekker, New York, 2004.
- [15] P. Heidelberger and P. Welch, *A spectral method for confidence interval generation and run length control in simulations*, *Communications of The ACM* **24** (1981), 233–245.
- [16] Philip Heidelberger and Peter Welch, *Simulation run length control in the presence of an initial transient*, *Operations Research* **31 - 6** (1983), 1109–1144.
- [17] McMillan and Creelman, *Detection Theory A User Guide*, first ed., Lawrence Erlbaum Associates, Mahwah, NJ, 2005.
- [18] P. Paolino, *Maximum likelihood estimation of models with beta-distributed dependent variables*, *Political Analysis* **9** (2001), 325–346.

- [19] Everson Philip and Morris Carl, *Inference for multivariate normal hierarchical models*, Royal Statistical Society **62** (2000), no. Part 2, 399 – 412.
- [20] Martyn Plummer, Nicky Best, Kate Cowles, and Karen Vines, *The coda package*, R Project, 2006.
- [21] SAS, *Sas/stat 9.2 users guide*, Cary, NC, 2008.
- [22] R.S. Silva and H. F. Lopes, *Copula, marginal distributions and model selection: A bayesian note*, Statistics and Computing **18** (2008), 313–320.
- [23] Smithson and Verkuilen, *A Better Lemon Squeezer?: Maximum-Likelihood Regression With Beta-Distributed Dependent Variables*, Psychological Methods **11** (2006), 54–71.
- [24] G. Smyth, *Generalized Linear Models with Varying Dispersion*, Journal of the Royal Statistical Society **51** (1989), 47–60.
- [25] D.J. Spiegelhalter, N.G. Best, B.P Carlin, and A. Linde, *Bayesian measures of model complexity and fit*, Journal of the Royal Statistical Society series B **64** (2002), 583–639.
- [26] W. N. Venables, *Exegeses on linear models*, S-PLUS users Conference (2000).
- [27] Zimprich, *Modeling Change in Skewed Variables Using Mixed Beta Regression Models*, Research in Human Development **7** (2010), 9–26.