

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS, A.C.



**ANÁLISIS DE LA TENDENCIA DEL INDICADOR DE OPORTUNIDAD DEL
REGISTRO ADMINISTRATIVO DE NACIMIENTOS EN
LA REPÚBLICA MEXICANA**

TESIS

Que para obtener el grado de
MAESTRO EN CIENCIAS EN ESTADÍSTICA OFICIAL

Presenta

Gustavo Aguilar Mata

Director de Tesis

DR. ROGELIO RAMOS QUIROGA

Guanajuato, Gto. Noviembre de 2010

DEDICATORIA

A Graciela Dolores, la persona más importante en mi existencia, a quien amo y quiero.

A mis padres Manuel y Concepción, cuyo esfuerzo y dedicación me dieron la oportunidad de estudiar. Gracias por todo el apoyo y comprensión.

A mis hijos Gail, Diego y Mariana, con todo mi cariño y con el sentimiento más noble y esperanzador.

A mi hermana Nora Elsa, con mucho cariño.

A mis hermanas Teresa, Diana y Claudia.

A mis hermanos Manuel, Fernando, Ricardo, Carlos y Humberto.

Al Centro de Investigación en Matemáticas Aplicadas, A.C.

Al Doctor Rogelio Ramos Quiroga, por su orientación, dirección, disposición, sugerencias y el tiempo invertido en este trabajo. Doctor, también le agradezco todos los documentos de investigación y docencia que ha elaborado, así como las rutinas de R, sin ellos no podría haber realizado este trabajo.

A mi sinodal Maestro Sergio Nava Muñoz, por compartir sin miramientos su conocimiento estadístico y matemático, muchas gracias Maestro Sergio.

Al mi sinodal Doctor Enrique Raúl Villa Diharce por su apoyo y tiempo dedicado a este trabajo, muchas gracias Doctor Enrique.

AGRADECIMIENTOS

Gracias especiales al Lic. Francisco Javier Solís Guerrero por su valioso apoyo.

Gracias a Antonio Pozzi P., Flavio de Alba G., Eduardo de Alba G., Ignacio Rodríguez B., Carlos Guzmán T., Víctor Rosado M., Rubén de la Torre I., Antonio Escobedo A., Toribio Sánchez N., José Vences R. y Miguel A. Suárez C., por toda su enseñanza profesional.

ÍNDICE

Capítulo I	Antecedentes	
	1.1 El sistema de estadísticas vitales.	1
	1.2 El proceso de Captación de las estadísticas vitales.	3
	1.3 Los modelos lineales de efectos mixtos.	7
Capítulo II	El indicador de oportunidad del registro y el modelo lineal mixto	
	2.1 El indicador de oportunidad del registro administrativo de nacimientos	13
	2.2 Los datos de panel	16
	2.3 El modelo lineal de efectos fijos	18
	2.4 El modelo lineal de efectos aleatorios	20
	2.5 El modelo lineal de efectos mixtos	24
	2.6 Nivel individual y multinivel del modelo lineal mixto	28
	2.7 Estimación máximo verosímil para el modelo lineal mixto	31
	2.8 Estimación máximo verosímil restringida para el modelo lineal mixto	41
	2.9 Estimación máximo verosímil en múltiples niveles de efectos aleatorios	43
Capítulo III	Estimación de los modelos de efectos fijos, aleatorios y mixtos	
	3.1 Los datos	48
	3.2 El modelo lineal simple	50
	3.3 El modelo efectos fijos	51
	3.4 El modelo de efectos mixtos	56
	3.5 Las funciones de varianza	60
	3.6 Las estructuras de correlación	67
Capítulo IV	El análisis de modelo y la proyección del IOR	
	4.1 El análisis del modelo	83
	4.2 Los pronósticos del modelo	92
Capítulo VI	Conclusiones	99
	Relación de Archivos	103
	Bibliografía	107
	Anexos	111

INTRODUCCIÓN

El Registro Civil es un organismo administrativo o servicio público, encargado de dejar constancia de los hechos o actos relativos al estado civil de las personas naturales, así como otros que las leyes le encomienden.

En el Registro Civil se inscriben los nacimientos, la filiación, el nombre y apellido de las personas, entendiéndose la filiación como el vínculo jurídico que existe entre dos personas donde una es descendiente de la otra, sea por un hecho natural o por un acto jurídico.

Es indudable que el tiempo transcurrido entre el nacimiento de las personas y su registro en las oficinas del Registro Civil está íntimamente ligado a la cultura de la sociedad mexicana y a la promoción y oferta de servicios que las mismas oficinas realizan en cada uno de los estados.

Dentro de las modalidades de tesis que ofrece el CIMAT para obtener el grado el presente trabajo se ubica en adaptar metodología específica para el análisis de problemas de interés del INEGI.

El objetivo principal de esta tesis es presentar el Indicador de Oportunidad del Registro y realizar la proyección del mismo para las treinta y dos entidades federativas del periodo 2008-2010 utilizando como herramienta principal el método estadístico conocido como Modelos Lineales Mixtos y la información triangular de los nacimientos registrados por entidad federativa del periodo 1994-2007.

Como justificación a este trabajo esta la necesidad por parte de las Oficinas Locales de Gobierno y de la Dirección de Registros Administrativos del INEGI de contar con un instrumento que permita conocer y sobre todo medir a través del tiempo el comportamiento de este fenómeno social.

El Indicador de oportunidad del Registro (IOR) se construye tomando en cuenta la información de los nacimientos registrados presentada en formato de triangulo run off, dando como resultado un indicador que mide la forma de la distribución de la ocurrencia de los nacimientos con respecto a su año de registro

Los Modelos Lineales Mixtos, por el cual los datos se analizan en forma longitudinal para obtener un solo modelo y que sirve para satisfacer diferentes necesidades de información, desde el nivel nacional hasta por entidad federativa.

Para facilitar el cumplimiento del objetivo principal se ha estructurado este trabajo en cinco capítulos.

CAPÍTULO I: ANTECEDENTES

En donde se describen los hechos históricos más importantes que han ocurrido en la historia del registro administrativo de nacimientos, así como los antecedentes del desarrollo estadístico y de computación de los modelos lineales mixtos. Reseñar y aprender del pasado es un buen punto de partida para analizar el futuro.

CAPÍTULO II: EL INDICADOR DE OPORTUNIDAD DEL REGISTRO Y EL MODELO LINEAL MIXTO

En donde se describe las características y alcances del Indicador de Oportunidad de Registro así como la notación y fundamentación teórica de los Modelos Lineales Mixtos.

CAPÍTULO III: ESTIMACIÓN DE LOS MODELOS DE EFECTOS FIJOS, ALEATORIOS Y MIXTOS

En donde se presentan particularmente los modelos de efectos fijos, aleatorios y mixtos, con el propósito de obtener estadísticamente el modelo más adecuado que se ajuste a los datos, examinando una gran variedad de estructuras de varianzas, pruebas ANOVA y análisis de residuales.

CAPÍTULO IV: EL ANÁLISIS DE MODELO Y LA PROYECCIÓN DEL IOR

En donde se presenta los componentes del modelo como son: parámetros, varianzas y covarianzas de efectos fijos y aleatorios, estructura de errores autocorrelacionados y con base en estos resultados obtener las proyecciones del Indicador de Oportunidad del Registro para el periodo 2008-2010.

CAPÍTULO V: CONCLUSIONES Y COMENTARIOS

En donde se presentan los comentarios y conclusiones generales.

CAPÍTULO I
ANTECEDENTES

1.1 El sistema de estadísticas vitales.

El Sistema de Estadísticas Vitales de Nacimientos consiste en un proceso que recaba información mediante el registro civil, sobre el registro de nacimiento de la persona, así como las características generales de los padres, además permite recopilar, procesar, analizar, evaluar, presentar y difundir esos datos en forma estadística.

En el país, el Sistema Nacional de Estadísticas Vitales de Nacimientos es una de las fuentes de información demográfica con mayor tradición, sus antecedentes inmediatos fueron los registros parroquiales que realizaba la Iglesia Católica en la Nueva España, con el propósito de registrar los bautizos.

Durante muchos años estos registros parroquiales fueron la única fuente de información demográfica en el país, proporcionada por los curatos y vicarías, para conocer, en forma aproximada, el crecimiento natural de la población.

Un suceso que contribuyó para la conformación del Sistema de las Estadísticas Vitales de Nacimientos, ocurrió en 1857 durante el gobierno de Ignacio Comonfort, cuando propone la creación de la Ley Orgánica del Registro Civil, la cual, a pesar de su importancia, no entró en vigor por contravenir con los intereses del clero.

Esta ley reconocía como actos del estado civil los nacimientos y adopciones, asimismo, marcaba un plazo para el registro y obligaba a los sacerdotes a informar diariamente a la autoridad civil, de todos los actos que celebraran. En caso de no hacerlo causaría una multa, y si reincidían en esta omisión, serían acusados ante la autoridad eclesiástica.

Como se mencionó, esta ley no entró en vigor en el año de su promulgación, lo cual fue posible hasta 1859, gracias a la obra legislativa del gobierno de Benito Juárez, como parte de las Leyes de Reforma, donde además de deslindar las funciones entre la Iglesia y el Estado, se exoneró a los sacerdotes de informar sobre los actos del estado civil.

El funcionamiento del Registro Civil, en sus inicios, fue bastante irregular debido a la situación política del país, especialmente durante el tiempo que duró la ocupación francesa, ya que fueron cerrados los juzgados, y reinstalados cuatro años después con el encargo adicional de proporcionar la información para la elaboración de las estadísticas.

Debido a las circunstancias mencionadas anteriormente en donde inclusive se aplicaron durante algún tiempo los códigos franceses para delitos penales, durante muchos años los datos provenientes del Registro Civil estuvieron muy por debajo de los datos registrados por las autoridades eclesiásticas, por lo que nuevamente por obra del Presidente Benito Juárez en 1871 se establece la obligatoriedad de los padres de inscribir a sus hijos en esta institución.

Posteriormente, ese este mismo año se expidió el reglamento de los juzgados del Registro Civil; y más tarde, en 1874, se decretó que estas leyes fueran adicionadas a la Constitución de 1857, para reglamentar el registro civil en México. Entre las disposiciones fundamentales de esta Ley, destaca que son actos del estado civil los nacimientos, las adopciones y los reconocimientos y serán encargados de la función registral únicamente los jueces del Registro Civil excluyendo de esta manera al Clero.

En 1883 se creó la Dirección General de Estadística, con la responsabilidad de generar las estadísticas del país, quien publicó varios documentos y boletines referidos tanto a los lineamientos que debería seguir la elaboración de estadísticas, como datos sobre el movimiento de la población. Esta información se publicaba en el Boletín Estadístico, el cual en 1893 fue sustituido por el Anuario Estadístico, en donde se prosiguió con la publicación de los datos de nacimientos, matrimonios, defunciones y nacidos muertos.

A partir de ese año se establece en México, el Sistema de Estadísticas Vitales basado en información recolectada por el registro civil de cada entidad federativa mediante la transcripción de la información de los actos del estado civil y los hechos vitales.

En 1917 se promulgó la Ley sobre Relaciones Familiares; en ésta se estableció que la familia debía sustentarse sobre bases más justas y racionales, lo que permitió la disolución del vínculo matrimonial a través del divorcio, entre otras disposiciones.

Por último, en 1974 se incorporaron como fuentes de información las Agencias del Ministerio Público, con el propósito de mejorar la cobertura sobre los nacidos muertos, a los cuales cambió su nomenclatura por el de Muertes Fetales.

Uno de los usos relevantes de las estadísticas vitales corresponde al análisis demográfico, y se le puede considerar como requisito previo para la planificación del desarrollo económico y social de un país. Estas estadísticas proporcionan información de la tasa y la tendencia del crecimiento natural de la población y sobre la conducta de sus componentes, mediante su agregación a lo largo del tiempo, sobre el tamaño de la población su estructura y distribución geográfica. También se usan para obtener estimaciones del tamaño de la población y realizar proyecciones de población.

Cabe señalar que los eventos demográficos están relacionados con factores económicos y sociales que abarcan aspectos importantes en el proceso de desarrollo del país, como son la urbanización, la educación, vivienda, ingreso, estructura familiar, trabajo, etcétera.

La estadística de nacimientos tiene su insumo en los registros administrativos, a través de las actas emitidas por el Registro Civil.

El acta de nacimiento del registro civil constituye una documentación jurídica esencial de su identidad y estado civil, a saber: nombre, fecha, hora, lugar de nacimiento, nacionalidad, edad de los padres, filiación, sexo y nacionalidad

(ciudadanía), de la que depende una amplia gama de derechos y actividades individuales y familiares.

Para fines administrativos, los registros de nacimientos constituyen la base de los programas de sanidad destinados al cuidado postnatal de la madre y el niño, y pueden utilizarse cuando sea necesario en los programas de vacunación e inmunización.

La información estadística sobre nacimientos aporta, además, datos sobre la edad en que las mujeres están concibiendo, en que edades se concentran más los nacimientos, la oportunidad del registro, la condición de salud en que se hace la comparecencia, el tamaño de la familia y la distribución de los nacimientos según el tamaño de la localidad de residencia habitual de la madre; y por último, la educación y nivel socioeconómico de los padres.

Una mayor información sobre las características del hecho, del registrado y de los padres, no solamente mejorará su comprensión, además proporcionará los elementos necesarios para la elaboración de programas sociales por parte de la administración pública.

1.2 El proceso de Captación de las estadísticas vitales.

Para el caso de la estadística de nacimientos la unidad de observación son los nacimientos registrados en las oficinas del Registro Civil, distribuidas en todo el territorio nacional.

La presentación de resultados, se realiza de acuerdo a tres ámbitos geográficos: lugar de registro, lugar de ocurrencia y lugar de residencia habitual de la madre, diferenciando en los dos últimos los acaecidos o residentes en el extranjero.

Los sistemas de información concebidos como el medio idóneo para estructurar, organizar y administrar la información, implican que su diseño y operación integren todos los conceptos, instrumentos y procedimientos necesarios para abordar y explicar la realidad.

Un aspecto importante para la operación de un sistema de información es la base sobre la cual se captan los datos; en este sentido, es necesario señalar que en el ámbito sociodemográfico existen tres fuentes de información básica:

- ❖ Censos
- ❖ Encuestas
- ❖ Registros administrativos

En el INEGI, la Dirección General de Estadísticas Sociodemográficas (DGES), es la responsable del Sistema Nacional de Estadísticas Vitales, entre las que se ubica la estadística de nacimientos.

Las oficinas del registro civil son las fuentes de información continua de esta estadística, y proporciona mensualmente copias de las actas o archivos electrónicos de los registros de nacimientos; para el año 2007 se contó con un total de 5 111 oficinas distribuidas en todo el territorio nacional.

Las actividades del proceso de generación de la estadística de nacimientos, abarcan desde el diseño de formatos de control de la información, hasta el diseño de los programas para el procesamiento electrónico y publicación de resultados.

La producción de las estadísticas vitales se desarrolla en las direcciones regionales del INEGI, estratégicamente ubicadas en el territorio nacional; cada una tiene a su cargo tres o cuatro entidades federativas, donde se ubican las coordinaciones estatales.

A continuación se describen, en términos generales, las fases del proceso de generación:

Oficina Central

1. Elabora el diseño conceptual de la estadística y define los criterios técnicos del proceso de generación.
2. Diseña o actualiza los formatos que se requieren para el seguimiento del proceso.
3. Prepara los requerimientos necesarios para solicitar información en medios magnéticos.
4. Capacita al personal involucrado en el proceso de generación, en los ámbitos estatal y regional.
5. Envía a Direcciones Regionales los formatos, requerimientos de información en medios magnéticos y los instructivos correspondientes.

Dirección Regional

6. Recibe y distribuye de acuerdo con las necesidades de cada entidad federativa de su jurisdicción, los formatos de control y seguimiento, los requerimientos con los que se podrá solicitar información en medios magnéticos y los instructivos con los criterios que se aplicarán durante la producción.

Coordinación Estatal

7. Mantiene actualizado el directorio de las fuentes informantes (en este caso de la Oficialías del Registro Civil), que son responsables de proporcionar la información que será integrada a la estadística.

8. Distribuye a las Oficialías del Registro Civil los formatos que servirán para el seguimiento de la producción, a partir de la recolección, así como los requerimientos de información en medios magnéticos.

9. Participa en la capacitación del personal de las fuentes informante, sobre la captación de los datos estadísticos así como en el llenado de los formatos de control.

10. Determina con la fuente informante el tiempo y forma en que será entregada la información, observando la normatividad establecida. Oficialía del Registro Civil.

11. Registra los nacimientos que le sean solicitados por la población.

12. Prepara la información que será entregada al INEGI con una copia impresa del acta o en archivos magnéticos.

13. Entrega al INEGI la información solicitada en el tiempo y forma establecidos con la Coordinación Estatal.

Coordinación Estatal

14. Recolectan la información en paquetes (impresa) o archivos electrónicos, de las fuentes informantes, la registran, se revisa el contenido y la congruencia de los datos asentados en los instrumentos de captación, se realiza el tratamiento manual e inicia el electrónico, posteriormente la envían a las direcciones regionales.

Dirección Regional

15. Supervisan la producción estatal y en su caso apoyan con la asesoría que sea necesaria para aplicar el procedimiento correcto en la producción.

16. Verifica que la recolección y tratamiento se realice de acuerdo con los lineamientos establecidos y elabora los reportes de avance mensual para reportar a oficina central.

17. Integra el archivo estatal, concluye el proceso electrónico, genera y revisa los reportes establecidos para garantizar su calidad y cobertura; una vez aprobados los remite a oficina central.

Oficina Central

18. Recibe los archivos estatales y concluye el proceso electrónico e integra revisa y libera el archivo nacional de la estadística.

19. Pone a disposición de la población los resultados, a través de medios impresos y electrónicos.

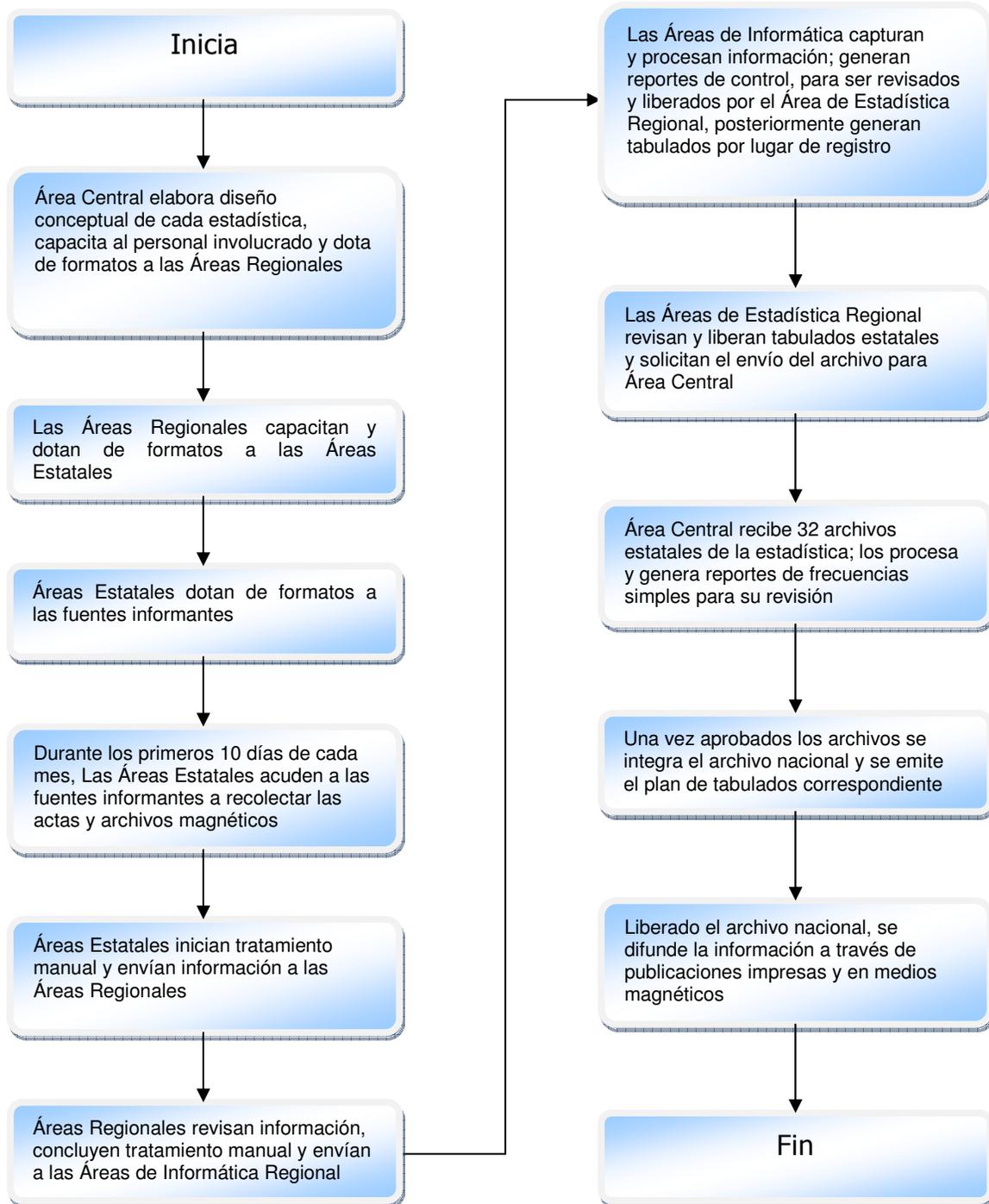


Figura 1.1 Diagrama de flujo del proceso de generación de la estadística de nacimientos

1.3 Los modelos lineales de efectos mixtos.

Esta sección tiene el propósito de presentar en forma sucinta las actividades más sobresalientes que históricamente se han desarrollado para sustentar los modelos lineales mixtos (MLM) de la actualidad. Se dividen los comentarios en dos partes: La primera presentada en un formato de línea del tiempo se refiere a la aportación que a través de los años han realizado los estadísticos profesionales y la segunda al desarrollo del software.

- ❖ 1861: La primera formulación del modelo de efectos aleatorios fue propuesto por George Bidell Airy recolectando observaciones astronómicas para diferentes noches separando la varianza de los efectos aleatorios nocturnos de los residuales obtenidos por estimaciones para cada noche. Bidell Airy fue director del observatorio astronómico de Greenwich en Inglaterra de 1835 a 1881.
- ❖ 1863: El profesor norteamericano William Chauvenet calcula la varianza de los efectos aleatorios de un modelo simple.
- ❖ 1925: El estadístico inglés Ronald Aylmer Fisher presenta su libro "Statistical Methods for Research Workers" considerado el libro del siglo XX con más influencia en los métodos estadísticos, el libro presenta métodos generales para estimar los componentes de la varianza tomando en cuenta la variación aleatoria para datos balanceados.
- ❖ 1927: El estadístico inglés George Udny Yule asume la dependencia explícita de los residuales y construye los modelos de correlación serial.
- ❖ 1931: El estadístico inglés Leonard Henry Caleb Tippett extendió los trabajos de Fisher de los métodos ANOVA para estimar la varianza de los efectos aleatorios.
- ❖ 1935: El matemático polaco Jerzy Neyman examinó la eficiencia de los bloques aleatorios y extendió el uso de los modelos lineales mixtos en sus trabajos.
- ❖ 1938: La séptima edición del libro de Fisher "Statistical Methods for Research Workers" incluye el tema de los coeficientes de correlación intra-clase (ICC).
- ❖ 1939: Jackson asume normalidad para los efectos aleatorios y es el primero en introducir el término "efecto" en el contexto de los MLM.
- ❖ 1940: El matemático austro húngaro Abraham Wald propone los intervalos de confianza para las razones de componentes de varianza.
- ❖ 1941: Ganguli aplicó técnicas ANOVA asociadas a efectos aleatorios para modelos mixtos anidados.
- ❖ 1946: Ganguli y Crump fueron los primeros en mencionar el problema que las técnicas ANOVA pueden producir estimaciones negativas de componentes de varianza asociadas con efectos aleatorios.

- ❖ 1947: El matemático norteamericano Churchill Eisenhart plantea la distinción entre modelos estadísticos de efectos fijos, aleatorios y mixtos y propone la nomenclatura básica aplicada actualmente, el escribe:

... analysis of variance can be, and is, used to provide solutions to problems of two fundamentally different types. These two distinct classes of problems are:

Class I: Detection and Estimation of Fixed (Constant) Relations Among Means of Sub-Sets of the Universe of Objects Concerned. This class includes all of the usual problems of estimating and testing to determine whether to infer the existence of, true differences among “treatment” means, among “variety” means, and, under certain conditions, among “place” means. Included in this class are all the problems of univariate and multivariate regression and of harmonic analysis. With respect to problems of estimation belonging to this class, analysis of variance is simply a form of the method of least squares

Class II: Detection and Estimation of Components of (Random) Variation Associated with a Composite Population. This class includes all problems of estimating, and testing to determine whether to infer the existence of components of variance ascribable to random deviation of the characteristics of individuals of a particular generic type from the mean values of these characteristics in the ‘population’ of all individuals of that generic type, etc. In this sense, *this is the true analysis of variance*, and the estimation of the respective components of the overall variance of a single observation requires further steps beyond the evaluations of the entries of the analysis-of-variance table itself. Problems of this class have received considerably less attention in the literature of the analysis of variance than have problems of Class I. . . .

... the mathematical models appropriate to problems of Class I differ from the mathematical models appropriate to problems of Class II and, consequently, so do the questions to be answered by the data.

Figura 1.2 Libro del matemático norteamericano Churchill Eisenhart

- ❖ 1950: El norteamericano Charles Roy Henderson propone la ecuación para obtener los “Best Linear Unbiased Predictions” de los modelos lineales mixtos.
- ❖ 1952: Anderson y Bancroft publican su libro “Statistical Theory in Research” donde se tratan los temas de estimación de la varianza para datos no balanceados en modelos de efectos mixtos anidados.
- ❖ 1967: Hartley y Rao demuestran que la estimación de los componentes de la varianza puede ser obtenida utilizando los métodos de máxima verosimilitud utilizando las ecuaciones que resultan de la representación matricial del modelo de efectos mixtos.
- ❖ 1971: La máxima verosimilitud restringida (REML) es presentada por Patterson y Thompson como un método para estimar los componentes de la varianza (sin asumir que los efectos fijos son conocidos) en un modelo lineal general con datos

no balanceados aunque sus métodos se desarrollan lentamente debido a que son necesarios métodos computacionales muy complejos para aquella época.

- ❖ 1972: Gabriel desarrolla el primer modelo autorregresivo de orden uno.
- ❖ 1986: Jennrich y Schluchter describen el uso de los modelos para diferentes patrones de covarianzas y describen como seleccionar el mejor modelo.
- ❖ 1987: Liang y Lee proponen la prueba de significancia para los componentes de varianza.
- ❖ 1994: Diggle, Liang, y Zeger distinguen tres tipos de componentes de varianza aleatoria: efectos aleatorios y coeficientes aleatorios, correlación serial y medidas de error aleatorio.

En cuanto al software se tienen los siguientes desarrollos claves para los modelos lineales mixtos:

- ❖ 1982: Bryk Y Raudenbush realizan el primer programa de modelos lineales anidados o modelos multinivel.
- ❖ 1988: Schluchter y Jennrich introducen las primeras rutinas para datos no balanceados.
- ❖ 1992: SAS introduce el proceso "Mixed".
- ❖ 1995: Stata libera su versión cinco e incluye su procedimiento xtreg para modelos con un solo factor aleatorio y su librería "xtgee" para análisis de modelos para datos tipo panel.
- ❖ 1998: Bates y Pinheiro desarrollan su librería lme para modelar funciones lineales de efectos mixtos para el software R.
- ❖ 2001: SPSS presenta su versión once que incluye modelos lineales mixtos.
- ❖ 2004: SPSS presenta su versión trece que incluye los modelos lineales generalizados mixtos (GLMM).
- ❖ 2005: Stata libera en su versión nueve su modulo "Longitudinal Panel Data" el cual ya incluye los (GLMM).
- ❖ 2005: Bates presenta su desarrollo lmer para R el cual también incluye los (GLMM).

CAPÍTULO II

EL INDICADOR DE OPORTUNIDAD DEL REGISTRO Y EL MODELO LINEAL MIXTO

2.1 El indicador de oportunidad del registro administrativo de nacimientos

El objetivo principal del Indicador de Oportunidad del Registro (IOR) es contar con una herramienta que permita medir numéricamente la oportunidad del registro administrativo de los nacimientos por entidad federativa con el propósito de conocer avances y rezagos de esta actividad, y de esta forma estar en posibilidades de dar a conocer estos indicadores a los gobiernos estatales para que fomenten vía recursos administrativos o de publicidad el registro oportuno de los nacimientos.

Para tener una idea de lo que el IOR pretende medir se presenta a continuación la distribución de los datos de nacimientos hipotéticos ocurridos para dos estados de la república para un año respectivo de registro y su distribución por fecha de ocurrido:



Fig. 2.1 Nacimientos hipotéticos registrados en el año 2007 y su distribución por fecha de ocurrido

La primera gráfica muestra que las personas que se registraron en las oficinas del registro civil en el año 2007 la mayoría nació en el mismo año 2007 mientras que la segunda gráfica muestra que la mayoría de las personas que se registraron en las oficinas del registro civil en el año 2007 nacieron en el año 2006, por tanto, se puede decir que la distribución de los datos de la primera gráfica es más "oportuna" en el registro que la segunda y el IOR pretende "medir" esta oportunidad.

El IOR se construye tomando en cuenta que la distribución del año de ocurrencia del evento (nacimiento) registrada en un año determinado, esta sesgada al año de registro, por lo que se calcula el tercer momento y se estandariza restando el cuadrado del valor esperado y dividiendo esta diferencia entre el promedio de las diferencias del año de referencia y los 9 años anteriores, dando como resultado un indicador que mide el alejamiento de la ocurrencia de los nacimientos con respecto a su año de registro.

La expresión matemática del indicador para el año de registro i -ésimo tomando diez años de ocurrencia del hecho, es la siguiente:

$$IOR_i = \frac{\sqrt{\sum_{j=i}^{i-9} (i-j+1)^3 \frac{N_{ij}}{N_i} - \left[\sum_{j=i}^{i-9} (i-j+1) \frac{N_{ij}}{N_i} \right]^2}}{\sum_{j=i}^{i-9} \frac{(i-j+1)}{10}}$$

Donde:

N_{ij} = Nacimientos ocurridos en el año j y registrados en el año i

N_i = Nacimientos registrados en el año i ocurridos los 10 últimos años

La dimensión del IOR es comparable para diferentes años y volumen de nacimientos y el rango de dominio se establece entre cero y 5.45, donde cero indica que todos los nacimientos ocurridos en el año se registran el mismo año, y por el contrario el valor 5.45 corresponde a que todos los nacimientos registrados en el año ocurrieron diez años antes.

La información se genero para cada una de las treinta y dos entidades federativas en formato de triangulo "Run-Off" muy utilizados por los Actuarios para estimar las Reservas de Siniestros Ocurridos pero no Reportados (IBNR) de las compañías de seguros.

La forma del triangulo "Run-Off" se muestra en la siguiente figura donde las filas representan la fecha cuando ocurrió el nacimiento y la columna el año en que se registro a la persona en la oficina del registro civil.

Año de Registro																Año de Ocurrencia
1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007		
22171	2838	133	55	53	28	37	22	28	8	5	6	9	2	2	1993	
	22741	2987	113	73	53	44	33	25	11	12	4		7		1994	
		22029	3207	124	65	67	57	47	16	8	3	10	4	3	1995	
			21254	2992	126	61	38	58	39	12	8	6	7	1	1996	
				21025	3122	110	64	64	47	18	15	8	7	7	1997	
					21723	307	101	86	52	47	28	11	9	6	1998	
						21448	2355	258	104	51	56	20	17	9	1999	
							21459	4095	177	82	55	59	34	14	2000	
								21923	3722	211	85	88	36	17	2001	
									22279	3884	222	144	56	29	2002	
										21012	3846	247	75	54	2003	
											21245	3616	210	95	2004	
												20628	3809	223	2005	
													20422	3866	2006	
														20932	2007	

Fig. 2.2 Información de nacimientos presentada en triángulos “Run-Off”

Un ejemplo gráfico de la distribución de frecuencias relativas del total de nacimientos registrado en el año 2007 para cada estado de la república se muestra a continuación, en ella se muestra diferentes distribuciones en el registro de nacimientos.

