

*Modelos Econométricos Espaciales: Una Perspectiva  
Bayesiana*

JORGE ARMANDO SICACHÁ PARADA  
ESTADÍSTICO



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
NOVIEMBRE DE 2017

*Modelos Econométricos Espaciales: Una Perspectiva  
Bayesiana*

JORGE ARMANDO SICACHÁ PARADA  
ESTADÍSTICO

TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE  
MAGISTER EN ESTADÍSTICA

DIRECTOR  
EDILBERTO CEPEDA C., PH.D.



UNIVERSIDAD NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE ESTADÍSTICA  
BOGOTÁ, D.C.  
NOVIEMBRE DE 2017

---

---

# General Index

---

---

<b>General Index</b>	<b>I</b>
<b>Index of tables</b>	<b>V</b>
<b>Índice de figuras</b>	<b>VII</b>
<b>Introducción</b>	<b>IX</b>
<b>1. Econometría Espacial</b>	<b>1</b>
1.1. Contexto . . . . .	1
1.1.1. Perspectiva Histórica . . . . .	1
1.1.2. Definiciones . . . . .	2
1.1.2.1. Econometría Espacial . . . . .	2
1.1.2.2. Dependencia Espacial . . . . .	2
1.1.2.3. Heterogeneidad Espacial . . . . .	3
1.2. Modelos Econométricos Espaciales . . . . .	3
1.3. Matrices de vecindad espacial . . . . .	5
1.4. Propuestas clásicas para el ajuste de modelos econométricos espaciales . . . .	8
1.4.1. Máxima verosimilitud . . . . .	8
1.4.2. Propuestas alternativas . . . . .	10
1.4.2.1. Variables instrumentales . . . . .	10
1.4.2.2. Estimación por el método de los momentos . . . . .	10
1.4.2.3. Regresión ponderada geográficamente . . . . .	11
1.4.2.4. Aplicación de filtros espaciales . . . . .	11
1.5. Propuestas Bayesianas para el ajuste de modelos econométricos espaciales .	11
1.5.1. Ajuste Bayesiano Clásico . . . . .	12
1.5.2. Ajuste a través de algoritmos MCMC . . . . .	12

---

1.5.3. INLA (Integrated Nested Laplace Approximation) . . . . .	14
<b>2. Modelos Econométricos Espaciales Bayesianos Homocedásticos</b>	<b>16</b>
2.1. Modelo SAR Bayesiano: Homocedástico . . . . .	16
2.1.1. Especificación . . . . .	16
2.1.2. Simulación . . . . .	18
2.1.3. Aplicación: Datos de crimen en Columbus . . . . .	19
2.1.3.1. Resultados: Ajuste con Kernel de Transición Uniforme . . . . .	20
2.1.3.2. Resultados: Ajuste con Kernel de Transición Normal . . . . .	21
2.2. Modelo SEM Bayesiano: Homocedástico . . . . .	25
2.2.1. Especificación . . . . .	25
2.2.2. Simulación . . . . .	26
2.2.3. Aplicación: Datos de crimen en Columbus . . . . .	27
2.2.3.1. Resultados: Ajuste con Kernel de Transición Uniforme . . . . .	28
2.2.3.2. Resultados: Ajuste con Kernel de Transición Normal . . . . .	29
2.3. Modelo General Bayesiano: Homocedástico . . . . .	33
2.3.1. Especificación . . . . .	33
2.3.2. Simulación . . . . .	34
2.3.3. Aplicación: Datos de crimen en Columbus . . . . .	36
2.3.3.1. Resultados: Ajuste con Kernel de Transición Uniforme . . . . .	36
2.3.3.2. Resultados: Ajuste con Kernel de Transición Normal . . . . .	38
<b>3. Modelos Econométricos Espaciales Bayesianos Heterocedásticos</b>	<b>43</b>
3.1. Modelo SAR Bayesiano: Heterocedástico . . . . .	43
3.1.1. Especificación . . . . .	43
3.1.1.1. Propuesta para $\gamma$ , [22], [24] . . . . .	44
3.1.1.2. Propuesta para $\rho$ . . . . .	45
3.1.2. Simulación . . . . .	46
3.1.3. Aplicación: Datos de Leucemia en Nueva York . . . . .	47
3.2. Modelo SEM Bayesiano: Heterocedástico . . . . .	51
3.2.1. Especificación . . . . .	51
3.2.1.1. Propuesta para $\gamma$ , [22], [24] . . . . .	52
3.2.1.2. Propuesta para $\lambda$ . . . . .	53
3.2.2. Simulación . . . . .	54
3.2.3. Aplicación: Datos de Leucemia en Nueva York . . . . .	55

3.3. Modelo General Bayesiano: Heterocedástico . . . . .	59
3.3.1. Especificación . . . . .	59
3.3.1.1. Propuesta para $\gamma$ , [22], [24] . . . . .	60
3.3.1.2. Propuesta para $\rho$ . . . . .	61
3.3.1.3. Propuesta para $\lambda$ . . . . .	61
3.3.2. Simulación . . . . .	62
3.3.3. Aplicación: Datos de Leucemia en Nueva York . . . . .	63
<b>4. Discusión y Conclusiones . . . . .</b>	<b>69</b>
4.1. Discusión . . . . .	69
4.1.1. Modelo SAR homocedástico . . . . .	69
4.1.2. Modelo SEM homocedástico . . . . .	69
4.1.3. Modelo SAR heterocedástico . . . . .	71
4.1.4. Modelo SEM heterocedástico . . . . .	71
4.2. Conclusiones y Extensiones . . . . .	71
<b>A. Demostraciones . . . . .</b>	<b>74</b>
A.1. Modelos homocedásticos . . . . .	74
A.1.1. Modelo general homocedástico . . . . .	74
A.1.1.1. Distribución condicional a posteriori de $\beta$ . . . . .	74
A.1.1.2. Distribución condicional a posteriori para $\sigma^2$ . . . . .	75
A.1.1.3. Distribución condicional a posteriori para $\rho$ . . . . .	75
A.1.1.4. Distribución condicional a posteriori para $\lambda$ . . . . .	76
A.1.2. Modelo SAR homocedástico . . . . .	77
A.1.2.1. Distribución condicional a posteriori de $\beta$ . . . . .	77
A.1.2.2. Distribución condicional a posteriori de $\sigma^2$ . . . . .	77
A.1.2.3. Distribución condicional a posteriori de $\rho$ . . . . .	77
A.1.3. Modelo SEM homocedástico . . . . .	77
A.1.3.1. Distribución condicional a posteriori de $\beta$ . . . . .	77
A.1.3.2. Distribución condicional a posteriori de $\sigma^2$ . . . . .	78
A.1.3.3. Distribución condicional a posteriori de $\lambda$ . . . . .	78
A.2. Modelos heterocedásticos . . . . .	78
A.2.1. Modelo general heterocedástico . . . . .	78
A.2.1.1. Distribución condicional a posteriori de $\beta$ . . . . .	78
A.2.1.2. Distribución condicional a posteriori de $\gamma$ . . . . .	79

---

A.2.1.3. Distribución condicional a posteriori para $\rho$ . . . . .	80
A.2.1.4. Distribución condicional a posteriori para $\lambda$ . . . . .	80
A.2.2. Modelo SAR heterocedástico . . . . .	81
A.2.2.1. Distribución condicional a posteriori de $\beta$ . . . . .	81
A.2.2.2. Distribución condicional a posteriori de $\gamma$ . . . . .	81
A.2.2.3. Distribución condicional a posteriori de $\rho$ . . . . .	81
A.2.3. Modelo SEM heterocedástico . . . . .	82
A.2.3.1. Distribución condicional a posteriori de $\beta$ . . . . .	82
A.2.3.2. Distribución condicional a posteriori de $\gamma$ . . . . .	82
A.2.3.3. Distribución condicional a posteriori de $\lambda$ . . . . .	82
<b>B. Implementación Computacional</b>	<b>83</b>
B.1. Función <i>hom_sar</i> . . . . .	83
B.2. Función <i>hom_sem</i> . . . . .	83
B.3. Función <i>hom_general</i> . . . . .	84
B.4. Función <i>hetero_sar</i> . . . . .	84
B.5. Función <i>hetero_sem</i> . . . . .	84
B.6. Función <i>hetero_general</i> . . . . .	85
<b>C. Bibliografía</b>	<b>86</b>

---

---

## Index of tables

---

---

2.1. Comparación de ajustes del modelo SAR Homocedástico . . . . .	18
2.2. Parámetros estimados del modelo SAR con kernel uniforme . . . . .	20
2.3. Parámetros estimados del modelo SAR con kernel normal . . . . .	22
2.4. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SAR homocedástico bajo las dos propuestas . . . . .	24
2.5. Comparación de ajustes del modelo SAR homocedástico . . . . .	27
2.6. Parámetros estimados del modelo SEM con kernel uniforme . . . . .	28
2.7. Parámetros estimados del modelo SEM Homocedástico con kernel normal . . . . .	30
2.8. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SEM homocedástico bajo las dos propuestas . . . . .	32
2.9. Comparación de ajustes del modelo general homocedástico . . . . .	35
2.10. Parámetros estimados del modelo general con kernel uniforme . . . . .	37
2.11. Parámetros estimados del Modelo General con kernel normal . . . . .	39
2.12. Medidas resumen y test de normalidad para los residuales del ajuste del modelo general homocedástico bajo las dos propuestas . . . . .	41
3.1. Comparación de ajustes del modelo SAR Heterocedástico . . . . .	46
3.2. Parámetros estimados del modelo SAR heterocedástico . . . . .	48
3.3. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SAR heterocedástico bajo las dos propuestas . . . . .	51
3.4. Comparación de ajustes del modelo SEM heterocedástico con distintos va- lores de $\lambda$ . . . . .	54
3.5. Parámetros estimados del modelo SEM heterocedástico . . . . .	55
3.6. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SEM heterocedástico bajo las dos propuestas . . . . .	58
3.7. Comparación de ajustes del modelo SEM Heterocedástico con distintos va- lores de $\rho$ y $\lambda$ . . . . .	62
3.8. Parámetros estimados del modelo general heterocedástico . . . . .	64

---

3.9. Medidas resumen y test de normalidad para los residuales del ajuste del modelo General heterocedástico bajo las dos propuestas . . . . .	67
---	----

---

---

## Índice de figuras

---

---

1.1. Clasificación de modelos presentada en [31] . . . . .	4
1.2. Estructura de vecindad espacial bajo el criterio Torre. Tomada de [2] . . . . .	6
1.3. Estructura de vecindad espacial bajo el criterio Alfil. Tomada de [2] . . . . .	6
1.4. Estructura de vecindad espacial bajo el criterio Reina. Tomada de [2] . . . . .	6
1.5. Estructuras de vecindad de segundo orden, bajo el criterio torre y el criterio reina. Tomado de [1]. . . . .	7
2.1. Cadenas de parámetros para el modelo SAR homocedástico con kernel de transición uniforme . . . . .	21
2.2. Cadenas de parámetros para el modelo SAR homocedástico con kernel de transición normal . . . . .	23
2.3. Residuales para el modelo ajustado con el kernel de transición uniforme para los datos de crimen en Columbus, [8] . . . . .	24
2.4. Residuales para el modelo ajustado con el kernel de transición normal para los datos de crimen en Columbus, [8] . . . . .	24
2.5. Cadenas Parámetros Modelo SEM Homocedástico con Kernel de Transición Uniforme . . . . .	29
2.6. Cadenas de los parámetros del modelo SEM homocedástico con kernel de transición normal . . . . .	31
2.7. Residuales para el modelo SEM homocedástico ajustado con el kernel de transición uniforme para los datos de crimen en Columbus, [8] . . . . .	32
2.8. Residuales para el modelo SEM homocedástico ajustado con el kernel de transición normal para los datos de crimen en Columbus, [8] . . . . .	32
2.9. Cadenas de los parámetros del modelo general homocedástico con kernel de transición uniforme . . . . .	38
2.10. Cadenas de los parámetros del modelo general homocedástico con kernel de transición normal . . . . .	40
2.11. Residuales para el modelo general homocedástico ajustado con el kernel de transición uniforme para los datos de crimen en Columbus, [8] . . . . .	41

---

2.12. Residuales para el modelo general homocedástico ajustado con el kernel de transición normal para los datos de crimen en Columbus, [8] . . . . .	41
3.1. Cadenas de los parámetros del modelo SAR heterocedástico con kernel de transición uniforme . . . . .	49
3.2. Cadenas de los parámetros del modelo SAR heterocedástico con kernel de transición normal . . . . .	49
3.3. Residuales para el modelo SAR heterocedástico ajustado con el kernel de transición uniforme para los datos de Leucemia en Nueva York, [62] . . . . .	50
3.4. Residuales para el modelo SAR heterocedástico ajustado con el kernel de transición normal para los datos de Leucemia en Nueva York, [62] . . . . .	50
3.5. Cadenas de los parámetros del modelo SEM heterocedástico con kernel de transición uniforme . . . . .	56
3.6. Cadenas de los parámetros del modelo SEM heterocedástico con kernel de transición normal . . . . .	57
3.7. Residuales para el modelo SEM heterocedástico ajustado con el kernel de transición uniforme para los datos de Leucemia en Nueva York, [62] . . . . .	58
3.8. Residuales para el modelo SEM heterocedástico ajustado con el kernel de transición normal para los datos de Leucemia en Nueva York, [62] . . . . .	58
3.9. Cadenas de los parámetros del modelo general heterocedástico con kernel de transición uniforme . . . . .	65
3.10. Cadenas de los parámetros del modelo general heterocedástico con kernel de transición normal . . . . .	66
3.11. Residuales para el modelo general heterocedástico ajustado con el kernel de transición uniforme para los datos de leucemia en Nueva York, [62] . . . . .	67
3.12. Residuales para el modelo general heterocedástico ajustado con el kernel de transición normal para los datos de leucemia en Nueva York, [62] . . . . .	67
4.1. Diferencia entre estimaciones y valores verdaderos para el modelo SAR Homocedástico . . . . .	70
4.2. Diferencia entre estimaciones y valores verdaderos para el modelo SEM homocedástico . . . . .	70
4.3. Diferencia entre estimaciones y valores verdaderos para el modelo SAR heterocedástico . . . . .	71
4.4. Diferencia entre estimaciones y valores verdaderos para el modelo SEM heterocedástico . . . . .	72

---

---

## Introducción

---

---

En este trabajo se buscan dos objetivos: Proponer un conjunto de modelos econométricos espaciales Bayesianos con el fin de explicar la asociación espacial entre unidades de área cuando existe rezago espacial, tanto de la variable dependiente, como en el término de error, asumiendo homocedasticidad y heterocedasticidad. De la misma manera se realiza el ajuste de los modelos propuestos mediante una extensión de la metodología presentada en [22] y [24], la cual está basada en el algoritmo de Metropolis Hastings, [56], [59].

Para el cumplimiento de estos objetivos se realizó una amplia revisión bibliográfica, comenzando por una revisión histórica de todo el contexto alrededor de la econometría espacial, [11], pasando por la especificación de los modelos econométricos espaciales, expuesta en [8] y [42], así como de distintas maneras de realizar la matriz de vecindades presente en los modelos, finalizando con las propuestas de ajuste clásico y Bayesiano, [42],[43].

Una vez concluida la revisión bibliográfica se concretó el primer objetivo a través del planteamiento de la taxonomía de los modelos econométricos espaciales Bayesianos. Para ajustar los modelos propuestos, se realizó el cálculo de las distribuciones a posteriori condicionales de todos sus parámetros, esto con el fin de emplear procedimientos MCMC, [33], tal como el algoritmo de Metropolis-Hastings, [56], [59]. Para el caso heterocedástico, cuando habían distribuciones condicionales a posteriori que resultaban intratables analíticamente, se optó por trabajar a través de aproximaciones de Taylor alrededor de las estimaciones del parámetro, esto con el fin de realizar una aproximación por variables de trabajo, la cual llevaría a la construcción de un kernel de transición con una forma analítica conocida, [22], [24].

Este documento consta de cuatro capítulos. Para iniciar, en el capítulo 1 se hace una presentación detallada de todos los avances en el campo de la econometría espacial. En la sección 1 se exponen los esfuerzos de los padres de esta área como Anselin y Paelinck, [10], quienes a través de distintas acciones e iniciativas nombraron, definieron y posicionaron la econometría espacial, tanto en el ámbito de la econometría misma, como en las más grandes publicaciones de economía a nivel mundial. En la sección 2 se presenta el conjunto de modelos econométricos espaciales, descrita en [8], [26] y [31]. En la sección 3 se describen distintas especificaciones para la matriz de vecindad espacial que se encuentran en la literatura de los modelos econométricos espaciales, [26], [34]. De igual manera se expone la importancia de una elección apropiada, y sus implicaciones en el ajuste de los modelos. En la sección 4 se presentan las metodologías de ajuste clásico

---

para este conjunto de modelos, cerrando en la sección 5 con diversas propuestas de ajuste Bayesiano que han sido descritas en múltiples publicaciones, partiendo de la presentación hecha en [43].

En el capítulo 2 se introduce el primer grupo de modelos Bayesianos propuestos, el cual corresponde a aquellos cuyos términos de error tienen variación homogénea a través del espacio. Se presenta la especificación de estos modelos, así como la derivación de las distribuciones condicionales a posteriori de los parámetros involucrados. Posteriormente son introducidas dos propuestas de ajuste Bayesiano para trabajar con las distribuciones condicionales a posteriori que no tienen una forma conocida. Finalmente, se presentan simulaciones y aplicaciones con datos reales con el fin de evidenciar el ajuste de este grupo de modelos con la propuesta realizada.

Después, en el capítulo 3 se expone el segundo grupo de modelos Bayesianos, el cual corresponde a aquellos cuyos términos de error tienen heterogeneidad de varianza a través del espacio. Se expone la especificación de estos modelos, así como la obtención de las distribuciones condicionales a posteriori de cada uno de los parámetros. Luego, se introducen dos propuestas de ajuste Bayesiano que surgen como solución para el muestreo de las distribuciones condicionales que no tienen una forma conocida. Por último, se aplica la metodología propuesta a conjuntos de datos simulados y reales.

Para terminar, en el capítulo 4 se discuten los resultados obtenidos en los capítulos 2 y 3. Así mismo, se exponen conclusiones sobre las propuestas de ajuste realizadas.

---

---

## Econometría Espacial

---

---

### 1.1. Contexto

#### 1.1.1. Perspectiva Histórica

La creación de la econometría espacial se remonta al Encuentro Anual de la Asociación Holandesa de Estadística, realizado en la ciudad de Tilburg en el año 1974, atendiendo la necesidad de crear un campo de la econometría enfocado en brindar herramientas para el trabajo de modelos urbanos y regionales,[10],[53]. Desde ese momento, los padres de la econometría espacial se enfocaron en posicionar esta rama en el entorno tanto de la econometría como de la economía. En [11] se destacan 5 dimensiones del cambio de la econometría espacial, ellas son: publicaciones en artículos, libros, software, investigación y ofertas de trabajo.

En el campo de los artículos, la econometría espacial se abrió espacio en célebres revistas, tales como *Journal of Regional Science* o *Journal of Urban Economics*, [11]. No obstante, en la parte de los libros, los resultados no son similares, puesto que ni en la década de los 80, ni en los 90, los libros de econometría tenían los efectos espaciales como tema recurrente, solo hasta después del año 2000 con autores como Baltagi,[14], [15], la idea de econometría espacial pudo tomar fuerza en la literatura econométrica.

La tercera dimensión mencionada en [11] corresponde al crecimiento del volumen de software disponible relacionado con econometría espacial. En los años 80, la falta de software era tomado como un limitante para el desarrollo de la econometría espacial, [8]. Sin embargo, en el año 1992 con la herramienta SpaceStat, [9], fue posible hacer análisis de modelos econométricos espaciales desde un programa diseñado específicamente para esto. Posteriormente, salió la herramienta S+Spacialstat. Hasta que al aproximarse el año 2000, autores de econometría espacial, tales como James P. Le Sage, [42],[43], y Roger Bivand, [17],[18], llevaron a cabo un avance desde el punto de vista de software en econometría espacial, con la programación de scripts y paquetes, en MATLAB, [47], y R, [54], respectivamente. En las secciones 1.4 y 1.5 se mencionan algunos de los paquetes disponibles en R para el ajuste de modelos econométricos espaciales.

Las últimas dos dimensiones del desarrollo de la econometría espacial son la investigación y el ámbito laboral, las cuales están ligadas tanto con la relación de esta rama de la econometría con otras áreas del conocimiento, como con su inserción en la sociedad, vista desde una necesidad de crear posiciones laborales afines a la econometría espacial.

Mientras se lograba posicionar la econometría espacial, diversos avances metodológicos fueron dándose alrededor de ella. El primer punto en el cual fue posible avanzar fue en la especificación de los modelos, esto incluye tanto la formulación, [8], [26], [31], como la construcción de matrices de vecindad, [13], [26], [34].

Además, la econometría espacial ha tenido avances en lo que respecta a las metodologías de ajuste de los modelos. En principio, la propuesta de ajuste presentada en [5],[8] y [51], consistía en un ajuste por máxima verosimilitud o por mínimos cuadrados en dos etapas. Propuestas posteriores contemplan el ajuste de estos modelos a través de estadística Bayesiana mediante algoritmos MCMC, [42], [43]. El detalle de estas propuestas será presentado en las secciones 1.4. y 1.5., respectivamente.

### 1.1.2. Definiciones

Una vez presentado el contexto histórico y las nuevas direcciones de la econometría espacial, es pertinente concluir esta sección con conceptos importantes en el contexto de este trabajo.

#### 1.1.2.1. Econometría Espacial

La econometría espacial es definida como una rama de la econometría que trata todo lo relacionado a la dependencia y la heterogeneidad espacial para modelos de regresión, [10].

Sin embargo, esta definición no deja claro el límite con la estadística espacial, específicamente aquella que hace manejo de datos *lattice* o de área. De hecho, [7] y [38] señalan que mientras el manejo de datos espaciales está orientado hacia los datos, la econometría espacial se enfoca en el modelo. Además la econometría espacial es más abierta a otras áreas, puesto que tiene aplicaciones en biología, ecología y salud, y por su parte la estadística espacial se dirige hacia estudios urbanos y regionales, [7] y [38].

#### 1.1.2.2. Dependencia Espacial

La dependencia espacial, se encuentra definida en [8] como la relación funcional que existe entre lo que sucede en un punto del espacio, y lo que sucede en otro lugar. Este concepto va muy ligado a lo enunciado por la primera ley de la geografía, [61], "todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes".

### 1.1.2.3. Heterogeneidad Espacial

En [8] se define la heterogeneidad espacial como un conjunto de aspectos que son recurrentes en el trabajo de ciencias regionales, estos incluyen la existencia de regiones más desarrolladas que otras, desde el punto de vista económico y de crecimiento urbano, o diferencias tan naturales como las diferencias de área y población de las distintas unidades de área consideradas en estos problemas.

Este es también ampliamente aplicado en otras ciencias, como la ecología, donde se entiende la heterogeneidad espacial como la existencia de distintas concentraciones de especies a lo largo del espacio, [63].

A continuación se presentarán varios modelos econométricos espaciales expuestos en la literatura.

## 1.2. Modelos Econométricos Espaciales

Este conjunto de modelos econométricos espaciales fue presentado en [8] y [51]. Estos modelos tienen una estructura general común, dada por:

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.1)$$

$$\boldsymbol{\varepsilon} = \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \boldsymbol{\mu} \quad (1.2)$$

donde  $\mathbf{y}$  corresponde al vector de respuestas de dimensión  $N$ , con información de cada una de las unidades espaciales.  $\mathbf{X}$  es una matriz  $N \times K$ , con los valores de las covariables para cada unidad espacial.  $\mathbf{W}_1$  y  $\mathbf{W}_2$  son las matrices de vecindad espacial apropiadamente definidas, de acuerdo a los lineamientos presentados en la sección 3. Por otra parte,  $\rho$  es el coeficiente de la variable respuesta rezagada espacialmente,  $\lambda$  es el coeficiente de la estructura autoregresiva de  $\boldsymbol{\varepsilon}$ , el término de error presentado en el modelo (1.1), y  $\boldsymbol{\mu} \sim N(0, \boldsymbol{\Omega})$  es un término de error normalmente distribuido con media  $\mathbf{0}$ , un vector de ceros, y matriz de varianzas y covarianzas diagonal,  $\boldsymbol{\Omega}$ , con elementos diagonales dados por  $\Omega_{ii} = h_i(\mathbf{Z}\boldsymbol{\gamma})$ . Estos términos muestran la heterocedasticidad como función de  $\boldsymbol{\gamma}$ . Si  $\boldsymbol{\gamma} = 0$ , entonces el modelo es homocedástico.

El modelo anterior es conocido como el modelo general. A partir de este se pueden obtener otros modelos, poniendo restricciones sobre los parámetros,  $\rho$  y  $\lambda$ . Esto es, por ejemplo si  $\lambda=0$ , entonces se tiene:

$$\mathbf{y} = \rho \mathbf{W}_1 \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.3)$$

este modelo es conocido como el modelo SAR (Spatial Autoregressive Model). De la misma manera, cuando se tiene  $\rho = 0$ , se obtiene el modelo SEM (Spatial Error Model). Dado por:

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \lambda \mathbf{W}_2)^{-1} \boldsymbol{\mu} \quad (1.4)$$

Bajo la especificación del modelo general, puede asumirse  $\mathbf{W}_1 = \mathbf{W}_2$ . Sin embargo, pueden presentarse problemas de identificabilidad, [12], [57]



- SDEM (*Spatial Durbin Error Model*) obtenido a partir del modelo GNS cuando  $\rho = 0$
- SLX (*Spatial Lag of X Model*) obtenido a partir del modelo GNS cuando  $\lambda = 0$  y  $\rho = 0$

En la sección 1.4. se demuestra la obtención de la función de verosimilitud del modelo más general de la taxonomía presentada en [8]. Así mismo, se presentan varias propuestas de ajuste que han sido realizadas para este modelo, partiendo del ajuste por máxima verosimilitud, la cual es la metodología de ajuste más clásica para estos modelos, y la cual puede ser aplicada en R , [54], a través de la función *spdep*.

Otro elemento determinante para el correcto ajuste de los modelos econométricos espaciales es la especificación de la matriz de vecindad espacial, algunas ideas para la construcción de esta matriz se presentan a continuación.

### 1.3. Matrices de vecindad espacial

Una parte fundamental en la especificación de un modelo de econometría espacial es aquella relacionada con la manera en que la asociación entre unidades espaciales es explicada, y para lograrlo se debe determinar de manera apropiada la matriz de vecindad espacial. Esta matriz, que es encontrada en el modelo (1.1) como  $\mathbf{W}_1$  y  $\mathbf{W}_2$ , está definida como una matriz cuadrada cuyos elementos,  $w_{ij}$ , hacen referencia a la intensidad de la asociación entre dos unidades espaciales,  $i$  y  $j$ . Estos pesos deben ser positivos y finitos, [51], [57]. Sin embargo, una duda recurrente al momento de formular un modelo econométrico espacial corresponde a cuál es la mejor manera de especificar esta matriz. Diversas propuestas han sido realizadas para construirla.

Las primeras ideas respecto a la construcción de esta matriz fueron elaboradas en el contexto de arreglos regulares de unidades espaciales,[26], [34]. Allí la matriz es construida asignando 1 a la entrada  $[i,j]$  si la unidad  $i$  y la unidad  $j$  son vecinas y 0 en otro caso. Para definir si dos unidades espaciales en un arreglo regular son vecinas, existen diversos criterios, el primero de ellos corresponde al criterio torre, el cual determina que dos unidades espaciales son vecinas si comparten algún lado, como se muestra en la figura 1.2.

Por otra parte, cuando las unidades espaciales comparten algún vértice y se asigna 1 en la matriz de contigüidad, se está aplicando el criterio alfil, como se presenta en la figura 1.3.

Finalmente, el criterio reina aplica cuando se consideran como vecinas dos unidades espaciales si comparten algún vértice o lado, ver figura 1.4.

Ahora bien, existen diversos ordenes de contigüidad en arreglos de unidades espaciales regulares, por ejemplo, bajo la construcción de una matriz de vecindad espacial de segundo orden y el criterio torre, se consideran vecinas aquellas unidades espaciales que tengan de por medio una unidad espacial con la cual comparten un lado, de la misma manera puede ser interpretada la contigüidad de segundo orden bajo el criterio reina, ver figura 1.5.

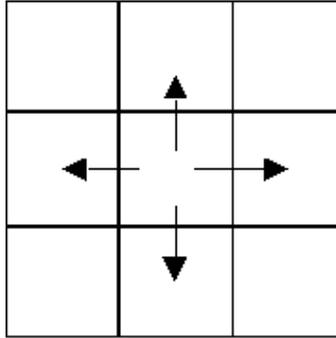


FIGURA 1.2. Estructura de vecindad espacial bajo el criterio Torre. Tomada de [2]

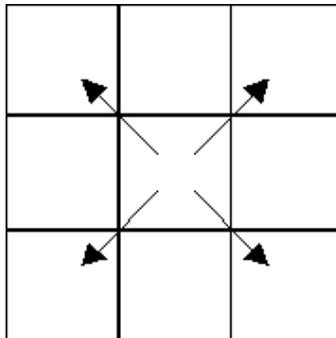


FIGURA 1.3. Estructura de vecindad espacial bajo el criterio Alfil. Tomada de [2]

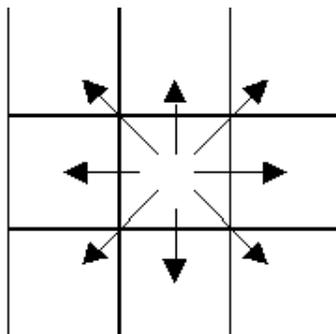


FIGURA 1.4. Estructura de vecindad espacial bajo el criterio Reina. Tomada de [2]

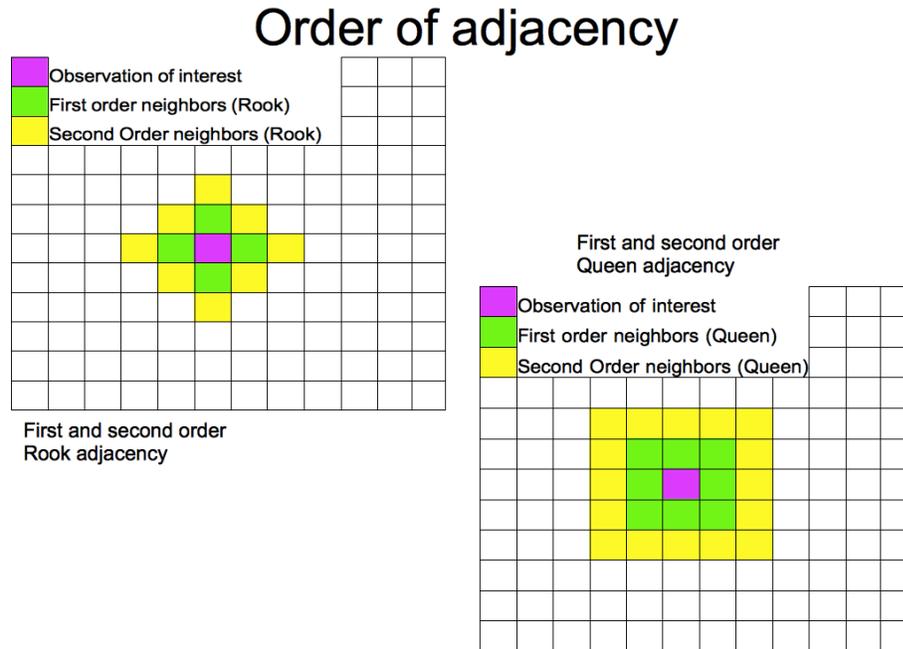


FIGURA 1.5. Estructuras de vecindad de segundo orden, bajo el criterio torre y el criterio reina. Tomado de [1]

No obstante, no todos los arreglos de unidades espaciales son regulares, de hecho, en la mayoría de las aplicaciones de datos de área, las unidades corresponden a países, estados, condados o barrios, los cuales no tienen una forma geométrica regular. Es por estas consideraciones que diversas propuestas han sido realizadas para la asignación de valores en estas matrices. La más elemental es asignar un valor de 1 en la entrada  $[i,j]$  de la matriz  $\mathbf{W}$ , si las dos unidades espaciales tienen alguna frontera o borde en común, [34], [51].

En [26] proponen una matriz de vecindades espaciales basada en la distancia entre centroides de unidades espaciales, la cual está dada por

$$W_{ij} = d_{ij}^{-a} \beta_{ij}^b$$

donde  $d_{ij}$  corresponde a la distancia entre los centroides  $i$  y  $j$ ,  $\beta_{ij}$  es la proporción de la frontera de la unidad espacial  $i$  que es ocupada por la unidad espacial  $j$  y  $a$  y  $b$  son parámetros para estimar, [5], [48], [57]. Así mismo, en [5], se propone una matriz de vecindad únicamente basada en la distancia entre centroides, la entrada  $[i,j]$  de la matriz está dada por

$$W_{ij} = \frac{1}{d_{ij}^2}$$

En busca de mayor interpretabilidad, se puede hacer que la suma de cada fila de la matriz  $\mathbf{W}$  sea 1, esto con el fin de que el término  $i$  del vector  $\mathbf{W}\mathbf{y}$  sea entendido como el promedio de los valores de la variable respuesta  $\mathbf{y}$  en aquellas áreas vecinas de la unidad  $i$ .

Otras discusiones y propuestas para la especificación de matrices de vecindad espacial pueden ser encontradas en [13], donde se presenta la importancia de realizar una

buena especificación a priori de la matriz de vecindad espacial. En [25] se propone la construcción de la matriz de vecindad espacial basada en herramientas como correlaciones parciales. Por otra parte, en [4] y [36], el enfoque está direccionado a la construcción de la matriz de vecindad espacial tomando como referencia el estadístico de autocorrelación espacial local de Getis, [37]. Otros trabajos que tienen en cuenta la matriz de vecindad espacial son [13] y [29].

## 1.4. Propuestas clásicas para el ajuste de modelos econométricos espaciales

### 1.4.1. Máxima verosimilitud

La propuesta clásica de ajuste de modelos econométricos espaciales corresponde al ajuste por máxima verosimilitud, esta ha sido presentada en [5], [8] y [51]. La idea es maximizar la función de verosimilitud derivada a continuación, como en [8] y [57], para los modelos (1.1) y (1.2). Estos modelos pueden ser reescritos como:

$$(\mathbf{I} - \rho \mathbf{W}_1) \mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.5)$$

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \lambda \mathbf{W}_2)^{-1} \boldsymbol{\mu} \quad (1.6)$$

De esta manera,  $\boldsymbol{\mu}$  está dado por:

$$\boldsymbol{\mu} = (\mathbf{I} - \lambda \mathbf{W}_2) [(\mathbf{I} - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}] \quad (1.7)$$

con distribución normal de media  $\mathbf{0}$  y matriz de covarianzas  $\boldsymbol{\Omega}$ . Luego, la función de verosimilitud para  $\boldsymbol{\mu}$  es:

$$L(\boldsymbol{\mu}) \propto |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \boldsymbol{\mu}' \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} \right]$$

Para obtener la función de verosimilitud de  $\mathbf{y}$ , se debe tomar en cuenta el teorema de transformación, [16],[19], y aplicar el jacobiano a la transformación en la ecuación 1.7.

