



# COLEGIO DE POSTGRADUADOS

---

---

Institución de Enseñanza e investigación  
en Ciencias Agrícolas

**CAMPUS MONTECILLO**

Socioeconomía, Estadística e Informática  
Estadística

## **Modelación Multivariada de la Pobreza Multidimensional**

**Alejandro Guzmán Rodríguez**

**T E S I S**

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS**

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2016

---

---



La presente tesis titulada: **Modelación Multivariada de la Pobreza Multidimensional**, realizada por el alumno: **Alejandro Guzmán Rodríguez**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

## MAESTRO EN CIENCIAS

### SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

#### CONSEJO PARTICULAR


CONSEJERO

  
Dr. Ciro Velasco Cruz

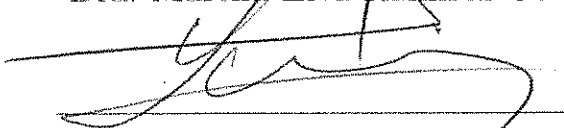
ASESOR

  
Dr. Sergio Pérez Elizalde

ASESOR

  
Dra. Martha Elva Ramírez Guzmán

ASESOR

  
Dr. Arturo Yamasaky Cruz

Montecillo, Texcoco, México, Mayo de 2016.



# MODELACIÓN MULTIVARIADA DE LA POBREZA MULTIDIMENSIONAL

Alejandro Guzmán Rodríguez, Mtro.  
Colegio de Postgraduados, 2016

A partir de un problema operativo sobre indicadores de pobreza en México, y una interpretación estadística de éste; se propone un método Bayesiano de modelación de variables binarias multivariadas. Se desarrolla un método MCMC basado en un modelo probit utilizando variables latentes; se recurre a una transformación para obtener un modelo jerárquico independiente que evita utilizar el algoritmo Metropolis, simplificando la simulación de datos y reduciendo el costo computacional.

**Palabras claves:** Modelo Probit Multivariado, Variables latentes, MCMC, Modelos jerárquicos, Pobreza multidimensional.

## MULTIVARIATE MODELLING OF MULTIDIMENSIONAL POVERTY

Alejandro Guzmán Rodríguez, M.S.  
Colegio de Postgraduados, 2016

Starting from an operative level problem about poverty indicators in Mexico, and a statistical interpretation of this problem, a Bayesian method for modeling multivariate binary variables is proposed. A MCMC with data augmentation, a method for a Multivariate Probit Model is developed; a transformation is used to apply a hierarchically independent model, this allow us to avoid the use of Metropolis algorithm, therefore simplifying the data simulation and reducing computing time.

**Key words:** Multivariate Probit Model, Data augmentation, MCMC, Hierarchical modeling, Multidimensional poverty.

# 武士道

義  
勇  
仁  
礼  
誠  
名  
誉  
忠  
義

## AGRADECIMIENTOS

Gracias a:

Cinthia,

mi mamá,

Ciro,

mis profesores,

el Colegio de Postgraduados,

y el CONACYT.



# Índice general

<b>1.</b>	<b>1</b>
1.1. Pobreza Binaria Multivariada . . . . .	2
1.2. Estado del Arte . . . . .	8
1.3. Problematización de la PM . . . . .	18
<b>2.</b>	<b>21</b>
2.1. Marco teórico . . . . .	21
2.2. Modelo Probit Multivariado . . . . .	27
2.3. Validación del Modelo . . . . .	32
<b>3.</b>	<b>49</b>
3.1. Modelación de la PM . . . . .	49
3.2. Conclusiones . . . . .	53
3.3. Carta sociológica . . . . .	56
<b>A.</b>	<b>61</b>
A.1. Desarrollo algebraico extendido . . . . .	61
<b>B.</b>	<b>71</b>
B.1. Detalles del Algoritmo . . . . .	71
<b>C.</b>	<b>85</b>
C.1. Resultados del ajuste de la PM . . . . .	85



# Capítulo 1

Durante 2013, antes de iniciar la maestría trabajaba en la delegación Querétaro de la Secretaría de Desarrollo Social (SEDESOL), en un puesto de análisis; una de las solicitudes específicas que me hizo el entonces delegado Manuel Pozo fue: ¿Dónde están los pobres?, ¿a dónde hay que llevar los programas de la secretaría para que éstos tengan impacto en los indicadores? El Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), es quien a partir de la definición oficial de pobreza mide la pobreza a nivel nacional y estatal cada dos años, mediante la Encuesta de Ingresos y Gastos en los Hogares (ENIGH) y el Módulo de Condiciones Socioeconómicas (MCS), que elabora y levanta el Instituto Nacional de Estadística y Geografía (INEGI). El problema es que la representatividad del diseño muestral y los resultados que presenta el CONEVAL se encuentran a nivel estatal; en aquel momento sabíamos que aproximadamente habitaban en Querétaro 707 mil personas en pobreza (multidimensional), pero a nivel operativo para la secretaría esta información era insuficiente: ¿Quiénes eran? ¿dónde estaban? ¿por qué estaban en situación de pobreza? Entonces fui incapaz de responder estas preguntas, lo que motivó el presente trabajo.

La motivación es operativa, pero la interpretación corresponde a las ciencias sociales. Cuando presenté el tema de investigación lo hice de la manera como comúnmente se hacen las solicitudes a las áreas de estadística: tengo estos datos, quiero saber esto, por qué no lo metes en el paquete y me das un p-valor. Sin embargo, la respuesta no es tan sencilla y requirió más de lo que pensaba.

## 1.1. Pobreza Binaria Multivariada

Las transiciones semánticas y por tanto léxicas entre disciplinas no son sencillas, este fue uno de los retos en el programa. Se replica esta transición a partir del concepto de Pobreza Multidimensional (PM) en este apartado.

### 1.1.1. Pobreza Multidimensional

La pobreza multidimensional consiste en una forma relativamente reciente de medir la pobreza en México. Es importante aproximarse a la PM porque es la definición oficial de pobreza en el país. A partir de esta definición y sus mediciones es como los gobiernos en México deberían de orientar las acciones públicas referentes a la pobreza.

El Comité Directivo del CONEVAL aprobó el día 26 de febrero de 2010 los lineamientos y criterios generales para la definición, identificación y medición de la pobreza, publicados en el Diario Oficial de la Federación (2010). De manera resumida presentamos en qué consiste la definición de PM, misma que considera las condiciones de vida de la población a partir de tres espacios, a saber:

**Bienestar económico:** se refiere a las necesidades asociadas a los bienes y servicios que puede adquirir la población mediante el ingreso.

**Derechos sociales:** trata de las carencias de la población en el ejercicio de sus derechos para el desarrollo social; en específico aquellas carencias asociadas a los indicadores mencionados en el artículo 36, fracciones II a la VII, de la Ley.

**Contexto territorial:** trata los aspectos que trascienden al ámbito individual (que pueden referirse a características geográficas, sociales y culturales, entre otras); en específico, aspectos asociados al grado de cohesión social, así como otros considerados relevantes para el desarrollo social.

La población en situación de PM será aquella cuyos ingresos sean insuficientes para adquirir los bienes y los servicios requeridos para satisfacer sus necesidades, y además presente carencia en al menos uno de los siguientes seis indicadores:

1. Rezago educativo

Figura 1.1: Dimensiones de la pobreza (CONEVAL, 2015)



- a) Población de 3 a 15 años que no cuenta con la educación secundaria terminada o no asiste a la escuela.
- b) Población nacida hasta 1981 que no cuenta con la educación primaria terminada.
- c) Población nacida a partir de 1982 que no cuenta con la educación secundaria terminada.

## 2. Acceso a los servicios de salud

- a) Población no afiliada o inscrita para recibir servicios de salud por parte del Seguro Popular, y que tampoco es derechohabiente de alguna institución de seguridad social.

## 3. Acceso a la seguridad social

- a) Población ocupada y asalariada que no recibe por parte de su trabajo las prestaciones de servicios médicos, incapacidad con goce de sueldo y SAR o Afore.
- b) Población ocupada y no asalariada que no recibe como prestación laboral o por contratación propia servicios médicos por parte de una institución pública de Seguridad Social y SAR o Afore.
- c) Población sin acceso a la seguridad social por alguno de los primeros dos criterios, que no goce de alguna jubilación o pensión, ni sea familiar directo de una persona dentro o fuera del hogar con acceso a la seguridad social.

- d)* Población de 65 años o más que no dispone de acceso a la seguridad social por alguno de los criterios anteriores ni es beneficiario de algún programa social de pensiones para adultos mayores.

#### 4. Calidad y espacios de la vivienda.

La población que resida en viviendas con al menos una de las siguientes características:

- a)* El material de la mayor parte de los pisos es tierra.
- b)* El material de la mayor parte del techo es lámina de cartón o desechos.
- c)* El material de la mayor parte de los muros es barro o bajareque; carrizo, bambú o palma; lámina de cartón, metálica o de asbesto; o material de desecho.
- d)* Con hacinamiento.

#### 5. Servicios básicos en la vivienda

La población que resida en viviendas con al menos una de las siguientes características:

- a)* El agua se obtiene de un pozo, río, lago, arroyo, pipa, o bien, el agua entubada la obtienen por acarreo de otra vivienda, o de la llave pública o hidrante.
- b)* No disponen de drenaje, o el desagüe tiene conexión a una tubería que va a dar a un río, lago, mar, barranca o grieta.
- c)* No disponen de energía eléctrica.

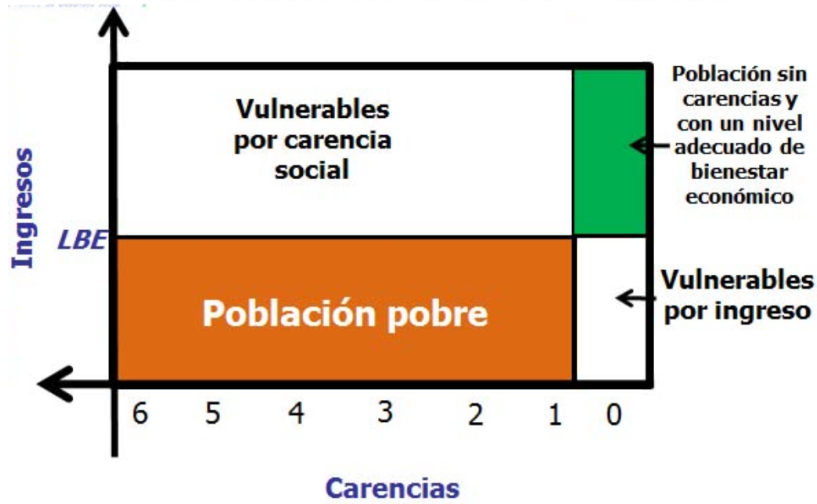
#### 6. Acceso a la alimentación.

- a)* Población en hogares con un grado de inseguridad alimentaria moderado o severo.

Para determinar si el ingreso de las personas es suficiente se consideran las líneas de bienestar:

**Línea del bienestar mínimo (LBM):** equivale al valor de la canasta alimentaria por persona al mes.

Figura 1.2: Identificación de la pobreza (CONEVAL, 2015)



**Línea de bienestar (LBE):** equivale al valor total de la canasta alimentaria y de la canasta no alimentaria por persona al mes.

Si el ingreso promedio en el hogar es menor a éstas líneas, se considera que las personas en el hogar no cuentan con el ingreso suficiente para satisfacer sus necesidades.

CONEVAL clasifica a partir de estos indicadores a las personas en 5 clases (ver Figura 1.2):

**Sin carencias y con un nivel adecuado de bienestar económico** Personas sin carencias sociales e ingreso superior a la línea de bienestar.

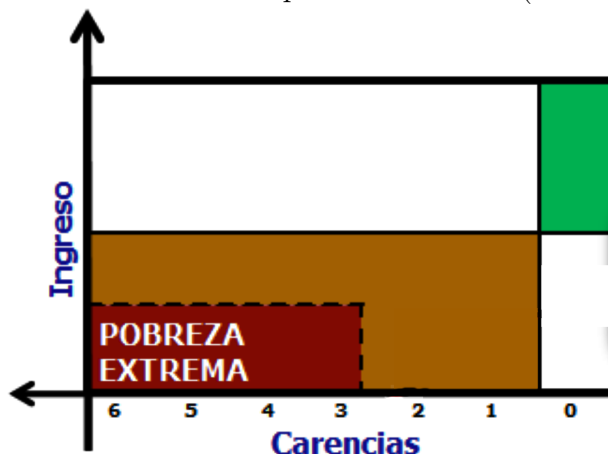
**Vulnerables por ingresos** Personas sin carencias sociales pero con un ingreso inferior a la línea de bienestar.

**Vulnerables por carencia social** Personas con al menos una carencia social pero con ingreso mayor a la LBE.

**Población pobre** Personas con al menos una carencia y un ingreso inferior a la LBE.

**Pobreza extrema** Es un subconjunto de la población pobre, que consiste en personas con al menos 3 carencias sociales y un ingreso inferior a la LBM (ver Figura 1.3).

Figura 1.3: Identificación de la pobreza extrema (CONEVAL,2015)



A partir de esta definición se realiza la medición de la pobreza según la normatividad oficial. La medición de la pobreza presentada por el CONEVAL (2013a) cada año desde 2010 está basada en estos criterios; y se calcula mediante la ENIGH-MCS elaborada por el INEGI. Esta herramienta de medición se levanta cada dos años, y permite un desglose nacional y estatal. La medición municipal de la pobreza se realiza cada 5 años mediante los censos y conteos de población y vivienda también realizados por el INEGI; aunque en 2010 no se realizó la medición de todos los componentes de la PM.

*Perfilando el problema:* La encuesta tiene una muestra reducida, que no permite inferir para los espacios operativamente relevantes. Sin embargo, el censo de población y vivienda 2010 (CPV2010) tiene una mayor cobertura, aunque no incluye toda la información necesaria para identificar la PM. El objetivo es utilizar la información del CPV2010 para poder estimar la PM en espacios más específicos.

### 1.1.2. Modelo Binario Multivariado

Los componentes de la PM se pueden agrupar en 1) Carencias y 2) Ingreso, como se muestra en la figura 1.1.

Mientras que las carencias se sistematizan a partir de la pregunta ¿tiene tal condición?, cuya respuesta es nominal del tipo binaria: sí/no; la variable ingreso es continua en  $\mathbb{R}^+$ , sin embargo se discretiza a una respuesta binaria, a partir de la LBE como se muestra más adelante, equivalente a la Figura 1.2.



Cada uno de los componentes de la PM se encuentra en un espacio binario  $\{0 = \text{no}, 1 = \text{sí}\}$ . Es posible caracterizar los componentes de la PM mediante un vector de respuestas binarias:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}.$$

La caracterización de las variables binarias correspondientes a las carencias es inmediata:

$$\begin{aligned} y_1 &= \text{Rezago educativo} \in \{0, 1\}, \\ y_2 &= \text{Acceso a servicios de salud} \in \{0, 1\}, \\ y_3 &= \text{Acceso a seguridad social} \in \{0, 1\}, \\ y_4 &= \text{Acceso a la alimentación} \in \{0, 1\}, \\ y_5 &= \text{Calidad y espacios en la vivienda} \in \{0, 1\}, \\ y_6 &= \text{Acceso a servicios públicos en la vivienda} \in \{0, 1\}. \end{aligned}$$

Se definen las variables dicotómicas correspondientes al ingreso (*ing*) mediante las líneas de bienestar, tal como lo realiza CONEVAL:

$$y_7 = \begin{cases} 1 & \text{ing} < LBE \\ 0 & \text{d.o.m} \end{cases}, \quad (1.1a)$$

$$y_8 = \begin{cases} 1 & \text{ing} < LMB \\ 0 & \text{d.o.m} \end{cases}. \quad (1.1b)$$

A partir de este vector de dimensión  $p = 8$ , podemos fácilmente, mediante combinaciones lineales llegar a la definición de PM y pobreza extrema (PE), tal como se presentan en las Figuras 1.2 y 1.3:

$$PM = \begin{cases} 1 & \sum_{i=1}^6 y_i \geq 1 \wedge y_7 = 1 \\ 0 & \text{d.o.m} \end{cases}, \quad (1.2)$$

$$PE = \begin{cases} 1 & \sum_{i=1}^6 y_i \geq 3 \wedge y_8 = 1 \\ 0 & \text{d.o.m} \end{cases}, \quad (1.3)$$

donde  $\wedge$  indica una conjunción lógica, i.e. que la persona tiene un número mínimo de carencias, y además un ingreso menor a la línea de bienestar correspondiente.

La pregunta sobre si las variables en  $\mathbf{y}$  están relacionadas, i.e. si son independientes, es relevante. Intuitivamente podemos asumir que algunas de ellas tienen relaciones directas: Servicios en la vivienda-Calidad y espacios en la vivienda, ingreso-alimentación. Mientras que otras relaciones se encuentran ampliamente documentadas (Levy Algazi, 2007) a partir de las condiciones laborales. Existen tendencias actuales de reconfiguración del Estado, que propician independencia entre algunas variables (Servicios de salud, seguridad social en ancianos), pero debido a la conformación histórica del estado mexicano todavía se encuentran relacionadas. Esto implica que existe una estructura de covarianza/correlación ( $\Sigma$ ) entre las variables de interés que no debe ser ignorada; los elementos en  $\mathbf{y}$  deben ser modelados conjuntamente y no por separado. La PM puede ser modelada como un vector de respuestas binarias correlacionadas.

Para la estimación de la pobreza en espacios específicos, aquellos donde la ENIGH-MCS no fue levantada, es necesario contar con información de la población en estos lugares, operativamente relevantes para los programas sociales; esta información la podemos encontrar en el Censo de Población y Vivienda 2010, en cuyo cuestionario no se incluyeron todas las variables necesarias para la medición de la pobreza, sin embargo se aplicó para la totalidad de la población del país. Se cuenta con un conjunto de covariables  $\mathbf{X}$  observadas que tienen relación con las variables de interés  $\mathbf{y}$ , por lo tanto este problema puede ser entendido como un problema de regresión:

$$E[\mathbf{y}] = f(\mathbf{X}),$$

donde  $f()$  es cierta función definida posteriormente.

El problema, originalmente operativo, se entiende ahora como un problema de regresión de una variable binaria multivariada; y cuya interpretación es de orden social.

## 1.2. Estado del Arte

En esta sección se presenta de manera resumida la situación actual del problema de interés; organizado en tres partes: 1) estudios sobre pobreza, 2) modelación de variables binarias multivariadas y 3) modelos probit multivariados.

### 1.2.1. Pobreza Multidimensional

El concepto de PM se ha popularizado en años recientes a pesar de que la inclusión de carencias en los estudios de pobreza no es nueva; se interpreta este aumento en oposición a la definición basada exclusivamente en el ingreso, como lo hacen Kakwani y Silber (2008):

Such a view makes a distinction between a first dimension of the capability space, where poverty is defined in absolute terms and refers to income levels which do not guarantee that basic physical needs will be covered, and a second dimension emphasizing a relative approach to poverty and considering those income levels which do not allow individuals to “function properly” in their social environment. The second aspect stresses evidently the by now very popular concept of social exclusion. Such a distinction is, however, not new.<sup>1</sup>

Hay un consenso más o menos generalizado sobre la existencia de múltiples dimensiones del fenómeno, pero no hay consenso sobre cuáles son éstas, cómo se manifiestan y miden:

...implementation of multidimensional poverty raises several conceptual and empirical issues. Among these, the choice of the most appropriate degree of aggregation for different indicators is one of the most debated.<sup>2</sup> (Coromaldi y Zoli, 2012)

Es sobre el problema de cómo definir a los pobres, que muchos de los estudios se centran; tratan de identificar las variables, y las funciones de éstas, que mejor describen la pobreza. Este es claramente un problema de aprendizaje no supervisado (Bishop, 2006, p.3).

---

<sup>1</sup>Este punto de vista hace una distinción entre una primera dimensión del espacio de capacidades, donde la pobreza se define en términos absolutos y se refiere a los niveles de ingreso que no garantiza que las necesidades físicas básicas serán cubiertas, y una segunda dimensión que enfatiza un enfoque relativista de la pobreza y toma en cuenta los niveles de ingresos que no permiten a las personas “funcionar adecuadamente” en su entorno social. El segundo aspecto destaca evidentemente el concepto por ahora muy popular de exclusión social. Tal distinción no es, sin embargo, nueva.

<sup>2</sup>... (la) implementación de la pobreza multidimensional plantea varios problemas conceptuales y prácticos. Entre estos, la elección de la construcción de los distintos indicadores es uno de los más debatidos.

En México, desde 2010 se manejan de manera oficial este tipo de indicadores, mediante el índice de marginación (CONAPO, 2012), y posteriormente mediante el índice de rezago social (CONEVAL, 2007); ambos son metodológicamente iguales. Utilizan análisis de componentes principales (PCA), a diferentes niveles de desagregación geográfica, la diferencia entre ellos radica en las variables que utilizan para la descomposición espectral; ambos índices son el primer componente principal de la correspondiente matriz de correlaciones. Aunque novedosos a nivel de administración pública nacional, estos índices muestran claros problemas como lo demuestran sus resultados en términos de valores propios<sup>3</sup>, además han sido poco exitosos a nivel operativo, como lo demuestra la estrategia 100x100 de la administración calderonista<sup>4</sup>.

Son dos los puntos centrales de las investigaciones metodológicas sobre PM, como menciona Pérez–Mayo (2005):

It is not easy to determine what and how many indicators should have been taken into account for deprivation measuring,<sup>5</sup>

así que la evaluación sobre cuáles indicadores deben ser incluidos ha sido una preocupación. El otro es cómo incluirlos en la medición de la PM :

There is a trade-off between synthesis and the best description. This issue has not been defined yet in the literature. Although, on one hand, joining all the attributes in an index offers the advantage of summarizing the complexity in a simple way, such aggregation causes a loss of information. Since a multidimensional phenomenon is studied, the search for a better description of such variety is an important goal<sup>6</sup> (Pérez–Mayo, 2005).

---

<sup>3</sup>En 2010, el índice de marginación municipal explicaba apenas el 50 % de la variabilidad de los indicadores que lo componen.

<sup>4</sup>En este sentido, resulta sorprendente que la innovación más reciente del CONAPO (2013) sea el índice de marginación absoluto, que consiste en un promedio de porcentajes, ignorando el problema que motivó el uso de PCA, y sin ninguna importancia estadística u operativa.

<sup>5</sup>No es fácil determinar cuales y cuántos indicadores deberían haberse tomado en cuenta para la medición de la privación,

<sup>6</sup>Hay un equilibrio entre la síntesis y la mejor descripción. Este problema no se ha definido todavía en la literatura. Aunque, por una parte, unir todas las características en un índice ofrece la ventaja de resumir la complejidad de una manera simple, tal reducción causa una pérdida de información. Dado que se estudia un fenómeno multidimensional, la búsqueda de una mejor descripción de este tipo es un objetivo importante

El PCA utilizado oficialmente en México se encuentra ya en desuso a nivel académico, las alternativas son variadas: non-linear principal component analysis (Coromaldi y Zoli, 2012), totally fuzzy and relative approach, information theory approach, axiomatic approach (D'Ambrosio et al., 2011), latent class analysis (Pérez-Mayo, 2005)<sup>7</sup>, hierarchical latent class, item response theory (Szeles y Fusco, 2013). Entre éstas es común la perspectiva de verosimilitudes y el uso del algoritmo EM (Dempster et al., 1977) .

Estos métodos pueden describirse como aprendizaje no supervisado, i.e. no existe una variable respuesta observable: la pobreza. Esto ha llevado a estudios sobre comparaciones entre la clasificación de los métodos (D'Ambrosio et al., 2011), así como a la complementación con otros métodos estadísticos para la evaluación del impacto de las variables/dimensiones/indicadores incluidos, ya que ambos problemas deben ser tratados conjuntamente; la manera de incluir las variables afecta el impacto que tienen en la definición de PM. Varios de los métodos utilizados construyen un índice (o variable latente), que conlleva el problema de definir el límite a partir del cual alguien se puede considerar pobre, ahí la discusión se encuentra aún más lejos de un consenso.

Los problemas de investigación aquí mencionados no coinciden con el de la presente investigación; en el caso mexicano, al menos a nivel de políticas públicas, este problema fue solucionado por decreto. La definición de PM descrita en el segundo apartado de este capítulo nos permite tener una variable respuesta de pobreza observable y bien definida. Esta definición de PM puede tener fallas y ser mejorable, pero tiene dos ventajas importantes: 1) al ser oficial, se obliga a las dependencias mexicanas orientar la política de desarrollo social a partir de ella, así como un financiamiento importante para su medición (ENIGH-MCS), 2) permiten a las instituciones/organizaciones orientar sus acciones a partir de indicadores bien definidos a modificar; y no en principios o valores abstractos: “hay que bajar los indicadores”.

Resulta interesante mencionar la investigación de Rojas (2008) por tres razones, 1) trata el caso mexicano, 2) lleva al extremo la definición de PM basándola en auto-percepción de la pobreza, y 3) propone un modelo multinomial (probit)<sup>8</sup>.

---

<sup>7</sup>Esta perspectiva fue recientemente adoptada por CONEVAL (2013c), aunque a manera de caja negra en la elaboración del índice de rezago social a nivel de AGEB.

<sup>8</sup>El problema de un vector de respuestas binario puede ser trasladado a un problema multinomial, donde cada posible combinación del vector se considera una alternativa de una variable multinomial, cuyo número de alternativas es  $2^p$ ; sin embargo, esto es fun-

### 1.2.2. Respuesta Binaria Multivariada

Los vectores de respuestas binarias han sido tratados de distintas maneras, de entre las primeras destacan los modelos log-lineales.

With discrete outcomes, the statistical literatura has focused on log-linear models for the joint distribution of a vector of responses. Here the natural parameters have interpretations in terms of the conditional distributions of a subset of variables given the others and not in terms of marginal distributions<sup>9</sup> (Liang et al., 1992).

Existe gran complejidad en la interpretación de los modelos log-lineales; de manera simplificada, la diferencia entre variables respuesta y variables independientes no está presente. Para tratar un vector de respuestas binarias es más atractivo considerar que las covariables tienen efectos en el vector respuesta. En la literatura podemos agrupar estas propuestas en dos tipos, aquellas que intentan modelar directamente las variables respuestas, categóricas, y las que lo hacen mediante variables latentes.

El principal problema que presentan los modelos que utilizan directamente la variable respuesta, consiste en la complejidad de la función de verosimilitud, que no es directamente maximizable.

Para resolver este problema se han propuesto distintas alternativas: *quadratic models* (Zhao y Prentice, 1990), que incluyen en la especificación del modelo términos cuadráticos y redundantes de la variable respuesta, los autores se refieren a este método como pseudo-máxima verosimilitud, pero está limitado por la dimensión del vector respuesta ( $p \leq 12$ ); en general, el cálculo de este tipo de propuestas implica la optimización en los  $2^p$  posibles resultados del vector respuesta, limitando así la obtención de los estimadores máximo verosímiles (MLE). Otra forma de resolver este problema consiste en recurrir a aproximaciones de las funciones de score (Zhao y Prentice, 1990), esta propuesta resulta en un conjunto de *Generalized Estimating Equations* (GEE)

---

cional para casos donde la dimensión ( $p$ ) es pequeña, en el caso del vector propuesto de dimensión  $p = 7$ , el número de alternativas de la variable multinomial correspondiente es de 128; aunque para nuestro problema se puede simplificar a 67 alternativas; pero variables multinomiales con tantas posibles alternativas no son fácilmente manejables ni conceptual y menos computacionalmente.