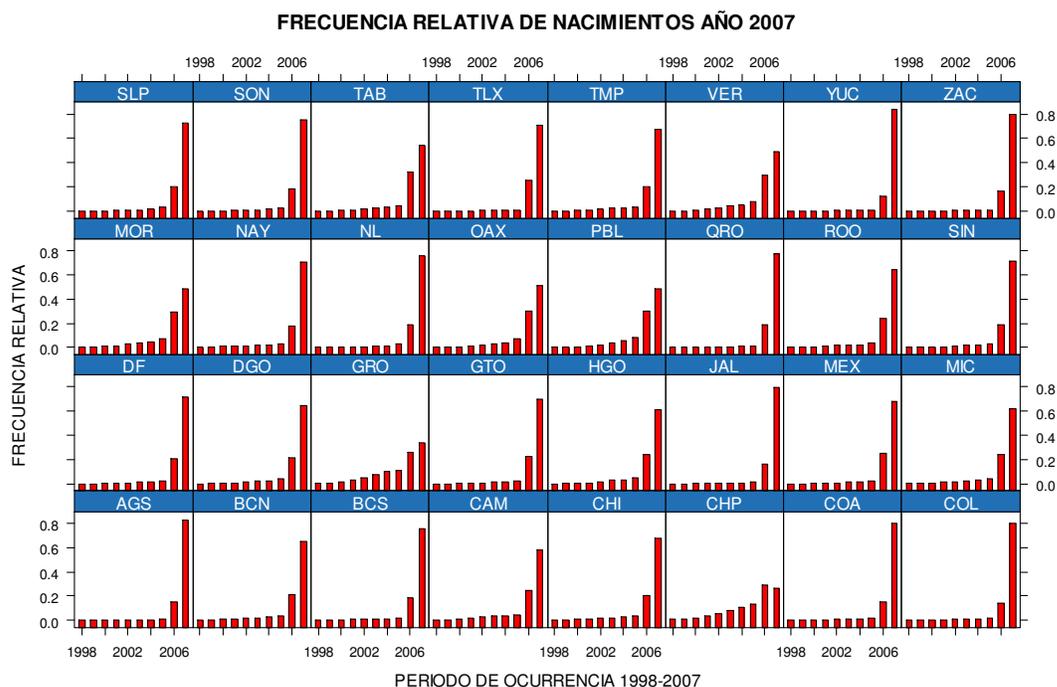


Fig. 2.3 Distribución de frecuencias relativas por estado

La información se obtuvo de las bases de datos de nacimientos que se encuentra en la página de internet del INEGI para los 32 estados que conforman la república mexicana teniendo como resultado un archivo de datos tipo “Panel” conformado por treinta y dos entidades federativas analizadas por un periodo de catorce años.

2.2 Los datos de panel

En estadística, el término de datos de panel o datos longitudinales se refiere a datos que combinan una dimensión temporal con otra transversal.

Un conjunto de datos de panel o longitudinal recoge observaciones sobre fenómenos a lo largo de determinados periodos de tiempo. La dimensión temporal enriquece la estructura de los datos y es capaz de aportar información que no aparece en un único corte. La información del IOR se ubica en la categoría de datos de panel.

Un ejemplo sencillo pero que ilustra muy claramente lo comentado anteriormente se muestra en la figura 2.4 (a) donde la capacidad de lectura es comparada con la edad en un estudio hipotético transversal de un conjunto de niños. La capacidad de lectura aparenta ser pobre entre niños de mayor edad. En la figura 2.4 (b) se asumen los mismos registros obtenidos por un estudio longitudinal en la cual cada niño fue medido en dos ocasiones. Ahora resulta claro que los niños a una mayor edad tienen un alto nivel de habilidad en la lectura.

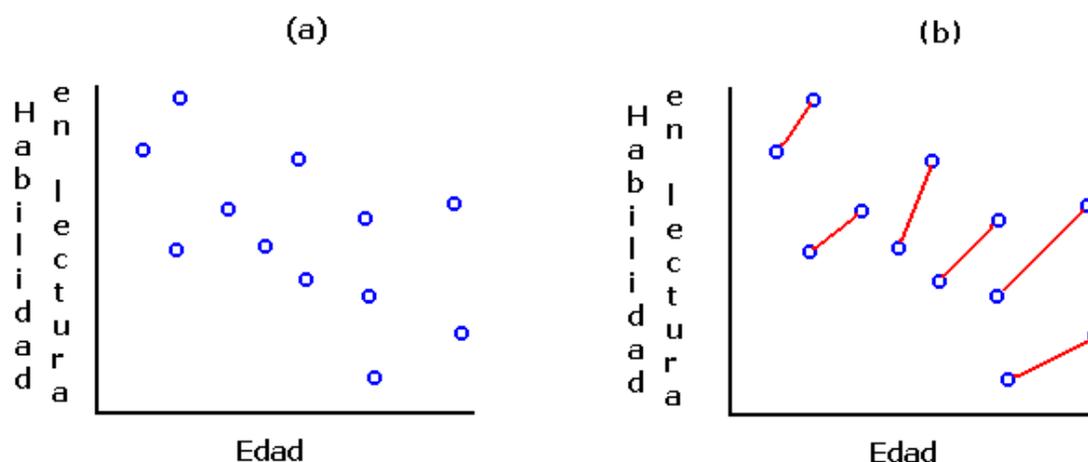


Fig. 2.4 Datos hipotéticos entre la habilidad de lectura y la edad.

Las unidades transversales (de tamaño N) pueden ser individuos, familias, regiones o estados de un país, países, empresas, etc. En general, se hablará de datos de panel cuando N es mayor al número de períodos de observación T_i (normalmente de 2 a 10 períodos en la mayoría de los casos y raras veces excediendo 20 observaciones temporales).

Si $T_i = T$, es decir, si se observa el mismo número de veces a todas las unidades transversales, se dirá que el panel de datos está completo o equilibrado (balanced); en otro caso se dirá que el panel es incompleto (unbalanced).

Entre los beneficios de usar datos de panel pueden incluirse:

- ❖ Control de la heterogeneidad individual: los datos transversales y temporales no son capaces, por si solos, de controlar la heterogeneidad inherente en el comportamiento de los individuos, empresas, regiones o países, corriéndose el riesgo de obtener estimaciones sesgadas cuando se utilizan datos de un tipo o de otro. Sin embargo, a través del uso de los datos de panel pueden controlarse estos efectos específicos, transversales o temporales, sean observables o no.
- ❖ Proporciona datos con mayor cantidad de información y con mayor grado de variabilidad.
- ❖ Son un medio adecuado para estudiar procesos dinámicos de ajuste ya que a partir de ellos se pueden analizar los cambios en el tiempo de las distribuciones transversales.
- ❖ Ayudan a identificar y medir efectos que no son detectables con datos puros de corte transversal o de series temporales.
- ❖ Permiten construir y contrastar modelos de comportamiento más complejos.

Por otro lado, entre las limitaciones cabe destacar las siguientes:

- ❖ Problemas de diseño muestral y de recolección de datos relacionados con inadecuadas tasas de cobertura, falta de respuesta, frecuencia y lapso temporal, período de referencia, etc.
- ❖ Distorsiones provocadas por errores de medida, que pueden aparecer por la falta de respuesta, respuestas incorrectas deliberadas, etc.
- ❖ Problemas de selección muestral tales como no aleatoriedad, auto-selección, no-respuesta inicial.

El conjunto de variables explicativas puede incluir:

- ❖ Variables que varían entre las unidades transversales y en el tiempo.
- ❖ Variables que no varían en el tiempo sino sólo en el corte transversal.
- ❖ Variables que sólo varían en la dimensión temporal pero no la transversal.

Un conjunto de datos de panel puede confundirse con una serie temporal. La diferencia entre unas y otras estriba en que las últimas:

- ❖ Suelen representar series más largas.
- ❖ El número de variables en cada corte es menor.
- ❖ Responden a problemas distintos.

El estudio de las series temporales depende en alto grado en las propiedades asintóticas de la dimensión temporal y, para ello, es necesario contar con un número suficiente de observaciones. No obstante, los datos de panel suelen corresponder a series más cortas. Por eso, desde el punto de vista metodológico, la manera en que la dimensión temporal forma parte del análisis de los datos de panel difiere de la habitual en el de las series temporales.

El conjunto de datos conformado por los IOR de cada entidad federativa son observaciones de un fenómeno social observado a lo largo del tiempo, catorce años para ser precisos y su aleatoriedad relevante es la que proporciona los elementos para su análisis en el tiempo.

2.3 El modelo lineal de efectos fijos

El enfoque más simple de analizar datos tipo panel es omitir las dimensiones del espacio y el tiempo de los datos agrupados y sólo calcular la regresión de mínimos cuadrados ordinarios usual. Un modelo de regresión básico para datos de corte transversal para una sola variable independiente viene dado por la expresión:

$$y_i = \alpha + \beta x_i + e_i \quad \text{para } i = 1, \dots, N \text{ unidades transversales} \quad (2.3.1)$$

Cuando cada unidad transversal se observa durante T_i períodos, el modelo básico resultante será:

$$y_{it} = \alpha + \beta x_{it} + e_{it} \quad (2.3.2)$$

para $i = 1, \dots, N$ unidades transversales y $t = 1, \dots, T_i$ periodos de observacion

Donde α y β son los parámetros a estimar.

En una situación como la que estamos analizando para los IOR, tenemos un factor agrupante, estado de la república, al que se asocian treinta y dos niveles, si incorporamos en el modelo (2.3.2) el **efecto** de cada grupo, un modelo lineal de efectos fijos será de la forma:

$$y_{it} = \alpha_i + \beta x_{it} + e_{it} \quad \text{para } i = 1, \dots, M, t = 1, \dots, T_i \quad (2.3.3)$$

El modelo de efectos fijos descrito anteriormente se caracteriza por pendientes constantes e interceptos que varían de un estado a otro.

Utilizando la técnica de "las variables dicotómicas de intersección diferencial" o también conocida como "modelo mínimo-cuadrático con variables ficticias" es posible modelar el carácter "individual" de cada estado suponiendo que las diferencias entre los estados son constantes o "fijas" y por ello se estima cada intercepto α_i .

Agrupando las observaciones temporales para cada unidad transversal o entidad federativa se tiene en su forma matricial y vectorial:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{(M-1)} \\ \alpha_M \end{bmatrix} + \begin{bmatrix} X_1 \\ \vdots \\ X_M \end{bmatrix} \beta + \begin{bmatrix} e_1 \\ \vdots \\ e_M \end{bmatrix} = D\alpha + X\beta + e = [D \ X] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + e \equiv W\gamma + e$$

El modelo anterior puede estimarse utilizando las ecuaciones normales $W^T W \lambda = W^T y$:

$$\begin{bmatrix} D^T D & D^T X & \Big| & D^T y \\ X^T D & X^T X & \Big| & X^T y \end{bmatrix} \sim \begin{bmatrix} D^T D & D^T X & \Big| & D^T y \\ o & X^T X - X^T D (D^T D)^{-1} D^T X & \Big| & X^T y - X^T D (D^T D)^{-1} D^T y \end{bmatrix}$$

$$\sim \begin{bmatrix} D^T D & D^T X & \Big| & D^T y \\ o & X^T M X & \Big| & X^T M y \end{bmatrix} \text{ donde } M = I - D (D^T D)^{-1} D^T$$

De aquí es fácil ver que:

$$\hat{\beta} = (X^T M X)^{-1} X^T M y \quad , \quad \hat{\alpha} = (D^T D)^{-1} D^T (y - X \hat{\beta})$$

M es una matriz de una estructura simple pero con el inconveniente que su manejo puede ser computacionalmente prohibitivo para un gran número de unidades.

De forma equivalente, para obtener estimaciones del vector β puede realizarse la regresión de $y_{it} - \bar{y}_i$ sobre el vector de regresores $x_{it} - \bar{x}_i$, es decir, se aplicará MCO a la ecuación:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$$

Y posteriormente, estimar los α_i a partir de las relaciones $\hat{\alpha}_i = \bar{y}_i - \bar{x}_i \hat{\beta}$.

Este modelo se conoce como "within group" porque todas las variables son centradas alrededor de la media de cada grupo.

2.4 El modelo lineal de efectos aleatorios

Aun cuando el modelo de efectos fijos toma en cuenta el efecto del factor agrupante, este no provee la suficiente representación del fenómeno cuando la información que se está analizando proviene de un muestreo, por lo que el modelo obtenido en este caso únicamente reflejará las particularidades de la muestra y no de la población de donde fue seleccionada. Otro inconveniente en el modelo de efectos fijos es el hecho que el número de parámetros en el modelo se incrementa linealmente con el número de grupos además de incrementar el número de grados de libertad.

Un modelo de efectos aleatorios solventa estos problemas al considerar el efecto de cada grupo como una variación aleatoria alrededor de la media poblacional asumiendo que las diferencias entre estados de la república no son fijas, sino aleatorias. Un desarrollo algebraico de este razonamiento de aleatoriedad se presenta a continuación:

$$y_{it} = \alpha_i + \beta x_{it} + e_{it} = \alpha_i + \beta x_{it} + e_{it} + (\bar{\alpha} - \bar{\alpha}) = \bar{\alpha} + (\alpha_i - \bar{\alpha}) + \beta x_{it} + e_{it}$$

donde $\bar{\alpha} = \frac{\sum_{i=1}^M \bar{\alpha}_i}{M}$

Si se considera a $(\alpha_i - \bar{\alpha})$ como una variable aleatoria cuya distribución se puede estimar se tendrá como resultado el modelo lineal de interceptos aleatorios:

$$y_{it} = (\alpha_0 + \zeta_{1i}) + \beta x_{it} + e_{it}$$

Donde α_0 es la media en el tiempo en la población de los grupos, ζ_i es la variable aleatoria que representa la desviación de la media en el tiempo del *i-ésimo* grupo con respecto a la media poblacional y e_{it} es la variable aleatoria que representa la desviación en el tiempo para la observación *it* del grupo *i* de la media en el tiempo del grupo *i*.

Según que se asuma que el efecto ζ_{1i} es un parámetro fijo o una variable aleatoria se tendrá el modelo de interceptos fijos o el modelo de interceptos aleatorios.

En el modelo lineal de interceptos aleatorios sólo el intercepto se supone aleatorio, mientras que los demás coeficientes se suponen fijos para todos los niveles, sin embargo, la relación entre las variables independientes y la variable respuesta también puede ser explicado por un modelo de pendientes aleatorias como muestra el siguiente desarrollo.

$$\begin{aligned} y_{it} &= (\alpha_0 + \zeta_{1i}) + \beta_i x_{it} + e_{it} = (\alpha_0 + \zeta_{1i}) + \beta_i x_{it} + e_{it} + (\bar{\beta} x_{it} - \bar{\beta} x_{it}) \\ &= (\alpha_0 + \zeta_{1i}) + \bar{\beta} x_{it} + (\beta_i - \bar{\beta}) x_{it} + e_{it} \end{aligned}$$

$$\text{donde } \bar{\beta} = \frac{\sum_{i=1}^M \beta_i}{M}$$

Si se considera a $(\beta_i - \bar{\beta})$ como una variable aleatoria cuya distribución se puede estimar se tendrá como resultado el modelo lineal de interceptos y pendientes aleatorios:

$$y_{it} = (\alpha_0 + \zeta_{1i}) + (\beta_1 + \zeta_{2i})x_{it} + e_{it}$$

Donde ζ_{1i} es el componente aleatorio de los interceptos y ζ_{2i} es el componente aleatorio de las pendientes, de aquí en adelante a ambos componentes los denominaremos b_i .

Los b_i son llamados efectos aleatorios porque están asociados con las unidades particulares del experimento (entidades federativas en nuestro caso) que fueron seleccionadas de la población en forma aleatoria o como en nuestro estudio que conforman la población de interés. Son denominados efectos porque ellos representan la desviación sobre la media general.

Las varianzas para las b_i son denotadas como ψ_b^2 y mide la variabilidad "entre grupos", la variabilidad de los e_{it} mide la variabilidad "dentro del grupo" y se denota por σ^2 .

Esto es:

$$b_i \sim N(0, \psi_b^2), \quad e_{it} \sim N(0, \sigma^2)$$

La estimación del modelo de interceptos aleatorios se describe en los siguientes párrafos:

La estructura de covarianza dentro de cada grupo o entidad federativa es:

$$\text{Cov}(y_{it}, y_{it'}) = \psi_b^2 + \sigma^2 \text{ si } t = t', \text{Cov}(y_{it}, y_{it'}) = \psi_b^2 \text{ si } t \neq t'$$

En notación matricial se tiene:

$$\text{Cov}(y_i) = \begin{bmatrix} \psi_b^2 + \sigma^2 & \psi_b^2 & \cdots & \psi_b^2 \\ \psi_b^2 & \psi_b^2 + \sigma^2 & \cdots & \psi_b^2 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_b^2 & \psi_b^2 & \cdots & \psi_b^2 + \sigma^2 \end{bmatrix}$$

La correlación entre las observaciones del mismo grupo tiene la forma:

$$\rho = \psi_b^2 / (\psi_b^2 + \sigma^2) .$$

Dependiendo de la naturaleza de los datos la estructura de covarianza dentro de cada grupo o entidad federativa puede ser diferente a la presentada anteriormente, como por ejemplo una estructura autoregresiva de primer orden:

$$Cov(y_i) = \begin{bmatrix} \sigma^2 & \rho^1 \sigma^2 & \dots & \rho^{t-1} \sigma^2 \\ \rho^1 \sigma^2 & \sigma^2 & \dots & \rho^{t-2} \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{t-1} \sigma^2 & \rho^{t-2} \sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

En el modelo de efectos fijos las b_i son tratadas como un conjunto de M coeficientes adicionales que se pueden estimar junto con las β y pueden tener cualquier valor. Por el contrario, en el modelo aleatorio se suponen que las b_i son variables aleatorias inobservables independientes de las x_{it} cuya distribución se puede estimar y sus valores de pueden predecir Es un error intentar estimar un efecto aleatorio ¿cómo saber donde está algo que se está moviendo?, lo que hay que intentar es predecir un efecto que es aleatorio.

El modelo de interceptos aleatorios se puede estimar utilizando la función de log verosimilitud:

$$l(\beta, \tau^2, V_0) = C - \frac{1}{2\tau^2} \sum_{i=1}^M (y_i - X_i \beta)^T V_0^{-1} (y_i - X_i \beta) - \frac{N}{2} \log(\tau^2) - \frac{M}{2} \log|V_0|$$

donde:

$$\tau^2 = \psi_b^2 + \sigma^2 \quad y \quad V_0 = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}$$

C es una constante de la forma:

$$N/2 + (N/2) * \log(2 * \pi) - (N/2) * \log(N)$$

Fijando V_0 , es decir "perfilando", los estimadores máximos verosímiles son:

$$\hat{\beta}(V_0) = \left(\sum_{i=1}^M X_i^T V_0^{-1} X_i \right)^{-1} \left(\sum_{i=1}^M X_i^T V_0^{-1} y_i \right)$$

$$\hat{\sigma}^2(V_0) = \frac{1}{N} SCE(V_0) = \frac{1}{N} (y_i - X_i \hat{\beta})^T V_0^{-1} (y_i - X_i \hat{\beta})$$

De aquí la log verosimilitud perfil es:

$$l(V_0) = C - \frac{N}{2} \log[SCE(V_0)] - \frac{M}{2} \log|V_0|$$

La estimación del modelo de interceptos aleatorios y pendientes aleatorios se tratará posteriormente.

¿Cuándo un factor es fijo o aleatorio? Un ejemplo ilustrativo puede ser el siguiente. Se tiene un modelo de regresión, por ejemplo para predecir el peso de las personas (y) en función de la edad (x). Si se eligen personas de 15 años, 10 años, etc. la variable edad es elegida por el investigador, no es aleatoria, fija se podría comentar en oposición a aleatoria. Diferente sería el caso si se elige personas al azar y luego le preguntamos la edad. En este caso la edad sería aleatoria.

Una representación de los modelos aleatorios se muestra en la siguiente figura:

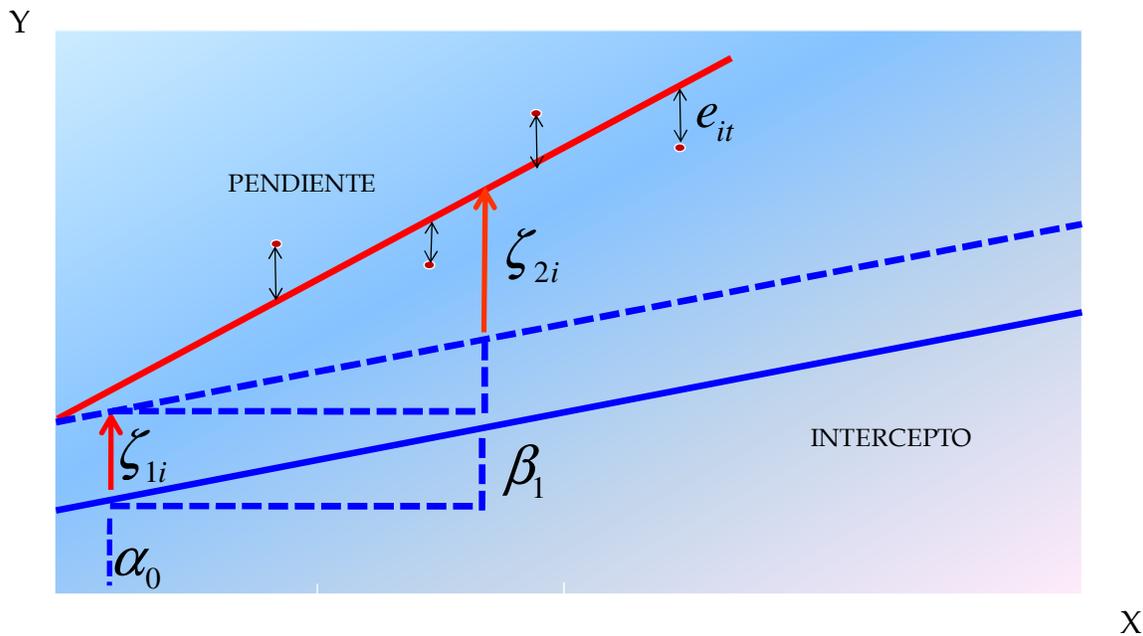


Fig. 2.5 Representación esquemática del modelo aleatorio

Hay que señalar la falta de terminología uniforme en toda la bibliografía al caracterizar los efectos mencionados. En general se dice que los efectos son fijos si son interesantes en sí mismos para modelar la media de la variable dependiente, y son aleatorios si el interés se centra en la población subyacente de donde se seleccionaron los datos o bien en analizar la varianza y covarianza de la variable dependiente.

2.5 El modelo lineal de efectos mixtos

Un modelo que incorpore tanto a los efectos fijos como los efectos aleatorios se denomina un modelo de efectos mixtos, Bates (2005), Littell (2000), Pinheiro y Bates (2000), Verbeke y Molenberghs (2000). No es éste el único nombre para estos modelos. Así tenemos: modelos de panel: Frees (2004), Wooldridge (2002); modelos longitudinales: Hand y Crowder (1999), Frees (2004), Singer y Willet (2003); modelos jerárquicos: Bryk y Raudenbush (1992), Gelman y Hill (2007); modelos multinivel: Goldstein (2003), Gelman y Hill (2007) y modelos para datos repetidos: Davis (2003).

Como la terminología no está completamente asentada y para ser concordantes con los temas vistos en la Maestría se conservara el término de modelos mixtos.

Para completar la notación del modelo incluyendo un mayor número de variables explicativas así como especificar la distribución teórica de las variables aleatorias se tiene que el modelo lineal mixto es de la forma:

$$y_{it} = x_{1it}\beta_1 + \dots + x_{pit}\beta_p + z_{1it}b_{i1} + \dots + z_{qit}b_{iq} + e_{it}$$

$$b_i \sim N(0, \psi_i^2)$$

$$\text{cov}(b_i, b_{i'}) = \psi_{ii'}$$

$$e_{it} \sim N(0, \sigma^2)$$

Donde:

y_{it} Son los valores de la variable respuesta para la t -ésima observación contenida en el i -ésimo de los M grupos.

β_1, \dots, β_p Son los coeficientes de los efectos fijos los cuales son idénticos para todos los M grupos.

x_{1it}, \dots, x_{pit} Son los regresores de los efectos fijos para la observación pit en el grupo i , el primer regresor usualmente toma el valor de uno para la constante del modelo.

b_{i1}, \dots, b_{iq} Son los coeficientes de los efectos aleatorios para el grupo i , asumiendo que están distribuidos en forma normal multivariada. Los efectos aleatorios, por tanto, varían por grupo. Los b_{ik} están considerados como variables aleatorias en el modelo como los errores e_{it} , pero no hay inconveniente alguno en estimar la distribución de dichas variables.

z_{1it}, \dots, z_{qit} Son los regresores de los efectos aleatorios.

ψ_i^2 Es la varianza y $\psi_{ii'}$ es la covarianza entre los efectos aleatorios, asumiendo ser constantes a través de los grupos.

e_{it} Es el error para la observación en el momento t del grupo i . Los errores para el grupo i se asumen distribuidos como una normal multivariada.

En forma alternativa pero equivalente, la forma matricial del modelo será de la forma:

$$\begin{aligned}
 y_i &= X_i\beta + e_i && \text{Para el modelo de efectos fijos} \\
 y_i &= Z_i b_i + e_i && \text{Para el modelo de efectos aleatorios} \\
 y_i &= X_i\beta + Z_i b_i + e_i && \text{Para el modelo de efectos mixtos} \\
 b_i &\sim N_q(0, \psi) \\
 e_i &\sim N_{n_i}(0, \sigma^2 I_i)
 \end{aligned}
 \tag{2.5.1}$$

Donde:

y_i Es el vector de respuestas para las observaciones del i -ésimo grupo de tamaño $n_i \times 1$.

X_i Es la matriz modelo o de diseño de los efectos fijos para las observaciones del i -ésimo grupo de dimensión $n_i \times p$.

β Es el vector de coeficientes de los efectos fijos de tamaño $p \times 1$.

Z_i Es la matriz modelo o de diseño de los efectos aleatorios para las observaciones del i -ésimo grupo de dimensión $n_i \times q$.

b_i Es el vector de coeficientes de los efectos aleatorios de tamaño $q \times 1$.

e_i Es el vector de errores para las observaciones del i -ésimo grupo.

ψ Es la matriz de covarianzas de los efectos aleatorios de dimensión $q \times q$.

$\sigma^2 I_i$ Es la matriz de covarianzas de los errores en el i -ésimo grupo de dimensión $n_i \times n_i$.

Condicionada en los efectos aleatorios b_i a la ecuación 2.5.1 se le conoce como el modelo condicional independiente con vector de medias $X_i\beta + Z_i b_i$ y matriz de covarianza $\sigma^2 I_i$, este modelo indica que las n_i respuestas del grupo i son independientes, condicionales en b_i y β .

Por otra parte este modelo de independencia condicional implica probablemente una simple estructura de covarianzas poco realista para el vector de respuestas y_{it} especialmente para modelos con un número reducido de efectos aleatorios, por lo anterior, lo asumido para la matriz de covarianzas de los errores puede relajarse en una forma que se permita modelar varianzas no constantes o bien estructuras especiales de correlación dentro de los grupos más apropiadas para los mismos.

Una gran variedad de modelos han sido propuestos para lo comentado anteriormente, estos modelos asumen que el vector e_i tiene varianza constante y puede ser descompuesto como $e_i = e_{(1)i} + e_{(2)i}$, donde $e_{(2)i}$ es un componente de correlación serial, esto es, parte de los perfiles observados para cada grupo es una respuesta a un proceso estocástico de variación en el tiempo. Este tipo de variación aleatoria resulta en una medida de correlación serial, la cual usualmente es una función decreciente en los intervalos de tiempo de estas medidas. Por lo anterior, el modelo lineal de efectos mixtos para este tipo de casos puede escribirse ahora como:

$$\begin{aligned}
 y_i &= X_i\beta + Z_ib_i + e_i = X_i\beta + Z_ib_i + e_{(1)i} + e_{(2)i} \\
 b_i &\sim N_q(0, \psi) \\
 e_i &\sim N_{n_i}(0, \sigma^2 \Lambda_{n_i}) \\
 e_{(1)i} &\sim N_{n_i}(0, \gamma_{e_{(1)}}^2 I_{n_i}) \\
 e_{(2)i} &\sim N_{n_i}(0, \gamma_{e_{(2)}}^2 \Omega_{n_i})
 \end{aligned} \tag{2.5.2}$$

Y el modelo es completado considerando una estructura específica para las $n_i \times n_i$ correlaciones que forman la matriz Ω_i . Tales estructuras son comúnmente estimadas utilizando el "Análisis de Series de Tiempo". Es habitual asumir que el efecto serial $e_{(2)i}$ es un fenómeno poblacional, independiente de los grupos. La matriz de correlación serial Ω_i entonces únicamente depende en i a través del número n_i de observaciones y a través de los puntos en el tiempo t_{ij} . Por otra parte, se asume que el (j, k) elemento h_{ijk} de Ω_i es modelado como $h_{ijk} = g|t_{ij} - t_{ik}|$ para una función decreciente $g(\cdot)$ con $g(0) = 1$. Esto significa que la correlación entre $e_{(2)ij}$ y $e_{(2)ik}$ únicamente depende de los intervalos de tiempo entre y_{ij} y y_{ik} y es decreciente cuando el tamaño de estos intervalos se incrementa.

Dos funciones usadas frecuentemente son las funciones de correlación serial Exponencial y Gaussiana definidas como:

$$g(u) = \exp(-\phi u) \quad \text{y} \quad g(u) = \exp(-\phi u^2)$$

Las matrices Λ_i son definidas positivas parametrizadas por un conjunto de parámetros λ , generalmente pequeños y fijos. Así, las Λ_i suponen independencia para diferentes grupos, e independencia de los efectos aleatorios b_i .

Como Λ_i es positiva definida, esta admite una matriz raíz invertible $\Lambda_i^{1/2}$, con inversa $\Lambda_i^{-1/2}$ tal que:

$$\Lambda_i = \left(\Lambda_i^{1/2}\right)^T \Lambda_i^{1/2} \quad y \quad \Lambda_i^{-1} = \Lambda_i^{-1/2} \left(\Lambda_i^{-1/2}\right)^T$$

Siendo:

$$y_i^* = \left(\Lambda_i^{-1/2}\right)^T y, \quad e_i^* = \left(\Lambda_i^{-1/2}\right)^T e_i, \quad X_i^* = \left(\Lambda_i^{-1/2}\right)^T X_i, \quad Z_i^* = \left(\Lambda_i^{-1/2}\right)^T Z_i$$

De esta forma (2.5.1) se puede reescribir como

$$y_i^* = X_i^* \beta + Z_i^* b_i + e_i^* \quad \text{con} \quad b_i \sim N_q(0, \Psi) \quad y \quad e_i^* \sim N_{ni}(0, \sigma^2 I)$$

Esto es y_i^* es descrita como un modelo básico de efectos mixtos.

Debido a que el diferencial de la transformación lineal $y_i^* = \left(\Lambda_i^{-1/2}\right)^T$ es simplemente $dy_i^* = \left|\Lambda_i^{-1/2}\right| dy$, la función de verosimilitud L correspondiente a un modelo de efectos lineales mixtos (2.5.1) es expresado como:

$$\begin{aligned} L(\beta, \theta, \sigma^2, \lambda | y) &= \prod_{i=1}^M p(y_i | \beta, \theta, \sigma^2, \lambda) = \prod_{i=1}^M p(y_i^* | \beta, \theta, \sigma^2, \lambda) \left|\Lambda_i^{-1/2}\right| \\ &= L(\beta, \theta, \sigma^2, \lambda | y^*) \prod_{i=1}^M \left|\Lambda_i^{-1/2}\right| \end{aligned}$$

Donde p denota la función de densidad de probabilidad.

La función de verosimilitud $L(\beta, \theta, \sigma^2, \lambda | y^*)$ corresponde a un modelo básico lineal de efectos mixtos y por tanto todos los resultados que se presentarán en la sección 2.7 también pueden ser aplicados a este modelo.

Un ejemplo del modelo lineal mixto puede ser el siguiente: Se selecciona una muestra al azar de profesores en un estudio experimental sobre la influencia de distintos estilos de enseñanza en el rendimiento de los alumnos, la variable profesor constituiría un factor aleatorio. Si comparamos en cambio la eficacia de tres métodos didácticos (A, B y C), éstos constituirían un factor de efectos fijos ya que se está específicamente interesado en estos métodos y no en otros. Se considera un modelo de efectos mixtos a aquél en el que uno o más factores son fijos y al menos uno es aleatorio. Una combinación de los dos estudios sería un ejemplo de diseño mixto que permitiría evaluar de forma conjunta tanto los métodos de enseñanza (A, B, C), como la variabilidad en el efecto de los profesores y la interacción entre ambos sobre el rendimiento de los alumnos.