De esta manera, el jacobiano de  $\mathbf{y}$  está dado por :

$$J = |\mathbf{I} - \lambda \mathbf{W}_2| |\mathbf{I} - \rho \mathbf{W}_1|$$

Luego, la función de verosimilitud de  $\mathbf{y}$  es:

$$L(\mathbf{y}) \propto |\mathbf{I} - \lambda \mathbf{W}_2| |\mathbf{I} - \rho \mathbf{W}_1| |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \boldsymbol{\mu}' \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} \right] \quad (1.8)$$

Finalmente, se reemplaza 1.7 en 1.8, para obtener:

$$L(\mathbf{y}) \propto |\mathbf{I} - \lambda \mathbf{W}_2| |\mathbf{I} - \rho \mathbf{W}_1| |\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} \boldsymbol{\mu}' \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} \right] \quad (1.9)$$

Si el modelo es homocedástico, entonces  $\boldsymbol{\Omega} = \sigma^2 \mathbf{I}$ , y por lo tanto, la función de verosimilitud podría ser reescrita de la forma

$$L(\mathbf{y}) \propto |\mathbf{I} - \lambda \mathbf{W}_2| |\mathbf{I} - \rho \mathbf{W}_1| (\sigma^{2n})^{-\frac{1}{2}} \exp \left[ -\frac{1}{2\sigma^2} \boldsymbol{\mu}' \boldsymbol{\mu} \right] \quad (1.10)$$

Para maximizar esta función de verosimilitud se debe calcular el vector de primeras derivadas, igualar a cero y resolver las ecuaciones resultantes. El vector de primeras derivadas de la función de verosimilitud, recordando que  $\boldsymbol{\Omega} = \sigma^2 \mathbf{I}$  está dado por:

$$\begin{aligned} \partial L / \partial \boldsymbol{\beta} &= \boldsymbol{\nu}' (\boldsymbol{\Omega}^{-1/2} (\mathbf{I} - \lambda \mathbf{W}_2) \mathbf{X}) \\ \partial L / \partial \rho &= -tr((\mathbf{I} - \rho \mathbf{W}_1)^{-1} \mathbf{W}_1) + \boldsymbol{\nu}' \boldsymbol{\Omega}^{-1/2} (\mathbf{I} - \lambda \mathbf{W}_2) \mathbf{W}_1 \mathbf{y} \\ \partial L / \partial \lambda &= -tr((\mathbf{I} - \lambda \mathbf{W}_2)^{-1} \mathbf{W}_2) + \boldsymbol{\nu}' \boldsymbol{\Omega}^{-1/2} \mathbf{W}_2 [(\mathbf{I} - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}] \\ \partial L / \partial \alpha_p &= -(1/2) tr(\boldsymbol{\Omega}^{-1} \mathbf{H}_p) + (1/2) \boldsymbol{\nu}' \boldsymbol{\Omega}^{-3/2} \mathbf{H}_p (\mathbf{I} - \lambda \mathbf{W}_2) [(\mathbf{I} - \rho \mathbf{W}_1) \mathbf{y} - \mathbf{X} \boldsymbol{\beta}] \end{aligned}$$

donde  $\mathbf{H}_p$  corresponde al hessiano del  $p$ -ésimo parámetro, con  $p = 1, \dots, P$ . Es evidente que este sistema de ecuaciones no tiene una solución cerrada, por lo que se deben emplear métodos numéricos apropiados para la maximización de esta función de verosimilitud. Tanto [8] como [42], dan una solución similar al problema en modelos homocedásticos. A continuación será presentada la modalidad trabajada en [42] y [57].

El primer paso consiste en realizar la estimación del modelo:

$$(\mathbf{I} - \lambda \mathbf{W}_2) \mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.11)$$

por mínimos cuadrados ponderados, asignando pesos dados por la matriz  $\mathbf{T} = (\mathbf{I} - \rho \mathbf{W}_1)' (\mathbf{I} - \rho \mathbf{W}_1)$ . Se obtiene:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{T} \mathbf{X})^{-1} \mathbf{X}' \mathbf{T} (\mathbf{I} - \lambda \mathbf{W}_2) \mathbf{y} \quad (1.12)$$

El segundo paso consiste en realizar el cálculo de los residuales dada la estimación de  $\boldsymbol{\beta}$ :

$$\hat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \lambda \mathbf{W}_2) \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}} \quad (1.13)$$

Posteriormente, se realiza la estimación de  $\sigma^2$  como:

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}{n} \quad (1.14)$$

Y por último, con las estimaciones de  $\boldsymbol{\beta}$  y  $\sigma^2$  se puede maximizar la función de verosimilitud en (1.10) por métodos numéricos para los parámetros  $\rho$  y  $\lambda$ .

Diversas metodologías para la maximización de funciones de verosimilitud han sido desarrolladas aprovechando la gran disponibilidad de herramientas computacionales, es por ello que este problema también puede ser conducido con la ayuda de funciones del paquete estadístico R, [54], como las encontradas en el paquete *optimx*, [49].

### 1.4.2. Propuestas alternativas

Si bien el ajuste por máxima verosimilitud es el más clásico en la literatura sobre econometría espacial, otras propuestas han sido realizadas para el ajuste de estos modelos, [11], tales como el uso de variables instrumentales, [5] y [8], y la estimación por el método de momentos, [60], [41] y [40]. Otras propuestas direccionadas hacia la solución del problema de heterogeneidad espacial corresponden a la regresión ponderada geográficamente, [32], y la aplicación de un filtro espacial para tratar el problema desde los modelos de regresión usuales, [35].

#### 1.4.2.1. Variables instrumentales

La propuesta del uso de variables instrumentales para el ajuste de modelos econométricos espaciales parte de suponer la existencia de un conjunto de instrumentos,  $\mathbf{Q}$ , fuertemente correlacionados con las variables originales del modelo, es decir,  $\mathbf{Z} = [\mathbf{y}, \mathbf{X}]$ , pero asintóticamente no correlacionado con el término de error. A partir de esta aproximación del problema, se obtiene el estimador

$$\boldsymbol{\theta} = [\mathbf{Q}'\mathbf{Z}]^{-1}\mathbf{Q}'\mathbf{y}$$

Otra idea, basada en el método de momentos, para la interpretación de la estimación de variables instrumentales, [8], consiste en la solución del sistema de ecuaciones dado por:

$$\mathbf{Q}'(\mathbf{y} - \mathbf{Z}\boldsymbol{\theta}) = \mathbf{0}$$

llevando al estimador:

$$\boldsymbol{\theta} = (\mathbf{Z}'\mathbf{P}_\mathbf{Q}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_\mathbf{Q}\mathbf{y} \quad (1.15)$$

con  $\mathbf{P}_\mathbf{Q} = \mathbf{Q}(\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'$ .

Esta propuesta para el ajuste de modelos econométricos espaciales es empleada principalmente para el ajuste del modelo SAR y del modelo SEM, y ofrece ventajas, ya que no requiere el supuesto de normalidad en el término de error, además el estimador es consistente. A pesar de no ser tan eficiente, este puede ser empleado como base para la aplicación de métodos de remuestreo, [3], tal como bootstrap, [27] y [30].

Si bien este estimador posee buenas propiedades, la estimación del parámetro  $\rho$  en el modelo SAR puede estar por fuera del límite teórico,  $(-1,1)$ , cuando el modelo tenga problemas de especificación, [3].

#### 1.4.2.2. Estimación por el método de los momentos

La estimación por el método de momentos fue presentada por [40] y [41], consiste en la formulación de un sistema de ecuaciones a partir del cual se obtienen las estimaciones del parámetro de rezago espacial y del parámetro de dispersión. En el caso del modelo SAR,

el sistema está dado por:

$$\begin{aligned} E[\mathbf{u}'\mathbf{u}/N] &= \sigma^2 \\ E[\mathbf{u}'\mathbf{W}'\mathbf{W}\mathbf{u}/N] &= \sigma^2(1/N)\text{tr}(\mathbf{W}'\mathbf{W}) \\ E[\mathbf{u}'\mathbf{W}\mathbf{u}/N] &= 0 \end{aligned}$$

donde  $\mathbf{u}$  es un vector de errores independientes e idénticamente distribuidos. Tomando  $\mathbf{u}$  como  $\varepsilon - \lambda\varepsilon$ , el sistema tiene como parámetros  $\lambda, \lambda^2$  y  $\sigma^2$ , [10]. En [41] se sugiere el uso de mínimos cuadrados no lineales para obtener la estimación de  $\lambda$ , y de esta manera estimar  $\beta$  y  $\sigma^2$ .

### 1.4.2.3. Regresión ponderada geográficamente

La regresión ponderada geográficamente, [32], consiste en obtener submuestras de datos para generar una estimación local en la cual se pondera por la distancia entre unidades espaciales, [44]. Se parte de un vector de pesos para cada unidad espacial, el cual depende mayormente de la distancia, pero que puede ser parametrizado de manera libre. Varias propuestas son presentadas en [44]. Una vez definidas las ponderaciones, se realiza la propuesta del modelo, el cual está dado por:

$$\mathbf{W}_i\mathbf{y} = W_iX\beta_i + \varepsilon_i$$

donde  $\mathbf{W}_i$  es el vector de ponderaciones, para el individuo  $i$ ,  $\mathbf{y}$  es el vector de la variable respuesta,  $\mathbf{X}$  es la matriz de variables explicativas,  $\beta_i$  es el vector de parámetros asociado a la observación  $i$  y  $\varepsilon_i$  es el término de error, el cual está normalmente distribuido.

### 1.4.2.4. Aplicación de filtros espaciales

[35] propuso la idea de ajustar un modelo de regresión sin dependencia espacial, partiendo de la aplicación de un filtro dado por:

$$x_i^* = \frac{x_i \left( \frac{\mathbf{W}_i}{n-1} \right)}{G_i(d)}$$

donde  $x_i \left( \frac{\mathbf{W}_i}{n-1} \right)$  es la proporción de unidades espaciales ubicadas a  $d$  unidades de la unidad espacial  $i$  y  $G_i(d)$  corresponde al estadístico  $G$  de Getis calculado para la  $i$ -ésima unidad espacial, [37]. Sin embargo, la construcción de esta nueva variable depende en gran manera de la distancia,  $d$ , respecto a la cual no existe una metodología clara de cálculo.

## 1.5. Propuestas Bayesianas para el ajuste de modelos econométricos espaciales

Las propuestas de ajuste clásico para modelos econométricos espaciales tienen algunas restricciones. Por ejemplo, el estimador de máxima verosimilitud parte de un supuesto de normalidad, así como un tamaño de muestra apto para asumir algunos resultados

asintóticos. Ante estas limitaciones, el paradigma Bayesiano surge como solución, ya que es útil para casos donde el tamaño de muestra es pequeño o hay observaciones atípicas, [43].

### 1.5.1. Ajuste Bayesiano Clásico

Una primera alternativa para el ajuste de modelos econométricos espaciales es presentada por [5],[6], [8] y [39]. El primer modelo trabajado es el SAR:

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \boldsymbol{\varepsilon} \quad (1.16)$$

con distribuciones a priori para los parámetros  $\rho$  y  $\sigma$ , dadas por:

$$\begin{aligned} \pi(\sigma) &\propto \sigma^{-1} & 0 < \sigma < \infty \\ \pi(\rho) &\propto cte & -1 < \rho < 1 \end{aligned}$$

Al realizar el producto entre las distribuciones a priori, bajo el supuesto de independencia entre los parámetros, y la función de verosimilitud normal para este modelo se obtiene la siguiente distribución a posteriori:

$$P(\rho, \sigma | \mathbf{y}) \propto |\mathbf{I} - \rho \mathbf{W}| \sigma^{-(N+1)} \times \exp\{-(2\sigma^2)^{-1} \mathbf{y}' (\mathbf{I} - \rho \mathbf{W})' (\mathbf{I} - \rho \mathbf{W}) \mathbf{y}\} \quad (1.17)$$

Posteriormente, se obtiene la distribución marginal a posteriori para  $\rho$ :

$$P(\rho | \mathbf{y}) \propto |\mathbf{I} - \rho \mathbf{W}| \{(-2\sigma^2)^{-1} \mathbf{y}' (\mathbf{I} - \rho \mathbf{W})' (\mathbf{I} - \rho \mathbf{W}) \mathbf{y}\} \quad (1.18)$$

A partir de (1.18), se puede obtener el valor esperado de  $\rho$ .

De igual manera fue trabajado el modelo SEM, las propuestas de distribuciones a priori eran:

$$\begin{aligned} \pi(\sigma) &\propto \sigma^{-1} & 0 < \sigma < \infty \\ \pi(\lambda) &\propto cte & -1 < \lambda < 1 \\ \pi(\boldsymbol{\beta}) &\propto cte \end{aligned}$$

Bajo el supuesto de independencia entre los parámetros, y al realizar el producto entre las distribuciones a priori y la función de verosimilitud normal de este modelo, se llega a la siguiente distribución conjunta a posteriori:

$$P(\lambda, \sigma, \boldsymbol{\beta}) \propto |\mathbf{I} - \lambda \mathbf{W}| \sigma^{-N} \exp\{-(1/2\sigma^2) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{I} - \lambda \mathbf{W})' (\mathbf{I} - \lambda \mathbf{W}) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} \quad (1.19)$$

De modo similar a como se hizo con el modelo SAR, se pueden obtener las distribuciones marginales para cada uno de los parámetros mediante integración numérica, la cual puede llegar a ser computacionalmente ineficiente, [8], [43].

### 1.5.2. Ajuste a través de algoritmos MCMC

Más adelante, con la mejora en tiempos y eficiencia de los computadores, los algoritmos MCMC, [33], se abrieron campo en el ajuste de modelos econométricos espaciales. Una de las propuestas más reconocidas fue presentada en [43], la cual estaba enfocada en el

ajuste del modelo SAR. Esta idea se realizó tanto para el caso homocedástico como para el heterocedástico bajo el entorno de un muestreador de Gibbs, [21], [55].

En el caso homocedástico el modelo está dado por

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (1.20)$$

donde se trabajan distribuciones a priori difusas, para  $\boldsymbol{\beta}$ ,  $\sigma^2$  y  $\rho$ . Las distribuciones condicionales a posteriori tanto para  $\boldsymbol{\beta}$  como para  $\sigma^2$  no supusieron ningún problema para el muestreador, ya que para  $\boldsymbol{\beta}$  puede ser expresada como una distribución normal multivariada, y para  $\sigma$  puede ser expresada como:

$$p(\sigma|\rho, \boldsymbol{\beta}) \propto \sigma^{-(n+1)} \exp \left\{ -\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) \right\}$$

lo que permite concluir que  $\sigma^2 \sim \chi^2(n)$ . Sin embargo, para  $\rho$  se obtuvo:

$$p(\rho|\boldsymbol{\beta}, \sigma, \mathbf{y}) \propto |\mathbf{I} - \rho \mathbf{W}| \exp \left\{ -\frac{1}{2\sigma^2}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}) \right\}$$

la cual no es una distribución sencilla de muestrear, por lo que se obtuvieron muestras de este parámetro a través de la razón de uniformes, [28], la cual parte del supuesto de que la función de distribución del parámetro  $\rho$  es acotada. De esta manera, se obtienen los valores  $b$ ,  $a_+$  y  $a_-$ , dados por:

$$\begin{aligned} b &\geq \sup \sqrt{p(\rho|\boldsymbol{\beta}, \sigma, \mathbf{y})} \\ a_- &\leq \inf \rho \sqrt{p(\rho|\boldsymbol{\beta}, \sigma, \mathbf{y})} \\ a_+ &\geq \sup \rho \sqrt{p(\rho|\boldsymbol{\beta}, \sigma, \mathbf{y})} \end{aligned}$$

Una vez obtenidos estos valores, la generación de  $\rho$  se hace a través del siguiente algoritmo:

1. Generar  $r$  como una realización de la variable aleatoria  $\mathbf{R} \sim U[0, b]$
2. Generar  $t$  como una realización de la variable aleatoria  $\mathbf{T} \sim U[a_-, a_+]$
3. Tomar  $\tilde{\rho} = \mathbf{R}/\mathbf{T}$
4. Repetir el algoritmo hasta que  $\mathbf{R}^2 \leq p(\tilde{\rho}|\boldsymbol{\beta}, \sigma, \mathbf{y})$

Por otra parte, se presentó el ajuste del modelo heterocedástico, dado por:

$$\begin{aligned} \mathbf{y} &= \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N(\mathbf{0}, \sigma^2 \mathbf{V}), \quad \mathbf{V} = \text{diag}(\nu_1, \dots, \nu_n) \end{aligned}$$

con distribuciones a priori difusas para  $\boldsymbol{\beta}$ ,  $\sigma$  y  $\rho$ , y distribuciones a priori independientes para  $\mathbf{V} = \nu_1, \dots, \nu_n$ , dadas por:

$$\begin{aligned} \pi(\nu_i^{-1}|q) &\sim \chi^2(q)/q \\ q &\sim \text{constante} \end{aligned}$$

Las distribuciones condicionales a posteriori de los parámetros  $\sigma$  y  $\nu_i$ ,  $i = 1, \dots, n$  son:

$$\begin{aligned} \sum_{i=1}^n (\varepsilon_i^2 / \nu_i) / \sigma^2 &\sim \chi^2(n) \\ (\sigma^{-2} \varepsilon_i^2 + q) / \nu_i &\sim \chi^2(q + 1) \end{aligned}$$

Para  $\beta$ , se obtiene una distribución normal multivariada. La distribución a posteriori para  $\rho$  no tiene una forma conocida, por lo que se utiliza de nuevo la razón de uniformes, [28], para obtener las respectivas muestras.

Más adelante, [42] hizo la presentación de la aplicación de la metodología Bayesiana MCMC, [33], a través del muestreador de Metropolis Hastings, [59],[56], y presentó el modelo econométrico espacial Bayesiano desde el cual se expondrá nuestra metodología de ajuste. En el caso del modelo SAR homocedástico, está dado por:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \varepsilon$$

donde  $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , con distribuciones a priori dadas por:

$$\begin{aligned} \pi(\beta, \sigma^2) &\sim \text{NormalGammaInversa}(b, B, a, b) \\ &= \pi(\beta | \sigma^2) \pi(\sigma^2) \\ &= N(c, \sigma^2 \mathbf{T}) \text{GammaInversa}(a, b) \\ \pi(\rho) &\sim U\left(\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}}\right) \end{aligned}$$

donde  $\lambda_{min}$  y  $\lambda_{max}$  corresponden al mínimo y al máximo valor propio de la matriz de vecindades,  $\mathbf{W}$ , sin estandarizar por filas, respectivamente. La construcción de las distribuciones condicionales a posteriori para los parámetros  $\beta$  y  $\sigma^2$  es similar a la realizada en el capítulo 2 de este trabajo. Sin embargo, el muestreo de  $\rho$  se hizo a través del algoritmo de Metropolis-Hastings, con una distribución propuesta normal y un procedimiento llamado caminata aleatoria tuneada, para producir candidatos de  $\rho$ . El muestreo de los valores de la distribución propuesta,  $\rho^*$ , están dados por:

$$\rho^* = \rho^{(c)} + d \cdot N(0, 1)$$

donde  $\rho^{(c)}$  es el valor actual de  $\rho$ . Se incluyó el parámetro  $d$ , llamado de tuneo, con el fin de que el muestreo recorra todo el espacio de la distribución condicional. Además, se propuso la idea de hacer el algoritmo adaptativo, para mantener una tasa de aceptación de entre el 40% y el 60%.

### 1.5.3. INLA (Integrated Nested Laplace Approximation)

La Aproximación de Laplace Anidada Integrada (INLA), por sus iniciales en inglés, [58], es una metodología que se emplea para obtener una aproximación de las distribuciones marginales a posteriori de los parámetros de interés. En [17],[45] se presenta la aplicación de esta metodología a los modelos econométricos espaciales presentados en [8] y [42].

Su propuesta consiste en realizar el ajuste del modelo condicionado sobre  $\rho$ , a través de la metodología *INLA*, [58]. Una vez obtenida la estimación de los parámetros del modelo,

---

se valen del intervalo sobre el cuál está definido  $\rho$ , para construir una grilla de valores entre sus límites teóricos, y de esta manera aproximar su distribución marginal a posteriori.

Esta metodología ofrece ventajas computacionales respecto a los muestreadores MCMC, [17], ya que es posible realizar el cálculo paralelo de los distintos estados de la cadena de Markov.

---

---

# Modelos Econométricos Espaciales Bayesianos Homocedásticos

---

---

Tal como se enunció en el capítulo anterior, de acuerdo a la especificación que se haga de la matriz de varianzas y covarianzas del término de error en el modelo general, (1.1), se puede obtener un modelo con homogeneidad de varianzas a través del espacio. Para obtener modelos econométricos espaciales homocedásticos, la matriz de varianzas y covarianzas debe ser expresada como  $\mathbf{\Omega} = \sigma^2 \mathbf{I}$ .

En este capítulo se presentarán alternativas Bayesianas para el ajuste de los modelos SAR, SEM y General homocedásticos, [8], basadas en algoritmos MCMC, [33], como el muestreador de Gibbs, [21], [55], y el algoritmo de Metropolis Hastings,[56], [59]. Para cada modelo se presentará su formulación Bayesiana, la derivación de sus funciones de distribución condicional a posteriori y la obtención de muestras a partir de ellas. Se realizarán dos aplicaciones, en la primera de ellas se ajustará un conjunto de datos obtenidos a partir de la simulación de una variable normal bajo distintos escenarios de los parámetros autoregresivos. La segunda aplicación se realizará con el conjunto de datos de crimen en 49 distritos de Columbia, trabajados en [8].

### 2.1. Modelo SAR Bayesiano: Homocedástico

#### 2.1.1. Especificación

El modelo SAR (Spatial AutoRegressive) se obtiene a partir del modelo general, (1.1), homocedástico, haciendo  $\lambda = 0$ . Es decir, está dado por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}$$

donde  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ . Las distribuciones a priori de este modelo son:

1.  $p(\boldsymbol{\beta}) \sim N(\mathbf{b}, \mathbf{B})$
2.  $p(\sigma^2) \sim \text{GammaInv}(\zeta, \vartheta)$
3.  $p(\rho) \sim U(-1/\omega_{max}, 1)$

donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindad espacial,  $\mathbf{W}$ , [8]. Por facilidad se asume que los parámetros son independientes a priori. Los términos  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\rho$  y  $\mathbf{W}$  corresponden a la especificación hecha en la sección 1.4.1.

Basados en las distribuciones a priori y la verosimilitud en (1.10) se obtiene la distribución a posteriori, dada por:

$$\pi(\boldsymbol{\beta}, \rho, \sigma^2) \propto (\sigma^2)^{-n/2} |\mathbf{A}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \frac{\zeta^\vartheta}{\Gamma(\vartheta)} (\sigma^2)^{-\vartheta-1} \exp \left( -\frac{\zeta}{\sigma^2} \right) \quad (2.1)$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$ . A partir de (2.1), las siguientes distribuciones condicionales a posteriori son obtenidas. Las demostraciones paso a paso de la derivación de estas distribuciones se encuentran en el apéndice A.

1. **Distribución condicional a posteriori para  $\boldsymbol{\beta}$ , [22], [23]**

Para el vector de parámetros  $\boldsymbol{\beta}$  se obtiene una distribución normal  $\pi(\boldsymbol{\beta} | \rho, \sigma^2) \sim N(\mathbf{b}^*, \mathbf{B}^*)$ , donde:

$$\mathbf{b}^* = \mathbf{B}^* \left( \frac{1}{\sigma^2} \mathbf{X}'\mathbf{y} - \frac{\rho}{\sigma^2} \mathbf{X}'\mathbf{W}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right)$$

$$\mathbf{B}^* = \left( \frac{1}{\sigma^2} \mathbf{X}'\mathbf{X} + \mathbf{B}^{-1} \right)^{-1}$$

2. **Distribución condicional a posteriori para  $\sigma^2$ , [22]**

La siguiente es la distribución condicional de  $\sigma^2$  dado  $(\boldsymbol{\beta}, \rho)$  es:

$$\pi(\sigma^2 | \boldsymbol{\beta}, \rho) \sim \text{GammaInv} \left( \zeta + \frac{n}{2}, \frac{k}{2} + \vartheta \right)$$

donde  $k = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

3. **Distribución condicional a posteriori para  $\rho$**

La distribución condicional a posteriori de  $\rho$  dado  $(\boldsymbol{\beta}, \sigma^2)$  es:

$$\pi(\rho | \boldsymbol{\beta}, \sigma^2) \propto |\mathbf{A}| \exp \left\{ -\frac{1}{2} \frac{1}{\sigma^2} \left( \rho - \frac{d}{c} \right)^2 \right\} \mathbf{I}_{(-1/\omega_{max}, 1)}(\rho) \quad (2.2)$$

con  $c = \frac{1}{\sigma^2} \mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}$  y  $d = \frac{1}{\sigma^2} \mathbf{y}'\mathbf{W}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Es evidente que la forma funcional de la distribución condicional a posteriori de  $\rho$  no corresponde a una función conocida, [42], es por esto que a continuación se proponen dos alternativas Bayesianas para el ajuste de este modelo:

- **Kernel de transición uniforme:** Una primera propuesta es hacer uso del muestreador de Metropolis-Hastings, [56], [59], empleando un kernel de transición unifor-

me para  $\rho$  dentro del intervalo  $(-1/\omega_{max}, 1)$ , donde  $\omega_{max}$  es el mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8].

- **Kernel de transición normal:** Tomando como referencia la distribución condicional a posteriori de  $\rho$ , 2.2, se puede proponer un kernel de transición normal,  $N(c^*, d^*)$ , con truncamiento en el intervalo  $(-1/\omega_{max}, 1)$ , donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8]. De esta manera se encuentra implementado en la función *hom\_sar* del paquete *BayesianSpatialReg* de R, [54]. Los parámetros  $c^*$  y  $d^*$  están dados por:

$$c^* = \frac{\mathbf{y}'\mathbf{W}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}} \qquad d^* = \frac{\sigma^2}{\mathbf{y}'\mathbf{W}'\mathbf{W}\mathbf{y}}$$

### 2.1.2. Simulación

Se desarrolló una simulación de 49 datos de una variable normal,  $\mathbf{y}$ , a partir del siguiente modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}$$

con  $\boldsymbol{\varepsilon} \sim N(0, 45\mathbf{I})$ . Para este modelo  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]; \boldsymbol{\beta} = [18, 0.478, -1.3]$ ;  $\sigma^2 = 45$  y hay tres distintos valores para  $\rho$ : 0.12, 0.52 y 0.90. Esto implica que la variable  $\mathbf{y}$  fue simulada con media  $(\mathbf{I} - \rho\mathbf{W})^{-1}(18 + 0.478\mathbf{x}_1 - 1.3\mathbf{x}_2)$  y matriz de varianzas y covarianzas dada por  $[(\mathbf{I} - \rho\mathbf{W})^{-1}]' \boldsymbol{\Sigma} [(\mathbf{I} - \rho\mathbf{W})^{-1}]$ , donde  $\boldsymbol{\Sigma}$  es una matriz diagonal con  $\Sigma_{i,i} = 45, i = 1, \dots, 49$ .

Por su parte, las realizaciones de las variables  $\mathbf{X}_1$  y  $\mathbf{X}_2$  fueron generadas a partir de distribuciones uniformes, definidas en los intervalos  $(0, 400)$  y  $(10, 23)$ , respectivamente. Finalmente, la matriz de vecindades  $\mathbf{W}$  es la misma utilizada en la aplicación de datos de crimen en Columbus, Ohio por [8].

Luego se ajustaron los datos con las dos metodologías propuestas y mediante Máxima Verosimilitud. Los resultados se presentan en la tabla 2.1, y son discutidos a continuación.

TABLA 2.1. Comparación de ajustes del modelo SAR Homocedástico

$\rho$ real	Método	Valor	$\beta_0$	$\beta_1$	$\beta_2$	$\rho$	$\sigma^2$	BIC	DIC
0,12	Kernel Uniforme	Estimación	-9,966	0,493	-0,957	0,660	110,536	761,129	2759,199
		Desviación estándar	6,872	0,012	0,367	0,196	24,113		
	Kernel Normal	Estimación	20,595	0,490	-1,429	0,081	54,229	258,551	894,800
		Desviación estándar	6,233	0,008	0,276	0,033	11,997		
	Máxima Verosimilitud	Estimación	21,658	0,489	-1,460	0,076	47,957	258,175	
		Desviación estándar	8,386	0,008	0,261	0,035	9,689		
0,52	Kernel Uniforme	Estimación	21,208	0,483	-1,313	0,664	44,218	1423,914	8179,361
		Desviación estándar	5,766	0,008	0,231	0,194	9,717		
	Kernel Normal	Estimación	20,388	0,483	-1,321	0,510	44,225	251,708	874,179
		Desviación estándar	5,523	0,008	0,231	0,020	9,625		
	Máxima Verosimilitud	Estimación	21,596	0,482	-1,340	0,506	38,971	251,284	
		Desviación estándar	5,613	0,008	0,219	0,022	7,878		
0,90	Kernel Uniforme	Estimación	38,366	0,497	-1,293	0,663	111,636	18686,82	93744,73
		Desviación estándar	31,723	0,015	0,416	0,196	42,465		
	Kernel Normal	Estimación	19,617	0,493	-1,332	0,895	97,175	302,857	1027,753
		Desviación estándar	13,237	0,013	0,374	0,012	21,272		
	Máxima Verosimilitud	Estimación	27,708	0,493	-1,393	0,888	86,669	302,293	
		Desviación estándar	17,383	0,012	0,368	0,016	17,603		

- Cuando  $\rho = 0.12$  se encuentra que la estimación hecha a través del kernel de transición uniforme presenta fuertes diferencias a los parámetros reales, especialmente para  $\beta_0$  y  $\sigma^2$ . Por otra parte, las estimaciones hechas a partir del kernel de transición normal y por Máxima Verosimilitud son muy similares, destacando que al emplear el kernel de transición normal se obtiene una estimación más cercana al verdadero valor de  $\rho$ , a pesar de tener un *BIC* mayor al obtenido mediante Máxima Verosimilitud.
- Se evidencia, cuando  $\rho = 0.52$ , que la estimación hecha a través del kernel de transición uniforme para  $\rho$  no se aleja de la vecindad de 0.66. Si bien el ajuste por Máxima Verosimilitud presenta errores estándar menores a los del ajuste con el kernel de transición normal, se encuentra que la propuesta Bayesiana es más acertada para los parámetros  $\rho$  y  $\sigma^2$ .
- Cuando el verdadero valor del parámetro  $\rho$  es 0,90 se evidencia que la propuesta bayesiana con kernel de transición normal se aproxima a la estimación con Máxima Verosimilitud, aportando errores estándar similares. La aproximación al verdadero valor de  $\rho$  es muy buena, tanto con la propuesta Bayesiana con kernel de transición normal, como con Máxima Verosimilitud. Para el ajuste Bayesiano con kernel de transición uniforme la estimación de  $\rho$  de nuevo es de 0.66.

También hay que resaltar que empleando el kernel de transición uniforme no se logran tasas de aceptación superiores al 3%, mientras que con el kernel normal las tasas se encuentran alrededor del 79%.

### 2.1.3. Aplicación: Datos de crimen en Columbus

Con el fin de aplicar la metodología propuesta para el modelo SAR homocedástico, se trabajó con los datos presentados en [8], que relacionan el nivel de crimen en 49 distritos de la ciudad de Columbus, Ohio con medidas de ingreso y valor de la vivienda.

Las variables de este conjunto de datos son:

- **CRIME:** Robos residenciales y de autos por cada 1000 hogares.
- **HOUSE:** Valor de la vivienda en 1000 dólares.
- **INC:** Ingreso del hogar en 1000 dólares.

El modelo a ajustar está dado por:

$$CRIME = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}CRIME + \varepsilon$$

con  $\mathbf{X} = [\mathbf{1}; HOUSE; CRIME]$ , donde  $\mathbf{1}$  es un vector de unos de longitud  $n = 49$ . La matriz de vecindades en este caso es de contigüidad binaria estandarizada por filas, es decir, si las unidades  $i$  y  $j$  comparten alguna frontera, entonces  $w_{ij} \neq 0$ . Ahora bien, al ser una matriz estandarizada por filas, se debe garantizar que  $\sum_j w_{ij} = 1$ . Esta matriz está disponible en el conjunto de datos *columbus* del paquete *spdep* de R, [54].

En este caso, se propuso un vector de ceros y una matriz diagonal con elementos

diagonales iguales a 10000 como valores a priori para el vector de medias y la matriz de varianzas y covarianzas de  $\beta$ , respectivamente. Mientras que se propuso un valor a priori de 0.01 para cada uno de los parámetros de la distribución a priori de  $\sigma^2$ . Se realizaron 50000 simulaciones, con un calentamiento de 2500 valores. Además, con el fin de minimizar el efecto de correlación entre las simulaciones se seleccionaron valores cada 10 simulaciones.

### 2.1.3.1. Resultados: Ajuste con Kernel de Transición Uniforme

En la tabla 2.2, se presentan las estimaciones para cada uno de los parámetros, así como su error estándar. Con el fin de comparar el presente trabajo frente a las metodologías clásicas, se expondrán los resultados descritos en [8], y se evaluarán las dos ideas a través del Criterio de Información Bayesiano (BIC), [50].