<sup>9</sup>En variables discretas, la literatura estadística se ha centrado en modelos log-lineales para la distribución conjunta de un vector de respuestas. Aquí los parámetros naturales tienen interpretaciones en términos de las distribuciones condicionales de un subconjunto de variables dadas el resto y no en términos de distribuciones marginales

desde la estadística clásica. Las GEE consisten en sustituir la optimización de la función de verosimilitud conjunta por la optimización de funciones, aproximadas a la función de score, y que son asintóticamente normales; para esto es necesario fijar ciertos valores de las funciones.

Otra propuesta desde la estadística frecuentista consiste en introducir los parámetros de correlación como efectos aleatorios de los modelos (*Rasch type model*). Entre estas propuestas resalta la de Agresti (1997) que define un modelo logístico con parámetros aleatorios; y utiliza el ajuste de un modelo log-lineal para estimar las probabilidades marginales; resultando en la doble complejidad de ajustar dos modelos de diferente clase, de manera independiente.

Una propuesta más reciente y acabada es la de Arminger et al. (2008). Ellos integran algunas de las ventajas de los modelos con parámetros aleatorios; caracterizando la correlación entre los elementos del vector respuestas mediante una variable latente categórica que requiere de modelación semiparamétrica. En resumen es un modelo logit con parámetros aleatorios, modelados mediante clases latentes, y que resultan en una distribución de mezcla.

La segunda alternativa para modelar datos binarios correlacionados se basa en la inclusión (*data augmentation*) de una variable latente continua multivariada. Aunque esta solución del problema no es reciente, recobró importancia partir del trabajo seminal de Albert y Chib (1993) -esta propuesta se detalla en el siguiente capítulo- y que actualmente se encuentra ampliamente documentada para distintos modelos de variables categóricas (Chib, 2000; Congdon, 2005; Jackman, 2009). Esta perspectiva puede utilizarse a partir únicamente de las verosimilitudes, pero su uso más fructífero ha sido en estadística Bayesiana. Una de las primeras aproximaciones al caso binario multivariado desde esta perspectiva revitalizada, fue la de Street y Czado (1996), partiendo de una variable latente Gaussiana; esta propuesta es ampliada y formalizada por Chib y Greenberg (1998), y se describe detalladamente en el siguiente capítulo.

Entre las propuestas de variables latentes utilizando métodos de Cadenas de Markov Monte Carlo (MCMC), la función probit es la más utilizada, ya que la simulación de datos es sencilla y la distribución Gaussiana multivariada tiene propiedades atractivas. Por ejemplo, tanto las distribuciones marginales como las condicionales son Gaussianas. Sin embargo, la distribución logística también tiene sus atractivos,

..., since the standard normal distribution does not have a closed

form, the regression coefficients are not interpretable as changes in a simple function of the category probabilities. Instead, the coefficients have a more esoteric interpretation in terms of changes in the underlying normal mean. For this reason, logistic models may be preferred<sup>10</sup> (O'Brien y Dunson, 2004).

Los parámetros de los modelos logísticos tienen una interpretación directa con respecto a las razones de momios. O'Brien y Dunson (2004) proponen un modelo cuya distribución conjunta es logística multivariada, mediante una aproximación a la distribución  $t$ -multivariada, y cuyas marginales también son logísticas. La simulación se realiza mediante un parámetro de varianza aleatorio con distribución gamma. Este modelo presenta la ventaja de poder especificar distribuciones a priori impropias no informativas o informativas. Sin embargo, para la simulación del parámetro de correlaciones es necesario recurrir a un paso Metropolis, aumentando el costo computacional del método.

Una ventaja adicional de la perspectiva de variables latentes es que permite transitar de manera sencilla hacia modelos similares, como modelos multinomiales, multinomiales ordinales e incluso multinomiales multivariados (McCulloch et al., 2000; O'Brien y Dunson, 2004; Imai y Van Dyk, 2005).

Aun cuando no se utiliza en el presente trabajo, vale la pena referirse a la propuesta de Swihart et al. (2014), que plantea que de hecho a partir de la definición de las distribuciones conjuntas de las variables latentes como distribuciones de cópula, es sencillo definir distribuciones marginales con propiedades deseables, y simular la distribución conjunta, ampliando y generalizando las posibles funciones liga de los modelos categóricos multivariados; adicionalmente a la incorporación de interceptos aleatorios a los modelos marginales, sin embargo computacionalmente aún no se encuentra en una eficiencia que las haga atractiva.

### 1.2.3. Modelo Probit Multivariado

En este punto se puede afirmar que un Modelo Probit Multivariado (MVP), es una alternativa viable para modelar la PM. En este apartado

---

<sup>10</sup> ..., dado que la distribución normal estándar no tiene una forma cerrada, los coeficientes de regresión no son interpretables como cambios en una función simple de probabilidades de las categorías. En cambio, los coeficientes tienen una interpretación más esotérica en términos de cambios en la media normal subyacente. Por esta razón, modelos logísticos suelen preferirse.



se realiza una breve revisión de las aplicaciones del MVP en distintas áreas de estudio.

En la agricultura Kassie et al. (2013) utilizaron un MVP<sup>11</sup> para estudiar la adopción de prácticas sustentables en áreas rurales de Tanzania. Las prácticas de interés sobre las que se investigó van desde rotación de cultivos de leguminosas, hasta introducción de semillas mejoradas. Concluyen que la introducción de la correlación, entre las prácticas adoptadas, mejora las estimaciones del análisis.

Gibbons y Lavigne (1998) utilizan un MVP para modelar la presencia de cuatro desordenes psiquiátricos en niños. El análisis está basado en la metodología propuesta por Bock y Gibbons (1996), y a partir de la definición de una variable latente, trasladan las probabilidades a una variable de distribución Gaussiana, y utilizan efectos aleatorios en la definición de la media de la variable latente. La identificación de los estimadores de máxima verosimilitud (MLE) se realizan mediante una combinación del algoritmo EM para establecer valores iniciales apropiados, para iniciar la optimización en la función de score. Para esto es necesario aproximar la integral que representa la probabilidad de cada observación, mediante el método Gauss-Hermite. Esta metodología está limitada a un tamaño máximo del vector respuesta de 8 elementos.

Calia y Ferrante (2013) estudian cómo las empresas manufactureras en Italia eligen estrategias o formas de internacionalización. El modelo utilizado incluye una matriz de correlaciones, el principal problema para la maximización de la verosimilitud consiste en la evaluación de una integral de alto orden, que resuelven utilizando simulación Monte Carlo (Greene, 2002); los errores estándar se estiman mediante el método Delta y prueban significancia mediante una prueba de Wald. La simulación de la función de verosimilitud evita, al menos para ciertos tamaños del vector respuesta, definir una estructura de correlación.

Este mismo método de estimación lo utiliza Nilsson (2012) en su estudio sobre la persistencia de la pobreza entre gemelos en Suecia. Donde la pregunta central es distinguir sobre si el permanecer en una situación de pobreza se debe a las condiciones que esta situación impone sobre el individuo, o sobre las características particulares del individuo. Es un estudio tipo panel,

---

<sup>11</sup>Aunque el estudio es interesante en cuanto al área de aplicación, carece de la descripción del método utilizado, y no es replicable, no especifica el modelo, ni el método de estimación; sirve así para recordar que los estudios deben incluir una descripción metodológica suficiente.

utilizando registros administrativos. El autor asume una estructura equicorrelacionada, i.e. la misma correlación entre todas las mediciones. Al tratarse de un estudio econométrico, adicionalmente a la estimación de probabilidades conjuntas, calcula estadísticos adicionales, como diferencia entre razones de promedios de probabilidades.

Hasta este punto se han presentado casos donde aún cuando se utilizaron variables latentes, sólo en los últimos dos casos se simulan muestras, i.e. se recurre a un método MCMC, pero aún en estos estudios la muestra simulada es utilizada para estimar MLE, a partir de los cual se supone una aproximación a la distribución normal. Si bien asintóticamente es correcta la aproximación, esta versión tiene el problema adicional de que utiliza únicamente el MLE para las predicciones de probabilidades. Esta situación no se presenta con la estimación Bayesiana, ya que utiliza para la predicción de nuevos valores toda la información de la distribución a posteriori de los parámetros.

El principal campo donde los modelos categóricos utilizan la perspectiva Bayesiana es en la modelación de decisiones (*choice behavior*). La estimación Bayesiana puede ser ampliada a casos más complejos de manera relativamente sencilla. Paap et al. (2005) utilizan un modelo de dos etapas para predicción del voto en Holanda. La primer etapa consiste en un MVP para estimar las alternativas entre las que los electores deciden su voto, un segundo modelo es ajustado para decidir entre dichas alternativas, en este caso se trata de un modelo Probit Multinomial. Si bien simplifican la inclusión de los modelos, al definirlos independientes entre sí, la implementación se basa en un método MCMC, con pasos Metropolis. Otro objetivo de este estudio es la inclusión de diferentes tipos de *No Respuesta*, tanto aquellos que no saben, no votarán, o no quisieron contestar la pregunta. La inclusión de este tipo de respuestas censuradas es sumamente sencillo para los modelos de variables latentes, en tanto únicamente implica remover restricciones de la distribución simulada.

De entre los problemas que el *choice behavior* puede tratar, el del marketing, entendido como lo que compra la gente, es el que ha tenido mayor desarrollo en términos de los modelos categóricos de variables latentes. En este campo referirse a Seetharaman et al. (2005), es relevante en tanto que sintetiza la investigación hasta el momento:

This paper presents a review of where we are, and where we should be going. We present an overview of multi-category models...<sup>12</sup>

---

<sup>12</sup> En este trabajo se presenta una revisión de dónde estamos y dónde deberíamos ir . Se presenta una visión general de los modelos de multiples categorías...

De entre el tipo de problemas que se han tratado en el marketing están:

- La compra conjunta de categorías de productos (complementos y sustitutos) .
- Tiempo entre las compras mediante un Modelo Multivariado de Riesgo, y variaciones de éste.
- Modelos de elección en paquete (*bundle*), que trata de explicar cómo un consumidor elige comprar en paquete o conjunto, ciertos tipos de productos.
- Elección de marcas entre diferentes categorías de productos, qué marca se elige de un tipo de producto dado que en otro tipo de producto se eligió cierta marca.
- Volumen de compra (cuánto se compra) de productos, entre categorías.

Si bien no todos los estudios incluidos se ajustan mediante modelos Bayesianos de variables latentes, la elección de Seetharaman et al. (2005) es clara:

Multi-category models are computationally burdensome in terms of the number of parameters to be estimated. Frequentist approaches to estimation would typically fail to handle this estimation burden unless one focuses on a limited (i.e., two or three) number of product categories and imposes theory-based restrictions on the estimable set of parameters... However, if one's interest, instead, is in estimating a flexible statistical model of purchase outcomes without imposing parameter restrictions (since there is little historical research wisdom on multi-category choice behavior of households to yield convenient a priori restrictions), Markov Chain Monte Carlo (MCMC) methods have to be invoked to enable the estimation of these models in the *Bayesian* realm...What underlies the MCMC estimation are (1) the *Bayes Theorem*, that says that sampling the posterior density of the model parameters must be the focus of estimation, and (2) the idea that one can use conditional distributions of model parameters to obtain their joint distribution. Generating correlated draws from

a Markov chain—whose limiting, invariant distribution is the target distribution - is the central idea behind the estimation.<sup>13</sup>

### 1.3. Problematización de la PM

A partir de lo hasta aquí tratado es posible caracterizar el presente trabajo de manera más formal. Si bien la motivación de la investigación es operativa, el problema original es de investigación social, el desarrollo de la investigación se da dentro de la modelación estadística. Así el planteamiento formal refleja la multiplicidad de espacios conceptuales en los que se ubica la investigación.

#### Problema

No existe una estimación de PM útil para implementar políticas focalizadas.

La definición oficial de PM reduce la complejidad de toma de decisiones dentro de las organizaciones encargadas del desarrollo social; pero al hacerlo aumenta la complejidad técnica de identificación de individuos en situación de pobreza. Se traslada entonces el problema al espacio metodológico.

#### Objetivo

Implementar un método que permita realizar estimaciones operativa-

---

<sup>13</sup>Los modelos de múltiples categorías son computacionalmente problemáticos en relación al número de parámetros a estimar. Los enfoques frecuentistas para estimar típicamente fracasan en esta tarea a menos que se centren un número reducido (dos o tres) de categorías de productos e impone restricciones teórica al conjunto de parámetros estimables... Sin embargo, si el interés es la estimación de un modelo estadístico flexible sobre las compras de productos sin imponer restricciones a los parámetros (dado que existe poca investigación previa sobre el comportamiento de las familias al elegir conjuntos de productos como para introducir restricciones a priori convenientes), Los métodos de cadena de Markov Monte Carlo (MCMC) deben ser introducidos para permitir la estimación de estos modelos en el *universo bayesiano*... lo que subyace en la estimación MCMC son (1) el *Teorema de Bayes*, que indica que el objetivo de la estimación debe ser el muestreo de la distribución a posteriori de los parámetros del modelo, y (2) la idea de que se pueden utilizar distribuciones condicionales de los parámetros para obtener su distribución conjunta. Generar muestras correlacionadas de una cadena de Markov -en cuya convergencia, la distribución estacionaria que es la distribución objetivo- es la idea central detrás de la estimación.

mente relevantes<sup>14</sup>. Esto significa proponer un método de modelación de la PM que además de realizar buenas estimaciones pueda implementarse de manera sencilla y eficiente.

### Hipótesis

La PM puede ser modelada mediante un MVP utilizando variables latentes.

Mediante una transformación el MVP puede ser tratado como un modelo jerárquico, reduciendo el costo computacional del ajuste.

---

<sup>14</sup>Mención “especial” merece la estimación de la pobreza municipal realizada por CONEVAL entre 2009 y 2010; el objetivo era similar al planteado aquí, estimar a nivel municipal la prevalencia de individuos en situación de PM. La metodología de este ejercicio (CONEVAL, 2011) explica que se convocó a un grupo de expertos nacionales e internacionales, a partir de este grupo se propusieron seis métodos, que fueron descartados. Así el CONEVAL decidió basarse de manera parcial en propuestas encontradas en el acerbo del Banco Mundial (Elbers et al., 2002, 2003; Demombynes et al., 2007), combinándolas con regresiones logísticas independientes, y estimaciones directas; obteniendo una metodología a pedazos, que tiene la ventaja de poderse implementar en Stata en lugar de R. Lo más relevante es que constantemente refiere para explicar la metodología a un documento que no existe o al menos no está disponible en la página del CONEVAL, y no se pudo rastrear en internet, por obvias razones no se incluye en las referencias.



# Capítulo 2

En este capítulo fundamentamos y describimos nuestra aproximación al problema de la modelación de las variables que componen la PM.

## 2.1. Marco teórico

### 2.1.1. Descomposición LDU

Para el desarrollo de la propuesta de modelo es necesario presentar el siguiente teorema, tomado de Shapiro (2015)<sup>1</sup>.

**Teorema 1.** *Let  $\mathbf{A}$  be a nonsingular matrix. Suppose it is possible to factor  $\mathbf{A}$  as  $\mathbf{A} = \mathbf{LDU}$ , where  $\mathbf{L}$  is lower triangular with ones on the diagonal, and  $\mathbf{D}$  is diagonal. Then the following holds.*

1.  $\mathbf{L}, \mathbf{D}$  and  $\mathbf{U}$  are uniquely determined by  $\mathbf{A}$ .
2.  $\mathbf{A}$  is symmetric if and only if  $\mathbf{U} = \mathbf{L}^t$ .
3.  $\mathbf{A}$  is Hermitian if and only if  $\mathbf{D}$  has real diagonal entries and  $\mathbf{U} = \mathbf{L}^*$ .
4.  $\mathbf{A}$  is positive definite if and only if  $\mathbf{D}$  has positive diagonal entries.

---

<sup>1</sup>Sea  $\mathbf{A}$  una matriz no singular. Se asume que es posible factorizar  $\mathbf{A}$  como  $\mathbf{A} = \mathbf{LDU}$ , donde  $\mathbf{L}$  es in matriz triangular inferior con unos en la diagonal, y  $\mathbf{D}$  es una matriz diagonal. De cual sigue lo siguiente.

- 1)  $\mathbf{L}, \mathbf{D}$  y  $\mathbf{U}$  son únicas y determinadas por  $\mathbf{A}$ .
- 2)  $\mathbf{A}$  es simétrica si y solo si  $\mathbf{U} = \mathbf{L}^t$ .
- 3)  $\mathbf{A}$  es Hermitiana si y solo si  $\mathbf{D}$  tiene entras diagonales reales y  $\mathbf{U} = \mathbf{L}^*$ .
- 4)  $\mathbf{A}$  es positiva definida si y solo si  $\mathbf{D}$  tiene elementos en la diagonal todos positivos.

**Corolario:** A partir de esta factorización es fácil mostrar que el primer elemento de la matriz  $\mathbf{A}$  es igual al primer elemento de la matriz diagonal, i.e.  $a_{11} = d_1$

$$a_{11} = [d_1 \ 0 \ \dots \ 0] \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = d_1.$$

Para el caso de  $p = 3$

$$\begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} 1 & u_{21} & u_{31} \\ 0 & 1 & u_{32} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} d_1 & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Esto se cumple únicamente para  $d_1$ .

## 2.1.2. Data augmentation

El uso de métodos Cadenas-de-Markov-Monte-Carlo (MCMC) en estadística Bayesiana utilizando variables latentes para modelar variables categóricas, tomó fuerza en los 90, principalmente a partir del trabajo de Albert y Chib (1993), que de manera sintética plantea lo siguiente:

Suppose that  $N$  independent binary random variables  $Y_1, \dots, Y_N$  are observed, where  $Y_i$  is distributed Bernoulli with probability of success  $p_i$ . The  $p_i$  are related to a set of co-variates that may be continuous or discrete. Define the binary regression model as  $p_i = H(\mathbf{x}_i^t \boldsymbol{\beta})$ ,  $i = 1, \dots, N$ , where  $\boldsymbol{\beta}$  is a  $k \times 1$  vector of unknown parameters,  $\mathbf{x}_i^t = [x_{i1}, \dots, x_{ik}]$  is a vector of known covariates, and  $H(\cdot)$  is a known cdf linking the probabilities  $p_i$  with the linear structure  $\mathbf{x}_i^t \boldsymbol{\beta}$ . The probit model is obtained if  $H$  is the standard Gaussian cdf, whereas the logit model is obtained if  $H$  is the logistic cdf<sup>2</sup> (Albert y Chib, 1993).

---

<sup>2</sup>Supongamos que se observan  $N$  variables aleatorias independientes binarias  $y_1, \dots, y_n$ , donde  $Y_i$  se distribuye Bernoulli con probabilidad de éxito  $p_i$ . Las  $p_i$  están relacionadas con un conjunto de covariables que pueden ser continuas o discretas. Definir el modelo de regresión binaria como  $p_i = H(\mathbf{x}_i^t \boldsymbol{\beta})$ ,  $i = 1, \dots, N$ , donde  $\boldsymbol{\beta}$  es un vector de dimension



Sea  $\pi(\boldsymbol{\beta})$  una distribución a priori, propia o impropia, entonces la distribución a posteriori de  $\boldsymbol{\beta}$  está dada por

$$\pi(\boldsymbol{\beta}|\text{data}) = \frac{\pi(\boldsymbol{\beta}) \prod_{i=1}^N H(\mathbf{x}_i^t; \boldsymbol{\beta})^{y_i} (1 - H(\mathbf{x}_i^t; \boldsymbol{\beta}))^{1-y_i}}{\int \pi(\boldsymbol{\beta}) \prod_{i=1}^N H(\mathbf{x}_i^t; \boldsymbol{\beta})^{y_i} (1 - H(\mathbf{x}_i^t; \boldsymbol{\beta}))^{1-y_i} d\boldsymbol{\beta}},$$

que no es manejable analíticamente; existen aproximaciones asintóticas que no trataremos aquí.

Suponiendo, que la función liga  $H$  corresponde a la función de distribución Gaussiana estándar, resultando en un modelo probit; la clave del data augmentation consiste en incluir  $N$  variables latentes independientes:  $Z_1, \dots, Z_N$ , donde  $Z_i \sim N(\mathbf{x}_i^t; \boldsymbol{\beta}, 1)$ , y

$$Y_i = \begin{cases} 1 & Z_i > 0, \\ 0 & Z_i \leq 0. \end{cases} \quad (2.2)$$

Observe that if the  $Z_i$ 's are known and a multivariate normal prior is chosen for  $\boldsymbol{\beta}$ , then the posterior distribution for  $\boldsymbol{\beta}$  can be derived using standard normal linear model results. The  $Z$ 's are of course unknown; however, given the data  $Y_i$ , the distribution of  $Z_i$  follows a truncated normal distribution. These principal observations, combined with the tool of Gibbs sampling, allow us to simulate from the exact posterior distribution of  $\boldsymbol{\beta}$  ... It can be easily shown that the  $Y_i$ 's are independent Bernoulli random variables with  $p_i = \Pr(Y_i = 1) = \Phi(\mathbf{x}_i^t; \boldsymbol{\beta})$  (Albert y Chib, 1993)<sup>3</sup>.

La distribución conjunta de  $\boldsymbol{\beta}$  y  $\mathbf{Z} = (Z_1, \dots, Z_N)$  dado los datos  $\mathbf{y} = (y_1, \dots, y_N)$  es

---

$k \times 1$  de parámetros desconocidos,  $\mathbf{x}_i^t = [x_{i1}, \dots, x_{ik}]$  es un vector de variables conocidas, y  $H(\cdot)$  es una función de distribución conocida que define la relación entre las probabilidades  $p_i$  y la estructura lineal  $\mathbf{x}_i^t; \boldsymbol{\beta}$ . Se obtiene el modelo probit si  $H$  es la función de distribución gaussiana estándar, mientras que se obtiene el modelo logit si  $H$  es la función de distribución logística

<sup>3</sup> Obsérvese que si los  $Z_i$  son conocidos y se elige como distribución a priori para  $\boldsymbol{\beta}$  una distribución normal multivariada, entonces la distribución posterior de  $\boldsymbol{\beta}$  puede ser derivada usando los resultados del modelo lineal estándar. Los  $Z_i$  son, por supuesto, desconocidos. Sin embargo, dado los datos  $y_i$ , la distribución de  $Z_i$  sigue una distribución normal truncada. Estas observaciones, junto con la herramienta de muestreo de Gibbs, nos permiten simular de la distribución posterior exacta de  $\boldsymbol{\beta}$  ... Se puede demostrar fácilmente que las  $y_i$  son variables Bernoulli independientes con  $p_i = \Pr(Y_i = 1) = \Phi(\mathbf{x}_i^t; \boldsymbol{\beta})$

$$\begin{aligned} \pi(\boldsymbol{\beta}, \mathbf{Z}|\mathbf{y}) = & \\ & C\pi(\boldsymbol{\beta}) \prod_{i=1}^N \{1(z_i > 0)1(y_i = 1) + 1(z \leq 0)1(y_i = 0)\} \\ & \times \phi(z_i; \mathbf{x}_i^T \boldsymbol{\beta}, 1). \end{aligned} \quad (2.3)$$

In (2.3),  $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  is the  $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  pdf,  $I(X \in A)$  is the indicator function that is equal to 1 if the random variable  $X$  is contained in the set  $A$ , and  $C$  here and henceforth is a generic proportionality constant. Note that this joint distribution is complicated in the sense that it is difficult to normalize and sample from directly (Albert y Chib, 1993)<sup>4</sup>.

Pero la estimación de la distribución marginal aposteriori de  $\boldsymbol{\beta}$ , utilizando Gibbs sampling, sólo requiere las distribuciones condicionales completas a posteriori de  $(\boldsymbol{\beta}|\mathbf{Z}, \mathbf{y})$  y  $(\mathbf{Z}|\boldsymbol{\beta}, \mathbf{y})$ , que tienen formas estándar.

A partir de (2.3) se tiene

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{y}, \mathbf{Z}) = C\pi(\boldsymbol{\beta}) \prod_{i=1}^N \phi(Z_i; \mathbf{x}_i^T \boldsymbol{\beta}, 1), \\ \boldsymbol{\beta}|\mathbf{y}, \mathbf{Z} \sim N_k(\hat{\boldsymbol{\beta}}, \hat{\mathbf{B}}^{-1}). \end{aligned} \quad (2.4)$$

This fully conditional posterior density is the usual posterior density for the regression parameter in the normal linear model  $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ , where  $\mathbf{X} = (\mathbf{x}_1^t, \dots, \mathbf{x}_N^t)^t$  and  $\varepsilon$  is distributed  $N_N(\mathbf{0}, \mathbf{I})$ , where  $\mathbf{I}$  is the identity matrix (Albert y Chib, 1993)<sup>5</sup>.

---

<sup>4</sup>En ( 2.3 ),  $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$  es función de distribución  $N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ ,  $I(X \in a)$  es la función indicadora que es igual a 1 si la variable aleatoria  $X$  está contenida en el conjunto  $A$  y  $C$  aquí y de ahora en adelante es una constante de proporcionalidad. Hay que tener en cuenta que esta distribución conjunta es complicada en el sentido de que es difícil de normalizar y simular directamente de esta

<sup>5</sup>Esta distribución posterior condicional completa es la distribución posterior habitual para el parámetro de regresión en el modelo de  $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$ , donde  $\mathbf{X} = (\mathbf{x}_1^t, \dots, \mathbf{x}_N^t)^t$  y  $\varepsilon$  se distribuye  $NM_N(\mathbf{0}, \mathbf{I})$ , donde  $\mathbf{I}$  es la matriz de identidad

Utilizando los resultados de un modelo lineal estándar, si  $\pi(\boldsymbol{\beta}) \propto 1$  entonces  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}(\mathbf{X}^t\mathbf{Z})$  y  $\hat{\mathbf{B}} = (\mathbf{X}^t\mathbf{X})^{-1}$ ; si  $\pi(\boldsymbol{\beta}) = N_k(\boldsymbol{\beta}^*, \mathbf{B}^*)$  entonces  $\hat{\boldsymbol{\beta}} = (\mathbf{B}^{*-1} + \mathbf{X}^t\mathbf{X})^{-1}(\mathbf{B}^{*-1}\boldsymbol{\beta}^* + \mathbf{X}^t\mathbf{Z})$  y  $\hat{\mathbf{B}} = (\mathbf{B}^{*-1} + \mathbf{X}^t\mathbf{X})^{-1}$ .