Una representación gráfica de los tres componentes estocásticos del modelo lineal mixto para los datos del IOR se muestra en la siguiente figura, la línea sólida representa la evolución promedio de la población (Nivel nacional), las líneas cortas representan la evolución de dos grupos (estados de la república). Los componentes residuales de la correlación serial y los errores clásicos son indicados con las líneas largas y los puntos respectivamente.

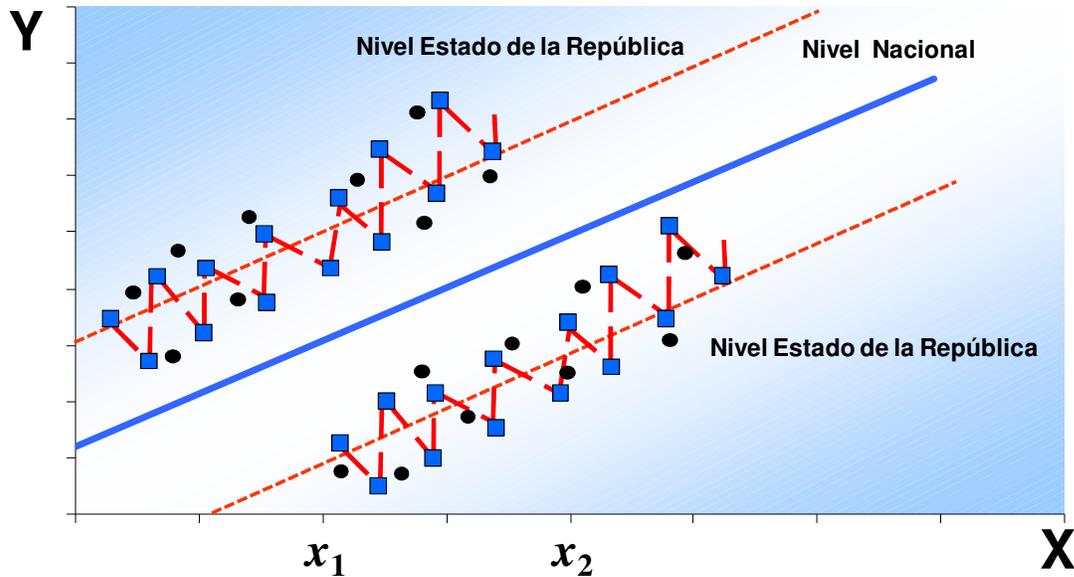


Fig. 2.6 Representación esquemática del modelo lineal mixto

2.6 Nivel individual y multinivel del modelo lineal mixto

El propósito de esta sección es presentar en términos muy generales la teoría de los modelos lineales de efectos mixtos en su formulación individual en la primera parte y posteriormente describir su extensión multinivel.

Para un nivel individual de agrupamiento, los modelos lineales de efectos mixtos propuestos por Laird y Ware en el año de 1982, expresan el vector de respuestas y_i de dimensión n_i para el i -ésimo grupo como:

$$y_i = X_i\beta + Z_i b_i + e_i; \quad i = 1, \dots, N; \quad b_i \sim N(0, \psi); \quad e_i \sim N(0, \sigma^2 I)$$

Donde la distribución de los efectos aleatorios b_i se define como una distribución normal con media cero y matriz de varianza y covarianza ψ , donde ψ debe ser una matriz simétrica y ser positiva semidefinida (todos los valores propios son positivos).

Cuando se obtiene la matriz Ψ es más conveniente computacionalmente expresar esta matriz utilizando un factor de precisión relativo Δ el cual satisface:

$$\frac{\Psi^{-1}}{\frac{1}{\sigma^2}} = \Delta^T \Delta$$

Si Ψ es definida-positiva, entonces Δ existe aunque no necesariamente es única. Posiblemente se elige el factor Cholesky de $\sigma^2 \Psi^{-1}$ como Δ .

La matriz Δ es llamada factor de precisión relativa porque es factor de precisión de la matriz Ψ^{-1} de los efectos aleatorios, expresados en términos de la precisión relativa del orden $\frac{1}{\sigma^2}$ de los errores e_i .

La formulación del modelo lineal de efectos mixtos utilizando los datos balanceados del IOR tendría las siguientes especificaciones:

$$N = 448, n_i = 14, i = 1, \dots, 32, p = 2, q = 2$$

$$Y_i = \begin{bmatrix} Y_{li} \\ \vdots \\ Y_{ni} \end{bmatrix} \quad \text{por otra parte } X_i = Z_i = \begin{bmatrix} 1 & 1994 \\ \vdots & \\ 1 & 2007 \end{bmatrix} \quad i = 1, \dots, 32$$

Cualquier raíz cuadrada de una matriz $\sigma^2 \Psi^{-1}$ de tamaño $p \times q$ puede ser usada como factor de precisión relativo.

La formulación de Laird y Ware para un nivel individual se pueden extender a un número mayor de niveles. Los modelos multinivel para modelos lineales de efectos mixtos jerarquizan los niveles. Para el caso de dos niveles jerarquizados de efectos aleatorios el vector de respuestas en su nivel de agrupamiento se escribe como $y_{ij}, i = 1, \dots, M, j = 1, \dots, M_i$ donde M es el número de grupos del primer nivel y M_i es el número de grupos del segundo nivel pertenecientes al grupo i del primer nivel. El tamaño de y_{ij} es n_{ij} .

Las matrices de los efectos fijos son X_{ij} de tamaño $n_{ij} \times p$. Usando los efectos aleatorios de primer nivel b_i de tamaño q_1 y los efectos aleatorios de segundo nivel b_{ij} de tamaño q_2 y sus correspondientes matrices modelo $Z_{i,j}$ de tamaño $n_i \times q_1$ y Z_{ij} de tamaño $n_i \times q_2$, se puede escribir el modelo como:

$$y_{ij} = X_{ij}\beta + Z_{i,j}b_i + Z_{ij}b_{ij} + e_{ij} \quad i = 1, \dots, M, j = 1, \dots, M_i$$

$$b_i \sim N(0, \Psi_1), \quad b_{ij} \sim N(0, \Psi_2), \quad e_{ij} \sim N(0, \sigma^2 I)$$

Los efectos aleatorios del primer nivel b_i se suponen independientes para diferente grupo i , los efectos aleatorios del segundo nivel b_{ij} se suponen independientes para diferentes i ó j y son independientes de los efectos aleatorios del primer nivel y los errores dentro del grupo e_{ij} se suponen independientes para diferentes i ó j , e independientes de los efectos aleatorios

Para extensiones con un número mayor de niveles se aplica el mismo patrón.

En un modelo con Q niveles de efectos aleatorios existen Q matrices de varianzas-covarianzas simétricas positivas definidas $\psi_k, k = 1, \dots, Q$. Para propósitos computacionales se expresa esto en términos de factores de precisión relativa $\Delta_k, k = 1, \dots, Q$ que satisface la igualdad:

$$\Delta_k^T \Delta_k = \sigma^2 \psi_k^{-1}, \quad k = 1, \dots, Q$$

Para optimizar la log-verosimilitud se expresa la matriz de varianza-covarianzas Ψ_k / σ^2 o su equivalente factor de precisión relativa Δ_k como un vector de parámetros no restringido $\theta_k, k = 1, \dots, Q$.

Para el caso general donde ψ_k / σ^2 es una matriz simétrica positiva definida de tamaño $q \times q$ se parametriza esta como *matriz logaritmo* y para definir esta parametrización, se tiene que cualquier matriz simétrica positiva definida A puede expresarse como la *matriz exponencial* de otra matriz simétrica B , esto es:

$$A = e^B = I + B + \frac{B^2}{2!} + \frac{B^3}{3!} + \dots$$

Y se dice que si A es la *matriz exponencial* de B , entonces B es la *matriz logaritmo* de A .

Suponiendo que A es una matriz simétrica y positiva definida de tamaño $q \times q$ una manera de evaluar la *matriz logaritmo* B es el calcular los valores y vectores propios de la descomposición:

$$A = U \Lambda U^T$$

Donde Λ es una matriz diagonal de tamaño $q \times q$ mientras que U es una matriz ortogonal de tamaño $q \times q$. Si A es positiva definida entonces todos los elementos de la diagonal de Λ deberán de ser positivos. La *matriz logaritmo* de Λ es la matriz diagonal cuyos elementos en la diagonal son los logaritmos de los elementos correspondientes de Λ . Denotando esto como $\log \Lambda$. Finalmente se tiene:

$$B = \log A = U (\log \Lambda) U^T$$

Y se definirá θ_k como los elementos que componen el triangulo superior de la matriz logaritmo de ψ_k / σ^2 . De esta forma se forma el vector de parámetros no restringido para ψ_k / σ^2 .

Según Piñeiro y Bates existen otras formas adicionales a la matriz logaritmo para parametrizar el factor Δ (Cholesky, log-Cholesky y Esférica).

2.7 Estimación máximo verosímil para el modelo lineal mixto

El propósito de esta sección es comentar la teoría sobre la estimación de los modelos lineales de efectos mixtos en su formulación individual y posteriormente describir su extensión multinivel con base en el libro de Piñeiro y Bates.

Considerando el modelo lineal mixto de nivel individual (2.5.1) que tiene un solo nivel de efectos aleatorios. Los parámetros del modelo son β y σ^2 y los parámetros determinados por Δ el factor de precisión relativa. Como se vio en la sección anterior se usará θ para representar el vector de parámetros no restringido que determinan a Δ . La función de verosimilitud para el modelo (2.5.1) es la densidad de probabilidad para los datos dados los parámetros, considerada una función de parámetros para datos fijos en lugar de una función de los datos con los parámetros fijos, es decir:

$$L(\beta, \theta, \sigma^2 | y) = p(y | \beta, \theta, \sigma^2)$$

Donde L es la función de verosimilitud, p es la función de densidad de probabilidad y y es la entrada del vector respuesta N -dimensional, $N = \sum_{i=1}^M n_i$. Al ser los efectos aleatorios b_i , $i = 1, \dots, M$ parte del modelo, se debe integrar la densidad condicional de los datos dados los efectos aleatorios con respecto a la densidad marginal de los efectos aleatorios para poder obtener la densidad marginal de los datos. Se puede usar la independencia de b_i y e_i , para expresar esto como:

$$L(\beta, \theta, \sigma^2 | y) = \prod_{i=1}^M p(y_i | \beta, \theta, \sigma^2) = \prod_{i=1}^M \int p(y_i | b_i, \beta, \sigma^2) p(b_i | \theta, \sigma^2) db_i \quad (2.7.1)$$

Donde la densidad condicional de y_i es una normal multivariada.

$$p(y_i | b_i, \beta, \sigma^2) = \frac{\exp\left(-\|y_i - X_i\beta - Z_i b_i\|^2 / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \quad (2.7.2)$$

Y la densidad marginal de b_i es también normal multivariada.

$$p(b_i | \theta, \sigma^2) = \frac{\exp(-b_i^T \Psi^{-1} b_i)}{(2\pi)^{\frac{q}{2}} \sqrt{|\Psi|}} = \frac{\exp(-\|\Delta b_i\|^2 / 2\sigma^2)}{(2\pi\sigma^2)^{\frac{q}{2}} \text{abs}|\Delta|^{-1}} \quad (2.7.3)$$

Donde $|\Delta|$ denota el determinante de la matriz Δ .

Substituyendo (2.7.2) y (2.7.3) en (2.7.1) la función de verosimilitud es:

$$\begin{aligned} L(\beta, \theta, \sigma^2 | y) &= \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \int \frac{\exp\left(-\left(\|y_i - X_i\beta - Z_i b_i\|^2 + \|\Delta b_i\|^2\right) / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{q}{2}}} db_i \\ &= \prod_{i=1}^M \frac{\text{abs}|\Delta|}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \int \frac{\exp\left(-\left(\|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i\|^2\right) / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{q}{2}}} db_i \end{aligned} \quad (2.7.4)$$

Donde:

$$\tilde{y}_i = \begin{bmatrix} y_i \\ 0 \end{bmatrix}, \quad \tilde{X}_i = \begin{bmatrix} X_i \\ 0 \end{bmatrix}, \quad \tilde{Z}_i = \begin{bmatrix} Z_i \\ \Delta \end{bmatrix} \quad (2.7.5)$$

Son vectores aumentados y matrices modelo.

Esta aproximación de cambio en la contribución de la distribución marginal de los efectos aleatorios en la filas adicionales de la variable respuesta y las matrices de diseño se conocen como "seudo datos" de aproximación porque crea el efecto de la distribución marginal por medio de "seudo observaciones" añadidas.

El exponente de la integral de (2.7.4) es de la forma de una suma de residuales al cuadrado. Se puede determinar el modelo condicional de efectos aleatorios, dados los datos, escribiendo \hat{b}_i , por la minimización de la suma de cuadrados residual. Este es un problema de mínimos cuadrados estándar cuya solución es:

$$\hat{b}_i = (\tilde{Z}_i^T \tilde{Z}_i)^{-1} \tilde{Z}_i^T (\tilde{y}_i - \tilde{X}_i \beta)$$

La norma al cuadrado puede expresarse como:

$$\begin{aligned}
 \|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i\|^2 &= \|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i + \tilde{Z}_i \hat{b}_i - \tilde{Z}_i \hat{b}_i\|^2 = \|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i \hat{b}_i\|^2 + \|\tilde{Z}_i (b_i - \hat{b}_i)\|^2 \\
 &= \|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i \hat{b}_i\|^2 + (b_i - \hat{b}_i)^T \tilde{Z}_i^T \tilde{Z}_i (b_i - \hat{b}_i)
 \end{aligned} \tag{2.7.6}$$

El primer término de (2.7.6) no depende de b_i , por lo que su exponencial puede salir de la integral, la integración exponencial del segundo término en (2.7.4) es equivalente a integrar una función de densidad normal multivariada. Por otro lado:

$$\begin{aligned}
 &\frac{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}}{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}} \int \frac{\exp\left(-\frac{(b_i - \hat{b}_i)^T \tilde{Z}_i^T \tilde{Z}_i (b_i - \hat{b}_i) / 2\sigma^2}{(2\pi\sigma^2)^{\frac{q}{2}}}\right) db_i}{(2\pi\sigma^2)^{\frac{q}{2}}} \\
 &= \frac{1}{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}} \int \frac{\exp\left(-\frac{(b_i - \hat{b}_i)^T \tilde{Z}_i^T \tilde{Z}_i (b_i - \hat{b}_i) / 2\sigma^2}{(2\pi\sigma^2)^{\frac{q}{2}} / \sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}}\right) db_i}{(2\pi\sigma^2)^{\frac{q}{2}} / \sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}} \\
 &= \frac{1}{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}} = \frac{1}{\sqrt{|Z_i^T Z_i + \Delta^T \Delta|}}
 \end{aligned} \tag{2.7.7}$$

Combinando (2.7.6) y (2.7.7) se puede expresar la integral de (2.7.4) como:

$$\int \frac{\exp\left(-\frac{\|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i b_i\|^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{\frac{q}{2}}} db_i = \frac{\exp\left(-\frac{\|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i \hat{b}_i\|^2}{2\sigma^2}\right)}{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}}$$

Para obtener:

$$L(\beta, \theta, \sigma^2 | y) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\sum_{i=1}^M \frac{\|\tilde{y}_i - \tilde{X}_i\beta - \tilde{Z}_i \hat{b}_i\|^2}{2\sigma^2}\right) \prod_{i=1}^M \frac{abs|\Delta|}{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}} \tag{2.7.8}$$

Esta última expresión puede utilizarse directamente por medio de una rutina en R para obtener los estimadores máximo verosímiles de β , θ y σ^2 . Sin embargo la rutina puede realizarse de una manera más fácil si se "perfila" la verosimilitud como función de θ únicamente. Esto es, se calculan los estimadores condicionales $\hat{\beta}$ y $\hat{\sigma}^2$ como los valores que maximizan a $L(\beta, \theta, \sigma^2)$ para un θ dado. Se observa que los términos de 2.7.8 que involucran a β y σ^2 en la verosimilitud son idénticos a los modelos de regresión lineal y por tanto $\hat{\beta}$ y $\hat{\sigma}^2$ se pueden estimar por la teoría de regresión estándar.

Es necesario tener precaución debido a que el estimador de mínimos cuadrados para β dependerá condicionalmente de b_i , y éste a su vez, dependerá de β , de este modo se deberá determinar el valor por mínimos cuadrados conjunto como solución a mínimos cuadrados de:

$$(\hat{b}_1, \dots, \hat{b}_M, \hat{\beta}^T)^T = \underset{(b_1, \dots, b_M, \beta)}{\operatorname{arg\,min}} \|y_e - X_e(b_1, \dots, b_M, \beta)^T\|^2, \text{ donde:}$$

$$X_e = \begin{bmatrix} Z_1 & 0 & \dots & 0 & X_1 \\ \Delta & 0 & \dots & 0 & 0 \\ 0 & Z_2 & \dots & 0 & X_2 \\ 0 & \Delta & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \Delta & Z_M & X_M \\ 0 & 0 & \dots & \Delta & 0 \end{bmatrix}, y_e = \begin{bmatrix} y_1 \\ 0 \\ y_2 \\ 0 \\ \vdots \\ y_M \\ 0 \end{bmatrix} \quad (2.7.9)$$

Conceptualmente se puede escribir:

$$(\hat{b}_1^T, \dots, \hat{b}_M^T, \hat{\beta}^T)^T = (X_e^T X_e)^{-1} X_e^T y_e$$

La teoría de la regresión lineal también ofrece el estimador máximo verosímil condicional para σ^2 :

$$\hat{\sigma}^2(\theta) = \frac{\|y_e - X_e(\hat{b}_1^T, \dots, \hat{b}_M^T, \hat{\beta}^T)^T\|^2}{N} \quad (2.7.10)$$

Se puede observar que este estimador es la suma de cuadrados de los residuales dividido por N y no por $N-p$

Sustituyendo este valor en la verosimilitud de (2.7.8) se tiene la verosimilitud perfil:

$$L(\theta) = L(\hat{\beta}(\theta), \theta, \hat{\sigma}^2(\theta)) = \frac{\exp(-N/2)}{(2\pi\hat{\sigma}^2(\theta))^{N/2}} \prod_{i=1}^M \frac{\operatorname{abs}|\Delta|}{\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|}} \quad (2.7.11)$$

No es necesario calcular los valores de $\hat{b}_1, \dots, \hat{b}_M$ o de $\hat{\beta}(\theta)$ para evaluar la verosimilitud perfil. Únicamente se necesita conocer la norma de los residuales del problema de mínimos cuadrados aumentado.

El modelo 2.5.1 puede expresarse como:

$$y_i = X_i\beta + Z_i b_i + e_i = X_i\beta + e_i^* \text{ donde } e_i^* = Z_i b_i + e_i$$

En virtud de que e_i^* es la suma de dos vectores aleatorios con distribución normal multivariada estos se distribuyen como una normal multivariada con media 0 y matriz de varianzas covarianzas $V_0 = \sigma^2 I + Z_i \psi Z_i^T$. De esta forma las observaciones y_i son vectores aleatorios independientes con media $X\beta$ y matriz de varianzas y covarianzas V_0 , Esto es:

$$p(y_i | \beta, \theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \exp\left(-\frac{(y_i - X\beta)^T V_0^{-1} (y_i - X\beta)}{2\sigma^2}\right) |V_0|^{-1/2}$$

Para un valor dado de θ , los valores de β y σ^2 que maximizan la verosimilitud pueden ser escritos como:

$$\hat{\beta}(\theta) = \left(\sum_{i=1}^M X_i^T V_0^{-1} X_i \right)^{-1} \left(\sum_{i=1}^M X_i^T V_0^{-1} y_i \right)$$

$$\hat{\sigma}^2(\theta) = \frac{1}{N} SCE(V_0) = \frac{1}{N} (y_i - X_i \hat{\beta}(\theta))^T V_0^{-1} (y_i - X_i \hat{\beta}(\theta))$$

Según Pinheiro y Bates la expresión anterior es más difícil de manejar computacionalmente, por esta razón se inclinan por la representación de los pseudo datos para el cálculo, sobre todo cuando dicha representación se combina con la descomposición ortogonal-triangular.

Para encontrar una solución al problema de mínimos cuadrados se usará la descomposición ortogonal-triangular de matrices rectangulares, método muy popular y preferido para resolver mínimos cuadrados. También es conocida como descomposición QR:

$$X = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = Q_t R$$

Donde X es una matriz de $n \times p$ ($n \geq p$) de rango p , Q es una matriz ortogonal de $n \times n$, R es una matriz triangular superior de tamaño $p \times p$ y Q_t (Q -truncada) consiste de las primeras p columnas de Q . Como Q es ortogonal, entonces $Q^T Q = Q Q^T = I$ y esto implica que $Q_t^T Q_t = I$.

Una propiedad importante de las matrices ortogonales es que preservan normas de vectores en la multiplicación sea Q ó Q^T , esto es, la transformación representada por Q es una generalización de una rotación o una reflexión en el plano, en particular:

$$\|Q^T y\|^2 = (Q^T y)^T Q^T y = y^T Q Q^T y = y^T y = \|y\|^2$$

Si se aplica esto al vector residual en el problema de mínimos cuadrados, se tiene:

$$\begin{aligned} \|y - X\beta\|^2 &= \|Q^T (y - X\beta)\|^2 = \|Q^T y - Q^T X\beta\|^2 = \left\| c - Q^T Q \begin{bmatrix} R \\ 0 \end{bmatrix} \beta \right\|^2 = \left\| c - \begin{bmatrix} R \\ 0 \end{bmatrix} \beta \right\|^2 \\ &= \|c_1 - R\beta\|^2 + \|c_2\|^2 \end{aligned}$$

Donde $c = (c_1^T c_2^T)^T = Q^T y$ es el vector de residuales rotado. Las componentes c_1 y c_2 son de tamaño p y $n - p$ respectivamente. Como X es de rango p , la matriz R de tamaño $p \times p$ es no singular y triangular superior.

La solución de mínimos cuadrados $\hat{\beta}$ es fácil de evaluar como la solución

$$R\hat{\beta} = c_1$$

Y la suma de cuadrados residual es $\|c_2\|^2$. La suma de cuadrados de los residuales puede ser calculada sin tener la estimación de $\hat{\beta}$.

La evaluación de la verosimilitud por medio de la descomposición se realiza descomponiendo la matriz ortogonal triangular aumentada \tilde{Z}_i de (2.7.5) como:

$$\tilde{Z}_i = Q_{(i)} \begin{bmatrix} R_{11(i)} \\ 0 \end{bmatrix}$$

Donde $Q_{(i)}$ es de $(n_{(i)} + q)(n_{(i)} + q)$ y $R_{11(i)}$ es de $q \times q$. Por lo que:

$$\|\tilde{y}_{(i)} - \tilde{X}_i \beta - \tilde{Z}_i b_i\|^2 = \|Q_{(i)}^T (\tilde{y}_{(i)} - \tilde{X}_i \beta - \tilde{Z}_i b_i)\|^2 = \|c_{1(i)} - R_{10(i)} \beta - R_{11(i)} b_i\|^2 + \|c_{0(i)} - R_{00(i)} \beta\|^2$$

Donde la matriz $R_{10(i)}$ es de $q \times p$, $R_{00(i)}$ es de $n_i \times p$, el vector $c_{1(i)}$ es de dimensión q y $c_{0(i)}$ es un vector de dimensión n_i , son definidos como:

$$\begin{bmatrix} R_{10(i)} \\ R_{00(i)} \end{bmatrix} = Q_{(i)}^T \tilde{X}_i \quad \begin{bmatrix} c_{1(i)} \\ c_{0(i)} \end{bmatrix} = Q_{(i)}^T \tilde{y}_i$$

Otro razonamiento sobre estas matrices es observándolas como una descomposición ortogonal triangular de la matriz aumentada:

$$\begin{bmatrix} Z_i & X_i & y_i \\ \Delta & 0 & 0 \end{bmatrix} = Q^{(i)} \begin{bmatrix} R_{11(i)} & R_{10(i)} & c_{1(i)} \\ 0 & R_{00(i)} & c_{0(i)} \end{bmatrix}$$

Donde la reducción a una forma triangular se detiene después de las primeras q columnas (El peculiar sistema de numeración de las submatrices y subvectores está diseñado para permitir una fácil extensión para un nivel más alto de efectos aleatorios.).

Retornando a la integral de (2.7.4) se puede remover el factor constante y reducir la expresión a la forma:

$$\int \frac{\exp\left(-\left(\|y_i - X_i\beta - Z_i b_i\|^2 + \|\Delta b_i\|^2\right) / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{1}{2}}} db_i =$$

$$\exp\left[\frac{\|c_{0(i)} - R_{00(i)}\beta\|^2}{-2\sigma^2}\right] \int \frac{\exp\left(\|c_{1(i)} - R_{10(i)}\beta - R_{11(i)}b_i\|^2 / -2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{q}{2}}} db_i \quad (2.7.12)$$

Porque $R_{11(i)}$ es no singular, se puede realizar un cambio de variable para $\phi_i = (c_{1(i)} - R_{10(i)}\beta - R_{11(i)}b_i) / \sigma$ con diferencial $d\phi_i = -\sigma^{-1}R_{11(i)}db_i$ y Jacobiano de la transformación $|J(b_i \rightarrow \phi_i)| = \sigma^{-q/2} abs |R_{11(i)}|$, y escribir la integral como:

$$\int \frac{\exp\left(-\|c_{1(i)} - R_{10(i)}\beta - R_{11(i)}b_i\|^2 / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{q}{2}}} db_i = \frac{1}{abs |R_{11(i)}|} \int \frac{\exp\left(-\|\phi_i\|^2 / 2\right)}{(2\pi)^{\frac{q}{2}}} d\phi_i$$

$$= \frac{1}{abs |R_{11(i)}|} \quad (2.7.13)$$

Y este es el mismo resultado de (2.7.8) debido a que:

$$\sqrt{|\tilde{Z}_i^T \tilde{Z}_i|} = \sqrt{\left| \begin{pmatrix} R_{11(i)}^T & 0 \end{pmatrix} Q_{(i)}^T Q_{(i)} \begin{bmatrix} R_{11(i)} \\ 0 \end{bmatrix} \right|}$$

$$= \sqrt{\left| \begin{pmatrix} R_{11(i)}^T & 0 \end{pmatrix} \begin{bmatrix} R_{11(i)} \\ 0 \end{bmatrix} \right|} = \sqrt{\left| \begin{pmatrix} R_{11(i)}^T & 0 \end{pmatrix} \begin{bmatrix} R_{11(i)} \\ 0 \end{bmatrix} \right|} = \sqrt{\left| \begin{pmatrix} R_{11(i)}^T & 0 \end{pmatrix} \begin{bmatrix} R_{11(i)} \\ 0 \end{bmatrix} \right|} = abs |R_{11(i)}|$$

Debido a que $R_{11(i)}$ es triangular, su determinante es simplemente el producto de los elementos de su diagonal.

Sustituyendo (2.7.13) en (2.7.12) y en (2.7.4) la función de verosimilitud es:

$$L(\beta, \theta, \sigma^2 | y) = \prod_{i=1}^M \frac{\exp\left(-\|c_{0(i)} - R_{00(i)}\beta\|^2 / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \text{abs}\left(\frac{|\Delta|}{|R_{11(i)}|}\right) =$$

$$\frac{\exp\left(-\sum_{i=1}^M \|c_{0(i)} - R_{00(i)}\beta\|^2 / 2\sigma^2\right)}{(2\pi\sigma^2)^{\frac{N}{2}}} \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|R_{11(i)}|}\right)$$

Donde el término del exponente tiene la forma de la suma de cuadrados residual para β ponderada sobre todos los grupos. Formando otra descomposición ortogonal triangular.

$$\begin{bmatrix} R_{00(1)} & c_{0(1)} \\ \vdots & \vdots \\ R_{00(M)} & c_{0(M)} \end{bmatrix} = Q_{(0)} \begin{bmatrix} R_{00} & c_0 \\ 0 & c_{-1} \end{bmatrix} \quad (2.7.14)$$

Produce la forma reducida:

$$L(\beta, \theta, \sigma^2 | y) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(\frac{\|c_{-1}\|^2 + \|c_0 - R_{00}\beta\|^2}{-2\sigma^2}\right) \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|R_{11(i)}|}\right) \quad (2.7.15)$$

Para un θ dado, los valores de β y σ^2 que maximizan la expresión anterior son:

$$\hat{\beta}(\theta) = R_{00}^{-1}c_0 \quad \text{y} \quad \hat{\sigma}^2(\theta) = \frac{\|c_{-1}\|^2}{N} \quad (2.7.16)$$

Los cuales resultan en la verosimilitud perfil:

$$L(\theta | y) = L(\hat{\beta}(\theta), \theta, \hat{\sigma}^2(\theta) | y) = \left(\frac{N}{2\pi\|c_{-1}\|^2}\right)^{\frac{N}{2}} \exp\left(-\frac{N}{2}\right) \prod_{i=1}^M \text{abs}\left(\frac{|\Delta|}{|R_{11(i)}|}\right) \quad (2.7.17)$$

O bien la log verosimilitud perfil:

$$\log(L(\theta / y)) = \frac{N}{2} [\log N - \log 2\pi - 1] - N \log \|c_{-1}\| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \quad (2.7.18)$$

La log-verosimilitud perfil (2.7.18) se maximiza con respecto a θ , produciendo el estimador de máxima verosimilitud $\hat{\theta}$. Los estimadores de máxima verosimilitud para $\hat{\beta}$ y $\hat{\sigma}^2$ se obtienen sustituyendo $\theta = \hat{\theta}$ en (2.7.16).

Aunque técnicamente los efectos aleatorios b_i no son parámetros para el modelo estadístico, frecuentemente se desea predecir sus valores. El modelo condicional de los efectos aleatorios, evaluado en el estimador condicional de β , es el mejor predictor lineal insesgado o BLUP de b_i , $i = 1, \dots, M$ y puede ser evaluado usando matrices de la descomposición ortogonal triangular como:

$$\hat{b}_i(\theta) = R_{11(i)}^{-1} (c_{1(i)} - R_{10(i)} \hat{\beta}(\theta)) \quad (2.7.19)$$

En la práctica, si θ es desconocido se reemplaza por su estimador de máxima verosimilitud $\hat{\theta}$, produciendo el estimador BLUP $\hat{b}_i(\hat{\theta})$.

La descomposición (2.7.14) es equivalente al cálculo de la descomposición QR de la matriz de gran tamaño X_e definida en (2.7.9). Si se determina la solución por mínimos cuadrados para (2.7.9) usando la descomposición ortogonal triangular:

$$X_e = Q_e \begin{bmatrix} R_e \\ 0 \end{bmatrix}$$

La parte triangular de la descomposición y la parte principal de la rotación, el vector de respuestas aumentadas deberá ser:

$$R_e = \begin{bmatrix} R_{11(1)} & 0 & \cdots & 0 & R_{10(1)} \\ 0 & R_{11(2)} & \cdots & 0 & R_{10(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & R_{11(M)} & R_{10(M)} \\ 0 & 0 & \cdots & 0 & R_{00} \end{bmatrix} \text{ y } c_1 = \begin{bmatrix} c_{1(1)} \\ c_{1(2)} \\ \vdots \\ c_{1(M)} \\ c_0 \end{bmatrix}$$

De esta forma $\hat{\beta}(\theta)$ y $\sigma^2(\theta)$ de (2.7.16) son los mismos que los obtenidos por (2.7.14) y (2.7.10). El vector c_{-1} es el vector residual en el sistema de coordenadas determinado por Q_e . Porque Q_e es ortogonal, $\|c_{-1}\|^2$ es la suma de los residuales al cuadrado del problema de mínimos cuadrados definido por X_e y y_e .

La log verosimilitud perfil (2.7.17) tiene la misma forma que (2.7.11) con tres componentes aditivos, una constante, una suma de cuadrados residuales en escala logarítmica y una suma de logaritmos de determinantes de proporciones.

2.8 Estimación máximo verosímil restringida para el modelo lineal mixto

La estimación por máxima verosimilitud de los componentes de varianza como σ^2 y σ_b^2 tienden a subestimar estos valores. Muchos analistas prefieren utilizar la estimación máxima verosímil restringida para la estimación de estos componentes.