TABLA 2.2. Parámetros estimados del modelo SAR con kernel uniforme

	$\hat{\theta}_B$	E.E.	$\hat{\theta}_{MV}$	E.E.
$\beta_0$	47.748	5.710	54.201	7.295
$\beta_1$	-0.308	0.099	-0.305	0.094
$\beta_2$	-1.087	0.330	-1.254	0.317
$\sigma^2$	119.64	26.348	107.46	10.366
$\rho$	0.664	0.193	0.315	0.134
BIC	317.049		327.51	
DIC	1165.337			
% Aceptación	17.18			

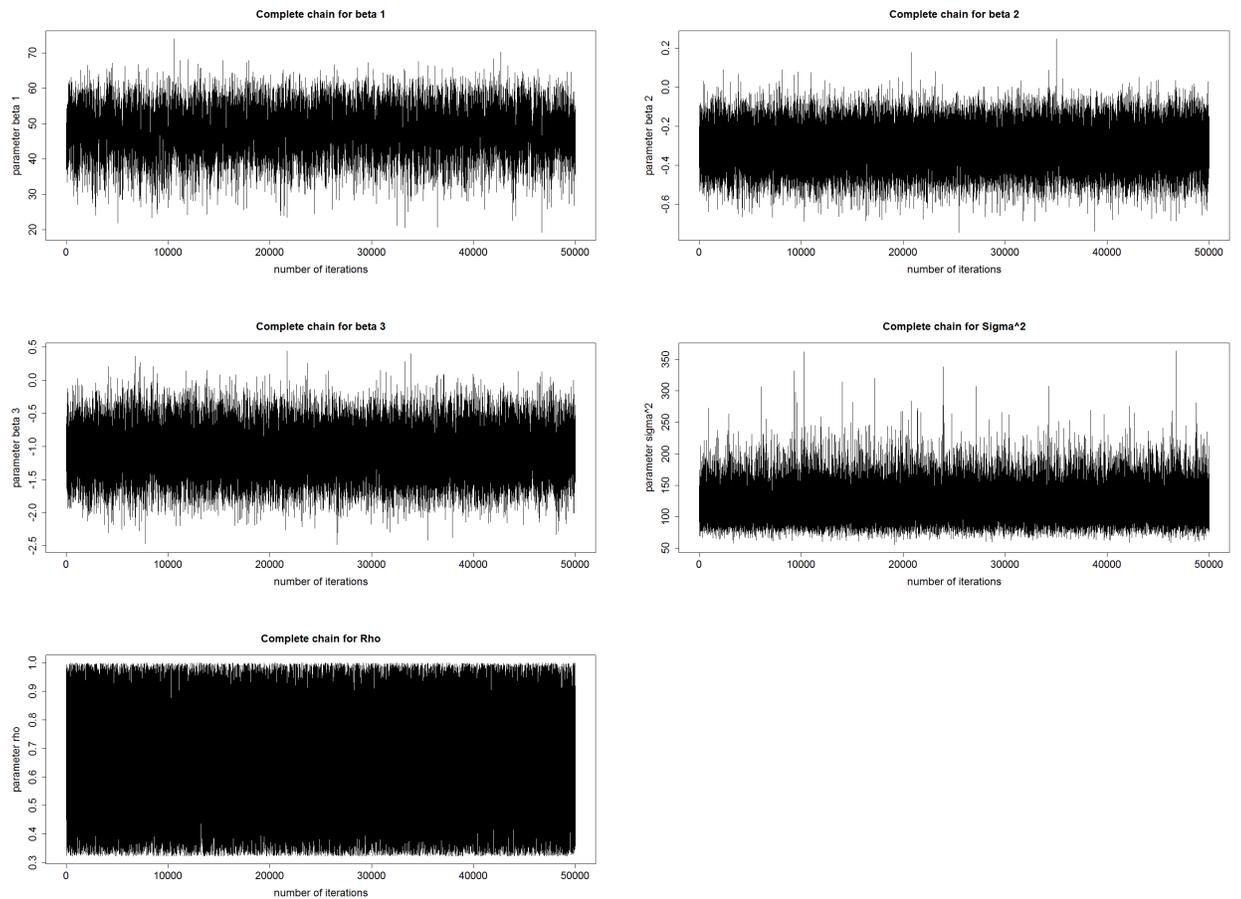
Los resultados reflejan que las estimaciones puntuales del vector de parámetros  $\beta$  son próximas puesto que están a alrededor de una desviación estándar. Además, los errores estándar de la propuesta Bayesiana son muy cercanos a los presentados en la estimación por Máxima Verosimilitud.

En el caso del parámetro  $\sigma^2$ , la estimación hecha con el kernel uniforme está por encima de la obtenida con Máxima Verosimilitud en más de una desviación estándar, mientras que la estimación de  $\rho$  es exactamente 0.664, la cual es mayor a la obtenida por Máxima Verosimilitud.

En cuanto al Criterio de Información Bayesiano, este es menor para las estimaciones obtenidas con el kernel de transición uniforme que con la metodología clásica, a pesar de tener una tasa de aceptación para el parámetro  $\rho$  por debajo del 20 %.

Las cadenas de cada parámetro del modelo SAR homocedástico se encuentran en la figura 2.1.

FIGURA 2.1. Cadenas de parámetros para el modelo SAR homocedástico con kernel de transición uniforme



En las cadenas de los cuatro primeros parámetros se evidencia que se alcanza rápida convergencia, además de tener estabilidad, a diferencia de lo observado para la cadena del parámetro  $\rho$  la cual muestra que no hay estabilidad, al oscilar entre 0.3 y 1.

### 2.1.3.2. Resultados: Ajuste con Kernel de Transición Normal

De manera análoga a como se hizo para el kernel uniforme, se ajusta un modelo Bayesiano empleando el kernel de transición normal en el mismo conjunto de datos. Los resultados son expuestos en la tabla 2.3.

TABLA 2.3. Parámetros estimados del modelo SAR con kernel normal

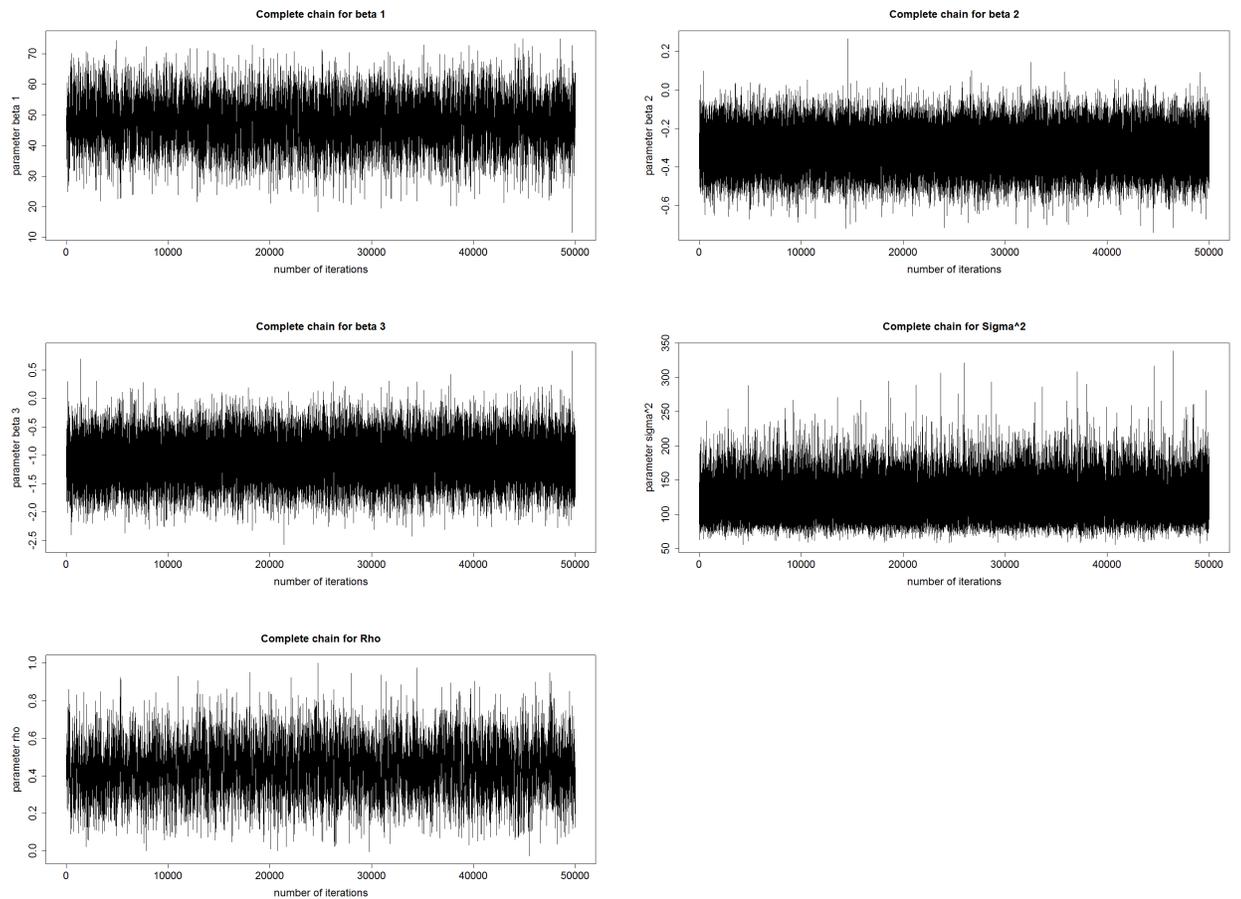
	$\hat{\theta}_B$	E.E.	$\hat{\theta}_{MV}$	E.E.
$\beta_0$	47.441	7.087	54.201	7.295
$\beta_1$	-0.308	0.099	-0.305	0.094
$\beta_2$	-1.079	0.343	-1.254	0.317
$\sigma^2$	120.672	26.660	107.460	10.366
$\rho$	0.441	0.125	0.315	0.134
BIC	300.203		327.51	
DIC	1010.982			
% Aceptación	62.88			

En este caso se observa que las estimaciones puntuales, obtenidas tanto con la metodología clásica como con la Bayesiana, son muy próximas para los parámetros del vector  $\beta$ , mientras que para  $\sigma^2$  y  $\rho$  las diferencias son de mayor magnitud. Vale la pena resaltar que el error estándar de la estimación de  $\rho$  es menor para la propuesta Bayesiana que para Máxima Verosimilitud.

Finalmente, el Criterio de Información Bayesiano (BIC), [50], es menor para la estimación a través del kernel de transición normal que para el ajuste por Máxima Verosimilitud. Además, la tasa de aceptación del parámetro  $\rho$  está alrededor del 63 %.

Las cadenas para cada parámetro del modelo SAR homocedástico obteniendo muestras de  $\rho$  mediante el kernel de transición normal se encuentran en la figura 2.2.

FIGURA 2.2. Cadenas de parámetros para el modelo SAR homocedástico con kernel de transición normal



Las cadenas muestran que el algoritmo logra una rápida convergencia hacia los valores estimados. De igual manera, las cadenas muestran buena estabilidad para todos los parámetros, lo que se refleja en bajos errores estándar.

Los histogramas con los residuales obtenidos al hacer el ajuste con el kernel de transición uniforme y normal se encuentran en las figuras 2.3 y 2.4, respectivamente. En ellos se observa una forma de campana con algunas observaciones atípicas para el caso normal. Por otra parte, se ve un fuerte sesgo en la estimación hecha a través del kernel de transición uniforme, puesto que se encuentra centrado en valores negativos alrededor de -20, contrario a lo sucedido con los residuales obtenidos mediante la otra propuesta, los cuales se encuentran centrados alrededor de cero.

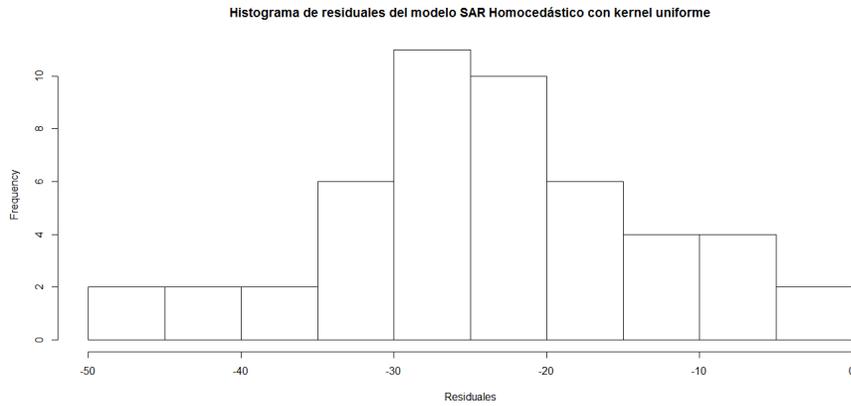


FIGURA 2.3. Residuales para el modelo ajustado con el kernel de transición uniforme para los datos de crimen en Columbus, [8]

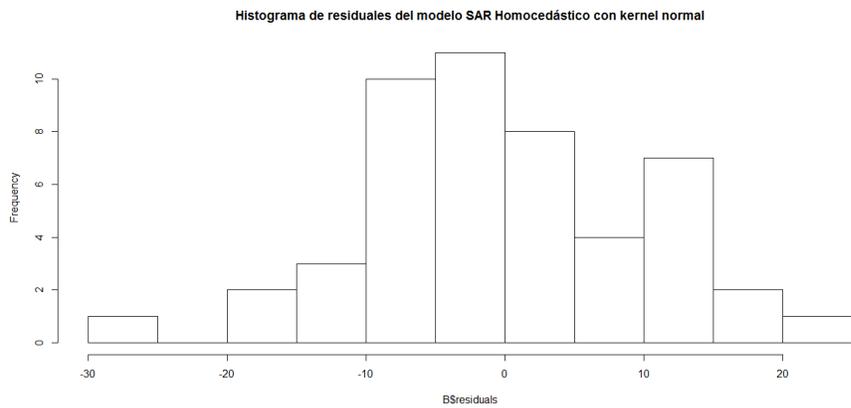


FIGURA 2.4. Residuales para el modelo ajustado con el kernel de transición normal para los datos de crimen en Columbus, [8]

Se evaluaron las medidas de tendencia central de estos residuos, así como si siguen una distribución normal, los resultados se encuentran en la tabla 2.4.

TABLA 2.4. Medidas resumen y test de normalidad para los residuos del ajuste del modelo SAR homocedástico bajo las dos propuestas

	Kernel Uniforme	Kernel Normal
Mínimo	-49,109	-26,680
Media	-23,763	-0,231
Mediana	-24,435	-1,123
Máximo	-3,163	24,564
p-valor (Shapiro Test)	0,579	0,781

Esta información da muestra del sesgo de las estimaciones obtenidas con el kernel de transición uniforme a pesar de tener residuos aparentemente normales.

## 2.2. Modelo SEM Bayesiano: Homocedástico

### 2.2.1. Especificación

El modelo SEM es obtenido cuando  $\rho = 0$  en el modelo general homocedástico. El modelo se escribe como sigue:

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu}\end{aligned}$$

donde  $\boldsymbol{\mu} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ . Las distribuciones a priori de cada uno de los parámetros están dadas por:

1.  $p(\boldsymbol{\beta}) \sim N(\mathbf{b}, \mathbf{B})$
2.  $p(\sigma^2) \sim \text{GammaInv}(\zeta, \vartheta)$
3.  $p(\lambda) \sim U(-1/\omega_{max}, 1)$

donde  $\omega_{max}$  es el mayor valor propio negativo, en valor absoluto, de la matriz de vecindad espacial,  $\mathbf{W}$ , [8]. Por facilidad se asume que los parámetros del modelo son independientes a priori. Los términos  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\lambda$ ,  $\mathbf{W}$  y  $\boldsymbol{\varepsilon}$  corresponden a la descripción realizada en la sección 1.4.1.

Partiendo de estas distribuciones a priori y la verosimilitud en (1.10), se obtiene la distribución a posteriori,  $\pi$ , dada por:

$$\begin{aligned}\pi(\boldsymbol{\beta}, \lambda, \sigma^2) \propto (\sigma^2)^{-n/2} |\mathbf{D}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' (\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \right\} \exp \left\{ -\frac{1}{2} (\mathbf{b} - \boldsymbol{\beta})' \mathbf{B}^{-1} (\mathbf{b} - \boldsymbol{\beta}) \right\} \\ \frac{\zeta^\vartheta}{\Gamma(\vartheta)} (\sigma^2)^{-\vartheta-1} \exp \left( -\frac{\zeta}{\sigma^2} \right) \quad (2.3)\end{aligned}$$

con  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ . A partir de 2.3, se obtienen las siguientes distribuciones condicionales a posteriori. La demostración del cálculo de estas distribuciones se presenta en el apéndice A.

#### 1. Distribución condicional a posteriori para $\boldsymbol{\beta}$ , [22]

Para el vector de parámetros  $\boldsymbol{\beta}$  se obtiene una distribución normal  $\pi(\boldsymbol{\beta}|\sigma^2, \lambda) \sim N(\mathbf{b}^*, \mathbf{B}^*)$ , donde:

$$\begin{aligned}\mathbf{b}^* &= \mathbf{B}^* \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \\ \mathbf{B}^* &= \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1}\end{aligned}$$

con  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ .

### 2. Distribución condicional a posteriori para $\sigma^2$

La siguiente es la distribución condicional obtenida para  $\sigma^2$  dado  $(\boldsymbol{\beta}, \lambda)$ :

$$\pi(\sigma^2 | \boldsymbol{\beta}, \lambda) \sim \text{InvGamma}\left(\zeta + \frac{n}{2}, \frac{k}{2} + \vartheta\right)$$

donde  $k = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}$ .

### 3. Distribución condicional a posteriori para $\lambda$

La distribución condicional a posteriori de  $\lambda$  dado  $(\boldsymbol{\beta}, \sigma^2)$ ,  $\pi(\lambda | \boldsymbol{\beta}, \sigma^2)$ , está dada por:

$$\pi(\lambda | \sigma^2, \boldsymbol{\beta}) \propto |\mathbf{D}| \exp\left\{-\frac{1}{2 \times \frac{1}{c}} \left(\rho - \frac{d}{c}\right)^2\right\} \mathbf{I}_{(-1/\omega_{max}, 1)}(\lambda) \quad (2.4)$$

con  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}$ ,  $c = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $d = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

La forma funcional de la distribución condicional a posteriori de  $\lambda$  no corresponde a una función de distribución conocida, [42]. Por tal razón serán presentadas dos propuestas para el ajuste de este modelo:

1. **Kernel de transición uniforme:** Se propone emplear el muestreador de Metropolis-Hastings, [56], [59], con kernel de transición uniforme, con  $\lambda$  entre  $-1/\omega_{max}$  y 1, donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8].
2. **Kernel de transición normal:** Basado en la distribución condicional a posteriori de  $\lambda$ , 2.4, se propone un kernel de transición normal,  $N(c^*, d^*)$ , truncado en el intervalo  $(-1/\omega_{max}, 1)$ , donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8]. Así se encuentra implementado en la función *hom.sem* del paquete *BSPADATA* de R, [54]. Los parámetros  $c^*$  y  $d^*$  están dados por:

$$c^* = \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \quad d^* = \frac{\sigma^2}{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}$$

#### 2.2.2. Simulación

Para evaluar el performance de las dos propuestas bayesianas en el ajuste del modelo SEM Homocedástico se elaboró la simulación de 49 datos de una variable normal,  $\mathbf{y}$ , a partir del siguiente modelo:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned}$$

con  $\boldsymbol{\mu} \sim N(0, 45\mathbf{I})$ . Para este modelo  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ ;  $\boldsymbol{\beta} = [18, 0.026, -0.4]$ ;  $\sigma^2 = 45$  y hay tres distintos valores para  $\lambda$ : 0.25, 0.55 y 0.85. Esto implica que la variable  $\mathbf{y}$  fue simulada con media  $18 + 0.026\mathbf{x}_1 - 0.4\mathbf{x}_2$  y matriz de varianzas y covarianzas  $[(\mathbf{I} - \lambda \mathbf{W})^{-1}]' \boldsymbol{\Sigma} [(\mathbf{I} - \lambda \mathbf{W})^{-1}]$ , donde  $\boldsymbol{\Sigma}$  es una matriz diagonal con  $\boldsymbol{\Sigma}_{i,i} = 45$ ,  $i = 1, \dots, 49$ .

Por otra parte, las realizaciones de las variables  $\mathbf{X}_1$  y  $\mathbf{X}_2$  fueron generadas a partir de dos distribuciones uniformes, una en el intervalo  $(0, 400)$  y la otra en  $(10, 23)$ ,

respectivamente. Finalmente, la matriz de vecindades  $\mathbf{W}$  es la misma utilizada en la aplicación de datos de crimen en Columbus, Ohio por [8].

Posteriormente, se ajustaron los datos con las dos metodologías propuestas y mediante Máxima Verosimilitud. Los resultados son presentados en la tabla 2.5.

TABLA 2.5. Comparación de ajustes del modelo SAR homocedástico

$\lambda$ real	Método	Valor	$\beta_0$	$\beta_1$	$\beta_2$	$\lambda$	$\sigma^2$	<i>BIC</i>	<i>DIC</i>
0,25	Kernel Uniforme	Estimación	22,363	0,020	-0,740	0,659	42,930	255,436	889,396
		Desviación estándar	5,143	0,008	0,276	0,195	9,313		
	Kernel Normal	Estimación	23,130	0,018	-0,768	0,383	43,240	249,349	869,568
		Desviación estándar	5,047	0,008	0,274	0,245	9,601		
	Máxima Verosimilitud	Estimación	23,989	0,017	-0,800	0,214	38,668	248,084	
		Desviación estándar	4,787	0,008	0,265	0,188	7,848		
0,55	Kernel Uniforme	Estimación	16,321	0,033	-0,431	0,666	46,095	256,177	892,473
		Desviación estándar	4,923	0,008	0,241	0,193	10,230		
	Kernel Normal	Estimación	15,858	0,031	-0,438	0,843	44,082	260,609	909,672
		Desviación estándar	6,941	0,007	0,235	0,132	9,607		
	Máxima Verosimilitud	Estimación	16,838	0,033	-0,447	0,520	41,691	254,807	
		Desviación estándar	4,454	0,007	0,232	0,142	8,638		
0,85	Kernel Uniforme	Estimación	14,256	0,026	-0,442	0,664	51,839	264,244	924,204
		Desviación estándar	6,599	0,009	0,287	0,196	11,872		
	Kernel Normal	Estimación	13,313	0,026	-0,426	0,902	48,816	269,141	931,869
		Desviación estándar	9,546	0,008	0,270	0,078	10,630		
	Máxima Verosimilitud	Estimación	14,918	0,025	-0,465	0,713	45,971	263,806	
		Desviación estándar	5,807	0,008	0,269	0,100	9,687		

- Para el caso  $\lambda = 0.25$ , la estimación de los parámetros del vector  $\beta$  son muy similares bajo las tres metodologías. Sin embargo, la estimación de  $\lambda$  es más precisa mediante Máxima Verosimilitud. Finalmente, se determina que  $\sigma^2$  es mejor estimado mediante el kernel de transición normal, a pesar de presentar errores estándar mayores a los encontrados bajo la metodología clásica.
- Cuando  $\lambda = 0.55$ , las estimaciones del vector  $\beta$  son similares bajo las tres propuestas, tanto en la estimación puntual como en error estándar. No obstante, la estimación de los parámetro  $\lambda$  y  $\sigma^2$  son más próximas a la realidad de los datos, mediante Máxima Verosimilitud y las propuestas Bayesianas, respectivamente
- Finalmente, para el caso  $\lambda = 0,85$ , se observa que las estimaciones del vector  $\beta$  son muy similares bajo los tres escenarios. Otras diferencias se observan en los parámetros  $\sigma^2$  y  $\lambda$ , puesto que las estimaciones mediante Máxima Verosimilitud son muy próximas al verdadero valor de  $\sigma^2$ , mientras que la propuesta Bayesiana con kernel de transición normal realiza la aproximación más cercana al verdadero valor de  $\lambda$ .

Al comparar las propuestas Bayesianas se observa que los valores del BIC y las tasas de aceptación de  $\lambda$  son más favorables para el kernel de transición uniforme cuando  $\lambda = 0,55$  y cuando  $\lambda = 0,85$ . En el caso en que  $\lambda = 0,25$  sucede lo contrario.

### 2.2.3. Aplicación: Datos de crimen en Columbus

Para aplicar la metodología propuesta para el modelo SEM homocedástico se trabajó con los datos presentados en [8], relacionados con el nivel de crimen en 49 distritos de la ciudad

de Columbus, Ohio. Las variables de esta base de datos están explicados en la sección 2.1.3.

El modelo a ajustar está dado por:

$$\begin{aligned} CRIME &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned}$$

con  $\mathbf{X} = [\mathbf{1}; HOUSE; CRIME]$ , donde  $\mathbf{1}$  es un vector de unos de longitud  $n = 49$ . La matriz de vecindades es de contigüidad binaria estandarizada por filas, es decir, si las unidades  $i$  y  $j$  comparten alguna frontera, entonces  $w_{ij} \neq 0$ . Ahora bien, al ser una matriz estandarizada por filas, se debe garantizar que  $\sum_j w_{ij} = 1$ . Esta matriz está disponible en el conjunto de datos *columbus* del paquete *spdep* de R, [54].

En este caso, se propuso un vector de ceros y una matriz con elementos diagonales iguales a 10000, como parámetros de la distribución a priori de  $\boldsymbol{\beta}$ , mientras que se asignó un valor a priori de 0,01 para cada uno de los parámetros a priori de la distribución de  $\sigma^2$ . Se generaon 50000 repeticiones, con un calentamiento de 2500 valores, para minimizar el efecto de correlación entre las simulaciones, se escogieron valores cada 10 repeticiones.

### 2.2.3.1. Resultados: Ajuste con Kernel de Transición Uniforme

En la tabla 2.6 se encuentran las estimaciones de los parámetros del modelo SEM homocedástico, así como su error estándar. Igualmente se exponen los resultados descritos en [8], obtenidos mediante Máxima Verosimilitud y se comparan los dos ajustes mediante el Criterio de Información Bayesiano (BIC), [50].

TABLA 2.6. Parámetros estimados del modelo SEM con kernel uniforme

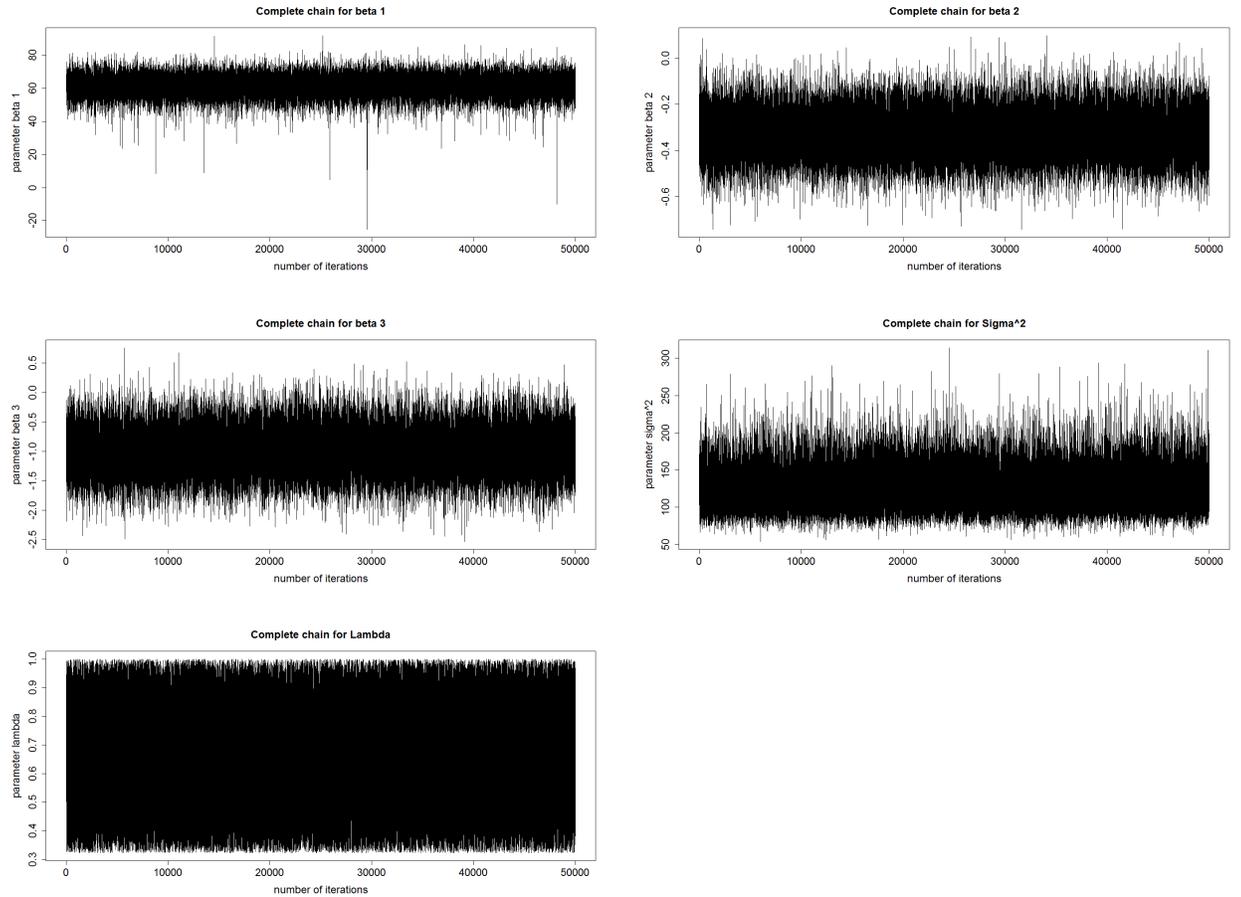
	$\hat{\theta}_B$	E.E.	$\hat{\theta}_{MV}$	E.E.
$\beta_0$	61,529	5,959	59,893	5,366
$\beta_1$	-0,327	0,095	-0,302	0,090
$\beta_2$	-0,998	0,371	-0,941	0,331
$\sigma^2$	124,541	27,385	98,575	19,8737
$\lambda$	0,663	0,196	0,562	0,134
BIC	305,65		304,26	
DIC	1040,22			
% Aceptación	47,24			

Para el ajuste con kernel uniforme, se observa que la estimación del vector  $\boldsymbol{\beta}$  es muy similar a la obtenida mediante Máxima Verosimilitud, logrando incluso diferencias por debajo de una desviación estándar. Para los parámetros,  $\sigma^2$  y  $\lambda$ , las estimaciones no son tan similares y poseen mayores errores estándar en la propuesta Bayesiana que en la clásica.

Al comparar los errores estándar de las dos metodologías se encuentra que el ajuste por Máxima Verosimilitud ofrece mejores resultados al arrojar valores de menor magnitud, y en términos del Criterio de Información Bayesiano, (BIC), la metodología Bayesiana aporta mejores resultados. Finalmente, se observa una tasa de aceptación para

el muestreo del parámetro  $\lambda$  de alrededor del 47%. Ahora, se presentan las cadenas de cada uno de los parámetros estimados.

FIGURA 2.5. Cadenas Parámetros Modelo SEM Homocedástico con Kernel de Transición Uniforme



Al revisar las cadenas de Markov para cada uno de los parámetros, se observa que el algoritmo propuesto conlleva a una rápida convergencia de los parámetros. Sin embargo, se observan algunas simulaciones atípicas para el intercepto y valores muy difusos en el intervalo  $(0, 1)$  para el parámetro  $\lambda$ , lo que se refleja en altos errores estándar.

### 2.2.3.2. Resultados: Ajuste con Kernel de Transición Normal

En la tabla 2.7 están expuestas las estimaciones de cada uno de los parámetros, así como sus errores estándar. De manera paralela se presentan los resultados obtenidos en [8] mediante Máxima Verosimilitud. Posteriormente, se compara el ajuste de las dos metodologías mediante el Criterio de Información Bayesiano (BIC), [50].

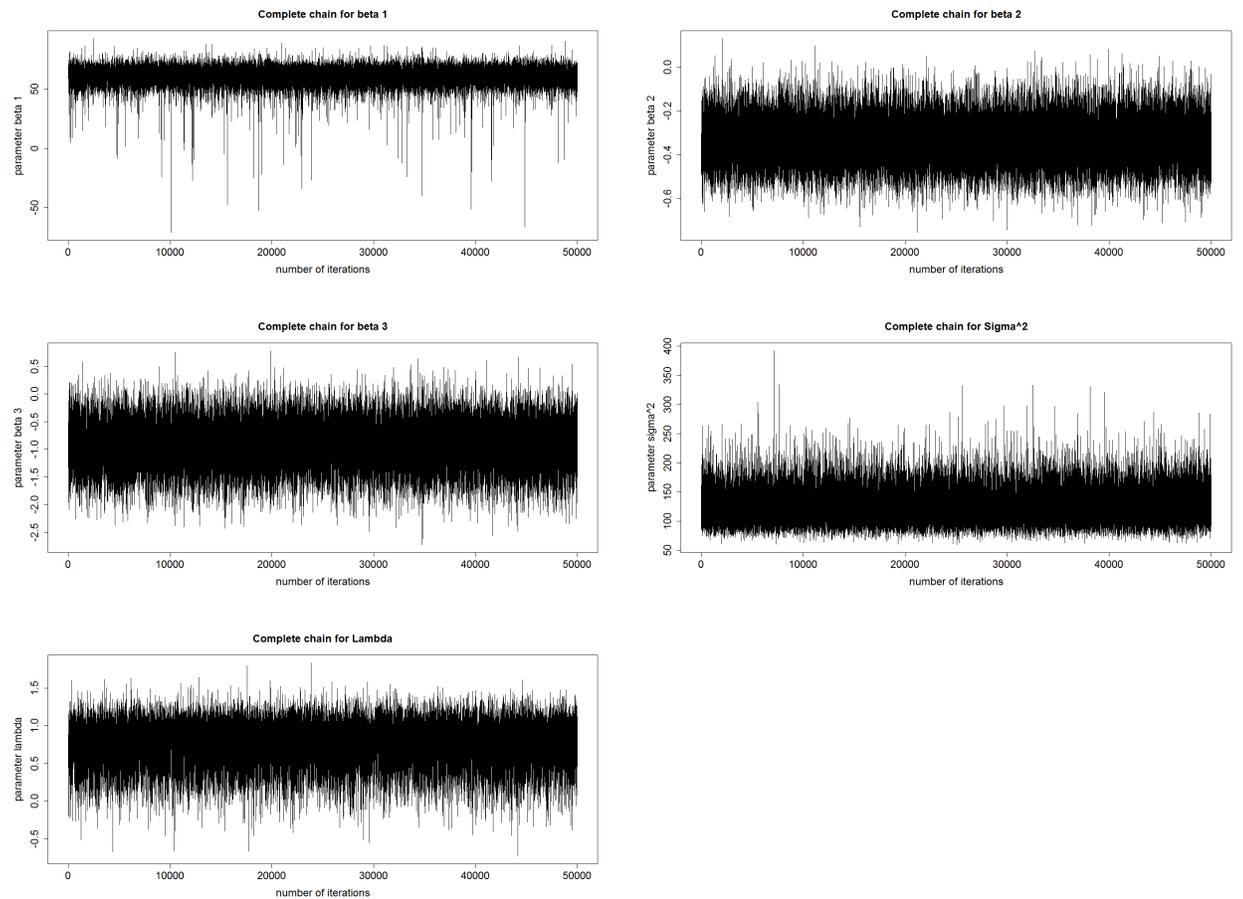
TABLA 2.7. Parámetros estimados del modelo SEM Homocedástico con kernel normal

	$\hat{\theta}_B$	E.E.	$\hat{\theta}_{MV}$	E.E.
$\beta_0$	60,484	8,068	59,893	5,366
$\beta_1$	-0,336	0,099	-0,302	0,090
$\beta_2$	-0,936	0,392	-0,941	0,331
$\sigma^2$	123,714	27,440	98,575	19,8737
$\lambda$	0,685	0,220	0,562	0,134
BIC	306,146		304,262	
DIC	1043,28			
% Aceptación	40,73			

En cuanto al ajuste del vector de parámetros  $\beta$ , las estimaciones puntuales son muy similares, a pesar de las diferencias de errores estándar. Para el ajuste de los parámetros  $\sigma^2$  y  $\lambda$ , las estimaciones presentan mayores diferencias, tanto en las estimaciones puntuales como en los errores estándar. Al comparar los valores del Criterio de Información Bayesiano, el ajuste Bayesiano arroja mejores resultados que el ajuste clásico. Hay que tener en cuenta que el ajuste por Máxima Verosimilitud presentó errores estándar más bajos, para todos los parámetros, que el realizado a través de métodos Bayesianos. Por último, la tasa de aceptación observada para el parámetro  $\lambda$  fue de 41 %.