La distribución para  $\mathbf{Z}|\boldsymbol{\beta}$  también tiene una forma sencilla. Las variables  $Z_1, \dots, Z_N$  son independientes y

$$\begin{aligned} Z_i|\mathbf{y}, \boldsymbol{\beta} &\sim N(\mathbf{x}_i^t\boldsymbol{\beta}, 1) \text{ truncada por la izquierda en } 0 \text{ si } y_i = 1, \\ Z_i|\mathbf{y}, \boldsymbol{\beta} &\sim N(\mathbf{x}_i^t\boldsymbol{\beta}, 1) \text{ truncada por la derecha en } 0 \text{ si } y_i = 0. \end{aligned} \quad (2.5)$$

In practice it is customary to assign a flat noninformative prior to  $\boldsymbol{\beta}$ . Given a previous value of  $\boldsymbol{\beta}$ , one cycle of the Gibbs algorithm would produce  $\mathbf{Z}$  and  $\boldsymbol{\beta}$  from the distributions (2.5) and (2.4)... Note that it is computationally easy to simulate from both the multivariate normal distribution (2.4) and the truncated normal distributions in (2.5) (Albert y Chib, 1993)<sup>6</sup>.

### 2.1.3. MVP de Chib & Greenberg

El modelo tratado en el apartado anterior implica que  $p = 1$ . Para generalizarlo a  $p > 1$  se utiliza la propuesta de Chib y Greenberg (1998).

Let  $Y_{ij}$  denote a binary 0/1 response on the  $i$ th observation unit and the  $j$ th variable, and let  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^t$  ( $1 \leq i \leq n$ ) denote the collection of responses on all  $p$  variables. According to the multivariate probit model, the probability that  $\mathbf{Y}_i = \mathbf{y}_i$  conditioned on parameters  $\boldsymbol{\beta}, \boldsymbol{\Sigma}$  and a set of covariates  $\mathbf{x}_{ij}$ , is given by<sup>7</sup>

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}) \equiv \Pr(\mathbf{y}_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int_{A_{t_p}} \dots \int_{A_{t_1}} \phi_p(\mathbf{t}|\mathbf{0}, \boldsymbol{\Sigma}) d\mathbf{t} \quad (2.6)$$

<sup>6</sup>En la práctica es habitual asignar a priori uniforme no informativa a  $\boldsymbol{\beta}$ . Dado un valor previo de  $\boldsymbol{\beta}$ , un ciclo del algoritmo de Gibbs produciría  $\mathbf{Z}$  y  $\boldsymbol{\beta}$  a partir de las distribuciones (2.5) y (2.4) ... Considerese que es computacionalmente fácil simular tanto de la distribución normal multivariante (2.4) y las distribuciones normales truncadas de (2.5)

<sup>7</sup>Sea  $Y_{ij}$  una respuesta binaria 0/1 en la  $i$ -ésima observación y el  $j$ -ésima variable y  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ip})^t$  ( $1 \leq i \leq n$ ) denote el conjunto de respuestas en todas las  $p$  variables. De acuerdo con el modelo probit multivariado, la probabilidad de que  $\mathbf{Y}_i = \mathbf{y}_i$  condicionada por  $\boldsymbol{\beta}, \boldsymbol{\Sigma}$  y el conjunto de covariables  $\mathbf{x}_{ij}$ , está dada dada por (2.6)

Donde  $\phi_p(\mathbf{t}|\mathbf{0}, \Sigma)$  es la función de densidad de una variable normal  $p$ -variada con vector de medias  $\mathbf{0}$  y matriz de correlaciones  $\Sigma = \{\sigma_{jk}\}$ , y  $A_{ij}$  es el intervalo de integración, definido como

$$A_{ij} = \begin{cases} (-\infty, \mathbf{x}'_{ij}\boldsymbol{\beta}_j) & \text{si } y_{ij} = 0, \\ [\mathbf{x}'_{ij}\boldsymbol{\beta}_j, \infty) & \text{si } y_{ij} = 1. \end{cases} \quad (2.7)$$

$\boldsymbol{\beta}' = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_J) \in \mathbb{R}^k$  es un vector de parámetros desconocidos, y  $k = \sum_{j=1}^p k_j$ . Se denotan los  $J = p(p-1)/2$  parámetros libres de  $\Sigma$  como  $\boldsymbol{\sigma} \equiv (\sigma_{12}, \sigma_{13}, \dots, \sigma_{p-1,p})$ .

It is important to note that  $\Sigma$  must be in correlation form for identifiability reasons. Suppose that  $(\boldsymbol{\gamma}, \Omega)$  is an alternative parameterisation, where  $\boldsymbol{\gamma}$  is the regression parameter vector and  $\Omega$  is the covariance matrix. Then it is easy to show that  $\Pr(\mathbf{y}_i|\boldsymbol{\gamma}, \Omega) = \Pr(\mathbf{y}_i|\boldsymbol{\beta}, \Sigma)$ , where  $\boldsymbol{\beta}_j = \omega_{jj}^{-1/2}\boldsymbol{\gamma}$ ,  $\Sigma = \mathbf{C}\Omega\mathbf{C}'$  and  $\mathbf{C} = \text{diag}\{\omega_{11}^{-1/2}, \dots, \omega_{pp}^{-1/2}\}$ . A parameterisation in terms of covariances is therefore not likelihood identified (Chib y Greenberg, 1998)<sup>8</sup>.

De manera equivalente al caso univariado, se plantea el MVP en términos de las variables latentes Gaussianas  $\mathbf{Z}'_i = (z_{i1}, \dots, z_{ip})$ , con distribución  $\mathbf{Z}_i \sim N_p(\mathbf{X}_i\boldsymbol{\beta}, \Sigma)$ , donde  $\mathbf{X}_i = \text{diag}(\mathbf{x}_{i1}^t, \dots, \mathbf{x}_{iJ}^t)$  es una matriz de covariables de dimensión  $p \times k$ , y sea  $Y_{ij}$  igual a 1 ó 0 de acuerdo al signo de  $z_{ij}$  como en (2.2):

$$y_{ij} = I(z_{ij} > 0) \quad (j = 1, \dots, p), \quad (2.8)$$

donde  $I(A)$  es una función indicadora del evento  $A$ . Dada esta formulación, la probabilidad descrita en (2.6) se puede expresar como

$$\int_{B_{ip}} \dots \int_{B_{i1}} \phi_p(\mathbf{Z}_i|\mathbf{X}_i\boldsymbol{\beta}, \Sigma) d\mathbf{Z}_i \quad (2.9)$$

<sup>8</sup>Es importante señalar que *igma* debe estar en forma de correlación por razones de identificabilidad. Supongamos que  $(\boldsymbol{\gamma}, \Omega)$  es una parametrización distinta, donde  $\boldsymbol{\gamma}$  es el vector de parámetros de regresión y  $\Omega$  es la matriz de covarianzas. Es fácil demostrar que  $\Pr(\mathbf{y}_i|\boldsymbol{\gamma}, \Omega) = \Pr(\mathbf{y}_i|\boldsymbol{\beta}, \Sigma)$ , donde  $\boldsymbol{\beta}_j = \omega_{jj}^{-1/2}\boldsymbol{\gamma}$ ,  $\Sigma = \mathbf{C}\Omega\mathbf{C}'$  y  $\mathbf{C} = \text{diag}\{\omega_{11}^{-1/2}, \dots, \omega_{pp}^{-1/2}\}$ . Por tanto una parametrización en términos de covarianzas no es identificable

where  $B_{ij}$  is the interval  $(0, \infty)$  if  $y_{ij} = 1$  and the interval  $(-\infty, 0]$  if  $y_{ij} = 0$ . We let  $B_i = B_{i1} \times B_{i2} \times \cdots \times B_{ip}$  and note that  $B_i$  is the set-valued inverse of the mapping in (2.8). It is important to bear in mind that  $B_i$  unlike the  $A_{ij}$ , depends only on the value of  $y_{ij}$  and not on the parameters. This latent variable representation and the inverse mapping from  $y_{ij}$  to  $z_{ij}$  form the basis of our posterior sampling method (Chib y Greenberg, 1998)<sup>9</sup>.

Hasta aquí es la propuesta de este documento coincide con la de Chib y Greenberg, ellos proponen la estimación de los valores de  $\sigma$  mediante un paso Metropolis, que debe además cumplir que la correspondiente matriz  $\Sigma$  sea positiva definida, i.e.  $\mathbf{x}^t \Sigma \mathbf{x} > 0 \forall \mathbf{x} \in \mathbb{R}^d$

## 2.2. Modelo Probit Multivariado Condicionalmente Independiente

La propuesta a continuación presentada evita pasos Metropolis para la simulación de los parámetros de correlación. Se ejemplifica con un vector respuesta de dimensión  $p = 3$ .

### 2.2.1. El modelo

La nomenclatura utilizada es diferente a la de Albert y Chib (1993) y Chib y Greenberg (1998), de tal manera que se define la variable latente de dimensión  $p$  como

$$\mathbf{w}_i = \mathbf{X}_i \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}_i^*,$$

donde  $\mathbf{X}_i$  corresponde a la matriz diseño de covariables correspondientes al  $i$ -ésimo individuo y  $\boldsymbol{\beta}^*$  es el vector de parámetros lineales de dimensión  $k$ . Es relevante señalar que  $\boldsymbol{\varepsilon}^* \sim N_p(\mathbf{0}, \Sigma)$ , donde  $\Sigma$  es una matriz de correlaciones. Este modelo corresponde al propuesto por Chib y Greenberg (1998).

---

<sup>9</sup>donde  $B_{ij}$  es el intervalo de  $(0, \infty)$  si  $y_{ij} = 1$  y el intervalo de  $(-\infty, 0]$  si  $y_{ij} = 0$ . Se define  $B_i = B_{i1} \times B_{i2} \times \cdots \times B_{ip}$ , notese que  $b_i$  es la inversa del mapeo en (2.8). Es importante tener en cuenta que  $B_i$  a diferencia de los  $A_{ij}$ , depende sólo del valor de  $y_{ij}$  y no de los parámetros. Esta representación de la variable latente y el mapeo inverso de  $y_{ij}$  a  $z_{ij}$  forman la base de nuestro método de muestreo a posteriori

Se define a partir del teorema 1, la matriz  $\mathbf{L}$  como

$$\boldsymbol{\Sigma} = \mathbf{L}^{-1}\mathbf{D}(\mathbf{L}^{-1})^t \quad (2.10)$$

Por propiedades de la distribución normal multivariada se tiene

$$\boldsymbol{\varepsilon}^* \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}), \quad (2.11)$$

$$\mathbf{L}\boldsymbol{\varepsilon}^* \sim N_p(\mathbf{0}, \mathbf{D}), \quad (2.12)$$

$$\boldsymbol{\varepsilon} = \mathbf{D}^{-1/2}\mathbf{L}\boldsymbol{\varepsilon}^* \sim N_p(\mathbf{0}, \mathbf{I}_{p \times p}), \quad (2.13)$$

donde  $\mathbf{I}_{p \times p}$  es una matriz identidad, obteniendo así variables de error independientes y de varianza 1, como se encuentran en el modelo de Albert y Chib (1993).

En general la transformación propuesta es:

$$(\mathbf{D}^{-1/2}\mathbf{L}) \mathbf{w}_i = (\mathbf{D}^{-1/2}\mathbf{L}) \mathbf{X}_i\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}_i,$$

para la parte del vector respuesta y la parte lineal, la transformación es un poco más complicada. Se define por simplicidad en notación  $\boldsymbol{\mu}_i^* = \mathbf{X}_i\boldsymbol{\beta}^*$ , y ejemplificando se tiene

$$\left( \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{\delta_2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\delta_3}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \right) \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \left( \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{\delta_2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{\delta_3}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \right) \begin{bmatrix} \mu_1^* \\ \mu_2^* \\ \mu_3^* \end{bmatrix},$$

que resulta en

$$\begin{bmatrix} w_1 \\ \frac{l_{21}w_1}{\sqrt{\delta_2}} + \frac{w_2}{\sqrt{\delta_2}} \\ \frac{l_{31}w_1}{\sqrt{\delta_3}} + \frac{l_{32}w_2}{\sqrt{\delta_3}} + \frac{w_3}{\sqrt{\delta_3}} \end{bmatrix} = \begin{bmatrix} \mu_1^* \\ \frac{l_{21}\mu_1^*}{\sqrt{\delta_2}} + \frac{\mu_2^*}{\sqrt{\delta_2}} \\ \frac{l_{31}\mu_1^*}{\sqrt{\delta_3}} + \frac{l_{32}\mu_2^*}{\sqrt{\delta_3}} + \frac{\mu_3^*}{\sqrt{\delta_3}} \end{bmatrix}.$$

Reparametrizando, tenemos que

$$\begin{bmatrix} z_1 \\ \gamma_{21}z_1 + z_2 \\ \gamma_{31}z_1 + \gamma_{32}z_2 + z_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \gamma_{21}\mu_1 + \mu_2 \\ \gamma_{31}\mu_1 + \gamma_{32}\mu_2 + \mu_3 \end{bmatrix}. \quad (2.14)$$

Donde

$$z_j = w_j / \sqrt[+]{\delta_j}; \quad (2.15a)$$

$$\gamma_{jk} = \frac{l_{jk}}{\sqrt[+]{\delta_j}} \sqrt[+]{\delta_k}, j > k; \quad (2.15b)$$

$$\mu_j = \mu_j^* / \sqrt[+]{\delta_j} = \mathbf{x}_j^t \boldsymbol{\beta}, \quad (2.15c)$$

aquí  $\sqrt[+]{\cdot}$  indica que se trata de la raíz cuadrada positiva, correspondiente a la desviación estándar en (2.12).

Debido a lo mostrado en el corolario del teorema 1,  $\delta_1 = 1$ . A partir de la parametrización en (2.14) es posible plantear un modelo jerárquico condicionalmente independiente:

$$\begin{aligned} z_1 &= \mu_1 + \varepsilon_1, \\ z_2|z_1 &= \mu_2 + \gamma_{21}(\mu_1 - z_1) + \varepsilon_2, \\ z_3|z_1, z_2 &= \mu_3 + \gamma_{31}(\mu_1 - z_1) + \gamma_{32}(\mu_2 - z_2) + \varepsilon_3. \end{aligned}$$

En forma compacta se puede expresar como

$$z_j | \mathbf{z}_{<j} = \mu_j + \boldsymbol{\gamma}_j^t (\boldsymbol{\mu}_{<j} - \mathbf{z}_{<j}) + \varepsilon_j, \quad (2.17)$$

donde  $\mathbf{z}_{<j}^t = (z_1, z_2, \dots, z_{j-1})$ , evidentemente  $\mathbf{z}_{<1} = \emptyset$ .

Definiendo

$$f(\mathbf{z}) = f_1(z_1) f_2(z_2|z_1) f_3(z_3|z_1, z_2) \dots f_p(z_p|z_1, \dots, z_{p-1}),$$

podemos expresar la probabilidad de conjunta como

$$\Pr(\mathbf{Y}_i = \mathbf{y}_i) = \int_{B_{i1}} \dots \int_{B_{ip}} f(\mathbf{z}_i) d\mathbf{z}_i,$$

la demostración de esta igualdad se encuentra en el anexo A.1.2.

### Sistema de ecuaciones

Dado que  $\delta_{ij} = 0$ ,  $i \neq j$ , el siguiente sistema tiene una solución única:

$$\begin{aligned}
\mathbf{D} &= \mathbf{L}\Sigma\mathbf{L}^t \\
&= \begin{bmatrix} 1 & & & & & & & & & l_{21} + \sigma_{21} \\ & l_{21} + \sigma_{21} & & & & & & & & l_{21}^2 + 2l_{21}\sigma_{21} + 1 \\ & l_{32}\sigma_{21} + l_{31} + \sigma_{31} & l_{21}l_{32}\sigma_{21} + l_{21}l_{31} + l_{31}\sigma_{21} + l_{21}\sigma_{31} + l_{32} + \sigma_{32} & & & & & & & \sigma_{21} + l_{31} + \sigma_{31} \\ & & & l_{21}l_{32}\sigma_{21} + l_{21}l_{31} + l_{31}\sigma_{21} + l_{21}\sigma_{31} + l_{32} + \sigma_{32} & & & & & & 2l_{31}l_{32}\sigma_{21} + l_{31}^2 + l_{32}^2 + 2l_{31}\sigma_{31} + 2l_{32}\sigma_{32} + 1 \\ & & & & & & & & & 0 \\ & & & & & & & & & 0 \\ & & & & & & & & & 0 \\ & & & & & & & & & 2l_{21}l_{31}l_{32} - l_{31}^2 - l_{32}^2 + 1 \end{bmatrix}, \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -l_{21}^2 + 1 & 0 \\ 0 & 0 & 2l_{21}l_{31}l_{32} - l_{31}^2 - l_{32}^2 + 1 \end{bmatrix},
\end{aligned}$$

todos los parámetros están dados en términos de los elementos de la matriz  $\mathbf{L}$ ; incluidos los parámetros de correlación, que están dados por

$$\begin{aligned}
\sigma_{21} &= -l_{21}, \\
\sigma_{31} &= l_{21}l_{32} - l_{31}, \\
\sigma_{32} &= l_{21}l_{31} - l_{32}.
\end{aligned}$$

Estos resultados, junto con los presentados en (2.15) permiten reducir el problema de la estimación de la correlación a la estimación de los parámetros  $\gamma$ .

### 2.2.2. Estimación

A partir de estos modelos se propone un algoritmo basado en muestreo Gibbs, similar al propuesto por Albert y Chib (1993), con un paso adicional.

#### Variable aleatoria $z$

Simulamos de una Normal Truncada ( $TN$ ) en 0, secuencialmente con varianza 1 y el parámetro de ubicación como se muestra en (2.17), i.e.

$$\begin{aligned}
z_j &\sim TN(\mu_j + \gamma_j^t(\boldsymbol{\mu}_{<j} - \mathbf{z}_{<j}), 1) \text{ truncada por la izquierda en 0 si } y_i = 1, \\
z_j &\sim TN(\mu_j + \gamma_j^t(\boldsymbol{\mu}_{<j} - \mathbf{z}_{<j}), 1) \text{ truncada por la derecha en 0 si } y_i = 0.
\end{aligned}$$



### Parámetros de la media

Para la simulación de  $\boldsymbol{\beta}$ , se define el modelo condicional

$$\boldsymbol{\Gamma} \mathbf{X}_i \boldsymbol{\beta} = \boldsymbol{\Gamma} \mathbf{z}_i. \quad (2.19)$$

Donde  $\boldsymbol{\Gamma}$  corresponde a la ordenación del vector  $\boldsymbol{\gamma}$  en forma de una matriz triangular inferior con unos en la diagonal principal, como se puede observar en el ejemplo de  $p = 3$ :

$$\begin{bmatrix} 1 & 0 & 0 \\ \gamma_{21} & 1 & 0 \\ \gamma_{31} & \gamma_{32} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1^t & \mathbf{0}^t & \mathbf{0}^t \\ \mathbf{0}^t & \mathbf{x}_2^t & \mathbf{0}^t \\ \mathbf{0}^t & \mathbf{0}^t & \mathbf{x}_3^t \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ \gamma_{21} & 1 & 0 \\ \gamma_{31} & \gamma_{32} & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{x}_1^t & \mathbf{0}^t & \mathbf{0}^t \\ \gamma_{21} \mathbf{x}_1^t & \mathbf{x}_2^t & \mathbf{0}^t \\ \gamma_{31} \mathbf{x}_1^t & \gamma_{31} \mathbf{x}_2^t & \mathbf{x}_3^t \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 + \gamma_{21} z_1 \\ z_3 + \gamma_{31} z_1 + \gamma_{32} z_2 \end{bmatrix}.$$

### Parámetros de correlación

En tanto que para el caso de  $\boldsymbol{\gamma}$ , el modelo condicional es simulando de manera independiente:

$$[\boldsymbol{\mu}_{<j} - \mathbf{z}_{<j}]^t \boldsymbol{\gamma}_j = z_j - \mu_j, \quad (2.20)$$

donde  $\boldsymbol{\gamma}_j^t = (\gamma_{j1}, \dots, \gamma_{j,(j-1)})$ ,  $j > 1$ . Para el caso de  $p = 3$

$$\begin{bmatrix} \mu_1 - z_1 & 0 & 0 \\ 0 & \mu_1 - z_1 & \mu_2 - z_2 \end{bmatrix} \begin{bmatrix} \gamma_{21} \\ \gamma_{31} \\ \gamma_{32} \end{bmatrix} = \begin{bmatrix} z_2 - \mu_2 \\ z_3 - \mu_3 \end{bmatrix}.$$

En este paso se encuentra la innovación de la propuesta, la transformación evita un paso Metropolis, permitiendo obtener muestras de distribuciones conocidas para todos los parámetros del modelo. Las simulaciones multivariadas son únicamente de distribuciones Gaussianas, que son muy sencillas de simular; al ser una adaptación jerárquica del algoritmo univariado de Albert y Chib (1993), la convergencia de la cadena de Markov es muy rápida; todo lo cual implica un costo computacional reducido, principal problema de otros métodos MCMC.

### 2.2.3. Distribuciones a priori

Para las distribuciones condicionales completas en (2.4), existen dos alternativas para definir distribuciones a priori para  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$ , asumiendo independencia entre los parámetros de la media ( $\boldsymbol{\beta}$ ) y los de correlación ( $\boldsymbol{\gamma}$ ). Se define  $\boldsymbol{\Theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{Z}\}$ , y  $\boldsymbol{\lambda}$  dependiendo del paso corresponde a  $\boldsymbol{\beta}$  ó  $\boldsymbol{\gamma}$ . Estas variables tienen distribuciones condicionadas completas:

$$\boldsymbol{\lambda} | \boldsymbol{\Theta}_{-\boldsymbol{\lambda}} \sim N_{k_\lambda}(\hat{\boldsymbol{\lambda}}, \boldsymbol{\Lambda}^{-1}).$$

Si  $\pi(\boldsymbol{\lambda}) \propto 1$ , entonces

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\Lambda}^{-1} \sum_{i=1}^n (\mathbf{X}_i^*)^t \mathbf{y}_i^*, \quad (2.21a)$$

$$\boldsymbol{\Lambda} = \sum_{i=1}^n (\mathbf{X}_i^*)^t \mathbf{X}_i^*. \quad (2.21b)$$

Si  $\pi(\boldsymbol{\lambda}) = N_{k_\lambda}(\boldsymbol{\lambda}_0, \boldsymbol{\Lambda}_0^{-1})$ , entonces

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{\Lambda}^{-1} \left( \boldsymbol{\Lambda}_0 \boldsymbol{\lambda}_0 + \sum_{i=1}^n (\mathbf{X}_i^*)^t \mathbf{y}_i^* \right), \quad (2.22a)$$

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_0 + \sum_{i=1}^n (\mathbf{X}_i^*)^t \mathbf{X}_i^*. \quad (2.22b)$$

Para los parámetros de la media,  $\boldsymbol{\lambda} = \boldsymbol{\beta}$ , acorde a (2.19),  $\mathbf{X}_i^* = \boldsymbol{\Gamma} \mathbf{X}_i$  y  $\mathbf{y}_i^* = \boldsymbol{\Gamma} \mathbf{z}_i$ . Mientras que si  $\boldsymbol{\lambda} = \boldsymbol{\gamma}_j$ , según (2.20), entonces  $\mathbf{X}_i^* = [\boldsymbol{\mu}_{i,<j} - \mathbf{z}_{i,<j}]^t$  y  $\mathbf{y}_i^* = z_{ij} - \mu_{ij}$ . Con de estas distribuciones condicionales se realizan las simulaciones en cada iteración.

## 2.3. Validación del Modelo

A partir de la propuesta desarrollada en la sección anterior, en ésta se presentan ejercicios que permiten visualizar la validez de la propuesta, y algunas de sus propiedades; esto mediante tres ejercicios de simulación y un ajuste del modelo a datos reales.

### 2.3.1. Estudio de Simulación

Para la simulación de datos utilizamos datos normales bivariados ( $p = 2$ ). Se define

$$\mathbf{w}_i \sim MN_2(\mathbf{X}_i\boldsymbol{\beta}^*, \boldsymbol{\Sigma}).$$

y la variable respuesta  $\mathbf{y}_i$  sigue de  $\mathbf{w}_i$  conforme a la ecuación (2.8). En los ejercicios de simulación se hicieron variar los parámetros  $\boldsymbol{\beta}^*$  y  $\boldsymbol{\Sigma}$ , según se explica en cada caso; los valores de las covariables  $\mathbf{X}$  se simularon de manera independiente de una distribución uniforme,  $x_{ij} \sim U(-1, 1)$ . Ambos elementos del vector respuesta tienen la misma estructura de medias, i.e. comparten la misma covariable:  $\mathbf{x}_{i1} = \mathbf{x}_{i2}$ , de dimensión 1, por lo que la matriz diseño correspondiente es

$$\mathbf{X}_i = \begin{bmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{bmatrix}.$$

En cada uno de los ejercicios con datos simulados se compara la propuesta con un modelo donde  $\boldsymbol{\Sigma} = \mathbf{I} \Leftrightarrow \boldsymbol{\gamma} = \mathbf{0}$ , i.e. los elementos de  $\mathbf{y}_i$  son independientes, éste es el caso particular tratado por Albert y Chib (1993). La comparación se realiza en términos de probabilidades estimadas; se define  $\hat{p}_{il} = \Pr(\mathbf{Y}_i = \mathbf{y}_l)$  como la probabilidad estimada por el MVP, y  $\tilde{p}_{il}$  la probabilidad estimada del modelo independiente. Sea  $\mathbf{y}_l$  la respuesta  $l$  de posibles valores que puede tomar la variable  $\mathbf{Y}_i$ , recordando que las variables binarias multivariadas tienen un número finito de posibles respuestas,  $l \in \{1, \dots, 2^p\}$ ; para el caso bivariado  $\mathbf{y}_l \in \{[0, 0]^t, [1, 0]^t, [0, 1]^t, [1, 1]^t\}$ , i.e. cuatro posibles respuestas.

Se realiza la comparación mediante dos métricas de ajuste. La primera medida, dado que se conoce la probabilidad real de los datos simulados, es una razón de errores cuadráticos medios (MSER):

$$MSER = \frac{\sum_l (\sum_{i=1}^n (\tilde{p}_{il} - p_{il})^2 / n)}{\sum_l (\sum_{i=1}^n (\hat{p}_{il} - p_{il})^2 / n)}, \quad (2.23)$$

donde  $\sum_l$  refiere a la sumatoria sobre cada una de las posibles respuestas de interés. El ajuste es mejor si el error es menor. Por lo tanto, si  $MSER > 1$ , el MVP es superior al modelo independiente.

El segundo estadístico corresponde a un Factor Bayes en escala logarítmica (LBF):

$$LBF = \sum_l \ln \left( \prod_{i=1}^{n_l} \frac{\hat{p}_{il}}{\tilde{p}_{il}} \right), \quad (2.24)$$

donde  $n_l$  corresponde al subconjunto de respuestas observadas iguales a la respuesta de interés  $y_l$ .

Debido a que lo deseable es que la probabilidad sea mayor dado los datos (Kass y Raftery, 1995), el MVP es superior si  $LBF > 0$ . La estimación de la probabilidad se realiza mediante método Monte Carlo a partir de una muestra aleatoria de las variables latentes utilizando los parámetros de la cadena resultante del ajuste.

