



UNIVERSIDAD NACIONAL DE COLOMBIA

# Estimación de un modelo SAR para datos panel con coeficientes espaciales específicos

**David Fernando Delgado Contreras**

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Estadística  
Bogotá, Colombia  
2019



# Estimación de un modelo SAR para datos panel con coeficientes espaciales específicos

**David Fernando Delgado Contreras**

Tesis presentada como requisito parcial para optar al título de:  
**Magister en Ciencias - Estadística**

Director:  
Ph.D. Oscar Orlando Melo Martínez

Línea de Investigación:  
Modelos Lineales para datos panel

Universidad Nacional de Colombia  
Facultad de Ciencias, Departamento de Estadística  
Bogotá, Colombia  
2019



## Dedicatoria

Este trabajo está completamente dedicado a mis padres, quienes siempre han estado ahí para soportar mis sueños y en creer en mis capacidades, los amo.

*“The only thing greater than the power of the mind is the courage of the heart”*

- John Nash



# Agradecimientos

Primero, agradezco la amplia e incondicional guía y ayuda del profesor Oscar Melo, quien fue fundamental para completar este trabajo y dotarlo de rigurosidad y sentido a través de sus comentarios precisos y valiosas enseñanzas, un método que comparte con sus alumnos desde el aula de clase.

A mi madre Sandra y mi padre Carlos Fernando, de quienes siempre he recibido apoyo y total confianza y han sido el motor de mis esfuerzos.

A mis amigos, profesores y compañeros que a través de esta experiencia han sido fuente de grandes lecciones, pero sobretodo a mi amiga Laura, quien siempre estuvo para darme ánimo e inspirar perseverancia.

A mi jefe Sandra, quien siempre valoró la importancia de este objetivo en mi vida, y a pesar de las circunstancias, nunca dejó de apoyarme y cubrirme cuando lo necesité; y a todos aquellos que de algún modo me encaminaron a no desfallecer, gracias.





## Resumen

Los modelos espaciales auto-regresivos permiten describir la dependencia espacial que surge cuando los valores que adopta una variable en una región o lugar están relacionados con las observaciones vecinas. Extensiones de estos modelos a estructuras de datos panel también han sido desarrolladas en la literatura espacial. El objetivo de este documento es presentar una propuesta que permita encontrar los estimadores de un modelo espacial auto-regresivo para datos panel con coeficientes espaciales específicos a través del método de máxima verosimilitud. La estrategia utilizada conlleva a obtener formas cerradas para la estimación de los parámetros asociados a las variables exógenas y a la varianza, mientras que resulta necesario el uso de métodos numéricos para calcular los coeficientes espaciales. Los resultados se aplican sobre un panel que tiene como objetivo explicar de forma lineal los componentes del IDH (Índice de Desarrollo Humano) dentro de una muestra intercontinental. Los experimentos dejaron notar que el cálculo de los coeficientes específicos resulta costoso computacionalmente, pero los resultados son significativamente diferentes a la especificación con un coeficiente espacial único y resulta en una mejor bondad de ajuste. En relación a la aplicación se resalta que mientras el modelo con único coeficiente espacial tiende a sobre-explicar el IDH por su componente de estándar de vida, el uso de esta propuesta atenúa la magnitud de tal parámetro asociado.

**Palabras clave:** Modelos espaciales auto-regresivos, datos panel, modelos lineales, desarrollo económico.

## Abstract

Spatial auto-regressive models allow to describe spatial dependence underlying when regional values are related to the neighbor observations. Extensions of this models to data panel structures have been developed in recent spatial literature, too. This paper has as objective to present a proposal that allows to find the estimators of a spatial auto-regressive model for panel data with specific spatial coefficients through the maximum likelihood methodology. Implemented strategy leads to obtain closed expressions for the parameter estimators associated to non-distance explanatory variables, and variance, while it is necessary the use numeric methodologies to compute the spatial coefficients. Results were applied over a panel which has as objective explaining in a linear way the HDI (Human Development Index) components within an intercontinental sample. Experiments let to see that estimation of specific coefficients is very computing expensive, but the results are statistically different respect

to the unique spatial coefficient specification and leads a better goodness-of-fit. Regarding to the application, we highlight the over-accounting in living standard component by the unique coefficient proposal, parameter attenuated using specific coefficients model.

**Keywords:** Spatial auto-regressive, panel data, linear models, development economics

# Contenido

<b>Agradecimientos</b>	<b>VII</b>
<b>Resumen</b>	<b>IX</b>
<b>1. Introducción</b>	<b>1</b>
<b>2. Evolución de los modelos espaciales para datos panel</b>	<b>4</b>
2.1. Especificación en modelos para datos panel . . . . .	5
2.2. Modelos espaciales auto-regresivos . . . . .	9
2.3. Modelos espaciales auto-regresivos para datos panel . . . . .	11
2.3.1. Heterogeneidad modelada mediante efectos fijos (EF) . . . . .	12
2.3.2. Heterogeneidad modelada mediante efectos aleatorios (EA) . . . . .	14
<b>3. Modelos SAR para datos panel con coeficientes espaciales específicos</b>	<b>16</b>
3.1. Forma matricial . . . . .	16
3.2. Estimación e inferencia por MV de los parámetros . . . . .	17
3.2.1. Estimadores y distribución asociada para $\hat{\beta}$ y $\hat{\sigma}^2$ . . . . .	19
3.2.2. Estimación e inferencia de $\mathbf{P}$ . . . . .	20
3.2.3. Construcción de intervalos de confianza para $\beta_j$ y $\rho_i$ . . . . .	22
3.2.4. Contraste de hipótesis para evaluar la necesidad de coeficientes es- pecíficos vs. coeficiente único . . . . .	23
<b>4. Un análisis al Índice de Desarrollo Humano y sus componentes descontando     efectos espaciales</b>	<b>25</b>
4.1. Descripción de los datos . . . . .	26
4.1.1. Matriz de pesos espaciales . . . . .	27
4.1.2. Prueba de Moran para validar correlación espacial . . . . .	28
4.2. Resultados . . . . .	29
4.2.1. Estimación bajo la metodología del SARP tradicional . . . . .	29
4.2.2. Estimación bajo la metodología del SARP con coeficientes espaciales específicos . . . . .	31

4.3. Consideraciones acerca de los resultados . . . . .	39
4.3.1. Tiempos computacionales y tamaño de la muestra en $N$ . . . . .	39
<b>5. Conclusiones y recomendaciones</b>	<b>41</b>
5.1. Conclusiones . . . . .	41
5.2. Recomendaciones . . . . .	42
<b>A. Anexo: Detalles matemáticos (estimadores SARP coeficientes específicos)</b>	<b>43</b>
A.1. Proceso Generador de Datos . . . . .	43
A.2. Función de Log-Verosimilitud . . . . .	43
A.3. Maximización de la verosimilitud: . . . . .	46
A.3.1. Insesgadez de $\hat{\beta}$ . . . . .	46
A.3.2. Estimador Máximo Verosímil de $\sigma^2$ . . . . .	47
A.3.3. Sesgo y corrección $\hat{\sigma}_{MV}^2$ . . . . .	47
A.3.4. Distribución de $\hat{\beta}$ . . . . .	48
<b>B. Anexo: Índice de Desarrollo Humano, definiciones y cálculos</b>	<b>52</b>
B.1. Índice de Desarrollo Humano . . . . .	52
B.2. Prueba de Moran de correlación espacial al IDH . . . . .	53
B.3. Resultados contraste de hipótesis coeficiente espacial único vs. específicos . .	53
<b>C. Anexo: Criterios de información y resultados pruebas a los residuales</b>	<b>56</b>
<b>D. Anexo: Matrices de pesos espaciales</b>	<b>61</b>
<b>E. Anexo: Código empleado en la aplicación</b>	<b>63</b>
<b>Bibliografía</b>	<b>80</b>

# Abreviaturas

Abreviatura	Término
MCO	Mínimos Cuadrados Ordinarios
MCG	Mínimos Cuadrados Generalizados
MV	Máxima Verosimilitud
PGD	Proceso Generador de Datos
SAR	Spatial Auto-Regressive
SARP	Spatial Auto-Regressive Panel
SEM	Spatial Error Model
FIML	Full Information Maximum Likelihood
MCOD	Mínimos Cuadrados Ordinarios con Dummies
EF	Efectos Fijos
EA	Efectos Aleatorios
IDH	Índice de Desarrollo Humano
IE	Índice Educativo
IEV	Índice de Esperanza de Vida
PIB	Producto Interno Bruto
IIB	Ingreso Interno Bruto

# 1. Introducción

Los modelos espaciales auto-regresivos (SAR por sus siglas en inglés), propuestos en Cliff & Ord (1973) han captado la atención de los investigadores en tanto permiten modelar la dependencia espacial que surge cuando los valores de una región o lugar están relacionados con las observaciones vecinas, reflejando tal dependencia a través de un parámetro parsimonioso auto-regresivo, asociado a la variable dependiente, dando solución al problema de sobre-especificación que surge al intentar especificar coeficientes exclusivos para cada unidad espacial en un conjunto de datos de corte transversal, un caso donde las observaciones no resultan suficientes para llevar a cabo la estimación. La estructura genérica de un SAR como el mencionado suele escribirse como:

$$y_i = \rho \sum_{j=1}^N w_{ij} y_j + \epsilon_i$$

donde  $i = 1, \dots, N$ ,  $w_{ij}$  corresponde al  $ij$ -ésimo elemento respectivo de la matriz de pesos espaciales relacionada, mientras que  $\rho$  representa el parámetro auto-regresivo espacial de primer orden,  $y_i$  es la observación  $i$ -ésima de la variable aleatoria de interés y  $N$  es la medida de la dimensión transversal del fenómeno a modelar.

Tal especificación ha sido extendida a la inclusión de variables explicativas adicionales (popularizado por Anselin (2013) como “*mixed-regressive, spatial autoregressive model*”):

$$y_i = \rho \sum_{j=1}^N w_{ij} y_j + \sum_{k=1}^K x_{ik} \beta_k + \epsilon_i$$

para las  $K$  variables explicativas  $x_k$  y sus respectivos parámetros  $\beta_k$ , con  $k = 1, \dots, K$ .

El objetivo de este tipo de modelos es encontrar parámetros corregidos por la dependencia espacial que pueda existir entre las unidades analizadas, y que no llega a ser explicada por las variables en  $\sum_{k=1}^K x_{ik} \beta_k$ , además de calcular la magnitud de tal dependencia.