Las cadenas del muestreo de la distribuciones condicionales a posteriori para cada uno de los parámetros del modelo, son presentadas en la figura 2.6.

FIGURA 2.6. Cadenas de los parámetros del modelo SEM homocedástico con kernel de transición normal



En las cadenas de Markov se encuentra, al igual que con la propuesta de kernel uniforme, que el algoritmo presenta una rápida convergencia. Sin embargo, hay una alta cantidad de valores que superan el umbral teórico de 1, lo que se refleja en una tasa de aceptación más baja para el ajuste con kernel de transición normal.

Los histogramas con los residuales calculados al ajustar el modelo SEM homocedástico mediante los kernels de transición uniforme y normal se muestran en las figuras 2.7 y 2.8, respectivamente. Las formas de ambos histogramas son muy similares, incluso en la presencia de valores alejados del centro de la distribución. No obstante, estos no tienen formas de campana apuntadas, a diferencia de lo observado para el caso del modelo SAR homocedástico. Ambas distribuciones parecen estar centradas en cero.

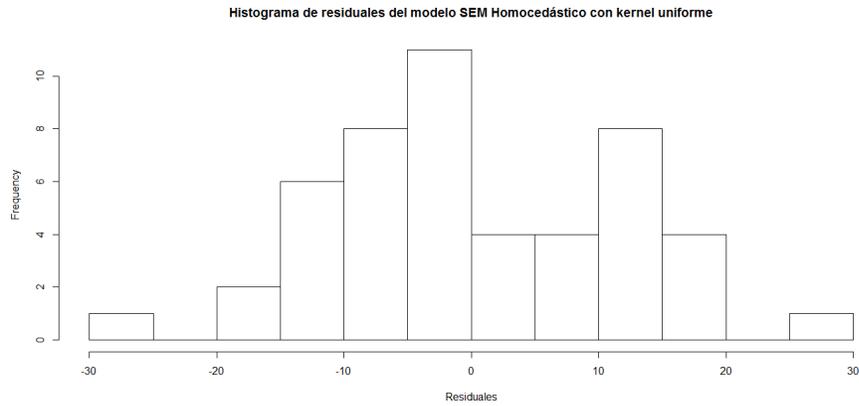


FIGURA 2.7. Residuales para el modelo SEM homocedástico ajustado con el kernel de transición uniforme para los datos de crimen en Columbus, [8]

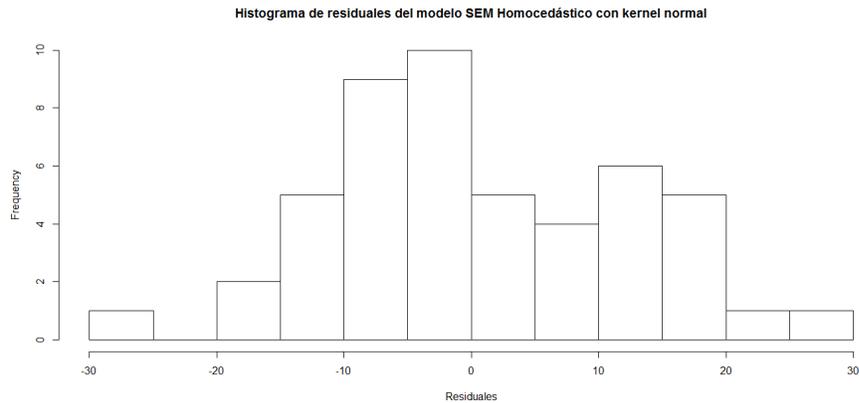


FIGURA 2.8. Residuales para el modelo SEM homocedástico ajustado con el kernel de transición normal para los datos de crimen en Columbus, [8]

Se evaluaron las medidas de tendencia de estos residuales, así como si siguen una distribución normal, los resultados se encuentran en la tabla 2.8.

TABLA 2.8. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SEM homocedástico bajo las dos propuestas

	Kernel Uniforme	Kernel Normal
Mínimo	-28,428	-27,166
Media	0,507	1,036
Mediana	-1,24	-0,619
Máximo	28,597	28,997
p-valor (Shapiro Test)	0,620	0,595

Se observa que las estimaciones logradas a partir de ambas metodologías no están centradas en cero a pesar de aparentemente provenir de distribuciones normales.

## 2.3. Modelo General Bayesiano: Homocedástico

### 2.3.1. Especificación

El modelo general homocedástico está dado por:

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}_1\mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda\mathbf{W}_2\boldsymbol{\varepsilon} + \boldsymbol{\mu}\end{aligned}$$

donde  $\boldsymbol{\mu} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ . Las distribuciones a priori de cada uno de los parámetros son:

1.  $p(\boldsymbol{\beta}) \sim N(\mathbf{b}, \mathbf{B})$
2.  $p(\sigma^2) \sim \text{GammaInv}(\zeta, \vartheta)$
3.  $p(\rho) \sim U(-1/\omega_{1_{max}}, 1)$
4.  $p(\lambda) \sim U(-1/\omega_{2_{max}}, 1)$

allí  $\omega_{1_{max}}$  y  $\omega_{2_{max}}$  corresponden a los mayores valores propios negativos, en valor absoluto, de las matrices de vecindad espacial,  $\mathbf{W}_1$  y  $\mathbf{W}_2$ , respectivamente, [8]. Por facilidad se asume independencia a priori entre los parámetros de este modelo. Los términos  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\rho$ ,  $\lambda$ ,  $\mathbf{W}_1$ ,  $\mathbf{W}_2$  y  $\boldsymbol{\varepsilon}$  son explicados en la sección 1.4.1.

Con las distribuciones a priori y la verosimilitud en (1.10), se obtiene la distribución a posteriori  $\pi$ , dada por:

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2, \rho, \lambda) &\propto (\sigma^2)^{-n/2} |\mathbf{A}| |\mathbf{D}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \frac{\zeta^\vartheta}{\Gamma(\vartheta)} (\sigma^2)^{-\vartheta-1} \exp \left( -\frac{\zeta}{\sigma^2} \right) \quad (2.5)\end{aligned}$$

Las siguientes distribuciones condicionales a posteriori son obtenidas al condicionar 2.5 para cada parámetro. Las demostraciones paso a paso se encuentran en el apéndice A. Hay que tener en cuenta que en este caso, las matrices  $\mathbf{A}$  y  $\mathbf{D}$  están dadas por  $\mathbf{I} - \rho\mathbf{W}_1\mathbf{y}$  y  $\mathbf{I} - \lambda\mathbf{W}_2\mathbf{y}$ , respectivamente:

1. **Distribución condicional a posteriori para  $\boldsymbol{\beta}$ , [22]**

La distribución condicional a posteriori,  $\pi(\boldsymbol{\beta}|\sigma^2, \rho, \lambda)$ , es una distribución normal  $N(\mathbf{b}^*, \mathbf{B}^*)$ , con:

$$\begin{aligned}\mathbf{b}^* &= \mathbf{B}^* \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \\ \mathbf{B}^* &= \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1}\end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

2. **Distribución condicional a posteriori para  $\sigma^2$**

La siguiente es la distribución a posteriori de  $\sigma^2$  dado  $(\boldsymbol{\beta}, \rho, \lambda)$ :

$$\pi(\sigma^2|\boldsymbol{\beta}, \rho, \lambda) \sim \text{GammaInv} \left( \zeta + \frac{n}{2}, \frac{k}{2} + \vartheta \right)$$

donde  $k = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{D}'\mathbf{D}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ ,  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

### 3. Distribución condicional a posteriori para $\rho$

Para el parámetro de rezago espacial  $\rho$ , se obtuvo la distribución  $\pi(\rho|\boldsymbol{\beta}, \sigma^2, \lambda)$ , dada por:

$$\pi(\rho|\boldsymbol{\beta}, \sigma^2, \lambda) \propto |\mathbf{A}| \exp \left\{ -\frac{1}{2 \times \frac{1}{c}} \left( \rho - \frac{d}{c} \right)^2 \right\} \mathbf{I}_{(-1/\omega_{1max}, 1)}(\rho) \quad (2.6)$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$ ,  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ ,  $c = \frac{1}{\sigma^2} \mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}$  y  $d = \frac{1}{\sigma^2} \mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

### 4. Distribución condicional a posteriori para $\lambda$

En el caso del parámetro  $\lambda$ , se obtuvo una distribución condicional a posteriori,  $\pi(\lambda|\boldsymbol{\beta}, \sigma^2, \rho)$ , dada por:

$$\pi(\lambda|\boldsymbol{\beta}, \sigma^2, \rho) \propto |\mathbf{D}| \exp \left\{ -\frac{1}{2 \times \frac{1}{c'}} \left( \rho - \frac{d'}{c'} \right)^2 \right\} \mathbf{I}_{(-1/\omega_{2max}, 1)}(\lambda) \quad (2.7)$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$ ,  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ ,  $c' = \frac{1}{\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2\mathbf{W}_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $d' = \frac{1}{\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

Dado que las distribuciones condicionales a posteriori, tanto de  $\rho$  como de  $\lambda$ , son analíticamente intratables, se realizarán dos propuestas para el ajuste de este modelo:

1. **Kernel de transición uniforme:** Una primera propuesta es hacer uso del muestreador de Metropolis-Hastings, [56], [59], empleando como kernels de transición, tanto para  $\rho$  como para  $\lambda$ , kernels uniformes en los intervalos  $(-1/\omega_{1max}, 1)$  y  $(-1/\omega_{2max}, 1)$ , respectivamente, donde  $\omega_{1max}$  y  $\omega_{2max}$  corresponden a los mayores valores propios negativos, en valor absoluto, de las matrices de vecindad espacial,  $\mathbf{W}_1$  y  $\mathbf{W}_2$ , respectivamente.
2. **Kernel de transición normal:** Partiendo de las distribuciones condicionales a posteriori de  $\rho$  y  $\lambda$ , (2.6) y (2.7), respectivamente, se proponen kernels de transición normal,  $N(c^*, d^*)$  y  $N(c'^*, d'^*)$ , truncados en los intervalos  $(-1/\omega_{1max}, 1)$  y  $(-1/\omega_{2max}, 1)$ , respectivamente, donde  $\omega_{1max}$  y  $\omega_{2max}$  corresponden a los mayores valores propios negativos, en valor absoluto, de las matrices de vecindades,  $\mathbf{W}_1$  y  $\mathbf{W}_2$ , respectivamente, [8]. Así se encuentra implementado en la función *hom\_general* del paquete *BSPADATA* de R, [54]. Los parámetros  $c^*$ ,  $d^*$ ,  $c'^*$  y  $d'^*$ , están dados por:

$$\begin{aligned} c^* &= \frac{\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}} & d^* &= \frac{\sigma^2}{\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}} \\ c'^* &= \frac{(\mathbf{A}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2(\mathbf{A}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{(\mathbf{A}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2\mathbf{W}_2(\mathbf{A}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} & d'^* &= \frac{\sigma^2}{(\mathbf{A}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2\mathbf{W}_2(\mathbf{A}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})} \end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

## 2.3.2. Simulación

Para evaluar el rendimiento de las dos propuestas realizadas para el ajuste del modelo General Homocedástico bajo distintas condiciones, se elaboró una simulación de 49 datos de una variable normal,  $\mathbf{y}$ , a partir del siguiente modelo:

$$\begin{aligned} y &= \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}_1\mathbf{y} + \varepsilon \\ \varepsilon &= \lambda\mathbf{W}_2\varepsilon + \boldsymbol{\mu} \end{aligned}$$

con  $\boldsymbol{\mu} \sim N(0, 45\mathbf{I})$ . Para este modelo  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3]$ ;  $\boldsymbol{\beta} = [18, 0.026, -0.4]$ ;  $\sigma^2 = 45$  y hay tres distintos valores para la pareja  $[\rho, \lambda]$ , estos son:  $[0.12; 0.48]$ ,  $[0.45; 0.65]$  y  $[0.75; 0.20]$ . Esto implica que la variable  $\mathbf{y}$  fue simulada con media  $(\mathbf{I} - \rho\mathbf{W}_1)^{-1}(18 + 0.026\mathbf{x}_1 - 0.4\mathbf{x}_2)$  y matriz de varianzas y covarianzas dada por  $[(\mathbf{I} - \rho\mathbf{W}_1)^{-1}(\mathbf{I} - \lambda\mathbf{W}_2)^{-1}]'\boldsymbol{\Sigma}[(\mathbf{I} - \rho\mathbf{W}_1)^{-1}(\mathbf{I} - \lambda\mathbf{W}_2)^{-1}]$ , donde  $\boldsymbol{\Sigma}$  es una matriz diagonal con  $\boldsymbol{\Sigma}_{i,i} = 45$ ,  $i = 1, \dots, 49$ .

Por su parte, las realizaciones de las variables  $\mathbf{X}_1$  y  $\mathbf{X}_2$  fueron generadas a partir de distribuciones uniformes en los intervalos  $(0, 400)$  y  $(10, 23)$ , respectivamente. Finalmente, las matrices de vecindades  $\mathbf{W}_1$  y  $\mathbf{W}_2$  se asumen iguales y corresponden a la matriz utilizada en la aplicación de datos de crimen en Columbus, Ohio por [8].

Posteriormente, se ajustaron los datos con las dos metodologías propuestas y mediante Máxima Verosimilitud. Los resultados son presentados en la tabla 2.9.

TABLA 2.9. Comparación de ajustes del modelo general homocedástico

$\rho$ real	$\lambda$ real	Método	Valor	$\beta_0$	$\beta_1$	$\beta_2$	$\rho$	$\lambda$	$\sigma^2$	BIC	DIC	
0,12	0,48	Kernel Uniforme	Estimación	7,643	0,036	-0,369	0,658	0,660	36,859	258,369	899,590	
			Desviación estándar	5,412	0,008	0,243	0,197	0,195	8,226			
		Kernel Normal	Estimación	14,720	0,029	-0,187	-0,534	0,965	31,284	262,911	906,309	
			Desviación estándar	22,306	0,007	0,212	0,368	0,038	6,871			
		Máxima Verosimilitud	Estimación	3,621	0,025	-0,122	0,588	-0,505	40,012	350,539		
			Desviación estándar	5,151	0,009	0,218	0,246	0,424	6,326			
0,45	0,65	Kernel Uniforme	Estimación	14,934	0,029	-0,192	0,662	0,660	80,050	296,289	1017,677	
			Desviación estándar	7,675	0,011	0,354	0,198	0,196	17,724			
		Kernel Normal	Estimación	36,749	0,023	-0,199	-0,520	0,970	63,706	301,149	1011,966	
			Desviación estándar	24,701	0,009	0,298	0,304	0,038	14,322			
		Máxima Verosimilitud	Estimación	28,216	0,033	-0,125	-0,364	0,802	55,568	371,454		
			Desviación estándar	11,080	0,009	0,288	0,322	0,120	7,454			
0,75	0,20	Kernel Uniforme	Estimación	34,780	0,023	-0,349	0,663	0,665	47,140	290,932	1067,967	
			Desviación estándar	8,311	0,008	0,306	0,194	0,196	10,373			
		Kernel Normal	Estimación	5,452	0,013	-0,131	-1,315	0,996	27,532	281,0626	949,748	
			Desviación estándar	31,946	0,006	0,234	0,265	0,004	6,077			
		Máxima Verosimilitud	Estimación	31,146	0,029	-0,351	0,519	0,398	37,657	345,717		
			Desviación estándar	23,344	0,007	0,229	0,360	0,442	6,137			

- Cuando  $\rho = 0,12$  y  $\lambda = 0,48$ , para los parámetros del vector  $\boldsymbol{\beta}$  la propuesta Bayesiana con kernel de transición normal es la más aproximada. A pesar de esto, vale la pena resaltar que el error estándar del intercepto es mayor al obtenido con las otras dos propuestas. En cuanto a los parámetros de rezago espacial, el estimador por Máxima Verosimilitud es el más acertado, sin ser cercano a los valores reales, ya que la propuesta normal toma valores muy cercanos a las fronteras de los parámetros y el kernel uniforme arroja valores muy próximos a 0.66. Finalmente, la estimación de  $\sigma^2$  es más proxima al emplear la propuesta clásica. Para comparar las dos propuestas Bayesianas se empleó el Deviance Information Criteria (DIC). En este caso señala que se obtienen mejores resultados a través del kernel de transición uniforme. En cuanto a tasas de aceptación, es más apropiado el ajuste empleando el kernel de transición normal.
- Al revisar el escenario en que  $\rho = 0,45$  y  $\lambda = 0,65$ , el ajuste por Máxima Verosimilitud ofrece los mejores resultados para el vector  $\boldsymbol{\beta}$  en términos de cercanía al verdadero valor. No obstante, ninguna de las tres metodología ofrece una buena aproximación para  $\rho$  y  $\lambda$ . Por último, la estimación de  $\sigma^2$  a través del kernel de transición uniforme es muy alta si se compara con el verdadero valor y la estimaciones de las otras dos metodologías, las cuales son más próximas. A pesar de tener mayor BIC, el ajuste Bayesiano hecho con el kernel de transición normal, es aparentemente más

apropiado que el realizado con el kernel de transición uniforme, puesto que tiene menor DIC y mayores tasas de aceptación para  $\rho$  y  $\lambda$ .

- Finalmente, cuando  $\rho = 0,75$  y  $\lambda = 0,20$ , las estimaciones por Máxima Verosimilitud del vector  $\beta$ , los parámetros de rezago espacial y  $\sigma^2$  son las más apropiadas. Entre las propuestas Bayesianas, la que aporta mejores resultados es la propuesta hecha con el kernel de transición normal.

Comparando las tasas de aceptación para las propuestas Bayesianas, se encuentra que al emplear el kernel de transición normal se aseguran mayores valores. El resultado de estas simulaciones permite concluir que ninguna de las metodologías trabajadas realiza aproximaciones tan exactas como para los otros dos modelos, esto a raíz de la complejidad en su proceso de estimación.

### 2.3.3. Aplicación: Datos de crimen en Columbus

Con el objetivo de implementar esta propuesta metodológica en el modelo General homocedástico, se trabajó con los datos presentados en [8], relacionados con el nivel de crimen en 49 distritos de la ciudad de Columbus, Ohio. La descripción de las variables presentes en esta base de datos se encuentra en la sección 2.1.3.

El modelo ajustado en este caso está dado por:

$$\begin{aligned} CRIME &= \mathbf{X}\beta + \rho\mathbf{W}_1\mathbf{CRIME} + \varepsilon \\ \varepsilon &= \lambda\mathbf{W}_2\varepsilon + \mu \end{aligned}$$

con  $\mathbf{X} = [\mathbf{1}; HOUSE; CRIME]$ , donde  $\mathbf{1}$  es un vector de unos de longitud  $n = 49$ . Se asumió  $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$ , la cual es una matriz de contigüidad binaria y estandarizada por filas, es decir, si dos unidades espaciales  $i$  y  $j$  son vecinas se asigna un valor diferente a 0 en la entrada  $[i, j]$ . Además, el hecho de ser estandarizada por filas implica que la suma de sus filas es igual a 1.

Como a priori se postuló un vector de ceros para la media de  $\beta$  y una matriz de varianzas y covarianzas a priori diagonal con 1000 en los términos de la diagonal. Así mismo, se propuso una distribución a priori gamma inversa con parámetros 0.01 y 0.01. Se realizaron 50000 simulaciones, con un calentamiento de 2500 valores, además, con el fin de minimizar el efecto de correlación entre las simulaciones, se escogieron valores cada 10 simulaciones.

#### 2.3.3.1. Resultados: Ajuste con Kernel de Transición Uniforme

A continuación se describen los resultados del proceso de estimación para cada uno de los parámetros del modelo general homocedástico. Para comparar el presente trabajo frente a las metodologías clásicas, se exponen los resultados descritos en [8] y se evalúan las dos ideas mediante el Criterio de Información Bayesiano (BIC), [50].

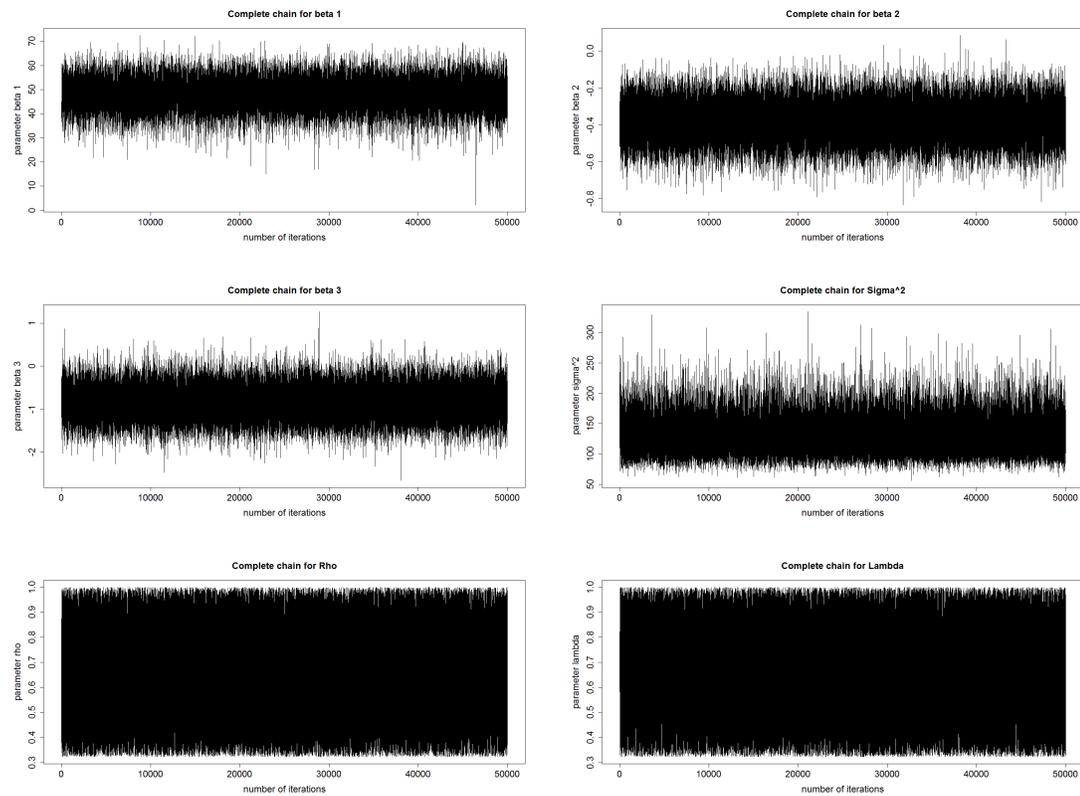
TABLA 2.10. Parámetros estimados del modelo general con kernel uniforme

	$\hat{\theta}_B$	E.E.	$\hat{\theta}_{MV}$	E.E.
$\beta_0$	48,134	6,016	47,784	9,903
$\beta_1$	-0,382	0,100	-0,282	0,090
$\beta_2$	-0,850	0,370	-1,026	0,326
$\sigma^2$	129,757	28,134	95,604	19,475
$\rho$	0,664	0,206	0,091	0,197
$\lambda$	0,663	0,279	0,167	0,297
BIC	329,801		342,849	
DIC	1113,841			
% Aceptación $\rho$	20,61			
% Aceptación $\lambda$	27,91			

Para el ajuste con kernel uniforme, se observa que la estimación del vector  $\beta$  es muy similar a la obtenida mediante Máxima Verosimilitud. Por otra parte, existen diferencias por encima de una desviación estándar para el parámetro  $\sigma^2$  y los parámetros de rezago espacial entre las dos propuestas. En términos de error estándar se observa que si bien ambos métodos generan errores estándar similares, el ajuste por Máxima Verosimilitud ofrece mejores resultados. En lo referente al Criterio de Información Bayesiano (BIC), con la propuesta de ajuste Bayesiano se obtiene un menor valor de este criterio. Finalmente, se observa una tasa de aceptación para el muestreo de los parámetros  $\rho$  y  $\lambda$ , del 20 % y del 27 %, respectivamente.

Ahora se presentan las cadenas de cada uno de los parámetros estimados:

FIGURA 2.9. Cadenas de los parámetros del modelo general homocedástico con kernel de transición uniforme



Al revisar las cadenas de Markov se infiere que el algoritmo propuesto conlleva a una rápida convergencia de los parámetros. Sin embargo, es evidente que tanto las estimaciones de  $\rho$ , como las de  $\lambda$ , oscilan entre los mismos valores, generando estimaciones muy similares y valores erróneos para los demás parámetros.

### 2.3.3.2. Resultados: Ajuste con Kernel de Transición Normal

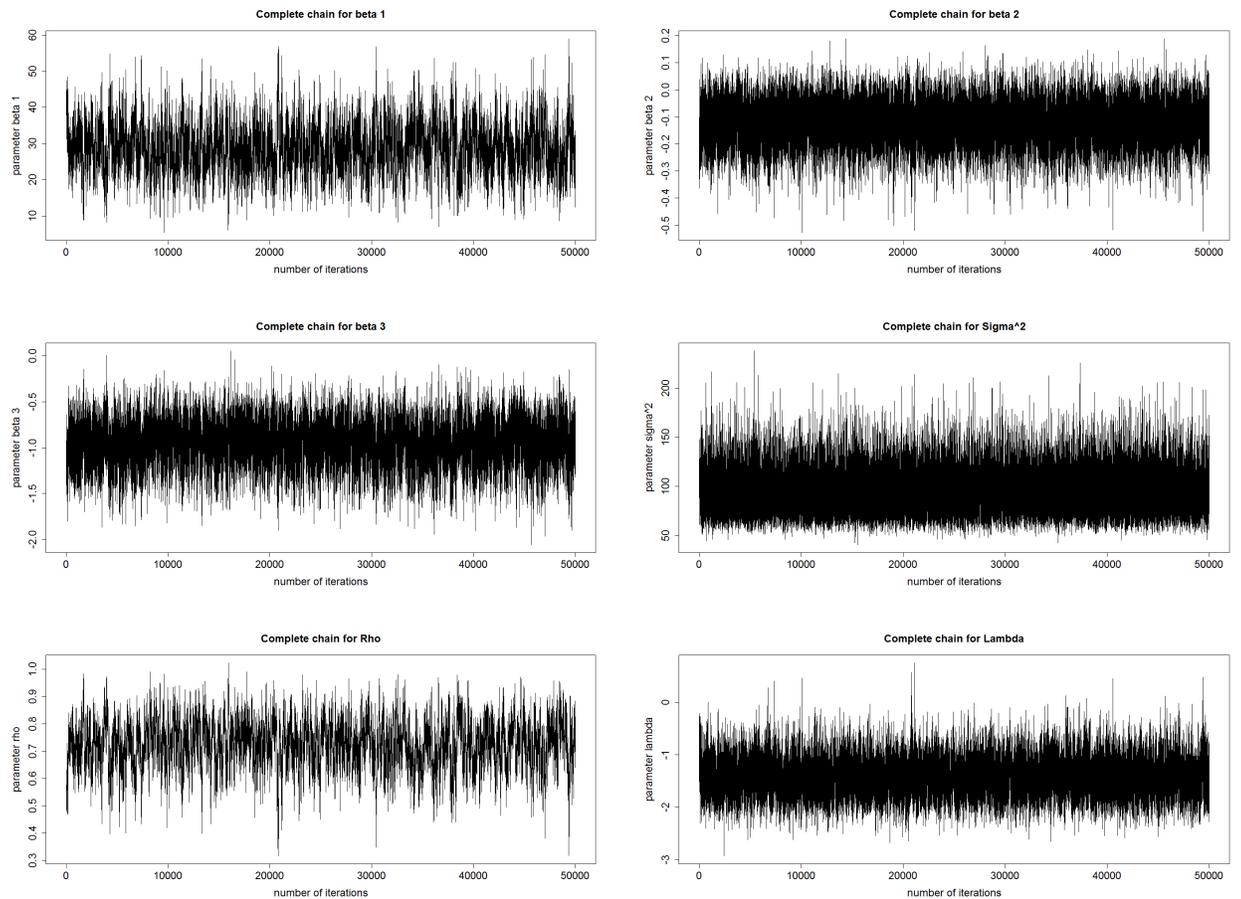
En la tabla 2.11, se presentan las estimaciones para cada uno de los parámetros y su respectivo error estándar. Nuevamente se compara el presente trabajo frente a las metodologías clásicas, presentadas en [8], y se comparan las dos ideas mediante el Criterio de Información Bayesiano (BIC), [50].

TABLA 2.11. Parámetros estimados del Modelo General con kernel normal

	$\hat{\theta}_B$	E.E.	$\hat{\theta}_{MV}$	E.E.
$\beta_0$	28,676	6,640	47,784	9,903
$\beta_1$	-0,125	0,076	-0,282	0,090
$\beta_2$	-0,963	0,232	-1,026	0,326
$\sigma^2$	93,504	20,753	95,604	19,475
$\rho$	0,726	0,087	0,091	0,197
$\lambda$	-1,408	0,335	0,167	0,297
BIC	321,595		342,849	
DIC	1069,836			
% Aceptación $\rho$	73,77			
% Aceptación $\lambda$	71,11			

Se encuentra que las estimaciones del vector  $\beta$  y el parámetro  $\sigma^2$  son muy similares, a excepción del intercepto. Por otra parte, existen diferencias por encima de una desviación estándar entre el estimador de Máxima Verosimilitud y el estimador Bayesiano para  $\rho$  y  $\lambda$ . En términos de error estándar, el ajuste Bayesiano tiene errores estándar menores a los presentados con el ajuste clásico, excepto para  $\sigma^2$  y  $\lambda$ . Adicionalmente, el Criterio de Información Bayesiano (BIC) de la propuesta de ajuste Bayesiano aporta mejores resultados. Finalmente, se observan buenas tasas de aceptación para el muestreo de los parámetros  $\rho$  y  $\lambda$ , las cuales están alrededor del 70 %.

FIGURA 2.10. Cadenas de los parámetros del modelo general homocedástico con kernel de transición normal



Las cadenas para los parámetros de este modelo alcanzan la convergencia de manera rápida. No obstante, se observa, especialmente para  $\beta_0$  y  $\rho$ , alta variabilidad.

El ajuste de este conjunto de datos mediante el modelo general homocedástico y los kernels de transición uniforme y normal arroja unos residuales que pueden ser observados en las figuras 2.11 y 2.12. Al igual que para el modelo SAR homocedástico, se observa que el ajuste con el kernel de transición uniforme no está centrado en cero y una distribución poco apuntada, mientras que los residuales obtenidos al ajustar con ayuda del kernel de transición normal están aparentemente centrados en cero.

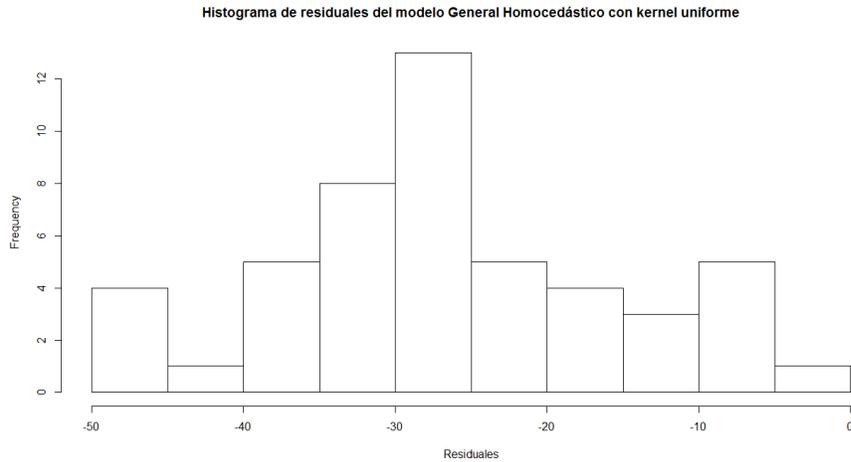


FIGURA 2.11. Residuales para el modelo general homocedástico ajustado con el kernel de transición uniforme para los datos de crimen en Columbus, [8]

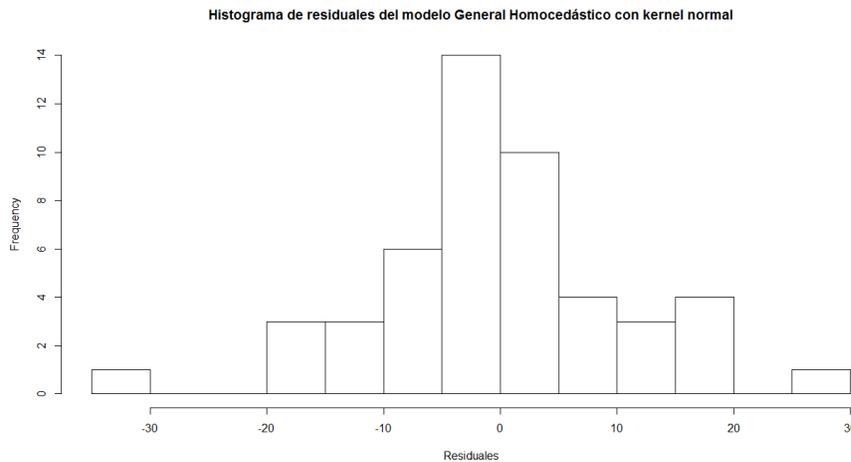


FIGURA 2.12. Residuales para el modelo general homocedástico ajustado con el kernel de transición normal para los datos de crimen en Columbus, [8]

Se evaluaron las medidas de tendencia central de estos residuales, así como si siguen una distribución normal, los resultados se encuentran en la tabla 2.12.

TABLA 2.12. Medidas resumen y test de normalidad para los residuales del ajuste del modelo general homocedástico bajo las dos propuestas

	Kernel Uniforme	Kernel Normal
Mínimo	-48,506	-32,489
Media	-26,343	-0,335
Mediana	-27,826	-0,638
Máximo	-3,142	25,043
p-valor (Shapiro Test)	0,396	0,378

Estos resultados muestran que las estimaciones logradas a partir del kernel de transición uniforme no están centradas en cero a pesar de aparentemente provenir de una distribución normal. Para el caso normal la distribución de los residuales parece ser normal y centrada en cero, a pesar de un conjunto de residuales aparentemente atípicos.

---



---

## Modelos Econométricos Espaciales Bayesianos Heterocedásticos

---



---

Ahora se abordará el caso en que la matriz de varianzas y covarianzas del modelo general, (1.1),  $\mathbf{\Omega}$ , es una matriz diagonal con término general dado por  $\Omega_{ii} = g(\mathbf{Z}\boldsymbol{\gamma})$ . Aquí  $\mathbf{Z}$  es una matriz de covariables,  $\boldsymbol{\gamma}$  es un vector de parámetros y  $g$  es una función apropiadamente definida.