Ambos métodos sin embargo, tienen el merito de estar basados en los principios de la verosimilitud tales como la consistencia, la normalidad asintótica y la eficiencia. Según Geert Verbeke y Geert Molenberghs la estimación máximo verosímil tiene la ventaja de proveer estimadores de los efectos fijos mientras que la estimación restringida por sí misma no lo hace y según los mismos autores la máxima verosimilitud subestima la varianza residual en estudios ANOVA clásicos o modelos de regresión solamente en los casos en que el rango de la matriz conformada de las variables explicativas sea menor o igual a cuatro pero este criterio es opuesto para la máxima verosimilitud restringida cuando el rango es mayor a cuatro.

Existen varias formas de definir el criterio de estimación de la estimación máximo verosímil restringida (RELM). Una definición que facilita los cálculos computacionales es la propuesta por Laird y Ware en el año 1982:

$$L_R(\theta, \sigma^2 / y) = \int L(\beta, \theta, \sigma^2 / y) d\beta$$

La cual dentro de los métodos bayesianos, corresponde en el asumir una distribución uniforme local para los efectos fijos β e integrarla fuera de la verosimilitud.

Usando (2.7.15) y usando el mismo cambio de variable como en (2.7.13) se tiene la función log verosimilitud restringida:

$$\begin{aligned} l_R(\theta, \sigma^2 / y) &= \log(L_R(\theta, \sigma^2 / y)) \\ &= -\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{\|c_{-1}\|^2}{2\sigma^2} - \log \text{abs} |R_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \end{aligned}$$

Esto genera el estimador condicional para σ^2

$$\hat{\sigma}_R^2(\theta) = \frac{\|c_{-1}\|^2}{N-p}$$

Del cual se genera la función log verosimilitud restringido perfilada:

$$\begin{aligned} l_r(\theta/y) &= l_r(\theta, \hat{\sigma}_R^2(\theta)/y) = -\frac{N-p}{2} \log(2\pi\sigma^2) - \frac{\|c_{-1}\|^2}{2\sigma^2} - \log \text{abs}|R_{00}| \\ &+ \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \\ &= -\frac{N-p}{2} \log(2\pi) - \frac{N-p}{2} \log(\sigma^2) - \frac{\|c_{-1}\|^2}{2\sigma^2} - \log \text{abs}|R_{00}| \\ &+ \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \\ &= -\frac{N-p}{2} \log(2\pi) - \frac{N-p}{2} [\log \|c_{-1}\|^2 - \log(N-p)] - \frac{N-p}{2} - \log \text{abs}|R_{00}| \\ &+ \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \\ &= -\frac{N-p}{2} \log(2\pi) - \frac{N-p}{2} (-\log(N-p)) - \frac{N-p}{2} - \frac{N-p}{2} [\log \|c_{-1}\|^2] - \log \text{abs}|R_{00}| \\ &+ \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \\ &= \frac{N-p}{2} \left[\log \left(\frac{N-p}{2\pi} \right) - 1 \right] - \frac{N-p}{2} [2 \log \|c_{-1}\|] - \log \text{abs}|R_{00}| \\ &+ \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \\ &= C - (N-p) \log \|c_{-1}\| - \log \text{abs}|R_{00}| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta|}{|R_{11(i)}|} \right) \end{aligned} \quad (2.8.1)$$

Los componentes de la función de la log verosimilitud restringida perfilada en (2.8.1) son similares a la log verosimilitud perfil (2.7.18) excepto que el log de la norma del vector residual tiene un múltiplo diferente y hay un término determinante extra:

$$\log abs|R_{00}| = \log \left| \sum_{i=1}^M X_i^T \Sigma_i^{-1} X_i \right| / 2$$

La evaluación de RELM es hecha por la optimización de la log verosimilitud restringido perfilada (2.8.1) con respecto a θ únicamente y usando el resultado del estimador REML $\hat{\theta}_R$ para obtener el estimador REML de $\sigma^2, \hat{\sigma}_R^2(\hat{\theta}_R)$. De una forma similar, el estimador REML de BLUP de los efectos aleatorios son obtenidos reemplazando θ con $\hat{\theta}_R$ en (2.7.19)

El criterio de REML solo depende de los parámetros θ y σ . Sin embargo esto es útil y probablemente aun sensible el evaluar "el mejor invitado" de β de (2.7.16) una vez que $\hat{\theta}_R$ ha sido determinada usando el criterio REML.

Una diferencia importante entre la función de verosimilitud y la función de verosimilitud restringida es que la primera es invariante ante reparametrizaciones uno a uno de los efectos fijos (por ejemplo, un cambio en el contraste representado por una variable categórica), mientras que la última no lo es.

Cambiando las matrices X_i resulta un cambio en $\log abs|R_{00}|$ y un cambio correspondiente en $l_R(\theta | y)$. Como una consecuencia, los modelos lineales de efectos mixtos con diferente estructura de efectos fijos apropiada usando REML no puede ser comparada sobre la base de su verosimilitud restringida. En particular, la razón de verosimilitud no es válida sobre estas circunstancias.

2.9 Estimación máximo verosímil en múltiples niveles de efectos aleatorios

La función de verosimilitud y la función de verosimilitud restringida para modelos multinivel con efectos mixtos pueden calcularse usando las mismas técnicas descritas en las páginas anteriores. Usando el modelo lineal de efectos mixtos, tenemos que la verosimilitud, se integra sobre ambos niveles de efectos aleatorios

$$L(\beta, \theta_1, \theta_2, \sigma^2 | y) = \prod_{i=1}^M \int \prod_{j=1}^{M_i} \left[\int p(y_{ij} | b_{ij}, b_i, \beta, \sigma^2) p(b_{ij} | \theta_2, \sigma^2) db_{ij} \right] p(b_i | \theta_1, \sigma^2) db_i \quad (2.9.1)$$

Al igual que un nivel de efectos aleatorios, se pueden simplificar las integrales de la expresión si se aumentan las matrices Z_{ij} con Δ_2 y se realiza la descomposición ortogonal triangular de estos arreglos aumentados. Esto permitirá evaluar las integrales interiores. Para evaluar las integrales exteriores se itera este proceso, es decir, primero se forman y se descomponen los arreglos:

$$\begin{bmatrix} Z_{ij} & Z_{i,j} & X_{ij} & y_{ij} \\ \Delta_2 & 0 & 0 & 0 \end{bmatrix} = Q^{(ij)} \begin{bmatrix} R_{22(ij)} & R_{21(ij)} & R_{20(ij)} & c_{2(ij)} \\ 0 & R_{11(ij)} & R_{10(ij)} & c_{1(ij)} \end{bmatrix} \quad (2.9.2)$$

Para $i=1,2,3,\dots,M, j=1,2,3,\dots,M_i$

La matriz $R_{22(ij)}$ es una matriz triangular superior de dimensión $q_2 \times q_2$. Los otros arreglos en el primer renglón de la descomposición en (2.9.2) se usan sólo si el estimador condicional $\hat{\beta}(\theta)$ o el modo condicional $\hat{b}_{ij}(\theta)$ y $\hat{b}_i(\theta)$ son requeridos. Los arreglos en el segundo renglón de la descomposición $R_{11(ij)}$, $R_{10(ij)}$ y $c_{1(ij)}$ tienen n_{ij} renglones cada uno. Para evaluar la integral exterior nuevamente se forma y se descompone el arreglo aumentado:

$$\begin{bmatrix} R_{11(i1)} & R_{10(i1)} & C_{1(i1)} \\ \vdots & \vdots & \vdots \\ R_{11(iM_i)} & R_{10(iM_i)} & C_{1(iM_i)} \\ \Delta_1 & 0 & 0 \end{bmatrix} = Q^{(i)} \begin{bmatrix} R_{11(i)} & R_{10(i)} & C_{1i} \\ 0 & R_{00(i)} & C_{0(i)} \end{bmatrix} \quad (2.9.3)$$

Usando las matrices y vectores generados en (2.9.2), (2.9.3) y (2.7.14) y siguiendo los mismos pasos que en el modelo de un solo nivel jerárquico se puede expresar la log verosimilitud perfil para θ_1 y θ_2 como:

$$l(\theta_1, \theta_2 / y) = \log L(\hat{\beta}(\theta_1, \theta_2), \theta_1, \theta_2, \hat{\sigma}^2(\theta_1, \theta_2) / y) = k - N \log \|c_{-1}\| + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta_1|}{|R_{11(i)}|} \right) \\ + \sum_{i=1}^M \sum_{j=1}^{M_i} \log \text{abs} \left(\frac{|\Delta_2|}{|R_{22(ij)}|} \right)$$

De manera similar el perfil de la verosimilitud restringida se expresa como:

$$l_R(\theta_1, \theta_2 / y) = \log L_R(\hat{\beta}_R(\theta_1, \theta_2), \theta_1, \theta_2, \hat{\sigma}_R^2(\theta_1, \theta_2) / y) = k - (N - p) \log \|c_{-1}\| - \log \text{abs} |R_{00}| \\ + \sum_{i=1}^M \log \text{abs} \left(\frac{|\Delta_1|}{|R_{11(i)}|} \right) + \sum_{i=1}^M \sum_{j=1}^{M_i} \log \text{abs} \left(\frac{|\Delta_2|}{|R_{22(ij)}|} \right)$$

Por último, los métodos de cálculo se extienden de igual forma para Q niveles jerárquicos de efectos aleatorios.

CAPÍTULO III

ESTIMACIÓN DE LOS MODELOS DE EFECTOS FIJOS, ALEATORIOS Y MIXTOS

El desarrollo de nuevas técnicas para analizar los datos longitudinales han tenido un gran desarrollo en los últimos veinticinco años y actualmente existen una gran variedad de herramientas que los estadísticos profesionales y empíricos utilizan para resolver problemas de muy diversa índole utilizando un gran número de paquetes informáticos. Dos brechas se están acortando en la actualidad, la primera entre el desarrollo de nuevas metodologías y su implementación casi inmediata en el software estadístico y la segunda, la disponibilidad y acceso al conocimiento ofrecido por medio del internet, conocimientos que anteriormente solo estaban disponibles para un reducido número de personas y con un espectro mínimo de aplicaciones.

Entre los paquetes estadísticos disponible para los modelos lineales de efectos mixtos se utilizo el denominado nlme para el Software R, de Pinheiro y Bates, de fecha 24 de Mayo del 2009. Entre las cualidades específicas de este paquete hay que mencionar sus representaciones gráficas y pruebas ANOVA, que permiten guiar la búsqueda del modelo más apropiado. Otro punto importante en la selección de la librería a utilizar es el hecho de que ambas personas son los autores del libro "Mixed Effects Models in S and S-Plus ". La versión de R utilizada en este trabajo es la número 2.8.1 del 22 de diciembre del año 2008.

En forma general se realizará el análisis del modelo estadístico bajo el siguiente procedimiento: (para un modelo con más variables predictoras y datos no balanceados es necesario realizar mucho más actividades, por ejemplo, para este trabajo no es necesario realizar análisis de colinealidad):

- ❖ Analizar en forma exploratoria la información por medio de gráficas que permitan observar en forma visual la estructura longitudinal de la información.
- ❖ Seleccionar en forma preliminar una estructura de efectos fijos.
- ❖ Seleccionar en forma preliminar una estructura de efectos aleatorios.
- ❖ Seleccionar el mejor modelo con base en los test estadísticos y análisis de residuales: autocorrelación, normalidad y varianza constante.
- ❖ Interpretación y aplicación del modelo.

En cuanto a la variabilidad de la información, la estructura de varianzas:

- ❖ Ayuda a explicar y entender la variabilidad aleatoria de los datos.
- ❖ Es altamente dependiente de los efectos fijos (la parte sistemática de la variabilidad de Y_i)
- ❖ Una estructura de covarianza adecuada conduce a mejores predicciones

Muy importante es encontrar una estructura de medias y varianzas apropiadas en virtud de que ambas son dependientes una de la otra. La estructura de medias se

entiende como la forma óptima en la selección e interacción de las variables independientes.

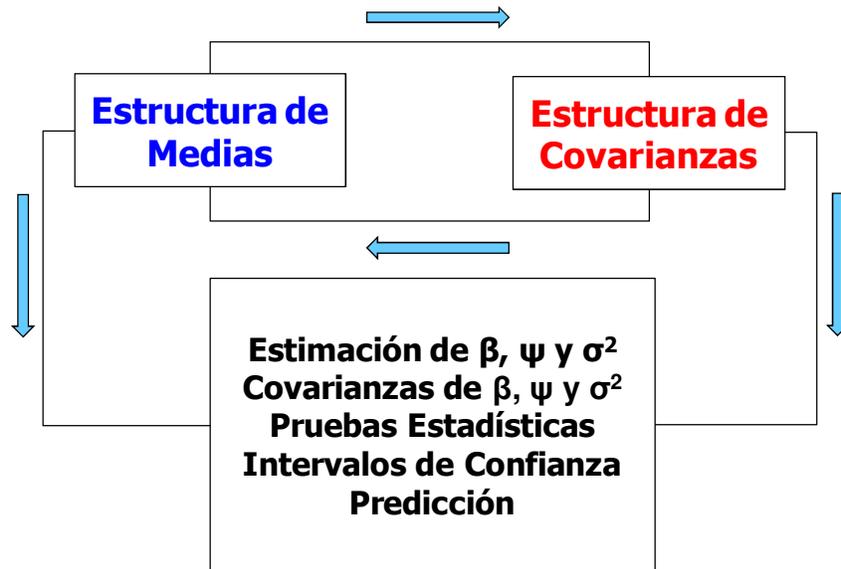


Fig. 3.1 Relación dependiente de las estructuras de los modelos lineales mixtos

3.1 Los datos

El cuadro de resultados de los IOR por entidad federativa se muestra en el anexo (A1) y su representación gráfica se muestra a continuación.

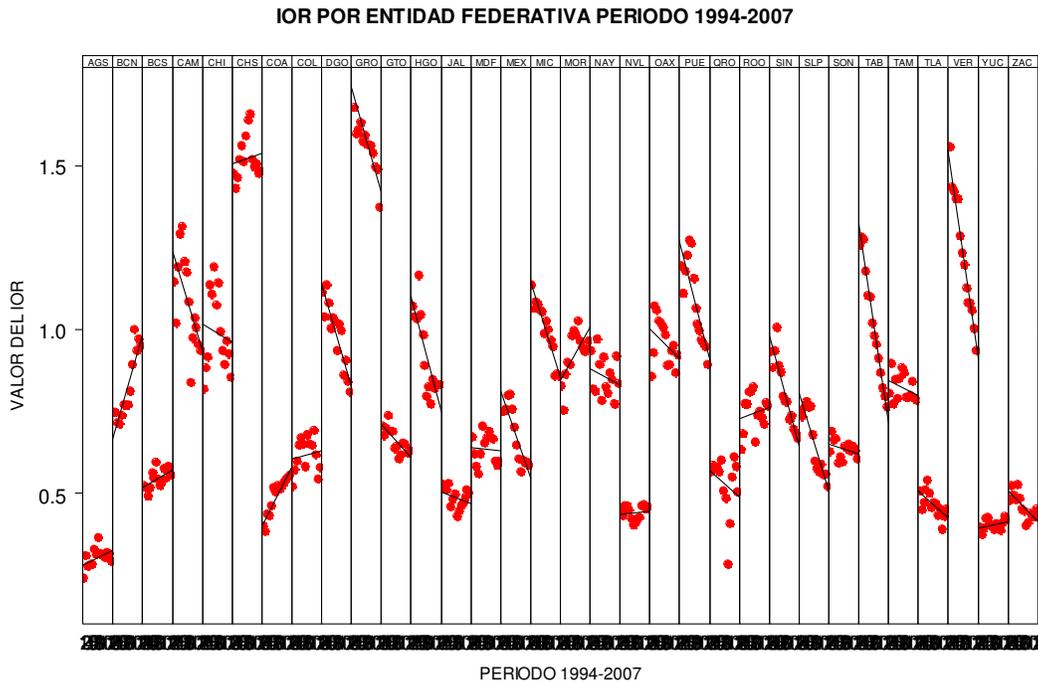


Fig. 3.2 Gráfica del IOR por entidad federativa 1994-2007

Se puede observar los diferentes valores que por entidad federativa tiene el IOR en el periodo 1994-2007 destacando que los valores más bajos corresponden al estado de Aguascalientes y se disputan los valores más elevados el estado de Chiapas y Guerrero.

Una gráfica tipo panel muestra información también interesante:

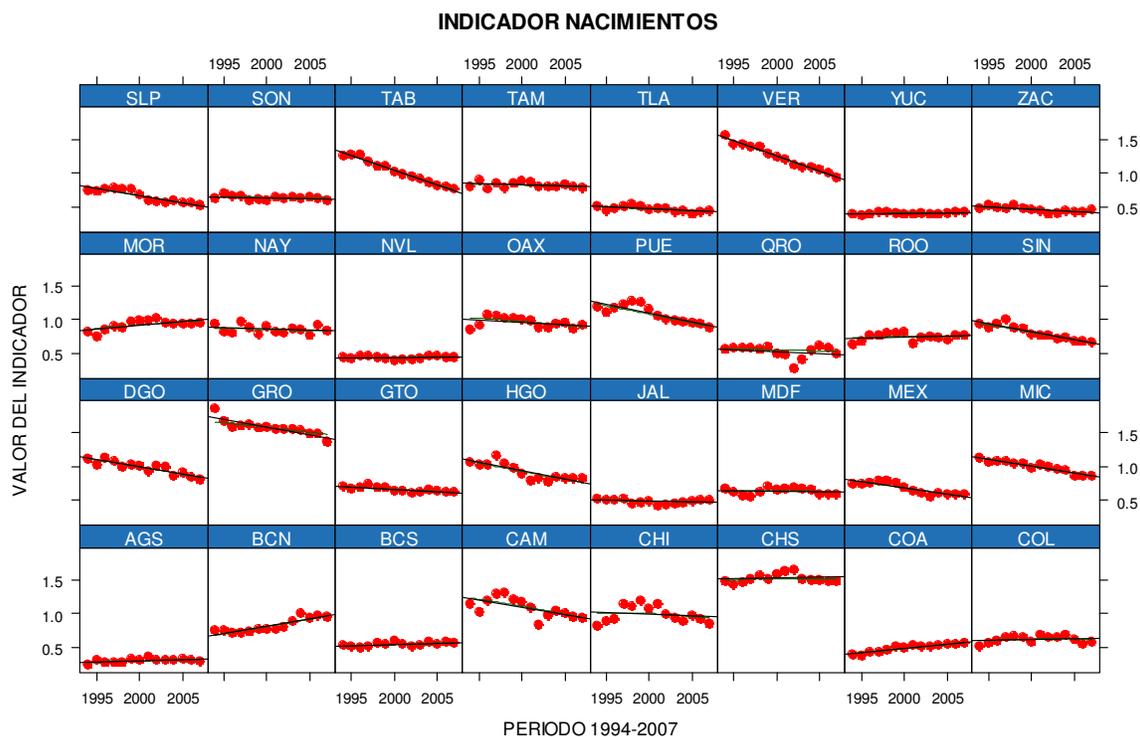


Fig. 3.3 Gráfica de dispersión del IOR por entidad federativa 1994-2007

Cuatro clasificaciones se podrían observar de la gráfica anterior:

- ❖ Estados con alto valor en su IOR y tendencia sin cambio: Chiapas.
- ❖ Estados con alto valor en su IOR y tendencia a la baja: Veracruz y Tabasco.
- ❖ Estados con valor intermedio en el IOR y tendencia a la alza: Morelos, Baja California Norte.
- ❖ Estados con valor intermedio en el IOR y tendencia a la baja: Puebla, Durango, Hidalgo, Michoacán, Sinaloa, San Luis Potosí y Campeche.

Una gráfica de perfiles del IOR por entidad federativa muestra la compactación que a lo largo del tiempo ha sufrido este indicador, se observa muy claramente el nivel descendente para los 32 estados y la conformación de tres grandes grupos en el año 2007:

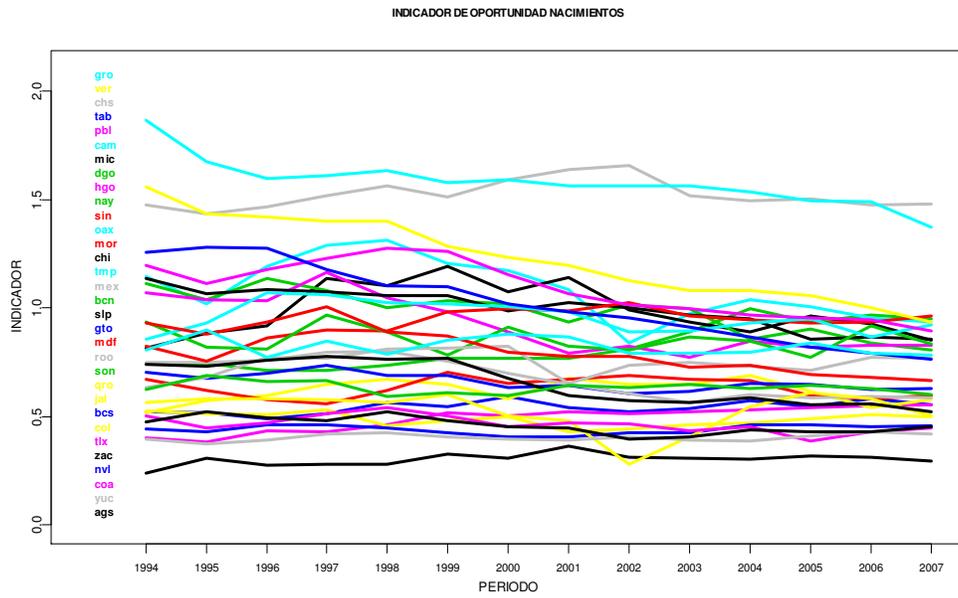


Fig. 3.4 Gráfica de perfiles del IOR por entidad federativa 1994-2007

3.2 El modelo lineal simple

Retomando los comentarios de la sección 2.3 el enfoque más simple de analizar datos tipo panel es omitir las dimensiones del espacio y el tiempo de los datos agrupados y sólo calcular la regresión de mínimos cuadrados ordinarios usual. Una gráfica de dispersión así como la recta ajustada por mínimos cuadrados ordinarios de la forma $y_{it} = \alpha + \beta x_{it} + e_{it}$ (modelo lineal simple) se muestra a continuación:

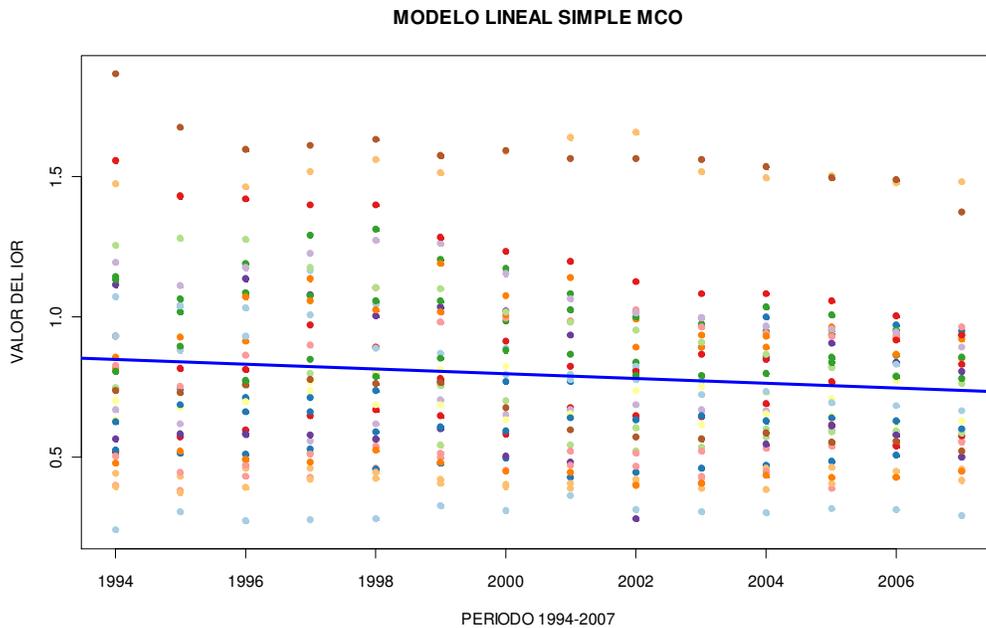


Fig. 3.5 Gráfica de dispersión del IOR por entidad federativa 1994-2007 y el modelo lineal simple

Los parámetros estimados y la varianza de los residuos del modelo lineal simple se presentan a continuación:

$$\alpha = 0.7926 \quad \beta = -0.0085 \quad \sigma^2 = (0.3136)^2$$

Un punto importante en el análisis es tomar en cuenta el factor agrupante estado de la república, de no tomarse en cuenta los residuos resultantes tendrán varianzas altas además de no estar centrados en cero como se muestra en la siguiente gráfica de residuos:

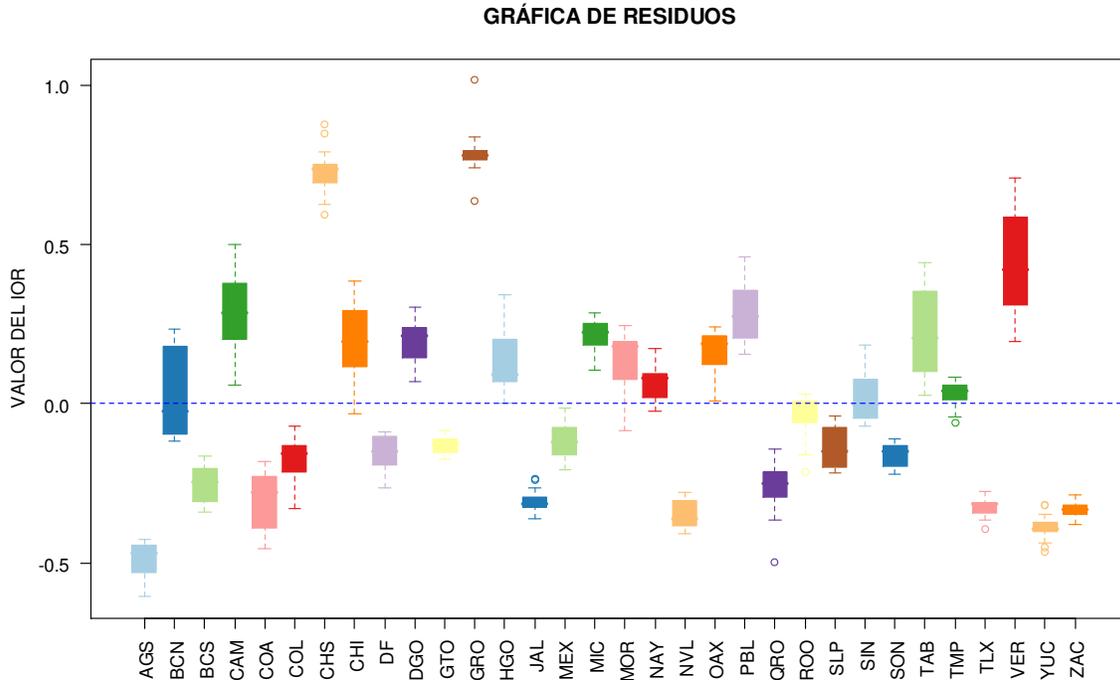


Fig. 3.6 Gráfica de Cajas de los residuos modelo lineal simple

3.3 El modelo efectos fijos

El comando `lmList` de la librería `nlme` de R permite ir un paso adelante con respecto al modelo lineal simple ya que permite obtener estimaciones de modelos lineales considerando el efecto de la entidad federativa, de forma tal, que se puede obtener pendientes y ordenadas al origen para cada entidad federativa.

La instrucción para llamar este comando tiene la forma:

$$lmList(\text{formula}, \text{data})$$

Donde la primera parte especifica el modelo lineal a ser estimado para cada subconjunto de los datos con la forma clásica de R $y \sim x_1 + \dots + x_n \mid g_i$; la segunda parte especifica el factor de agrupación.

El anexo (A2) contiene algunos de los más importantes comandos comando de la función *lmList*.

Los parámetros estimados tomando en cuenta la agrupación para el subíndice *i* para un modelo de efectos fijos de la forma:

$$y_{it} = \alpha_i + \beta_i x_{it} + e_{it}$$

se presentan en la tabla (A3) de los anexos.

Por lo que se refiere a la varianza de los residuos estos son del orden de $\sigma^2 = (0.0545)^2$, solamente una quinta parte del modelo lineal simple.

La representación gráfica de estos 32 modelos lineales estimados se muestra a continuación:

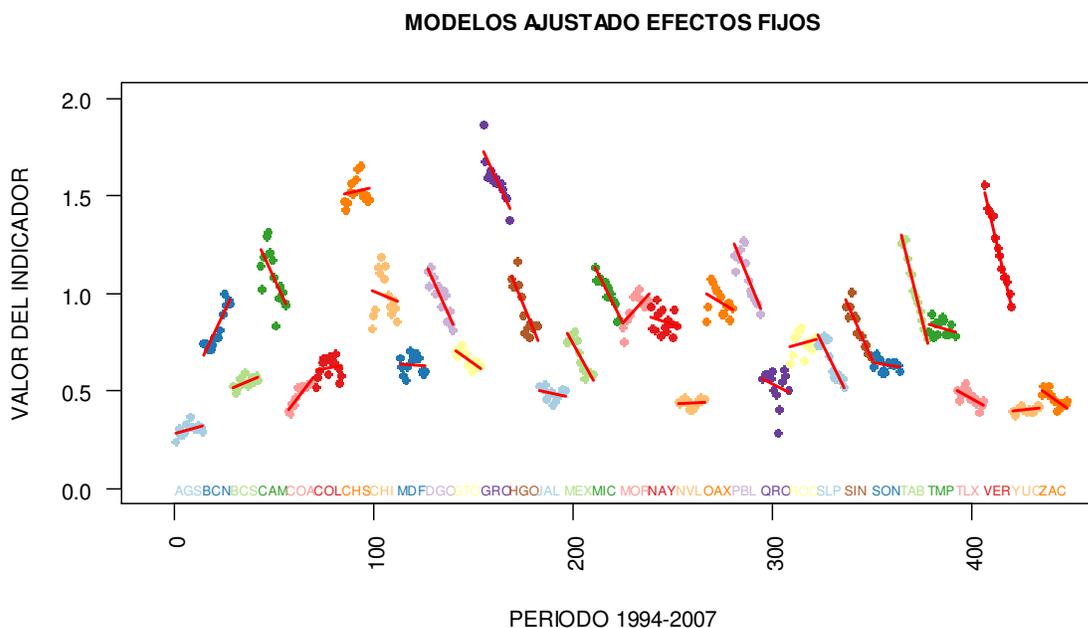


Fig. 3.7 Gráfica de líneas ajustadas modelo de efectos fijos

Por otra parte la gráfica de los residuos del modelo de efectos fijos muestra que están centrados en cero, aunque persiste gran variabilidad entre las entidades:

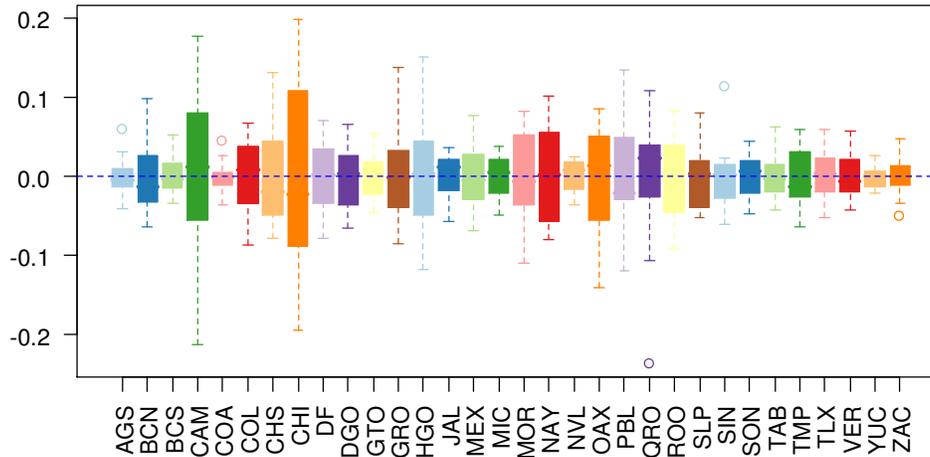
BOX PLOT RESIDUOS EFECTOS FIJOS

Fig. 3.8 Gráfica de Cajas de los residuos del modelo de efectos fijos

Una forma inicial de visualizar un posible modelo de efectos aleatorios es realizar una gráfica de dispersión de los valores de las pendientes e interceptos obtenidos utilizando la función *pairs*:

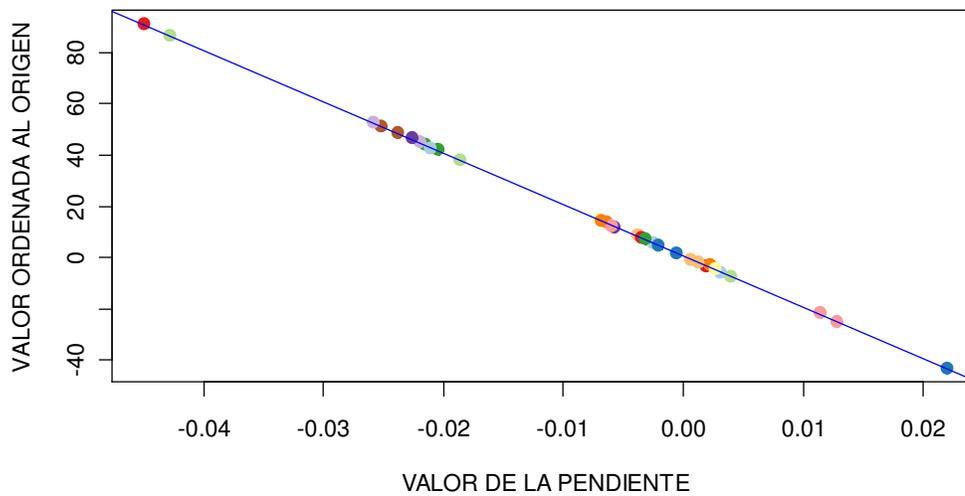
DIAGRAMA DE DISPERSIÓN PENDIENTES-ORDENADAS AL ORIGEN

Fig. 3.9 Gráfica de Dispersión de pendientes y ordenadas al origen del modelo de efectos fijos

Debido a que los datos no están centrados para la variable independiente se observa una correlación entre el valor de la pendiente y la ordenada al origen, esto ocurre porque todos los datos fueron recolectados entre los años 1994 y 2007, pero la intersección de la recta con el eje de las ordenadas representa el valor del IOR al año cero. La extrapolación al año cero dará como resultado un coeficiente de determinación alto (alrededor de 0,99) entre las estimaciones de las pendientes y la estimación de los interceptos.