En línea con lo anterior, el reciente auge de bases de datos de corte transversal observadas en sucesivos periodos de tiempo (conocidas como datos panel), también ha recibido atención en el campo de los modelos espaciales, ya que ofrecen más grados de libertad y mayor variabilidad a través de la dimensión transversal. De este modo, extensiones equivalentes en

datos panel para modelos SAR y SEM (modelo de error espacial, por sus siglas en inglés) en su forma más parsimoniosa han sido descritas y caracterizadas anteriormente como puede encontrarse en Elhorst (2003). Por ejemplo, un modelo SAR para datos panel (SARP) adopta una estructura como la que se muestra a continuación:

$$\mathbf{Y} = \rho(\mathbf{I}_T \otimes \mathbf{W}_N)\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

donde  $\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{1N}, Y_{21}, \dots, Y_{TN})^t$  es un vector aleatorio de dimensión  $NT \times 1$  que contiene las variables aleatorias respuesta,  $\mathbf{W}$  es una matriz de dimensiones  $N \times N$  de pesos espaciales, y  $\mathbf{X}$  es una matriz  $NT \times (K+1)$  que contiene las observaciones de las  $K$  variables explicativas diferentes a la distancia, y un intercepto. Mientras tanto,  $\rho$  sigue representando un único valor real y  $\boldsymbol{\beta}$  es el vector asociado a los efectos marginales correspondientes a las  $K$  covariables, y el intercepto. Finalmente  $\mathbf{I}_T$  es una matriz identidad de dimensión  $T \times T$ , acorde a la ventana de tiempo observada.

A partir de esos primeros avances, la literatura relacionada con modelos espaciales para datos panel se ha expandido ampliamente con nuevas propuestas que más que modificar, complementan las especificaciones base. Como ejemplos importantes cabe resaltar a Baltagi, Song & Koh (2003) donde se considera una prueba para evaluar la dependencia espacial sobre datos panel en el componente de error, cuando este se caracteriza de forma auto-regresiva, en Kapoor, Kelejian & Prucha (2007) puede encontrarse un adelanto teórico para la implementación de modelos espaciales en datos panel flexibles a la especificación de efectos aleatorios provenientes de las unidades espaciales. Lee & Yu (2010a) proponen modelos espaciales para datos panel con efectos fijos, y errores auto-regresivos, trabajo que va en línea con Mutl & Pfaffermayr (2008), quienes realizaron los desarrollos necesarios para la estimación de modelos de datos panel con rezagos espaciales bajo propuestas tanto de efectos fijos como aleatorios, además de proponer una prueba de Hausman para encontrar la especificación adecuada del modelo.

Ahora bien, es importante destacar que en los documentos mencionados anteriormente se han encontrado una amplia gama de aplicaciones, como por ejemplo: economía de la agricultura (Druska & Horrace 2004), transporte (Frazier & Kockelman 2005), políticas económicas públicas (Egger, Pfaffermayr & Winner 2005) y demanda de bienes (Baltagi & Li 2006), entre otros.

No obstante, aunque existen avances en los cuales se incorpora la estimación de vectores con efectos (tanto fijos como aleatorios, ver Baltagi, Egger & Pfaffermayr (2013) y Kapoor et al. (2007)) propios de las unidades espaciales sobre la variable dependiente, aún no existen en la literatura propuestas que permitan modelar efectos auto-regresivos espaciales de cada unidad sobre las respuestas.

En vista de tal vacío, esta investigación tiene como objetivo el planteamiento y desarrollo del proceso de estimación de los parámetros para una forma funcional que ofrezca la inclusión de los efectos específicos de las unidades espaciales estudiadas sobre las demás, así como la descripción de tal proceso de estimación e inferencia.

En este orden de ideas, la organización del documento está dada por el orden descrito a continuación: en el Capítulo 2, se encuentra una revisión literaria relacionada con una breve descripción de los modelos para datos panel, el proceso de especificación y estimación de modelos SAR y los avances más relevantes en torno a modelos espaciales para datos panel en relación a esta investigación. Posteriormente, en el Capítulo 3, se presenta el desarrollo del proceso de estimación e inferencia para modelos SAR para datos panel con coeficientes espaciales específicos, acompañado (Capítulo 4) de una aplicación donde se modela el Índice de Desarrollo Humano como función de sus componentes (llevados a una forma lineal) y la segregación del elemento espacial. Finalmente, el Capítulo 5 termina con algunas conclusiones y recomendaciones respecto a los resultados encontrados.



## 2. Evolución de los modelos espaciales para datos panel

Esta sección tiene como objetivo brindar al lector una contextualización general respecto a los modelos de datos panel, la especificación y estimación de modelos SAR, y la evolución de la literatura relacionada con la convergencia existente entre estas dos áreas de la estadística. En línea con lo anterior, es importante resaltar que mientras por un lado la especificación y estimación de modelos espaciales cobija una porción importante de la atención de los investigadores, debido a la gran cantidad de fenómenos que pueden envolver en alguna medida dependencia espacial entre las unidades analizadas; por otro lado, el hecho que los datos panel ofrezcan conjuntos de observaciones usualmente más informativos que los datos unidimensionales como los de corte transversal o las series de tiempo (gracias a que contienen mayor variabilidad y menos colinealidad entre las variables), hace de los modelos espaciales para datos panel un área de la estadística de amplio y creciente interés.

De esta forma, la necesidad de ampliar los límites de la literatura para esta clase de modelos surge del problema de variable omitida que se presenta cuando existe alguna dependencia espacial entre las unidades observadas que no se especifica o aísla del resto de parámetros estimados, y termina por quedar inmersa en el componente de error, lo que usualmente trae como consecuencia sesgo en los demás parámetros estimados y construcción equívoca de los intervalos de confianza (Elhorst 2003).

Sumado a lo anterior, está la potencial heterogeneidad que deriva de las diferentes magnitudes y sentidos en que las relaciones espaciales se presenten entre las unidades de estudio. Los primeros adelantos respecto a la estimación de efectos (heterogeneidad en parámetros) se trata en trabajos como Pesaran & Smith (1995).

Tal heterogeneidad es una característica propia de los conjuntos de datos tipo panel, y que esencialmente genera un atractivo entre los investigadores al momento de decidir si trabajar con la mera serie de tiempo de una sola unidad longitudinal, o  $n$ -plicar el esfuerzo construyendo un panel con  $n$  individuos, el cual permitiría aislar los efectos relativos a cada unidad más allá de las relaciones que puedan tener con otras variables.

Por todo lo anterior, es que la combinación o convergencia de los modelos espaciales con las

estructuras de datos panel resulta en una herramienta estadística que permite el modelamiento en simultáneo de la mayor variabilidad y menor colinealidad propia de los conjuntos de observaciones bidimensionales con las relaciones espaciales presentes entre las unidades longitudinales.

## 2.1. Especificación en modelos para datos panel

En términos generales, los datos de tipo panel son agrupaciones de información que organizan las observaciones de diferentes variables para diferentes unidades (corte transversal) a través del tiempo (ver la estructura presentada en la tabla 2-1). Tal bidimensionalidad en la información suele ser aprovechada mediante especificaciones que agrupen mediante parámetros únicos los efectos marginales de cada variable para todos los individuos a través de todos los periodos analizados; un ejemplo clásico de tal especificación para una respuesta  $y_{it}$  es:

$$y_{it} = \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_k x_{k,it} + e_{it} \quad (2-1)$$

donde las unidades transversales están indexadas por  $i = 1, \dots, N$ , la dimensión temporal por  $t = 1, \dots, T$ ,  $e_{it} \sim i.i.d.$  (esta notación implica distribución idéntica e independiente a través de todas las variables aleatorias inmersas), y cada  $\beta_j$  es un parámetro que agrupa los efectos transversales y temporales conjuntos de las variables explicativas  $x_k$  sobre  $y_{it}$ .

A pesar que la estructura presentada en la tabla 2-1 suele ser la de uso más común en la mayor parte de aplicaciones para datos panel, para estimaciones relacionadas con modelos espaciales la organización de los datos necesariamente suele estar organizada a través de la dimensión temporal. Las tablas 2-2 y 2-3 presentan un ejemplo de un panel con 4 periodos, 3 individuos, la variable dependiente y dos explicativas que permiten ilustrar tal organización. No obstante, aunque existe la posibilidad de encontrar efectos longitudinales sobre la respuesta, la especificación anterior (ecuación 2-1) no contempla la posibilidad que cada unidad transversal  $i$  tenga un efecto  $c_i$  no homogéneo sobre la respuesta  $y_{it}$ , es decir, un modelo como el que se muestra en la ecuación (2-2).

$$y_{it} = c_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_k x_{k,it} + e_{it} \quad (2-2)$$

donde  $c_i$  suele ser conocido como componente no observado, variable latente o heterogeneidad no observada, (Wooldridge 2010). En este sentido, uno de los primeros aspectos a revisar en estos modelos es sí tales parámetros de heterogeneidad deben ser incluidos, o no. Con el fin de encontrar un criterio de decisión para tal aspecto, pueden usarse dos pruebas que

**Tabla 2-1.:** Estructura típica de una base de datos tipo panel. En esta notación,  $i$  corresponde a la unidad transversal,  $t$  al periodo de observación, y  $N$  y  $T$  a las dimensiones transversales y temporales, respectivamente.

$i$	$t$	$y$	$x_1$	$\dots$	$x_K$
$i_1$	1	$y_{11}$	$x_{1,11}$	$\dots$	$x_{K,11}$
$i_1$	2	$y_{12}$	$x_{1,12}$	$\dots$	$x_{K,12}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$i_1$	$T$	$y_{1T}$	$x_{1,1T}$	$\dots$	$x_{K,1T}$
$i_2$	1	$y_{21}$	$x_{1,21}$	$\dots$	$x_{K,21}$
$i_2$	2	$y_{22}$	$x_{1,22}$	$\dots$	$x_{K,22}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$i_2$	$T$	$y_{2T}$	$x_{1,2T}$	$\dots$	$x_{K,2T}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$i_N$	1	$y_{N1}$	$x_{1,N1}$	$\dots$	$x_{K,N1}$
$i_N$	2	$y_{N2}$	$x_{1,N2}$	$\dots$	$x_{K,N2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$i_N$	$T$	$y_{NT}$	$x_{1,NT}$	$\dots$	$x_{K,NT}$

**Tabla 2-2.:** Panel organizado por dimensión transversal

ID	Periodo	Dependiente	Explicativa 1	Explicativa 2
1	2001	$y_{1,2001}$	$x_{1,1,2001}$	$x_{2,1,2001}$
1	2002	$y_{1,2002}$	$x_{1,1,2002}$	$x_{2,1,2002}$
1	2003	$y_{1,2003}$	$x_{1,1,2003}$	$x_{2,1,2003}$
1	2004	$y_{1,2004}$	$x_{1,1,2004}$	$x_{2,1,2004}$
2	2001	$y_{2,2001}$	$x_{1,2,2001}$	$x_{2,2,2001}$
2	2002	$y_{2,2002}$	$x_{1,2,2002}$	$x_{2,2,2002}$
2	2003	$y_{2,2003}$	$x_{1,2,2003}$	$x_{2,2,2003}$
2	2004	$y_{2,2004}$	$x_{1,2,2004}$	$x_{2,2,2004}$
3	2001	$y_{3,2001}$	$x_{1,3,2001}$	$x_{2,3,2001}$
3	2002	$y_{3,2002}$	$x_{1,3,2002}$	$x_{2,3,2002}$
3	2003	$y_{3,2003}$	$x_{1,3,2003}$	$x_{2,3,2003}$
3	2004	$y_{3,2004}$	$x_{1,3,2004}$	$x_{2,3,2004}$