### Ejercicio 1: $\sigma$ fija

En este ejercicio, el parámetro de correlación  $\sigma_{21} = \sigma$  es fijo, y los valores de  $\beta$  seleccionados de tal manera que el intercepto contenga toda la información de la media, y ambos elementos tengan la misma media (2.25),

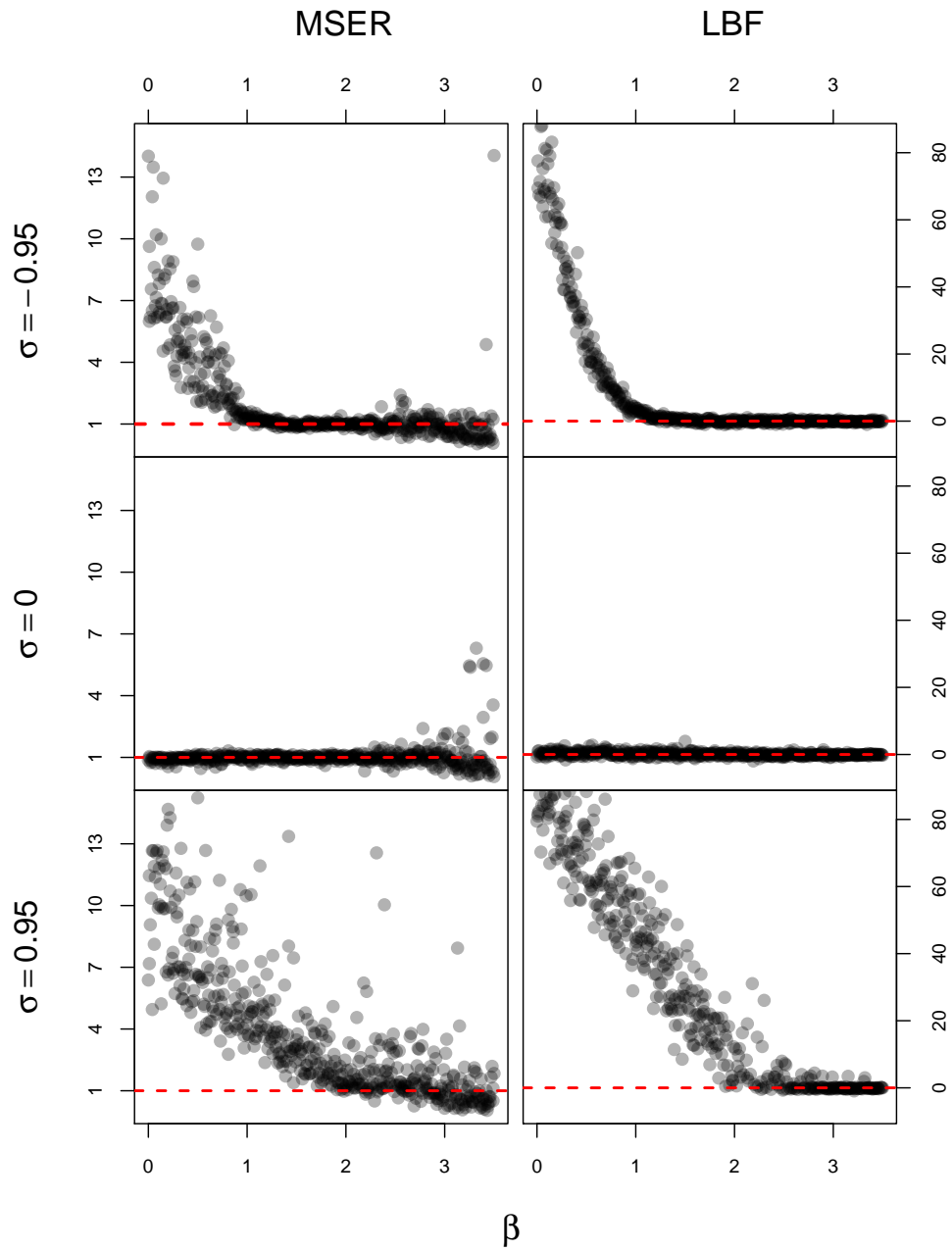
$$\beta = \begin{bmatrix} \beta_t \\ 0 \\ \beta_t \\ 0 \end{bmatrix}, \quad (2.25)$$

de esta manera toda la información de la media está dada por el parámetro  $\beta_t$ , marginalmente  $w_{ij} \sim N(\beta_t, 1)$ . Los valores seleccionados para los parámetros en este ejercicio fueron:

- $\sigma \in \{-0.95, 0, 0.95\}$ ,
- $\beta_t \in \{0, \dots, 3.5\}$ ,  $t \in \{1, 2, \dots, 500\}$ .

Se evaluó el ajuste para tres valores de  $\sigma$  y quinientos valores de  $\beta_t$ , en el intervalo  $[0, 3.5]$ . En cada simulación se generaron muestras de tamaño  $n = 250$ . Se ajustó con una cadena de 6000 iteraciones, y en la estimación de las probabilidades se utilizó la información de las últimas 5000 iteraciones.

Los resultados de las medidas de ajuste (MSER y LBF) se muestran en la figura 2.1. Ambas muestran comportamientos similares, aunque con mayor variabilidad del MSER. Cuando existe una correlación fuerte ( $|\sigma| = 0.95$ ) se

Figura 2.1: MSER y LBF,  $\sigma$  fija.

Cuadro 2.1: Impacto de correlación en la probabilidad según la media

	$\mu_1 = \mu_2 = 0.5$		$\mu_1 = \mu_2 = 3.5$	
	$\sigma = 0$	$\sigma = 0.95$	$\sigma = 0$	$\sigma = 0.95$
$\Pr(w_1 \leq 0 \cap w_2 \leq 0)$	0.0952	0.2640	0.0000	0.0001
$\Pr(w_1 \leq 0 \cap w_2 > 0)$	0.2133	0.0446	0.0002	0.0001
$\Pr(w_1 > 0 \cap w_2 \leq 0)$	0.2133	0.0446	0.0002	0.0001
$\Pr(w_1 > 0 \cap w_2 > 0)$	0.4781	0.6469	0.9995	0.9997

observa que, al aumentar la media ( $\beta$ ) la diferencia entre los ajustes disminuye. Mientras que el LBF converge a 0, el MSER continúa disminuyendo.

Este comportamiento es acorde a lo esperado; al trasladarse la distribución lejos del 0 -aumento o disminución de la media-, el impacto que puede tener la correlación es despreciable en términos de  $\Pr(\mathbf{w}_i)$ , esto es fácilmente ejemplificable.

El cuadro 2.1 muestra las probabilidades de interés  $\Pr(\mathbf{Y}_i = \mathbf{y}_l)$  en términos de la variable latente  $\mathbf{w}$ ,  $w_j \in \mathbf{w}$ , en el caso bivariado, cuando el vector de medias tiene un valor pequeño,  $\mu_j = 0.5$ , la diferencia de las probabilidades cuando la correlación va de 0 a 0.95 es de 0.17,

$$|\Pr(\mathbf{Y}_i = \mathbf{y}_l | \sigma = 0) - \Pr(\mathbf{Y}_i = \mathbf{y}_l | \sigma = 0.95)| = 0.1688;$$

mientras que cuando la distribución conjunta es trasladada a una media mayor,  $\mu_j = 3.5$ , la diferencia es de apenas 0.0001, siendo despreciable en términos prácticos, sugiriendo que la variable respuesta está bien explicada con la información de la media.

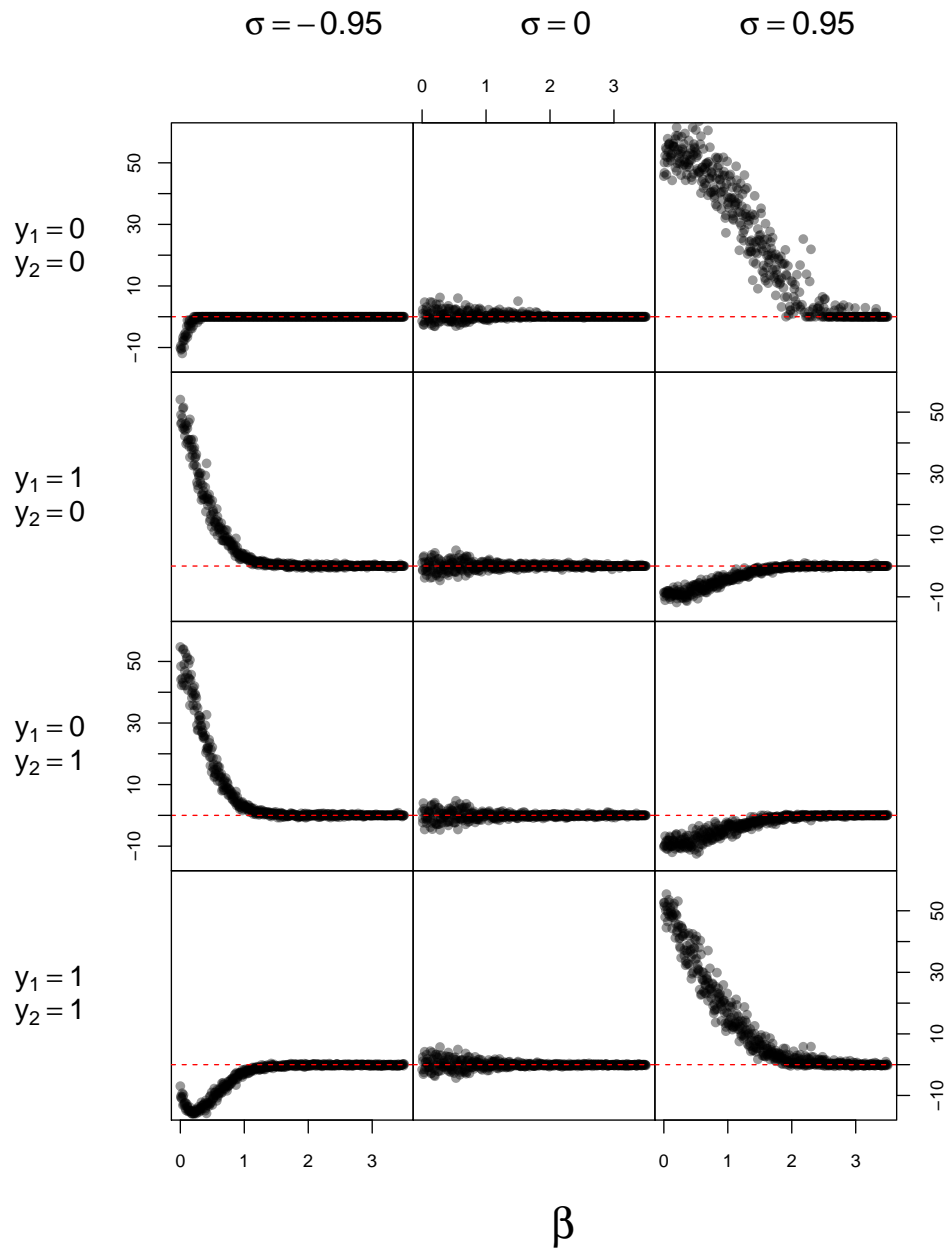
En el caso independiente ( $\sigma = 0$ ) ambos ajustes son equivalentes, como se puede observar en la segunda fila de la Figura 2.1, aún cuando el MSE muestra un aumento de variabilidad conforme la media aumenta.

### *LBF por casos*

Es posible separar los estadísticos de ajuste por casos, dado que  $p = 2$  el vector respuesta sólo puede tomar cuatro valores. Se calculó el LBF para cada uno de estos casos<sup>10</sup>, los resultados se muestran en la Figura 2.2, únicamente se incluye el LBF.

<sup>10</sup>Estos casos corresponden al índice  $l$  en la definición del MSER y LBF presentadas en las expresiones (2.23) y (2.24)

Figura 2.2: LBF por casos,  $\sigma$  fija.



Los ajustes donde  $\sigma = 0$  muestran el mismo comportamiento sin importar la respuesta observada  $\mathbf{y}_i$ , en todos los casos el  $LBF \approx 0$ . Cuando la correlación es positiva (tercera columna de la figura), el MVP muestra mejor ajuste cuando las respuestas son las mismas para ambos componente de  $\mathbf{y}_i$ , i.e. ambos son 0 ó 1 (1er y 4er fila de la figura), mientras que donde no coinciden las respuestas observadas, una es 0 y la otra 1, el modelo independiente muestra un mejor ajuste. Con correlación negativa (primera columna), la tendencia es inversa, resultando un mejor ajuste del MVP para los casos donde las medias no son coincidentes (2a y 3er fila).

Para todos los casos las diferencias de ajuste disminuyen conforme aumenta el valor de la media, situación previamente explicada. El LBF es una medida de la probabilidad dado los datos, cuando la correlación es fuerte en sentido positivo una respuesta no concordante es menos probable que una concordante y viceversa, véase el cuadro 2.1, específicamente cuando  $\mu_j = 0.5$  y  $\sigma = 0.95$ . Los casos menos probables dados los parámetros de la simulación son también aquellos donde el modelo independiente muestra mejor ajuste; esta situación está incluida en el BF mediante la parte aditiva de su construcción, estos casos son más escasos y no suman tanto en el BF conjunto (Fig. 2.1).

Este ejercicio nos permite observar que el MVP es preferible al independiente cuando las medias no son tan grandes como para que la covarianza sea no significativa, i.e. cuando las medias sea de suficiente magnitud como para explicar las probabilidades, o cuando no se incluyan covariables significativas; el caso contrario con una media grande, los modelos son casi equivalentes, los estadísticos de ajuste tienden a los valores de decisión:  $MSER = 1$  y  $LBF = 0$ . En términos de análisis de datos significa que cuando las covariables definen casi por si mismas la variable respuesta, i.e. que las covariables tienen un gran impacto en probabilidad marginal de cada elemento de la variable respuesta -sus medias ( $|\mu_j|$ ) son grandes - dejan poca información sin explicar, de tal forma que los parámetros de correlación son poco significativos.

### Ejercicio 2: $\beta$ fija

El segundo ejercicio considera los siguientes valores para los parámetros:



- $\beta \in \left\{ \begin{bmatrix} -0.5 \\ 0 \\ -0.5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0 \\ -0.5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0 \\ 0.5 \\ 0 \end{bmatrix} \right\},$
- $\sigma_t \in (-1, \dots, 1), t \in \{1, 2, \dots, 500\};$

el parámetro de correlación es evaluado en todo el intervalo de valores posibles, en quinientos valores distintos. El parámetro  $\beta$  únicamente toma tres valores. Esta conformación del vector de medias permite que nuevamente toda la información de la media se encuentre contenida en el parámetro  $\beta$ . Las covariables no intervienen en la probabilidad; sin embargo al realizar el ajuste del modelo sí se estiman coeficientes para estas covariables. Para cada combinación de parámetros se simuló una muestra de tamaño  $n = 250$ . Los resultados se muestran en la Figura 2.3.

Los valores de  $\beta_j \neq 0 \Rightarrow \beta_j \in \{-0.5, 0.5\}$  se eligieron debido a que son lo suficientemente pequeños para permitir que el parámetro de correlaciones tenga impacto en la probabilidad; pero lo suficientemente grande como para intervenir significativamente en la definición de la probabilidad.

Las medidas de ajuste muestran comportamientos similares, con una mayor variabilidad del MSER. El resultado de las simulaciones es el esperado en términos de la correlación. Cuando  $\sigma \rightarrow 0$  ambos modelos muestran ajustes similares, aglutinándose entorno al valor de decisión ( $MSER = 1, LBF = 0$ ). Mientras que cuando  $|\sigma| \rightarrow 1$  las medidas de ajuste aumentan, i.e. el ajuste del MVP es mejor que el del modelo independiente.

Cuando el sentido de la correlación es apoyado por el vector de medias, i.e. con una correlación positiva y ambos elementos de  $\beta$  son del mismo signo, el ajuste del MVP es mejor que en el caso contrario. Aún en los casos cuando la relación entre los elementos del vector de medias es contrario a la planteada por la correlación, el MVP muestra mejor ajuste. Esto explica la falta de simetría en las gráficas de la Figura 2.3.

#### *LBF por casos*

De manera similar al ejercicio 1, se calculó el BF por casos. En la Figura 2.4 se observa que cuando la correlación tiende a cero los modelos muestran ajustes equivalentes ( $LBF = 0$ ).

La variable respuesta observada ( $\mathbf{y}_t$ ) determina la dirección de la curva, no el valor de la media; cuando el signo de la correlación coincide con la

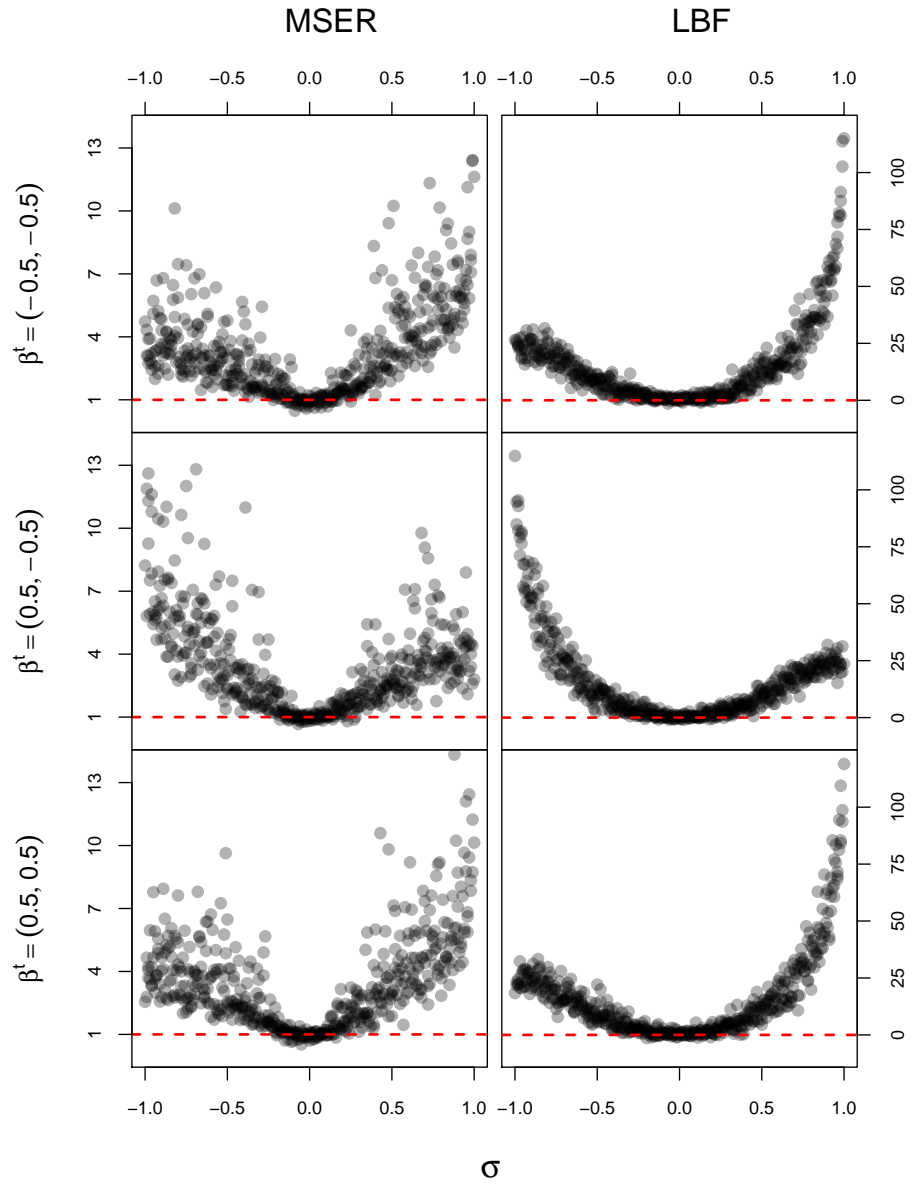
Figura 2.3: MSER y LBF,  $\beta$  fija

Figura 2.4: LBF por casos,  $\beta$  fija.

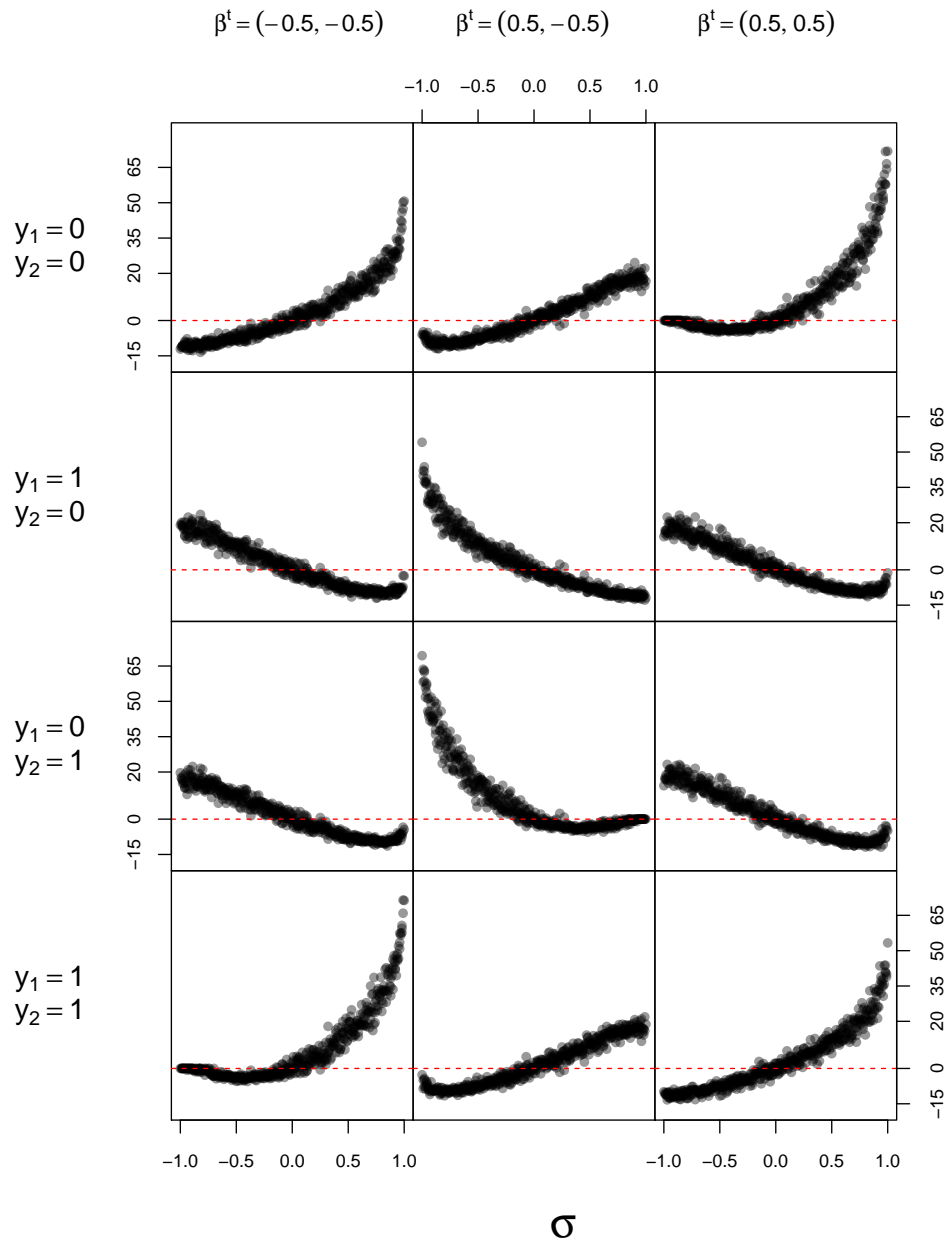
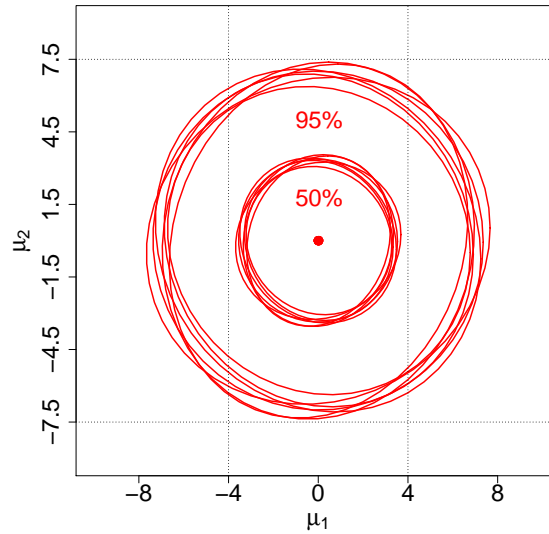


Figura 2.5: Espacio muestral de las medias.



combinación de respuestas observadas ( $\sigma > 0$  si  $y_1 = y_2$  ó  $\sigma < 0$  si  $y_1 \neq y_2$ ) el  $LBF > 0$ . Mientras que en caso contrario, el modelo independiente muestra mejor ajuste.

En los dos primeros ejercicios de simulación, los modelos de los datos fueron simplificados de tal manera que las covariables no se incluyen en la definición de la media, por tanto las observaciones simuladas en cada corrida provienen de la misma distribución. Sin embargo, en casos prácticos esto rara vez sucede.

### Ejercicio 3: $\beta$ aleatoria, $\sigma$ seleccionada

En el último ejercicio, el vector de medias de cada observación es aleatorio, y los valores de  $\sigma$  se fijan a partir de valores de interés:

- $\beta = \begin{bmatrix} 0 \\ \beta_1 \\ 0 \\ \beta_2 \end{bmatrix}$ , donde  $\beta_j \sim N(0, 5)$ ,  $j = 1, 2$ ;
- $\sigma \in \{-0.9, -0.6, -0.4, 0, 0.2, 0.5, 0.75\}$ .

La parte lineal de la simulación no contiene intercepto, quedando en función de los parámetros  $\beta_j$  y la covariable  $x_i$ ,  $\mu_{ij} = x_i\beta_j$ . Se simulan en cada ajuste  $n = 250$  observaciones independientes, y para cada valor seleccionado de  $\sigma$  se realizan 250 simulaciones distintas.

Dado esta estructura, las medias se encuentran en el espacio mostrado en la Figura 2.5; contiene una elipse para cada cada valor de  $\sigma$  (7 elipses), ajustada cada una a partir de  $250 \times 250$  vectores de medias. Se observa que la mitad de las medias se encuentran en un círculo de contenido en el cuadrado  $(-4, 4)$ , y el 95% de las medias en el círculo de radio 7 y centrado en  $(0, 0)$  [que corresponde a  $\boldsymbol{\mu}_i \sim MN_2(\mathbf{0}, \sigma_1^2 = \sigma_2^2 = 2.5, \rho = 0)$ ]. Aproximadamente la mitad de los valores tienen una media que, según lo observado en los dos ejercicios previos, permite que la correlación sea relevante. Se eligió esta estructura para emular casos más cercanos a los que se espera encontrar en la realidad.