Centrando los datos, los parámetros estimados muestran un comportamiento totalmente diferente:

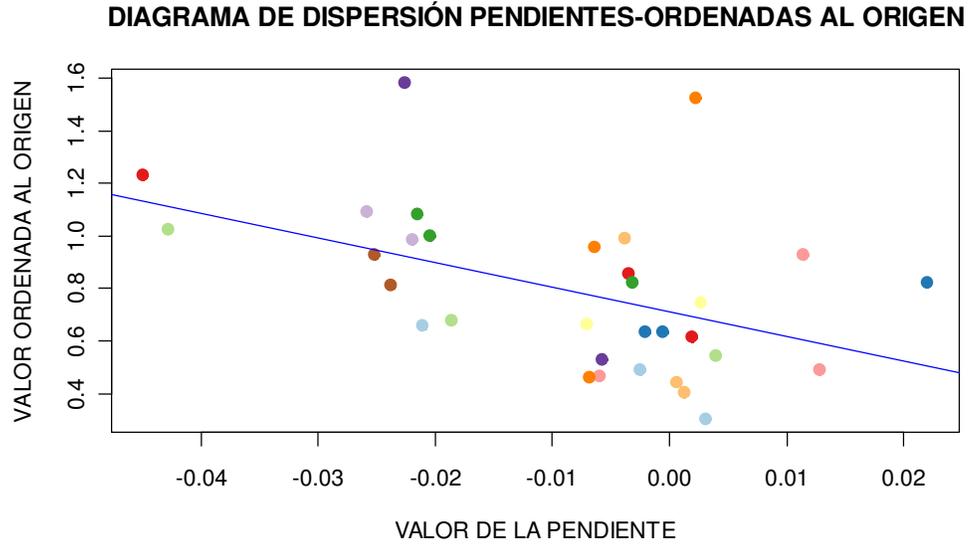


Fig. 3.10 Gráfica de Dispersión de pendientes y ordenadas al origen del modelo de efectos fijos datos centrados

La gráfica anterior indica posibilidades de incorporar un modelo aleatorio, una grafica de intervalos utilizando la función *intervals* ayudará de una forma más directa.

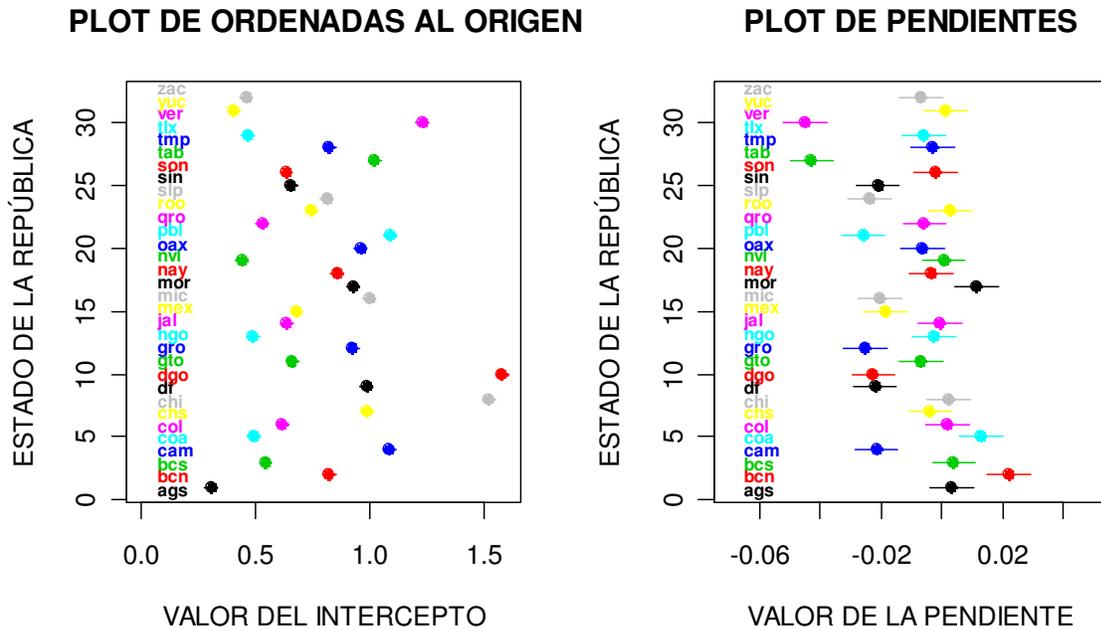


Fig. 3.11 Gráfica de de pendientes y ordenadas al origen del modelo de efectos fijos

Se observa que hay poco solapamiento entre los intervalos generados para el valor del intercepto. Por lo que se refiere a los intervalos de las pendientes si bien existe una mayor sobre posición en ellos los resultados indican que es posible realizar un modelo de efectos aleatorios.

Continuando con el análisis del modelo de efectos fijos, una gráfica de cuantiles teóricos y observados utilizando la función *qqnorm* se ilustra a continuación:

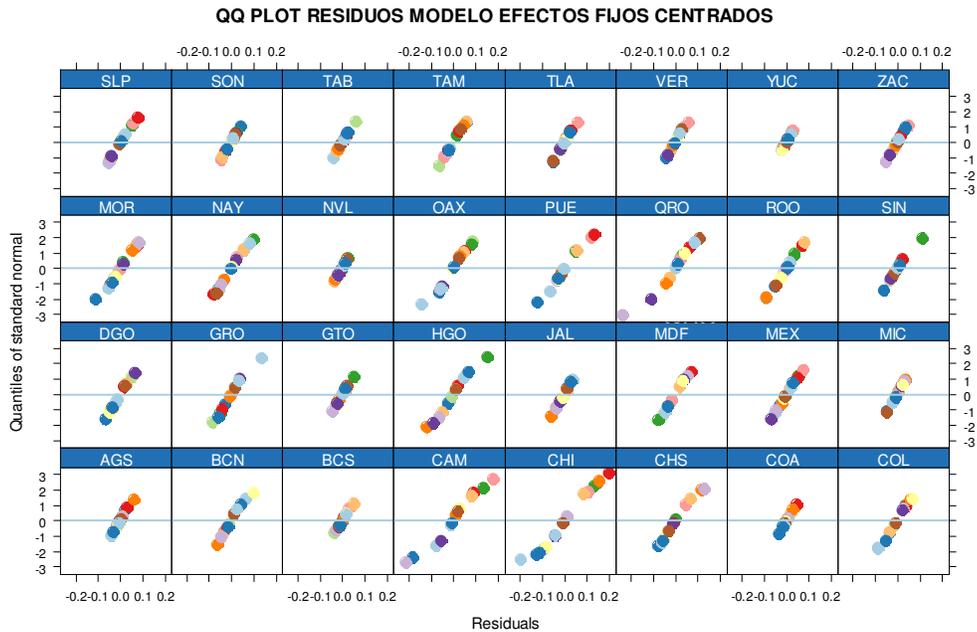


Fig. 3.12 Gráfica de residuos del modelo de efectos fijos

Se observa que los residuos para algunas entidades no se comportan como una distribución normal.

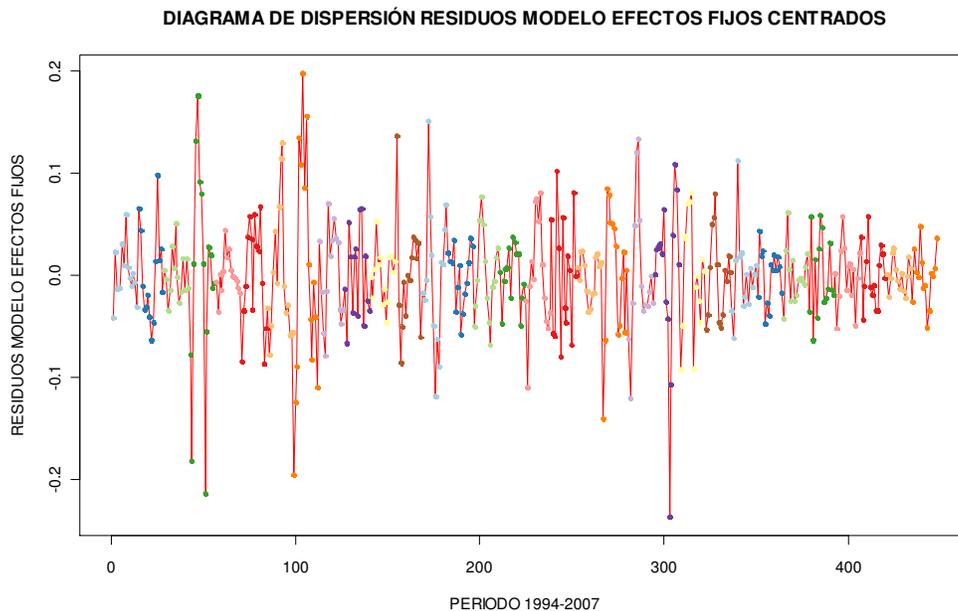


Fig. 3.13 Gráfica de residuos del modelo de efectos fijos

La gráfica anterior muestra residuales heterocedasticos. Por otra parte la función *acf* muestra la problemática de la autocorrelación en los residuos.

FUNCIÓN DE AUTOCORRELACIÓN EFECTOS FIJOS CENTRADOS

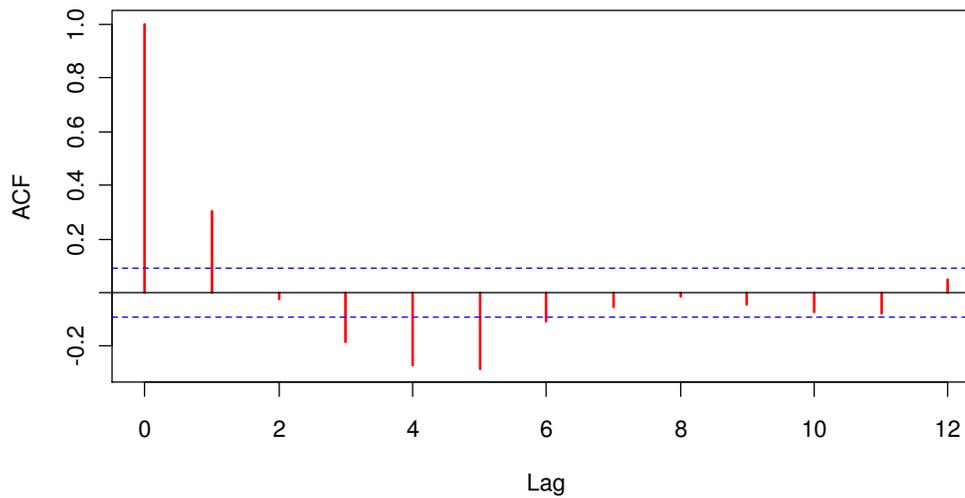


Fig. 3.14 Gráfica de autocorrelación empírica de los residuos del modelo de efectos fijos

De no presentarse la heterocedasticidad, la no normalidad y la autocorrelación de los residuos en el modelo de efectos fijos éste parecería ser una opción viable para ajustar los datos, sin embargo, la teoría en la que se sustenta el modelo de efectos fijos no incluye una forma de poder remediar esta situación, por lo anterior es necesario utilizar otras herramientas para resolver esta problemática.

3.4 El modelo de efectos mixtos

El modelo de efectos fijos y efectos aleatorios tiene en el paquete nlme, la formulación:

$$Lme(\text{fixed}, \text{data}, \text{random})$$

Donde la primera parte especifica el modelo lineal en sus dos lados clásicos especificándose los efectos fijos del modelo; la segunda parte especifica el archivo que contienen los datos que deberán tener formato *data.frame*; la tercera parte se especifica con la notación de relación lineal de un solo sentido los efectos aleatorios y la estructura de agrupación. El anexo (A4) contiene algunas de los más importantes comandos comando de la función *lme*.

Como se observo en la gráfica (3.11) es probable que los datos se ajusten mejor con un modelo de efectos aleatorios.

El modelo de Hausman tiene como propósito principal el apoyar la decisión de seleccionar entre un modelo de efectos fijos o uno aleatorio.

La idea básica del estadístico de Hausman radica en comparar los estimadores de los coeficientes de regresión bajo ambos modelos utilizando el estadístico de prueba:

$$W = (\hat{\beta}_F - \hat{\beta}_A)^T \hat{\Sigma} (\hat{\beta}_F - \hat{\beta}_A)$$

Bajo una hipótesis nula de que W se distribuye como una *ji-cuadrada* con k grados de libertad (número total de variables regresoras), se tiene el siguiente resultado:

$$W = (-0.00853573 - (-0.00853573))^2 / (0.00000091406 - 0.0000009118626) \approx 0$$

Con un *p-valor* asociado aproximadamente igual a uno no se rechaza la hipótesis nula y se recomienda usar el modelo de efectos aleatorios.

Dos modelos básicos se analizaron:

- ❖ Modelo de interceptos aleatorios y pendientes fijas
- ❖ Modelo Mixto

El modelo de interceptos aleatorios (ialea) arroja la siguiente gráfica donde se observa que para ciertos estados de la república no es conveniente mantener fijo el valor de las pendientes.

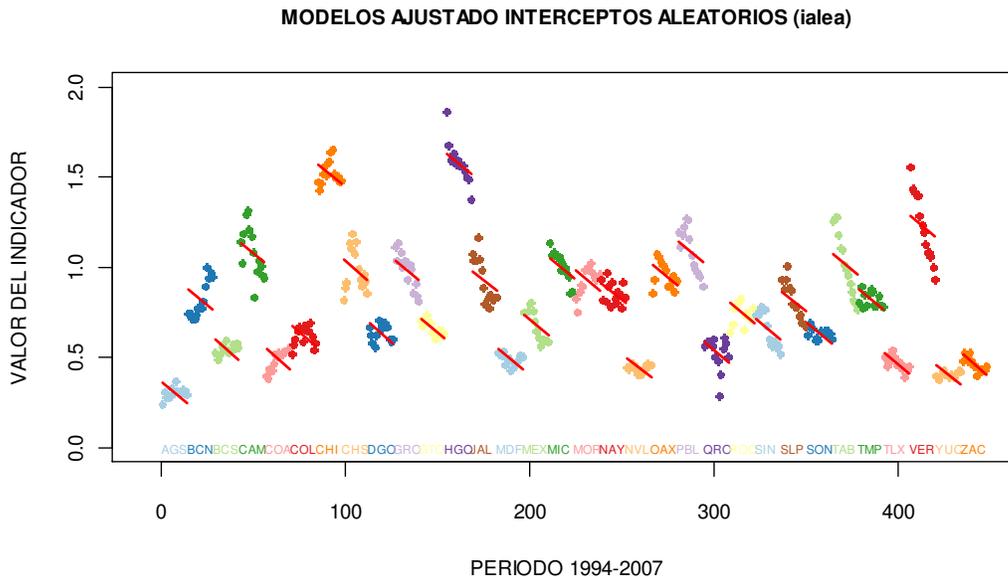


Fig. 3.15 Gráfica de líneas ajustadas modelo de efectos aleatorios

Un método general para comparar dos modelos que utilizan la misma información lo proporciona el comando *anova*, este comando suministra los valores estimados de la *logverosimilitud* de los modelos, el *Criterio de Información Akaike (AIC)* (Sakamoto, Ishiguoy Kitagawa.1986), el *Criterio de Información Bayesiana (BIC)* (Schwarz. 1978) y la *Prueba de Razón de Verosimilitud (Test L. Ratio)*.

Los criterios de evaluación se determinan de la siguiente forma:

$$AIC = -2(\log L) + 2(p) \quad BIC = -2(\log L) + (p) \log(N)$$

Donde $\log L$ es la log verosimilitud obtenida por el modelo, p es el número de parámetros y N es el número total de observaciones.

Para el modelo mixto se tiene por ejemplo:

$$AIC = -2(528.4576) + 2(6) = -1,044.9152$$

Los criterios de evaluación AIC y BIC consideran el principio “*más pequeño es mejor*”. Esto es, si se utiliza el criterio AIC para comparar dos modelos o más, se selecciona el modelo que tenga el valor más pequeño.

El *Test de Razón de Verosimilitud (TLR)* considera si $L2$ es la verosimilitud del modelo general y $L1$ es la del modelo restringido se cumple que $L2 > L1$ y esta desigualdad implica que $\log L2 > \log L1$ por ser la verosimilitud mayor o igual a cero. En la terminología de modelos anidados este principio se refiere a que si un modelo se dice anidado si y solo si representa un caso especial de otro modelo. Considerando $M1$ y $M2$ dos modelos estadísticos tal que $M1$ está anidado en $M2$, con verosimilitudes $L1$ y $L2$, respectivamente. Entonces, se debe tener $L2 > L1$, y podemos comparar modelos anidados usando el test de razón de verosimilitudes. El TLR se define como:

$$2(\log(L2/L1)) = 2[\log(L2) - \log(L1)]$$

Si p_i es el número de parámetros a ser estimados para el modelo i , entonces la distribución asintótica de la distribución de los TLR bajo la hipótesis nula de que el modelo restringido es estadísticamente igual al modelo general tendrá una distribución *ji-cuadrada* con $p_2 - p_1$ grados de libertad.

Para el comparativo del modelo mixto y el modelo de interceptos aleatorios se tiene:

$$TLR = 2[528.4576 - 403.4031] = 250.109$$

El p -valor para 250.109 bajo una *ji-cuadrada* con $(6-4=2)$ grados de libertad tiene el valor de cero, por lo tanto se rechaza la hipótesis nula que el modelo restringido es estadísticamente igual al modelo general.

De acuerdo al comparativo obtenido con el comando *anova* entre el modelo de interceptos aleatorios (*ialea*) y el modelo mixto (*aleau*), un mejor ajuste lo muestra el modelo mixto:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
	ialea	1 4	-798.8062	-782.387	403.4031			
	aleau	2 6	-1044.9151	-1020.286	528.4576	1 vs 2	250.1090	<.0001

La gráfica del modelo mixto se muestra a continuación (aleau):

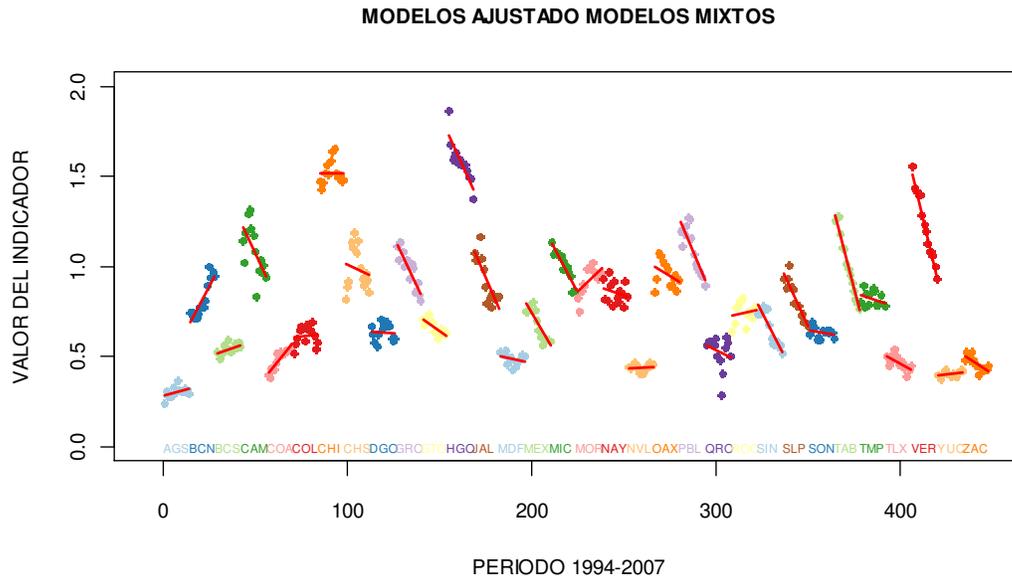


Fig. 316 Gráfica de líneas ajustadas modelo de efectos mixtos

Su ajuste es muy similar al realizado por el modelo de efectos fijos visto en el apartado 3.3 bajo el comando *lmList*, un comparativo de los coeficientes del modelo mixto y el de efectos fijos muestra que los interceptos son prácticamente iguales y los coeficientes de las pendientes tienen poca diferencia:

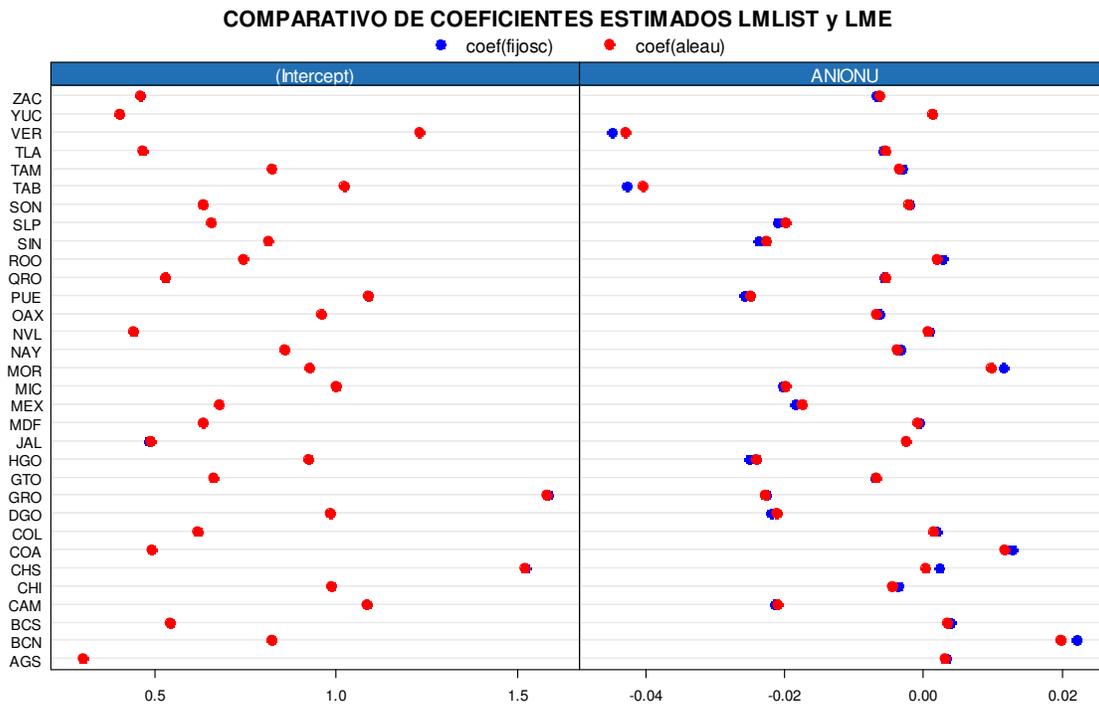


Fig. 3.17 Gráfica de coeficientes modelo de efectos fijos y aleatorios

Una prueba *anova* entre el modelo de efectos fijos y el modelo mixto muestra que el valor del AIC se inclina por el modelo de efectos fijos, mientras que el valor del BIC indica que es el modelo mixto el más apropiado estadísticamente, el *TLR* indica que los modelos son estadísticamente diferentes:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
aleau	1	6	-1044.915	-1020.286	528.4576			
fijosc	2	65	-1273.092	-1006.281	701.5463	1 vs 2	346.1774	<.0001

Un punto importante se comenta en las notas del Dr. Rogelio Ramos Q.

"Por un lado, el modelo con efectos fijos es costoso en términos de grados de libertad usados para la estimación de efectos individuales y, para estudios longitudinales con un gran número de unidades, el modelo de efectos aleatorios es más intuitivo."

Un refuerzo a favor del modelo de efectos aleatorios es la prueba de *Hausman* que indica que el modelo mixto es más aceptable que el modelo de efectos fijos, por esta razón, se continuará el desarrollo del modelo mixto para ajustar los datos del IOR.

3.5 Las funciones de varianza

Continuando con el modelo mixto la gráfica de los residuos se muestra a continuación:

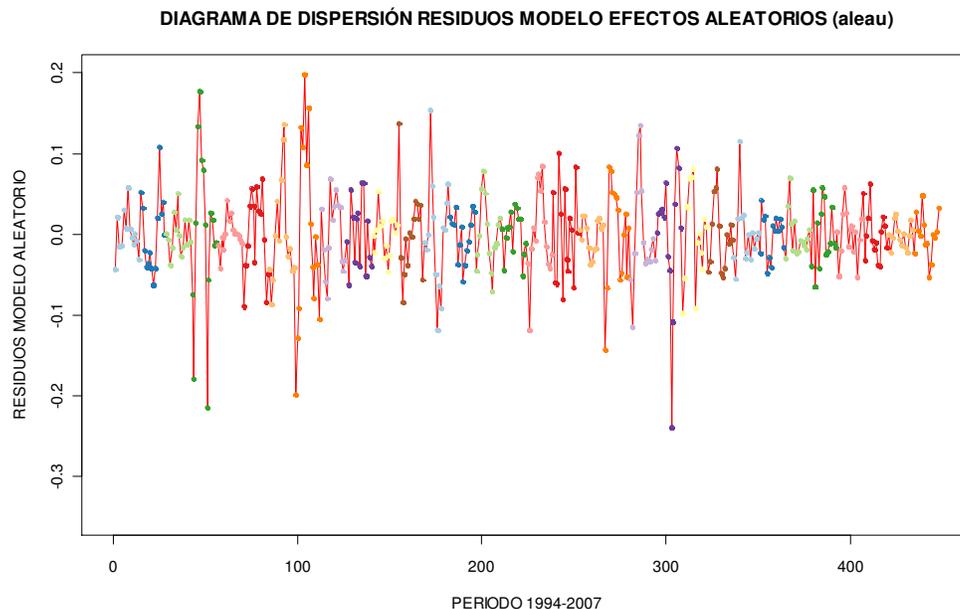


Fig. 3.18 Gráfica de residuos modelo de efectos aleatorios

El modelo de efectos mixtos considera los errores dentro de los grupos independientes e idénticamente distribuidos con media cero y varianza constante, sin embargo existen muchas aplicaciones donde se involucran datos agrupados para los cuales los errores dentro de los grupos son heterocedásticos como muestra la gráfica anterior.

Otra herramienta para detectar la heterocedasticidad de los errores dentro de los grupos es la gráfica de los residuos estandarizados contra los valores estimados y alguna de las covariables de varianza, en el modelo mixto *aleau* se observa que la variabilidad decrece conforme crece el tiempo t , como únicamente se tiene una covariable de varianza se utiliza esta en la gráfica de residuos:

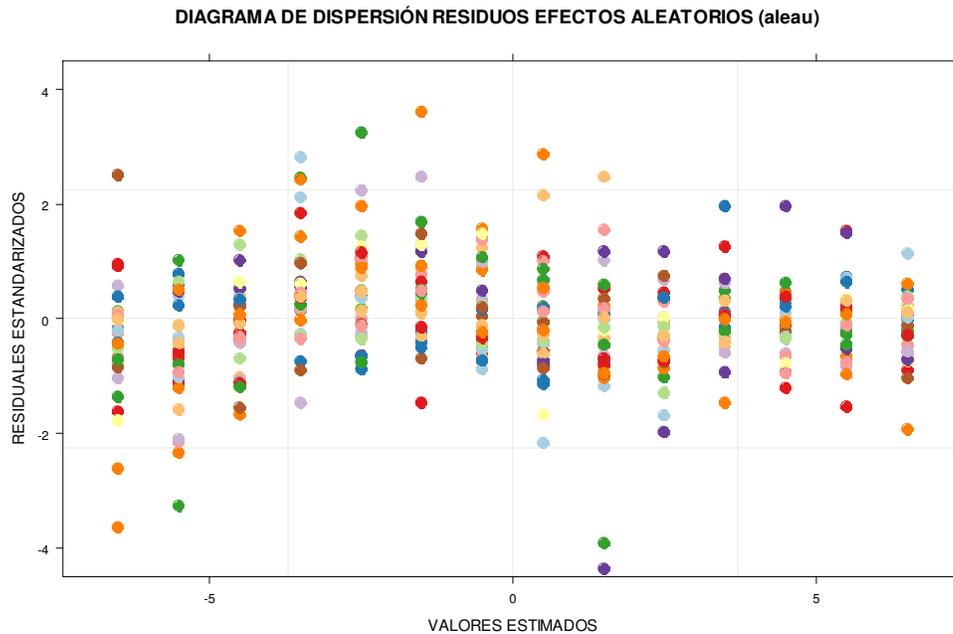


Fig. 3.19 Gráfica de residuos heterocedásticos modelo de efectos aleatorios

Para solventar la problemática anterior se utilizarán las funciones de varianza para modelar la estructura de varianza de los errores dentro de los grupos.

Se define la función general de varianzas para los errores dentro de los grupos como:

$$\text{Var}(e_{it} / b_i) = \sigma^2 g^2(\mu_{it}, v_{it}, \delta)$$

Donde $\mu_{it} = E[y_{ij} / b_i]$ (respuesta media), v_{it} es un vector de covariables de varianza, δ es un vector de parámetros de varianza y $g()$ es la función de varianza asumiendo continuidad en δ .

En la librería *nlme* existe un conjunto de clases de funciones de varianza para especificar la estructura de varianza dentro de los grupos: *varFixed*, *varIdent*, *varPower*, *varExp*, *varConsPower* y *varComb*

varFixed

Esta clase representa la función de varianzas sin parámetros y una covariable de varianza simple, utilizada cuando la varianza dentro de los grupos es conocida, por ejemplo, si se asume que la varianza dentro de los grupos se incrementa linealmente con la covariable $A\tilde{N}O$, se tiene:

$$Var(e_{it}) = \sigma^2 A\tilde{N}O_{it}$$

Correspondiendo una función de varianza de la forma:

$$g(A\tilde{N}O_{it}) = \sqrt{A\tilde{N}O_{it}}$$

varIdent

Esta clase representa un modelo de varianzas con diferentes varianzas para cada nivel o variable de estratificación S , tomando valores en el conjunto $\{1, 2, 3, \dots, S\}$.

$$Var(e_{it}) = \sigma^2 \delta_{S_{it}}^2$$

Correspondiendo una función de varianza de la forma:

$$g(s_{it}, \delta) = \delta_{it}$$

El modelo usa $S+1$ parámetros para representar S varianzas y por tanto no son identificables, para conseguir esto, se imponen restricciones en el parámetro de varianzas δ . Usando $\delta_1 = 1$ de tal forma que $\delta_l, l = 2, \dots, S$ representa la razón entre las desviaciones estándar del l -ésimo estrato y el primer estrato.

varPower

El modelo de varianza representado por esta clase es:

$$Var(e_{it}) = \sigma^2 |v_{it}|^{2\delta} \tag{3.5.1}$$

Correspondiendo una función de varianza de la forma:

$$g(v_{it}, \delta) = |v_{it}|^\delta$$

La cual es una potencia del valor absoluto de la covariable de varianza. El parámetro δ es irrestricto, lo cual significa que puede tener cualquier valor en el campo de los reales y por tanto (3.5.1) se puede modelar para casos donde la varianza se incrementa o decrece con el valor absoluto de la covariable de varianza.