**Tabla 2-3.:** Panel organizado por dimensión temporal

ID	Periodo	Dependiente	Explicativa 1	Explicativa 2
1	2001	$y_{1,2001}$	$x_{1,1,2001}$	$x_{2,1,2001}$
2	2001	$y_{2,2001}$	$x_{1,2,2001}$	$x_{2,2,2001}$
3	2001	$y_{3,2001}$	$x_{1,3,2001}$	$x_{2,3,2001}$
1	2002	$y_{1,2002}$	$x_{1,1,2002}$	$x_{2,1,2002}$
2	2002	$y_{2,2002}$	$x_{1,2,2002}$	$x_{2,2,2002}$
3	2002	$y_{3,2002}$	$x_{1,3,2002}$	$x_{2,3,2002}$
1	2003	$y_{1,2003}$	$x_{1,1,2003}$	$x_{2,1,2003}$
2	2003	$y_{2,2003}$	$x_{1,2,2003}$	$x_{2,2,2003}$
3	2003	$y_{3,2003}$	$x_{1,3,2003}$	$x_{2,3,2003}$
1	2004	$y_{1,2004}$	$x_{1,1,2004}$	$x_{2,1,2004}$
2	2004	$y_{2,2004}$	$x_{1,2,2004}$	$x_{2,2,2004}$
3	2004	$y_{3,2004}$	$x_{1,3,2004}$	$x_{2,3,2004}$

ayudan a decidir acerca de esta inclusión de efectos longitudinales en los modelos (ver Greene (2000)). Por un lado está la prueba  $F$  que evalúa la hipótesis nula donde el componente no observado es constante a través de los individuos ( $H_0 : c_i = c$ ), así, el estimador agregado sería eficiente, al no sobre-especificar el modelo:

$$F = \frac{(R_{MCO}^2 - R_{Agrupado}^2)/(N - 1)}{(1 - R_{MCO}^2)/(NT - N - k)} \sim F_{(N-1, NT-N-k)}$$

donde se entienden los  $R^2$  como las medidas respectivas de bondad de ajuste para el modelo *MCO* (Mínimos Cuadrados Ordinarios con dummies), que se refiere a la ecuación (2-2) estimada mediante MCO con dummies, donde  $c_i$  varía en  $i = 1, \dots, N$  para las  $N - 1$  dummies y el intercepto; y el modelo *Agrupado* a la función planteada en (2-1) con una constante única  $c$  adicional.

Por el otro lado, la prueba de Breusch & Pagan (1980), está basada en un test de multiplicador tipo Lagrange en el que se evalúa si los  $c_i$  de (2-2) son aleatorios y por ende, existe variabilidad en ellos, de forma que:

$$H_0 : \sigma_c^2 = 0 \quad \text{vs} \quad H_1 : \sigma_c^2 > 0$$

donde  $\sigma_c^2$  es la varianza del parámetro que caracteriza la heterogeneidad no observable.

Ahora bien, si las pruebas arrojan evidencia significativa que soporte la existencia de efectos en los modelos, el problema se convierte en determinar si  $c_i$  debería ser tratado como un parámetro o como una variable aleatoria. En el primer caso, la especificación suele llamarse modelo de efectos fijos, mientras el segundo, modelo de efectos aleatorios (para profundizar en estos aspectos remitirse a Baltagi (2001)).

Usualmente, la prueba de Hausman (1978), que evalúa ortogonalidad entre el modelo de efectos aleatorios y los regresores permite realizar selección entre efectos fijos y aleatorios. La disyuntiva consiste en que al estimar por efectos fijos, cada dummy (parámetro individual) consume un grado de libertad, y en caso que los efectos no tengan correlación con los regresores, se llega a un estimador ineficiente en contraste con el de MCG (Mínimos Cuadrados Generalizados) que se usa en la estimación por efectos aleatorios. Sin embargo, cuando existe correlación entre los regresores y los efectos (endogeneidad) el estimador de MCG para efectos fijos resulta ser inconsistente, (ver Kmenta & Rafailzadeh (1997)).

Aunque existen muchos otros tipos de avances relacionados con el modelamiento de estructuras de datos panel, como propuestas para dimensiones con  $T$  grande (Maddala & Wu (1999) y Kao (1999)), Modelos Lineales Generalizados para datos panel (Chintagunta, Jain & Vilcassim (1991) y Maddala (1987)), panel dinámico (Anselin & Kelejian (1997) y Anselin (2013)), no se ahondará en tales adelantos ya que no son relevantes para los propósitos

específicos de este trabajo.

## 2.2. Modelos espaciales auto-regresivos

Por su lado, el estudio de fenómenos sujetos a dependencia espacial (una situación donde los valores observados en una región o lugar dependen de los valores de las observaciones vecinas, LeSage & Pace (2009)) dio paso a una buena cantidad de modelos que intentan ofrecer alternativas de especificación para tales fenómenos.

En este sentido, una propuesta donde todas las relaciones entre las unidades espaciales pudiesen ser explicadas por separado sería la más informativa, y a la vez la más compleja de estimar dada la cantidad de información con la que usualmente se cuenta. Por ejemplo, cuando se supone una situación donde existan tres unidades espaciales, las cuales tienen efectos significativamente diferentes a cero, cualquiera sobre todas las demás, además de otras variables explicativas adicionales, tal fenómeno podría describirse mediante el siguiente sistema de ecuaciones:

$$\begin{aligned} y_i &= \alpha_{i,j}y_j + \alpha_{i,k}y_k + \mathbf{x}_i^t\boldsymbol{\beta} + \epsilon_i \\ y_j &= \alpha_{j,i}y_i + \alpha_{j,k}y_k + \mathbf{x}_j^t\boldsymbol{\beta} + \epsilon_j \\ y_k &= \alpha_{k,i}y_i + \alpha_{k,j}y_j + \mathbf{x}_k^t\boldsymbol{\beta} + \epsilon_k \end{aligned} \quad (2-3)$$

donde  $\epsilon_s \sim i.i.d.$  para los  $s = i, j, k$  y  $\mathbf{x}_s$  incluye los  $K_s$  regresores no espaciales correspondientes a cada unidad transversal, i.e.,  $\mathbf{x}_{s_i}^t = (x_{1,i}, x_{2,i}, \dots, x_{K_s,i})$ .

El sistema (2-3) resulta muy ilustrativo, ya que deja ver que aún cuando esta es una especificación muy atractiva para un investigador (por la cantidad de relaciones que permite explicar), también se muestra poco útil en la práctica, puesto que incluso en caso que  $\boldsymbol{\beta}$  no contuviera parámetros significativos, la estimación de los  $\alpha$  (coeficientes auto-regresivos) se vuelve imposible en tanto se tendrían más parámetros que observaciones.

Bajo tal restricción, la solución más evidente es agrupar los efectos espaciales, de modo que se sacrifique la especificidad en pro de dar paso a alternativas más parsimoniosas que solucionen el problema de sobre-parametrización. En los trabajos de Cliff & Ord (1973) y Ord (1975) se propuso una primera aproximación donde un solo parámetro auto-regresivo agrupa la dependencia espacial existente entre todas las unidades:

$$y_i = \rho \sum_{j=1}^N w_{ij}y_j + \epsilon_i \quad (2-4)$$

donde  $w_{ij}$  es un elemento de la matriz de pesos espaciales  $\mathbf{W}$  que representa el efecto o ponderación espacial de la unidad  $j$  sobre  $i$ . Así,

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & & w_{2N} \\ \vdots & & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{bmatrix}$$

Con base en lo anterior, el término  $\sum_{j=1}^N w_{ij}y_j$  suele llamarse *rezago espacial* en tanto que resulta ser una combinación lineal de los valores asociados a la unidad  $j$  con los pesos correspondientes de la matriz  $\mathbf{W}$ . Tal estructura del *rezago espacial*, permite la estimación de un único  $\rho$  que intuitivamente agrega el efecto espacial del sistema objeto de estudio. Finalmente, se supone que  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ , lo que resulta muy útil cuando se opta por el uso del método de máxima verosimilitud.

Cuando la estructura anterior (2-4) es combinada con un modelo de regresión convencional se obtiene lo que se conoce desde Anselin (2013) como “*mixed-regressive, spatial-autoregressive model*”:

$$\mathbf{Y} = \rho \mathbf{WY} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2-5)$$

donde (2-5) tiene una forma idéntica a (2-4) al adicionarse  $\mathbf{X}$  que contiene la matriz de variables explicativas adicionales ( $\mathbf{X}$  de dimensión  $NT \times K$ ), y  $\boldsymbol{\beta}$ , el vector de coeficientes asociado (de dimensión  $K \times 1$ ). Respecto a los errores se asume independencia entre los mismos y homocedasticidad, de forma que  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ .

En cuanto a la estimación, el uso de MCO resulta en parámetros inconsistentes tanto del  $\rho$ , como de los errores estándar, mientras que en contraste, la estimación por MV (Máxima Verosimilitud) suele ser una alternativa adecuada (Lee 2004).

De esta manera, la función de log-verosimilitud asociada al PGD (proceso generador de datos) de (2-5) puede escribirse como:

$$\ln(\mathcal{L}) = -(N/2) \ln(\pi\sigma^2) + \ln |\mathbf{I} - \rho \mathbf{W}| - \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{2\sigma^2} \quad (2-6)$$

donde  $\rho \in (\min(\phi)^{-1}, \max(\phi)^{-1})$ ,  $\phi$  es un vector  $N \times 1$  con los valores propios de  $\mathbf{W}$  (Barry & Pace 1999).

Con lo anterior en cuenta, es claro que el método de MV se concentra en la estimación de  $\rho$ ,  $\boldsymbol{\beta}$  y  $\sigma^2$ . Sin embargo, no es trivial encontrar soluciones cerradas para todos los parámetros. En respuesta a esta situación, Pace & Barry (1997) sugieren una aproximación conveniente

que consiste en enfocarse en obtener expresiones para  $\hat{\beta}$  y  $\hat{\sigma}^2$ , mientras que se busca una función donde se maximice  $\rho$  mediante el uso de algún método numérico. Así, se llega a:

$$\begin{aligned}\hat{\beta}_O &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{Y} \\ \hat{\beta}_d &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{W} \mathbf{X}^t \mathbf{Y} \\ \epsilon_O &= \mathbf{Y} - \mathbf{X} \hat{\beta}_O \\ \epsilon_d &= \mathbf{W} \mathbf{Y} - \mathbf{X} \hat{\beta}_d \\ \hat{\sigma}^2 &= N^{-1} \epsilon^t \epsilon\end{aligned}$$

donde  $\hat{\beta}_O$  resulta en el componente “ordinario” del estimador agregado  $\hat{\beta}$ , mientras que  $\hat{\beta}_d$ , es el componente que corrige la dependencia espacial en los parámetros, y los errores de tales componentes resultan ser  $\epsilon_O$  y  $\epsilon_d$ , respectivamente.