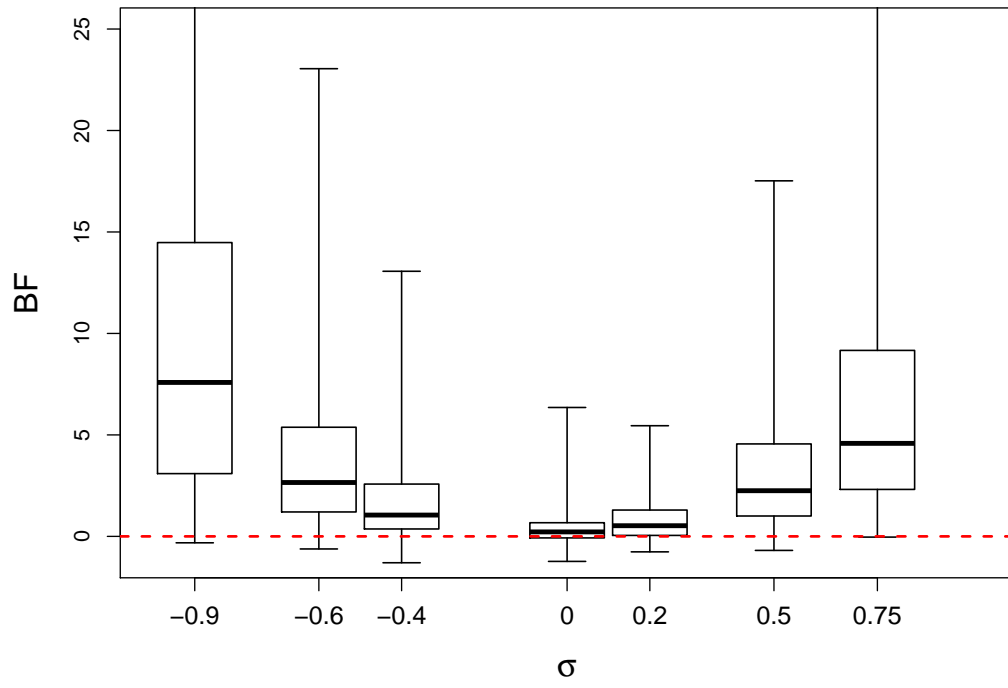
La selección de valores de  $\sigma$  se realizó con la intención de representar el espacio paramétrico de la correlación. Únicamente se utilizaron siete valores ya que el comportamiento es simétrico en  $\sigma \in (-1, 0)$  y  $\sigma \in (0, 1)$ . Para cada ajuste se calculó el LBF que comprara el MVP con el modelo independiente. La Figura 2.6 muestra en forma de gráficas de cajas y bigotes, la distribución de la medida de ajuste para cada valor seleccionado de  $\sigma$ .

Según lo esperado se observa que para todos los casos la mayor parte de las muestras simuladas se encuentran en  $BF > 0$ , indicando un mejor ajuste del MVP. Incluso cuando  $\sigma = 0$  el ajuste resultó mejor, se espera que conforme aumente el número de simulaciones, la distribución del LBF en estos casos tenderá a centrarse en 0.

Conforme el valor de la correlación se aleja de 0, la distribución del BF también se aleja del valor de decisión 0; es de resaltar que los valores más alejados de la media de cada distribución se encuentra en  $LBF > 0$ . Se concluye que cuando la correlación ( $|\sigma|$ ) es grande, el MVP tiene mejor ajuste; incluso en casos donde  $\sigma = 0$ , el ajuste es equivalente al modelo independiente. Aunque el modelo independiente es un caso particular del MVP (véase anexo B.1.2), en casos donde se identifique una correlación no significativa debería de preferirse el modelo más sencillo.

### 2.3.2. Estudio de seis ciudades

El presente ejemplo consiste en un estudio longitudinal sobre los efectos de la contaminación del aire en infantes. Los datos se toma de Chib y Greenberg

Figura 2.6: Intervalos,  $\sigma$  específicos.

(1998), quienes lo utilizan como ejemplo para su método; otro análisis de estos datos también se encuentra en Street y Czado (1996).

The data ... contain repeated binary measures of the wheezing status (1 = yes, 0 = no) for each of 537 children from Stuebenville, Ohio, at ages 7, 8, 9 and 10 years. The objective of the study is to model the probability of wheezing status over time as a function of a binary indicator variable representing the mother's smoking habit during the first year of the study and the age of the child (Chib y Greenberg, 1998)<sup>11</sup>.

Se ajustaron dos modelos, el completamente multivariado y el modelo independiente (i.e.  $\sigma_{ij} = 0$ ). Las medias de cada elemento del vector respuesta se especificaron como

$$\mu_{ij} = \beta_{j0}^* + \beta_{j1}^*x_{i1} + \beta_{j2}^*x_{i2} + \beta_{j3}^*x_{i3},$$

según la parametrización de Chib y Greenberg (1998), i.e. previa transformación. Donde

- $x_{1i}$  es la edad del  $i$ -ésimo niño centrada en 9,
- $x_{2i}$  es la condición de fumadora de la madre del  $i$ -ésimo niño (1=sí, 0=no),
- $x_{3i}$  es la interacción entre la condición de fumar de la madre y la edad del niño.

Para ambos modelos se utilizaron los siguientes hiperparámetros a priori

$$\beta_0 = \mathbf{0} \quad , \quad \mathbf{B}_0^{-1} = 10 \times \mathbf{I}_{16 \times 16};$$

$$\gamma_0 = \mathbf{0} \quad , \quad \mathbf{\Gamma}_0^{-1} = 10 \times \mathbf{I}_{6 \times 6}.$$

Se ajustaron los modelos utilizando una cadena de 10500 iteraciones, y para el análisis se utilizaron las últimas 10000. Los resultados se presentan

---

<sup>11</sup>Los datos ... contienen observaciones binarias repetidas de la presencia de sibilancia (1=sí, 0=no) para cada uno de los 537 niños de Stuebenville, Ohio, con edades de 7, 8, 9 y 10 años . El objetivo del estudio es modelar la probabilidad de la situación de sibilancia en el tiempo como una función de una variable indicadora binaria que representa el hábito de fumar de la madre durante el primer año del estudio y la edad del niño

Cuadro 2.2: Valores estimados de  $\Sigma$ 

	HPD 0.95			
	Media	Mediana	Inferior	Superior
$\sigma_{21}$	0.319	0.322	0.198	0.427
$\sigma_{31}$	0.274	0.276	0.151	0.390
$\sigma_{32}$	0.420	0.421	0.311	0.531
$\sigma_{41}$	0.287	0.288	0.147	0.428
$\sigma_{42}$	0.300	0.305	0.098	0.486
$\sigma_{43}$	0.303	0.308	0.087	0.498

en la parametrización original  $(\beta^*, \Sigma)$ , en lugar de la parametrización de la transformación  $(\beta, \gamma)$ , con la finalidad de interpretación del análisis y comparación con los resultados obtenidos por Chib y Greenberg (1998).

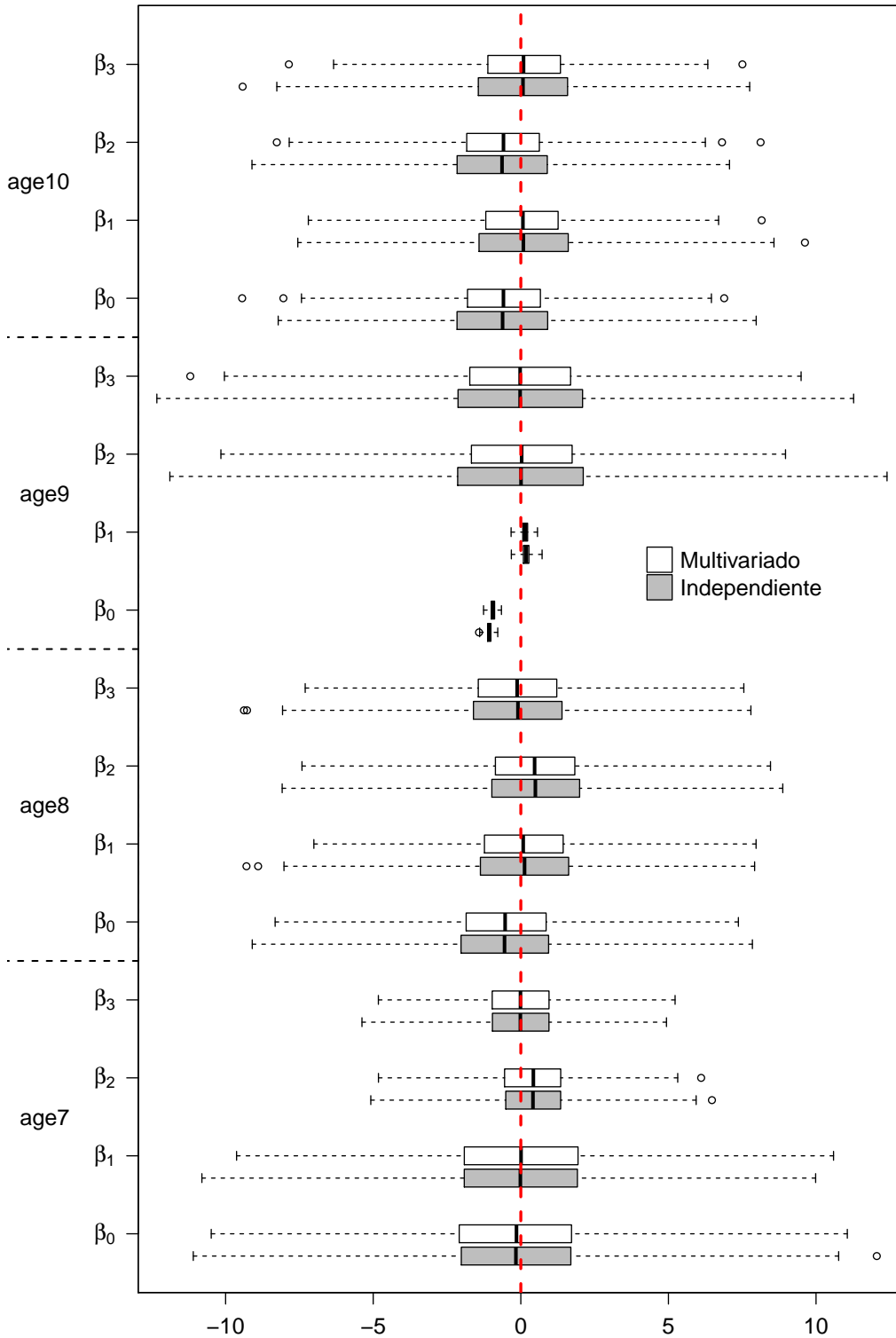
La figura 2.7 muestra los resultados para los parámetros de media  $(\beta^*)$ , y compara el modelo independiente con el multivariado. Se observa que las distribuciones a posteriori de casi todos los parámetros (exceptuando el intercepto para la edad de 9 años) se encuentran prácticamente centradas en cero, lo cual indica que estas covariables no aportan información relevante para la probabilidad de presentar la condición. Estos resultados coinciden con lo presentado por Chib y Greenberg (1998). Comparativamente se observa que consistentemente la variabilidad de los parámetros en el modelo multivariado es menor a la presentada en el modelo independiente.

Las estimaciones de los parámetros de correlación se muestran en el cuadro 2.2, ahí se muestran la media, la mediana y los límites de los intervalos de High Posterior Density (HPD) al 95 %. Para cada uno de los parámetros de correlación, la probabilidad de ser mayor a 0 es de al menos 0.95, que indica existe correlación positiva entre los elementos del vector respuesta. Los resultados difieren en magnitud a los obtenidos por Chib y Greenberg (1998), esto se debe a que ellos ajustaron un caso particular del modelo; los parámetros de lineales  $(\beta_0^*, \dots, \beta_3^*)$  son constantes entre los años y los parámetros de correlación son iguales entre elementos ( $\sigma_{21} = \sigma_{31} = \dots = \sigma_{43}$ ); estos casos aún no están implementados en esta propuesta (vease anexo B.1.2).

Para evaluar el ajuste entre ambos modelos se utiliza el LBF. En este caso, no sólo se evaluó el ajuste para 1) todas las posibles combinaciones de respuesta ( $\mathbf{y}_l$ ), i.e. las probabilidades conjuntas, también se evaluaron ajustes para tres preguntas de posible interés: 2) Las probabilidades marginales para



Figura 2.7: Valores estimados de  $\beta$



Cuadro 2.3: LBF en estudio de seis ciudades

	<b>LBF</b>
1) Conjunta	94.624
2) Marginal	1.206
3) >1 año	79.670
4) Consecutivas	81.858

cada edad [ $\Pr(y_{ij} = 1)$ ], 3) la probabilidad de presentar en más de un año el síntoma [ $\Pr(\mathbf{1}^t \mathbf{y}_i > 1)$ ] y, 4) la probabilidad de presentar en al menos dos años consecutivos el síntoma [ $\Pr(y_{i,j} = 1, y_{i,j+1} = 1)$ ]. Cada una de estas probabilidades son subconjuntos de las posibles respuestas conjuntas ( $\mathbf{y}_i$ ). Los LBF calculados para cada uno de estos conjuntos se presentan en el cuadro 2.3.

Los resultados de la comparación de ajustes, mostrados en el cuadro 2.3, indican que en todos los casos, el modelo multivariado presenta un mejor ajuste que el modelo independiente. El caso donde la diferencia entre ajustes es menor, es el de las probabilidades marginales, como es de esperarse (vease anexo A.1.3). Estos conjuntos de posibilidades son de interés no solo hipotéticamente para el estudio en niños, sino que preguntas similares se utilizan para caracterizar la pobreza multidimensional en el próximo apartado. Al incluirlos en esta sección se buscó mostrar en un caso real que el modelo es una alternativa útil para este tipo de preguntas.

Los estudios de simulación y el estudio de caso presentado en este capítulo permiten concluir que el método propuesto para ajustar un modelo MVP tiene propiedades deseables. Comparativamente el MVP en términos de ajuste es al menos igual de bueno que el modelo independiente. En cuanto a la interpretación de las estimaciones, el método permite regresar a la parametrización original en términos de una distribución Gaussiana Multivariada, pero sin necesidad de recurrir a ningún paso Metropolis. Esto hace de la propuesta una alternativa atractiva para modelar datos binarios multivariados dado que es computacionalmente eficiente.

# Capítulo 3

## 3.1. Modelación de la PM

La PM es estimada a nivel estatal cada dos años mediante la Encuesta de Ingreso y Gasto en los Hogares y el Módulo de Condiciones Socioeconómicas (ENIGH-MCS); en el ajuste aquí presentado se utiliza la base de datos correspondiente al levantamiento realizado en 2010 (INEGI, 2011), así como la base sistematizada que realiza el CONEVAL para la estimación de la pobreza (CONEVAL, 2013b). Únicamente se utilizó la información para el estado de Querétaro y las personas de 15 años y más, que representan 5,293 casos completos.

### Variables

Se integró una tabla utilizando variables de ambas fuentes, las covariables incluidas en el modelo se muestran en el cuadro 3.1, donde “X” indica que la variable fue incluida como variable nominal y “\*” indica que la variable fue incluida como variable continua, sea originalmente continua u ordinal; el modelo incluye en total 56 coeficientes de regresión  $\beta$ , para un vector respuesta de dimensión siete. La descripción de cada una de las variables en el cuadro se puede consultar en el anexo C.1,

### Distribuciones apriori

Para  $\beta$  se utilizó una distribución apriori difusa:

$$\pi(\beta) = MN_{56}(\mathbf{0}, \mathbf{I} \times 10).$$

Cuadro 3.1: Covariables incluidas en el modelo de PM

Covariable	Vector respuesta						Ingreso
	ic serviv	ic educa	ic salud	ic calviv	ic segsoc	ic alimenta	
eco					x	x	
drenaje_d				x			x
piso_d						x	
piso							*
agua_d				x		x	x
tamaño	x			*	*		x
habitantes	*			*		*	*
sexo_jefe				x	x	x	x
educa_jefe	*			*		*	*
salud							x
asistencia						x	
educa					x		*
indigena		x	x	x	x	x	x
edad		*	*		*	*	
sexo		x	x		x		

Cuadro 3.2: Información a priori de la correlación

		Media	$q$	$\Pr(\sigma < q)$
ic educa	ic serviv	0	0.95	0.832
ic salud	ic serviv	0	0.95	0.832
ic salud	ic educa	0.3	0	0.3
ic calviv	ic serviv	0.8	0.2	0.15
ic calviv	ic educa	0	0.95	0.832
ic calviv	ic salud	0	0.95	0.832
ic segsoc	ic serviv	0.8	0.2	0.15
ic segsoc	ic educa	0.4	0.2	0.2
ic segsoc	ic salud	0.2	0	0.3
ic segsoc	ic calviv	0.4	0.2	0.2
ic alimenta	ic serviv	0.8	0.2	0.15
ic alimenta	ic educa	0.6	0	0.2
ic alimenta	ic salud	0.2	0	0.3
ic alimenta	ic calviv	0.6	0	0.2
ic alimenta	ic segsoc	0.7	0	0.2
Ingreso	ic serviv	0.8	0.2	0.15
Ingreso	ic educa	0.4	0.2	0.2
Ingreso	ic salud	0.4	0.2	0.2
Ingreso	ic calviv	0.7	0	0.2
Ingreso	ic segsoc	0.8	0.2	0.15
Ingreso	ic alimenta	0.6	0	0.2

Mientras que para los parámetros de correlación  $\gamma$ , se utilizaron distribuciones a priori informativas basadas en conocimiento intuitivo del problema, que corresponden a las características conocidas de los parámetros  $\sigma_{jk}$  mostradas en el cuadro 3.2. Suponiendo independencia a priori de los parámetros  $\gamma_{jk}$ ; sobre este modo de inclusión de información a priori en la distribución para  $\gamma$  se puede consultar el anexo B.1.1.

## Resultados

Se generó una cadena de 11000 iteraciones, y se utilizaron las últimas 10000 para las estimaciones. Estadísticos de las distribuciones a posteriori

de los parámetros pueden ser consultados en el anexo C.1. Adicionalmente se ajustó un modelo que no contenía covariables, únicamente intercepto. Al calcular el LBF para la probabilidad de ser definido como pobre, i.e.

$$\Pr(y^{pm} = 1) = \Pr\left(\sum_{j=1}^6 y_j > 0 \wedge y_7 = 1\right);$$

que equivale a lo descrito en la Figura 1.2, i.e. tener al menos una carencia y un ingreso menor a la línea de bienestar. Al comparar estas probabilidades mediante el factor Bayes se obtuvo:  $LBF = 363.5$ , indicando que las covariables en conjunto tienen impacto significativo en el ajuste.

Utilizando las distribuciones a posteriori del modelo ajustado se estimó la probabilidad de estar en situación de pobreza para cada una de las observaciones del Censo de Población y Vivienda 2010 (CPV2010) en Querétaro disponibles en la página del INEGI<sup>1</sup>.

A partir de la estimación de la probabilidad de PM para cada uno de los casos, se relacionó la probabilidad en tablas de frecuencia para cada uno de los 18 municipios del estado. Para definir si una persona se identifica como pobre, fue necesario determinar un límite de la probabilidad. Para esto se fijó que la estimación coincidiera con la estimación puntual de población pobre mayor a 14 años en Querétaro según la ENIGH-MCS 2010, i.e. 38.5 % de la población; para las probabilidades estimadas esto equivale a

$$\Pr(y_i^{pm} = 1) > 0.547 \Rightarrow \hat{y}_i = 1,$$

i.e. si la probabilidad de ser pobre es mayor a 0.547 se definió como pobre; esta restricción se basó en lo realizado por CONEVAL (2011)<sup>2</sup>. Utilizando el límite citado, se obtuvieron estimaciones municipales, las cuales se comparan con las estimaciones para pobreza municipal de población total de CONEVAL, los resultados se encuentran en el cuadro 3.3.

En general los porcentajes son similares, pero existen diferencias importantes en algunos casos. Sin embargo, este ejercicio muestra que el MVP es una alternativa viable para la estimación de la PM.

<sup>1</sup>Se utilizó la información en forma de tabla de frecuencia de casos completos, obtenida mediante la consulta interactiva; se tratan de 279,310 casos diferentes.

<sup>2</sup>Sin embargo, si el objetivo es una estimación operativamente funcional, en lugar de este procedimiento que busca “cuadrar” cifras oficiales, se puede recurrir a tablas de decisión, a partir de costos operativos y costos simbólicos; sin embargo, la información requerida depende de los requerimientos operativos que se le quiera dar a la estimación.

Cuadro 3.3: Estimación de población en situación de pobreza, 2010

Municipio	% Población total (CONEVAL)	% Población $\geq 15$
1 Amealco de Bonfil	76.60	80.63
2 Pinal de Amoles	84.60	79.37
3 Arroyo Seco	72.10	80.92
4 Cadereyta de Montes	72.90	76.41
5 Colón	55.60	69.20
6 Corregidora	18.70	21.77
7 Ezequiel Montes	62.70	68.60
8 Huimilpan	67.30	64.10
9 Jalpan de Serra	63.40	62.46
10 Landa de Matamoros	81.20	77.81
11 El Marqués	51.10	54.12
12 Pedro Escobedo	47.70	62.14
13 Peñamiller	70.10	80.66
14 Querétaro	29.50	21.37
15 San Joaquín	75.80	68.99
16 San Juan del Río	36.20	64.04
17 Tequisquiapan	55.40	53.77
18 Tolimán	69.30	83.41

Adicionalmente CONEVAL acostumbra presentar estimaciones adicionales: número de carencias promedio, pobreza extrema, vulnerables por carencia social, vulnerables por ingreso, no pobres y no vulnerables, y estimación para cada una de las seis carencias; el MNP a partir del mismo ajuste permite realizar estas estimaciones, exceptuando el de la pobreza extrema para el cual sería necesario realizar un ajuste adicional, o incluir un elemento adicional en el vector respuesta (línea de bienestar mínimo).

## 3.2. Conclusiones

En general el método propuesto para ajustar un MVP, presenta características atractivas, siendo la más relevante que este método MCMC tiene en cada iteración una probabilidad de aceptación de 1, haciéndolo compu-

tacionalmente eficiente. El MVP es una alternativa para modelar la PM en México y realizar estimaciones a niveles operativamente relevantes, sin embargo aún hay trabajo por realizar para poder obtener modelos más precisos y a la vez a niveles más locales como AGEBS, que depende de la información disponible.

### 3.2.1. Trabajo futuro

Esta primera aproximación a la modelación de la PM mediante un MVP, evidencia su viabilidad para abordar el problema, sin embargo para una mejor y más detallada implementación del modelo a la problemática queda aún trabajo pendiente, y problemas particulares que requieren atención:

- La fuente de la información para el ajuste del modelo, la ENIGH-MCS, tienen un diseño muestral complejo, multietápico, que no fue considerado para este primer ajuste. Para realizar una mejor estimación es importante incluir el diseño muestral en el ajuste (véase anexo B.1.3). Así mismo, se consideró a personas dentro de una misma vivienda como observaciones independientes, cuando es evidente que no es el caso.
- La implementación presentada se encuentra a nivel municipal, sin embargo operativamente este nivel de desagregación geográfica es poco útil, la siguiente unidad de utilidad es el área geoestadística básica (AGEB), si bien el CPV2010 recabó información hasta este nivel, no se encuentra disponibles<sup>3</sup>, lo que sí se encuentra disponible, con sus trámites burocráticos de por medio, es la muestra censal, que además de contener variables adicionales a las del censo, contiene información de la ubicación geográfica de los datos. Si bien fue diseñada para realizar estimaciones municipales puede ser utilizada para realizar las estimaciones de los AGEBS en muestra; o implementar una versión del modelo que incluya correlación espacial para poder realizar estimaciones para espacios cuya información no esté disponible en la muestra.
- Incluir información geográfica en el modelo, es decir considerar que existe una correlación espacial entre los puntos, permitiría realizar dos

---

<sup>3</sup>El INEGI tiene políticas de disponibilidad de información kafkianas, por alguna razón restringe esta información incluso para instituciones públicas federales, contribuyendo así que la información se utilice poco, claro que también se quejan en el instituto de que la información se utiliza poco.



tipos de estimaciones adicionales, 1) cada dos años a partir de la información de la ENIGH-MCS se podría establecer estimaciones para todo el territorio nacional, sin depender de los censos de población; 2) a partir de la información de las muestras censales cada 5 años se puede realizar estimación para todo el territorio nacional, a nivel de localidades rurales o AGEBs urbanos, unidades que son operativamente relevantes.

- Debido a la definición oficial de PM dependiendo de la edad hay criterios diferentes para algunos de los indicadores de carencias, pero también existe problema para las covariables incluidas en el modelo, particularmente aquellas sobre escolaridad y condición de actividad económica, no son relevantes para niños; esta complejidad adicional es lo que nos llevó a realizar este primer ajuste únicamente para personas de 15 años o más. Es necesario realizar definiciones y transformaciones de las covariables, para poder incluir a toda la población de manera coherente.
- La implementación de un proceso, automático o manual, de selección de variables mejoraría el ajuste. Sin embargo, debido a la cantidad de información disponible la implementación del método es computacionalmente demandante; en esta primera aproximación el ajuste implicó aproximadamente 10 hrs. de cómputo con un procesador intel core i7; mientras que el cálculo de la probabilidad estimada requirió aproximadamente 20 GB de memoria RAM. Por lo que es necesario no solamente contar con equipo de cómputo suficientemente potente sino además utilizar un método de selección de variables que requiera pocos ajustes, los métodos automáticos tradicionales no son una buena opción.
- En el mismo sentido, cuando se trabajan grandes volúmenes de datos, es factible implementar una versión de cómputo paralelo automático; se realizó una versión manual de esto, partiendo la cadena de 10000 requerida en tres partes más su correspondiente burnin; recortando así el tiempo de cómputo, sin embargo una versión automática haría más atractiva la implementación.
- De entre los elementos binarios del vector respuesta existe un caso particular, el ingreso. El modelo considera que existen variables latentes

que determinan las variables binarias, en el caso del ingreso en realidad existe una variable continua dicotomizada, que es visible en los datos de la encuesta y a partir de cierto límite es dicotomizada, la línea de bienestar o la línea mínima de bienestar en el caso de la pobreza extrema. Sin embargo, al dicotomizarse la variable, un tipo de censura, se pierde información; utilizar la variable dicotómica en lugar de la continua, implica desaprovechar información. Es posible incluir la variable continua en el ajuste del modelo, como valores fijos, estandarizando acorde a la línea de bienestar correspondiente. La modificación del algoritmo es mínima y podría traer mejores ajustes; sin embargo sabemos que típicamente el ingreso, y más en un país como México, no sigue una distribución normal, sino una con colas más pesadas (desigualdad), y esto debería ser tomado en cuenta.

### 3.3. Carta sociológica

Este documento es suficiente para los objetivos planteados, pero presenta carencias en dos sentidos: el impacto y alcance de la investigación primero y segundo sobre su posición relativa en términos de perspectiva epistemológica; esta última sección busca solventar estas carencias.