Se muestra a continuación una gráfica de residuos tipo panel del modelo *aleau* en donde se aprecia que varias entidades tiene residuos heterocedásticos:

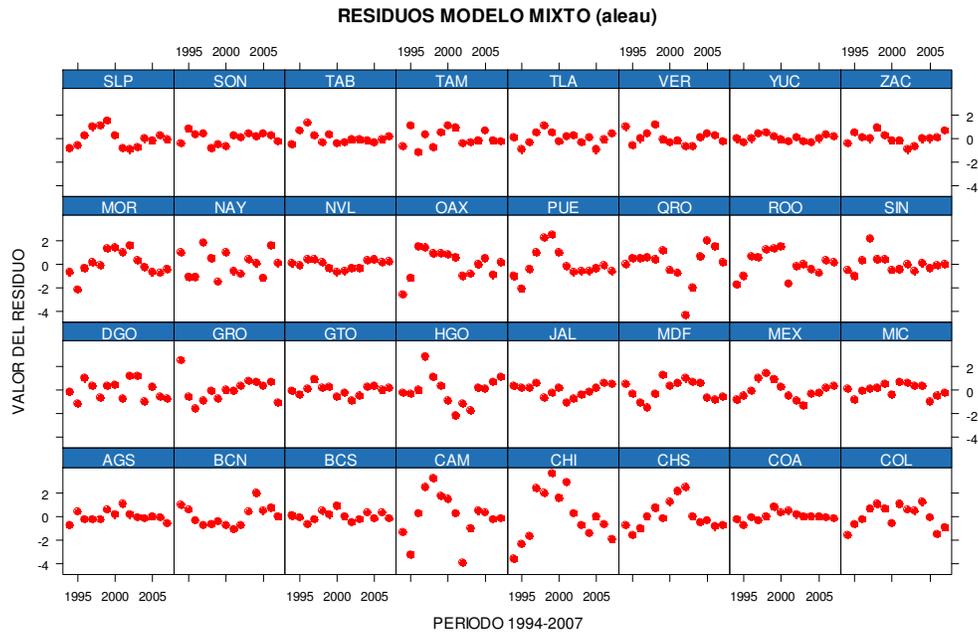


Fig. 3.20 Gráfica de panel residuos heterocedásticos modelo de efectos aleatorios

Utilizando la clase *varPower* para la función de covarianzas se genera un nuevo modelo *aleaw* y se observa que los valores de los residuos se ubican en un rango más pequeño:

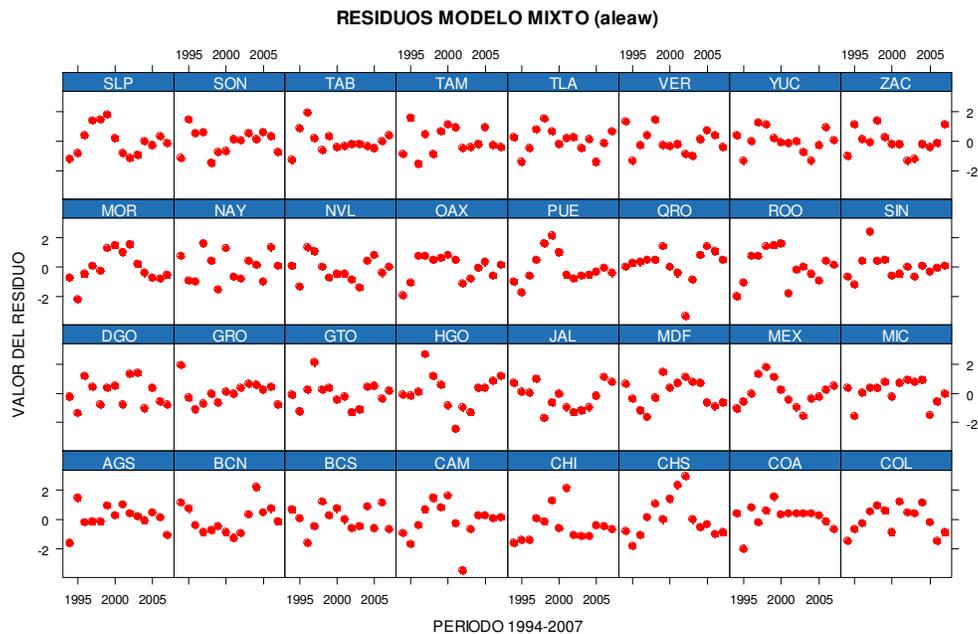


Fig. 3.21 Gráfica de panel residuos modelo de efectos aleatorios varPower

Una gráfica de líneas muestra varianzas razonablemente más homogéneas:

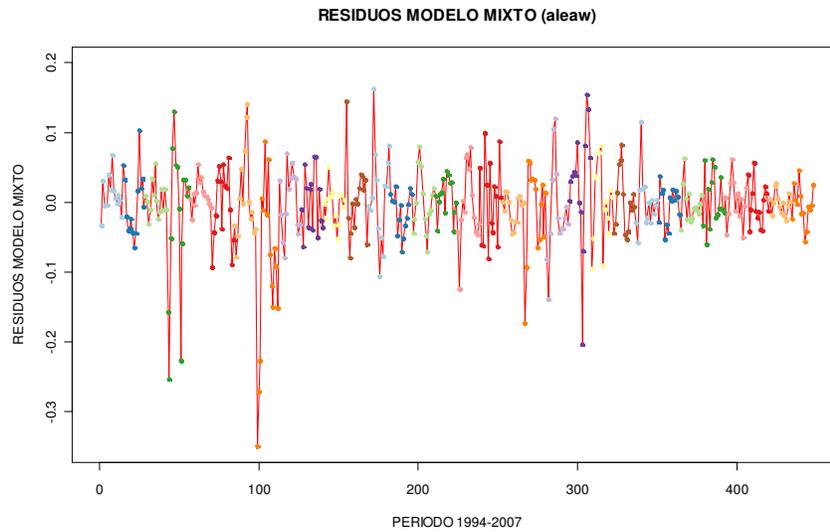


Fig. 3.22 Gráfica de líneas residuos modelo de efectos aleatorios varPower

La gráfica utilizando la covariable de varianza también muestra mejora en la variabilidad de los residuos:

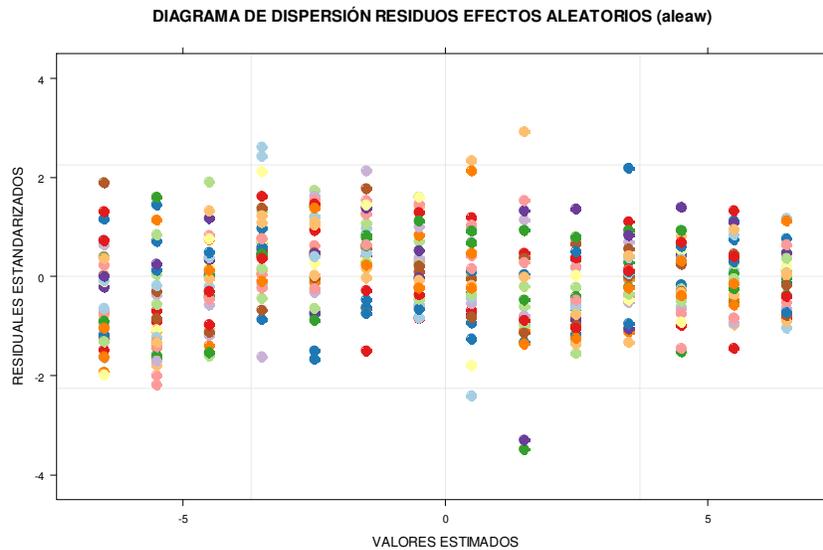


Fig. 3.23 Gráfica de residuos modelo de efectos aleatorios varPower

La prueba *anova* entre los dos modelos se muestra a continuación:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
aleau	1	6	-1044.915	-1020.2864	528.4576			
aleaw	2	38	-1135.004	-979.0218	605.5020	1 vs 2	154.0888	<.0001

De la prueba estadística se concluye que si es conveniente incluir el modelo homocedástico ya que al incorporar la función de varianza *varPower* se incrementa la log-verosimilitud.

Para tener una idea de los parámetros obtenidos del modelo de varianza la siguiente gráfica muestra en color rojo la tendencia "suave" de los residuos del estado de Coahuila y la línea azul la tendencia de la función de varianza con parámetro $\hat{\delta} = -0.8095$. Se observa claramente que las tendencias son muy parecidas.

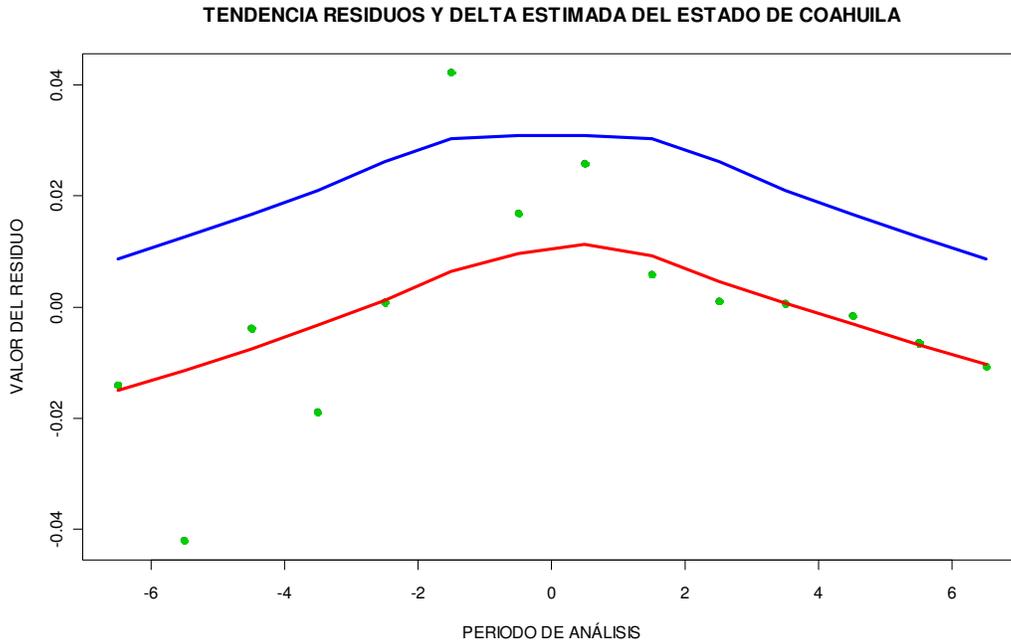


Fig. 3.24 Gráfica de residuos del estado de Coahuila efectos aleatorios aleaw

El mismo delta se utiliza para "normalizar" los residuos: $r_{ij} = (y_{ij} - \hat{y}_{ij}) / \hat{\sigma} x_{ij}^{\hat{\delta}}$

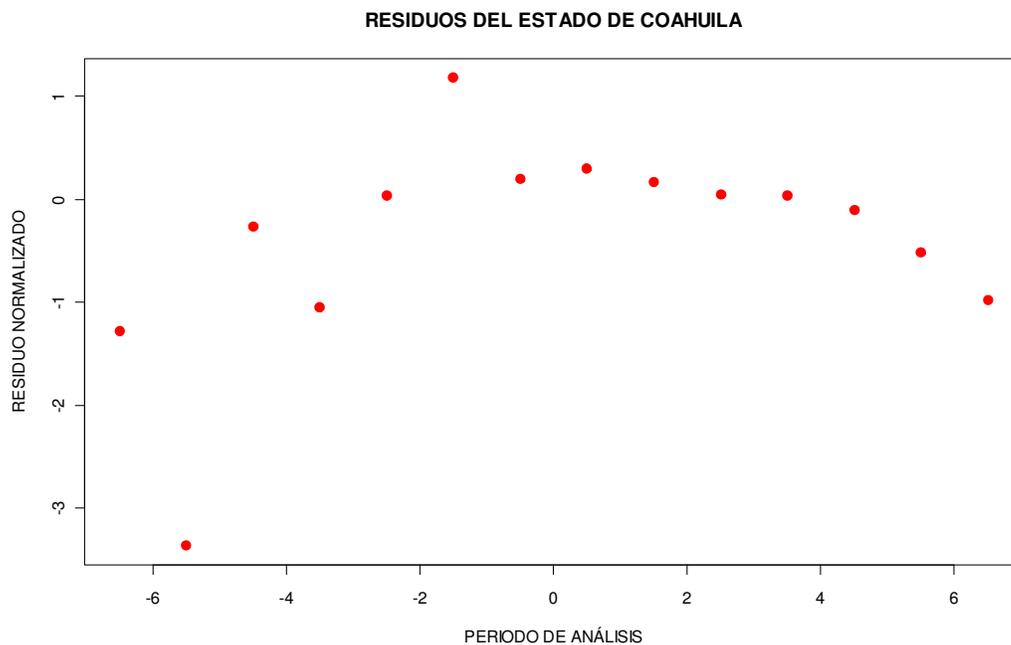


Fig. 3.25 Gráfica de residuos normalizados del estado de Coahuila efectos aleatorios aleaw

En cuanto al comportamiento en la normalidad, esta refleja un comportamiento adecuado:

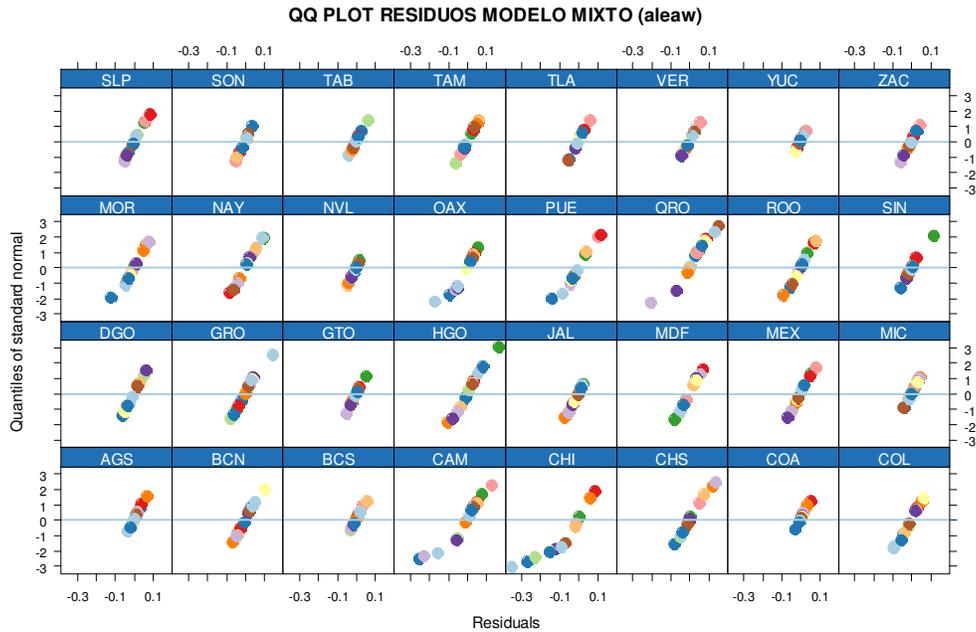


Fig. 3.26 Gráfica qqnorm residuos efectos aleatorios aleaw

Sin embargo se tiene nuevamente la problemática de la autocorrelación en los residuos para el modelo aleaw. Utilizando el comando *ACF* se tiene la siguiente gráfica:

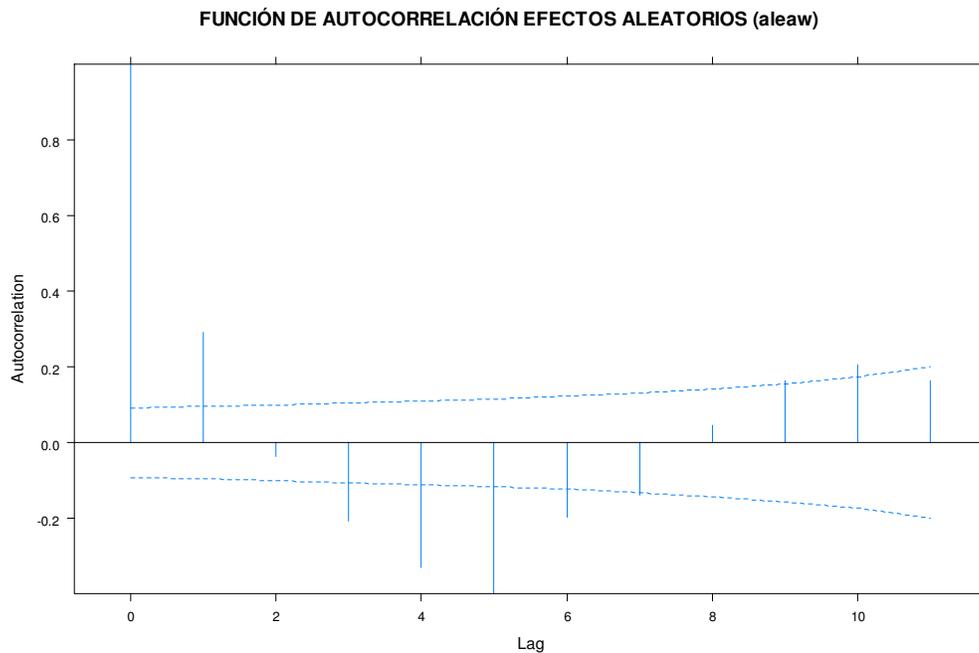


Fig. 3.27 Gráfica de autocorrelación de los residuos efectos aleatorios aleaw

3.6 Las estructuras de correlación

Las estructuras de correlación se utilizan para modelar la dependencia entre los valores observados. En el contexto de los modelos lineales mixtos estas estructuras son utilizadas para modelar la dependencia de los errores dentro de los grupos. Históricamente las estructuras de correlación han sido desarrolladas para dos principales tipos de datos: las series de tiempo y los datos espaciales, las primeras generalmente asociadas a una variable de tiempo y las segundas referidas principalmente a la localización en un plano real de dos dimensiones.

Para establecer en forma general las estructuras de correlación se asume que los valores e_{it} están asociados con vectores de posición p_{it} . Para datos de series de tiempo los p_{it} son típicamente valores escalares enteros, mientras que para los datos espaciales son los elementos del vector situado en un plano de dos dimensiones

Las estructuras de correlación se suponen isotrópicas, esto es, la correlación entre dos errores dentro del mismo grupo e_{it} , $e_{it'}$ asumen dependencia en sus correspondientes vectores posicionales únicamente a través de la distancia que hay entre ellos y no de los valores o magnitudes o direcciones que tengan.

La estructura de correlación para un nivel simple de agrupamiento se expresa como:

$$Cor(e_{it}, e_{it'}) = h(d(e_{it}, e_{it'}), \rho)$$

Donde d es la distancia entre los valores, ρ es un vector de parámetros de correlación y $h()$ es la función de correlación que toma valores entre -1 y 1 para ρ continua.

Estructura de Correlación Serial

La estructura de correlación serial es utilizada para modelar la dependencia para datos observados a través del tiempo vistos como vectores unidimensionales.

La correlación serial general se define como:

$$Cor(e_{it}, e_{it'}) = h(|(e_{it} - e_{it'})|, \rho)$$

Que en el contexto de series de tiempo la función $h()$ toma la figura de la *Función de Autocorrelación*. La función de autocorrelación empírica es un estimador no paramétrico de la *Función de Autocorrelación* definida como:

$$\hat{\rho}(l) = \frac{\sum_{i=1}^N \sum_{t=1}^{n_i-l} r_{it} r_{i(t+l)} / M(l)}{\sum_{i=1}^N \sum_{t=1}^{n_i} r_{it}^2 / M(0)}$$

Donde:

$$r_{it} = (y_{it} - \hat{y}_{it}) / \hat{\sigma}_{ij} \text{ son los residuales estandarizados del modelo mixto estimado}$$

$$\text{con } \sigma_{ij}^2 = \text{Var}(e_{ij})$$

$M(l)$ es el numero de pares residuales usados en la distancia l (lag)

Se describen a continuación dos estructuras de correlación más usadas en la práctica y que la librería nlme ofrece a sus usuarios.

❖ *Componente Simétrico*

Esta es la estructura más simple para los errores dentro de los grupos la cual asume igualdad en las correlaciones y su modelo se escribe como:

$$\text{Cor}(e_{it}, e_{it'}) = \rho \quad \forall t \neq t', \quad h(k, \rho) = \rho \quad k = 1, \dots$$

Donde ρ es referido generalmente como el coeficiente de correlación intraclase.

❖ *Componente de promedios móviles autoregresivos*

Esta familia de estructura de correlaciones incluye varios modelos lineales estacionarios: los modelos autoregresivos, los modelos de promedios móviles y los modelos mixtos ARMA.

Los modelos autoregresivos expresan la observación actual como una función lineal de observaciones previas independientes más un término de ruido homocedástico:

$$e_t = \phi_1 e_{t-1} + \dots + \phi_p e_{t-p} + a_t \quad \text{donde } p \text{ es llamado el orden del modelo}$$

El $AR(1)$ es el modelo autoregresivo más simple pero uno de los más usados y su función de correlación decrece exponencialmente con las distancias (*lag*):

$$h(k, \phi) = \phi^k \quad k = 0, 1, \dots$$

Para un orden mayor la función de correlación tiene la forma:

$$h(k, \phi) = \phi_1 h(|k-1|, \phi) + \dots + \phi_p h(|k-p|, \phi) \quad k = 0, 1, \dots$$

Para el modelo de promedios móviles se asume que la observación actual es una función lineal de términos de ruido independientes e idénticamente distribuidos:

$$e_t = \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} + a_t \quad \text{donde } q \text{ es llamado el orden del modelo}$$

Existen q parámetros de correlación en el modelo, dados por $\theta = (\theta_1, \dots, \theta_q)$.

La estructura de correlación de promedios móviles se define como:

$$h(k, \phi) = \begin{cases} \frac{\theta_k + \theta_1 \theta_{k-1} + \dots + \theta_{k-q} \theta_k}{\theta_1^2 + \dots + \theta_q^2} & k = 1, \dots, q \\ 0 & k = q+1, q+2, \dots \end{cases}$$

Para observaciones con más de q unidades de tiempo son incorrelacionadas.

Los modelos ARMA son los formados por la combinación de los modelos autoregresivos y de promedios móviles:

$$e_t = \sum_{i=1}^p \phi_i e_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + a_t$$

Hay $p+q$ parámetros de correlación ρ en un modelo $ARMA(p, q)$.

La estructura de correlación se define como:

$$h(k, \rho) = \begin{cases} \phi_1 h(|k-1|, \phi) + \dots + \phi_p h(|k-p|, \phi) + \theta_1 \psi(k-1, \rho) + \dots + \theta_q \psi(k-q, \rho) & k = 1, \dots, q \\ \phi_1 h(|k-1|, \phi) + \dots + \phi_p h(|k-p|, \phi) & k = q+1, q+2, \dots \end{cases}$$

Donde:

$$\psi(k, \phi, \theta) = E[e_{t-k} a_t] / \text{var}(e_t) \text{ para valores } k = 1, 2, 3, \dots \psi(k, \phi, \theta) = 0$$

Y en este caso e_{t-k} y a_t son independientes y $E(a_t) = 0$.

Estructura de Correlación Espacial

Las estructuras de correlación espacial fueron diseñadas para modelar la dependencia de los datos por medio de vectores en un plano de dos dimensiones.

Por simplicidad se denota e_x a la observación con posición $x = (x_1, \dots, x_r)^T$. Cualquier distancia métrica isotrópica podrá ser usada para la estructura de correlación espacial, las más comunes son las siguientes:

$$\text{Euclidiana} = d(e_x, e_y) = \sqrt{\sum_{i=1}^r (x_i - y_i)^2}$$

$$\text{Manhattan} = d(e_x, e_y) = \sum_{i=1}^r |x_i - y_i|$$

$$\text{Maxima} = d(e_x, e_y) = \max_{i=1, \dots, r} |x_i - y_i|$$

Las estructuras de correlación espacial con una función distancia $d(\cdot)$ son representadas por los semivariogramas que tienen la siguiente forma:

$$\gamma[d(e_x, e_y), \lambda] = \frac{1}{2} \text{Var}(e_x - e_y) = \frac{1}{2} \text{E}(e_x - e_y)^2$$

Los errores dentro de los grupos pueden estandarizarse para que tengan varianza unitaria sin cambiar la estructura de correlación, si se asume esto la función γ dependerá únicamente del parámetro de correlación ρ y se tendrá la siguiente igualdad:

$$\gamma[s, \rho] = 1 - h(s, \rho)$$

Un estimador del semivariograma tiene la forma:

$$\hat{\gamma}(s) = \frac{1}{2M(s)} \sum_{i=1}^N \sum_{d(p_{it}, p_{it'})=s} (r_{it} - r_{it'})^2$$

Donde:

$r_{it} = (y_{it} - \hat{y}_{it}) / \hat{\sigma}_{ij}$ son los residuales estandarizados del modelo mixto estimado
 con $\sigma_{ij}^2 = \text{Var}(e_{ij})$

$M(s)$ es el numero de pares residuales a una distancia s entre un par y otro.

Un estimador robusto que toma en cuenta los valores atípicos tiene la forma:

$$\bar{\gamma}(s) = \left[\frac{1}{2M(s)} \sum_{i=1}^N \sum_{d(p_{it}, p_{it'})=s} \sqrt{|r_{it} - r_{it'}|} \right]^4 / (0.457 + 0.494 / M(s))$$

La librería *nlme* cuenta con varias estructuras de correlación espacial:

Exponencial $\gamma[s, \rho] = 1 - \exp(-s / \rho)$

Gausiana $\gamma[s, \rho] = 1 - \exp(-(s / \rho)^2)$

Lineal $\gamma[s, \rho] = 1 - (s / \rho) I_{s < \rho}$

Cuadratica	$\gamma[s, \rho] = (s / \rho)^2 / [1 + (s / \rho)^2]$
Esferica	$\gamma[s, \rho] = 1 - [1 - 1.5(s / \rho) + 0.5(s / \rho)^3] I_{s < p}$

Donde $I_{s < p}$ es una variable indicadora que toma el valor de 1 cuando se cumple que $s < p$ y cero en otro caso.

Retornando al problema de los residuos correlacionados del modelo *aleaw* se buscará un modelo que proporcione un mejor ajuste a los datos tomando en cuenta cuatro puntos importantes a observar:

❖ Punto uno

Observar el comportamiento de los coeficientes de autocorrelación y autocorrelación parcial para determinar en base a su comportamiento un modelo AR, AM o ARMA.

❖ Punto dos

Observar las medidas de bondad de ajuste "Akaike's information criterion (AIC)" propuestas por Hirotugu Akaike o bien la medida AICC que es la misma AIC mas un ajuste por muestra pequeñas.

❖ Punto tres

Observar los errores estándar de los coeficientes obtenidos y verificar que sean estadísticamente diferentes de cero.

❖ Punto cuatro

Observar el comportamiento de los p-valores de la prueba de test de independencia de Ljung- Box donde valores pequeños de los mismos indican dependencia.

De acuerdo a los puntos anteriores el modelo que mejor ajusta los datos es un modelo $ARMA(p=2, q=2)$.

La bondad de ajuste (AIC), los coeficientes y sus errores estándar para corroborar la significancia estadística se presentan a continuación:

	ar1	ar2	ma1	ma2	intercept	AIC
	1.3635	-0.5959	-1.1913	0.2631	-0.0175	
s.e.	0.1096	0.0999	0.1337	0.1353	0.0124	1120.45

Con un alfa de 5% tres son estadísticamente diferentes de cero (por milésimas el coeficiente *ma2* no rechaza la hipótesis nula de igualdad cero).

En cuanto a los p-valores de Ljung-Box y la autocorrelación de los residuales mostraron el siguiente resultado:

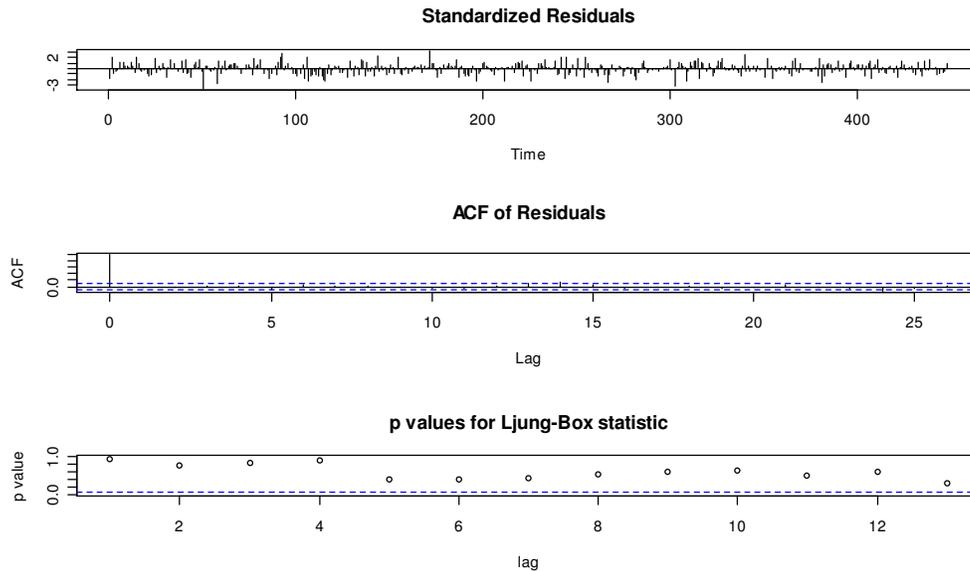


Fig. 3.28 Modelo autoregresivo y promedios móviles aleawarma

Incorporando la estructura de correlación ARMA(2,2) al modelo *aleaw* se tiene un nuevo modelo *aleawarma*. Utilizando el comando *intervals* se pueden obtener los resultados de la estimación de los parámetros y su intervalo de confianza:

	lower	est.	upper
Phi1	0.8378536	1.35458578	1.2291102
Phi2	-0.7276269	-0.53192724	-0.2561583
Theta1	-0.7606323	-0.88048748	0.9165424
Theta2	-0.2370988	0.05354637	0.3353935

El cuadro anterior de cifras indica que solamente ϕ_1 y ϕ_2 son estadísticamente diferente de cero. Realizando una especie de "step wise" se tiene el valor calculado para la estructura de correlación ARMA(1,0):

	lower	est.	upper
Phi	0.514175	0.6748564	0.7897726

Realizando la prueba *anova* entre los modelos *aleaw* y *aleawarma* se tiene:

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
aleaw		1	38	-1135.004	-979.0218	605.5020		
aleawarma		2	39	-1241.764	-1081.6772	659.8821	1 vs 2	108.7602
								<.0001

El bajo valor de la prueba indica que la incorporación del modelo ARMA(1,0) tiene como resultado un mejor ajuste a los datos.

Utilizando la función de autocorrelación empírica se puede confirmar visualmente lo que nos indica la prueba *anova*:

FUNCIÓN DE AUTOCORRELACIÓN EFECTOS ALEATORIOS (aleawarma)

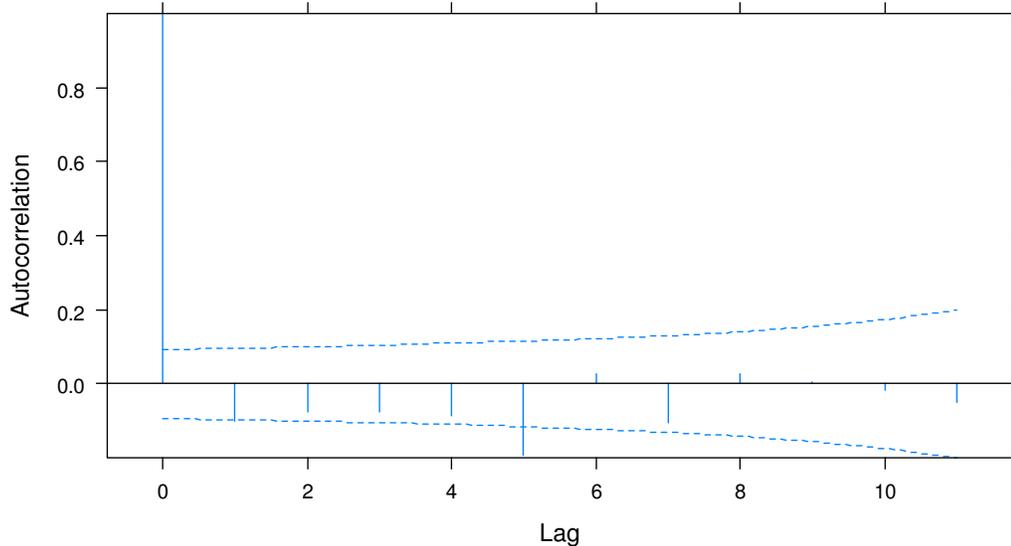


Fig. 3.29 Gráfica de autocorrelación empírica de los residuos efectos aleatorios aleawarma

Utilizando el método *anova* y *update* se realiza un comparativo de *corARMA* con varias estructuras de correlación. El cuadro siguiente proporciona evidencia donde se muestra que el modelo *aleawarma* es el que estima los valores más pequeños para los criterios AIC y BIC. Por otra parte el modelo *corExp* y el modelo *corARMA(1,0)* son iguales, la igualdad se obtiene definiendo a $\phi = \exp(-1/\rho)$ y tomando en cuenta que la función de correlación asociada con una estructura exponencial se define como $h(s, \phi) = \phi^s$.