Así, la propuesta de Pace & Barry (1997) está centrada en minimizar mediante algún método numérico  $\epsilon = \epsilon_O - \rho \epsilon_d$  para diferentes valores de  $\rho$ .

Aunque la literatura relacionada con modelos espaciales es amplia, y existe una buena cantidad de modelos que complementan a los SAR, como los SEM (*Spatial Error Model*) popularizados en Anselin (2001), donde el componente espacial es modelado en el segmento aleatorio del error, los SARMA espaciales (*Spatial Autoregressive Moving Average*, ver Lee & Yu (2010b)) que incluyen un componente auto-regresivo espacial y un control espacial en el error, entre otros tipos de modelos con múltiples matrices de pesos espaciales o interacciones espaciales con las variables explicativas (una buena guía bien actualizada de los mismos puede encontrarse en Elhorst (2003)), tales tópicos no se profundizarán en este documento, puesto no tienen mayor relación con la metodología y propuesta objetivo de esta investigación. No obstante, podrían ser de interés para desarrollos posteriores que también incluyan la estimación de coeficientes específicos.

## 2.3. Modelos espaciales auto-regresivos para datos panel

Anteriormente se hizo una revisión de los principales y más generales tópicos relativos a los modelos para datos panel y a los modelos espaciales auto-regresivos. Ahora bien, esto permite pasar a revisar los aspectos más relevantes (respecto al enfoque de este trabajo) de los modelos espaciales auto-regresivos para datos panel.

La primera aproximación de los modelos espaciales a datos panel fue una sencilla extensión de (2-5) a los  $T$  periodos de tiempo en que se observan las  $N$  unidades espaciales:



$$\mathbf{Y} = \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2-7)$$

La forma de (2-7) es una especificación ampliamente usada en aplicaciones de interacción social-espacial de las regiones de estudio a través del tiempo (Brueckner 2006).

Sin embargo, aunque (2-7) resulta parsimoniosa y conveniente en torno a la forma de estimación (puesto que en términos generales sigue la misma lógica que en su forma longitudinal), se sujeta a varios supuestos bastante restrictivos, como que la estructura y relación espacial no cambia a través del tiempo, de ahí que tanto  $\rho$  como  $\mathbf{W}$  no sean función de  $t$ .

Por otro lado, a pesar que ya no se tiene solamente una observación por cada unidad transversal, (2-7) no muestra alguna diferenciación entre los posibles efectos específicos que pueda tener cada  $i$  sobre la variable respuesta, ni en forma de intercepto, ni a través de la dependencia espacial, lo que en suma con los supuestos anteriores se conoce como “*homogeneidad extrema*” (llamado así en Anselin, Le Gallo & Jayet (2008)).

Ahora bien, a la hora de proponer especificaciones que incluyan algún tipo de heterogeneidad espacial y/o temporal, podría llegarse a un gran número de combinaciones (incluso algunas que podrían resultar no estimables por problemas de identificación, e.g. una en que se buscase estimar parámetros específicos para los efectos longitudinales y temporales para cada unidad transversal y temporal, respectivamente, lo que implicaría una situación similar a la expuesta en (2-3).

Por lo anterior, a continuación se mencionan solamente aquellos avances relacionados con especificaciones que modelen heterogeneidad en el componente longitudinal del panel, dejando de lado las alternativas para la estimación del panel con efectos temporales, que si bien no dejan de ser relevantes, no guardan mayor relación con la metodología y aporte de este documento.

### 2.3.1. Heterogeneidad modelada mediante efectos fijos (EF)

Los modelos tradicionales (sin componentes espaciales) incluían parámetros específicos para cada unidad longitudinal que capturaban la heterogeneidad no observada, lo que se modelaba como el  $c_i$  de (2-2) para los  $i = 1, \dots, N$  (ver Baltagi (2001)).

Para la extensión al componente espacial auto-regresivo, tales parámetros se podrían estimar dentro de un modelo de la forma:

$$\mathbf{Y} = \rho(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + (\mathbf{1}_T \otimes \boldsymbol{\alpha}) + \boldsymbol{\epsilon} \quad (2-8)$$

donde  $\boldsymbol{\alpha}$  es un vector  $N \times 1$  que contiene los parámetros individuales de los efectos fijos.

En este caso,  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$ . Así,  $\boldsymbol{\alpha}$  captura la heterogeneidad no observada de cada unidad transversal, aprovechando mejor la información del panel (en relación con 2-7) y “limpiando” a  $\boldsymbol{\epsilon}$  de un ruido adicional.

No obstante, el problema que puede verse en este caso, tiene que ver con la consistencia de  $\boldsymbol{\alpha}$ , puesto que la estimación por efectos fijos no suele ser posible cuando  $N \rightarrow \infty$  (ver Arellano (2003)), lo que resulta ser una complicación debido a que los modelos espaciales recaen asintóticamente en la dimensión transversal para dotar de consistencia y normalidad asintótica a los estimadores.

A pesar de lo anterior, en Elhorst (2003) se argumenta que como muchas veces el interés recae en obtener estimaciones consistentes de  $\boldsymbol{\beta}$ , la estimación por MV estándar resulta apropiada. Además, en Lee (2004) también se presenta una sección dedicada a coeficientes fijos específicos para modelos SAR en datos panel donde existe no solamente un intercepto por unidad espacial, sino también un parámetro exclusivo por variable explicativa y unidad longitudinal:

$$\begin{bmatrix} 1 & -\delta_{21} & \cdots & -\delta_{N1} \\ -\delta_{12} & 1 & \cdots & -\delta_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ -\delta_{1N} & -\delta_{2N} & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (2-9)$$

De manera equivalente, (2-9) puede escribirse como  $\mathbf{\Gamma Y} = \mathbf{XB} + \boldsymbol{\epsilon}$  en donde los  $\delta_{ij}$  especifican los parámetros auto-regresivos, y los  $\beta_i$  los parámetros exclusivos para la dimensión longitudinal, no sobra aclarar que  $\mathbf{X}$  es la matriz que agrupa las variables explicativas, en  $\mathbf{\Gamma}$  se encuentran los parámetros auto-regresivos y  $\mathbf{B}$  es un vector con los  $\beta_i$ . En forma uniecuacional se llega a una estructura como a continuación:

$$Y_i = \sum_{j \neq i} \delta_{ji} Y_j + X_i \beta_i + \varepsilon_i$$

Por otro lado, se supone que  $E(\varepsilon_i \varepsilon_{i'}) = 0$  y que  $E(\varepsilon_i^2) = \sigma_i^2$ . Para la estimación de este tipo de modelos, puede usarse una función de log-verosimilitud, como la provista por Hausman (1975) que permite la maximización de las condiciones de primer orden para sistemas de ecuaciones simultáneos (FIML, Máxima Verosimilitud con Información Completa, por sus siglas en inglés).

El estimador  $\boldsymbol{\eta}_i$  por FIML resulta en una expresión:

$$\boldsymbol{\eta}_i = \begin{bmatrix} \hat{\delta}_i \\ \hat{\beta}_i \end{bmatrix} = (\hat{\mathbf{Z}}' \mathbf{Z}_i)^{-1} \hat{\mathbf{Z}}' \mathbf{Z}_i \quad (2-10)$$

donde  $\mathbf{Z}_i$  y  $\boldsymbol{\eta}_i$  resultan de la vectorización de  $\boldsymbol{\Gamma}$  y  $\mathbf{B}$ , respectivamente, y  $\hat{\boldsymbol{\delta}}_i$  es  $[\delta_{i1}, \dots, \delta_{i-1}, \delta_{i+1}, \dots, \delta_{iN}]^t$ . De forma que cada ecuación puede verse como:

$$Y_i = [Y_1 \quad \dots \quad Y_{i-1} \quad Y_{i+1} \quad \dots \quad Y_N \quad X_i] \times \begin{bmatrix} \delta_{i1} \\ \vdots \\ \delta_{i-1} \\ \delta_{i+1} \\ \vdots \\ \delta_{iN} \\ \beta_i \end{bmatrix} + \varepsilon_i \equiv \mathbf{Z}_i \boldsymbol{\eta}_i + \varepsilon_i$$

Y con  $\hat{\mathbf{Z}} = \boldsymbol{\Gamma}^{-1} \mathbf{X} \mathbf{B}$ . Por su lado, el estimador de  $\sigma_{ii}$  necesario para estimar la matriz de covarianza resulta estar dado por:

$$\hat{\sigma}_{ii} = \frac{(Y_i - \mathbf{Z}_i \boldsymbol{\eta}_i)^t (Y_i - \mathbf{Z}_i \boldsymbol{\eta}_i)}{T}$$

Es importante resaltar que  $\hat{\mathbf{Z}}$  sigue siendo función de  $\hat{\boldsymbol{\eta}}_i$ , y no hay forma cerrada para despejar el sistema de ecuaciones, por lo que es necesario el uso de costosos métodos numéricos para lograr la estimación de los  $\delta_i$ .

### 2.3.2. Heterogeneidad modelada mediante efectos aleatorios (EA)

Cuando se modela heterogeneidad espacial como variable aleatoria, se sigue una lógica parecida a la mencionada en la sección 2.1, que consiste en tratar el  $c_i$  de (2-2) como una variable aleatoria, un caso donde se debe asegurar que tal  $c_i$  no guarde correlación con ningún otro de los regresores. No obstante, como se describe en Lee & Yu (2010b), en el caso de los EA para panel espacial, la estructura de la especificación del efecto aleatorio no se limita a un cambio independiente sobre  $y_{it}$ , sino que los efectos aleatorios resultan en un SAR adicional que no varía en el tiempo, y que incluso puede incluir una matriz de pesos espaciales exclusiva, es decir:

$$\mathbf{Y} = \mathbf{1}_T \otimes (\mathbf{1}_N^t \mathbf{b}_0 + \mathbf{z}^t \boldsymbol{\eta}_0) + \rho_1 (\mathbf{I}_T \otimes \mathbf{W}_1) \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{1}_T \otimes \mathbf{C}_N^{-1} \mathbf{c} + [\mathbf{I}_{NT} - \rho_2 (\mathbf{I}_T \otimes \mathbf{W}_2)]^{-1} \boldsymbol{\varepsilon} \quad (2-11)$$

Con  $\mathbf{u} = \rho_3 \mathbf{W}_3 \mathbf{u} + \mathbf{c}$  que resulta ser el SAR (invariante en el tiempo) para modelar el efecto aleatorio espacial dimensionado al panel mediante  $\mathbf{C}_N = \mathbf{I}_N - \rho_3 \mathbf{W}_3$ , y un término de error que también puede o no contener una estructura espacial auto-regresiva (para el segundo caso, la estimación simplemente se sigue de  $\boldsymbol{\epsilon} = \boldsymbol{\varepsilon}$ ), siendo  $\boldsymbol{\epsilon} = \rho_2 (\mathbf{I}_T \otimes \mathbf{W}_2) \boldsymbol{\epsilon} + \boldsymbol{\varepsilon}$ .