Fuera de las Ciencias Sociales (CS) las consideraciones epistémicas se encuentran aisladas de las preocupaciones prácticas entre quien hace ciencia, al grado que en las instituciones científicas la reflexión epistemológica está ausente; canonicamente se adopta una perspectiva maniquea hace décadas detractada por sus mismo propositores (Carnap, 1989), el llamado *Método Científico* es principio de autoridad. En contraparte, en México gran parte de la comunidad de profesionales de las CS adopta un discurso basado en dialéctica erística (Schopenhauer, 2002), dónde la evidencia empírica es relegada; reducida a métodos antropológicos, ejecutados de manera poco rigurosa, o llevados in extremis a meros anecdóticos; o en otros casos la ideología como sustituto de la evidencia.

Estos extremos cada vez más alejados han permitido que la Economía ocupe un lugar central; con base en la exclusión del resto de las CS, que ya no la considera CS, mientras que muchos economistas no se consideran tampoco parte de las CS. Esta exclusión no tiene fundamento metodológico, sino meramente técnico, la gloria de la Economía contemporánea es la econometría; que gana presencia ante la inclusión de técnicas matemáticas

ante la carencia de reflexión epistémica y el miedo de los *científicos* sociales a utilizar estas técnicas, permitiendo una opacidad y falta de rigor en los trabajos y aplicaciones de buena parte de la aportación econométrica; aunado a la exigencia de la ciencia institucionalizada de publicación sobre cualquier otro criterio; esto sin desconocer la presencia de grandes investigaciones econométricas.

En este contexto el problema de interés en este estudio no es la pobreza, es la modelación de la pobreza multidimensional, como orientador de la operación pública de desarrollo social, i.e. el enfoque los recursos materiales, humanos y financieros desde la administración pública.

Se decidió obviar la importancia de la pobreza en México, no solo por que no este el problema, además se corría el riesgo de involucrarse en una discusión que ha atrapado durante décadas a las CS en México, con pocos resultados. Se optó por asumir una estrategia pragmática, tomar los lineamientos del Coneval y la legislatura mexicana, debido a que en este espacio y con esta definición está ya dada y a partir de ella se puede actuar; se enfocó el problema como un problema técnico que requería una solución acorde.

Sin embargo es pertinente delinear nuestra posición respecto a la definición de PM de Coneval como satisfactoria. La definición de las dimensiones basada en carencias permite abarcar buena parte de las problemáticas relacionadas con la pobreza en México, la clasificación de pobre/no-pobre es bastante flexible y no está únicamente determinada por el ingreso. Aunque la discusión y el desarrollo conceptual deben continuarse, basados en la evolución del fenómeno en el tiempo tomando en cuenta las reacciones y mecanismos del Estado a partir de la definición actual. Justamente en la intervención del aparato estatal radica la principal virtud de esta definición de, está diseñada para ser medida y con un instrumento correspondiente: la ENIGH; depende poco de interpretaciones subjetivas, i.e tiene carga objetiva fuerte; no está basada en principios abstractos o valores, sino en distinciones sencillas, principalmente dicotómicas.

La definición está dirigida a organizaciones-instituciones: burocracia y burócratas. Considerando que el problema de la burocracia (Weber, 1992) mexicana trasciende distintos ámbitos, toda persona que ha vivido suficiente tiempo en el país puede dar cuenta de ello. Partimos del reconocimiento que las instituciones no están formadas por personas, sino operan más allá de los individuos que la conforman (Parsons, 1988) y que semanticamente son sistemas autopoiéticos (Luhmann, 1997) con los cuales la comunicación no puede conformarse por principios o valores propios de la política o el humanismo;

sino a partir de mecanismos comunicativos que son propios de estos sistemas organizacionales. Esto significa que la ejecución de proyectos a partir de las organizaciones, y de las públicas por excelencia, tienen problemas de ejecución de origen al tratarlas como un conjunto de personas con criterios de lógica individuales o políticos, se debe recurrir a sus procedimientos, formatos y formas; acudir a una dependencia y esperar que las cosas sean lógicas o razonables es al menos en México una receta para la frustración.

Así comunicativamente la PM es una propuesta interesante; en México históricamente se intentó solucionar el problema de la pobreza mediante grandes ideas y principios: revolucionarios, postrevolucionarios o neoliberales; sin embargo estas cargas semánticas son incomprensibles a nivel institucional, se requieren decisiones y objetivos; los indicadores dicotómicos, como los de la PM, permiten trasladar lineamientos y exigencias del entorno a los sistemas organizacionales burocráticos, logrando acoplamientos estructurales exitosos. A partir de la teoría de sistemas esto implica que la complejidad del entorno político puede ser reinterpretada para en el sistema organizacional-burocrático mediante indicadores; aumentando la complejidad técnica dentro de estos, en específico la definición y medición de la PM, a partir de la cual es posible elaborar formatos y cuestionarios; pero la medición, identificación y estimación de la pobreza son problemas de alta complejidad interna, aún por solventarse. Nuestra propuesta busca reducir la complejidad de la decisión de dónde dirigir los programas públicos basado en un modelo estadístico.

Además de la selección de un mejor modelo, es necesario que el algoritmo propuesto sea útil a niveles operativos-burocráticos. Esto implica la presentación en un formato que el personal en las dependencias pueda consultar e interpretar; pretender que el usuario entienda el algoritmo no es realista.

La idea de que no es factible pensar en la interpretación del modelo para su uso generalizado, lleva a otro punto de reflexión, de la relación entre las CS y la estadística. En disciplinas con carga hermeneútica fuerte, la comprensión del problema es el objetivo de las investigaciones. Al mismo tiempo sin embargo la justificación para no utilizar herramientas estadísticas es que los fenómenos sociales son muy complejos e incluyen muchas variables, que además no se pueden medir. Se busca interpretar cada parámetro en los modelos pero también incluir todas las variables posibles, así la interpretación de variables es una tarea onerosa que en muchos casos lleva a contradicciones, es ilegible; o en caso contrario lleva a modelos simplistas. Considérese la modelación de la PM en este documento, una interpretación exhaustiva de cada parámetro implica 56 covariables explicativas en relación con la varia-

ble dicotómica correspondiente, asumiendo que el resto de las variables son constantes; adicionalmente 21 parámetros de correlación entre las variables dependientes, i.e. 67 parámetros; si el modelo se amplía incluyendo factores aleatorios y variables adicionales, el número de parámetro a interpretar es aún mayor.

En el contexto que motivó esta investigación la interpretación de covariables no es relevante; dado el número de covariables la interpretación de cada parámetro traería más confusión que claridad. Las CS típicamente no se encuentran cómodas con este tipo de procesos. La propuesta antepone la predicción (estimación) del fenómeno a la explicación de sus mecanismos; es más cercana al machine learning que al ethos de las CS; consideramos que este cambio paradigmático es útil en otros campos de las CS, sobrepone la medición e intervención a las discusiones académicas cerradas, que pueden llegar a lo bizantino.



# Anexo A

## A.1. Desarrollo algebraico extendido

### A.1.1. Parametrización original

Una de las dificultades del método propuesto es la interpretabilidad, y la dificultad de regresar a la parametrización original de una Gaussiana multivariada; sin embargo, esto es un problema que requiere un tratamiento algebraico sencillo, pero relativamente sencillo. Para calcular la matriz  $\Sigma$ , es necesario obtener las matrices  $\mathbf{L}$  y  $\mathbf{D}$ , de la definición en la ecuación (2.10):

$$\Sigma = \mathbf{L}^{-1}\mathbf{D}(\mathbf{L}^{-1})^t,$$

mientras que para los parámetros de la media  $\beta_j^* = \beta\sqrt{\delta_j}$ ; y  $l_{jk} = \gamma_{jk}\sqrt{\delta_j}/\sqrt{\delta_k}$ ; el aspecto complicado es obtener los valores de  $\delta_j$  a partir de  $\gamma$ .

Para el manejo algebraico se utilizó el software de capacidades algebraicas Sage (The Sage Developers, 2015); recordando que  $\delta_1 = 1$ , en términos de  $l_{jk}$ ,

$$\delta_2 = -l_{21}^2 + 1,$$

expresando  $\gamma_{jk}^* = \gamma_{jk}\sqrt{\delta_k} = l_{jk}/\sqrt{\delta_j}$  es fácil mostrar que

$$1/\delta_2 = (\gamma_{jk}^*)^2 + 1,$$

de manera similar

$$\begin{aligned}\delta_3 &= 2 l_{21} l_{31} l_{32} - l_{31}^2 - l_{32}^2 + 1, \\ \delta_3^{-1} &= -2 l_{21} \gamma_{31}^* \gamma_{32}^* + (\gamma_{31}^*)^2 + (\gamma_{32}^*)^2 + 1.\end{aligned}$$

Este mismo patrón se repite para todos los valores subsecuentes de  $\delta_j$ , así obteniendo de manera jerárquica, para el caso extendido cuando  $p = 7$  presentamos las siguientes expresiones:

$$\delta_1 = 1,$$

$$\delta_2 = -l_{21}^2 + 1,$$

$$\delta_3 = 2 l_{21} l_{31} l_{32} - l_{31}^2 - l_{32}^2 + 1,$$

$$\begin{aligned}\delta_4 &= -2 l_{21} l_{32} l_{41} l_{43} - 2 l_{21} l_{31} l_{42} l_{43} + 2 l_{21} l_{41} l_{42} + 2 l_{31} l_{41} l_{43} \\ &\quad + 2 l_{32} l_{42} l_{43} - l_{41}^2 - l_{42}^2 - l_{43}^2 + 1,\end{aligned}$$

$$\begin{aligned}\delta_5 &= 2 l_{21} l_{32} l_{43} l_{51} l_{54} + 2 l_{21} l_{31} l_{43} l_{52} l_{54} + 2 l_{21} l_{32} l_{41} l_{53} l_{54} + 2 l_{21} l_{31} l_{42} l_{53} l_{54} \\ &\quad - 2 l_{21} l_{32} l_{51} l_{53} - 2 l_{21} l_{31} l_{52} l_{53} - 2 l_{21} l_{42} l_{51} l_{54} - 2 l_{31} l_{43} l_{51} l_{54} - 2 l_{21} l_{41} l_{52} l_{54} \\ &\quad - 2 l_{32} l_{43} l_{52} l_{54} - 2 l_{31} l_{41} l_{53} l_{54} - 2 l_{32} l_{42} l_{53} l_{54} + 2 l_{21} l_{51} l_{52} + 2 l_{31} l_{51} l_{53} \\ &\quad + 2 l_{32} l_{52} l_{53} + 2 l_{41} l_{51} l_{54} + 2 l_{42} l_{52} l_{54} + 2 l_{43} l_{53} l_{54} - l_{51}^2 - l_{52}^2 - l_{53}^2 - l_{54}^2 + 1,\end{aligned}$$



$$\begin{aligned}
\delta_6 = & -2 l_{21} l_{32} l_{43} l_{54} l_{61} l_{65} - 2 l_{21} l_{31} l_{43} l_{54} l_{62} l_{65} - 2 l_{21} l_{32} l_{41} l_{54} l_{63} l_{65} \\
& - 2 l_{21} l_{31} l_{42} l_{54} l_{63} l_{65} - 2 l_{21} l_{32} l_{43} l_{51} l_{64} l_{65} - 2 l_{21} l_{31} l_{43} l_{52} l_{64} l_{65} \\
& - 2 l_{21} l_{32} l_{41} l_{53} l_{64} l_{65} - 2 l_{21} l_{31} l_{42} l_{53} l_{64} l_{65} + 2 l_{21} l_{32} l_{43} l_{61} l_{64} + 2 l_{21} l_{31} l_{43} l_{62} l_{64} \\
& + 2 l_{21} l_{32} l_{41} l_{63} l_{64} + 2 l_{21} l_{31} l_{42} l_{63} l_{64} + 2 l_{21} l_{32} l_{53} l_{61} l_{65} + 2 l_{21} l_{42} l_{54} l_{61} l_{65} \\
& + 2 l_{31} l_{43} l_{54} l_{61} l_{65} + 2 l_{21} l_{31} l_{53} l_{62} l_{65} + 2 l_{21} l_{41} l_{54} l_{62} l_{65} + 2 l_{32} l_{43} l_{54} l_{62} l_{65} \\
& + 2 l_{21} l_{32} l_{51} l_{63} l_{65} + 2 l_{21} l_{31} l_{52} l_{63} l_{65} + 2 l_{31} l_{41} l_{54} l_{63} l_{65} + 2 l_{32} l_{42} l_{54} l_{63} l_{65} \\
& + 2 l_{21} l_{42} l_{51} l_{64} l_{65} + 2 l_{31} l_{43} l_{51} l_{64} l_{65} + 2 l_{21} l_{41} l_{52} l_{64} l_{65} + 2 l_{32} l_{43} l_{52} l_{64} l_{65} \\
& + 2 l_{31} l_{41} l_{53} l_{64} l_{65} + 2 l_{32} l_{42} l_{53} l_{64} l_{65} - 2 l_{21} l_{32} l_{61} l_{63} - 2 l_{21} l_{31} l_{62} l_{63} \\
& - 2 l_{21} l_{42} l_{61} l_{64} - 2 l_{31} l_{43} l_{61} l_{64} - 2 l_{21} l_{41} l_{62} l_{64} - 2 l_{32} l_{43} l_{62} l_{64} - 2 l_{31} l_{41} l_{63} l_{64} \\
& - 2 l_{32} l_{42} l_{63} l_{64} - 2 l_{21} l_{52} l_{61} l_{65} - 2 l_{31} l_{53} l_{61} l_{65} - 2 l_{41} l_{54} l_{61} l_{65} - 2 l_{21} l_{51} l_{62} l_{65} \\
& - 2 l_{32} l_{53} l_{62} l_{65} - 2 l_{42} l_{54} l_{62} l_{65} - 2 l_{31} l_{51} l_{63} l_{65} - 2 l_{32} l_{52} l_{63} l_{65} - 2 l_{43} l_{54} l_{63} l_{65} \\
& - 2 l_{41} l_{51} l_{64} l_{65} - 2 l_{42} l_{52} l_{64} l_{65} - 2 l_{43} l_{53} l_{64} l_{65} + 2 l_{21} l_{61} l_{62} + 2 l_{31} l_{61} l_{63} \\
& + 2 l_{32} l_{62} l_{63} + 2 l_{41} l_{61} l_{64} + 2 l_{42} l_{62} l_{64} + 2 l_{43} l_{63} l_{64} + 2 l_{51} l_{61} l_{65} + 2 l_{52} l_{62} l_{65} \\
& + 2 l_{53} l_{63} l_{65} + 2 l_{54} l_{64} l_{65} - l_{61}^2 - l_{62}^2 - l_{63}^2 - l_{64}^2 - l_{65}^2 + 1,
\end{aligned}$$

$$\begin{aligned}
\delta_7 = & 2 l_{21} l_{32} l_{43} l_{54} l_{65} l_{71} l_{76} + 2 l_{21} l_{31} l_{43} l_{54} l_{65} l_{72} l_{76} + 2 l_{21} l_{32} l_{41} l_{54} l_{65} l_{73} l_{76} \\
& + 2 l_{21} l_{31} l_{42} l_{54} l_{65} l_{73} l_{76} + 2 l_{21} l_{32} l_{43} l_{51} l_{65} l_{74} l_{76} + 2 l_{21} l_{31} l_{43} l_{52} l_{65} l_{74} l_{76} \\
& + 2 l_{21} l_{32} l_{41} l_{53} l_{65} l_{74} l_{76} + 2 l_{21} l_{31} l_{42} l_{53} l_{65} l_{74} l_{76} + 2 l_{21} l_{32} l_{43} l_{54} l_{61} l_{75} l_{76} \\
& + 2 l_{21} l_{31} l_{43} l_{54} l_{62} l_{75} l_{76} + 2 l_{21} l_{32} l_{41} l_{54} l_{63} l_{75} l_{76} + 2 l_{21} l_{31} l_{42} l_{54} l_{63} l_{75} l_{76} \\
& + 2 l_{21} l_{32} l_{43} l_{51} l_{64} l_{75} l_{76} + 2 l_{21} l_{31} l_{43} l_{52} l_{64} l_{75} l_{76} + 2 l_{21} l_{32} l_{41} l_{53} l_{64} l_{75} l_{76} \\
& + 2 l_{21} l_{31} l_{42} l_{53} l_{64} l_{75} l_{76} - 2 l_{21} l_{32} l_{43} l_{54} l_{71} l_{75} - 2 l_{21} l_{31} l_{43} l_{54} l_{72} l_{75} \\
& - 2 l_{21} l_{32} l_{41} l_{54} l_{73} l_{75} - 2 l_{21} l_{31} l_{42} l_{54} l_{73} l_{75} - 2 l_{21} l_{32} l_{43} l_{51} l_{74} l_{75} \\
& - 2 l_{21} l_{31} l_{43} l_{52} l_{74} l_{75} - 2 l_{21} l_{32} l_{41} l_{53} l_{74} l_{75} - 2 l_{21} l_{31} l_{42} l_{53} l_{74} l_{75} \\
& - 2 l_{21} l_{32} l_{43} l_{64} l_{71} l_{76} - 2 l_{21} l_{32} l_{53} l_{65} l_{71} l_{76} - 2 l_{21} l_{42} l_{54} l_{65} l_{71} l_{76} \\
& - 2 l_{31} l_{43} l_{54} l_{65} l_{71} l_{76} - 2 l_{21} l_{31} l_{43} l_{64} l_{72} l_{76} - 2 l_{21} l_{31} l_{53} l_{65} l_{72} l_{76} \\
& - 2 l_{21} l_{41} l_{54} l_{65} l_{72} l_{76} - 2 l_{32} l_{43} l_{54} l_{65} l_{72} l_{76} - 2 l_{21} l_{32} l_{41} l_{64} l_{73} l_{76} \\
& - 2 l_{21} l_{31} l_{42} l_{64} l_{73} l_{76} - 2 l_{21} l_{32} l_{51} l_{65} l_{73} l_{76} - 2 l_{21} l_{31} l_{52} l_{65} l_{73} l_{76} \\
& - 2 l_{31} l_{41} l_{54} l_{65} l_{73} l_{76} - 2 l_{32} l_{42} l_{54} l_{65} l_{73} l_{76} - 2 l_{21} l_{32} l_{43} l_{61} l_{74} l_{76} \\
& - 2 l_{21} l_{31} l_{43} l_{62} l_{74} l_{76} - 2 l_{21} l_{32} l_{41} l_{63} l_{74} l_{76} - 2 l_{21} l_{31} l_{42} l_{63} l_{74} l_{76} \\
& \dots
\end{aligned}$$

$\delta_7 = \dots$ 

$$\begin{aligned}
& - 2 l_{21} l_{32} l_{53} l_{61} l_{75} l_{76} - 2 l_{21} l_{42} l_{54} l_{61} l_{75} l_{76} - 2 l_{31} l_{43} l_{54} l_{61} l_{75} l_{76} \\
& - 2 l_{21} l_{31} l_{53} l_{62} l_{75} l_{76} - 2 l_{21} l_{41} l_{54} l_{62} l_{75} l_{76} - 2 l_{32} l_{43} l_{54} l_{62} l_{75} l_{76} \\
& - 2 l_{21} l_{32} l_{51} l_{63} l_{75} l_{76} - 2 l_{21} l_{31} l_{52} l_{63} l_{75} l_{76} - 2 l_{31} l_{41} l_{54} l_{63} l_{75} l_{76} \\
& - 2 l_{32} l_{42} l_{54} l_{63} l_{75} l_{76} - 2 l_{21} l_{42} l_{51} l_{64} l_{75} l_{76} - 2 l_{31} l_{43} l_{51} l_{64} l_{75} l_{76} \\
& - 2 l_{21} l_{41} l_{52} l_{64} l_{75} l_{76} - 2 l_{32} l_{43} l_{52} l_{64} l_{75} l_{76} - 2 l_{31} l_{41} l_{53} l_{64} l_{75} l_{76} \\
& - 2 l_{32} l_{42} l_{53} l_{64} l_{75} l_{76} + 2 l_{21} l_{32} l_{43} l_{71} l_{74} + 2 l_{21} l_{31} l_{43} l_{72} l_{74} + 2 l_{21} l_{32} l_{41} l_{73} l_{74} \\
& + 2 l_{21} l_{31} l_{42} l_{73} l_{74} + 2 l_{21} l_{32} l_{53} l_{71} l_{75} + 2 l_{21} l_{42} l_{54} l_{71} l_{75} + 2 l_{31} l_{43} l_{54} l_{71} l_{75} \\
& + 2 l_{21} l_{31} l_{53} l_{72} l_{75} + 2 l_{21} l_{41} l_{54} l_{72} l_{75} + 2 l_{32} l_{43} l_{54} l_{72} l_{75} + 2 l_{21} l_{32} l_{51} l_{73} l_{75} \\
& + 2 l_{21} l_{31} l_{52} l_{73} l_{75} + 2 l_{31} l_{41} l_{54} l_{73} l_{75} + 2 l_{32} l_{42} l_{54} l_{73} l_{75} + 2 l_{21} l_{42} l_{51} l_{74} l_{75} \\
& + 2 l_{31} l_{43} l_{51} l_{74} l_{75} + 2 l_{21} l_{41} l_{52} l_{74} l_{75} + 2 l_{32} l_{43} l_{52} l_{74} l_{75} + 2 l_{31} l_{41} l_{53} l_{74} l_{75} \\
& + 2 l_{32} l_{42} l_{53} l_{74} l_{75} + 2 l_{21} l_{32} l_{63} l_{71} l_{76} + 2 l_{21} l_{42} l_{64} l_{71} l_{76} + 2 l_{31} l_{43} l_{64} l_{71} l_{76} \\
& + 2 l_{21} l_{52} l_{65} l_{71} l_{76} + 2 l_{31} l_{53} l_{65} l_{71} l_{76} + 2 l_{41} l_{54} l_{65} l_{71} l_{76} + 2 l_{21} l_{31} l_{63} l_{72} l_{76} \\
& + 2 l_{21} l_{41} l_{64} l_{72} l_{76} + 2 l_{32} l_{43} l_{64} l_{72} l_{76} + 2 l_{21} l_{51} l_{65} l_{72} l_{76} + 2 l_{32} l_{53} l_{65} l_{72} l_{76} \\
& + 2 l_{42} l_{54} l_{65} l_{72} l_{76} + 2 l_{21} l_{32} l_{61} l_{73} l_{76} + 2 l_{21} l_{31} l_{62} l_{73} l_{76} + 2 l_{31} l_{41} l_{64} l_{73} l_{76} \\
& + 2 l_{32} l_{42} l_{64} l_{73} l_{76} + 2 l_{31} l_{51} l_{65} l_{73} l_{76} + 2 l_{32} l_{52} l_{65} l_{73} l_{76} + 2 l_{43} l_{54} l_{65} l_{73} l_{76} \\
& + 2 l_{21} l_{42} l_{61} l_{74} l_{76} + 2 l_{31} l_{43} l_{61} l_{74} l_{76} + 2 l_{21} l_{41} l_{62} l_{74} l_{76} + 2 l_{32} l_{43} l_{62} l_{74} l_{76} \\
& + 2 l_{31} l_{41} l_{63} l_{74} l_{76} + 2 l_{32} l_{42} l_{63} l_{74} l_{76} + 2 l_{41} l_{51} l_{65} l_{74} l_{76} + 2 l_{42} l_{52} l_{65} l_{74} l_{76} \\
& + 2 l_{43} l_{53} l_{65} l_{74} l_{76} + 2 l_{21} l_{52} l_{61} l_{75} l_{76} + 2 l_{31} l_{53} l_{61} l_{75} l_{76} + 2 l_{41} l_{54} l_{61} l_{75} l_{76} \\
& + 2 l_{21} l_{51} l_{62} l_{75} l_{76} + 2 l_{32} l_{53} l_{62} l_{75} l_{76} + 2 l_{42} l_{54} l_{62} l_{75} l_{76} + 2 l_{31} l_{51} l_{63} l_{75} l_{76} \\
& + 2 l_{32} l_{52} l_{63} l_{75} l_{76} + 2 l_{43} l_{54} l_{63} l_{75} l_{76} + 2 l_{41} l_{51} l_{64} l_{75} l_{76} + 2 l_{42} l_{52} l_{64} l_{75} l_{76} \\
& + 2 l_{43} l_{53} l_{64} l_{75} l_{76} - 2 l_{21} l_{32} l_{71} l_{73} - 2 l_{21} l_{31} l_{72} l_{73} - 2 l_{21} l_{42} l_{71} l_{74} - 2 l_{31} l_{43} l_{71} l_{74} \\
& - 2 l_{21} l_{41} l_{72} l_{74} - 2 l_{32} l_{43} l_{72} l_{74} - 2 l_{31} l_{41} l_{73} l_{74} - 2 l_{32} l_{42} l_{73} l_{74} - 2 l_{21} l_{52} l_{71} l_{75} \\
& - 2 l_{31} l_{53} l_{71} l_{75} - 2 l_{41} l_{54} l_{71} l_{75} - 2 l_{21} l_{51} l_{72} l_{75} - 2 l_{32} l_{53} l_{72} l_{75} - 2 l_{42} l_{54} l_{72} l_{75} \\
& - 2 l_{31} l_{51} l_{73} l_{75} - 2 l_{32} l_{52} l_{73} l_{75} - 2 l_{43} l_{54} l_{73} l_{75} - 2 l_{41} l_{51} l_{74} l_{75} - 2 l_{42} l_{52} l_{74} l_{75} \\
& - 2 l_{43} l_{53} l_{74} l_{75} - 2 l_{21} l_{62} l_{71} l_{76} - 2 l_{31} l_{63} l_{71} l_{76} - 2 l_{41} l_{64} l_{71} l_{76} - 2 l_{51} l_{65} l_{71} l_{76} \\
& - 2 l_{21} l_{61} l_{72} l_{76} - 2 l_{32} l_{63} l_{72} l_{76} - 2 l_{42} l_{64} l_{72} l_{76} - 2 l_{52} l_{65} l_{72} l_{76} - 2 l_{31} l_{61} l_{73} l_{76} \\
& - 2 l_{32} l_{62} l_{73} l_{76} - 2 l_{43} l_{64} l_{73} l_{76} - 2 l_{53} l_{65} l_{73} l_{76} - 2 l_{41} l_{61} l_{74} l_{76} - 2 l_{42} l_{62} l_{74} l_{76} \\
& - 2 l_{43} l_{63} l_{74} l_{76} - 2 l_{54} l_{65} l_{74} l_{76} - 2 l_{51} l_{61} l_{75} l_{76} - 2 l_{52} l_{62} l_{75} l_{76} - 2 l_{53} l_{63} l_{75} l_{76} \\
& - 2 l_{54} l_{64} l_{75} l_{76} + 2 l_{21} l_{71} l_{72} + 2 l_{31} l_{71} l_{73} + 2 l_{32} l_{72} l_{73} + 2 l_{41} l_{71} l_{74} + 2 l_{42} l_{72} l_{74} \\
& + 2 l_{43} l_{73} l_{74} + 2 l_{51} l_{71} l_{75} + 2 l_{52} l_{72} l_{75} + 2 l_{53} l_{73} l_{75} + 2 l_{54} l_{74} l_{75} + 2 l_{61} l_{71} l_{76} \\
& + 2 l_{62} l_{72} l_{76} + 2 l_{63} l_{73} l_{76} + 2 l_{64} l_{74} l_{76} + 2 l_{65} l_{75} l_{76} \\
& - l_{71}^2 - l_{72}^2 - l_{73}^2 - l_{74}^2 - l_{75}^2 - l_{76}^2 + 1.
\end{aligned}$$

A partir de las expresiones anteriores podemos regresar a la parametrización original, y ampliar el procedimiento a un tamaño  $p$  razonable, es computacionalmente posible, no se encontró una expresión simplificada de estas soluciones al sistema de ecuaciones. Es evidente que la implementación es complicada, aunque hasta  $p = 5$  es fácilmente implementable. Esta es la única desventaja del método propuesto, pero para solventar esto se propone, si se requiere la parametrización original, generar nuevamente las cadenas, volver a ajustar el modelo, pero cambiando el orden del vector respuesta; considerando que es fácilmente obtenible para los primeros cinco elementos, hasta obtener todos los parámetros requeridos. Se insiste en que esto es necesario únicamente para  $p > 5$ , y sólo si se requieren estimaciones de los parámetros originales para interpretación; la alternativa es conveniente cuando el requerimiento computacional no es tan alto, i.e. un tamaño de muestra no tan grande,  $n < 7000$  aproximadamente.