Model	df	AIC	BIC	logLik
aleawarma	1 39	-1241.764	-1081.6772	659.8821
aleaw	2 38	-1135.004	-979.0218	605.5020
aleawRAT	3 39	-1235.196	-1075.1093	656.5981
aleawSPH	4 39	-1235.153	-1075.0663	656.5766
aleawLIN	5 39	-1214.326	-1054.2390	646.1630
aleawGAU	6 39	-1219.389	-1059.3016	648.6943
aleawEXP	7 39	-1241.764	-1081.6772	659.8821

[¡Y por fin se ha llegado a un modelo estadísticamente aceptable!](#)

Como se comento anteriormente la estructura de medias y la estructura de varianzas son dependientes una de la otra y esto obliga a revisar nuevamente la homocedasticidad de los errores del último modelo *aleawarma*. La siguiente grafica de residuos del modelo *aleawarma* no difiere de la obtenida por el modelo *aleaw*:

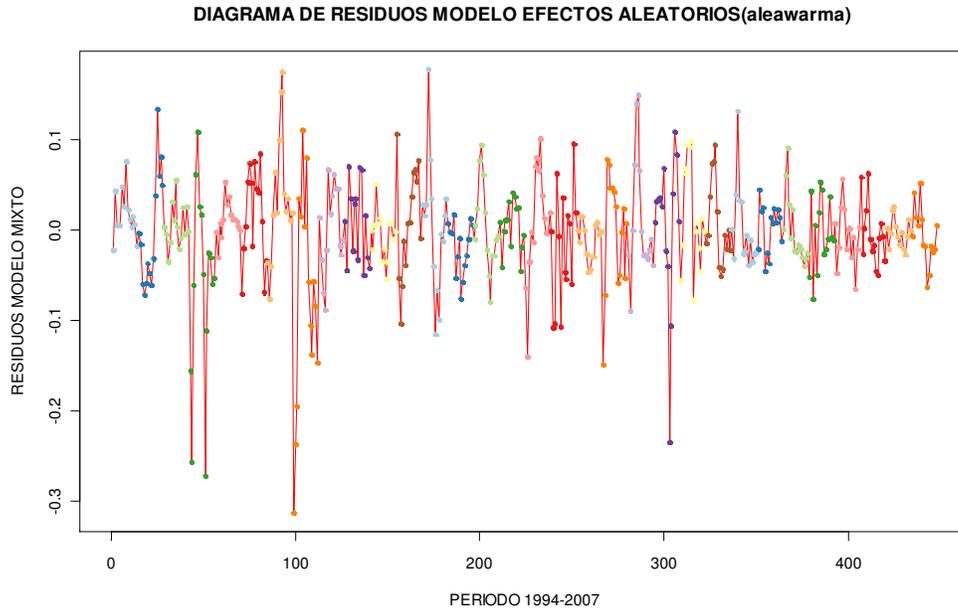


Fig. 3.30 Gráfica de líneas de los residuos efectos aleatorios aleawarma

En cuanto a la normalidad se tienen algunos ligeros cambios con respecto al modelo *aleaw* como muestra a continuación la siguiente gráfica:

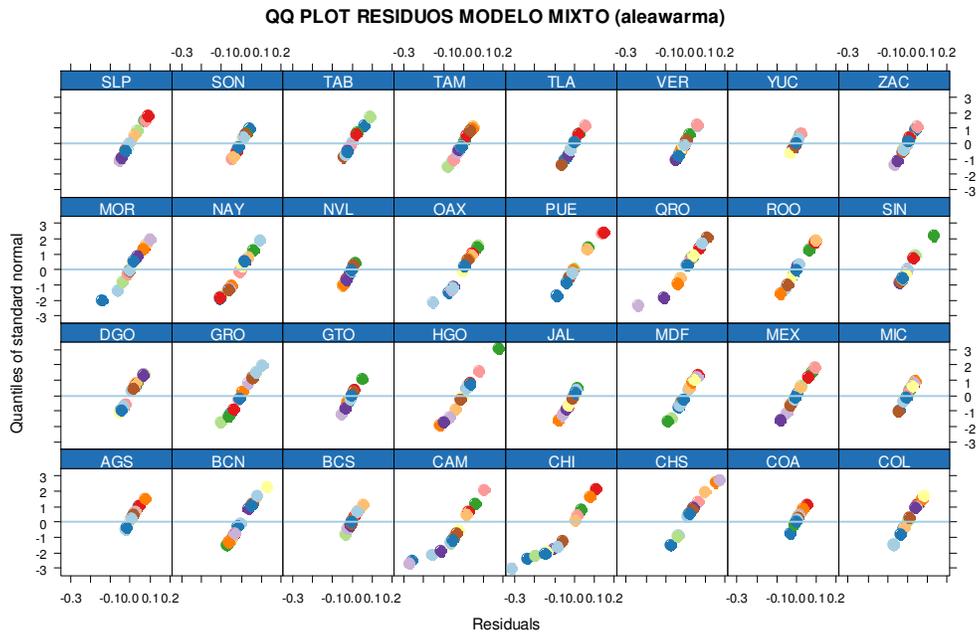


Fig. 3.31 Gráfica qqnorm residuos efectos aleatorios aleawarma

Se observa que los residuos para algunas entidades no se comportan como una distribución normal, sin embargo la mayoría están distribuidos simétricamente y de acuerdo a José C. Pinheiro y Douglas M. Bates las estimaciones del modelo mixto no debe cambiar sustancialmente.

Realizando la prueba de normalidad propuesta por Samuel Shapiro y Martin Wilk donde se establece la hipótesis nula que los residuos tiene una distribución normal, los resultados indican que cuatro de los treinta y dos estados no cumplen el supuesto de normalidad (ver anexo A5).

Los cuatros estados que no cumplen el supuesto de normalidad son: Puebla, Querétaro, Sinaloa y Tabasco, sin embargo, la situación no es tan grave en virtud de que una gran mayoría de los estados si cumple el supuesto, por lo anterior, no existe inconveniente en utilizar el modelo *aleawarma*.

Por otra parte cuando la instrucción *anova* únicamente incluye un solo modelo se obtiene resultados similares a la tabla *anova* clásica como son: la suma de cuadrados, los grados de libertad del numerador y denominador, los f valores y el p-valor del contraste de Wald, la instrucción *anova* para el modelo *aleawarma* se muestra a continuación:

	numDF	denDF	F-value	p-value
(Intercept)	1	415	232.99733	<.0001
ANIONU	1	415	10.18108	0.0015

El resultado de la prueba de Wald indica que hay suficiente evidencia estadística para rechazar la hipótesis nula que establece igualdad en el comportamiento medio de los IOR para cada entidad federativa.

En cuanto a los efectos aleatorios la librería RLRsim de R proporciona una prueba estadística basada en una distribución de muestreo finita y asintótica de razones de verosimilitud propuesta por Crainiceanu y Ruppert en el año 2004 y que es utilizada como la distribución nula para la prueba estadística:

$$H_0: Var(b) = 0 \quad \text{contra} \quad H_1: Var(b) > 0$$

Utilizando el comando *exactLRT* con 10,000 simulaciones se obtiene el siguiente resultado:

```
LRT = 1550.229, p-value < 2.2e-16
```

Y por tanto, hay suficiente evidencia estadística para rechazar la hipótesis nula de que las varianzas de los efectos aleatorios son iguales a cero.

Se muestra a continuación los semivariogramas del penúltimo modelo *aleaw* y el modelo definitivo *aleawarma*:

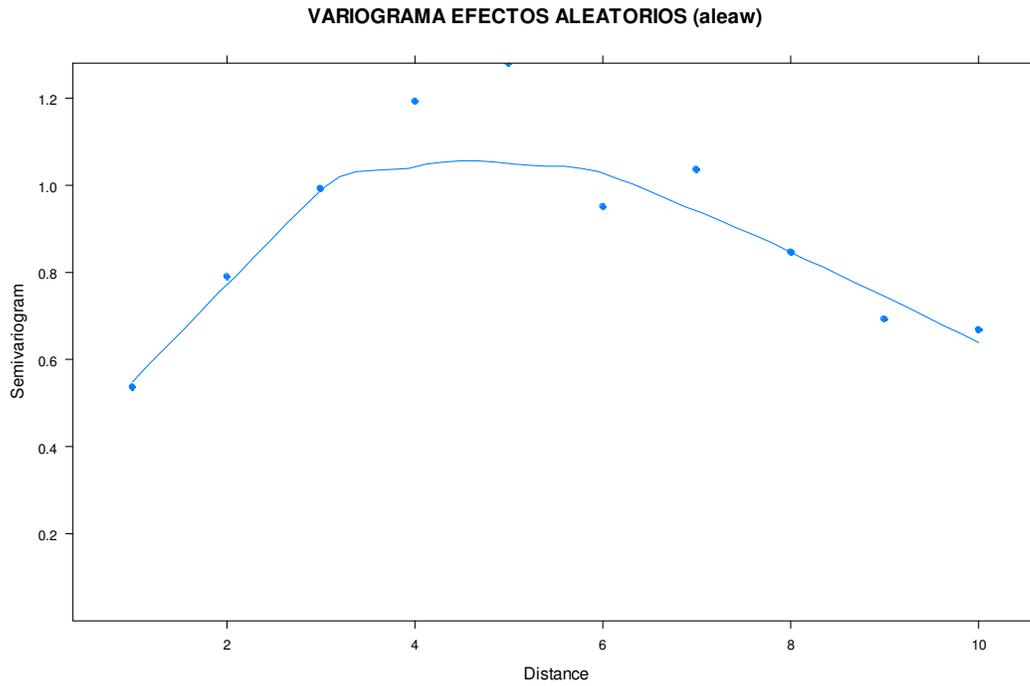


Fig. 3.32 Gráfica de autocorrelación espacial de los residuos efectos aleatorios *aleaw*

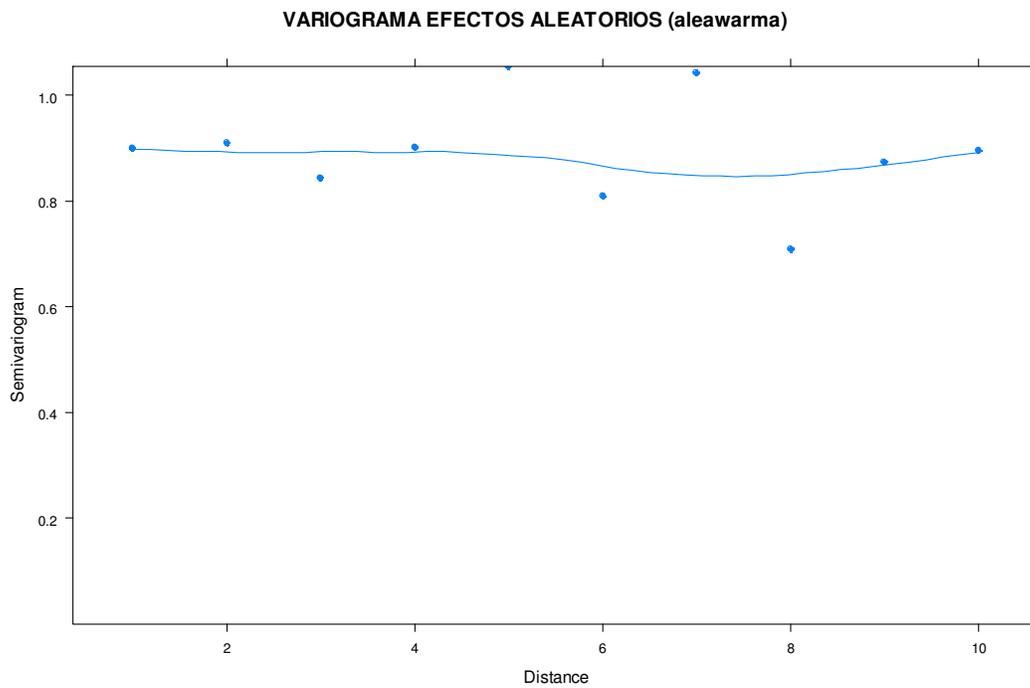


Fig. 3.33 Gráfica de autocorrelación espacial de los residuos efectos aleatorios *aleawarma*

Analizando los variogramas se observa que los residuos normalizados son correlacionados para el modelo *aleaw* mientras que para el modelo *aleawarma* son incorrelacionados (se observan constantes) y por lo tanto nuevamente se confirma que el modelo *aleawarma* es el más adecuado.

La gráfica comparativa de coeficientes del modelo mixto (*aleawama*) y el modelo de efectos fijos (*fijosc*) muestra valores diferentes a los presentados en la página 59 para las entidades de la república.

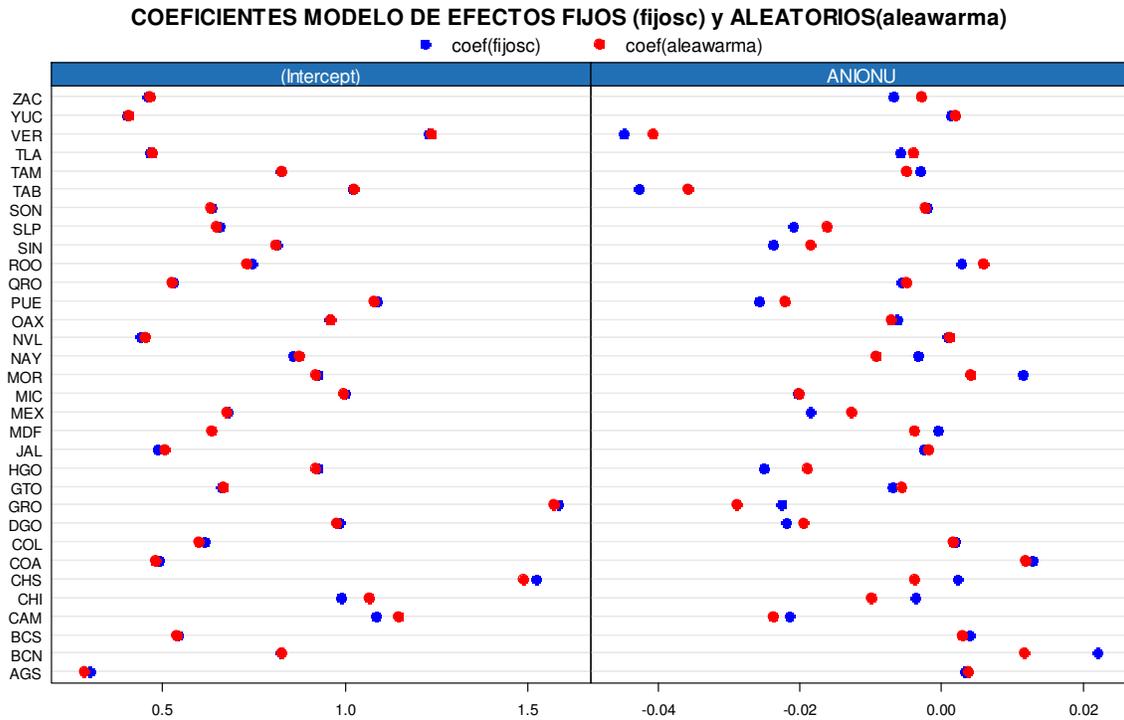


Fig. 3.34 Gráfica de coeficientes modelo de efectos fijos y aleatorios

Un cuadro resumen de los modelos se presenta a continuación:

Modelo	Nombre	G. L.	AIC	BIC	LogLik
Efectos Fijos	fijosc	65	-1273.0	-1006.2	701.5
Interceptos Aleatorios	ialea	4	-798.8	-782.3	403.4
Modelo Mixto	aleau	6	-1044.9	-1020.2	528.4
Modelo Mixto y Función de Varianza	aleaw	38	-1135.0	-979.0	605.5
Modelo Mixto y Estructura de Correlación	aleawarma	39	-1241.7	-1081.6	659.8

Cuadro 3.1 Resumen de los modelos

En forma general el modelo lineal mixto encontrado y expresado en su forma matricial sería:

$$\begin{bmatrix} y_{(1)1} \\ \cdot \\ \cdot \\ y_{(1)14} \\ y_{(2)1} \\ \cdot \\ \cdot \\ y_{(2)14} \\ \vdots \\ y_{(32)1} \\ \cdot \\ \cdot \\ y_{(32)14} \end{bmatrix} = \begin{bmatrix} 1 & x_{(1)1} \\ 1 & \cdot \\ 1 & \cdot \\ 1 & x_{(1)14} \\ 1 & x_{(2)1} \\ \vdots & \cdot \\ 1 & x_{(2)14} \\ 1 & \vdots \\ 1 & x_{(32)1} \\ 1 & \cdot \\ 1 & \cdot \\ 1 & x_{(32)14} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} + \begin{bmatrix} 1 & z_{(1)1} & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdots & \cdot & \cdot \\ 1 & z_{(1)14} & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 1 & z_{(2)1} & \cdots & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdots & \cdot & \cdot \\ 0 & 0 & 1 & z_{(2)14} & \cdots & \cdots & 0 & 0 \\ \cdot & \cdot & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \ddots & 0 & 1 & z_{(32)1} \\ \cdot & \cdot & \cdot & \cdot & \ddots & \ddots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ddots & \ddots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & z_{(32)14} \end{bmatrix} \begin{bmatrix} \hat{b}_{(1)1} \\ \hat{b}_{(1)2} \\ \hat{b}_{(2)1} \\ \hat{b}_{(2)2} \\ \cdot \\ \cdot \\ \cdot \\ \hat{b}_{(32)1} \\ \hat{b}_{(32)2} \end{bmatrix} + \begin{bmatrix} e_{(1)1} \\ \cdot \\ \cdot \\ e_{(1)14} \\ e_{(2)1} \\ \cdot \\ \cdot \\ e_{(2)14} \\ \vdots \\ e_{(32)1} \\ \cdot \\ \cdot \\ e_{(32)14} \end{bmatrix}$$

El comando *extract.lmeDesign* de R ofrece información detallada de:

Vr: Matriz de covarianzas estimada de los efectos aleatorios dividida por la varianza estimada de los residuales.

X: La matriz modelo o de diseño de los efectos fijos

Z: La matriz modelo o de diseño de los efectos aleatorios

Sigmasq: La varianza de los residuales

Lambda: La proporción de las varianzas de los efectos aleatorios y la varianza de los residuales

Y: La variable respuesta

La gráfica de los puntos observados (color verde) y los modelos lineales estimados (color rojo solido nivel nacional y color negro punteado nivel de grupo) por el modelo mixto *aleawarma* se muestra a continuación:

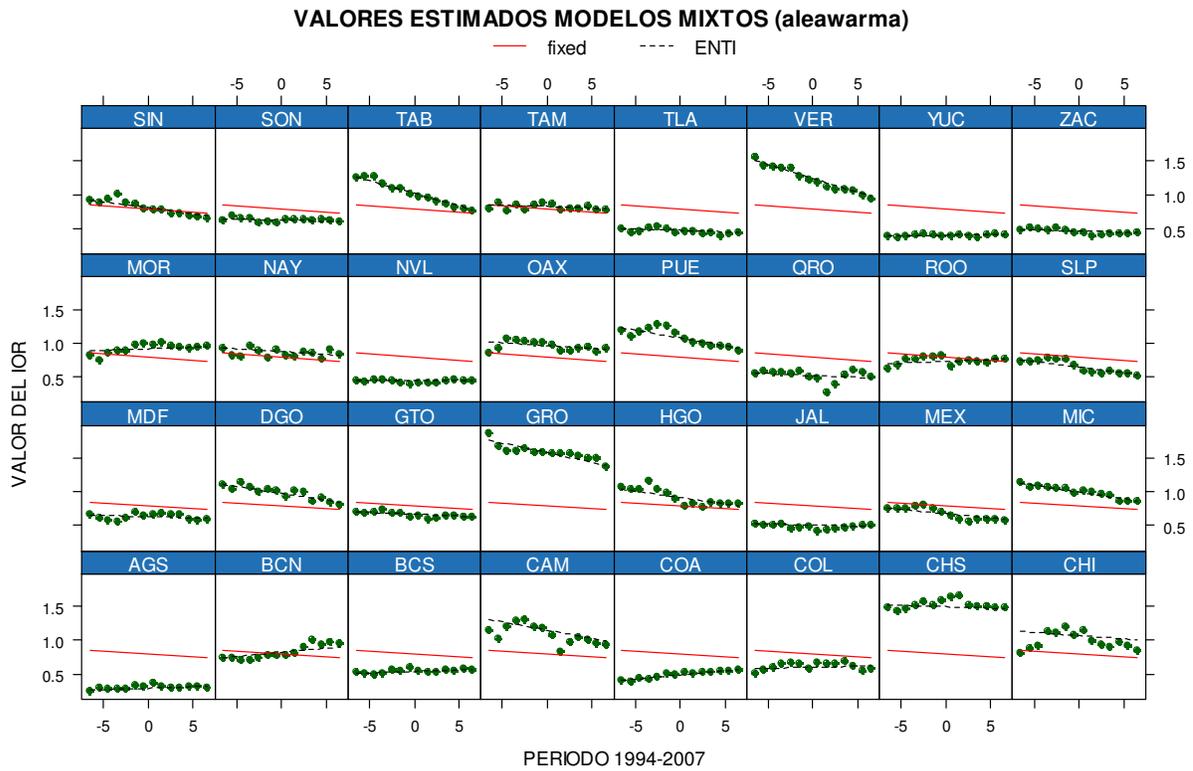


Fig. 3.35 Gráfica de panel valores estimados modelo de efectos mixtos aleawarma

Considerando que la línea solida de color rojo es la que representa el promedio nacional, la gráfica es muy ilustrativa:

Tabasco tiende a la baja y se ubica cerca de nivel nacional.

Veracruz a un ritmo más pronunciado a la baja que Tabasco de igual forma se ubica en el año 2007 cerca del promedio nacional.

Se observa el caso del estado de Coahuila con tendencia creciente y acercándose al promedio nacional.

En sentidos opuestos con respecto al nivel nacional se ubican los estados de Aguascalientes y el estado de Chiapas.

CAPÍTULO IV

EL ANÁLISIS DEL MODELO Y LA PROYECCIÓN DEL IOR

El análisis y proyección del IOR permitirá conocer si el esfuerzo de las entidades federativas refleja una tendencia a la baja o da señales de crecimiento en la oportunidad del registro de los nacimientos. Si se observa que la tendencia de este fenómeno social en el tiempo no ha tenido cambios bruscos entonces es posible obtener un modelo lineal mixto que ajuste los datos de una forma razonable.

4.1 El análisis del modelo

Los efectos fijos y sus errores estándar se obtienen con el comando *summary*, se observa que ambos coeficientes son estadísticamente diferentes del cero:

$$\hat{\beta}$$

Los efectos fijos son:

```
Fixed effects: INDI ~ ANIONU
              Value Std.Error DF t-value p-value
(Intercept)  0.7945836 0.05431690 415 14.62866  0.0000
ANIONU       -0.0085724 0.00268663 415 -3.19078  0.0015
```

Y sus intervalos al 95% de confianza:

```
              lower      est.      upper
(Intercept)  0.68805162  0.794583587  0.901115551
ANIONU       -0.01384175 -0.008572447 -0.003303148
```

Los efectos aleatorios y sus errores estándar se obtienen con el comando *intervals*, se observa que su predicción es estadísticamente diferente de cero:

Los efectos aleatorios son:

```
Random effects:
Formula: INDI ~ ANIONU | ENTI
              StdDev   Corr
(Intercept)  0.30503459 (Intr)
ANIONU       0.01383084 -0.584
```

Y sus intervalos de confianza:

```
              lower      est.      upper
sd((Intercept))  0.23808846  0.30503459  0.39080477
sd(ANIONU)       0.01017238  0.01383084  0.01880506
```

Con la información del cuadro anterior se forma la matriz de covarianzas de los efectos aleatorios o bien con la instrucción $var(ranef(aleawarma))$, en ella se observa que la varianza más grande corresponde a los interceptos $(0.3050)^2$:

$$\hat{\psi} = \begin{pmatrix} (0.3050)^2 & (-0.584)(0.3050)(0.0138) \\ (-0.584)(0.3050)(0.0138) & (0.0138)^2 \end{pmatrix}$$

La varianza de los errores se presenta a continuación:

$$\hat{\sigma}^2 = (0.06574)^2$$

Y su intervalo de confianza:

lower	est.	upper
0.05264827 ²	0.06574080 ²	0.08208916 ²

En cuanto a la estructura de correlación, la estimación del coeficiente del modelo autoregresivo ARMA(1,0) con su respectivo intervalo de confianza es:

$$\hat{\phi}$$

Correlation structure:			
	lower	est.	upper
Phi	0.5154411	0.6748563	0.7891235

Con la instrucción $ranef$ se obtiene la predicción de los efectos aleatorios para cada una de las entidades mismos que se presentan en el anexo A6. A manera de ilustración se presentan los valores predichos para algunos estados de la república:

$$\hat{b}_i = \left(\hat{b}_{1,1} = -0.50780, \hat{b}_{2,1} = 0.012323 \right), \quad \hat{b}_{i=15=MEX} = \left(\hat{b}_{1,15} = -0.11917, \hat{b}_{2,15} = -0.00426 \right), \quad \hat{b}_{i=32=ZAC} = \left(\hat{b}_{1,32} = -0.32849, \hat{b}_{2,32} = 0.00570 \right)$$

Las siguientes gráficas muestran la distribución de los valores predichos para interceptos y pendientes aleatorios, en ellas se aprecia su comportamiento normal.

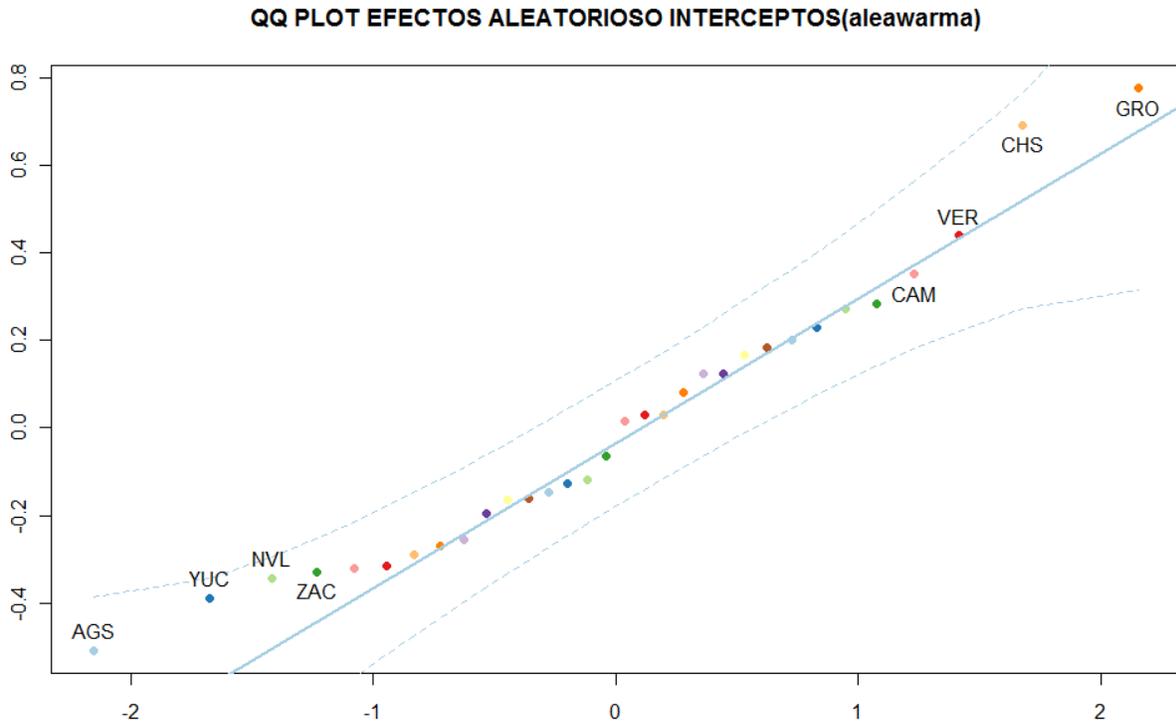


Fig. 4.1 Gráfica de ordenadas al origen efectos aleatorios

En ambas gráficas se señalan algunos valores alejados de la media.

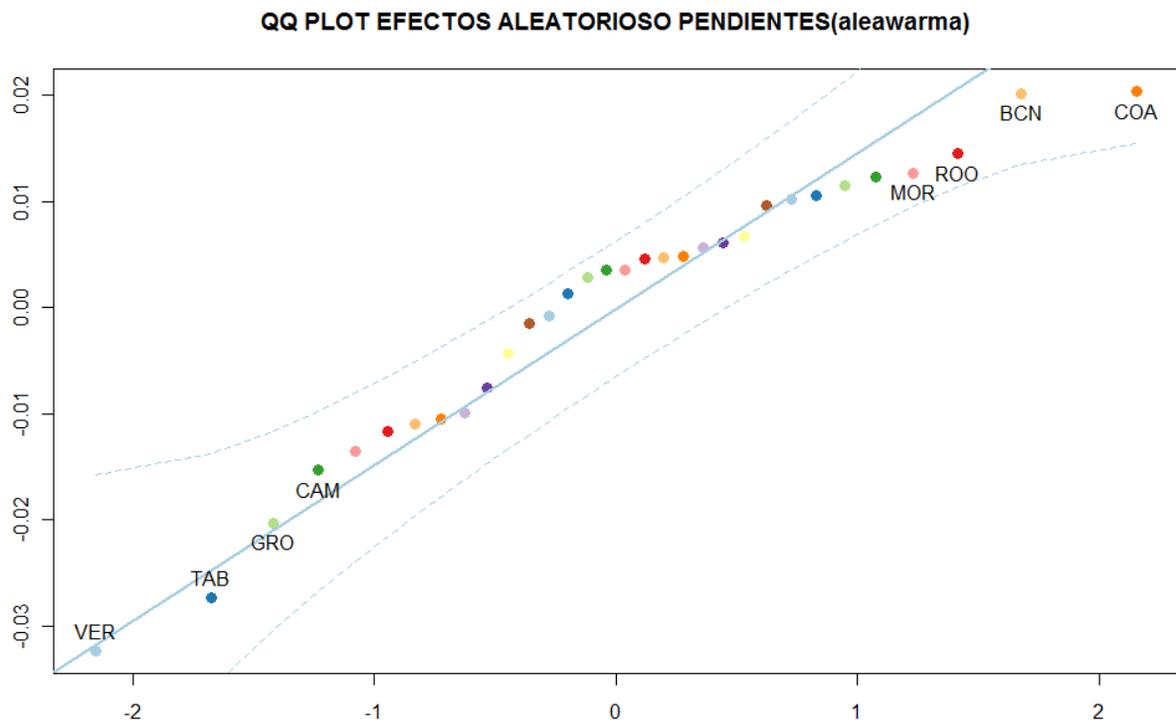


Fig. 4.2 Gráfica de pendientes efectos aleatorios

R no cuenta con alguna instrucción directa para obtener las varianzas de los efectos aleatorios predichos por entidad federativa, por esta razón se utiliza la versión 8 de SAS para consultar su bibliografía de cómo obtener los intervalos de confianza de los valores predichos de los efectos aleatorios.