Además en (2-11),  $\mathbf{b}_0$  es el vector de coeficientes constantes en espacio y tiempo,  $\boldsymbol{\eta}_0$  es el vector de parámetros asociado a los regresores que no varían en el tiempo  $\mathbf{z}$ ,  $\mathbf{c} \sim N(\mathbf{0}, \sigma_c^2 \mathbf{I}_N)$  y  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_N)$  con  $E(\mathbf{c}, \mathbf{u}) = 0$ , y finalmente  $\mathbf{1}_M$  es un vector de unos de tamaño  $M$ . La estructura de (2-11) puede estimarse por MV mediante la función de log-verosimilitud planteada en Baltagi et al. (2013) o mediante MCG como propone Kapoor et al. (2007). Se puede notar que el uso de una estimación adecuada de la especificación de (2-11) conlleva a varias ventajas y limitantes:

- Puede modelarse y filtrarse del error (que finalmente viene siendo  $\boldsymbol{\varepsilon}$ ) la correlación espacial que llegue a existir entre las unidades espaciales.
- Incluso tal correlación puede especificarse con varias matrices de pesos espaciales según convenga, y arrojar múltiples  $\rho$  que agreguen el efecto espacial según el componente.
- Los efectos aleatorios resultan específicos para cada  $i$ .
- Se contempla la existencia e inclusión de variables estáticas pero heterogéneas transversalmente que tienen algún efecto significativo sobre  $y_{it}$ .
- Aunque se tienen efectos aleatorios específicos por unidad espacial, tales efectos relacionados a  $\mathbf{u}$  no se ponderan en las unidades espaciales, es decir, no es posible conocer en esta especificación el efecto de algún  $i$  sobre un  $j$ .
- El cálculo de los determinantes que estarían asociados al proceso de estimación de los parámetros resulta computacionalmente más complejo, en tanto las dimensiones del panel (en especial la dimensión longitudinal) aumenten.

### 3. Modelos SAR para datos panel con coeficientes espaciales específicos

En algunas ocasiones puede resultar del interés de un investigador conocer el efecto específico que tiene en agregado cada unidad espacial sobre las demás en un sistema particular, algo similar a una respuesta con la siguiente forma funcional:

$$y_{it} = \sum_{j=1}^N \rho_j w_{ij} y_{jt} + \sum_{k=1}^K \beta_k x_{k,it} + \epsilon_{it} \quad (3-1)$$

donde  $i = 1, \dots, N$ ,  $t = 1, \dots, T$  y  $\epsilon_{it} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ .

En este caso,  $\rho_j$  representa el efecto espacial agregado que tiene la unidad  $j$  sobre las demás, y que al ser ponderado por el rezago espacial de la variable respuesta, arroja el resultado específico sobre  $i$  en cada  $t$ .

Este capítulo está dedicado a desarrollar el proceso de estimación e inferencia de un modelo como el presentado en (3-1).

#### 3.1. Forma matricial

Se parte de la estructura para un  $t$  arbitrario cualquiera y omitiendo el componente exógeno  $\mathbf{X}\beta$ :

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & & w_{2N} \\ \vdots & & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{bmatrix} \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & 0 & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \rho_N \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{Nt} \end{bmatrix}$$

De forma que si se define  $\mathbf{Y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})^t$ ,  $\mathbf{W}$  tiene la misma estructura que la presentada en (2-4), y  $\mathbf{P} = \text{diag}(\rho_1, \dots, \rho_N)$ , el modelo anterior puede escribirse como:

$$\mathbf{Y}_t = \mathbf{W}\mathbf{P}\mathbf{Y}_t + \boldsymbol{\epsilon}_t \quad \text{donde} \quad \mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_T)$$

Ahora extendiéndolo a todo el panel:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_T \end{bmatrix} = \begin{bmatrix} \mathbf{W}\mathbf{P}\mathbf{Y}_1 \\ \mathbf{W}\mathbf{P}\mathbf{Y}_2 \\ \vdots \\ \mathbf{W}\mathbf{P}\mathbf{Y}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix} = \begin{bmatrix} \mathbf{W} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{P}\mathbf{Y}_1 \\ \mathbf{P}\mathbf{Y}_2 \\ \vdots \\ \mathbf{P}\mathbf{Y}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix}$$

donde  $\mathbf{0}$  son matrices de ceros de la dimensión adecuada.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{W} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W} \end{bmatrix} \begin{bmatrix} \mathbf{P} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{P} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_T \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix}$$

Las matrices anteriores pueden escribirse entonces como:

$$\mathbf{Y} = (\mathbf{I}_T \otimes \mathbf{W})(\mathbf{I}_T \otimes \mathbf{P})\mathbf{Y} + \boldsymbol{\epsilon}$$

Usando propiedades del producto Kroneker se puede reorganizar:

$$\mathbf{Y} = (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})\mathbf{Y} + \boldsymbol{\epsilon}$$

Insertando el componente exógeno  $\mathbf{X}\boldsymbol{\beta}$  con  $\mathbf{X}$  de dimensiones  $NT \times q$  y  $\boldsymbol{\beta}$  es un vector  $q \times 1$

$$\mathbf{Y} = (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3-2)$$

Finalmente se obtiene (3-2) que resulta más conveniente para los desarrollos posteriores.

Dado que el proceso tiene un componente auto-regresivo espacial, el proceso generador de datos se puede escribir como a continuación:

$$\mathbf{Y} = [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \quad (3-3)$$

donde  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT})$

## 3.2. Estimación e inferencia por MV de los parámetros

Dado el supuesto sobre el vector de errores  $\boldsymbol{\epsilon}$ , la función de log-verosimilitud asociada puede escribirse como (ver A.2 para más detalles de la prueba):

$$\begin{aligned} \ln(\mathcal{L}) = & -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln | \mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P}) | \\ & - \frac{1}{2\sigma^2} \{ ([\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t ([\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \} \end{aligned} \quad (3-4)$$

**Prueba:**

Primero se encuentra la media,

$$\begin{aligned} E(\mathbf{Y}) &= E\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\} \\ &= [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]^{-1}\mathbf{X}\boldsymbol{\beta} \end{aligned} \quad (3-5)$$

y luego la varianza,

$$\begin{aligned} Var(\mathbf{Y}) &= Var\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\} \\ &= Var\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]^{-1}\boldsymbol{\epsilon}\} \end{aligned} \quad (3-6)$$

Sea  $\mathbf{B} = [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]$ , entonces  $Var(\mathbf{Y}) = \mathbf{B}^{-1}Var(\boldsymbol{\epsilon})(\mathbf{B}^{-1})^t = \sigma^2\mathbf{B}^{-1}(\mathbf{B}^{-1})^t = \boldsymbol{\Sigma}$

Teniendo en cuenta que  $\mathbf{Y} \sim N([\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})]^{-1}\mathbf{X}\boldsymbol{\beta}; \boldsymbol{\Sigma})$ ,

$$\mathcal{L}(\boldsymbol{\beta}, \mathbf{P} | \mathbf{Y}) = \frac{1}{(2\pi)^{\frac{NT}{2}} | \boldsymbol{\Sigma} |^{\frac{1}{2}}} \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \quad (3-7)$$

Como el denominador puede escribirse:

$$(2\pi)^{\frac{NT}{2}} | \boldsymbol{\Sigma} |^{\frac{1}{2}} = (2\pi)^{\frac{NT}{2}} | \sigma^2\mathbf{B}^{-1}(\mathbf{B}^{-1})^t |^{\frac{1}{2}}$$

Se inserta en la verosimilitud (3-7):

$$\mathcal{L}(\boldsymbol{\beta}, \mathbf{P} | \mathbf{Y}) = \frac{1}{(2\pi\sigma^2)^{\frac{NT}{2}} | \mathbf{B}^{-1} |} \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \quad (3-8)$$

Se aplica logaritmo a (3-8) para trabajar con la log-verosimilitud

$$\begin{aligned} \ln(\mathcal{L}) &= \ln \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{NT}{2}} | \mathbf{B}^{-1} |} \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \right\} \\ &= \ln(1) - \ln \left\{ (2\pi\sigma^2)^{\frac{NT}{2}} | \mathbf{B}^{-1} | \right\} + \ln \left\{ \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \right\} \\ &= -\frac{NT}{2} \ln\{2\pi\sigma^2\} + \ln | \mathbf{B} | - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t [\mathbf{B}^{-1}(\mathbf{B}^{-1})^t]^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \end{aligned} \quad (3-9)$$

Expandiendo  $\mathbf{B}$  de (3-10) se llega al resultado de (3-4):

$$\ln(\mathcal{L}) = -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln |\mathbf{B}| - \frac{1}{2\sigma^2}(\mathbf{BY} - \mathbf{X}\beta)^t(\mathbf{BY} - \mathbf{X}\beta) \quad (3-10)$$

De la anterior función de log-verosimilitud es claro que el objetivo es encontrar estimadores para  $\beta$ ,  $\mathbf{P}$  y  $\sigma^2$ .

### 3.2.1. Estimadores y distribución asociada para $\hat{\beta}$ y $\hat{\sigma}^2$

El procedimiento común (encontrando los estimadores por MV, ver A.3) tan solo permite hallar las formas cerradas de  $\hat{\beta}$  y  $\hat{\sigma}^2$  en función de  $\mathbf{P}$ , mientras tanto, la estimación de  $\mathbf{P}$  necesita de algún método numérico enfocado en calcular su diagonal. No obstante, en esta sección se presentarán los estimadores para  $\beta$  y  $\sigma^2$  suponiendo que  $\mathbf{P}$  es conocido. Claramente, aunque en la práctica es necesario estimar  $\mathbf{P}$  antes que  $\beta$  y  $\sigma^2$ , es necesario suponer a  $\mathbf{P}$  conocido para insertar los resultados de tales estimadores en (3-4), de manera que las formas cerradas ya encontradas permitan obtener la función que arroje los  $\rho_j$  de (3-1). Siendo así, las condiciones de primer orden necesarias para hallar los estimadores de  $\beta$  y  $\sigma^2$ , los cuales se presentan a continuación:

$$\frac{\partial(\ln(\mathcal{L}))}{\partial\beta} = \frac{\partial\left(-\frac{NT}{2} \ln(2\pi\sigma^2) + \ln(|\mathbf{B}|) - \frac{1}{2\sigma^2}[\mathbf{BY} - \mathbf{X}\beta]^t[\mathbf{BY} - \mathbf{X}\beta]\right)}{\partial\beta} = 0$$

De donde se obtiene:

$$\hat{\beta} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]\mathbf{Y} \quad (3-11)$$

Con  $\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{B}\mathbf{B}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1})$ , ver A.3.4 para mayores detalles.