A continuación se expondrán propuestas Bayesianas para el ajuste de los modelos SAR, SEM y General, [8], en su versión heterocedástica, basadas en los algoritmos MCMC, [33], como en el muestreador de Gibbs, [21] y [55], y el algoritmo de Metropolis Hastings, [56], [59], y en la construcción de variables de trabajo, [22], [24].

Cada modelo será formulado desde una perspectiva Bayesiana. Se obtendrán las funciones de distribución condicionales a posteriori para todos los parámetros, y posteriormente se propondrán alternativas para obtener muestras de aquellas funciones condicionales a posteriori con formas desconocidas. La fase final de cada sección se compone de una aplicación de las metodologías expuestas en datos de leucemia en Nueva York, que se encuentra en [62], y de una simulación con una variable normal en la que se evalúan las ideas de este trabajo.

### 3.1. Modelo SAR Bayesiano: Heterocedástico

#### 3.1.1. Especificación

El modelo SAR se obtiene a partir del modelo general heterocedástico haciendo  $\lambda = 0$ . Es decir, el modelo está dado por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}$$

donde  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{\Omega})$ , con  $\mathbf{\Omega}$  una matriz diagonal con término general dado por  $g(\Omega_{ii}) = \mathbf{Z}\boldsymbol{\gamma}$ , con  $g$  una función apropiadamente definida,  $\mathbf{Z}$  la matriz que incluye las covariables que explican la dispersión y  $\boldsymbol{\gamma}$  su vector de coeficientes. Además, los términos  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\rho$ ,  $\mathbf{W}$  y  $\boldsymbol{\varepsilon}$  siguen la descripción dada en la sección 1.4.1. La definición del modelo se completa con las siguientes distribuciones a priori:

1.  $p(\boldsymbol{\beta}) \sim N(\mathbf{b}, \mathbf{B})$
2.  $p(\boldsymbol{\gamma}) \sim N(\mathbf{g}, \mathbf{G})$
3.  $p(\rho) \sim U(-1/\omega_{max}, 1)$

donde  $\omega_{max}$  es el mayor valor propio negativo, en valor absoluto, de la matriz de vecindad espacial,  $\mathbf{W}$ , [8]. Por facilidad, se asume que estas distribuciones a priori son independientes.

Con en el modelo especificado, la distribución a posteriori,  $\pi$ , está dada por:

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho) \propto |\boldsymbol{\Sigma}|^{-1/2} |\mathbf{A}| \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})' \mathbf{G}^{-1} (\boldsymbol{\gamma} - \mathbf{g}) \right\} \quad (3.1)$$

donde  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}$ . A partir de (3.1) se obtienen las siguientes distribuciones condicionales. La derivación de estas distribuciones se encuentra en el apéndice A.

#### 1. Distribución condicional a posteriori de $\boldsymbol{\beta}$ , [22]

Para el vector de parámetros  $\boldsymbol{\beta}$  se obtiene una distribución normal  $\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \rho) \sim N(\mathbf{b}^*, \mathbf{B}^*)$  con:

$$\begin{aligned} \mathbf{b}^* &= \mathbf{B}^* (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b}) \\ \mathbf{B}^* &= (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{B}^{-1})^{-1} \end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}$ .

#### 2. Distribución condicional a posteriori de $\boldsymbol{\gamma}$

La siguiente es la distribución condicional obtenida para el vector  $\boldsymbol{\gamma}$ ,  $\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \rho)$ :

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \rho) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})' \mathbf{G}^{-1} (\boldsymbol{\gamma} - \mathbf{g}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}$ .

#### 3. Distribución condicional a posteriori de $\rho$

La siguiente es la distribución obtenida al condicionar 3.1 respecto a  $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ :

$$\pi(\rho|\boldsymbol{\beta}, \boldsymbol{\gamma}) \propto |\mathbf{A}| \exp \left\{ -\frac{1}{2 \times \frac{1}{c}} \left( \rho - \frac{d}{c} \right)^2 \right\} I_{(-1/\omega_{max}, 1)}(\rho)$$

con  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}$ ,  $c = \mathbf{y}' \mathbf{W}' \boldsymbol{\Sigma}^{-1} \mathbf{W} \mathbf{y}$  y  $d = \mathbf{y}' \mathbf{W}' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

Al revisar las distribuciones condicionales a posteriori de  $\boldsymbol{\gamma}$  y  $\rho$  se encuentra que no tienen formas funcionales conocidas, y por lo tanto no resultan sencillas de generar. Es por esto que las siguientes propuestas son realizadas para muestrear estos parámetros.

##### 3.1.1.1. Propuesta para $\boldsymbol{\gamma}$ , [22], [24]

Para obtener muestras de  $\boldsymbol{\gamma}$  se propone un kernel de transición basado en [22] y [24], donde se construyen propuestas tomando como referencia el algoritmo de Fisher Scoring,

[52].

En este caso las variables para aproximar la transformación de las observaciones alrededor de las estimaciones del parámetro parten del modelo observado:

$$t_i = (y_i - \rho^{(c)} \mathbf{W}'_i \mathbf{y} - x'_i \boldsymbol{\beta}^{(c)})^2 \sim \sigma_i^2 \chi_1^2 \quad \text{para } i = 1, 2, \dots, n$$

Estas variables tienen media  $E(t_i) = \sigma_i^2$ , varianza  $Var(t_i) = 2\sigma_i^4$  y cumplen  $g[E(t_i)] = z'_i \boldsymbol{\gamma}$ .

Como  $g$  es una función diferenciable, se puede emplear la aproximación lineal alrededor de  $E(t_i)$ :

$$g(t_i) \simeq g[E(t_i)] + g'[E(t_i)][t_i - E(t_i)] = \tilde{y}_i$$

y

$$Var[g(t_i)] = (g'[E(t_i)])^2 Var(t_i)$$

en alguna vecindad de  $E(t_i) = \sigma_i^2$ . Luego, con  $\boldsymbol{\beta}^{(c)}$  y  $\boldsymbol{\gamma}^{(c)}$  los valores actuales de  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$ , el valor apropiado de la observación es:

$$\tilde{y}_i = z'_i \boldsymbol{\gamma}^{(c)} + g'[g^{-1} z'_i \boldsymbol{\gamma}^{(c)}][(y_i - \rho^{(c)} \mathbf{W}'_i \mathbf{y} - x'_i \boldsymbol{\beta}^{(c)})^2 - g^{-1}(z'_i \boldsymbol{\gamma}^{(c)})]$$

con varianza dada por:

$$Var[\tilde{y}_i] = 2\{g'[(z'_i \boldsymbol{\gamma}^{(c)})]g^{-1}(z'_i \boldsymbol{\gamma}^{(c)})\}^2$$

Si  $g = \log$ , el vector de las observaciones sería  $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$ :

$$\tilde{y}_i = z'_i \boldsymbol{\gamma}^{(c)} + \frac{(y_i - \rho^{(c)} \mathbf{W}'_i \mathbf{y} - x'_i \boldsymbol{\beta}^{(c)})^2}{\exp(z'_i \boldsymbol{\gamma}^{(c)})} - 1, \quad i = 1, 2, \dots, n$$

con varianza igual a 2.  $\tilde{y}_i$  será en este caso la variable de trabajo para generar muestras de  $\boldsymbol{\gamma}$ . Por lo tanto, el kernel de transición obtenido como la mixtura entre la distribución a priori de  $\boldsymbol{\gamma}$  y la distribución de la variable de trabajo, está dado por:

$$q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(c)}, \boldsymbol{\gamma}^{(n)}) = N(\mathbf{g}^*, \mathbf{G}^*)$$

donde  $\mathbf{g}^* = \mathbf{G}^*(\mathbf{G}^{-1}\mathbf{g} + 0.5\mathbf{Z}'\tilde{\mathbf{Y}})$  y  $\mathbf{G}^* = (\mathbf{G}^{-1} + 0.5\mathbf{Z}'\mathbf{Z})^{-1}$

### 3.1.1.2. Propuesta para $\rho$

Para obtener muestras de  $\rho$ , las ideas respecto al kernel de transición corresponden a las mismas del caso homocedástico.

- **Kernel de transición uniforme:** Una primera propuesta es hacer uso del muestreador de Metropolis-Hastings, [56], [59], con kernel de transición uniforme para  $\rho$  en el intervalo  $(-1/\omega_{max}, 1)$ , donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8].
- **Kernel de transición normal:** Teniendo en cuenta la distribución condicional a posteriori de  $\rho$ , calculada en el apéndice A, se propone un kernel de transición normal  $N(c^*, d^*)$ , donde:

$$c^* = \frac{\mathbf{y}'\mathbf{W}'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\mathbf{y}'\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}} \qquad d^* = \frac{1}{\mathbf{y}'\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}}$$

Esta distribución normal es truncada en el intervalo  $(-1/\omega_{max}, 1)$ , donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8]. De esta manera se encuentra implementado en la función *hetero\_sar* del paquete *BSPADATA* de R, [54].

### 3.1.2. Simulación

Se elaboró una simulación de 49 datos de una variable normal,  $\mathbf{y}$ , generada a partir del siguiente modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}$$

con  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ . Para este modelo  $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2]$ ,  $\boldsymbol{\beta} = [-35, 0.35, -1.7]$  y la matriz  $\boldsymbol{\Omega}$  es diagonal y está dada por:

$$\log(\boldsymbol{\Omega}_{ii}) = \mathbf{Z}\boldsymbol{\gamma} \quad i = 1, \dots, 49$$

con  $\mathbf{Z} = [\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_3]$  y  $\boldsymbol{\gamma} = [-8, 0.026, -0.4]$ . Esto implica que la variable  $\mathbf{y}$  fue simulada con media  $(\mathbf{I} - \rho\mathbf{W})^{-1}(-35\mathbf{x}_0 + 0.35\mathbf{x}_1 - 1.7\mathbf{x}_2)$  y matriz de varianzas y covarianzas dada por  $[(\mathbf{I} - \rho\mathbf{W})^{-1}]'\boldsymbol{\Omega}[(\mathbf{I} - \rho\mathbf{W})^{-1}]$ .

Por su parte, la variable  $\mathbf{X}_0$  es un vector de unos de longitud 49 y las realizaciones de las variables  $\mathbf{X}_1$ ,  $\mathbf{X}_2$  y  $\mathbf{X}_3$  son generadas a partir de distribuciones uniformes, en los intervalos  $(0, 400)$ ,  $(10, 23)$  y  $(0, 10)$ , respectivamente. Finalmente, la matriz de vecindades  $\mathbf{W}$  es la misma utilizada en la aplicación de datos de crimen en Columbus, Ohio por [8].

Se hicieron simulaciones para tres valores de  $\rho$ : 0.12, 0.52 y 0.75. Los datos fueron ajustados mediante las dos metodologías, y los resultados se encuentran en la tabla 3.1.

TABLA 3.1. Comparación de ajustes del modelo SAR Heterocedástico

$\rho$ verdadero	Método	Valor	$\beta_0$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\rho$	BIC	DIC
0,12	Kernel Uniforme	Estimación	-40,195	0,345	-1,479	2,952	-0,002	0,098	0,662	219,863	1565,629
		Desviación estándar	3,065	0,006	0,069	0,604	0,002	0,069	0,199		
	Kernel Normal	Estimación	-34,975	0,350	-1,700	-9,201	0,026	-0,226	0,120	-155,610	-369,127
		Desviación estándar	0,020	0,002	0,001	0,616	0,002	0,073	0,000		
0,52	Kernel Uniforme	Estimación	-35,055	0,350	-1,695	-5,455	0,018	-0,193	0,661	27,978	4165024
		Desviación estándar	0,142	0,000	0,006	0,995	0,008	0,290	0,197		
	Kernel Normal	Estimación	-35,021	0,350	-1,696	-5,786	0,016	-0,094	0,516	-73,282	-90,000
		Desviación estándar	0,069	0,003	0,005	0,597	0,000	0,106	0,001		
0,75	Kernel Uniforme	Estimación	-34,807	0,350	-1,710	-4,464	0,012	0,024	0,645	-14,074	193016,6
		Desviación estándar	0,174	0,000	0,010	0,693	0,002	0,094	0,195		
	Kernel Normal	Estimación	-34,806	0,350	-1,710	-4,541	0,012	0,031	0,749	-16,782	49,117
		Desviación estándar	0,175	0,000	0,010	0,680	0,002	0,094	0,001		

- Cuando  $\rho = 0.12$  se encuentra que la estimación hecha con el kernel de transición uniforme presenta diferencias respecto a los parámetros verdaderos. Además presenta errores estándar muy altos si se compara con los resultados del ajuste con el kernel normal, los cuales reflejan que los parámetros son recuperados de manera eficiente. Por otra parte, el Criterio de Información Bayesiano (BIC) y el Deviance Information Criteria (DIC) aportan mejores resultados para el ajuste con kernel de transición normal, así como mayores tasas de aceptación para los parámetros  $\boldsymbol{\gamma}$  y  $\rho$ , los cuales están alrededor de 50 % y 75 %, respectivamente, mientras que con la otra alternativa fueron del 45 % y 2 %.

- La segunda situación sucede cuando  $\rho = 0.52$ , allí la estimación realizada a través del kernel de transición uniforme tiene el mismo problema del caso anterior, puesto que la estimación de  $\rho$  no se aleja de 0.66, lo que propicia diferencias en las estimaciones de los demás parámetros y mayores errores estándar. La simulación realizada empleando el kernel de transición normal arroja mejores resultados, puesto que son próximos a los verdaderos valores, los errores estándar son menores y tanto el Criterio de Información Bayesiano (BIC) como el Deviance Information Criteria (DIC) son menores.
- Por último, con un valor de  $\rho = 0.75$  las estimaciones hechas mediante ambas propuestas son muy similares, al igual que los errores estándar. No obstante, la propuesta realizada con el kernel de transición normal es mejor al aportar mayores tasas de aceptación y un valor más próximo a la realidad de  $\rho$ .

### 3.1.3. Aplicación: Datos de Leucemia en Nueva York

Para aplicar y evaluar la propuesta metodológica de esta sección se utilizaron los datos presentados en [62], los cuales constan de las siguientes variables:

- **PROPLEU**: Proporción de pacientes con leucemia en la población en riesgo. Esta proporción fue transformada en [62] para poder emplear modelos donde se suponen términos de error normales.
- **PCTAGE65P**: Porcentaje de personas en cada unidad espacial con 65 años de edad o más.
- **PCTOWNHOME**: Porcentaje de personas en cada unidad espacial con vivienda propia.
- **PEXPOSURE**: Exposición potencial definida como la distancia inversa entre cada unidad espacial y el sitio más cercano donde se encuentre tricloroetileno.

Estos datos corresponden al conjunto de datos *NY\_data* disponibles en la librería *spdep* de R, [54]. El modelo a ajustar está dado por:

$$PROPLEU = \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}PROPLEU + \boldsymbol{\varepsilon}$$

con  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ , donde  $\boldsymbol{\Omega}$  es una matriz diagonal con término general de la diagonal dado por  $\boldsymbol{\Omega}_{ii} = \exp(\gamma_0 + \gamma_1 PEXPOSURE_i)$ ,  $i = 1, \dots, n$ . Además,  $\mathbf{X} = [\mathbf{1}; PCTAGE65P; PCTOWNHOME]$ , donde  $\mathbf{1}$  es un vector de unos de longitud  $n = 281$ . La matriz de vecindad espacial  $\mathbf{W}$  es obtenida a partir del objeto *listw\_NY* del conjunto de datos *NY\_data* de R, [54]. Es un matriz de contigüidad binaria, estandarizada por filas, es decir, si un par de unidades espaciales,  $i$  y  $j$ , son vecinas, entonces la entrada  $[i, j]$  de la matriz es diferente de 0 y se garantiza que la suma de cada fila de la matriz es 1.

En esta situación se propuso un vector de ceros y una matriz diagonal con elementos diagonales iguales a 10000, como valores a priori para la media y la matriz de varianzas y covarianzas de  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$ , respectivamente. Se generaron 50000 simulaciones, con calentamiento de 2500 valores y selección de observaciones cada 10 simulaciones, con el objetivo de reducir el efecto de correlación entre muestras sucesivas.

Los resultados del ajuste de este modelo se encuentran en la tabla 3.2, allí se encuentran las estimaciones puntuales de cada parámetro acompañado de su error estándar, el Criterio de Información Bayesiano, BIC, [50] y las tasas de aceptación de los parámetros  $\gamma$  y  $\rho$ . Para efectos de comparación, se presentan de manera conjunta los resultados obtenidos mediante el kernel de transición uniforme y el kernel de transición normal.

TABLA 3.2. Parámetros estimados del modelo SAR heterocedástico

	Kernel Uniforme	SD Kernel Uniforme	Kernel Normal	SD Kernel Normal
$\beta_0$	-0,395	0,141	-0,397	0,143
$\beta_1$	3,520	0,592	3,622	0,617
$\beta_2$	-0,367	0,167	-0,418	0,171
$\gamma_0$	-0,619	0,160	-0,594	0,164
$\gamma_1$	-0,152	0,076	-0,164	0,078
$\rho$	0,655	0,203	0,298	0,099
BIC	66,427		64,903	
DIC	1248,819		1059,569	
% Aceptación $\gamma$	50,24		50,92	
% Aceptación $\rho$	10,54		60,65	

Las estimaciones puntuales de los vectores de parámetros  $\beta$  y  $\gamma$  presentan similitudes bajo las dos propuestas. No obstante, hay diferencias significativas entre las estimaciones del parámetro de rezago espacial,  $\rho$ , ya que en la propuesta uniforme este no se aleja del punto medio del intervalo  $(-1/\omega_{max}, 1)$  donde  $\omega_{max}$  es el mayor valor propio, en valor absoluto, de la matriz  $\mathbf{W}$ , [8].

Por otra parte, se observa que los errores estándar obtenidos bajo el kernel de transición normal son muy similares a los obtenidos con el uniforme a excepción del parámetro  $\rho$ , el cual en el modelo estimado con el kernel de transición uniforme tiene un error estándar más grande. No obstante, el Criterio de Información Bayesiano (BIC), [50], es muy similar bajo las dos propuestas, las diferencias se marcan para el Deviance Information Criteria (DIC), [52], el cual es menor para la propuesta con kernel de transición normal. Finalmente, se encontró que la tasa de aceptación de las propuestas hechas para  $\rho$  con el kernel de transición normal son mayores que con la otra propuesta.

A continuación, en las figuras 3.1 y 3.2 se presentan las cadenas, tanto de las estimaciones hechas usando el kernel de transición uniforme como el normal.

FIGURA 3.1. Cadenas de los parámetros del modelo SAR heterocedástico con kernel de transición uniforme

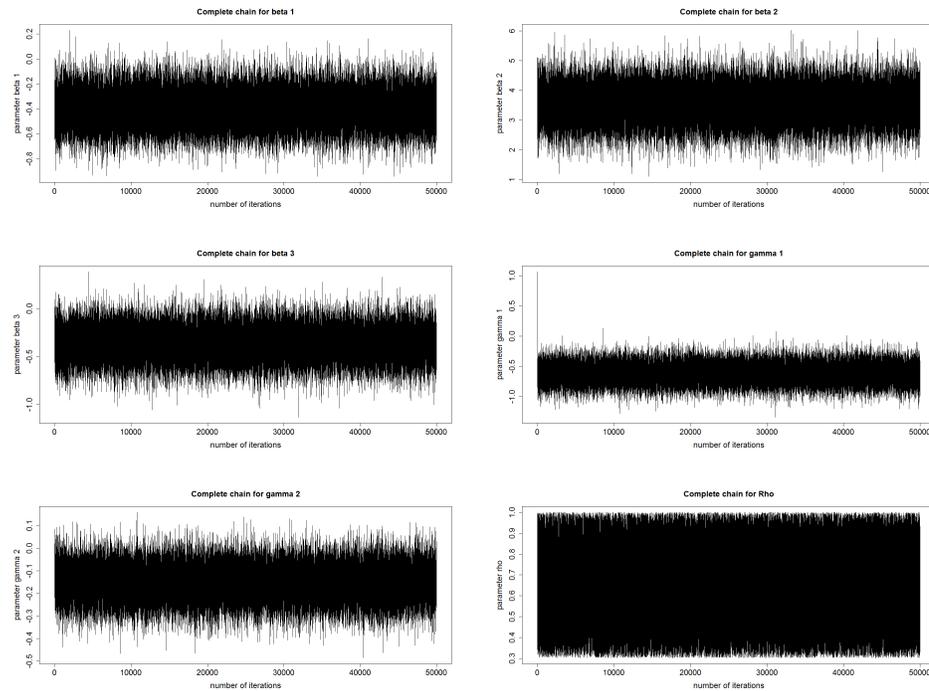
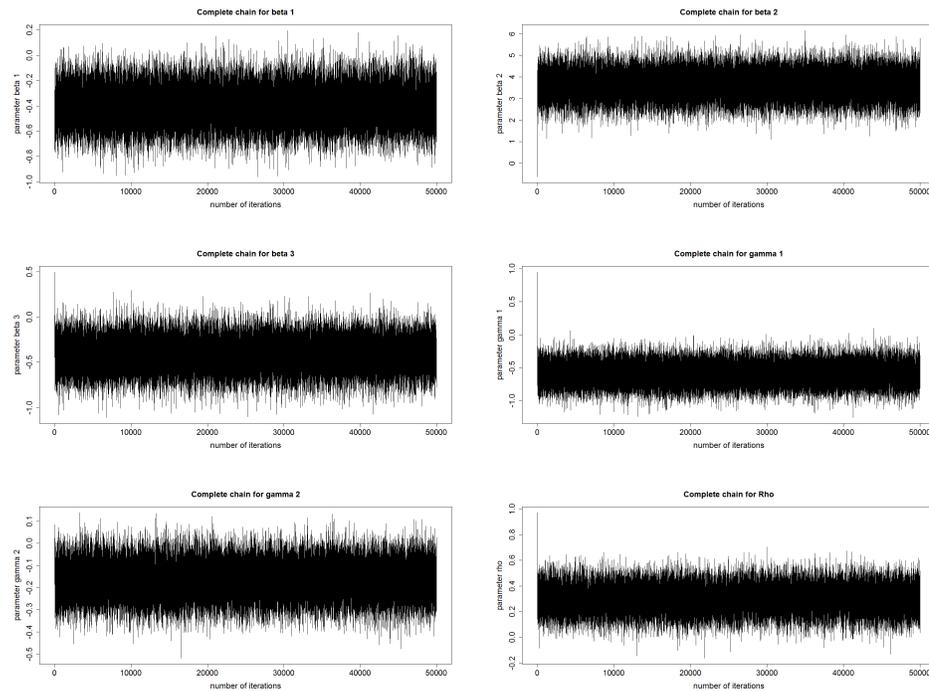


FIGURA 3.2. Cadenas de los parámetros del modelo SAR heterocedástico con kernel de transición normal



Se observa que los parámetros logran una rápida convergencia y tienen formas similares. No obstante, existen grandes diferencias en las cadenas del parámetro  $\rho$ , puesto que en la estimación mediante el kernel de transición uniforme la variabilidad es muy alta, tomando

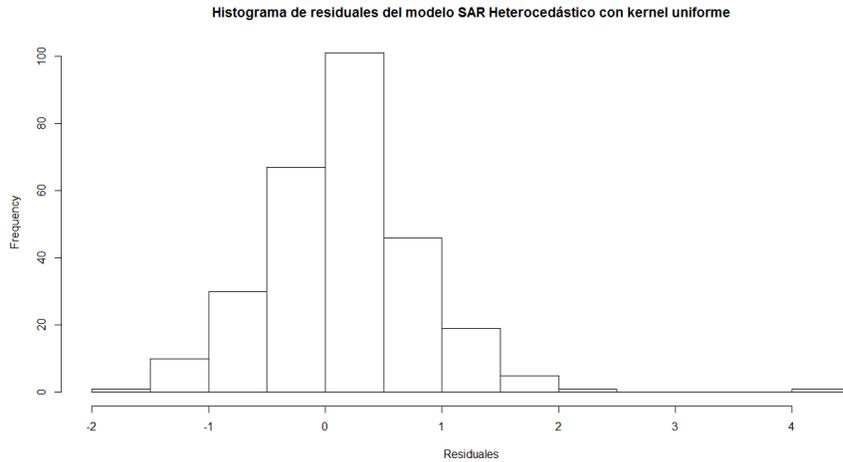


FIGURA 3.3. Residuales para el modelo SAR heterocedástico ajustado con el kernel de transición uniforme para los datos de Leucemia en Nueva York, [62]

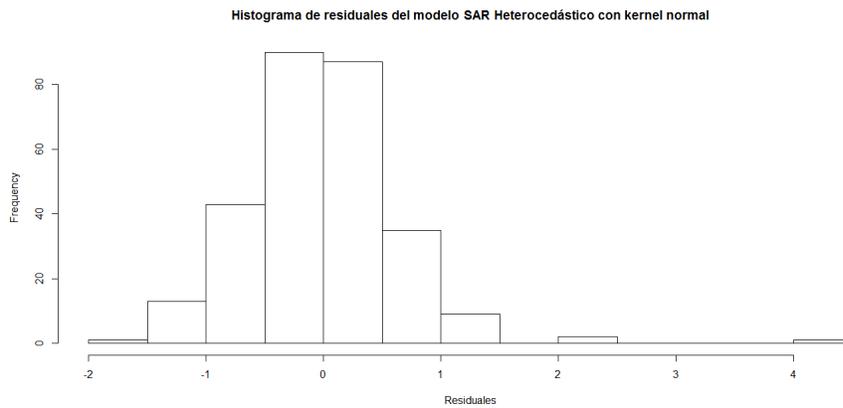


FIGURA 3.4. Residuales para el modelo SAR heterocedástico ajustado con el kernel de transición normal para los datos de Leucemia en Nueva York, [62]

valores entre 0.25 y 1.

Los histogramas con los residuales obtenidos al hacer el ajuste con los kernels de transición uniforme y normal se encuentran en las figuras 3.3 y 3.4, respectivamente. En ellos se observan formas de campana con algunas observaciones atípicas. Además, para las dos propuestas se observa que el histograma se encuentra centrado en cero. Se evaluaron las medidas de tendencia de estos residuales, así como si siguen una distribución normal, los resultados se encuentran en la tabla 3.3.

TABLA 3.3. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SAR heterocedástico bajo las dos propuestas

	Kernel Uniforme	Kernel Normal
Mínimo	-1,721	-1,747
Media	0,165	-0,006
Mediana	0,148	-0,040
Máximo	4,061	4,111
p-valor (Shapiro Test)	1,43E-06	1,69E-08

Esta información señala que los residuales poseen medias próximas a cero, más para los residuales producto del ajuste mediante el kernel de transición normal. De igual manera indica que los residuales parecen no seguir una distribución normal, esto es generado por los residuales atípicos.

### 3.2. Modelo SEM Bayesiano: Heterocedástico

#### 3.2.1. Especificación

El modelo SEM heterocedástico se obtiene a partir del modelo general, 1.1, cuando  $\rho = 0$ . Es decir, el modelo está dado por:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned}$$

donde  $\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ , con  $\boldsymbol{\Omega}$  una matriz diagonal con término general dado por  $g(\boldsymbol{\Omega}_{ii}) = \mathbf{Z}\boldsymbol{\gamma}$ , con  $g$  una función apropiadamente definida. Por otra parte, los términos  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\lambda$ ,  $\mathbf{W}$ ,  $\boldsymbol{\varepsilon}$  y  $\boldsymbol{\mu}$  corresponden a la descripción realizada en la sección 1.4.1, y  $\mathbf{Z}$  y  $\boldsymbol{\gamma}$  a lo expuesto en la sección 3.1.1. La especificación del modelo se completa con las siguientes distribuciones a priori:

1.  $p(\boldsymbol{\beta}) \sim N(\mathbf{b}, \mathbf{B})$
2.  $p(\boldsymbol{\gamma}) \sim N(\mathbf{g}, \mathbf{G})$
3.  $p(\lambda) \sim U(-1/\omega_{max}, 1)$

donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindad espacial,  $\mathbf{W}$ , [8]. Por facilidad, se asume que estas distribuciones a priori son independientes.

Teniendo en cuenta las distribuciones a priori y la función de verosimilitud del modelo SEM heterocedástico, se obtiene la distribución a posteriori,  $\pi$ , dada por:

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda) \propto |\boldsymbol{\Sigma}|^{-1/2} |\mathbf{D}| \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \\ \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})' \mathbf{G}^{-1} (\boldsymbol{\gamma} - \mathbf{g}) \right\} \quad (3.2) \end{aligned}$$

donde  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ . A partir de 3.2 las siguientes distribuciones condicionales a posteriori son obtenidas. Las demostraciones paso a paso de la obtención de estas distribuciones se encuentran en el apéndice A.

**1. Distribución condicional a posteriori para  $\beta$ , [22],[24]**

La distribución condicional a posteriori del vector de parámetros  $\beta$  está dada por:

$$\pi(\beta|\gamma, \lambda) \sim N(\mathbf{b}^*, \mathbf{B}^*)$$

con:

$$\begin{aligned} \mathbf{b}^* &= \mathbf{B}^*(\mathbf{X}'\mathbf{D}'\Sigma^{-1}\mathbf{D}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b}) \\ \mathbf{B}^* &= (\mathbf{X}'\mathbf{D}'\Sigma^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1})^{-1} \end{aligned}$$

donde  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ .

**2. Distribución condicional a posteriori para  $\gamma$**

La siguiente es la distribución condicional a posteriori del vector de parámetros  $\gamma$ :

$$\pi(\gamma|\beta, \lambda) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}' \Sigma^{-1} \mathbf{D} (\mathbf{y} - \mathbf{X}\beta) \right\} \exp \left\{ -\frac{1}{2}(\gamma - \mathbf{g})' \mathbf{G}^{-1} (\gamma - \mathbf{g}) \right\}$$

donde  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$

**3. Distribución condicional a posteriori de  $\lambda$**

La distribución a posteriori de  $\lambda$ , condicionando 3.2 por los demás parámetros, toma la siguiente forma:

$$\pi(\lambda|\beta, \gamma) \propto |\mathbf{D}| \exp \left\{ -\frac{1}{2 \times \frac{1}{c}} \left( \lambda - \frac{d}{c} \right)^2 \right\} \mathbf{I}_{(-1/\omega_{max}, 1)}(\lambda)$$

donde  $c = (\mathbf{y} - \mathbf{X}\beta)' \mathbf{W}' \Sigma^{-1} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)$ ,  $d = (\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ .

Al revisar las distribuciones condicionales a posteriori de  $\gamma$  y  $\lambda$  se encuentra que no tienen formas conocidas, y por lo tanto no resultan sencillas de generar. Es por esto que las siguientes propuestas son realizadas para muestrear estos parámetros:

**3.2.1.1. Propuesta para  $\gamma$ , [22], [24]**

Para obtener muestras de  $\gamma$  se propone un kernel de transición de manera similar a como se encuentra en [22] y [24], donde se construyen propuestas basadas en el algoritmo de Fisher Scoring, [52].

En este caso las variables para aproximar la transformación de las observaciones alrededor de las estimaciones del parámetro, parten del modelo observado:

$$t_i = (y_i - \lambda^{(c)} \mathbf{W}'_i \mathbf{y} - x'_i \beta^{(c)} + \lambda^{(c)} \mathbf{W}'_i x'_i \beta^{(c)})^2 \sim \sigma_i^2 \chi_1^2 \quad \text{para } i = 1, 2, \dots, n$$

Estas variables tienen media  $E(t_i) = \sigma_i^2$ , varianza  $Var(t_i) = 2\sigma_i^4$  y cumplen  $g[E(t_i)] = z'_i \gamma$ .

Como  $g$  es una función diferenciable, se puede emplear la aproximación lineal alrededor de  $E(t_i)$ :

$$g(t_i) \simeq g[E(t_i)] + g'[E(t_i)][t_i - E(t_i)] = \tilde{y}_i$$

y

$$\text{Var}[g(t_i)] = g'[E(t_i)]^2 \text{Var}(t_i)$$

en alguna vecindad de  $E(t_i) = \sigma_i^2$ . Luego, con  $\beta^{(c)}$  y  $\gamma^{(c)}$ , los valores actuales de  $\beta$  y  $\gamma$ , el valor apropiado de la observación es:

$$\tilde{y}_i = z_i' \gamma^{(c)} + g'[g^{-1}(z_i' \gamma^{(c)})][(y_i - \lambda^{(c)} \mathbf{W}_i' \mathbf{y} - x_i' \beta^{(c)} + \lambda^{(c)} \mathbf{W}_i' x_i' \beta^{(c)})^2 - g^{-1}(z_i' \gamma^{(c)})]$$

con varianza:

$$\text{Var}[\tilde{y}_i] = 2\{g'(z_i' \gamma^{(c)})g^{-1}(z_i' \gamma^{(c)})\}^2$$

Si  $g = \log$  el vector de las observaciones sería  $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$ , con:

$$\tilde{y}_i = z_i' \gamma^{(c)} + \frac{(y_i - \lambda^{(c)} \mathbf{W}_i' \mathbf{y} - x_i' \beta^{(c)} + \lambda^{(c)} \mathbf{W}_i' x_i' \beta^{(c)})^2}{\exp(z_i' \gamma^{(c)})} - 1, \quad i = 1, 2, \dots, n$$

con varianza igual a 2. El vector  $\tilde{\mathbf{y}}$  corresponde a la variable de trabajo.

Luego, el kernel de transición puede ser obtenido como la mixtura entre la distribución a priori de  $\gamma$  y la distribución de las variables de trabajo. Este está dado por:

$$q_\gamma(\gamma^{(c)}, \gamma^{(n)}) = N(\mathbf{g}^*, \mathbf{G}^*)$$

donde  $\mathbf{g}^* = \mathbf{G}^*(\mathbf{G}^{-1} \mathbf{g} + 0.5 \mathbf{Z}' \tilde{\mathbf{Y}})$  y  $\mathbf{G}^* = (\mathbf{G}^{-1} + 0.5 \mathbf{Z}' \mathbf{Z})^{-1}$

### 3.2.1.2. Propuesta para $\lambda$

Para obtener muestras de  $\lambda$ , las propuestas de kernel de transición corresponden a las mismas del caso homocedástico.

- **Kernel de transición uniforme:** Una primera propuesta es hacer uso del muestreador de Metropolis-Hastings, [56], [59], empleando como kernel de transición para  $\lambda$  un kernel uniforme entre  $-1/\omega_{max}$  y 1, donde  $\omega_{max}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}$ , [8].
- **Kernel de transición normal** Tomando como base la distribución condicional a posteriori de  $\lambda$ , calculada en el apéndice A, se propone un kernel de transición normal,  $N(c^*, d^*)$ , donde:

$$c^* = \frac{(\mathbf{y} - \mathbf{X}\beta)' \Sigma^{-1} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)}{(\mathbf{y} - \mathbf{X}\beta)' \mathbf{W}' \Sigma^{-1} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)} \quad d^* = \frac{1}{(\mathbf{y} - \mathbf{X}\beta)' \mathbf{W}' \Sigma^{-1} \mathbf{W} (\mathbf{y} - \mathbf{X}\beta)}$$

Este kernel de transición fue realizado teniendo en cuenta la distribución condicional a posteriori de  $\lambda$ . Esta distribución normal se encuentra truncada en el intervalo  $(-1/\omega_{max}, 1)$ . Esta consideración está implementada en la función *hetero\_sem* de la librería *BSPADATA* de R, [54].