Para esto sin embargo no es posible utilizar la definición original de la matriz  $\Sigma$ , en su lugar requiere utilizar las igualdades de  $\sigma_{ik}$  individuales:

$$\begin{aligned}
\sigma_{21} &= -l_{21}, \\
\sigma_{31} &= l_{21}l_{32} - l_{31}, \\
\sigma_{32} &= l_{21}l_{31} - l_{32}, \\
\sigma_{41} &= -l_{21}l_{32}l_{43} + l_{21}l_{42} + l_{31}l_{43} - l_{41}, \\
\sigma_{42} &= -l_{21}l_{31}l_{43} + l_{21}l_{41} + l_{32}l_{43} - l_{42}, \\
\sigma_{43} &= -l_{21}l_{32}l_{41} - l_{21}l_{31}l_{42} + l_{31}l_{41} + l_{32}l_{42} - l_{43}, \\
\sigma_{51} &= l_{21}l_{32}l_{43}l_{54} - l_{21}l_{32}l_{53} - l_{21}l_{42}l_{54} - l_{31}l_{43}l_{54} + l_{21}l_{52} + l_{31}l_{53} + l_{41}l_{54} - l_{51}, \\
\sigma_{52} &= l_{21}l_{31}l_{43}l_{54} - l_{21}l_{31}l_{53} - l_{21}l_{41}l_{54} - l_{32}l_{43}l_{54} + l_{21}l_{51} + l_{32}l_{53} + l_{42}l_{54} - l_{52}, \\
\sigma_{53} &= l_{21}l_{32}l_{41}l_{54} + l_{21}l_{31}l_{42}l_{54} - l_{21}l_{32}l_{51} - l_{21}l_{31}l_{52} - l_{31}l_{41}l_{54} - l_{32}l_{42}l_{54} \\
&\quad + l_{31}l_{51} + l_{32}l_{52} + l_{43}l_{54} - l_{53}, \\
\sigma_{54} &= l_{21}l_{32}l_{43}l_{51} + l_{21}l_{31}l_{43}l_{52} + l_{21}l_{32}l_{41}l_{53} + l_{21}l_{31}l_{42}l_{53} - l_{21}l_{42}l_{51} - l_{31}l_{43}l_{51} \\
&\quad - l_{21}l_{41}l_{52} - l_{32}l_{43}l_{52} - l_{31}l_{41}l_{53} - l_{32}l_{42}l_{53} + l_{41}l_{51} + l_{42}l_{52} + l_{43}l_{53} - l_{54}.
\end{aligned}$$

Esto implica sin embargo que para regresar a la parametrización original, si  $p = 6$  se requiere ajustar 3 veces el modelo, lo mismo para  $p = 7$ , para  $p = 8$  son necesarias 4 veces, y 6 veces para  $p = 9$ . Sin embargo, esto es significativo para la interpretación, lo que implica que habría que interpretar 36 parámetros de correlación, conforme aumenta el número de covariables el interés de interpretación disminuye en términos prácticos.

### A.1.2. Probabilidades conjuntas

La intención de este apartado es mostrar que el modelo propuesto efectivamente modela la probabilidad de interes  $\mathbf{Y}_i = \mathbf{y}_i$

Para el caso de  $p = 2$ , a partir de la ecuación (2.9):

$$\begin{aligned}
\Pr(\mathbf{Y}_i = \mathbf{y}_i) &= \int_{B_{i2}} \int_{B_{i1}} \phi_2(\mathbf{W}_i | \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}) d\mathbf{W}_i \\
&= \int_{B_{i2}} \int_{B_{i1}} \phi(W_1 | \mu_1^*, 1) \phi(W_2 | \mu_2^* + \sigma_{21}(W_1 - \mu_1^*), (1 - \sigma_{21}^2)) dW_1 dW_2 \\
&= \int_{B_{i1}} \phi(W_1 | \mu_1^*, 1) \int_{B_{i2}} \phi(W_2 | \mu_2^* + \sigma_{21}(W_1 - \mu_1^*), \delta_2) dW_2 dW_1 \\
&= \int_{B_{i1}} \phi(Z_1 | \mu_1, 1) \Phi^* \left( \frac{\mu_2^* + \sigma_{21}(W_1 - \mu_1^*)}{\sqrt{\delta_2}} \right) dZ_1 \\
&= \int_{B_{i1}} \phi(Z_1 | \mu_1, 1) \Phi^* (\mu_2 + \gamma_{21}(Z_1 - \mu_1)) dZ_1 \\
&= \int_{B_{i2}} \int_{B_{i1}} \phi(Z_1 | \mu_1, 1) \phi(Z_2 | \mu_2 + \gamma_{21}(Z_1 - \mu_1), 1) d\mathbf{Z}_i
\end{aligned}$$

Para el caso de  $p = 3$

$$\begin{aligned}
\Pr(\mathbf{Y}_i = \mathbf{y}_i) &= \int_{B_{i2}} \int_{B_{i1}} \int_{B_{i3}} \phi_3(\mathbf{W}_i | \boldsymbol{\mu}_i^*, \boldsymbol{\Sigma}) d\mathbf{W}_i \\
&= \int_{B_{i1}} \phi(W_{i1} | \mu_1^*, 1) \int_{B_{i2}} f_{W_2|W_1}(w_2 | w_1) \\
&\quad \int_{B_{i3}} f_{W_3|W_1, W_2}(w_3 | w_1, w_2) d\mathbf{W}_i \\
&= \int_{B_{i1}} \phi(Z_{i1} | \mu_1, 1) \int_{B_{i2}} f_{W_2|W_1}(w_2 | w_1) \\
&\quad \int_{B_{i3}} \phi\left(w_3 \left| \mu_3^* + \frac{l_{31}(\mu_1^* - w_1) + l_{32}(\mu_2^* - w_2)}{\delta_3}, 1/\delta_3\right.\right) dW_3 dW_2 dZ_1 \\
&= \int_{B_{i1}} \phi(Z_{i1} | \mu_1, 1) \int_{B_{i2}} f_{W_2|W_1}(w_2 | w_1) \\
&\quad \Phi^*\left(\mu_3^* + \frac{l_{31}(\mu_1^* - w_1) + l_{32}(\mu_2^* - w_2)}{\delta_3} \middle/ 1/\delta_3\right) dW_2 dZ_1 \\
&= \int_{B_{i1}} \phi(Z_{i1} | \mu_1, 1) \int_{B_{i2}} f_{W_2|W_1}(w_2 | w_1) \\
&\quad \Phi^*(\mu_3 + \gamma_{31}(\mu_1 - z_1) + \gamma_{32}(\mu_2 - z_2)) dW_2 dZ_1 \\
&= \int_{B_{i1}} f_{Z_1}(z_1) \int_{B_{i2}} \phi(w_2 | \mu_2^* + \sigma_{21}(w_1 - \mu_1^*), \delta_2) \\
&\quad \int_{B_{i3}} f_{Z_3|Z_1, Z_2}(z_3 | z_1, z_2) dZ_3 dW_2 dZ_1 \\
&= \int_{B_{i1}} f_{Z_1}(z_1) \int_{B_{i2}} f_{W_2|W_1}(w_2 | w_1) \\
&\quad \int_{B_{i3}} f_{Z_3|Z_1, Z_2}(z_3 | z_1, z_2) dZ_3 dW_2 dZ_1 \\
&= \int_{B_{i1}} f_{Z_1}(z_1) \int_{B_{i2}} \sqrt[4]{\delta_2} f_{W_2|Z_1}\left(\sqrt[4]{\delta_2} z_2 | z_1\right) \\
&\quad \int_{B_{i3}} f_{Z_3|Z_1, Z_2}(z_3 | z_1, z_2) dZ_3 dZ_2 dZ_1 \\
&= \int_{B_{i1}} f_{Z_1}(z_1) \int_{B_{i2}} \phi(z_2 | \mu_2 + \gamma_{21}(\mu_1 - z_1), 1) \\
&\quad \int_{B_{i3}} f_{Z_3|Z_1, Z_2}(z_3 | z_1, z_2) dZ_3 dZ_2 dZ_1 \\
&= \int_{B_{i1}} \int_{B_{i2}} \int_{B_{i3}} f_{Z_1}(z_1) f_{Z_2|Z_1}(z_2 | z_1) f_{Z_3|Z_1, Z_2}(z_3 | z_1, z_2) d\mathbf{Z}
\end{aligned}$$

Esta es la probabilidad estimada en nuestra propuesta. A partir del modelo es inmediato identificar que  $W_1 = Z_1$ , además definimos:

$$\Phi^*(a/b) = \begin{cases} \Phi(\frac{0-a}{b}) & y_i = 0 \\ 1 - \Phi(\frac{0-a}{b}) & y_i = 1 \end{cases},$$

donde  $\Phi(\cdot)$  es la *cdf* de  $N(0, 1)$ .

Este mismo procedimiento puede ser extendido a cualquier  $p$ .

### A.1.3. Probabilidades marginales

De lo anterior se concluye que algunas probabilidades pueden ser directamente estimadas de la muestra generada por el algoritmo propuesto:

- La probabilidad marginal del primer elemento del vector respuesta,  $\Pr(Y_1 = 1) = \Pr(Z_1 > 0)$ ;
- las probabilidades conjuntas

$$\Pr(\mathbf{Y} = \mathbf{y}_l) = \int_{B_p} \dots \int_{B_1} f_{Z_1}(z_1) f_{Z_2|Z_1}(z_2|z_1) \dots \dots f_{Z_p|Z_1, \dots, Z_{p-1}}(z_p|z_1, \dots, z_{p-1}) d\mathbf{Z};$$

- y marginalmente, para cualquier subconjunto consecutivo desde  $y_1$ , i.e.  $\mathbf{Y}^* = [Y_1, \dots, Y_{p-a}]^t$ ,  $a \in (1, \dots, p-1)$ ,

$$\Pr(\mathbf{Y}^* = \mathbf{y}_l^*) = \int_{B_{p-a}} \dots \int_{B_1} f_{Z_1}(z_1) f_{Z_2|Z_1}(z_2|z_1) \dots \dots f_{Z_{p-a}|Z_1, \dots, Z_{p-a-1}}(z_p|z_1, \dots, z_{p-a-1}) d\mathbf{z}^*.$$

Sin embargo también estas igualdades se sostienen para las probabilidades marginales de cada elemento de  $\mathbf{Y}$ , i.e.

$$\Pr(Y_j = 1) = \Pr(Z_j | \mathbf{Z}_{<j} > 0).$$

Se muestra a continuación para el caso de  $j = 2$ :

$$\begin{aligned}
\Pr(Y_2 = 1) &= \Pr(W_2 > 0) \\
&= \Pr(W_2 > 0 | W_1 \leq 0) \Pr(W_1 \leq 0) \\
&\quad + \Pr(W_2 > 0 | W_1 > 0) \Pr(W_1 > 0) \\
&= \frac{\Pr(W_2 > 0, W_1 \leq 0)}{\Pr(W_1 \leq 0)} \Pr(W_1 \leq 0) \\
&\quad + \frac{\Pr(W_2 > 0, W_1 > 0)}{\Pr(W_1 > 0)} \Pr(W_1 > 0) \\
&= \Pr(W_2 > 0, W_1 \leq 0) + \Pr(W_2 > 0, W_1 > 0) \\
&= \int_{\mathbb{R}^+} \int_{\mathbb{R}^-} f_{Z_1}(z_1) f_{Z_2|Z_1}(z_2|z_1) dz_1 dz_2 \\
&\quad + \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} f_{Z_1}(z_1) f_{Z_2|Z_1}(z_2|z_1) dz_1 dz_2 \\
&= \int_{\mathbb{R}^+} \int_{\mathbb{R}} f_{Z_1}(z_1) f_{Z_2|Z_1}(z_2|z_1) dz_1 dz_2 \\
&= \int_{\mathbb{R}} f_{Z_1}(z_1) \int_{\mathbb{R}^+} f_{Z_2|Z_1}(z_2|z_1) dz_2 dz_1 \\
&= \int_{\mathbb{R}^+} f_{Z_2|Z_1}(z_2|z_1) dz_2 \\
&= \Pr(Z_2 | Z_1 > 0).
\end{aligned}$$

Este mismo procedimiento puede ser fácilmente extendido a cualquier  $j$  y cualquier  $p$ . Esto implica que a partir de la cadena generada con el algoritmo propuesto se puede estimar también la probabilidad marginal para cualquier elemento  $j$ , de la variable respuesta; sin necesidad de ajustar un modelo independiente para cada elemento. Recordando que la probabilidad marginal de cada elemento no es afectada por la correlación, dado las propiedades de la distribución Normal.

#### A.1.4. Equivalencia entre $\mathbf{LDL}^t$ y $\mathbf{TT}^t$

La descomposición  $\mathbf{A} = \mathbf{LDL}^t$  que se utiliza en la transformación de las variables, tiene equivalencia con la descomposición de Cholesky, i.e.  $\mathbf{A} = \mathbf{TT}^t$ , es sencillo mostrar dicha equivalencia:

$$\begin{aligned}
\mathbf{A} &= \mathbf{T}\mathbf{T}^t \\
&= \mathbf{T}\mathbf{I}\mathbf{T}^t \\
&= \mathbf{T}\mathbf{T}_d^{-1}\mathbf{T}_d\mathbf{T}_d^t(\mathbf{T}_d^{-1})^t\mathbf{T}^t \\
&= (\mathbf{T}\mathbf{T}_d^{-1})\mathbf{T}_d^2(\mathbf{T}\mathbf{T}_d^{-1})^t \\
&= \mathbf{L}\mathbf{D}\mathbf{L}^t
\end{aligned}$$

Definiendo las matrices  $\mathbf{B}_d$  como una matriz diagonal cuyos elementos de la diagonal principal son los de la diagonal principal de la matriz  $\mathbf{B}$ , utilizando el producto Hadamard ( $\circ$ ) se define:  $\mathbf{B}_d = \mathbf{B}_{k \times k} \circ \mathbf{I}_{k \times k}$ .

Quedando las equivalencias entre las matrices de las descomposiciones como,  $\mathbf{L} = \mathbf{T}\mathbf{T}_d^{-1}$ ,  $\mathbf{D} = \mathbf{T}_d^2$ , esto garantiza que los elementos de la diagonal principal de  $\mathbf{L}$  sean 1, y que la descomposición sea única. Este resultado es útil para la implementación del algoritmo del MVP aquí propuesto, en tanto que en la mayoría de paquetes donde se pudiera implementar el algoritmo se encuentra disponible la descomposición de Cholesky; en particular este resultado se utiliza para la implementación de la definición a priori de  $\gamma$  a partir de información sobre  $\Sigma$ , que se utilizó en la modelación de la PM, y que se encuentra descrita en el anexo B.1.1.



# Anexo B

## B.1. Detalles del Algoritmo

### B.1.1. Selección a priori de $\pi(\boldsymbol{\gamma})$

Si bien proponer una distribución informativa de los parámetros de media  $\boldsymbol{\beta}$ , es difícil en tanto que se desconocen previo al ajuste las propiedades de la variable latente  $\mathbf{z}$ ; en el caso de los parámetros de correlación  $\boldsymbol{\Sigma}$ , y por tanto de  $\boldsymbol{\gamma}$ , sí es posible tener información previa valiosa de qué tan (cor)relacionados están los elementos del vector respuesta -este es el caso de la modelación de la PM-, sin embargo no es inmediato la incorporación de esta información al modelo mediante los parámetros  $\boldsymbol{\gamma}$ .

Para tal fin necesitamos definir la matriz

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ \gamma_{21} & 1 & 0 & \dots & 0 \\ \gamma_{31} & \gamma_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \gamma_{p1} & \gamma_{p2} & \dots & \gamma_{p,p-1} & 1 \end{bmatrix},$$

y cada elemento se define a partir de las matrices  $\mathbf{L}$  y  $\mathbf{D}$ , como  $\gamma_{jk} = l_{jk}\sqrt{\delta_k}/\sqrt{\delta_j}$ . Así

$$\boldsymbol{\Gamma} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{1/2}.$$

A partir de esta ecuación es sencillo calcular la matriz  $\boldsymbol{\Gamma}$  correspondiente a una  $\boldsymbol{\Sigma}$  específica.

Para poder especificar los parámetros a priori de  $\pi(\boldsymbol{\gamma})$  es necesario únicamente definir la media a priori de  $\sigma$  y un cuantil por cada parámetro de

correlación, i.e.  $E[\sigma_{jk}]$ ,  $\Pr(\sigma < q)$

Aún cuando en la modelación de la PM presentada en este documento, se consideraron los parámetros  $\gamma_{jk}$  como mutuamente independientes, esto trae consecuencias en la definición a priori, una mejor manera consiste en establecer conjuntamente por renglones de la matriz  $\mathbf{\Gamma}$ , es decir  $\boldsymbol{\gamma}_j$  como se encuentra definida para el modelo condicional en la ecuación (2.20).

### B.1.2. Casos particulares

#### Correlaciones conocidas, $\boldsymbol{\sigma}_j = \mathbf{0}$

Una de las propiedades de un MVP es que  $\boldsymbol{\Sigma}$  es positiva definida, esto para que la distribución asociada esté bien definida. Sin embargo, esto impone condiciones sobre los posibles valores de los elementos  $\sigma_{jk}$ ; ya que para ser una matriz de covarianzas debe ser positiva definida, se impone la restricción para las magnitudes conjuntas de los parámetros de restricción.

Esta condición es relevante en el modelo, dado que al suponer un valor fijo de un parámetro en 0, permite que el resto de los valores sean de mayor magnitud. Sin embargo, debido a la transformación utilizada no es posible imponer esta restricción a elementos individuales  $\sigma_{jk}$ , sin embargo es sencillo fijar subconjuntos de parámetros a cero:

$$\boldsymbol{\gamma}_j = \mathbf{0} \Rightarrow \boldsymbol{\sigma}_j = \mathbf{0}.$$

Para poder utilizar esta propiedad del modelo es indispensable considerar previamente el orden en que se incluyen los elementos del vector respuesta. También implica que no cualquier combinación de valores fijados en 0 es posible. Es una limitante del modelo que se debe a la transformación utilizada para mejorar la eficiencia del cómputo. Esta forma de fijar valores se encuentra implementada en la función computacional en R, mediante el parámetro `fix`.

#### Mismo efecto de covariables, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_j = \dots = \boldsymbol{\beta}_p$

Existe la posibilidad de restringir el vector de parámetros lineales sea igual para cada elemento de la variable respuesta, esto tiene sentido para ciertos estudios; p.ej. en el caso de estudios longitudinales donde las mismas covariables son medidas en diferentes puntos en el tiempo, es una de las

condiciones que Chib y Greenberg (1998) impusieron en el análisis de los datos de seis ciudades. Podemos expresar cómo:

$$\begin{aligned}\mu_{i1} &= \mathbf{x}_{i1}^t \boldsymbol{\beta}^*, \\ &\vdots \\ \mu_{ij} &= \mathbf{x}_{ij}^t \boldsymbol{\beta}^*, \\ &\vdots \\ \mu_{ip} &= \mathbf{x}_{ip}^t \boldsymbol{\beta}^*.\end{aligned}$$

Debido a la transformación utilizada  $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^* / \sqrt{\delta_j}$ , la implementación de esta forma particular del modelo requiere el cálculo en cada paso de  $\delta_j$ ; que como se mostró en el anexo A.1.1 además de aumentar el costo computacional del ajuste a partir de  $\delta_7$  es difícil de implementar.

Sin embargo, una alternativa para un modelo similar es fijar los efectos de los parámetros de regresión transformados

$$\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \dots = \boldsymbol{\beta}_p.$$

Si bien este modelo no tiene una interpretación marginal, en términos condicionales puede ser eficiente y sencillo de implementar.

Tanto para el caso de los parámetros  $\boldsymbol{\beta}$  como para los transformados, aún cuando las covariables  $\mathbf{x}_{ij}$  no sean las mismas para cada medición  $j$ , sí es necesario que los vectores sean de la misma dimensión. La implementación de este modelo únicamente requiere redefinir la matriz  $\mathbf{X}_i$  incluida en la ecuación (2.19) de la siguiente manera:

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_{i1}^t \\ \vdots \\ \mathbf{x}_{ij}^t \\ \vdots \\ \mathbf{x}_{ip}^t \end{bmatrix}.$$

Esta misma implementación puede extenderse a casos intermedios, donde solo algunos de los parámetros  $\beta_l$  sean los mismos para ciertos elementos del vector respuesta.

**Un solo parámetro de correlación,**  $\sigma_{21} = \dots = \sigma_{jl} = \dots = \sigma_{p,p-1}$

Este tipo de modelos son conocidos como equicorrelacionados, y es la otra condición impuesta en el análisis de Chib y Greenberg (1998) sobre los datos de niños. Sin embargo, nuevamente debido a la transformación utilizada, la implementación no es tan directa, para que se cumpla la matriz  $\mathbf{L}$  tiene la siguiente estructura, ejemplificada para el caso  $p = 5$ :

$$\begin{bmatrix} 1 & & & & \\ l & & & & \\ l/(1-l) & l/(1-l) & & & \\ l/(1-2l) & l/(1-2l) & l/(1-2l) & & \\ l/(1-3l) & l/(1-3l) & l/(1-3l) & l/(1-3l) & 1 \end{bmatrix}$$

i.e.  $l_{jk} = l/(1-l(j-2))$ . Si bien a partir de esta ecuación se puede implementar esta versión del modelo, aún es necesario para llegar a los parámetros  $\gamma$  una transformación adicional, lo que implica el cálculo de  $\delta_j$ , con los problemas ya mencionados que esto conlleva.

La pregunta consiguiente es casi automática, qué clase de modelo en términos marginales, tenemos si  $\gamma_{21} = \dots = \gamma_{p,p-1}$ . Ejemplificando para  $p = 3$  se tiene que

$$\Sigma = \begin{bmatrix} 1 & & & \\ -\gamma/\sqrt{\delta_2} & & & \\ (\gamma^2 - \gamma)/\sqrt{\delta_3} & 1 & & \\ (\gamma^2 - \gamma\delta_2)/\sqrt{\delta_2\delta_3} & (\gamma^2 - \gamma\delta_2)/\sqrt{\delta_2\delta_3} & 1 & \end{bmatrix}.$$

Donde  $\delta_2^{-1} = 1 + \gamma^2$  y  $\delta_3^{-1} = 1 - \gamma^3 + \gamma^2 \left( \frac{\gamma^2 + 2}{\gamma^2 + 1} \right)$ , formas de las cuales no se encontró una interpretación útil. Por lo tanto el modelo equicorrelacionado no es factible a partir de la transformación propuesta.

Un modelo autorregresivo, fijo

$$\Sigma = \begin{bmatrix} 1 & & & \\ \sigma & 1 & & \\ \sigma^2 & \sigma & 1 & \\ \sigma^3 & \sigma^2 & \sigma & 1 \end{bmatrix},$$

representa aún más dificultades, y por tanto tampoco es implementable bajo la propuesta. Los mismo sucede con modelos estructurados o semi estructurados, i.e.  $\sigma_{jk} = \sigma_{lm}$ . Así se concluye que la propuesta es útil bajo dos casos,

el modelo no estructurado, y el modelo con combinaciones de variables independientes,  $\sigma_j = \mathbf{0}$ .

### B.1.3. Diseños muestrales complejos

Se proponen dos alternativas para incluir la información de un diseño muestral complejo:

1. Utilizar una función de los factores de expansión como pesos, en el cálculo de los parámetros a posteriori en cada iteración, fácilmente implementable a partir de las ecuaciones (2.21) y (2.22).
2. En lugar de utilizar los factores de expansión, se utiliza el diseño muestral, representado por las unidades primarias de muestreo (UPM), y los estratos. Para cada uno de estos se incorpora un nuevo nivel jerárquico que permita estimar la varianza dentro del grupo (Hoff, 2009; Congdon, 2010). La mejor manera de realizarlo es incluyendo estos parámetros en los modelos independientes ya transformados, implicaría sin embargo que la interpretación del parámetro no sea sencilla, además de aumentar el costo computacional, sobre todo al incluirlo a nivel de vivienda. Esta misma aproximación puede utilizarse para modelar dependencia espacial.

Ninguna de estas alternativas se encuentra implementada en este momento, sin embargo la implementación es relativamente sencilla.

### B.1.4. Descripción de las funciones en R

Esta sección describe el uso de las funciones que implementan la propuesta; se encuentran programadas en R (R Core Team, 2015). La parte de simulación intensiva se realiza en C++ utilizando la librería *Armadillo* (Sanderson, 2010), integradas a R con el paquete *Rcpp* (Eddelbuettel, 2013). Estas funciones se encuentran disponibles para su uso en:

<https://github.com/likankura/MVP>

La sección se encuentra en inglés en pos del uso generalizado del mismo.

## Function index

---

<code>design_mvp</code>	Design Matrices for regression
-------------------------	--------------------------------

---

### Description

Creates a design object to be used in the MVP fit via `mvp`. This object is also used in the estimation of latent variables and probabilities.

### Usage

```
design_mvp(list.formula,data)
```

### Arguments

`list.formula` A list of formulas, whose length is equal to the length of the response vector. For each response variable a marginal model must be specified.

`data` A dataframe containing the variables specified in `list.formula`.

### Details

The MVP can use different independent variables for each element of the response vector, so each set of variables must be specified.

Due to the algorithm limitations, in order to to return to the original multivariate Gaussian parameterization, it is important to keep record in which the elements of the response vector are included, particularly when  $p > 5$ .

These objects are also used in estimating predicted values via `mvp_predict`.

### Value

A class type `mvp.design` is yet to be implemented. At this moment the returned object is a list that includes: design matrices, the response matrix, variable names, and some fixed vales used by the algorithm. Yet to be implemented without the response variable for new data.

### Example

For some data yet to be included.

```

# A p=2 example with
# same independent variables for each response
des <- design_mvp(list(y1~x1+x2,y2~x1+x2) , our.data)

# More complicated example
des <- designg_mvp(list(
  y2~x3+x2+x4,
  y1~x4,+x1,
  y3~x1+x2+x3+x4), our.data)

```

---

**mvp**


---

 Fit Multivariate Probit Model
 

---

#### Description

Fits a MVP via transformation to achieve conditional independence. Returns a MCMC sample for inference. The returned parameters are transformations of the original probit model parameters, original parameterization estimates can be calculated via `mvp_original`.