De lo anterior, resulta claro que  $\hat{\beta}_j \sim N(\beta_j, a_{jj}\sigma^2)$ , con  $a_{jj}$  el elemento  $j$ -ésimo de la diagonal de  $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{B}\mathbf{B}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}$ .

Mientras tanto, para  $\sigma^2$  la estimación por MV:

$$\frac{\partial(\ln(\mathcal{L}))}{\partial\sigma^2} = \frac{\partial\left[-\frac{NT}{2} \ln(2\pi\sigma^2)\right]}{\partial\sigma^2} + \frac{\partial[\ln |\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})|]}{\partial\sigma^2} + \frac{\partial\left[\frac{1}{2\sigma^2}(\mathbf{BY} - \mathbf{X}\beta)^t(\mathbf{BY} - \mathbf{X}\beta)\right]}{\partial\sigma^2} = 0$$

Deja como resultado:

$$\hat{\sigma}_{MV}^2 = \frac{(\mathbf{BY} - \mathbf{X}\hat{\beta})^t(\mathbf{BY} - \mathbf{X}\hat{\beta})}{NT} \quad (3-12)$$

Como suele ocurrir con el estimador por MV de la varianza, el resultado es sesgado, por lo que al realizar la corrección del sesgo se obtiene (ver A.3.3 para la demostración respectiva):



$$\hat{\sigma}_{ins}^2 = \frac{1}{NT - K} (\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})^t (\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (3-13)$$

donde  $\frac{(NT-K)\hat{\sigma}_{ins}^2}{\sigma^2} \sim \chi_{(NT-K)}^2$ .

### 3.2.2. Estimación e inferencia de $\mathbf{P}$

Dado que  $\mathbf{P}$  es una matriz condicionada a que todos los elementos fuera de la diagonal son cero, la optimización debe realizarse sólo sobre la diagonal de  $\mathbf{P}$ , es importante tener tal condición en cuenta a la hora de escribir las funciones de optimización en cualquier aplicación. Lo anterior también resulta importante teóricamente, ya que la matriz de covarianzas necesaria para hacer inferencia sobre los parámetros es de dimensión  $N \times N$  y no  $NN \times NN$ . Con esto en cuenta, se define  $\mathbf{P}^* = \text{diag}(\mathbf{P})$ , de forma que  $\mathbf{P}^*$  es un vector de dimensión  $N \times 1$  que contiene los efectos auto-regresivos espaciales específicos.

Para encontrar la función donde se maximice  $\mathbf{P}^*$ , se insertan los resultados de  $\hat{\boldsymbol{\beta}}$  y  $\hat{\sigma}_{MV}^2$  en la función de verosimilitud:

$$f(\mathbf{P}^*) = -\frac{NT}{2} \ln(2\pi\hat{\sigma}^2) + \ln |\mathbf{B}| - \frac{1}{2\hat{\sigma}^2} \left[ (\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})^t (\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right]$$

Ahora la optimización se debe concentrar sobre  $\mathbf{P}^*$ , así que se usa la siguiente notación para separarlo de  $\mathbf{B}$  y dejarlo más simplificado:

$$\begin{aligned} \mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{BY} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{BY} = (\mathbf{I}_{NT} - \mathbf{H})\mathbf{BY} \\ &= (\mathbf{I}_{NT} - \mathbf{H})[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{P})(\mathbf{I}_T \otimes \mathbf{W})]\mathbf{Y} \end{aligned}$$

donde  $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ .

Como hay constantes, definiendo  $C = -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2}$ , la función final resulta en:

$$\begin{aligned} f(\mathbf{P}^*) &= C - \frac{NT}{2} \ln([(\mathbf{I}_{NT} - \mathbf{H})[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{P}^*)(\mathbf{I}_T \otimes \mathbf{W})]\mathbf{Y}]^t \\ &\quad (\mathbf{I}_{NT} - \mathbf{H})[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{P}^*)(\mathbf{I}_T \otimes \mathbf{W})]\mathbf{Y}) + \ln |\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{P}^*)(\mathbf{I}_T \otimes \mathbf{W})| \end{aligned}$$

En términos generales, la maximización de la función se puede llevar a cabo con cualquier algoritmo de optimización multivariado que el usuario considere conveniente. Sin embargo, a continuación se muestra una de las aproximaciones que resultará útil, y suele ser el punto de partida de métodos numéricos de optimización más robustos (el método de Newton adaptado a este caso particular):

- Descomponga  $f(\mathbf{P}^*)$  mediante una serie de Taylor y conserve los tres primeros sumandos (segundo orden):

$$f(\mathbf{P}^*) \approx f(\mathbf{P}^{*(0)}) + (\mathbf{P}^* - \mathbf{P}^{*(0)})^t \nabla f(\mathbf{P}^*) + \frac{1}{2}(\mathbf{P}^* - \mathbf{P}^{*(0)})^t \mathbf{H}(\mathbf{P}^* - \mathbf{P}^{*(0)})$$

donde  $\mathbf{H}$  es la matriz Hessiana correspondiente y  $\mathbf{P}^{*(0)}$  es el vector de valores iniciales que se provee al algoritmo.

- Se buscan los puntos críticos diferenciando e igualando a cero

$$\frac{\partial f(\mathbf{P}^*)}{\partial \mathbf{P}^*} = 0$$

Para el siguiente resultado es útil tener claro que para una forma cuadrática de la forma:  $q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^t \mathbf{A} \mathbf{x} + \mathbf{b}^t \mathbf{x} + \mathbf{c}$ , el gradiente  $\nabla q(\mathbf{x}) = \mathbf{A} \mathbf{x} + \mathbf{b}$

Por ende,

$$\frac{\partial f(\mathbf{P}^*)}{\partial \mathbf{P}^*} = \mathbf{H}(\mathbf{P}^* - \mathbf{P}^{*(0)}) + \nabla f(\mathbf{P}^*) = 0$$

$$\mathbf{P}^* = \mathbf{P}^{*(0)} - \mathbf{H}^{-1} \nabla f(\mathbf{P}^*)$$

- Realice las  $m$  iteraciones necesarias hasta que de con la tolerancia deseada (los paquetes usualmente usan 1e-6 como valor por defecto) usando la siguiente expresión:

$$\mathbf{P}^{*(m+1)} = \mathbf{P}^{*(m)} - \mathbf{H}^{-1} \nabla f(\mathbf{P}^{*(m)})$$

Además, con la función  $f(\mathbf{P}^*)$  bien definida, mediante el cálculo del Hessiano, se puede obtener una estimación de la matriz de información de Fisher observada, la cual produce asintóticamente los mismos resultados que la matriz de información de Fisher teórica, y de la cual es posible calcular los errores estándar de los  $\rho_i$ . Así, la información de Fisher de los parámetros  $\rho_i$  para  $i = 1, \dots, N$  es:

$$\mathbf{I}(\mathbf{P}^*) = \left[ \frac{\partial^2}{\partial \rho_i \partial \rho_j} f(\mathbf{P}^*) \right]$$

Para  $1 \leq i, j \leq N$ . Y la información de Fisher observada ( $\mathbf{I}_O(\mathbf{P}^*)$ ), i.e., la matriz de información evaluada en las estimaciones máximo verosímiles:

$$\mathbf{I}_O(\mathbf{P}^*) = -\mathcal{H}(f(\mathbf{P}^*))$$

Con lo anterior, aprovechando las propiedades asintóticas subyacentes de la Ley de los grandes números y el teorema del límite central de la función de score y de la distribución del estimador MV (descritas a detalle en el texto de Pawitan (2001), capítulos 8 y 9), se puede usar  $\mathbf{I}_O(\mathbf{P}^*)$  como un estimador asintótico de la matriz de covarianza con condiciones de normalidad:

$$\hat{\mathbf{P}}_{MV}^* \stackrel{a}{\sim} N(\mathbf{P}^*, [\mathbf{I}_O(\mathbf{P}^*)]^{-1})$$

Así, el error estándar del elemento  $\rho_j$  corresponde a:

$$ee(\rho_i) = \sqrt{k_{ii}}$$

donde  $k_{ii}$  es el elemento  $i$ -ésimo de la diagonal de la matriz  $\mathbf{I}_O(\mathbf{P}^*)^{-1}$ .

### 3.2.3. Construcción de intervalos de confianza para $\beta_j$ y $\rho_i$

Combinando los resultados de la distribución de  $\hat{\beta}$  y la ecuación (3-11), para la construcción de un intervalo al  $(1 - \alpha)\%$  de confianza para  $\beta_j$ , se escribe:

$$\frac{\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{a_{jj}}}}{1/\sigma \sqrt{SCE/(NT - K - 1)}} \sim t_{(NT - K - 1)}$$

Incluyendo  $SCE = \hat{\sigma}^2(NT - K - 1)$

Un intervalo al  $(1 - \alpha)\%$  de confianza para  $\beta_j$  está dado por:

$$\hat{\beta}_j \pm t_{(NT - K - 1, \alpha/2)} \sqrt{a_{jj}} \hat{\sigma}$$

Por otro lado, con las propiedades asintóticas discutidas en la sección anterior, relativas a la distribución de  $\mathbf{P}_{MV}^*$ , para cada  $\rho_i$  se tiene:

$$\hat{\rho}_i \stackrel{a}{\sim} N(\rho_i, k_{ii})$$

De modo que para probar  $H_0 : \rho_i = \rho_0$  vs.  $H_1 : \rho_i \neq \rho_0$  resulta útil el uso del estadístico de Wald:

$$z = \frac{\hat{\rho}_i - \rho_0}{ee(\rho_i)} = \frac{\hat{\rho}_i - \rho_0}{\sqrt{k_{ii}}}$$

Y un intervalo al  $(1 - \alpha)\%$  de confianza para  $\rho_i$  resulta en:

$$\hat{\rho}_i \pm z_{\alpha/2} \sqrt{k_{ii}}$$

### 3.2.4. Contraste de hipótesis para evaluar la necesidad de coeficientes específicos vs. coeficiente único

En general, el proceso de optimización de la diagonal de la matriz  $\mathbf{P}$  es más costoso computacionalmente que la estimación de un único  $\rho$  (el SARP tradicional), ya que requiere optimizar sobre  $N - 1$  dimensiones adicionales, además que este último es más parsimonioso en términos del número de parámetros estimados.

Por tales razones, para un investigador puede resultar de interés evaluar la hipótesis:

$$H_0 : \rho = \rho_i$$

para  $i = 1, \dots, N$ .

Para efectos de construir el procedimiento de contraste, se reescribe tal hipótesis como a continuación:

$$H_0 : \begin{bmatrix} \rho \\ \rho \\ \vdots \\ \rho \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_N \end{bmatrix} \quad (3-14)$$

donde se define al vector  $\mathbf{P}_u^* = (\rho, \rho, \dots, \rho)^t$  y  $\mathbf{P}^* = (\rho_1, \rho_2, \dots, \rho_N)^t$ .