### 3.2.2. Simulación

Posteriormente se hizo la simulación de 49 datos de una variable normal,  $\mathbf{y}$ , generada a partir del siguiente modelo:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned}$$

con  $\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ . Para este modelo  $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2]$ ,  $\boldsymbol{\beta} = [-35, 0.35, -1.7]$  y la matriz  $\boldsymbol{\Omega}$  es diagonal y está dada por:

$$\log(\boldsymbol{\Omega}_{ii}) = \mathbf{Z}'\boldsymbol{\gamma} \quad i = 1, \dots, 49$$

con  $\mathbf{Z} = [\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_3]$  y  $\boldsymbol{\gamma} = [-8, 0.026, -0.4]$ . Esto implica que la variable  $\mathbf{y}$  fue simulada con media  $-35\mathbf{x}_0 + 0.35\mathbf{x}_1 - 1.7\mathbf{x}_2$  y matriz de varianzas y covarianzas  $[(\mathbf{I} - \lambda\mathbf{W})^{-1}]'\boldsymbol{\Omega}[(\mathbf{I} - \lambda\mathbf{W})^{-1}]$ .

Por su parte, la variable  $\mathbf{X}_0$  es un vector de unos de longitud 49 y las realizaciones de las variables  $\mathbf{X}_1$ ,  $\mathbf{X}_2$  y  $\mathbf{X}_3$  fueron generadas a partir de distribuciones uniformes en los intervalos (0, 400), (10, 23) y (0, 10), respectivamente. Finalmente, la matriz de vecindades  $\mathbf{W}$  es la misma utilizada en la aplicación de datos de crimen en Columnus, Ohio por [8].

Se hicieron simulaciones para tres valores de  $\lambda$ : 0.25, 0.55 y 0.75. Los datos fueron ajustados mediante las dos metodologías, y los resultados se presentan en la tabla 3.4.

TABLA 3.4. Comparación de ajustes del modelo SEM heterocedástico con distintos valores de  $\lambda$

$\lambda$ real	Método	Valor	$\beta_0$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\lambda$	BIC	DIC
0,25	Kernel Uniforme	Estimación	-34,944	0,350	-1,704	-5,240	0,016	-0,323	0,658	147,538	764,526
		Desviación estándar	0,039	0,000	0,002	0,702	0,002	0,079	0,197		
	Kernel Normal	Estimación	-35,026	0,350	-1,699	-6,571	0,020	-0,400	0,164	-118,510	-254,793
		Desviación estándar	0,018	0,000	0,001	0,668	0,002	0,078	0,019		
0,55	Kernel Uniforme	Estimación	-35,004	0,350	-1,701	-5,428	0,015	-0,136	0,664	-3,713	160,414
		Desviación estándar	0,094	0,000	0,005	0,568	0,002	0,075	0,197		
	Kernel Normal	Estimación	-35,001	0,350	-1,700	-5,371	0,015	-0,154	0,361	-59,370	-76,095
		Desviación estándar	0,091	0,000	0,005	0,578	0,002	0,075	0,056		
0,75	Kernel Uniforme	Estimación	-35,014	0,350	-1,700	-6,134	0,016	-0,257	0,661	-143,26	-303,201
		Desviación estándar	0,034	0,000	0,001	0,689	0,002	0,111	0,194		
	Kernel Normal	Estimación	-35,017	0,350	-1,700	-6,224	0,016	-0,249	0,671	-143,175	-329,136
		Desviación estándar	0,035	0,000	0,001	0,677	0,002	0,110	0,065		

- Cuando  $\lambda = 0.25$  se encuentran similitudes entre las dos propuestas para la estimación y los errores estándar de los vectores  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$ . No obstante, la diferencia se marca en la estimación del parámetro de rezago espacial,  $\lambda$ , el cual es más próximo a la realidad estimando a través del kernel de transición normal. Esto se refleja en el Criterio de Información Bayesiano (BIC), el Deviance Information Criteria (DIC) y en la tasa de aceptación de  $\lambda$ , la cual es de alrededor del 80 % con el kernel de transición normal y de casi el 54 % con el kernel uniforme.
- La estimación de  $\lambda$  cuando su valor verdadero es 0.55, realizada a través del kernel de transición uniforme, no se aleja de 0.66, propiciando diferencias en las estimaciones de los demás parámetros y mayores errores estándar. Mientras que el trabajo realizado con el kernel de transición normal refleja mayor proximidad de las estimaciones a los valores reales, mayores tasas de aceptación y menores valores para los Criterios de Información BIC y DIC.

- Cuando  $\lambda = 0.75$ , las estimaciones obtenidas mediante el kernel de transición normal son mejores en términos de proximidad a los valores reales, con menores errores estándar, menores valores de los Criterios de Información, BIC y DIC, así como mayores tasas de aceptación.

### 3.2.3. Aplicación: Datos de Leucemia en Nueva York

Para aplicar la metodología propuesta para el modelo SEM heterocedástico se trabajó con los datos presentados en [62], relacionados con la cantidad de casos de Leucemia en 281 unidades espaciales de Nueva York. Las variables de este conjunto de datos son explicadas en la sección 3.1.3. El modelo a ajustar está dado por:

$$\begin{aligned} PROPLEU &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned}$$

con  $\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ , donde  $\boldsymbol{\Omega}$  es una matriz diagonal con término general en la diagonal dado por  $\boldsymbol{\Omega}_{ii} = \exp(\gamma_0 + \gamma_1 PEXPOSURE_i)$ ,  $i = 1, \dots, 281$ . Además,  $\mathbf{X} = [\mathbf{1}; PCTAGE65P; PCTOWNHOME]$ , donde  $\mathbf{1}$  es un vector de unos de longitud  $n = 281$ . La matriz de vecindad espacial  $\mathbf{W}$  es obtenida a partir del objeto *listw\_NY* del conjunto de datos *NY\_data* de R, [54]. Es una matriz de contigüidad binaria, estandarizada por filas, es decir, si un par de unidades espaciales,  $i$  y  $j$ , son vecinas, entonces la entrada  $[i, j]$  de la matriz es diferente de 0 y se garantiza que la suma de cada fila de la matriz es 1.

En esta situación se propuso un vector de ceros y una matriz diagonal con elementos diagonales iguales a 10000, como valores a priori para la media y la matriz de varianzas y covarianzas de  $\boldsymbol{\beta}$  y  $\boldsymbol{\gamma}$ , respectivamente. Se generaron 50000 simulaciones con calentamiento de 2500 valores y selección de observaciones cada 10 simulaciones, esto con el objetivo de reducir el efecto de correlación entre muestras sucesivas.

Los resultados del ajuste de este modelo se encuentran en la tabla 3.5, allí se encuentran las estimaciones puntuales de cada parámetro acompañado de sus errores estándar, el Criterio de Información Bayesiano, BIC, [50], el Deviance Information Criteria (DIC), [52], y las tasas de aceptación de los parámetros  $\boldsymbol{\gamma}$  y  $\lambda$ . Para efectos de comparación, se presentan de manera conjunta los resultados obtenidos mediante el kernel de transición uniforme y el kernel de transición normal.

TABLA 3.5. Parámetros estimados del modelo SEM heterocedástico

	Kernel Uniforme	SD Kernel Uniforme	Kernel Normal	SD Kernel Normal
$\beta_0$	-0,482	0,167	-0,458	0,163
$\beta_1$	3,841	0,649	3,887	0,634
$\beta_2$	-0,413	0,205	-0,464	0,200
$\gamma_0$	-0,593	0,162	-0,570	0,163
$\gamma_1$	-0,163	0,077	-0,175	0,077
$\lambda$	0,653	0,203	0,372	0,128
BIC	96,613		69,909	
DIC	1183,26		1087,139	
% Aceptación $\boldsymbol{\gamma}$	50,62		50,90	
% Aceptación $\lambda$	12,764		41,15	

Los resultados reflejan que las estimaciones puntuales de los vectores de parámetros  $\beta$  y  $\gamma$  son muy similares bajo las dos propuestas. No obstante, hay diferencias significativas entre las estimaciones del parámetro de rezago espacial,  $\lambda$ , ya que en la propuesta uniforme este no se aleja del punto medio del intervalo  $(-1/\omega_{max}, 1)$ , donde  $\omega_{max}$  es el mayor valor propio, en valor absoluto, de la matriz  $\mathbf{W}$ , [8].

Por otra parte, se observa que los errores estándar obtenidos bajo el kernel de transición normal son mayores que los obtenidos con el uniforme para los parámetros  $\gamma$  y  $\lambda$ . No obstante, el Criterio de Información Bayesiano (BIC), [50], y el Deviance Information Criteria, [52], fueron menores al emplear el kernel de transición normal. Finalmente, se encontró que la tasa de aceptación de las propuestas hechas para  $\lambda$  con el kernel de transición normal son mejores que con la otra metodología.

A continuación, se presentan las cadenas, tanto de las estimaciones hechas usando el kernel de transición uniforme como el normal, para obtener muestras de  $\lambda$ .

FIGURA 3.5. Cadenas de los parámetros del modelo SEM heterocedástico con kernel de transición uniforme

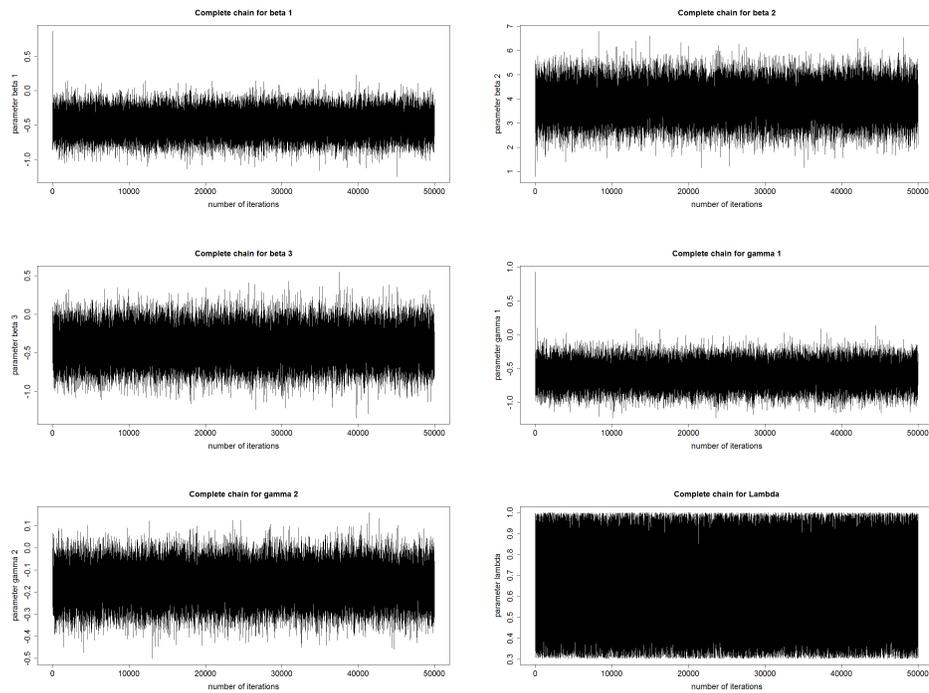
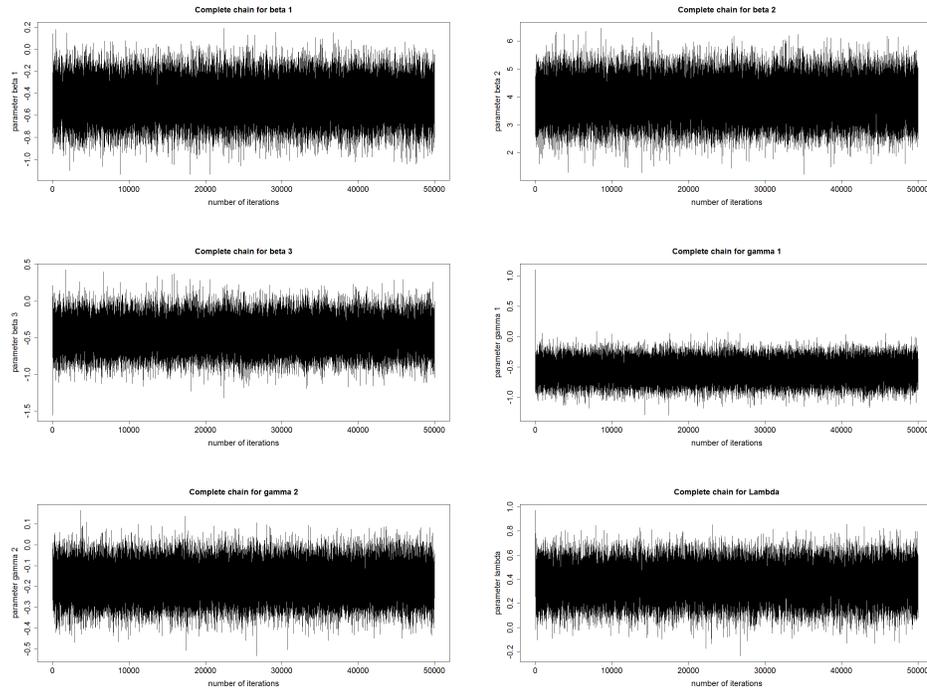


FIGURA 3.6. Cadenas de los parámetros del modelo SEM heterocedástico con kernel de transición normal



Se observa que las cadenas de todos los parámetros alcanzan una rápida convergencia. Vale la pena resaltar el caso de la cadena del parámetro  $\lambda$ , la cual oscila en gran parte del intervalo  $(0, 1)$ , a diferencia de lo observado para el mismo parámetro con el kernel de transición normal.

Los histogramas con los residuales calculados al ajustar el modelo SEM heterocedástico mediante los kernels de transición uniforme y normal se muestran en las figuras 3.7 y 3.8, respectivamente. Las formas de ambos histogramas son muy similares, incluso en la presencia de valores alejados del centro de la distribución, estos tienen formas de campana apuntadas. Ambas distribuciones parecen estar centradas en cero.

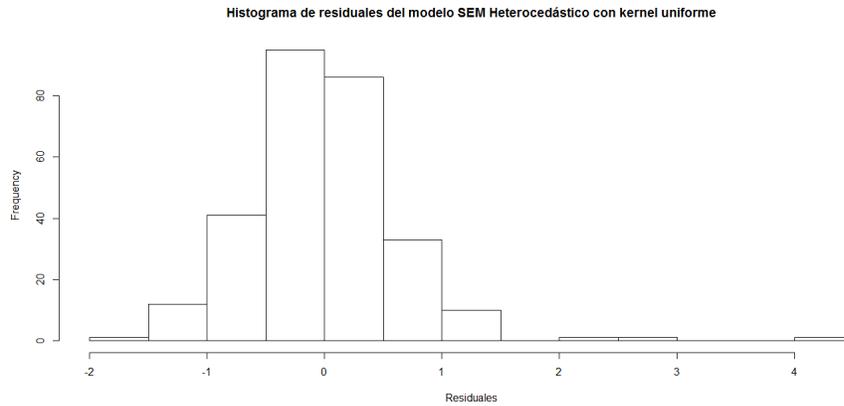


FIGURA 3.7. Residuales para el modelo SEM heterocedástico ajustado con el kernel de transición uniforme para los datos de Leucemia en Nueva York, [62]

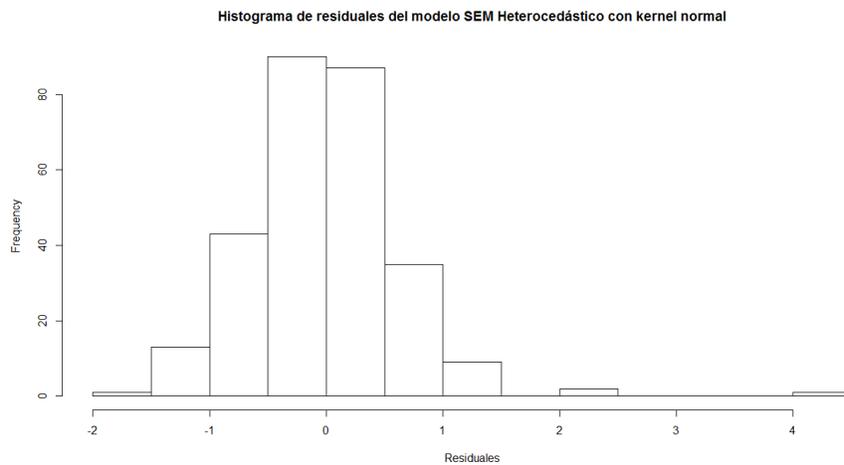


FIGURA 3.8. Residuales para el modelo SEM heterocedástico ajustado con el kernel de transición normal para los datos de Leucemia en Nueva York, [62]

Se evaluaron las medidas de tendencia de estos residuales, así como si siguen una distribución normal, los resultados se encuentran en la tabla 3.6.

TABLA 3.6. Medidas resumen y test de normalidad para los residuales del ajuste del modelo SEM heterocedástico bajo las dos propuestas

	Kernel Uniforme	Kernel Normal
Mínimo	-1,721	-1,750
Media	-0,018	-0,018
Mediana	-0,079	-0,077
Máximo	4,119	4,105
p-valor (Shapiro Test)	9,79E-09	1,08E-08

Estos resultados muestran que las estimaciones logradas a partir de ambas metodologías producen residuales cercanos a cero. También se revela que las distribuciones de estos residuales no son normales, probablemente por la influencia que tienen los residuales atípicos.

### 3.3. Modelo General Bayesiano: Heterocedástico

#### 3.3.1. Especificación

Este modelo incluye la asociación espacial, tanto de la variable respuesta, como del término de error. Además de esto, tiene en cuenta la heterogeneidad de varianzas, a través del espacio, de la variable respuesta. El modelo está dado por:

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \rho\mathbf{W}_1\mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda\mathbf{W}_2\boldsymbol{\varepsilon} + \boldsymbol{\mu}\end{aligned}$$

donde  $\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ , con  $\boldsymbol{\Omega}$  matriz diagonal con término general dado por  $g(\Omega_{ii}) = \mathbf{Z}\boldsymbol{\gamma}$ , donde  $g$  es una función apropiadamente definida. Los términos  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\boldsymbol{\beta}$ ,  $\rho$ ,  $\lambda$ ,  $\mathbf{W}$ ,  $\boldsymbol{\varepsilon}$  y  $\boldsymbol{\mu}$  corresponden a la descripción dada en la sección 1.4.1.  $\mathbf{Z}$  y  $\boldsymbol{\gamma}$  están explicados en la sección 3.1.1. La especificación del modelo concluye con la definición de las siguientes distribuciones a priori:

1.  $p(\boldsymbol{\beta}) \sim N(\mathbf{b}, \mathbf{B})$
2.  $p(\boldsymbol{\gamma}) \sim N(\mathbf{g}, \mathbf{G})$
3.  $p(\rho) \sim U(-1/\omega_{1_{max}}, 1)$
4.  $p(\lambda) \sim U(-1/\omega_{2_{max}}, 1)$

donde  $\omega_{1_{max}}$  y  $\omega_{2_{max}}$  corresponden a los mayores valores propios negativos, en valor absoluto, de las matrices de vecindad espacial,  $\mathbf{W}_1$  y  $\mathbf{W}_2$ , [8]. Por facilidad, se asume que las distribuciones a priori son independientes. Con la información a priori y la función de verosimilitud en (1.10), se construye la distribución a posteriori,  $\pi$ , dada por:

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda) &\propto |\boldsymbol{\Sigma}|^{-1/2} |\mathbf{A}| |\mathbf{D}| \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ &\quad \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})' \mathbf{G}^{-1} (\boldsymbol{\gamma} - \mathbf{g}) \right\} \quad (3.3)\end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ . Suponiendo conocidos los demás parámetros, se obtienen las distribuciones condicionales a posteriori a partir de (3.3) y son presentadas a continuación. Las demostraciones se encuentran en el apéndice A.

#### 1. Distribución condicional a posteriori para $\boldsymbol{\beta}$ , [22],[24]

La distribución a posteriori del vector de parámetros  $\boldsymbol{\beta}$  está dada por:

$$\pi(\boldsymbol{\beta} | \boldsymbol{\gamma}, \rho, \lambda) \sim N(\mathbf{b}^*, \mathbf{B}^*)$$

con:

$$\begin{aligned}\mathbf{b}^* &= \mathbf{B}^* (\mathbf{X}' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b}) \\ \mathbf{B}^* &= (\mathbf{X}' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} \mathbf{X} + \mathbf{B}^{-1})^{-1}\end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

## 2. Distribución condicional a posteriori para $\gamma$

La siguiente es la forma funcional de la distribución condicional a posteriori del vector de parámetros  $\gamma$ :

$$\pi(\gamma|\beta, \rho, \lambda) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{A}\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}' \Sigma^{-1} \mathbf{D}(\mathbf{A}\mathbf{y} - \mathbf{X}\beta) \right\} \exp \left\{ -\frac{1}{2}(\gamma - \mathbf{g})' \mathbf{G}^{-1}(\gamma - \mathbf{g}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

## 3. Distribución condicional a posteriori de $\rho$

Para el parámetro de rezago espacial  $\rho$  se obtuvo la distribución condicional a posteriori,  $\pi(\rho|\beta, \gamma, \lambda)$ :

$$\pi(\rho|\beta, \gamma, \lambda) \propto |\mathbf{A}| \exp \left\{ -\frac{1}{2 \times \frac{1}{c}} \left( \rho - \frac{d}{c} \right)^2 \right\} \mathbf{I}_{(-1/\omega_{1max}, 1)}(\rho)$$

con  $A = \mathbf{I} - \rho\mathbf{W}_1$ ,  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ ,  $c = \mathbf{y}'\mathbf{W}'_1\mathbf{D}'\Sigma^{-1}\mathbf{D}\mathbf{W}_1\mathbf{y}$  y  $d = (\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}'\Sigma^{-1}\mathbf{D}\mathbf{W}_1\mathbf{y}$

## 4. Distribución condicional a posteriori de $\lambda$

En el caso del parámetro  $\lambda$  se obtuvo una distribución condicional a posteriori,  $\pi(\lambda|\beta, \gamma, \rho)$ , dada por:

$$\pi(\lambda|\beta, \gamma, \rho) \propto |\mathbf{D}| \exp \left\{ -\frac{1}{2 \times \frac{1}{c}} \left( \rho - \frac{d}{c} \right)^2 \right\} \mathbf{I}_{(-1/\omega_{2max}, 1)}(\lambda)$$

con  $A = \mathbf{I} - \rho\mathbf{W}_1$ ,  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ ,  $c = (\mathbf{A}\mathbf{y} - \mathbf{X}\beta)' \mathbf{W}'_2 \Sigma^{-1} \mathbf{W}_2 (\mathbf{A}\mathbf{y} - \mathbf{X}\beta)$  y  $d = (\mathbf{A}\mathbf{y} - \mathbf{X}\beta)' \mathbf{W}'_2 \Sigma^{-1} (\mathbf{A}\mathbf{y} - \mathbf{X}\beta)$

Al analizar las distribuciones condicionales a posteriori de  $\gamma$ ,  $\rho$  y  $\lambda$ , se encuentra que no tienen formas funcionales conocidas, y por lo tanto no resultan sencillas de generar. Por tal razón se propuso la siguiente metodología para muestrear estos parámetros.

### 3.3.1.1. Propuesta para $\gamma$ , [22], [24]

Para obtener muestras de  $\gamma$  se construye un kernel de transición basado en [22] y [24], donde se realizan propuestas basadas en el algoritmo de Fisher Scoring, [52]. En este caso las variables para aproximar la transformación de las observaciones alrededor de las estimaciones del parámetro, parten del modelo observado:

$$t_i = (y_i - \rho^{(c)} \mathbf{W}'_{1i} \mathbf{y} - \lambda^{(c)} \mathbf{W}'_{2i} \mathbf{y} + \lambda^{(c)} \rho^{(c)} \mathbf{W}'_{2i} \mathbf{W}_{1i} \mathbf{y} - x'_i \beta^{(c)} + \lambda^{(c)} \mathbf{W}'_{2i} x'_i \beta^{(c)})^2 \sim \sigma_i^2 \chi_1^2$$

para  $i = 1, \dots, n$ . Estas observaciones tienen media  $E(t_i) = \sigma_i^2$ , varianza  $Var(t_i) = 2\sigma_i^4$  y cumplen  $g[E(t_i)] = z'_i \gamma$ . Al ser  $g$  una función diferenciable, se puede hacer uso de la aproximación lineal alrededor de  $E(t_i)$ :

$$g(t_i) \simeq g[E(t_i)] + g'[E(t_i)][t_i - E(t_i)] = \tilde{y}_i$$

con varianza:

$$Var[g(t_i)] = g'[E(t_i)]^2 Var(t_i)$$

en alguna vecindad de  $E(t_i) = \sigma_i^2$ . Luego, con  $\beta^{(c)}$  y  $\gamma^{(c)}$  los valores actuales de  $\beta$  y  $\gamma$ , respectivamente, el valor apropiado de la observación es:

$$\tilde{y}_i = z_i' \gamma^{(c)} + g' [g^{-1} z_i' \gamma^{(c)}] [(y_i - \rho^{(c)} \mathbf{W}'_{1i} \mathbf{y} - \lambda^{(c)} \mathbf{W}'_{2i} \mathbf{y} + \lambda^{(c)} \rho^{(c)} \mathbf{W}'_{2i} \mathbf{W}_{1i} \mathbf{y} - x_i' \beta^{(c)} + \lambda^{(c)} \mathbf{W}'_{2i} x_i' \beta^{(c)})^2 - g^{-1} (z_i' \gamma^{(c)})]$$

con varianza:

$$Var[\tilde{y}_i] = 2\{g'(z_i' \gamma^{(c)})g^{-1}(z_i' \gamma^{(c)})\}^2$$

Cuando  $g = \log$ , el vector de las observaciones sería  $\tilde{\mathbf{y}} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)$ , con  $\tilde{y}_i$ :

$$\tilde{y}_i = z_i' \gamma^{(c)} + \frac{(y_i - \rho^{(c)} \mathbf{W}'_{1i} \mathbf{y} - \lambda^{(c)} \mathbf{W}'_{2i} \mathbf{y} + \lambda^{(c)} \rho^{(c)} \mathbf{W}'_{2i} \mathbf{W}_{1i} \mathbf{y} - x_i' \beta^{(c)} + \lambda^{(c)} \mathbf{W}'_{2i} x_i' \beta^{(c)})^2}{\exp(z_i' \gamma^{(c)})} - 1$$

para  $i = 1, \dots, n$ . El vector  $\tilde{\mathbf{y}}$  es conocido como la variable de trabajo y tiene varianza igual a 2. El kernel de transición es obtenido como la mixtura entre la distribución a priori, y la de la variable de trabajo, ambas normales. Obteniendo:

$$q_\gamma(\gamma^{(c)}, \gamma^{(n)}) = N(\mathbf{g}^*, \mathbf{G}^*)$$

donde  $\mathbf{g}^* = \mathbf{G}^*(\mathbf{G}^{-1} \mathbf{g} + 0.5\mathbf{Z}'\tilde{\mathbf{Y}})$  y  $\mathbf{G}^* = (\mathbf{G}^{-1} + 0.5\mathbf{Z}'\mathbf{Z})^{-1}$

### 3.3.1.2. Propuesta para $\rho$

Para obtener muestras de  $\rho$ , los kernel de transición corresponden a los mismos del caso homocedástico.

- **Kernel de transición uniforme:** La primera aproximación se basa en hacer uso del muestreador de Metropolis-Hastings, [56], [59], generando muestras a partir de una distribución uniforme entre  $-1/\omega_{1_{max}}$  y 1, donde  $\omega_{1_{max}}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}_1$ , [8].
- **Kernel de Transición Normal:** Partiendo de la distribución condicional a posteriori de  $\rho$ , calculada en el apéndice A, se propone un kernel de transición normal  $N(c^*, d^*)$ , donde:

$$c^* = \frac{\mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} (\mathbf{y} - \mathbf{X} \beta)}{\mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}} \quad d^* = \frac{1}{\mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}}$$

con  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}_2$ . Esta distribución normal es truncada en el intervalo  $(-1/\omega_{1_{max}}, 1)$ , donde  $\omega_{1_{max}}$  corresponde al mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}_1$ , [8]. Así se encuentra implementado en la función *hetero-general* de la librería *BSPADATA* de R, [54].

### 3.3.1.3. Propuesta para $\lambda$

Para simular la distribución condicional a posteriori de  $\lambda$ , la metodología es la misma respecto al modelo general homocedástico.

- **Kernel de transición uniforme:** Mediante el muestreador de Metropolis-Hastings, [56], [59], se obtienen muestras de  $\rho$  a través de un kernel de transición uniforme, con límites  $1/\omega_{2_{max}}$  y 1. En este caso  $\omega_{2_{max}}$  es el mayor valor propio negativo, en valor absoluto, de la matriz  $\mathbf{W}_2$ , [8].

- **Kernel de transición normal:** Haciendo uso del algoritmo Metropolis-Hastings, [56], [59], se obtienen muestras a partir de un kernel de transición normal  $N(c^*, d^*)$ , donde:

$$c^* = \frac{(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}'_2 \boldsymbol{\Sigma}^{-1} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}'_2 \boldsymbol{\Sigma}^{-1} \mathbf{W}_2 (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \quad d^* = \frac{1}{(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}'_2 \boldsymbol{\Sigma}^{-1} \mathbf{W}_2 (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}$$

con  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1$ . Esta distribución normal es truncada en el intervalo  $(-1/\omega_{2_{max}}, 1)$ , donde  $\omega_{2_{max}}$  es el mayor valor propio negativo, en valor absoluto, de la matriz de vecindades,  $\mathbf{W}_2$ , [8]. Así se encuentra implementado en la función *hetero\_general* de la librería *BSPADATA* de R, [54].

### 3.3.2. Simulación

Se generaron 49 datos de una variable normal,  $\mathbf{y}$ , a partir del siguiente modelo:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}_1 \mathbf{y} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &= \lambda \mathbf{W}_2 \boldsymbol{\varepsilon} + \boldsymbol{\mu} \end{aligned}$$

con  $\boldsymbol{\mu} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ . Para este modelo  $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2]$ ,  $\boldsymbol{\beta} = [-35, 0.35, -1.7]$  y la matriz  $\boldsymbol{\Omega}$  es diagonal y está dada por:

$$\log(\boldsymbol{\Omega}_{ii}) = \mathbf{Z}\boldsymbol{\gamma} \quad i = 1, \dots, 49$$

con  $\mathbf{Z} = [\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_3]$  y  $\boldsymbol{\gamma} = [-8, 0.026, -0.4]$ . Esto implica que la variable  $\mathbf{y}$  fue simulada con media  $(\mathbf{I} - \rho \mathbf{W}_1)^{-1}(-35\mathbf{x}_0 + 0.35\mathbf{x}_1 - 1.7\mathbf{x}_2)$  y matriz de varianzas y covarianzas dada por  $[(\mathbf{I} - \rho \mathbf{W}_1)^{-1}(\mathbf{I} - \lambda \mathbf{W}_2)^{-1}]' \boldsymbol{\Omega} [(\mathbf{I} - \rho \mathbf{W}_1)^{-1}(\mathbf{I} - \lambda \mathbf{W}_2)^{-1}]$ .

Por su parte, la variable  $\mathbf{X}_0$  es un vector de unos de longitud 49 y las realizaciones de las variables  $\mathbf{X}_1$ ,  $\mathbf{X}_2$  y  $\mathbf{X}_3$  son generadas a partir de distribuciones uniformes en los intervalos  $(0, 400)$ ,  $(10, 23)$  y  $(0, 10)$ , respectivamente. Finalmente, las matrices de vecindad  $\mathbf{W}_1$  y  $\mathbf{W}_2$  se asumen iguales y corresponden a la matriz utilizada en la aplicación de datos de crimen en Columbus, Ohio por [8].

Se hicieron simulaciones para tres parejas de valores de  $\rho$  y  $\lambda$ :  $(0.12, 0.48)$ ;  $(0.45, 0.65)$  y  $(0.70, 0.20)$ . Los datos fueron ajustados mediante las dos metodologías, y los resultados se encuentran en la tabla 3.7.

TABLA 3.7. Comparación de ajustes del modelo SEM Heterocedástico con distintos valores de  $\rho$  y  $\lambda$

$\rho$ real	$\lambda$ real	Método	Valor	$\beta_0$	$\beta_1$	$\beta_2$	$\gamma_1$	$\gamma_2$	$\gamma_3$	$\rho$	$\lambda$	BIC	DIC
0,12	0,48	Kernel Uniforme	Estimación	-44,840	0,379	-1,847	1,221	0,001	0,057	0,666	0,663	1014,778	3825,295
			Desviación estándar	2,041	0,003	0,093	0,572	0,002	0,067	0,197	0,196		
		Kernel Normal	Estimación	-34,736	0,349	-1,712	-7,252	0,023	-0,301	0,120	0,392	-92,669	-187,788
			Desviación estándar	0,095	0,000	0,004	0,602	0,002	0,070	0,001	0,030		
0,45	0,65	Kernel Uniforme	Estimación	-34,621	0,349	-1,706	-0,403	-0,001	-0,205	0,663	0,657		33397,18
			Desviación estándar	0,557	0,000	0,018	0,816	0,002	0,090	0,195	0,196		
		Kernel Normal	Estimación	-33,517	0,349	-1,706	-0,813	-0,001	-0,176	0,445	0,953	2,616	102,289
			Desviación estándar	-33,517	0,349	-1,706	-0,813	-0,001	-0,176	0,445	0,953		
0,70	0,20	Kernel Uniforme	Estimación	-35,385	0,350	-1,673	-2,674	0,003	-0,326	0,662	0,663		97901,02
			Desviación estándar	0,128	0,000	0,006	0,608	0,002	0,088	0,194	0,197		
		Kernel Normal	Estimación	-35,382	0,350	-1,675	-2,629	0,004	-0,373	0,699	0,231	-86,014	-166,748
			Desviación estándar	0,115	0,000	0,005	0,594	0,002	0,083	0,001	0,090		

- Cuando  $\rho = 0.12$  y  $\lambda = 0.48$  se encuentra que ambas propuestas se aproximan a los valores reales del vector  $\beta$ , mientras que las estimaciones del vector  $\gamma$  y los parámetros  $\rho$  y  $\lambda$  son más aproximadas a la realidad mediante el kernel de transición normal. Además se observan menores errores estándar para las estimaciones obtenidas a través de esta propuesta, así como un menor valor del Criterio de Información Bayesiano (BIC), el Deviance Information Criteria (DIC) y mayores tasas de aceptación.
- Por otra parte, en el caso en que  $\rho = 0.45$  y  $\lambda = 0.65$  se observa un fenómeno similar al anterior, resultados más favorables para el muestreo con el kernel de transición normal, ya que las estimaciones obtenidas con el kernel uniforme no se alejan de la vecindad de 0.66. No obstante, la estimación del parámetro  $\lambda$  no es tan próxima a su verdadero valor a diferencia de las estimaciones de los demás parámetros. Los valores del Criterio de Información Bayesiano (BIC) y el Deviance Information Criteria (DIC) no son tan altos con el kernel de transición normal como con el kernel de transición uniforme.
- Finalmente, cuando  $\rho = 0.70$  y  $\lambda = 0.20$ , se ratifica la superioridad en la estimación de la propuesta hecha con el kernel de transición normal. Esto se evidencia con estimaciones más acertadas y precisas, así como menores valores del Criterio de Información Bayesiano (BIC), el Deviance Information Criteria (DIC) y mayores tasas de aceptación para todos los parámetros, a excepción del vector de parámetros  $\gamma$ , para el cual se obtienen tasas muy similares bajo las dos propuestas.