SAS obtiene la varianza de los efectos aleatorios por grupo de la siguiente manera:

$$\begin{aligned} \text{Var}(b_i(\theta)) &= \psi - \left\{ \psi Z_i^T \left[V_i^{-1} - V_i^{-1} X_i \left(\sum_{i=1}^M X_i^T V_i^{-1} X_i \right)^{-1} X_i^T (V_i^{-1})^T \right] Z_i \psi \right\} \\ &= \psi - \psi (Z_i^T V_i^{-1} Z_i) \psi + \psi (Z_i^T V_i^{-1} X_i) \left(\sum_{i=1}^M X_i^T V_i^{-1} X_i \right)^{-1} (X_i^T (V_i^{-1})^T Z_i) \psi \end{aligned}$$

Donde:

$$V_i = Z_i \psi Z_i^T + \sigma^2 \Lambda_i$$

Estas varianzas se pueden consultar en el anexo A6.

Por otra parte si se suma los efectos fijos con los efectos aleatorios se puede obtener el modelo lineal estimado para cada entidad federativa. Así, por ejemplo, para Aguascalientes se tendría con el comando:

fixef (aleawarma) [1:2]+ ranef (aleawarma) [1:1,1:2]:

$$\hat{y}_{AGS} = \hat{\alpha}_{AGS} + \hat{\beta}_{AGS}$$

	(Intercept)	ANIONU
AGS	0.2867801	0.003750963

Con la ayuda de la instrucción coef (aleawarma), se obtiene las estimaciones de los interceptos y pendientes de las líneas de regresión de todas las entidades de la república, estos valores se muestran en el anexo A7.

La siguiente gráfica muestra los valores ordenados de menor a mayor de los interceptos de las líneas de regresión para todos los estados, se observa que tienen en principio un valor bajo para su IOR como es el estado de Aguascalientes y Yucatán y otros en sentido contrario como son los estados de Chiapas y Guerrero (la línea azul es a nivel nacional).

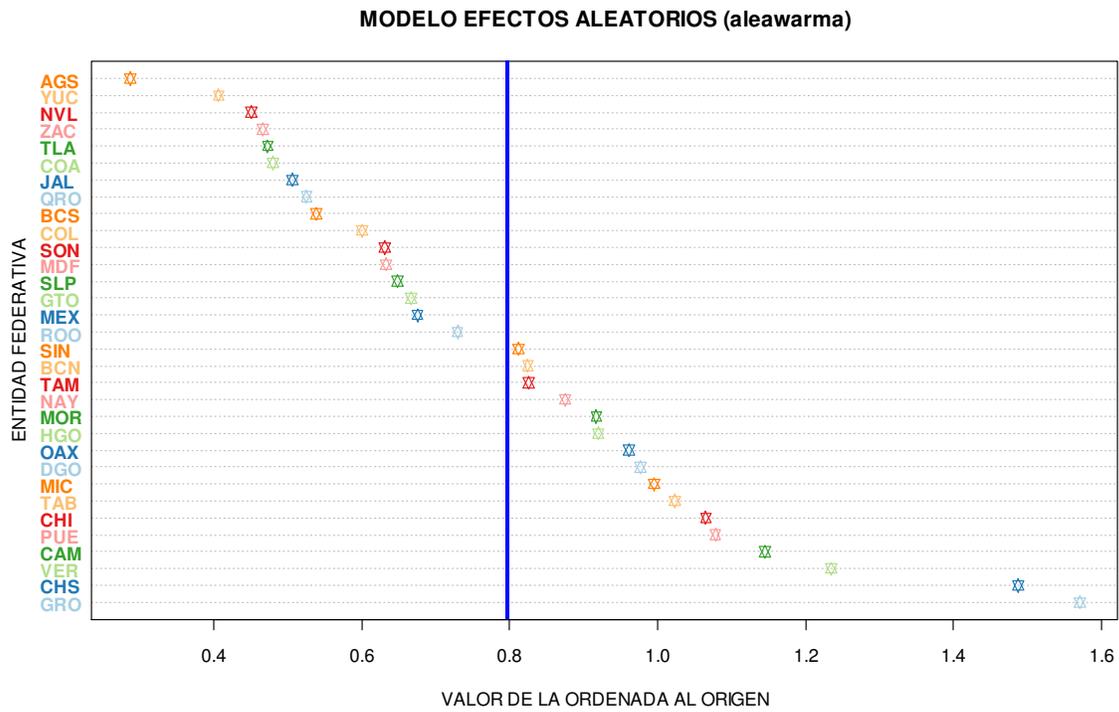


Fig. 4.3 Gráfica de ordenadas al origen modelo mixto aleawarma

La gráfica de pendientes ordenada de menor a mayor se muestra a continuación (la línea azul es a nivel nacional y la línea roja es cero).

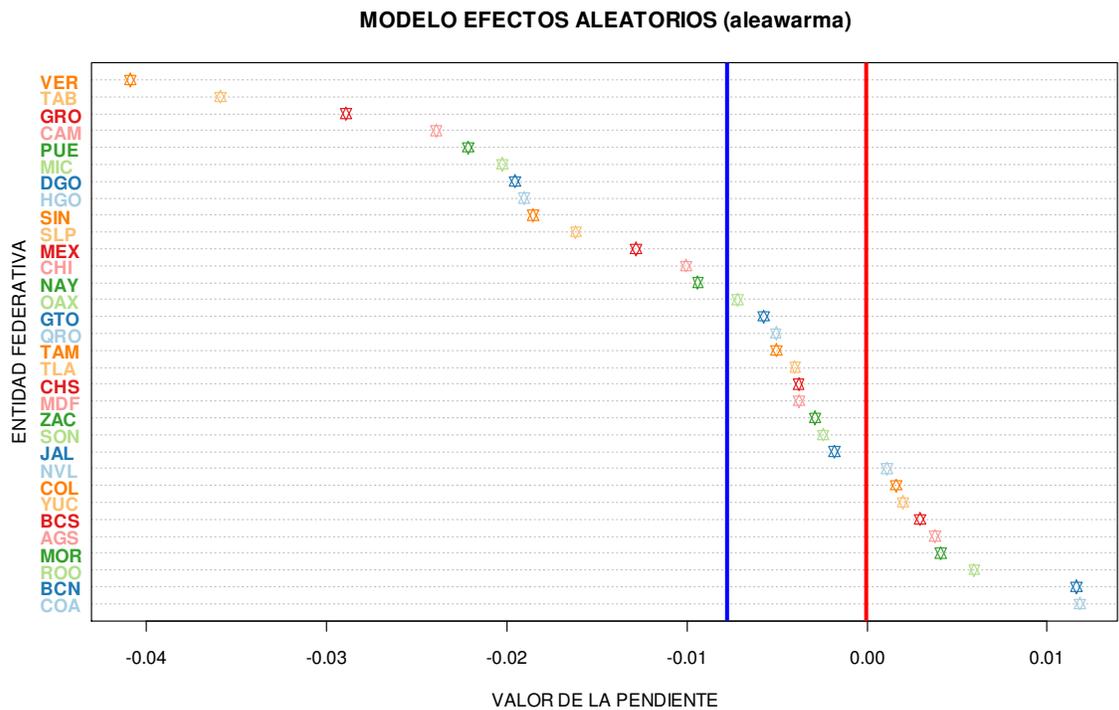


Fig. 4.4 Gráfica de pendientes modelo mixto aleawarma

En la gráfica anterior se observa que la mayoría de las entidades tienen pendientes negativas, lo que significa que se están realizando acciones diversas para mejorar el registro de nacimiento: Veracruz, Tabasco y Guerrero son los tres estados más sobresalientes en este sentido.

Analizando los interceptos y las pendientes en forma conjunta es importante señalar que si bien los estados de Chiapas y Guerrero tienen las ordenadas más altas de toda la república, no se puede negar que en los últimos catorce años el estado de Guerrero ha superado con creces al estado de Chiapas en el mejoramiento de su IOR.

Referente a la función de varianza, la estimación de los parámetros $\hat{\delta}_i$ así como sus intervalos de confianza para cada una de las entidades se encuentran en el anexo A8. A manera de ilustración se presentan los valores estimados para algunos estados de la república:

$$\hat{\delta}_i$$

$$\hat{\delta}_{i=1=AGS} = -0.3202, \quad \hat{\delta}_{i=15=MEX} = -0.5843, \quad \hat{\delta}_{i=32=ZAC} = -0.3514$$

Los estados con comportamiento de función de varianza estadísticamente diferente de cero son aquellos que no incluyen al cero en su intervalo de confianza (ver anexo A8), para el caso de que el intervalo de confianza contenga al cero, se deberá interpretar que su función de varianza no se comporta como un modelo de potencia (varPower).

Los intervalos de confianza resultantes de los componentes de la varianza y de los efectos fijos se basan en una aproximación normal de la distribución de los estimadores de máxima verosimilitud.

Sea df_i los grados de libertad del denominador de la prueba t condicional, correspondiente al i -ésimo efecto fijo basado en el valor estimado de $\hat{\beta}_i$, un intervalo de confianza al nivel $1-\alpha$ para $\hat{\beta}_i$ es:

$$\hat{\beta}_i \pm t_{df_i}(1-\alpha/2) \hat{\sigma}_{ML} \sqrt{[R_{00}^{-1} R_{00}^T]_{ii}}$$

Donde la matriz R_{00} es calculada al momento de estimar el valor de θ .

Una expresión para los intervalos de confianza al nivel $1-\alpha$ para σ es (Ver detalles en el libro *Mixed Effects Models in S and S-PLUS* de José C. Pinheiro y Douglas M. Bates):

$$\left[\hat{\sigma} \exp\left(-z(1-\alpha/2) \sqrt{[t^{-1}]_{\sigma\sigma}}\right), \hat{\sigma} \exp\left(z(1-\alpha/2) \sqrt{[t^{-1}]_{\sigma\sigma}}\right) \right]$$

De acuerdo al modelo descrito por (2.5.2) la matriz de covarianzas de los errores $e_{2(i)}$ para cada entidad federativa sigue un comportamiento de autocorrelación autoregresivo de orden uno que se puede modelar bajo el siguiente proceso:

$$\text{Tiempo 1: } e_{2(i)1} = a_{i1}$$

$$\text{Tiempo 2: } e_{2(i)2} = \phi e_{2(i)1} + a_{i2} = \phi a_{i1} + a_{i2}$$

$$\text{Tiempo 3: } e_{2(i)3} = \phi e_{2(i)2} + a_{i3} = \phi(\phi a_{i1} + a_{i2}) + a_{i3}$$

$$\text{Tiempo 4: } e_{2(i)4} = \phi e_{2(i)3} + a_{i4} = \phi[\phi(\phi a_{i1} + a_{i2}) + a_{i3}] + a_{i4}$$

⋮

$$\text{Tiempo } t: e_{2(i)t} = \phi e_{2(i)(t-1)} + a_{it} \text{ donde } a_{it} \sim N(0, \gamma_a^2)$$

Por lo anterior se tiene:

$$e_{2(i)t} = a_{it} + \phi a_{i(t-1)} + \phi^2 a_{i(t-2)} + \dots + \phi^{k-1} a_{i(t-k)} + \phi^k e_{2(i)(t-k)} \text{ para } k > 0$$

De esta forma se tiene:

$$\begin{aligned} \text{Var}(e_{2(i)t}) &= \gamma_a^2 + \phi^2 \gamma_a^2 + \phi^{2(2)} \gamma_a^2 + \dots + \phi^{2(k-1)} \gamma_a^2 + \phi^{2k} \text{Var}(e_{2(i)(t-k)}) \\ &= [\gamma_a^2] [1 + \phi^2 + \phi^4 + \dots + \phi^{2(k-1)}] + \phi^{2k} \text{Var}(e_{2(i)(t-k)}) \end{aligned}$$

Por ser $|\phi| < 1$ y t con tendencia a menos infinito y k con tendencia a infinito se tiene:

$$\text{Var}(e_{2(i)t}) = [\gamma_a^2] \left[\frac{1}{1 - \phi^2} \right] = \frac{\gamma_a^2}{1 - \phi^2}$$

Cumpliendo la igualdad al tiempo t :

$$\begin{aligned} \text{Var}(e_{2(i)t}) &= \text{Var}(\phi e_{2(i)(t-1)} + a_{it}) \\ &= \phi^2 \text{Var}(e_{2(i)(t-1)}) + \gamma_a^2 \\ &= \phi^2 \frac{\gamma_a^2}{1 - \phi^2} + \gamma_a^2 = \frac{\gamma_a^2}{1 - \phi^2} = \gamma_{e_{2(i)t}}^2 \end{aligned}$$

Por otra parte:

$$\text{Cov}(e_{2(i)t}, e_{2(i)(t-k)}) = E \left[\left(e_{2(i)t} - \mu_{2(i)t} \right) \left(e_{2(i)(t-k)} - \mu_{2(i)(t-k)} \right) \right]$$

$$\begin{aligned}
 &= E\left[\left(e_{2(i)t}\right)\left(e_{2(i)(t-k)}\right)\right] \\
 &= E\left[\left(a_{it} + \phi a_{i(t-1)} + \phi^2 a_{i(t-2)} + \dots + \phi^{k-1} a_{i(t-(k-1))} + \phi^k e_{2(i)(t-k)}\right)\left(e_{2(i)(t-k)}\right)\right] \\
 &= E\left[\left(\phi^k e_{2(i)(t-k)}\right)\left(e_{2(i)(t-k)}\right)\right] = \phi^k \left[E\left[\left(e_{2(i)(t-k)}\right)^2\right]\right] \\
 &= \phi^k \left[E\left[\left(e_{2(i)(t-k)}\right)^2\right] - E\left[\left(e_{2(i)(t-k)}\right)\right]^2\right] = \phi^k \left[Var\left(e_{2(i)(t-k)}\right)\right] \\
 &= \phi^k \frac{\gamma_a^2}{1-\phi^2} \text{ para } k > 0
 \end{aligned}$$

De esta forma las covarianzas se forman de la siguiente manera:

$$\begin{aligned}
 Cov(e_{2(i)t}, e_{2(i)(t-1)}) &= \phi \frac{\gamma_a^2}{1-\phi^2} = \phi \gamma_{e_{2(i)t}}^2 \\
 Cov(e_{2(i)t}, e_{2(i)(t-2)}) &= \phi^2 \frac{\gamma_a^2}{1-\phi^2} = \phi^2 \gamma_{e_{2(i)t}}^2 \\
 &\vdots \\
 Cov(e_{2(i)t}, e_{2(i)(t-k)}) &= \phi^{t-k} \frac{\gamma_a^2}{1-\phi^2} = \phi^{t-k} \gamma_{e_{2(i)t}}^2
 \end{aligned}$$

Teniendo la igualdad:

$$Corr(e_{2(i)t}, e_{2(i)(t-k)}) = \frac{Cov(e_{2(i)t}, e_{2(i)(t-k)})}{\sqrt{var(e_{2(i)t}) var(e_{2(i)(t-k)})}}$$

Las correlaciones se forman de la siguiente manera:

$$\begin{aligned}
 Corr(e_{2(i)t}, e_{2(i)(t-1)}) &= \frac{\phi \frac{\gamma_a^2}{1-\phi^2}}{\sqrt{\frac{\gamma_a^2}{1-\phi^2} \frac{\gamma_a^2}{1-\phi^2}}} = \phi \\
 Corr(e_{2(i)t}, e_{2(i)(t-2)}) &= \frac{\phi^2 \frac{\gamma_a^2}{1-\phi^2}}{\sqrt{\frac{\gamma_a^2}{1-\phi^2} \frac{\gamma_a^2}{1-\phi^2}}} = \phi^2
 \end{aligned}$$

⋮

$$\text{Corr}(e_{2(i)t}, e_{2(i)(t-k)}) = \frac{\phi^{t-k} \frac{\gamma_a^2}{1-\phi^2}}{\sqrt{\frac{\gamma_a^2}{1-\phi^2} \frac{\gamma_a^2}{1-\phi^2}}} = \phi^{t-k}$$

La matriz de varianzas y covarianzas para cada uno de los grupos tomando a f_i de la forma $\phi^{|t-k|}$ se establece como:

$$\gamma_{e_{2(i)}}^2 \Omega_i$$

$$\text{Var}(e_{2(i)}) = \gamma_{e_{2(i)}}^2 \Omega_i = \gamma_{e_{2(i)}}^2 \begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{13} \\ \phi & 1 & \phi & \dots & \phi^{12} \\ \phi^2 & \phi & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 & \phi \\ \phi^{13} & \phi^{12} & \dots & \phi & 1 \end{pmatrix} = \frac{\gamma_a^2}{1-\phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{13} \\ \phi & 1 & \phi & \dots & \phi^{12} \\ \phi^2 & \phi & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 & \phi \\ \phi^{13} & \phi^{12} & \dots & \phi & 1 \end{pmatrix}$$

Obtenida la matriz de varianzas de los errores ahora se puede estimar la matriz de varianzas y covarianzas de la variación de los datos longitudinales:

$$\text{Var}(y_i) = \underbrace{Z_i \psi Z_i^T}_{\text{Efecto Aleatorios}} + \underbrace{\gamma_{e_{2(i)}}^2 \Omega_i}_{\text{Correlacion Serial}} + \underbrace{\gamma_{e_{1(i)}}^2 I}_{\text{Error Clasico}}$$

Donde γ_a^2 tiene varianzas de la forma $(1-\phi^2)\sigma^2\phi^2$ y $\gamma_{e_{1(i)}}^2$ tiene varianzas de la forma $(1-\phi^2)\sigma^2$.

Con los siguientes comandos de R se obtienen las diferentes matrices de varianzas y covarianzas de la variable respuesta para cada grupo en particular:

Para el modelo condicional:

$$y_i / b_i \sim N(X_i \beta + Z_i b_i, \gamma_{e_{2(i)}}^2 \Omega_i + \gamma_{e_{1(i)}}^2 I)$$

```
matd1<-getVarCov(aleawarma, individual="AGS", type="conditional")
```

Para el modelo marginal:

$$y_i \sim N(X_i \beta, Z_i \psi Z_i^T + \gamma_{e_{2(i)}}^2 \Omega_i + \gamma_{e_{1(i)}}^2 I)$$

```
matd2<-getVarCov(aleawarma, individual="AGS", type="marginal")
```

4.2 Los pronósticos del modelo

El ajuste del modelo a los datos del IOR se calcula utilizando información previa de catorce años y se pronostica un máximo de tres años para un futuro teórico.

En los modelos mixtos los pronósticos y los valores predichos pueden obtenerse para diferentes niveles de jerarquía o bien a nivel poblacional. A nivel poblacional los pronósticos estiman el valor esperado marginal de la variable respuesta. Por ejemplo sea x_h un vector de covariables de efectos fijos, el valor esperado marginal para la correspondiente respuesta y_h es:

$$E(y_h) = x_h^T \beta$$

Para un nivel de jerarquía k los pronósticos se obtienen de la esperanza condicional de la respuesta esperada dados los efectos aleatorios de niveles menores o iguales a k . Por ejemplo, sea el vector $z_h(i)$ de covariables de efectos aleatorios asociados con el i -ésimo grupo a un nivel jerárquico, el *nivel-1* del pronóstico estimado es:

$$E(y_h(i) | b_i) = x_h^T \beta + z_h(i)^T b_i$$

Para una jerarquía posterior sea el vector $z_h(i,j)$ de covariables de efectos aleatorios asociados con el i -ésimo grupo a un nivel jerárquico, el *nivel-2* del pronóstico estimado es:

$$E(y_h(i) | b_i, b_{ij}) = x_h^T \beta + z_h(i)^T b_i + z_h(i,j)^T b_{ij}$$

De manera similar se opera para obtener los pronósticos y estimaciones del resto de los niveles jerárquicos.

Por otra parte los “*Best Linear Unbiased Predictors*” se obtienen utilizando en las expresiones anteriores los correspondientes estimadores (en la práctica, el vector de parámetros θ es reemplazado por los estimadores máximos verosímiles):

$$\hat{y}_h(i) = x_h^T \hat{\beta}(\theta) + z_h(i)^T \hat{b}_i(\theta)$$

Utilizando el comando *predict* de R se obtienen las cifras del pronóstico las cuales se pueden observar en el anexo A9.

Con un proceso AR(1) los pronósticos para una respuesta y_h^* tienen la forma:

$$E(y_h^*(i) | b_i) = x_h^T \beta + z_h^*(i)^T b_i + E(e_{(2)h^*} | y_h) = x_h^T \beta + z_h^*(i)^T b_i + \phi^{|(h^*)-t(h)|} [y_h - x_h^T \beta - z_h(i)^T b_i]$$

La gráfica de los últimos 14 años del IOR marcada en color verde así como los tres años de pronóstico (2008,2009 y 2010) marcados con color rojo muestra el efecto de la estructura de correlación en el modelo lineal mixto:

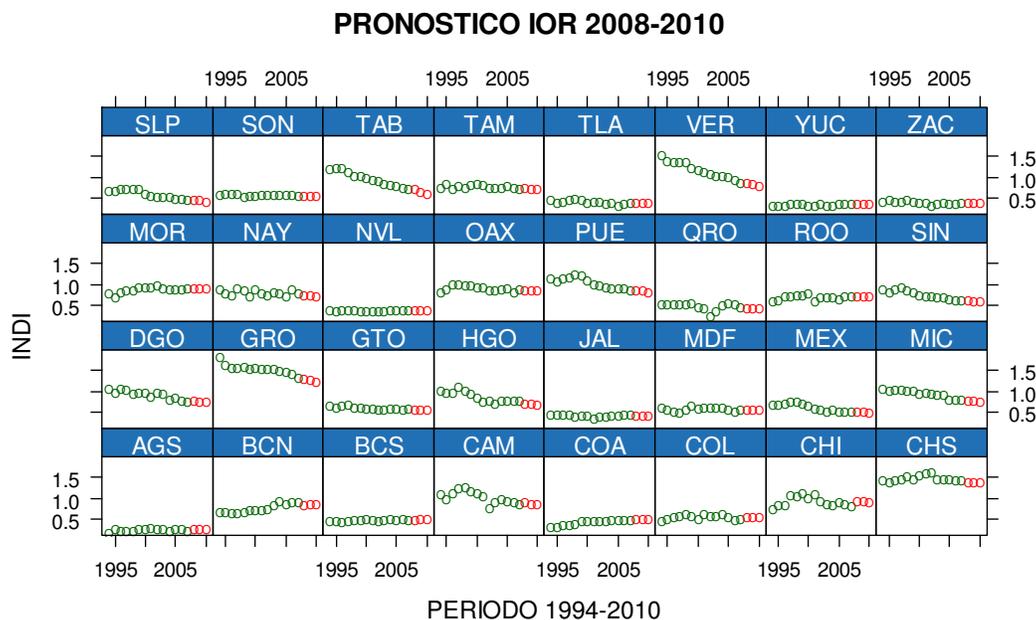


Fig. 4.5 Gráfica de pronósticos del modelo mixto aleawarma

Una gráfica de perfiles muestra la compactación de las entidades a excepción de Chiapas y Guerrero y la tendencia a la baja en el IOR de la mayoría de las entidades. El comportamiento de Veracruz es sobresaliente, de tener el segundo IOR más alto en 1994 su tendencia lo ubica en un mejor lugar para el año 2010.

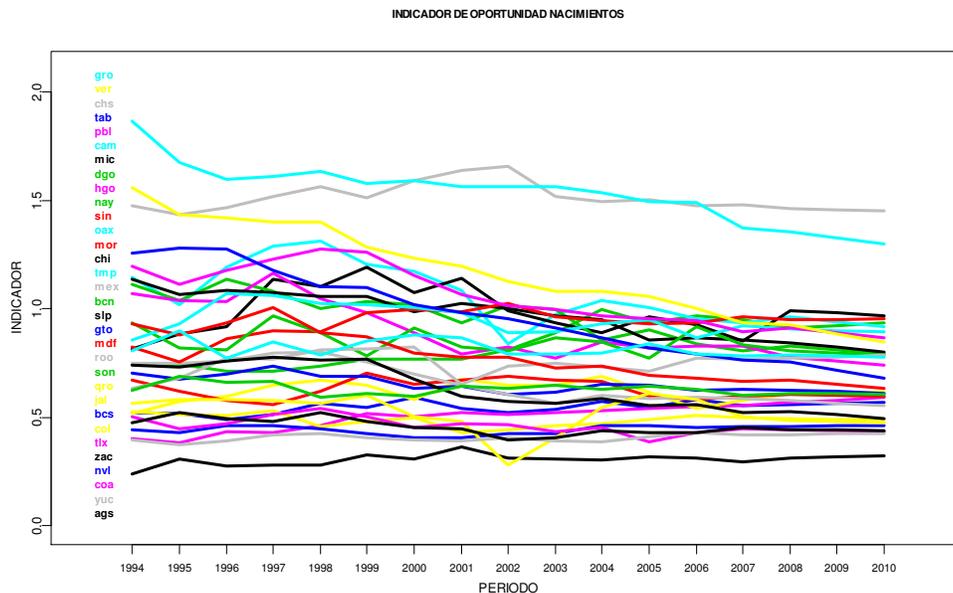


Fig. 4.6 Gráfica de perfiles de los pronósticos del modelo mixto aleawarma

Analizando los valores observados del periodo 1994-2007 y de tendencia 2008-2010 se pueden realizar el siguiente cuadro:

Periodo 1994-2007				Periodo 2008-2010			
DECREMENTO IOR		INCREMENTO IOR		INCREMENTO IOR		DECREMENTO IOR	
VER	-40.06%	COA	38.90%	<u>CHI</u>	13.75%	HGO	-11.03%
TAB	-39.18%	BCN	27.45%	AGS	9.99%	TAB	-10.77%
SLP	-29.51%	AGS	21.73%	COL	6.69%	VER	-9.46%
SIN	-28.48%	ROO	20.85%	COA	6.38%	MIC	-6.43%
DGO	-27.51%	MOR	16.60%	ROO	2.47%	<u>NAY</u>	-5.70%
GRO	-26.33%	COL	11.22%	BCS	1.77%	GRO	-5.65%
PUE	-25.11%	OAX	7.34%	YUC	1.61%	<u>MEX</u>	-5.48%
MIC	-24.39%	BCS	6.56%	NVL	0.98%	<u>SLP</u>	-4.99%
HGO	-22.51%	YUC	5.84%	<u>SON</u>	0.74%	<u>QRO</u>	-4.96%
MEX	-21.79%	CHI	4.42%			SIN	-4.68%
CAM	-18.20%	NVL	3.13%			<u>TLA</u>	-3.34%
QRO	-11.30%	CHS	0.39%			<u>OAX</u>	-3.10%
TLA	-11.07%					<u>ZAC</u>	-3.02%
MDF	-10.99%					PUE	-2.96%
NAY	-10.80%					<u>GTO</u>	-2.56%
GTO	-10.66%					<u>JAL</u>	-2.12%
ZAC	-5.35%					<u>CHS</u>	-2.06%
JAL	-4.90%					<u>DGO</u>	-1.99%
SON	-3.58%					CAM	-1.91%
TAM	-2.97%					BCN	-1.54%
						MOR	-0.76%
						<u>TAM</u>	-0.54%
						<u>MDF</u>	-0.02%
NACIONAL DECREMENTO DEL 13.10%				NACIONAL DECREMENTO DEL 3.48%			

Cuadro 4.1 Comparativo de los datos observados 1994-2007 y los datos pronosticados 2008-2010

De la tabla anterior los estados subrayados en el periodo de proyección deben de analizarse con cuidado debido a que sus pendientes predichas para el modelo lineal se consideran estadísticamente iguales a cero, para estos estados es recomendable únicamente considerar el valor de la ordenada al origen como medida de análisis.

De los estados que tienen pendiente estadísticamente diferente de cero en su parte de efectos aleatorios resaltan las entidades con tendencia a la alza del periodo de pronóstico, los estados de Aguascalientes y Colima que históricamente ha tenido el IOR con valores bajos tienen el reto de mantener este buen comportamiento, por otra parte Baja California Sur, tiene un valor bajo en su IOR, sin embargo, el problema de este estado radica en el valor positivo de su pendiente (ver gráfica de pendientes). Por otra parte, los estados de Baja California y Coahuila son los dos estados que tienen un panorama difícil debido a que son los dos estados con mayor pendiente,

adicionalmente a esta problemática se suma una más para el estado de Coahuila donde aparece con porcentaje de crecimiento en el análisis de ambos periodos.

Los estados con buenos resultados en el comportamiento de su IOR son:

DECREMENTO IOR		
1	HGO	-11.03%
2	TAB	-10.77%
3	VER	-9.46%
4	MIC	-6.43%
5	GRO	-5.65%
6	SIN	-4.68%
7	PUE	-2.96%
8	CAM	-1.91%
9	BCN	-1.54%
10	MOR	-0.76%

Cuadro 4.2 Estados con decremento en su IOR periodo 2007-2010

Se presenta a continuación los valores de la ordenada al origen de los estados con pendiente estadísticamente igual a cero en sus efectos aleatorios y un comparativo con el valor del intercepto más bajo de la república mexicana que en este caso es el estado de Aguascalientes (0.2867):

Entidad	Intercepto	Comparativo
CHS	1.4869	418.50%
CHI	1.0653	271.48%
DGO	0.9768	240.62%
OAX	0.9611	235.15%
NAY	0.7946	177.07%
TAM	0.7946	177.07%
MEX	0.6754	135.51%
GTO	0.6669	132.56%
SLP	0.6486	126.18%
MDF	0.6333	120.81%
SON	0.6307	119.92%
QRO	0.5255	83.26%
JAL	0.5062	76.52%
TLA	0.4730	64.94%
ZAC	0.4661	62.53%

Cuadro 4.3 Comparativo de interceptos año 2010.

Del cuadro anterior el estado de Chiapas es el que tiene el reto más difícil de todos los estados debido a que tienen el IOR más alto y sin tendencia de mejora.

De la tabla anterior los estados subrayados en el periodo de proyección deben de analizarse igualmente con cuidado debido a que sus pendientes y ordenadas al origen predichas por los efectos aleatorios se consideran estadísticamente iguales a cero, para estos estados es recomendable considerar el valor de la ordenada y pendiente de los efectos fijos como medida de análisis.

Los estados de Veracruz y Guerrero han realizado una labor encomiable para tener un registro más oportuno en el registro de los nacimientos.

Por otra parte es necesario comentar que a nivel nacional el IOR tiene tendencia a la baja en ambos periodos de análisis.

CAPÍTULO VI
CONCLUSIONES Y COMENTARIOS

Un punto a reflexionar acerca de los modelos lineales mixtos se refiere al coeficiente de determinación R^2 utilizado en los modelos lineales clásicos para determinar qué porcentaje de la varianza es explicado por el modelo. Para un modelo de efectos mixtos es definitiva la pregunta ¿Porcentaje de cual varianza?

Otro punto a reflexionar es la posibilidad de automatizar el cálculo de los IOR a nivel municipal y de esta forma estar en condiciones de identificar variaciones de cada entidad federativa en particular.

El método de los modelos lineales mixtos vistos en este trabajo responde satisfactoriamente a las necesidades de conocer a un plazo de tres años la tendencia del Indicador de Oportunidad del Registro de nacimientos de los estados de la República Mexicana tomando como insumo los registros administrativos de nacimientos sin necesidad de realizar ajustes a esta información, respetando lo máximo posible las cifras proporcionadas por las oficinas del registro civil, las cuales en los últimos años, ha mejorado en forma importante su administración en la mayoría de los estados.

Este trabajo de tesis utiliza en su desarrollo herramientas estadísticas suficientes para poder obtener proyecciones del IOR óptimas y coherentes que pueden ser útiles para la toma de decisiones por parte de las autoridades de los estados.

Existe el debate de que tan necesarias son las mediciones del IOR y las proyecciones del mismo utilizando los elementos estadísticos vistos en este trabajo. No obstante, aunque existan diferentes propuestas con diferentes niveles de complejidad y personas técnicamente preparadas y con una gran experiencia, no se debe ser ilusorio en creer que estos elementos otorgarán una posibilidad real de conocer en forma exacta el comportamiento de este fenómeno social.

Al tratar el futuro se debe adoptar una actitud realista ante las ventajas y limitaciones de las proyecciones del IOR, en virtud de que es una tarea siempre difícil pronosticar datos a futuro, como lo es también el hecho de realizar pronósticos de cualquier tipo, aún con las herramientas estadísticas que utilizan la probabilidad y los intervalos de confianza.

No obstante lo anterior, el método aquí propuesto puede ser de utilidad como una herramienta estadística poderosa para conocer la tendencia a nivel nacional, estatal y municipal del IOR a través del tiempo; y la creación de nuevas metodologías y análisis será el reto que enfrentarán los actuales y futuros estadísticos.