Con estas especificaciones, se aclara que el contraste de hipótesis se realiza sobre las estimaciones espaciales auto-regresivas de (2-7) respecto de (3-1).

Así, para evaluar (3-14), es decir,  $H_0 : \mathbf{P}_u^* = \mathbf{P}^*$  o  $H_0 : \mathbf{P}_u^* - \mathbf{P}^* = \mathbf{0}$  asumiendo que ambos vectores aleatorios son independientes el uno del otro y definiendo  $\mathbf{d} = \mathbf{P}_u^* - \mathbf{P}^*$  puede usarse a  $\mathbf{d} \sim N_N(\mathbf{0}, \Sigma_{\mathbf{P}_u^*} + \Sigma_{\mathbf{P}^*})$  para realizar el contraste. En este caso,  $\Sigma_{\mathbf{P}^*} = \mathbf{I}_O(\mathbf{P})^{-1}$  y  $\Sigma_{\mathbf{P}_u^*} = \sigma^2 \mathbf{J}_N$ , con  $\mathbf{J}$  una matriz cuadrada de unos de la dimensión indicada.

Aunque en este caso  $\Sigma_{\mathbf{P}_u^*}$  y  $\Sigma_{\mathbf{P}^*}$  no son conocidas pero si estimables de la MV de donde se obtienen los  $\rho$ , la suma de tales matrices con la independencia que se presupone resulta en  $\hat{\Sigma}_{\mathbf{P}_u^*} + \hat{\Sigma}_{\mathbf{P}^*} \sim \mathcal{W}_N(NT, \Sigma_{\mathbf{P}_u^*} + \Sigma_{\mathbf{P}^*})$ , ver Kumar (2014).

Bajo las condiciones anteriores, puede usarse el estadístico  $T^2$  de Hotelling para matrices de covarianza heterogeneas no conocidas, con el fin de evaluar el contraste de hipótesis de interés (ver Bibby, Kent & Mardia (1979)):

$$T^2 = (\hat{\mathbf{P}}_{\mathbf{u}}^* - \hat{\mathbf{P}}^*)^t [NT(\hat{\Sigma}_{\mathbf{P}_u^*} + \hat{\Sigma}_{\mathbf{P}^*})]^{-1} (\hat{\mathbf{P}}_{\mathbf{u}}^* - \hat{\mathbf{P}}^*)$$

$$T^2 \approx \frac{vN}{v - N + 1} F_{(N, v-N+1)}$$

donde,

$$v = \frac{N + N^2}{\sum_{i=1}^2 \left\{ tr \left[ \left( \frac{1}{NT} \hat{\Sigma}_i \hat{\Sigma}_i^{-1} \right)^2 \right] + tr \left( \left[ \frac{1}{NT} \hat{\Sigma}_i \hat{\Sigma}_i^{-1} \right] \right)^2 \right\}}$$

donde  $tr(\cdot)$  se refiere a la traza de la matriz,  $\hat{\Sigma} = \frac{1}{NT}(\hat{\Sigma}_{\mathbf{P}_u^*} + \hat{\Sigma}_{\mathbf{P}^*})$  e  $i = 1, 2$ , asociados a las matrices  $\mathbf{P}^*, \mathbf{P}_u^*$ , respectivamente.

Una vez estimados los cálculos anteriores, el valor-p para decidir acerca de  $H_0$  puede obtenerse como:

$$\text{valor p} = P \left( F(\cdot) > \frac{v - N + 1}{vN} T^2 \right) = 1 - F_{(N, v-N+1)} \left( \frac{v - N + 1}{vN} T^2 \right)$$

## 4. Un análisis al Índice de Desarrollo Humano y sus componentes descontando efectos espaciales

El IDH (Índice de Desarrollo Humano), diseñado como una medida que permitiera medir el desarrollo económico y la calidad de vida de una población más allá de indicadores de crecimiento económico ajustados al tamaño población (como el Producto Interno Bruto per cápita o el Producto Nacional Bruto per cápita), fue introducido por el Programa de Desarrollo de las Naciones Unidas en 1990, y se ha convertido en una medida popular entre los académicos a la hora de evaluar la condición de una sociedad o comparar la calidad de vida de las mismas jerárquicamente.

De acuerdo a lo anterior, el IDH no tiene solamente en cuenta la producción agregada de riqueza en proporción al tamaño de la población como dimensión de medida para el desarrollo humano, sino que también incorpora longevidad y salud, y educación dentro de su composición, como se describe en la tabla resumen [4-1](#).

**Tabla 4-1.:** Composición del IDH

Dimensión	Longevidad y vida saludable	Educación	Estándar de vida decente
<b>Indicadores</b>	Esperanza de vida al nacer	Promedio y esperanza de años de escolaridad	Ingreso Interno Bruto (IIB) per capita
<b>Índice</b>	Índice de esperanza de vida	Índice Educativo	Índice IIB

No obstante, y a pesar que el objetivo era ajustar en proporciones “equivalentes”, la influencia de cada una de las dimensiones sobre el resultado final del índice, i.e., el peso de cada uno de los componentes, no es necesariamente constante o igual a través de los países, o los años de medición, precisamente por la metodología de construcción del índice (para ver detalles

acerca de tal metodología remítase al Anexo B).

Además de la variabilidad transversal y temporal en la influencia de las dimensiones sobre el resultado del IDH, existe literatura que señala que el desarrollo humano contiene un componente de correlación espacial inmerso en su naturaleza (i.e. estar más cerca o más lejos de países o regiones más desarrollados puede influenciar el IDH de una nación o localidad). En línea con lo anterior, se encuentra el trabajo de Qiu, Sung, Davis & Tchernis (2018), donde se propone un modelo CAR (Auto-regresivo Condicional por sus siglas en inglés) espacial con variable latente que resulta en una mejoría en la precisión de las estimaciones al incluir el componente espacial. Otros adelantos como Letkovičová et al. (2014), Celemín & Velázquez (2017) y Shaker (2015) corroboran mediante pruebas de correlación espacial (generalmente la de Moran (1950)) y con mapas de calor, la existencia de un componente espacial dentro de los indicadores relacionados con el desarrollo humano (siempre se incluye al IDH como punto de referencia). Sin embargo, nunca es claro que de algún modo se descuenta o cuantifique la magnitud del efecto espacial sobre las estimaciones.

Por tanto, en vista de la variabilidad temporal y transversal de los pesos o influencia de los índices que componen la medida, la evidencia de correlación espacial que trabajos anteriores señalan existe en el IDH por regiones; el vacío de literatura que descuenta efectos espaciales sobre la influencia de los componentes o variables explicativas de los índices de desarrollo humano, y la disponibilidad de datos para construir un panel balanceado, se hace pertinente estimar las relaciones estructurales entre el IDH y sus componentes a nivel agregado (“muestra global”) descontando o limpiando los parámetros asociados de la variabilidad espacial. Se realizará la estimación comparando el SAR Panel tradicional con la propuesta de este trabajo (SARP con coeficientes espaciales específicos), de forma que puedan ilustrarse las ventajas y limitaciones de la misma frente a trabajos previos.

## 4.1. Descripción de los datos

Con el fin de cubrir la mayor parte de las regiones posibles, la muestra de datos que se escogió para el panel contiene 12 países (3 para cada uno de los continentes de interés; LATAM, Norte-América, Europa y Asia<sup>1</sup>) en un rango de tiempo de 1990 al 2017 con frecuencia anual. La fuente de las variables dependiente y explicativas fue la página web del Programa de Desarrollo de las Naciones Unidas<sup>2</sup>.

Específicamente, la variable dependiente a analizar es el IDH, mientras que las explicativas

<sup>1</sup>África fue excluido debido a que la disponibilidad de datos es escasa y no lograba cumplir con la misma ventana temporal de los otros grupos

<sup>2</sup>hdr.undp.org

son el Índice Educativo (IE), el Índice de Esperanza de Vida (IEV) y una medida de ingreso o riqueza material (en este caso se toma el logaritmo natural del IIB).

La tabla 4-2 muestra los promedios por país de las observaciones en el periodo especificado, lo que da una idea para comparar los niveles de las unidades transversales.

**Tabla 4-2.:** Promedios de los índices en el periodo de estudio.

País	IDH	IE	ln(IIB)	IEV
Austria	0,855	0,765	10,576	0,907
Brasil	0,693	0,585	9,407	0,786
Canadá	0,884	0,841	10,509	0,922
China	0,633	0,525	8,539	0,816
Colombia	0,675	0,567	9,139	0,795
India	0,528	0,422	8,028	0,674
Italia	0,839	0,722	10,459	0,928
Corea del Sur	0,834	0,800	10,031	0,884
México	0,716	0,578	9,602	0,841
Perú	0,687	0,632	8,919	0,789
Suiza	0,895	0,811	10,881	0,933
USA	0,895	0,875	10,745	0,883

#### 4.1.1. Matriz de pesos espaciales

La matriz de pesos espaciales fue construida a partir de las distancias euclidianas entre los centroides de los países incluidos en las estimaciones. Las distancias para construir la matriz se obtuvieron del proyecto GADM<sup>3</sup> y luego se realizaron ciertas transformaciones que adecuaron las magnitudes de los datos y estandarizaron el producto final (la matriz inicial y el producto final puede ser consultado en el Anexo D):

- Se estandariza en filas la matriz de distancias inicial. Esta transformación se realiza debido a que si se aplica una función de potencia negativa directamente a las distancias se obtienen números muy “pequeños” que pueden implicar pérdida en la eficacia de las

<sup>3</sup>[www.gadm.org](http://www.gadm.org), GADM es un proyecto que busca condensar los datos geográficos de todas las áreas administrativas a nivel global y ponerlas a disposición libre de usuarios con fines académicos y/o no comerciales



estimaciones debido a la precisión por decimales que está implícita en la configuración computacional de los sistemas.

- Se aplica una función de potencia negativa a la matriz de distancias estandarizada, de forma que se invierta el efecto de la cercanía/lejanía entre las unidades espaciales y guarde coherencia con la agregación auto-regresiva:

$$w_{ij} = \frac{d_{ij}^{-\alpha}}{\sum_j d_{ij}^{-\alpha}}$$

donde  $w_{ij}$  es el elemento correspondiente de la matriz de pesos espaciales y  $d_{ij}$  es el elemento correspondiente a la matriz de distancias euclidianas estandarizada. Finalmente, se tomó  $\alpha = 1$  en relación con no agregar mayor desproporción en la inversión del efecto.

Aquí es importante notar que la matriz final también resulta estandarizada, lo que mejora la eficiencia de los cálculos dada la cantidad de decimales provenientes de las magnitudes “grandes” de las distancias.

#### 4.1.2. Prueba de Moran para validar correlación espacial

Antes de presentar los resultados de las estimaciones, tiene sentido corroborar la presencia de correlación espacial en los datos a trabajar. Sin embargo, dado que se está ante datos tipo panel, la prueba debe realizarse para los  $T$  periodos en que se observan las unidades transversales.