#### Usage

```
mvp(design, iter=6000, prior=1, fix=FALSE)
```

#### Arguments

**design** An object generated by `mvp_design` containing the design and response matrices.

**iter** Specifies the length of the Markov Chain to be simulated.

**prior** Either 1 to specify prior densities are proportional to one; or a list with the following structure

```
list( beta=list(mean=c(mean.vector.beta),var=prior.
  matrix.beta),
  gamma=list(mean=c(mean.vector.gamma),
  var=list(matrix.cor.2,matrix.cor.3,...,matrix.
  cor.p)))
```

which correspond to the prior parameters of normal distributions. The prior variance matrix for the `gamma` is a diagonal block matrix, each element of the list corresponds to the covariance matrix for each row in the  $\mathbf{\Gamma}$  matrix. A simplified form is yet to be implemented.

**fix** `FALSE` to specify that all the correlation parameter are free, or a list of two vectors, the first one specifies the  $\gamma$  parameters to fix, and the second specifies the values this parameters are set equal to, default is 0.

#### Details

The algorithm fits a MVP model via latent variables, a transformation and a hierarchical model. As a result the estimated parameters are transformations of the original Gaussian model, and have no direct interpretation in terms of the marginal models for each variable in the response vector.

Heavy computations are done in C++ via `RcppArmadillo`, its an efficient Gibbs sampler.

#### Value

An object class is yet to be implemented, at the moment it returns a list that contains the MCMC samples for  $\beta$  and  $\gamma$ , in matrix forms.

#### Example

```
# A simple example with no additional parameters for priors
# or fixed correlation values.
      #Data simulation
      beta <- runif(4,-beta_root[k],beta_root[k])
      X <- cbind(rep(1,n),runif(n,-1,1))
      med <- as.matrix( cbind(X,X)%*%sparseMatrix(i
=1:4,j=c(rep(1,2),rep(2,2)),x=beta) )
      z <- matrix(rnorm(p*n),nrow=n)%*%rho+med
      y <- z>0
      dat <- data.frame(x1=X[,2],y1=y[,1],y2=y[,2])
      # Model fitting
      dis <- design_mvp(list(y1~x1,y2~x1),dat)
```



```

# Fitting the PM data
# Prior information and design matrix
load('mcs.Rdata')
lineales <- list(
  ic_serviv ~ tamaño_cont + habitantes +
  educa_jefe_cont,
  ic_educacion ~ indigena + edad + sexo,
  ic_salud ~ indigena + edad + sexo,
  ic_calviv ~ drenaje_d + agua_d +
  tamaño_cont + habitantes + sexo_jefe + educa_jefe_cont
+ indigena,
  ic_segsoc ~ eco + tamaño_cont +
  sexo_jefe + educa + indigena + edad + sexo,
  ic_alimenta ~ eco + piso_d + agua_d +
  habitantes + sexo_jefe + educa_jefe_cont + asistencia +
  indigena + edad,
  ic_ingreso ~ drenaje_d + piso_cont +
  agua_d + tamaño + habitantes + sexo_jefe +
  educa_jefe_cont + salud + educa_cont + indigena)

design <- design_mvp(lineales,mcs)

apriori <- list( beta=list(mean=rep(0,k), var=diag
(k)*10 ), gamma= prior_gamma_fun(p=7,read.csv('prior
gamma cubos.csv')))

adj <- mvp(design,prior= apriori,iter=11000)

```

---

**mvp\_predict**

 Generates MCMC samples of the latent variables
 

---

#### Description

Based in MCMC samples of fitted values, it returns a sample of the latent variable (predictive distribution), for either the data used in the fitting (adjusted values) or new data (predicted values).

## Usage

```
mvp_predict(MCMC, design='original', burnin=1000, iter='
original', prob=FALSE)
```

## Arguments

- MCMC** An MCMC object with samples of the regression parameters obtained via `mvp`.
- design** Either a design object for new data, obtained via `mvp_design`, or the default value `original` that uses the same data used in fitting the MVP.
- burnin** A positive integer defining the number of iterations to be burned in the estimation, default value is 1000.
- iter** Number of iterations used in prediction, if equal to `original` the complete sample (minus the burnin) is used. If set to different positive integer value, a bootstrap sample from the MCMC sample is used.
- prob** If `FALSE` no probability is estimated; if set to `all` the probability of each the possible response ( $2^p$ ) is calculated; if  $p$  and  $n$  are big this might take some time and considerable RAM. Yet to be implemented an additional option to compute specific responses probabilities.

## Details

Probabilities are estimated by a Monte Carlo estimation based on the latent distribution sample.

## Value

List containing the latent variables sample for each observation in `desing`, and if `prob='all'` a  $n \times 2^p$  matrix containing the probability of each possible response.

## Example

```
# Simple example
multi <- mvp(dis,iter=6000)
```

```

    ad_mult <-.mvp_predict(multi)

# PM prediction for new data
  load(CPV.Rdata)
  lineales <- list(
    ic_serviv ~      tamaño_cont + habitantes +
    educa_jefe_cont,
    ic_educacion ~  indigena + edad + sexo,
    ic_salud ~      indigena + edad + sexo, #
    eco probablemente
    ic_calviv ~     drenaje_d + agua_d +
    tamaño_cont + habitantes + sexo_jefe + educa_jefe_cont
    + indigena,
    ic_segsoc ~     eco + tamaño_cont +
    sexo_jefe + educa + indigena + edad + sexo,
    ic_alimenta ~   eco + piso_d + agua_d +
    habitantes + sexo_jefe + educa_jefe_cont + asistencia +
    indigena + edad,
    ic_ingreso ~    drenaje_d + piso_cont +
    agua_d + tamaño + habitantes + sexo_jefe +
    educa_jefe_cont + salud + educa_cont + indigena)
    dis <- design.mvp(lineales, CPV)

    load('adj.Rdata') #loads the model a posterior
    distributions
    pred <-.mvp_predict( MCMC=adj, design=dis, prob='
    all')
```

---

`mvp_original`

Computes original Gaussian parameters

---

#### Description

Based in the MCMC sample of  $\beta$  and  $\gamma$ , computes the correspondent sample for the Gaussian parameters  $\mathbf{w}_i \sim MN_p(\mathbf{X}_i\beta^*, \Sigma)$ .

#### Usage

`mvp_original(MCMC)`

### Arguments

**MCMC**            A MCMC object resulting from fitting a MVP via `mvp`.

### Details

Back transforms the mean and variance parameters of a multivariate Gaussian distribution from the transformed parameters used in the `mvp` fit. Very long equations must be used, the equations complexity increases as  $p$  does. At the moment it is just implemented for  $p \leq 5$ . For higher dimensions more fits are required. However this algorithm is yet to be implemented, but it can be done manually, by just changing the order the binary responses are specified in the model. Even if the dimension of the response vector is greater than five only the first ten correlation parameters will be returned, i.e.  $(\sigma_{21}, \sigma_{31}, \sigma_{32}, \sigma_{41}, \sigma_{42}, \sigma_{43}, \sigma_{51}, \sigma_{52}, \sigma_{53}, \sigma_{54})$ , along with the  $\beta^*$  subset corresponding to the first five elements of  $\mathbf{w}_i$ . To be integrated with a summary type function in the near future.

### Value

A list containing matrices of the MCMC sample but transformed into Gaussian parameters.

### Example

```
multi <- mvp(dis,iter=6000)
original_param <- mvp_original(multi)
```

---

`prior_gamma`

Educated correlation apriori definitions

---

### Description

Even when prior information about the correlation parameters is available, due to the transformation used in fitting the model via `mvp`, this information cannot be directly introduced into the model. This function helps making the appropriate transformations based on prior knowledge about the correlation matrix, expressed in terms of quantiles.

## Usage

```
prior_gamma(p, prior.data)
```

## Arguments

**p** A positive integer corresponding to the response vector dimension.

**prior.data** Data frame of  $(p^2 - p)/2$  rows, and three columns: mean, quantile= $q$ ,  $\Pr(\sigma_{jk} < q)$ , containing the corresponding marginal information about  $\Sigma$ .

## Details

In the current implementation, marginal properties are used; but for a better prior distribution a block matrix type is yet to be implemented. The function uses `chol` to transform  $\sigma \rightarrow \gamma$  and `qnorm()` to compute the corresponding variance(s), better specification would include multivariate quantile function, but will be less intuitive.

## Value

A list of matrices that can be directly introduced into `mvp()` in the prior specification.

## Example

```
# for the PM fit
quant_info <- data.frame(
  mean=c(0,0,0.3,0.8,0,0,0.8,0.4,0.2,0.4,0.8 ,0.6,
  0.2,0.6,0.7,0.8,0.4,0.4,0.7,0.8,0.6),
  quantile = c(0.95,0.95,0,0.2,0.95,0.95,0.2,
  0.2,0,0.2,0.2,0,0,0,0.2,0.2,0.2,0,0.2,0
  prob=c( 0.832,0.832,0.3,0.15,0.832,0.832,
  0.15,0.2,0.3,0.2,0.15,0.2,0.3,0.2,0.2,
  0.15,0.2,0.2,0.2,0.15,0.2) )
  k <- 56
  apriori <- list( beta=list(mean=rep(0,k), var=diag
(k)*10 ),
```

```
      gamma= prior_gamma(p=7,prior.data=
quant_info)
    )
adj <-.mvp(design,prior= apriori,iter=11000)
```

Aún queda pendiente refinar algunas funciones, proponer una función tipo `summary`, e integrar como paquete de R, ya se ha avanzado hasta un punto donde se puede visualizar como una posible paquetería funcional.

# Anexo C

## C.1. Resultados del ajuste de la PM

### C.1.1. Variables en el modelo

Se presenta la descripción de las variables incluidas en el ajuste de la PM para Querétaro. El valor entre paréntesis indica la base de datos de donde se obtuvo la variable, CONEVAL o la ENIGH-MCS, en el caso de las variables discretas se presentan los niveles:

**agua\_d** (CONEVAL) Acceso a servicios de agua.

- 1=tiene acceso,
- 2=no tiene acceso.

**asistencia** (CONEVAL) Condición de asistencia escolar.

- 1=Asiste a la escuela,
- 2=No asiste a la escuela.

**drenaje\_d** (CONEVAL) Disponibilidad de drenaje.

- 1=drenaje,
- 2=sin drenaje.

**eco** (CONEVAL) Condición de actividad económica.

- 1=Desocupada,
- 2=Ocupada,
- 3= Población No Económicamente activa.

**edad** (CONEVAL) Edad en años cumplidos.

**educa** (ENIGH-MCS) Nivel educativo.

- 1=menos de primaria,
- 2=primaria,
- 3=secundaria,
- 4=media superior,
- 5=superior.

**educa\_jefe** (ENIGH-MCS) Nivel educativo del jefe del hogar.

- 1=sin escolaridad,
- 2=educación básica,
- 3=educación media superior,
- 4=educación superior.

**habitantes** (ENIGHMCS) Número de habitantes en la vivienda.

**ic\_alimenta** (CONEVAL) Indicador de carencia por acceso a la alimentación.

- 0=no presenta carencia,
- 1=presenta carencia.

**ic\_calviv** (CONEVAL) Indicador de carencia por calidad y espacios de la vivienda.

- 0=no presenta carencia,
- 1=presenta carencia.

**ic\_educ** (CONEVAL) Indicador de carencia por rezago educativo.

- 0=no presenta carencia,
- 1=presenta carencia.

**ic\_salud** (CONEVAL) Indicador de carencia por acceso a servicios de salud.

- 0=no presenta carencia,
- 1=presenta carencia.

**ic\_segso** (CONEVAL) Indicador de carencia por acceso a la seguridad social.

- 0=no presenta carencia,
- 1=presenta carencia.

**ic\_serviv** (CONEVAL) Indicador de carencia por servicios básicos en la vivienda.



0=no presenta carencia,  
1=presenta carencia.

**indigena** (CONEVAL) Hablante de lengua indígena.

0=no hablante,  
1=hablante.

**Ingreso** (CONEVAL) Población con un ingreso menor a la línea de bienestar.

0=Ingreso mayor a la LB,  
1=Ingreso menor a la LB.

**piso** (ENIGH-MCS) Principal material del piso.

1=tierra,  
2=cemento,  
3=mosaico.

**piso\_d** (CONEVAL) Condición de piso de tierra.

1=no tierra,  
2=tierra.

**sexo** (CONEVAL) Sexo.

1=hombre,  
2=mujer.

**sexo\_jefe** (ENIGH-MCS) Sexo del jefe del hogar.

1=hombre,  
2=mujer.

**tamaño** (CONEVAL) Tamaño de la localidad.

1=hasta 2499 habitantes,  
2=hasta 14999 habitantes,  
3=hasta 99999 habitantes,  
4=al menos 100000 habitantes.

### C.1.2. Estimación de los parámetros

Se presentan los resultados de los ajustes para los parámetros del modelo.

Cuadro C.1: Parámetros de correlación  $\gamma_{jk}$ 

	Mean	Límite inferior	Límite superior
ic_educacion & ic_serviv	-0.10	-0.15	-0.06
ic_salud & ic_serviv	-0.01	-0.05	0.04
ic_salud & ic_educacion	-0.03	-0.07	0.01
ic_calviv & ic_serviv	-0.04	-0.11	0.02
ic_calviv & ic_educacion	-0.08	-0.14	-0.03
ic_calviv & ic_salud	0.02	-0.03	0.08
ic_segsoc & ic_serviv	-0.10	-0.14	-0.05
ic_segsoc & ic_educacion	-0.04	-0.09	0.00
ic_segsoc & ic_salud	-0.46	-0.51	-0.42
ic_segsoc & ic_calviv	-0.03	-0.07	0.01
ic_alimenta & ic_serviv	-0.03	-0.08	0.01
ic_alimenta & ic_educacion	-0.06	-0.11	-0.02
ic_alimenta & ic_salud	0.02	-0.03	0.07
ic_alimenta & ic_calviv	-0.03	-0.07	0.02
ic_alimenta & ic_segsoc	-0.04	-0.09	-0.00
ic_ingreso & ic_serviv	0.00	-0.05	0.05
ic_ingreso & ic_educacion	-0.00	-0.05	0.04
ic_ingreso & ic_salud	0.05	-0.01	0.11
ic_ingreso & ic_calviv	-0.04	-0.08	0.01
ic_ingreso & ic_segsoc	-0.21	-0.25	-0.17
ic_ingreso & ic_alimenta	-0.12	-0.16	-0.08

Cuadro C.2: Parámetros de media  $\beta_{jl}$ 

	Media	Límite inferior	Límite superior
ic_serviv\$(Intercept)	0.87	0.61	1.13
ic_serviv\$tamaño_cont	-0.67	-0.73	-0.61
ic_serviv\$habitantes	-0.01	-0.03	0.01
ic_serviv\$educa_jefe_cont	-0.25	-0.36	-0.14
ic_educacion\$(Intercept)	-1.66	-2.00	-1.27
ic_educacion\$indigenano lengua	-0.03	-0.37	0.33
ic_educacion\$edad	0.02	0.02	0.03

Cuadro C.2: Parámetros de media  $\beta_{jl}$ 

	Media	Límite inferior	Límite superior
ic_educacion\$sexo2	0.08	0.00	0.16
ic_salud\$(Intercept)	-0.71	-1.11	-0.32
ic_salud\$indigenano lengua	0.23	-0.15	0.61
ic_salud\$edad	-0.00	-0.01	-0.00
ic_salud\$sexo2	-0.13	-0.20	-0.06
ic_calviv\$(Intercept)	-1.38	-1.89	-0.89
ic_calviv\$drenaje_dsin drenaje	0.48	0.30	0.66
ic_calviv\$agua_dsin agua	0.29	0.11	0.46
ic_calviv\$tamaño_cont	-0.07	-0.12	-0.01
ic_calviv\$habitantes	0.14	0.11	0.16
ic_calviv\$sexo_jefeJefes mujeres	0.07	-0.07	0.21
ic_calviv\$educa_jefe_cont	-0.21	-0.30	-0.12
ic_calviv\$indigenano lengua	-0.28	-0.73	0.17
ic_segsoc\$(Intercept)	3.27	2.75	3.76
ic_segsoc\$ecoocupada	-0.55	-0.75	-0.33
ic_segsoc\$ecoPNEA	-0.81	-1.03	-0.60
ic_segsoc\$tamaño_cont	-0.28	-0.32	-0.25
ic_segsoc\$sexo_jefeJefes mujeres	0.08	-0.01	0.18
ic_segsoc\$educaprimaria	0.08	-0.09	0.25
ic_segsoc\$educasecundaria	-0.25	-0.44	-0.06
ic_segsoc\$educamedia_sup	-0.32	-0.53	-0.11
ic_segsoc\$educasuperior	-0.42	-0.63	-0.23
ic_segsoc\$indigenano lengua	-0.54	-0.97	-0.15
ic_segsoc\$edad	-0.02	-0.02	-0.02
ic_segsoc\$sexo2	-0.07	-0.15	0.02
ic_alimenta\$(Intercept)	-0.91	-1.51	-0.36
ic_alimenta\$ecoocupada	-0.33	-0.50	-0.16
ic_alimenta\$ecoPNEA	-0.33	-0.51	-0.16
ic_alimenta\$ piso_dtierra	0.70	0.44	0.98
ic_alimenta\$agua_dsin agua	0.12	-0.03	0.26
ic_alimenta\$habitantes	0.06	0.04	0.08
ic_alimenta\$sexo_jefeJefes mujeres	0.00	-0.10	0.10
ic_alimenta\$educa_jefe_cont	-0.25	-0.30	-0.19
ic_alimenta\$asistenciaNo asiste a la escuela	-0.03	-0.16	0.10

Cuadro C.2: Parámetros de media  $\beta_{jl}$ 

	Media	Límite inferior	Límite superior
ic_alimenta\$indigenano lengua	0.67	0.17	1.19
ic_alimenta\$edad	-0.00	-0.00	0.00
ic_ingreso\$(Intercept)	1.31	0.88	1.74
ic_ingreso\$drenaje_dsin drenaje	0.37	0.21	0.54
ic_ingreso\$ piso_cont	-0.36	-0.45	-0.27
ic_ingreso\$agua_dsin agua	0.76	0.59	0.93
ic_ingreso\$tamañoh14999	0.51	0.40	0.63
ic_ingreso\$tamañoh99999	0.29	0.12	0.45
ic_ingreso\$tamaño100000+	0.45	0.33	0.56
ic_ingreso\$habitantes	0.05	0.04	0.07
ic_ingreso\$sexo_jefeJefes mujeres	0.05	-0.05	0.15
ic_ingreso\$educa_jefe_cont	-0.33	-0.40	-0.27
ic_ingreso\$saludSin derechohabiencia	0.00	-0.14	0.13
ic_ingreso\$educa_cont	-0.19	-0.23	-0.15
ic_ingreso\$indigenano lengua	0.02	-0.35	0.38

# Referencias

- Agresti, A. (1997). A Model for Repeated Measurements of a Multivariate Binary Response. *Journal of the American Statistical Association*, 92(437):315.
- Albert, J. H. y Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. 88(88:422):669–679.
- Arminger, G., Clogg, C. C., y Cheng, T. (2008). Regression Analysis of Multivariate Binary Response Variables Using Rasch-Type Models and Finite-Mixture Methods. *Sociological Methodology*, 30(1):1–26.
- Bates, D. y Maechler, M. (2015). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.2-0.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer Science+Business Media.
- Bock, R. D. y Gibbons, R. D. (1996). High-Dimensional Multivariate Probit Analysis. *Biometrics*, 52(4):1183–1194.
- Calia, P. y Ferrante, M. R. (2013). How do firms combine different internationalisation modes? A multivariate probit approach. *Review of World Economics*, 149(4):663–696.
- Carnap, R. (1989). El carácter metodológico de los conceptos teóricos. Siglo XXI, México.
- Chib, S. (2000). Bayesian Methods for Correlated Binary Data. In Dey, D., editor, *Generalized Linear Models A Bayesian Perspective*, chapter 7. Marcel Dekker, New York.

- Chib, S. y Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, (85):347–361.
- Chopin, N. (2011). Fast simulation of truncated Gaussian distributions. *Statistics and Computing*, 21(2):275–288.
- Congdon, P. (2005). *Bayesian models for categorical data*. John Wiley & Sons.
- Congdon, P. D. (2010). *Applied Bayesian Hierarchical Methods*. Chapman and Hall/CRC.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2007). Los mapas de Pobreza en México, Anexo técnico metodológico. Technical report, Consejo Nacional de Evaluación de la Política de Desarrollo Social, México.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2010). Lineamientos y criterios generales para la definición, identificación y medición de la pobreza.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2011). Metodología para la medición de la pobreza en los municipios de México, 2010. Technical report, Consejo Nacional de Evaluación de la Política Social.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2013a). Anexo estadístico para entidades federativas de la medición de la pobreza 2010-2012.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2013b). Estimación de la pobreza 2010-2012.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2013c). Nota técnica del Rezago Social en las AGEB urbanas en México. Technical report, Consejo Nacional de Evaluación de la Política de Desarrollo Social, México.
- Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL (2015). Medición de la pobreza en México y en las entidades federativas 2014, resumen ejecutivo.

- Consejo Nacional de Población, CONAPO (2012). *Índice de marginación urbana 2010*. Consejo Nacional de Población, México.
- Consejo Nacional de Población, CONAPO (2013). *Índice absoluto de marginación 2000-2010*. Consejo Nacional de Población, México.
- Coromaldi, M. y Zoli, M. (2012). Deriving Multidimensional Poverty Indicators: Methodological Issues and an Empirical Analysis for Italy. *Social Indicators Research*, 107(1):37–54.
- Demombynes, G., Elbers, C., Lanjouw, J. O., y Lanjouw, P. (2007). *How Good A Map? Putting Small Area Estimation To The Test*. The World Bank.
- Dempster, A. P., Laird, N. M., y Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1).
- D’Ambrosio, C., Deutsch, J., y Silber, J. (2011). Multidimensional approaches to poverty measurement: an empirical analysis of poverty in Belgium, France, Germany, Italy and Spain, based on the European panel. *Applied Economics*, 43(February 2015):951–961.
- Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York. ISBN 978-1-4614-6867-7.
- Eddelbuettel, D. y Sanderson, C. (2014). Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational Statistics and Data Analysis*, 71:1054–1063.
- Elbers, C., Lanjouw, J. O., y Lanjouw, P. (2002). *Micro-Level Estimation of Welfare*. The World Bank.
- Elbers, C., Lanjouw, J. O., y Lanjouw, P. (2003). Micro–Level Estimation of Poverty and Inequality. *Econometrica*, 71(1):355–364.
- Gibbons, R. y Lavigne, J. (1998). Emergence of childhood psychiatric disorders: A multivariate probit analysis. *STATISTICS IN MEDICINE*, 17(21):2487–2499.
- Greene, W. H. (2002). *Econometrics Analysis*. Prentice Hall, 5th edition.

- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer-Verlag, New York.
- Imai, K. y Van Dyk, D. a. (2005). A Bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124:311–334.
- Instituto Nacional de Estadística y Geografía, INEGI (2011). *Modulo de Condiciones Socioeconómicas 2010, Encuesta Nacional de Ingresos y Gastos de los Hogares, Descripción de la base de datos*.
- Jackman, S. S. U. (2009). *Bayesian Analysis for the Social Sciences*. John Wiley & Sons, 1st edition.
- Kakwani, N. y Silber, J. (2008). Introduction: Multidimensional Poverty Analysis: Conceptual Issues, Empirical Illustrations and Policy Implications. *World Development*, 36(6):987–991.
- Kass, R. E. y Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773.
- Kassie, M., Jaleta, M., Shiferaw, B., Mmbando, F., y Mekuria, M. (2013). Adoption of interrelated sustainable agricultural practices in smallholder systems: Evidence from rural Tanzania. *Technological Forecasting and Social Change*, 80(3):525–540.
- Levy Algazi, S. (2007). *¿Qué sigue después de Progres-Oportunidades?* Inter-American Development Bank.
- Liang, K.-Y., Zeger, S. L., y Qaqish, B. (1992). Multivariate regression analyses for categorical data.
- Luhmann, N. (1997). Organización y Decisión. Autopoiesis, Acción y Entendimiento comunicativo. In *Organización y Decisión. Autopoiesis, Acción y Entendimiento comunicativo*, chapter 1, page 98. Anthropos, Barcelona.
- McCulloch, R. E., Polson, N. G., y Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99:173–193.
- Nilsson, W. (2012). Heterogeneity or true state dependence in poverty: The tale told by twins. *Review of Income and Wealth*, 58(1):1–23.



- O'Brien, S. M. y Dunson, D. B. (2004). Bayesian multivariate logistic regression. *Biometrics*, 60(3):739–46.
- Paap, R., van Nierop, E., van Heerde, H. J., Wedel, M., Franses, P. H., y Alsem, K. J. (2005). Consideration sets, intentions and the inclusion of “don't know” in a two-stage model for voter choice. *International Journal of Forecasting*, 21(1):53–71.
- Parsons, T. (1988). *El sistema social*. Alianza.
- Pérez-Mayo, J. (2005). Identifying deprivation profiles in Spain: a new approach. *Applied Economics*, 37(February 2015):943–955.
- Pirani, E. (2013). Evaluating contemporary social exclusion in Europe: A hierarchical latent class approach. *Quality and Quantity*, 47(2):923–941.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rojas, M. (2008). Experienced Poverty and Income Poverty in Mexico: A Subjective Well-Being Approach. *World Development*, 36(6):1078–1093.
- Sanderson, C. (2010). Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments. Technical report, NICTA.
- Schopenhauer, A. (2002). *El arte de tener razón :expuesto en 38 estratagemas*. Alianza, México.
- Seetharaman, P. B., Chib, S., Ainslie, A., Boatwright, P., Chan, T., Gupta, S., Mehta, N., Rao, V., y Strijnev, A. (2005). Models of multi-category choice behavior. *Marketing Letters*, 16(3-4):239–254.
- Shapiro, H. (2015). *Linear Algebra and Matrices: Topics for a Second Course*. American Mathematical Soc.
- Street, K. y Czado, C. (1996). Multivariate Probit Analysis of Binary Time Series Data with Missing Responses. *Time*, 23:1–16.
- Swihart, B. J., Caffo, B. S., y Crainiceanu, C. M. (2014). A unifying framework for marginalised random-intercept models of correlated binary outcomes. *International Statistical Review*, 82(2):275–295.

- Szeles, M. R. y Fusco, A. (2013). Item response theory and the measurement of deprivation: Evidence from Luxembourg data. *Quality and Quantity*, 47(3):1545–1560.
- The Sage Developers (2015). *Sage Mathematics Software (Version 6.9)*. <http://www.sagemath.org>.
- Weber, M. (1992). *Economía y Sociedad: esbozo de sociología comprensiva*. FCE, México.
- Zhao, L. P. y Prentice, R. L. (1990). Correlated binary regression using a generalized quadratic model. *Biometrika*, 77(3):642–648.