### 3.3.3. Aplicación: Datos de Leucemia en Nueva York

Para aplicar la metodología propuesta para el modelo general heterocedástico se trabajó con los datos presentados en [62], que se relacionan con la cantidad de casos de Leucemia en 281 unidades espaciales de la ciudad de Nueva York. Las variables de este conjunto de datos están explicados en la sección 3.1.3. El modelo a ajustar es:

$$\begin{aligned} PROPLEU &= \mathbf{X}\beta + \rho \mathbf{W}_1 PROPLEU + \varepsilon \\ \varepsilon &= \lambda \mathbf{W}_2 \varepsilon + \mu \end{aligned}$$

con  $\mu$  un término de error normal de media  $\mathbf{0}$  y matriz de varianzas diagonal,  $\Omega$ , con término diagonal dado por  $\Omega_{ii} = \exp(\gamma_0 + \gamma_1 PEXPOSURE_i)$ ,  $i = 1, \dots, 281$ . Además,  $\mathbf{X} = [\mathbf{1}; PCTAGE65P; PCTOWNHOME]$ , donde  $\mathbf{1}$  es un vector de unos de longitud  $n = 281$ . En este caso se asume  $\mathbf{W}_1 = \mathbf{W}_2$ , donde la matriz  $\mathbf{W}_1$  es obtenida a partir del objeto *listw\_NY* del conjunto de datos *NY\_data* de R, [54]. Esta es una matriz de contigüidad binaria, estandarizada por filas, es decir, si un par de unidades espaciales,  $i$  y  $j$ , son vecinas, entonces la entrada  $[i, j]$  de la matriz es diferente de 0 y se garantiza que la suma de cada fila de la matriz es 1.

Se propuso un vector de ceros y una matriz diagonal con elementos diagonales iguales a 10000 como valores a priori para el vector de medias y la matriz de varianzas y covarianzas, respectivamente, tanto de  $\beta$  como de  $\gamma$ . Se realizaron 50000 simulaciones, con un calentamiento de 2500 valores, además, con el fin de minimizar el efecto de correlación entre las simulaciones se seleccionaron valores cada 10 simulaciones.

En la tabla 3.8 se presentan las estimaciones para cada uno de los parámetros, así como su error estándar, su Criterio de Información Bayesiano (BIC) y sus tasas de aceptación, para los parámetros  $\gamma$ ,  $\rho$  y  $\lambda$ . Se exponen simultáneamente los resultados de las dos metodologías con el objetivo de compararlas.

TABLA 3.8. Parámetros estimados del modelo general heterocedástico

	Propuesta	Kernel Uniforme	SD Kernel Uniforme	Kernel Normal	SD Kernel Normal
$\beta_0$		-0,473	0,168	-0,345	4,872
$\beta_1$		3,695	0,657	2,493	0,525
$\beta_2$		-0,273	0,206	-0,071	0,197
$\gamma_0$		-0,610	0,161	-1,044	0,162
$\gamma_1$		-0,127	0,076	-0,176	0,077
$\rho$		0,648	0,202	-1,816	0,134
$\lambda$		0,648	0,202	0,996	0,005
BIC		209,042			
DIC		1527,007			
% Aceptación $\gamma$		50,30		50,22	
% Aceptación $\rho$		5,45		0,06	
% Aceptación $\lambda$		5,95		22,31	

En este caso, las estimaciones puntuales bajo las dos propuestas ya no son las mismas. Se encuentra que los errores estándar bajo el kernel de transición normal son menores o iguales a los calculados con el kernel uniforme, a excepción del parámetro  $\beta_0$ . Además, el Criterio de Información Bayesiano (BIC), es mayor con la obtención de muestras de  $\rho$  y  $\lambda$  mediante el kernel de transición normal. Finalmente, se encuentra que las tasas de aceptación de  $\gamma$ ,  $\rho$  y  $\lambda$  son menores bajo el kernel de transición uniforme, excepto para el parámetro  $\rho$ .

En las figuras 3.9 y 3.10 se muestran las cadenas de los parámetros del modelo general heterocedástico cuando la generación de muestras de  $\rho$  y  $\lambda$  se hace con el kernel de transición uniforme y con el kernel de transición normal, respectivamente.

FIGURA 3.9. Cadenas de los parámetros del modelo general heterocedástico con kernel de transición uniforme

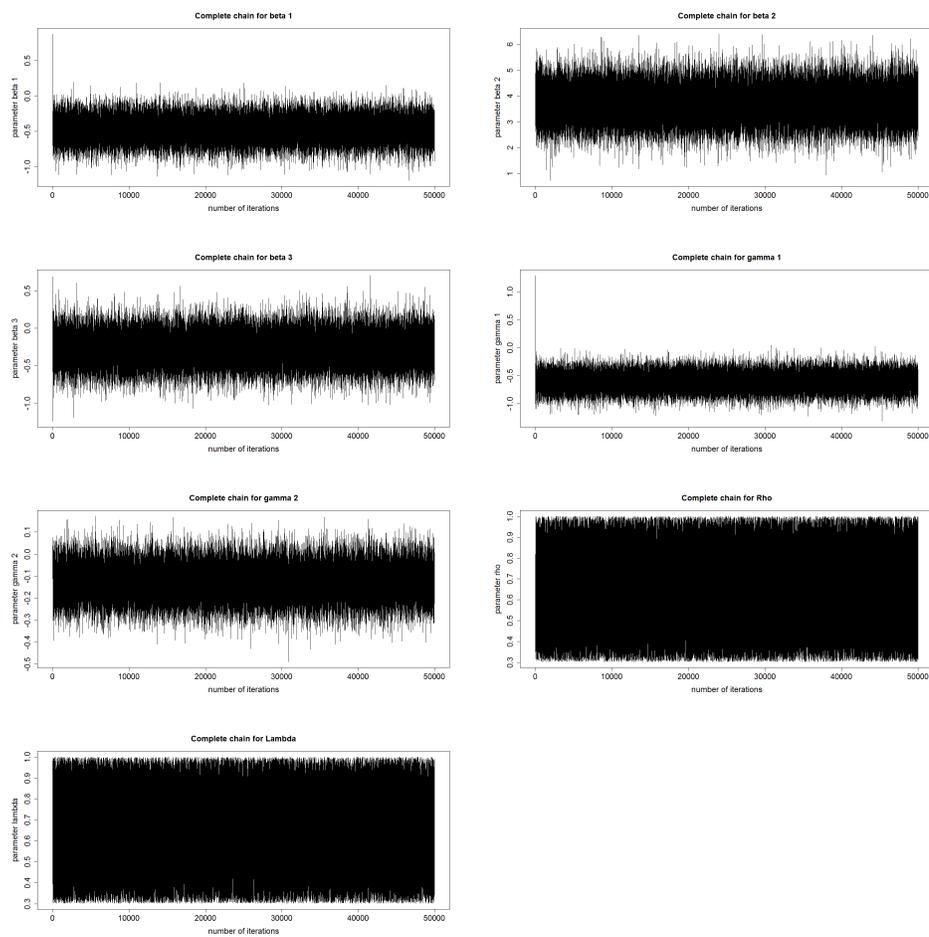
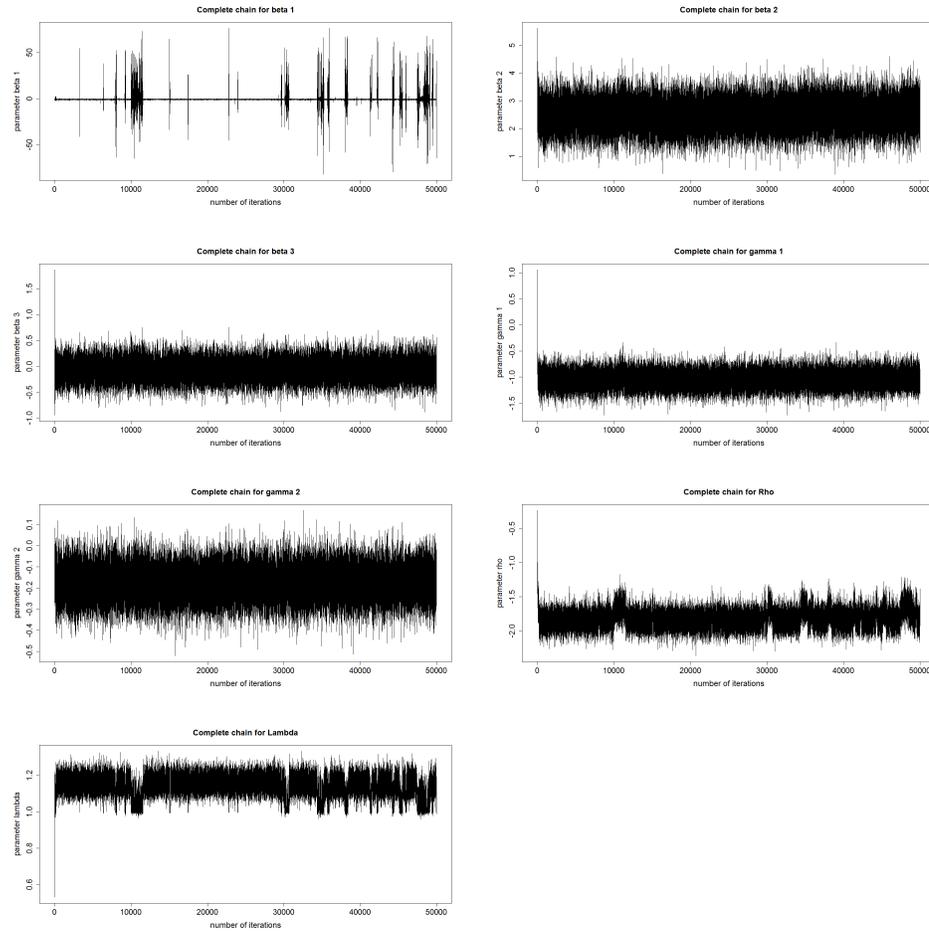


FIGURA 3.10. Cadenas de los parámetros del modelo general heterocedástico con kernel de transición normal



De nuevo se observa la rápida convergencia del algoritmo bajo las dos propuestas realizadas. A pesar de los valores irregulares de las cadenas de  $\rho$  y  $\lambda$  con el kernel de transición uniforme, donde el rango de las cadenas está entre 0.1 y 1, y con el kernel de transición normal, donde la cadena toma valores por fuera del intervalo  $(-1, 1)$ .

Los histogramas con los residuales calculados al ajustar el modelo general heterocedástico mediante los kernels de transición uniforme y normal se muestran en las figuras 3.11 y 3.12, respectivamente. Mientras que para el ajuste realizado con el kernel de transición normal se observan muchos residuales alejados de cero y una distribución apuntada. Para los residuales obtenidos con el kernel uniforme se observa una forma de campana con presencia de un valor atípico.

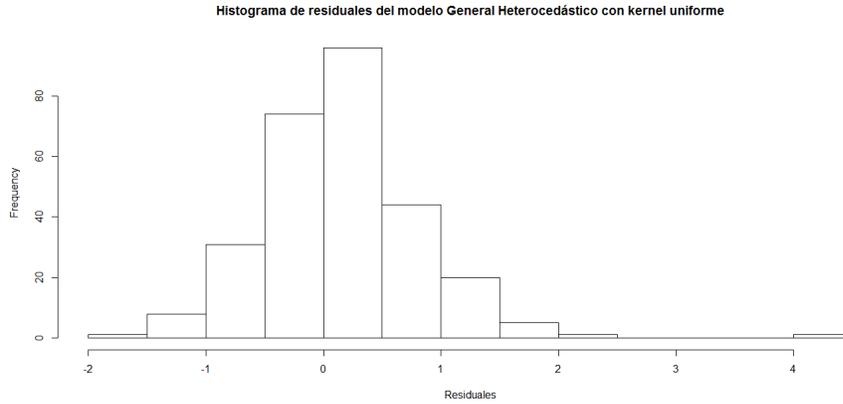


FIGURA 3.11. Residuales para el modelo general heterocedástico ajustado con el kernel de transición uniforme para los datos de leucemia en Nueva York, [62]

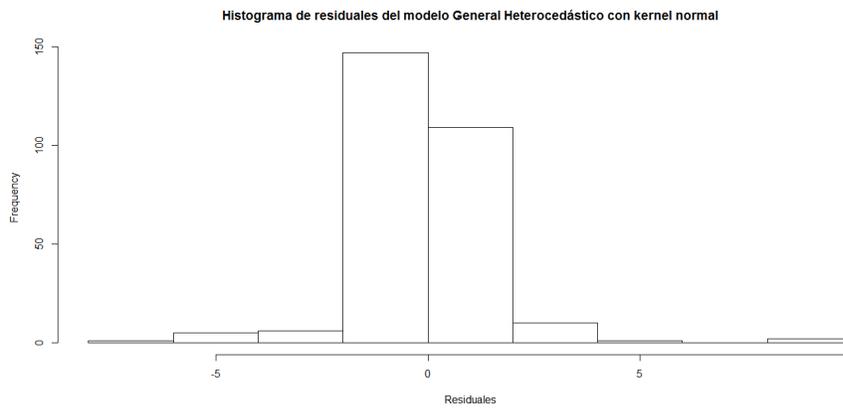


FIGURA 3.12. Residuales para el modelo general heterocedástico ajustado con el kernel de transición normal para los datos de leucemia en Nueva York, [62]

Se evaluaron las medidas de tendencia central de estos residuales, así como si siguen una distribución normal, los resultados se encuentran en la tabla 3.9

TABLA 3.9. Medidas resumen y test de normalidad para los residuales del ajuste del modelo General heterocedástico bajo las dos propuestas

	Kernel Uniforme	Kernel Normal
Mínimo	-1,699	-6,074
Media	0,151	-0,144
Mediana	0,129	-0,126
Máximo	4,038	9,752
p-valor (Shapiro Test)	1,32E-06	<2,2E-16

Estos resultados muestran que las estimaciones logradas a partir de ambas metodologías produce algunos residuales que se alejan de cero. También se revela que las distribuciones de estos residuales no son normales, en el caso normal es originado por la presencia de

valores atípicos y el nivel de apuntamiento de la distribución de los residuales, por otra parte los residuales obtenidos por el kernel de transición uniforme presentan un mejor comportamiento, con residuales más cercanos a cero y un p-valor únicamente influenciado por un valor atípico.

---

---

## Discusión y Conclusiones

---

---

### 4.1. Discusión

En los capítulos 2 y 3 se desarrollaron dos propuestas de ajuste Bayesiano para los modelos econométricos espaciales Bayesianos. Estas tenían como motivación las dificultades que presentan las metodologías clásicas de ajuste de modelos econométricos como Máxima Verosimilitud cuando hay tamaños de muestra pequeños u observaciones atípicas. Ahora se verá si los estimadores obtenidos en las simulaciones de este trabajo son insesgadas. La revisión se hizo para los modelos SAR y SEM, tanto en su versión homocedástica como heterocedástica. Se realizaron 100 simulaciones de los datos, bajo los mismos parámetros de las simulaciones de los capítulos 2 y 3. Los modelos fueron ajustados para cada uno de estos datos simulados y las estimaciones obtenidas fueron comparadas respecto a los verdaderos valores de los parámetros y puestas en boxplots. Los resultados por modelo son presentados y discutidos a continuación.

#### 4.1.1. Modelo SAR homocedástico

Las diferencias entre las 100 estimaciones y los valores verdaderos de los parámetros están plasmadas en la figura 4.1.

Se evidencia la posible existencia de sesgo tanto para el parámetro  $\beta_0$ , como para  $\sigma^2$ , puesto que la mediana se encuentra alejada de cero, lo contrario sucede con los parámetros  $\beta_1$ ,  $\beta_2$  y  $\rho$ .

#### 4.1.2. Modelo SEM homocedástico

Los boxplot de las diferencias entre estimaciones y valores reales de los parámetros están en la figura 4.2.

Este caso muestra de nuevo que las distancias para los parámetros  $\beta_1$ ,  $\beta_2$  y  $\lambda$  se encuentran muy próximas a 0. Por otra parte, las medianas de las diferencias de las estimaciones de  $\beta_0$  y  $\sigma^2$  son cercanas a cero, a pesar de las altas dispersiones que reflejan los boxplot. También se evidencia la presencia de diferencias atípicas para cada parámetro del modelo.

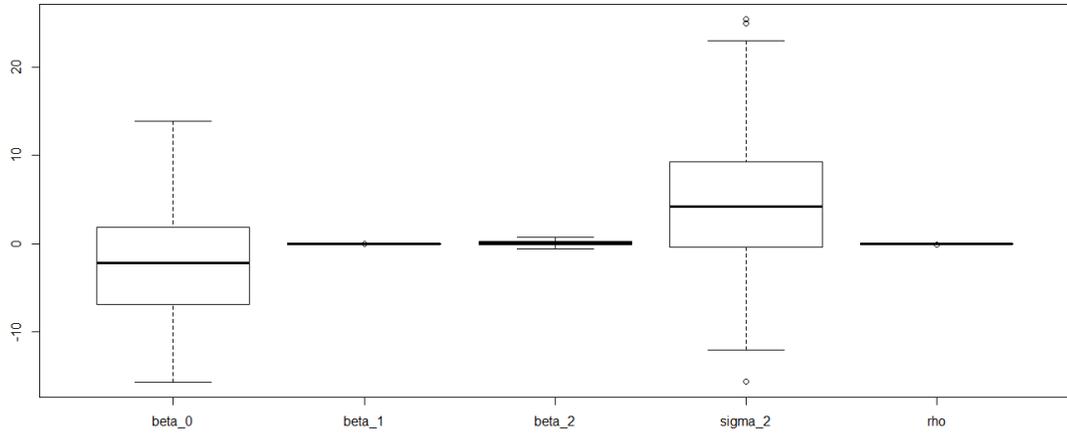


FIGURA 4.1. Diferencia entre estimaciones y valores verdaderos para el modelo SAR Homocedástico

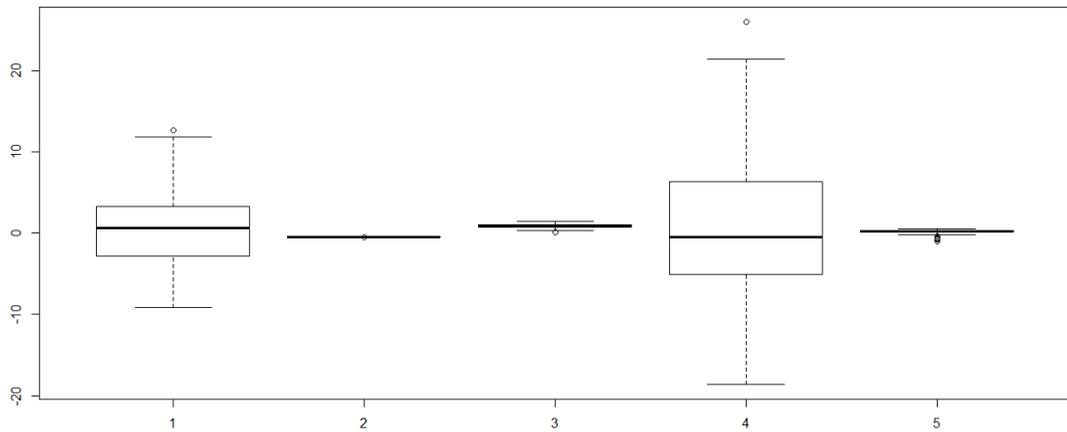


FIGURA 4.2. Diferencia entre estimaciones y valores verdaderos para el modelo SEM homocedástico

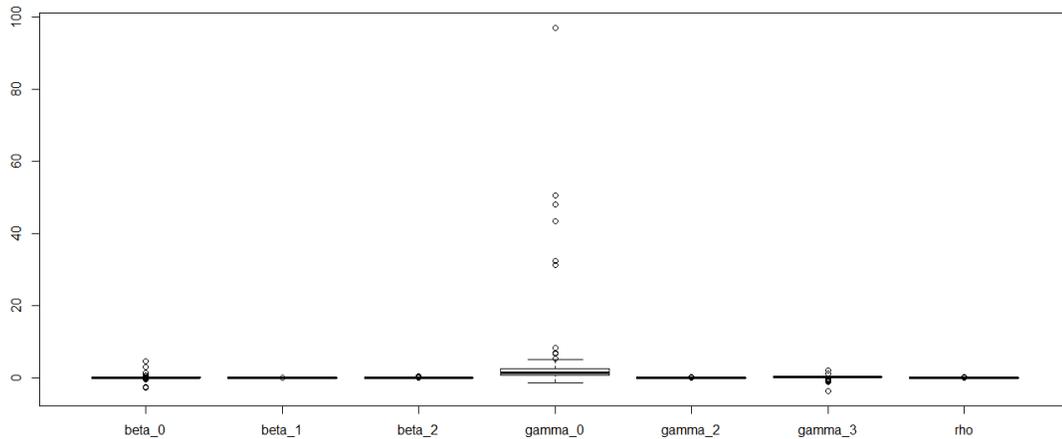


FIGURA 4.3. Diferencia entre estimaciones y valores verdaderos para el modelo SAR heterocedástico

#### 4.1.3. Modelo SAR heterocedástico

En la figura 4.3 se presentan las diferencias entre las 100 estimaciones y los valores verdaderos de los parámetros del modelo SAR heterocedástico.

Los boxplot denotan resultados mucho más favorables que los de su par homocedástico, ya que las diferencias de todos los parámetros se encuentran próximas a 0. No obstante, es alto el número de estimaciones atípicas, especialmente para el parámetro  $\gamma_0$ , cuyas diferencias son las más alejadas de 0 y las más variables de los siete parámetros estimados.

#### 4.1.4. Modelo SEM heterocedástico

La figura 4.4 presenta las diferencias entre las 100 estimaciones y los valores verdaderos de los parámetros del modelo SEM heterocedástico.

Para el modelo SEM heterocedástico, las medianas de las diferencias entre las estimaciones y los valores reales de los parámetros se encuentran muy próximas a cero, a excepción de  $\gamma_0$ , para el cual se observa alta variabilidad, además de encontrarse muy distante de cero.

## 4.2. Conclusiones y Extensiones

1. En este trabajo se especificó un grupo de modelos econométricos espaciales Bayesianos, tomando como referencia los modelos propuestos en [8]. Mientras unos modelos suponían homocedasticidad, los otros tenían en cuenta que en la dinámica econométrica es frecuente que haya heterogeneidad a través del espacio. La especificación de los modelos Bayesianos incluyó distribuciones a priori para los modelos homocedásticos, siguiendo las propuestas hechas en [42]. Mientras que para los modelos heterocedásticos las distribuciones a priori fueron realizadas siguiendo lo hecho en [22] y [24].

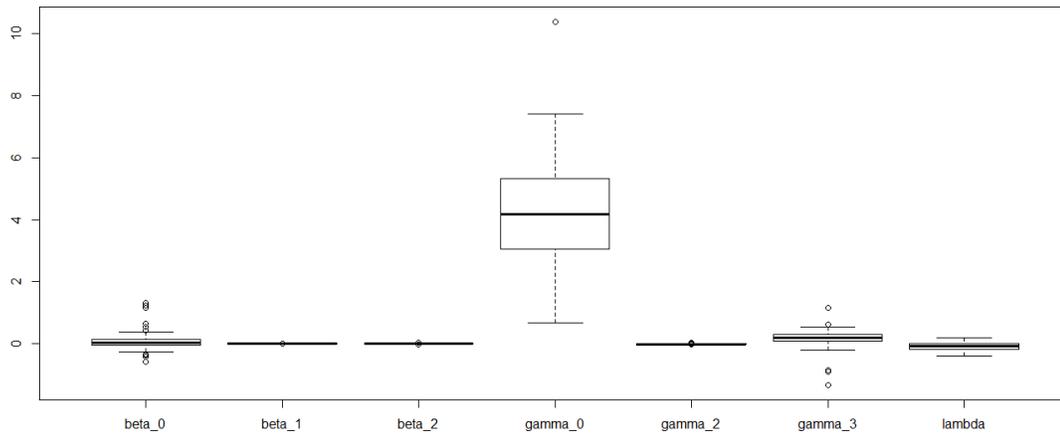


FIGURA 4.4. Diferencia entre estimaciones y valores verdaderos para el modelo SEM heterocedástico

2. Se realizó una extensión de la propuesta metodológica en [22] y [24] para el ajuste de modelos econométricos espaciales. Primero, se demostró el cálculo de las distribuciones condicionales a posteriori para cada uno de los parámetros involucrados en los modelos propuestos. Posteriormente, para los parámetros de rezago espacial con distribuciones condicionales a posteriori difíciles de muestrear se propuso un kernel de transición uniforme y uno normal. La propuesta uniforme está basada en la propuesta empírica de [8] y la normal en la forma de la distribución. En los modelos heterocedásticos se hizo una propuesta basada en variables de trabajo que aproxima la distribución condicional a posteriori de  $\gamma$ . Esta propuesta está basada en [22] y [24].
3. Se elaboró la librería *BSPADATA* en R, la cual consta de 6 funciones encargadas del ajuste por las metodologías propuestas de los modelos SAR, SEM y General en sus versiones homocedástica y heterocedástica. Este paquete es importante, teniendo en cuenta que no existe ninguna librería o función de R, [54], dedicada al ajuste mediante algoritmos MCMC de este grupo de modelos.
4. Con el fin de evaluar la propuesta metodológica, se realizaron aplicaciones con datos reales y simulaciones, las cuales aportaron muy buenos resultados en términos de estimación puntual de parámetros, errores estándar, Criterio de Información Bayesiano (BIC), Deviance Information Criterion (DIC) y tasas de aceptación. Además, el algoritmo desarrollado para cada uno de los modelos arrojó muy buenos resultados en términos de convergencia y eficiencia.
5. La propuesta de ajuste para los modelos econométricos espaciales arroja resultados muy similares a los obtenidos mediante Máxima Verosimilitud, en términos de errores estándar de las estimaciones y del Criterio de Información Bayesiano (BIC).
6. Con el trabajo realizado con los datos reales y simulados se encontró que a pesar de que el ajuste con el kernel de transición uniforme es más conservador, ya que garantiza desde el comienzo que las estimaciones estarán al interior del intervalo teórico  $(-1/\omega_{max}, 1)$ , el kernel de transición normal ofrece mejores resultados, al

generar menores errores estándar, estimaciones más similares a las obtenidas por Máxima Verosimilitud y más próximas a la realidad de los datos simulados. Además, la estimación de los parámetros de rezago espacial a través del kernel de transición uniforme depende únicamente de la matriz de contigüidad espacial, mientras que la propuesta con kernel de transición normal incorpora la información de la muestra, haciendo más funcional esta propuesta.

7. Es posible realizar este algoritmo con la ayuda de matrices dispersas, tal como lo exhibe [42], teniendo en cuenta la cantidad de ceros al interior de las matrices de vecindad espacial.
8. En [31] se hace referencia a un conjunto de modelos más general respecto al acá trabajado. Es posible extender la propuesta Bayesiana realizada para los modelos que no fueron trabajados. A pesar de que este trabajo esté diseñado para modelos normales, puede ser extendido para modelos con otros supuestos distribucionales.
9. En términos de sesgo, la propuesta Bayesiana empleando un kernel de transición normal arroja mejores resultados para los modelos heterocedásticos, ya que las diferencias entre estimaciones y valores reales son muy próximas a cero para gran parte de los parámetros de los modelos.

---



---

## Demostraciones

---



---

### A.1. Modelos homocedásticos

#### A.1.1. Modelo general homocedástico

La distribución condicional a posteriori se encuentra en (2.5), y está dada por:

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma^2, \rho, \lambda) \propto (\sigma^2)^{-n/2} |\mathbf{A}| |\mathbf{D}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \\ \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \frac{\zeta^\vartheta}{\Gamma(\vartheta)} (\sigma^2)^{-\vartheta-1} \exp \left( -\frac{\zeta}{\sigma^2} \right) \quad (\text{A.1}) \end{aligned}$$

con  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}_2$ .

A continuación se obtendrán las distribuciones condicionales a posteriori de  $\boldsymbol{\beta}$ ,  $\sigma^2$ ,  $\rho$  y  $\lambda$ .

##### A.1.1.1. Distribución condicional a posteriori de $\boldsymbol{\beta}$

Al condicionar (2.5) respecto a  $\sigma^2$ ,  $\rho$  y  $\lambda$  se obtiene:

$$\pi(\boldsymbol{\beta} | \sigma^2, \rho, \lambda) \propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}_2$ . Entonces,

$$\begin{aligned} \pi(\boldsymbol{\beta} | \sigma^2, \rho, \lambda) \propto \exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma^2} (-\mathbf{y}' \mathbf{A}' \mathbf{D}' \mathbf{D} \mathbf{X} \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \boldsymbol{\beta}' \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} \boldsymbol{\beta}) \right. \right. \\ \left. \left. - (\boldsymbol{\beta}' \mathbf{B}^{-1} \boldsymbol{\beta} - \boldsymbol{\beta}' \mathbf{B}^{-1} \mathbf{b} - \mathbf{b}' \mathbf{B}^{-1} \boldsymbol{\beta}) \right] \right\} \\ \propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right) \boldsymbol{\beta} - \left( \frac{1}{\sigma^2} \mathbf{y}' \mathbf{A}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{b}' \mathbf{B}^{-1} \right) \boldsymbol{\beta} \right. \right. \\ \left. \left. - \boldsymbol{\beta}' \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \right] \right\} \end{aligned}$$

Ahora, multiplicando por  $\exp \left\{ -\frac{1}{2} \left[ \left( \frac{1}{\sigma^2} \mathbf{y}' \mathbf{A}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{b}' \mathbf{B}^{-1} \right) \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \right] \right\}$  se obtiene:

$$\begin{aligned} \pi(\boldsymbol{\beta} | \sigma^2, \rho, \lambda) &\propto \exp \left\{ -\frac{1}{2} \left[ \left( \frac{1}{\sigma^2} \mathbf{y}' \mathbf{A}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{b}' \mathbf{B}^{-1} \right) \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \right. \right. \\ &\quad \left. \left. \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \right] \right\} \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right) \boldsymbol{\beta} \right. \right. \\ &\quad \left. \left. - \left( \frac{1}{\sigma^2} \mathbf{y}' \mathbf{A}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{b}' \mathbf{B}^{-1} \right) \boldsymbol{\beta} - \boldsymbol{\beta}' \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' - \left( \frac{1}{\sigma^2} \mathbf{y}' \mathbf{A}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{b}' \mathbf{B}^{-1} \right) \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \right] \right. \\ &\quad \left. \left[ \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right) \right] \left[ \boldsymbol{\beta} - \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \right] \right\} \end{aligned}$$

Luego, la distribución condicional a posteriori de  $\boldsymbol{\beta}$  es normal,  $N(\mathbf{b}^*, \mathbf{B}^*)$ , con:

$$\begin{aligned} \mathbf{b}^* &= \mathbf{B}^* \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right) \\ \mathbf{B}^* &= \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}_2$ .

#### A.1.1.2. Distribución condicional a posteriori para $\sigma^2$

A partir de (2.5) se obtiene la distribución condicional a posteriori de  $\sigma^2$ , la cual está dada por:

$$\pi(\sigma^2 | \boldsymbol{\beta}, \rho, \lambda) \propto (\sigma^2)^{-n/2} \exp \left\{ -\frac{k}{2\sigma^2} \right\} (\sigma^2)^{-\zeta-1} \exp \left\{ -\frac{\nu}{\sigma^2} \right\}$$

donde  $k = (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})$ ,  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1$ ,  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}_2$ .  $\zeta$  y  $\nu$  son los parámetros de la distribución a priori de  $\sigma^2$ . Luego,

$$\pi(\sigma^2 | \boldsymbol{\beta}, \rho, \lambda) \propto (\sigma^2)^{-n/2-\zeta-1} \exp \left\{ -\frac{1}{\sigma^2} \left( \frac{k}{2} + \zeta \right) \right\}$$

Entonces,

$$\sigma^2 | \boldsymbol{\beta}, \rho, \lambda \sim \text{GammaInv} \left( \frac{n}{2} + \zeta, \frac{k}{2} + \nu \right)$$

#### A.1.1.3. Distribución condicional a posteriori para $\rho$

En principio, la distribución condicional a posteriori de  $\rho$  está dada por:

$$\pi(\rho | \boldsymbol{\beta}, \sigma^2, \lambda) \propto |\mathbf{A}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{A} \mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

Entonces,

$$\begin{aligned}
 \pi(\rho|\boldsymbol{\beta}, \sigma^2, \lambda) &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ \frac{1}{2\sigma^2} [\mathbf{y}'\mathbf{A}'\mathbf{D}'\mathbf{D}\mathbf{A}\mathbf{y} - \mathbf{y}'\mathbf{A}'\mathbf{D}'\mathbf{D}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{D}'\mathbf{D}\mathbf{A}\mathbf{y}] \right\} \\
 &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2\sigma^2} [\mathbf{y}'(\mathbf{I} - \rho\mathbf{W}_1)\mathbf{D}'\mathbf{D}(\mathbf{I} - \rho\mathbf{W}_1)\mathbf{D}'\mathbf{D}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{D}'\mathbf{D}(\mathbf{I} - \rho\mathbf{W}_1)\mathbf{y}] \right\} \\
 &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2\sigma^2} [-\rho\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{y} - \rho\mathbf{y}'\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y} + \rho^2\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y} \right. \\
 &\quad \left. + \rho\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{X}\boldsymbol{\beta} + \rho\boldsymbol{\beta}'\mathbf{X}'\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}] \right\} \\
 &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2\sigma^2} [\rho^2(\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}) - 2\rho(\mathbf{y}'\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y})] \right\} \\
 &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2\sigma^2} [\rho^2(\mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}) - 2\rho((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y})] \right\} \\
 &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{c}{2\sigma^2} \left[ \rho^2 - 2\rho\frac{d}{c} + \frac{d^2}{c^2} - \frac{d^2}{c^2} \right] \right\} \\
 &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{c}{2\sigma^2} \left( \rho - \frac{d}{c} \right)^2 \right\}
 \end{aligned}$$

donde  $c = \mathbf{y}'\mathbf{W}'_1\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}$  y  $d = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{D}'\mathbf{D}\mathbf{W}_1\mathbf{y}$ .

#### A.1.1.4. Distribución condicional a posteriori para $\lambda$

Condicionando por  $\boldsymbol{\beta}$ ,  $\sigma^2$  y  $\rho$ , se obtiene la siguiente distribución a posteriori:

$$\pi(\lambda|\boldsymbol{\beta}, \sigma^2, \rho) \propto |\mathbf{D}| \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{D}'\mathbf{D}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ . Luego,

$$\pi(\lambda|\boldsymbol{\beta}, \sigma^2, \rho) \propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ \frac{1}{2\sigma^2} [(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{I} - \lambda\mathbf{W}_2 - \lambda\mathbf{W}'_2 + \lambda^2\mathbf{W}'_2\mathbf{W}_2)(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] \right\}$$

Definiendo  $\mathbf{e} = \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ , se obtiene:

$$\pi(\lambda|\boldsymbol{\beta}, \sigma^2, \rho) \propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ \frac{1}{2\sigma^2} [\lambda^2(\mathbf{e}'\mathbf{W}'_2\mathbf{W}_2\mathbf{e}) - 2\lambda(\mathbf{e}'\mathbf{W}_2\mathbf{e})] \right\}$$

Si se toma  $c = \mathbf{e}'\mathbf{W}'_2\mathbf{W}_2\mathbf{e}$  y  $d = \mathbf{e}'\mathbf{W}_2\mathbf{e}$ , se tiene:

$$\begin{aligned}
 \pi(\lambda|\boldsymbol{\beta}, \sigma^2, \rho) &\propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ -\frac{c}{2\sigma^2} \left[ \lambda^2 - 2\lambda\frac{d}{c} + \frac{b^2}{c^2} - \frac{d^2}{c^2} \right] \right\} \\
 &\propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ -\frac{c}{2\sigma^2} \left( \lambda - \frac{d}{c} \right)^2 \right\}
 \end{aligned}$$

donde  $c = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2\mathbf{W}_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $d = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

Con estos resultados para el modelo general homocedástico se pueden obtener las fórmulas de los modelos SAR y SEM en su versión homocedástica, cuando  $\rho = 0$  y  $\lambda = 0$ , respectivamente.

### A.1.2. Modelo SAR homocedástico

#### A.1.2.1. Distribución condicional a posteriori de $\beta$

Tomando los parámetros de la distribución normal obtenida en el modelo general y haciendo  $\lambda = 0$ , se obtiene una distribución normal  $N(\mathbf{b}^*, \mathbf{B}^*)$ , donde:

$$\mathbf{b}^* = \mathbf{B}^* \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{A} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right)$$

$$\mathbf{B}^* = \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{X} + \mathbf{B}^{-1} \right)^{-1}$$

con  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}$ .

#### A.1.2.2. Distribución condicional a posteriori de $\sigma^2$

Los parámetros de la distribución Gamma Inversa,  $GammaInv(\zeta^*, \nu^*)$ , obtenidos para el modelo general, son los siguientes cuando  $\lambda = 0$ :

$$\zeta^* = \frac{n}{2} + \zeta$$

y

$$\nu^* = \frac{k}{2} + \nu$$

donde  $k = (\mathbf{A} \mathbf{y} - \mathbf{X} \beta)' (\mathbf{A} \mathbf{y} - \mathbf{X} \beta)$  y  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}$ .  $\zeta$  y  $\nu$  son los parámetros de la distribución a priori de  $\sigma^2$ .

#### A.1.2.3. Distribución condicional a posteriori de $\rho$

Cuando  $\lambda = 0$  se obtiene la siguiente distribución condicional a posteriori de  $\rho$ :

$$\pi(\rho | \beta, \sigma^2, \lambda) \propto |\mathbf{I} - \rho \mathbf{W}_1| \exp \left\{ -\frac{c^*}{2\sigma^2} \left( \rho - \frac{d^*}{c^*} \right)^2 \right\}$$

donde  $c^* = \mathbf{y}' \mathbf{W}' \mathbf{W} \mathbf{y}$  y  $d^* = (\mathbf{y} - \mathbf{X} \beta)' \mathbf{W} \mathbf{y}$ .

### A.1.3. Modelo SEM homocedástico

#### A.1.3.1. Distribución condicional a posteriori de $\beta$

Tomando los parámetros de la distribución normal obtenida en el modelo general y haciendo  $\rho = 0$ , se obtiene una distribución normal,  $N(\mathbf{b}^*, \mathbf{B}^*)$ , donde:

$$\mathbf{b}^* = \mathbf{B}^* \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{y} + \mathbf{B}^{-1} \mathbf{b} \right)$$

$$\mathbf{B}^* = \left( \frac{1}{\sigma^2} \mathbf{X}' \mathbf{D}' \mathbf{D} \mathbf{X} + \mathbf{B}^{-1} \right)^{-1}$$

con  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}$ .

### A.1.3.2. Distribución condicional a posteriori de $\sigma^2$

Los parámetros de la distribución Gamma Inversa,  $GammaInv(\zeta^*, \nu^*)$ , obtenidos para el modelo general, son los siguientes cuando  $\rho = 0$ :

$$\zeta^* = \frac{n}{2} + \zeta$$

y

$$\nu^* = \frac{k}{2} + \nu$$

donde  $k = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \mathbf{D} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}$ .  $\zeta$  y  $\nu$  son los parámetros de la distribución a priori de  $\sigma^2$ .

### A.1.3.3. Distribución condicional a posteriori de $\lambda$

Cuando  $\rho = 0$  se obtiene la siguiente distribución condicional a posteriori de  $\lambda$ :

$$\pi(\lambda | \boldsymbol{\beta}, \sigma^2, \rho) \propto |\mathbf{I} - \lambda \mathbf{W}| \exp \left\{ -\frac{c^*}{2\sigma^2} \left( \lambda - \frac{d^*}{c^*} \right)^2 \right\}$$

donde  $c^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W}' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $d^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{W} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$

## A.2. Modelos heterocedásticos

### A.2.1. Modelo general heterocedástico

La distribución completa a posteriori se encuentra en (3.3) y está dada por:

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho, \lambda) \propto |\boldsymbol{\Sigma}|^{-1/2} |\mathbf{A}| |\mathbf{D}| \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\}$$

$$\exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})' \mathbf{G}^{-1} (\boldsymbol{\gamma} - \mathbf{g}) \right\} \quad (\text{A.2})$$

donde  $\mathbf{A} = \mathbf{I} - \rho \mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda \mathbf{W}_2$ . A continuación se obtendrán las distribuciones condicionales a posteriori de  $\boldsymbol{\beta}$ ,  $\boldsymbol{\gamma}$ ,  $\rho$  y  $\lambda$ :

#### A.2.1.1. Distribución condicional a posteriori de $\boldsymbol{\beta}$

Al condicionar (3.3) respecto a  $\boldsymbol{\gamma}$ ,  $\rho$  y  $\lambda$  se obtiene:

$$\pi(\boldsymbol{\beta} | \boldsymbol{\gamma}, \rho, \lambda) \propto \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta} - \mathbf{b})' \mathbf{B}^{-1} (\boldsymbol{\beta} - \mathbf{b}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ . Entonces,

$$\begin{aligned} \pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \rho, \lambda) &\propto \exp \left\{ -\frac{1}{2} \left[ (-\mathbf{y}'\mathbf{A}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X}\boldsymbol{\beta}) \right. \right. \\ &\quad \left. \left. - (\boldsymbol{\beta}'\mathbf{B}^{-1}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{B}^{-1}\mathbf{b} - \mathbf{b}'\mathbf{B}^{-1}\boldsymbol{\beta}) \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right) \boldsymbol{\beta} - \left( \mathbf{y}'\mathbf{A}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{b}'\mathbf{B}^{-1} \right) \boldsymbol{\beta} \right. \right. \\ &\quad \left. \left. - \boldsymbol{\beta}' \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \right] \right\} \end{aligned}$$

Ahora, multiplicando por  $\exp \left\{ -\frac{1}{2} \left[ \left( \mathbf{y}'\mathbf{A}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{b}'\mathbf{B}^{-1} \right) \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \right] \right\}$ , se obtiene:

$$\begin{aligned} \pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \rho, \lambda) &\propto \exp \left\{ -\frac{1}{2} \left[ \left( \mathbf{y}'\mathbf{A}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{b}'\mathbf{B}^{-1} \right) \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \right. \right. \\ &\quad \left. \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \right\} \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right) \boldsymbol{\beta} \right. \right. \\ &\quad \left. \left. - \left( \mathbf{y}'\mathbf{A}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{b}'\mathbf{B}^{-1} \right) \boldsymbol{\beta} - \boldsymbol{\beta}' \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}' - \left( \mathbf{y}'\mathbf{A}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{b}'\mathbf{B}^{-1} \right) \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \right] \right. \\ &\quad \left. \left[ \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right) \right] \left[ \boldsymbol{\beta} - \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \right] \right\} \end{aligned}$$

Luego, la distribución condicional a posteriori de  $\boldsymbol{\beta}$  es normal,  $N(\mathbf{b}^*, \mathbf{B}^*)$ , con:

$$\begin{aligned} \mathbf{b}^* &= \mathbf{B}^* \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \\ \mathbf{B}^* &= \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \end{aligned}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ .

#### A.2.1.2. Distribución condicional a posteriori de $\boldsymbol{\gamma}$

A partir de (3.3) se obtiene la distribución condicional a posteriori de  $\boldsymbol{\gamma}$ , la cual está dada por:

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \rho, \lambda) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{D}' \boldsymbol{\Sigma}^{-1} \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})' \mathbf{G}^{-1} (\boldsymbol{\gamma} - \mathbf{g}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{B} = \mathbf{I} - \lambda\mathbf{W}_2$ . Se observa que esta distribución condicional a posteriori no corresponde a ninguna forma funcional conocida, es por esto que se realizó

la propuesta basada en variables de trabajo, [22], [24], con el fin de obtener muestras para estimar  $\gamma$ .

### A.2.1.3. Distribución condicional a posteriori para $\rho$

En principio, la distribución condicional a posteriori de  $\rho$  está dada por:

$$\pi(\rho|\beta, \gamma, \lambda) \propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2}(\mathbf{A}\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}' \Sigma^{-1} \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\beta) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ . Entonces,

$$\begin{aligned} \pi(\rho|\beta, \sigma^2, \lambda) &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2} [\mathbf{y}' \mathbf{A}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{A} \mathbf{y} - \mathbf{y}' \mathbf{A}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{X} \beta - \beta' \mathbf{X}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{A} \mathbf{y}] \right\} \\ &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2} [\mathbf{y}' (\mathbf{I} - \rho\mathbf{W}_1) \mathbf{D}' \Sigma^{-1} \mathbf{D} (\mathbf{I} - \rho\mathbf{W}_1) \mathbf{y} - \mathbf{y}' (\mathbf{I} - \rho\mathbf{W}_1) \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{X} \beta \right. \\ &\quad \left. - \beta' \mathbf{X}' \mathbf{D}' \Sigma^{-1} \mathbf{D} (\mathbf{I} - \rho\mathbf{W}_1) \mathbf{y}] \right\} \\ &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2} [-\rho \mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{y} - \rho \mathbf{y}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y} \right. \\ &\quad \left. + \rho^2 \mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y} + \rho \mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{X} \beta + \rho \beta' \mathbf{X}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}] \right\} \\ &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2} [\rho^2 (\mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}) - 2\rho (\mathbf{y}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y} \right. \\ &\quad \left. - \beta' \mathbf{X}' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y})] \right\} \\ &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{1}{2} [\rho^2 (\mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}) - 2\rho ((\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y})] \right\} \\ &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{c}{2} \left[ \rho^2 - 2\rho \frac{d}{c} + \frac{b^2}{c^2} - \frac{d^2}{c^2} \right] \right\} \\ &\propto |\mathbf{I} - \rho\mathbf{W}_1| \exp \left\{ -\frac{c}{2} \left( \rho - \frac{d}{c} \right)^2 \right\} \end{aligned}$$

donde  $c = \mathbf{y}' \mathbf{W}'_1 \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}$  y  $d = (\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}' \Sigma^{-1} \mathbf{D} \mathbf{W}_1 \mathbf{y}$ .

### A.2.1.4. Distribución condicional a posteriori para $\lambda$

Condicionando por  $\beta$ ,  $\gamma$  y  $\rho$ , se obtiene la siguiente distribución a posteriori:

$$\pi(\lambda|\beta, \gamma, \rho) \propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ -\frac{1}{2}(\mathbf{A}\mathbf{y} - \mathbf{X}\beta)' \mathbf{D}' \Sigma^{-1} \mathbf{D} (\mathbf{A}\mathbf{y} - \mathbf{X}\beta) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}_1$  y  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}_2$ . Luego,

$$\pi(\lambda|\beta, \gamma, \rho) \propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ -\frac{1}{2} [(\mathbf{A}\mathbf{y} - \mathbf{X}\beta)' (\Sigma^{-1} - \lambda \Sigma^{-1} \mathbf{W}_2 - \lambda \mathbf{W}'_2 \Sigma^{-1} + \lambda^2 \mathbf{W}'_2 \Sigma^{-1} \mathbf{W}_2) (\mathbf{A}\mathbf{y} - \mathbf{X}\beta)] \right\}$$

Definiendo  $\mathbf{e} = \mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ , se obtiene:

$$\pi(\lambda|\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho) \propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ \frac{1}{2} \left[ \lambda^2 (\mathbf{e}'\mathbf{W}'_2\boldsymbol{\Sigma}^{-1}\mathbf{W}_2\mathbf{e}) - 2\lambda (\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{W}_2\mathbf{e}) \right] \right\}$$

Si se toma  $c = \mathbf{e}'\mathbf{W}'_2\boldsymbol{\Sigma}^{-1}\mathbf{W}_2\mathbf{e}$  y  $d = \mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{W}_2\mathbf{e}$ , se tiene:

$$\begin{aligned} \pi(\lambda|\boldsymbol{\beta}, \boldsymbol{\gamma}, \rho) &\propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ -\frac{c}{2} \left[ \lambda^2 - 2\lambda \frac{d}{c} + \frac{b^2}{c^2} - \frac{d^2}{c^2} \right] \right\} \\ &\propto |\mathbf{I} - \lambda\mathbf{W}_2| \exp \left\{ -\frac{c}{2} \left( \lambda - \frac{d}{c} \right)^2 \right\} \end{aligned}$$

donde  $c = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}'_2\boldsymbol{\Sigma}^{-1}\mathbf{W}_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $d = (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}\mathbf{W}_2(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ .

Con estos resultados para el modelo general heterocedástico se pueden obtener las fórmulas de los modelos SAR y SEM heterocedásticos.

## A.2.2. Modelo SAR heterocedástico

### A.2.2.1. Distribución condicional a posteriori de $\boldsymbol{\beta}$

Tomando los parámetros de la distribución normal obtenida en el modelo general heterocedástico y haciendo  $\lambda = 0$ , se obtiene una distribución normal  $N(\mathbf{b}^*, \mathbf{B}^*)$ , donde:

$$\begin{aligned} \mathbf{b}^* &= \mathbf{B}^* \left( \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right) \\ \mathbf{B}^* &= \left( \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1} \end{aligned}$$

con  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$ .

### A.2.2.2. Distribución condicional a posteriori de $\boldsymbol{\gamma}$

A partir de (3.3), y haciendo  $\lambda = 0$ , se obtiene la distribución condicional a posteriori de  $\boldsymbol{\gamma}$ , la cual está dada por:

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \rho) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{A}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2} (\boldsymbol{\gamma} - \mathbf{g})'\mathbf{G}^{-1}(\boldsymbol{\gamma} - \mathbf{g}) \right\}$$

donde  $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$ . Se observa que esta distribución condicional a posteriori no corresponde a ninguna forma funcional conocida.

### A.2.2.3. Distribución condicional a posteriori de $\rho$

Cuando  $\lambda = 0$  se obtiene la siguiente distribución condicional a posteriori de  $\rho$  a partir de (3.3):

$$\pi(\rho|\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda) = |\mathbf{I} - \rho\mathbf{W}| \exp \left\{ -\frac{c^*}{2} \left( \rho - \frac{d^*}{c^*} \right)^2 \right\}$$

donde  $c = \mathbf{y}'\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}$  y  $d^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}$ .

### A.2.3. Modelo SEM heterocedástico

#### A.2.3.1. Distribución condicional a posteriori de $\boldsymbol{\beta}$

Tomando los parámetros de la distribución normal obtenida en el modelo general y haciendo  $\rho = 0$ , se obtiene una distribución normal  $N(\mathbf{b}^*, \mathbf{B}^*)$ , donde:

$$\mathbf{b}^* = \mathbf{B}^* \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{y} + \mathbf{B}^{-1}\mathbf{b} \right)$$

$$\mathbf{B}^* = \left( \mathbf{X}'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}\mathbf{X} + \mathbf{B}^{-1} \right)^{-1}$$

con  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ .

#### A.2.3.2. Distribución condicional a posteriori de $\boldsymbol{\gamma}$

A partir de (3.3), y haciendo  $\rho = 0$ , se obtiene la distribución condicional a posteriori de  $\boldsymbol{\gamma}$ , la cual está dada por:

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta}, \rho) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{D}'\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right\} \exp \left\{ -\frac{1}{2}(\boldsymbol{\gamma} - \mathbf{g})'\mathbf{G}^{-1}(\boldsymbol{\gamma} - \mathbf{g}) \right\}$$

donde  $\mathbf{D} = \mathbf{I} - \lambda\mathbf{W}$ . Se observa que esta distribución condicional a posteriori no corresponde a ninguna forma funcional conocida.

#### A.2.3.3. Distribución condicional a posteriori de $\lambda$

Cuando  $\rho = 0$  se obtienen la siguiente distribución condicional a posteriori de  $\lambda$ :

$$\pi(\lambda|\boldsymbol{\beta}, \sigma^2) = |\mathbf{I} - \lambda\mathbf{W}| \exp \left\{ -\frac{c^*}{2} \left( \lambda - \frac{d^*}{c^*} \right)^2 \right\}$$

donde  $c^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}'\boldsymbol{\Sigma}^{-1}\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$  y  $d^* = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ . Esta forma funcional justifica la propuesta de un kernel de transición normal.

---



---

## Implementación Computacional

---



---

Para implementar la propuesta de este trabajo, se diseñó el paquete *BSPADATA* en el programa *R*, [54]. La aplicación del ajuste presentado en este trabajo se puede hacer a través de las seis funciones de este paquete, una por cada modelo presentado. A continuación se presenta la sintaxis de cada función.

### B.1. Función *hom\_sar*

La sintaxis de la función *hom\_sar* está dada por:

```
hom_sar(y,X,W,nsim,burn,step,b_pri,B_pri,r_pri,lambda_pri,beta_0,sigma2_0,
rho_0,kernel,plot)
```

donde  $\mathbf{y}$ ,  $\mathbf{X}$  y  $\mathbf{W}$  corresponden a la variable respuesta, las variables explicativas y la matriz de vecindad espacial, respectivamente. Por otra parte **nsim**, **burn** y **step**, son el número de simulaciones, el período de calentamiento de la cadena y la longitud del salto de la cadena para el cálculo de las estimaciones. Mientras que **b\_pri**, **B\_pri**, **r\_pri** y **lambda\_pri** corresponden a los parámetros de las distribuciones a priori de  $\beta$  y  $\sigma^2$ . Finalmente, **beta\_0**, **sigma2\_0** y **rho\_0** son los valores de inicio de las cadenas para la estimación de los parámetros. Dependiendo de la distribución escogida para el kernel de transición de  $\rho$ , se puede optar por **kernel**=“uniform” o por **kernel**=“normal”.

### B.2. Función *hom\_sem*

La sintaxis general de esta función es:

```
hom_sem(y,X,W,nsim,burn,step,b_pri,B_pri,r_pri,lambda_pri,beta_0,sigma2_0,
lambda_0,kernel,plot)
```

donde  $\mathbf{y}$ ,  $\mathbf{X}$  y  $\mathbf{W}$  corresponden a la variable respuesta, las variables explicativas y la matriz de vecindad espacial, respectivamente. Por otra parte **nsim**, **burn** y **step**, son el número de simulaciones, el período de calentamiento de la cadena y la longitud del salto de la cadena para el cálculo de las estimaciones. Mientras que **b\_pri**, **B\_pri**, **r\_pri** y **lambda\_pri** corresponden a los parámetros de las distribuciones a priori de  $\beta$  y  $\sigma^2$ . Finalmente, **beta\_0**, **sigma2\_0** y **lambda\_0** son los valores de inicio de las cadenas para la estimación de los parámetros. Dependiendo de la propuesta elegida para el kernel de transición de  $\lambda$ , se puede escoger la opción **kernel**=“uniform” o **kernel**=“normal”.

### B.3. Función *hom\_general*

La sintaxis general de la función es:

```
hom_general(y,X,W1,W2,nsim,burn,step,b_pri,B_pri,r_pri,lambda_pri,beta_0,
sigma2_0,rho_0,lambda_0,kernel,plot,mateq)
```

donde  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\mathbf{W1}$  y  $\mathbf{W2}$  corresponden a la variable respuesta, las variables explicativas y las matrices de vecindad espacial, asumidas iguales por defecto, respectivamente. Por otra parte **nsim**, **burn** y **step**, son el número de simulaciones, el período de calentamiento de la cadena y la longitud del salto de la cadena para el cálculo de las estimaciones. Mientras que **b\_pri**, **B\_pri**, **r\_pri** y **lambda\_pri** corresponden a los parámetros de las distribuciones a priori de  $\beta$  y  $\sigma^2$ . Finalmente, **beta\_0**, **sigma2\_0** y **lambda\_0** son los valores de inicio de las cadenas para la estimación de los parámetros. Dependiendo de la propuesta elegida para el kernel de transición de  $\lambda$ , se puede escoger la opción **kernel**=“uniform” o **kernel**=“normal”. El parámetro **mateq** indica si se asume  $\mathbf{W1} = \mathbf{W2}$ .

### B.4. Función *hetero\_sar*

La sintaxis general de la función *hetero\_sar* está dada por:

```
hetero_sar(y,X,Z,W,nsim,burn,step,b_pri,B_pri,g_pri,G_pri,beta_0,gamma_0,
rho_0,kernel,plot)
```

donde  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\mathbf{Z}$  y  $\mathbf{W}$  corresponden a la variable respuesta, las variables explicativas de la media, las variables explicativas de la varianza y la matriz de vecindad espacial, respectivamente. Por otra parte **nsim**, **burn** y **step**, son el número de simulaciones, el período de calentamiento de la cadena y la longitud del salto de la cadena para el cálculo de las estimaciones. Mientras que **b\_pri**, **B\_pri**, **g\_pri** y **G\_pri** corresponden a los parámetros de las distribuciones a priori de  $\beta$  y  $\gamma$ . Finalmente, **beta\_0**, **gamma\_0** y **rho\_0** son los valores de inicio de las cadenas para la estimación de los parámetros. Dependiendo de la distribución escogida para el kernel de transición de  $\rho$ , se puede optar por **kernel**=“uniform” o **kernel**=“normal”.

### B.5. Función *hetero\_sem*

La sintaxis general de esta función es:

```
hetero_sem(y,X,Z,W,nsim,burn,step,b_pri,B_pri,g_pri,G_pri,beta_0,gamma_0,
lambda_0,kernel,plot)
```

donde  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\mathbf{Z}$  y  $\mathbf{W}$  corresponden a la variable respuesta, las variables explicativas de la media, las variables explicativas de la varianza y la matriz de vecindad espacial, respectivamente. Por otra parte **nsim**, **burn** y **step**, son el número de simulaciones, el período de calentamiento de la cadena y la longitud del salto de la cadena para el cálculo de las estimaciones. Mientras que **b\_pri**, **B\_pri**, **g\_pri** y **G\_pri** corresponden a los parámetros de las distribuciones a priori de  $\beta$  y  $\gamma$ . Finalmente, **beta\_0**, **gamma\_0** y **lambda\_0** son los valores de inicio de las cadenas para la estimación de los parámetros. Dependiendo de la distribución escogida para el kernel de transición de  $\lambda$ , se puede escoger la opción **kernel**=“uniform” o **kernel**=“normal”.

## B.6. Función *hetero\_general*

La sintaxis general de la función *hetero\_general* es:

```
hetero_general(y,X,Z,W1,W2,nsim,burn,step,b_pri,B_pri,g_pri,G_pri,beta_0,  
gamma_0,rho_0,lambda_0,kernel,plot,mateq)
```

donde  $\mathbf{y}$ ,  $\mathbf{X}$ ,  $\mathbf{Z}$ ,  $\mathbf{W1}$  y  $\mathbf{W2}$  corresponden a la variable respuesta, las variables explicativas de la media, las variables explicativas de la varianza y las matrices de vecindad espacial, asumidas por defecto iguales. Por otra parte **nsim**, **burn** y **step** son el número de simulaciones, el período de calentamiento de la cadena y la longitud del salto de la cadena para el cálculo de las estimaciones. Mientras que **b\_pri**, **B\_pri**, **g\_pri** y **G\_pri** corresponden a los parámetros de las distribuciones a priori de  $\beta$  y  $\gamma$ . Finalmente, **beta\_0**, **gamma\_0**, **rho\_0** y **lambda\_0** son los valores de inicio de las cadenas para la estimación de los parámetros. Dependiendo de la distribución escogida para el kernel de transición de  $\lambda$ , se puede escoger **kernel**=“uniform” o **kernel**=“normal”. El parámetro **mateq** está vinculado a si se asume  $\mathbf{W1} = \mathbf{W2}$ .

---



---

## Bibliografía

---



---

- [1] *Exploratory spatial data analysis*, [https://rpubs.com/corey\\_sparks/105700](https://rpubs.com/corey_sparks/105700), Accessed: 2016-09-18.
- [2] *Global spatial autocorrelation indices - moran's i, geary's c and the general cross-product statistic*, <http://www.lpc.uottawa.ca/publications/moransi/moran.htm>, Accessed: 2016-09-18.
- [3] *Spatial dependence models: estimation and testing*, [https://www.uam.es/personal\\_pdi/economicas/coro/docencia/doctorado/spateconUPC/Slides/Session7\\_SpatialDepModel\\_Notes.pdf](https://www.uam.es/personal_pdi/economicas/coro/docencia/doctorado/spateconUPC/Slides/Session7_SpatialDepModel_Notes.pdf), Accessed: 2016-09-18.
- [4] J. Aldstadt and A. Getis, *Using amoeba to create a spatial weights matrix and identify spatial clusters*, *Geographical Analysis* **38** (2006), 327–343.
- [5] L. Anselin, *Estimation methods for spatial autoregressive structures*, Regional Science Dissertation and Monograph **8** (1980).
- [6] L. Anselin, *A note on small sample properties of estimators in a first-order spatial autoregressive model*, *Environment and Planning A* **14** (1982), 1023–1030.
- [7] L. Anselin, *Some further notes on spatial models and regional science*, *Journal of Regional Science* **26** (1986), 799–802.
- [8] ———, *Spatial Econometrics: Methods and Models*, Kluwer Academic, Boston, 1988.
- [9] ———, *SpaceStat Tutorial: A Workbook for using SpaceStat in the Analysis of Spatial Data*, (1992).
- [10] ———, *Spatial econometrics*, A companion to theoretical econometrics (2001), 310–330.
- [11] ———, *Thirty years of spatial econometrics*, *Papers in Regional Science* **89** (2010), no. 1, 3–25.
- [12] L. Anselin and A. Bera, *Spatial dependence in linear regression models with an introduction to spatial econometrics*, pp. 237–289, Taylor & Francis, New York, 1998.
- [13] R. Assuncao and E. Krainski, *Neighborhood dependence in bayesian spatial models*, *Biometrical Journal* **51** (2009), no. 5, 851–869.
- [14] B. Baltagi, *A companion to theoretical econometrics*, 1 ed., Blackwell Companions to Contemporary Economics, Wiley-Blackwell, 2001.

- 
- [15] B. Baltagi and I. Kelejian, H. and Prucha, *Analysis of spatially dependent data*, Journal of Econometrics **140** (2007), 1–4.
- [16] P. Billingsley, *Probability and Measure. 3rd Edition*, Wiley, 1995.
- [17] R. Bivand, V. Gomez-Rubio, and H. Rue, *Approximate Bayesian inference for spatial econometrics models*, Spatial Statistics **9** (2014), 146–165.
- [18] R. Bivand, E. Pebesma, and V. Gomez-Rubio, *Applied spatial data analysis with r*, 1 ed., Springer New York, 2008.
- [19] L. Blanco, *Probabilidad*, Universidad Nacional de Colombia, Facultad de Ciencias, 2004.
- [20] R. Capello, *Spatial spillovers and regional growth: A cognitive approach*, European Planning Studies **17** (2009), 639–658.
- [21] G. Casella and E. George, *Explaining the Gibbs Sampler*, The American Statistician (1992), no. 3, 167–174.
- [22] E. Cepeda-Cuervo, *Modelagem da Variabilidade em Modelos Lineales Generalizados*, (Unpublished PhD thesis. Univerisidade Federal de Rio de Janeiro, Brazil) (2001).
- [23] E. Cepeda-Cuervo and D. Gamerman, *Bayesian Modeling of Variance Heterogeneity in Normal Regression Models*, Brazilian Journal of Probability and Statistics **14** (2001), 207–221.
- [24] ———, *Bayesian methodology for modeling parameters in the two parameters exponential family*, Revista Estadística **57** (2005), 93–105.
- [25] Y. Chen, *On the four types of weight functions for spatial contiguity matrix*, Letters in Spatial and Resource Sciences **5** (2012), 65–72.
- [26] A. Cliff and J. Ord, *Spatial Process. Models and Applications*, Ed. Pion, 1981.
- [27] Hinkley D. Davison, A., *Bootstrap methods and their application*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1997.
- [28] L. Devroye, *Non-uniform random variate generation. the random bit model*, pp. 768–783, 1986.
- [29] L. Duncan, *CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors*, Journal of Statistical Software **55** (2013), no. 13, 1–24.
- [30] B. Efron and R. Tibshirani, *An introduction to bootstrap*, 1 ed., Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Chapman & Hall, 1994.
- [31] P. Elhorst and S. Vega, *On spatial econometric models, spillover effects, and W*, ERSA conference papers (2013), no. 2, 1–28.
- [32] S. Fotheringham, *Trends in quantitative methods i: stressing the local*, Progress in Human Geography **21** (1997), 88–96.
- [33] D. Gamerman, *Markov Chains Monte Carlo: Stochastic Simulation for bayesian Inference*, Chapman and Hall, 1997.

- 
- [34] R.C. Geary, *The Contiguity Ratio and Statistical Mapping*, Journal of the Royal Statistical Society B **5** (1954), 115–145.
- [35] A. Getis, *Spatial filtering in a regression framework: Examples using data on urban crime, regional inequality, and government expenditures*, pp. 191–202, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [36] A. Getis and J. Aldstadt, *Constructing the spatial weights matrix using a local statistic*, Geographical Analysis **36** (2004), 90–104.
- [37] A. Getis and J. Ord, *The analysis of spatial association by use of distance statistics*, Geographical Analysis **24** (1992), no. 3, 189–206.
- [38] R. Haining, *Spatial models and regional science: A comment on anselin's paper and research directions*, Journal of Regional Science **26** (1986), 793–798.
- [39] L. Hepple, *Bayesian analysis of the linear model with spatial dependence*, pp. 179–199, Springer Netherlands, Dordrecht, 1979.
- [40] H. Kelejian and I. Prucha, *A generalized moments estimator for the autoregressive parameter in a spatial model*, International Economic Review **40** (1999), no. 2, 509–533.
- [41] Harry H. Kelejian and Ingmar R. Prucha, *A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances*, The Journal of Real Estate Finance and Economics **17** (1998), no. 1, 99–121.
- [42] J. Le Sage and K. Pace, *Introduction to Spatial Econometrics*, Chapman & Hall/CRC, Boca Raton, 2009.
- [43] J. LeSage, *Bayesian Estimation of Spatial Autoregressive Models*, Journal of Composite Materials **16** (1997), no. 4, 928–940.
- [44] J.P. LeSage, *A family of geographically weighted regression models*, pp. 241–264, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.
- [45] F. Lindgren and H. Rue, *Bayesian Spatial Modelling with R-INLA*, Journal of Statistical Software **63** (2015), no. 19.
- [46] T. Ma, T. Hong, and H. Zhang, *Tourism spatial spillover effects and urban economic growth*, Journal of Business Research **68** (2015), 74–80.
- [47] MATLAB, *version 7.10.0 (r2010a)*, The MathWorks Inc., Natick, Massachusetts, 2010.
- [48] P. Moran, *The interpretation of statistical maps*, Journal of the Royal Statistical Society Series B (Methodological) **10** (1948), 243–251.
- [49] J. Nash, *On Best Practice Optimization Methods in R*, Journal of Statistical Software **60** (2014), no. 2, 1–14.
- [50] A. Neath and J. Cavanaugh, *The bayesian information criterion: background, derivation, and applications*, Wiley Interdisciplinary Reviews: Computational Statistics **4** (2012), 0–203.

- 
- [51] K. Ord, *Estimation methods for models of spatial interaction*, Journal of the American Statistical Association **70** (1975).
- [52] M. Osborne, *Fisher's method of scoring*, International Statistical Review **60** (1992), 99–117.
- [53] J. Paelinck and L. Klaasen, *Spatial Econometrics*, Saxon House, 1979.
- [54] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [55] P. Resnik and E. Hardisty, *Gibbs Sampling for the Uninitiated*, Tech. report, University of Maryland, Department of Computer Science, 2010.
- [56] C. Robert, *The Metropolis-Hastings algorithm*, Tech. report, Universite Paris-Dauphine, 2015.
- [57] D. Rodriguez, *Modelar la Concentracion de la Tierra en Colombia Mediante Modelos Econometricos Espaciales*, Master's thesis, Universidad Nacional de Colombia, Mayo 2010.
- [58] H. Rue, S. Martino, and N. Chopin, *Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **71** (2009), 319–392.
- [59] C. Siddhartha and E. Greenberg, *Understanding the Metropolis-Hastings Algorithm*, The American Statistician **49** (1995), no. 4, 327–335.
- [60] Conley T.G., *Gmm estimation with cross sectional dependence*, Journal of Econometrics **92** (1999), no. 1, 1 – 45.
- [61] W. R. Tobler, *A Computer Movie Simulating Urban Growth in the Detroit Region*, Economic Geography **46** (1970), 234–270.
- [62] L. Waller and C. Gotway, *Applied spatial statistics for public health data*, 1 ed., Wiley Series in Probability and Statistics, John Wiley & Sons, 2004.
- [63] J. Wang, T. Zhang, and B. Fu, *A measure of spatial stratified heterogeneity*, Ecological Indicators **67** (2016).
- [64] N. Yu, M. de Jong, S. Storm, and J. Mi, *Spatial spillover effects of transport infrastructure: evidence from chinese regions*, Journal of Transport Geography **28** (2013), 56–66.