El estadístico  $I$  a usar, según la propuesta de Moran (1950) ajustado a la notación que se está trabajando tiene la forma:

$$I = \frac{N}{\sum_{i,j=1}^N w_{ij}} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

El cual puede contrastarse mediante la distribución normal estándar con la siguiente transformación:

$$z_I = \frac{I - E(I)}{\sqrt{V(I)}}$$

donde  $E(I) = -1/(N - 1)$  y  $V(I) = E(I^2) - E(I)^2$ .

Los resultados de las pruebas en los  $T$  periodos arrojaron evidencia significativa que sustenta la presencia de correlación espacial en la variable dependiente (el IDH). Tales resultados pueden visualizarse en el la tabla **B-1** del Anexo **B**.

En tanto todos los resultados de la estadística parecen ser más o menos estables a través del tiempo, y la estadística  $z$  es siempre positiva, se puede concluir que los datos reflejan una distribución donde los valores de un IDH alto o bajo están más agrupados espacialmente de lo que se esperaría si el proceso espacial subyacente fuera aleatorio, es decir, los países con un IDH alto tienden a estar cerca de otros países con un IDH similar. Lo anterior hace pensar, dado que las muestras se tomaron en vecindades segregadas por continente, que las características propias de las localidades continentales explican de alguna manera el nivel de desarrollo de las naciones y da paso a los experimentos por componentes descontando efectos espaciales.

## 4.2. Resultados

Como se observa a continuación, aunque en términos de pronóstico la metodología de componentes espaciales específicos parece mejorar el ajuste (lo que se corrobora de forma gráfica y a través de las medidas de bondad de ajuste que se proponen en el Anexo **A**), es muy interesante el hecho de la diferencia en los parámetros estimados asociados a las dimensiones que componen la medida, a favor de la existencia de alguna ganancia informativa al estimar los coeficientes específicos. Principalmente se destaca el cambio que se tiene en la poderación obtenida del componente de riqueza de una estimación a otra.

### 4.2.1. Estimación bajo la metodología del SARP tradicional

En este caso, se estima el modelo relacionado a la ecuación (2-7), donde  $\rho$  es único, el vector  $\mathbf{Y}$  está asociado al IDH y la matriz  $\mathbf{X}$  contiene las variables explicativas y una dummy para cada una de las unidades transversales, de modo que se puede corregir el problema de variable omitida por los niveles específicos que se observarían en los residuos y en las gráficas de dispersión. Los parámetros asociados se encuentran en la tabla **4-3**.

Con el fin de ilustrar la bondad del ajuste y contrastar posteriormente con la metodología de coeficientes espaciales específicos, los diagramas de dispersión del IDH observado y estimado contra las variables explicativas acordes a estas estimaciones se presentan en las figuras **4-1**, **4-2** y **4-3**.

Los resultados bajo esta metodología muestran los signos esperados en relación a las variables explicativas diferentes a las indicativas por país. No obstante, llama la atención que el

**Tabla 4-3.:** Estimación usando SARP tradicional

Parámetro	Valor	Estadístico	Valor-p	IC 95 %	
<b>Coefficiente espacial</b>		<b>z</b>			
$\rho$	-0,877	-62106	0,000	-0,877	-0,876
<b>Variables explicativas</b>		<b>t</b>			
IE	0,519	25,8	0,000	0,480	0,559
IEV	0,057	19,2	0,000	0,052	0,063
ln(IIB)	1,028	31,3	0,000	0,963	1,092
Constante	-0,347	-12,7	0,000	-0,400	-0,293
BR	-0,300	13,3	0,000	-0,349	-0,251
CA	-0,419	-28,5	0,000	-0,472	-0,367
CN	-0,299	9,4	0,000	-0,347	-0,250
CO	-0,300	12,3	0,000	-0,348	-0,251
IN	-0,166	29,5	0,000	-0,210	-0,122
IT	-0,356	-4,7	0,000	-0,411	-0,301
KR	-0,437	-29,9	0,000	-0,487	-0,386
MX	-0,299	14,2	0,000	-0,351	-0,247
PE	-0,314	7,4	0,000	-0,360	-0,267
CH	-0,390	-29,5	0,000	-0,445	-0,335
US	-0,416	-20,1	0,000	-0,468	-0,364
$R^2 = 99,85 \%$					
$R^2$ Ajustado = 99,56 %					
$F = 724235,6$		Valor - p = 0,000			

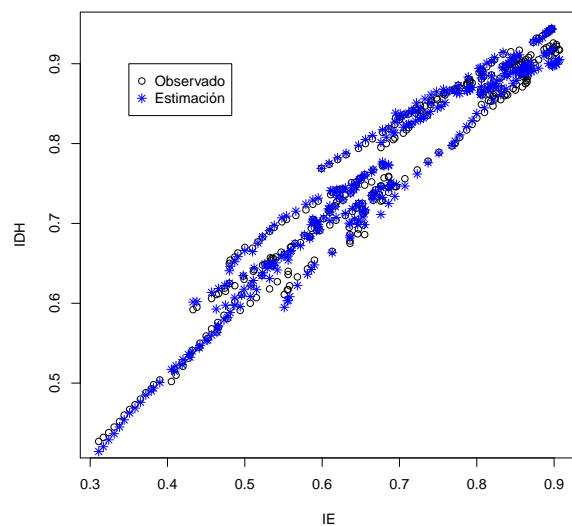
parámetro relacionado con el estándar de vida (el  $\ln(\text{IIB})$ ) es de magnitud muy superior a los otros dos componentes, mientras que el relacionado con el de salud y longevidad tiene una magnitud reducida. Esta es una debilidad que surge de la forma de construcción del IDH en relación a la uniformidad que se buscaba en relación a la ponderación de los componentes. La debilidad se explica por dos frentes. Uno, dado que la dimensión de estándar de vida se construye a partir del ingreso per cápita de la economía objetivo, las magnitudes tienden a ser “grandes” respecto de las otras dos dimensiones. Y dos, la variabilidad proveniente de esta dimensión es mucho mayor en tanto el IIB de una economía se iguala en cuentas nacionales al Producto Interno Bruto (PIB), que a su vez cambia en función de un grupo muy amplio de factores (e.g. políticas económicas vigentes, ciclos económicos, precios internacionales de las materias primas, choques externos, cambios tecnológicos, etc.), mientras que las dimensiones de educación, y salud y longevidad no varían de manera drástica a menos que exista un cambio estructural en las políticas públicas (lo que es poco frecuente en contraste). Por otro lado, se observa una fuerte correlación negativa (-91 %) entre el estimado del parámetro indicativo por país y el promedio de su IDH (ver tabla 4-2). Tal correlación señala un par de aspectos clave; primero, que existe un efecto proveniente de la dimensión transversal, es decir, que la forma en que se relacionan los componentes del IDH con el mismo varía de país a país, y segundo, que es necesaria una compensación de las unidades transversales que tienen una pendiente mayor de mejora año a año (haciendo más negativo su intercepto).

#### 4.2.2. Estimación bajo la metodología del SARP con coeficientes espaciales específicos

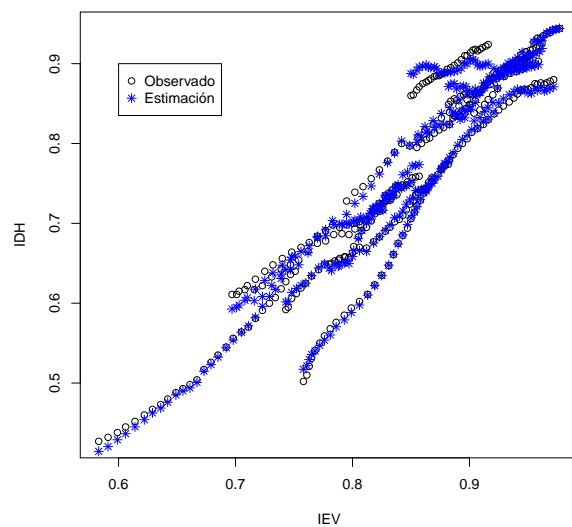
Para la estimación con coeficientes espaciales específicos (acorde a la forma propuesta en (3-2)) se incluyeron las mismas variables y transformaciones en  $\mathbf{W}$ ,  $\mathbf{X}$  y  $\mathbf{Y}$  respecto de las incorporadas en el SARP tradicional, con el fin de hacer de la comparación lo más limpia posible, por tanto los resultados tienen solamente diferencia en cantidad e interacción de los coeficientes espaciales (hablando de la estructura del modelo propuesto) en lo que a especificación paramétrica concierne. Los resultados se presentan en la tabla 4-4.

Las gráficas con las dispersiones entre el IDH observado y estimado y las variables explicativas se pueden visualizar en las figuras 4-6, 4-7 y 4-8.

Para resaltar respecto a los resultados de esta metodología, destaca la importante disminución en la magnitud del parámetro relacionado con la dimensión de estándar de vida, lo que sugiere que la inclusión de coeficientes espaciales específicos ayuda a suavizar el efecto que tiene la variabilidad y magnitud del indicador construido a partir del comportamiento del IIB per cápita, sin que este pierda significancia.



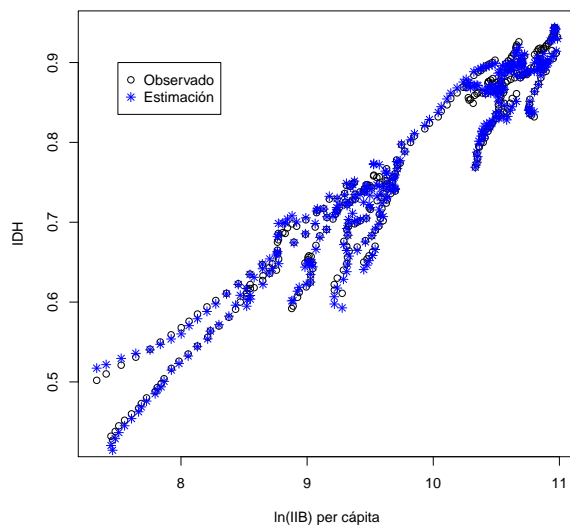
**Figura 4-1.:** Dispersión de IDH vs. IE (Dimensión de conocimiento) - Coeficiente único.



**Figura 4-2.:** Dispersión de IDH vs. IEV (Dimensión de salud y longevidad) - Coeficiente único.

**Tabla 4-4.:** Estimación usando SARP coeficientes específicos.

Parámetro	Valor	Estadístico	Valor-p	IC 95 %	
<b>Coeficientes espaciales</b>		<b>z</b>			
Constante	-0,241	-69,3	0,000	-0,033	-0,031
BR	0,033	76,4	0,000	0,032	0,034
CA	-0,011	-2,9	0,011	-0,018	-0,004
CN	0,027	41,6	0,000	0,026	0,028
CO	0,007	1,0	0,506	-0,007	0,021
IN	0,024	30,1	0,000	0,022	0,026
IT	0,015	7,3	0,000	0,011	0,019
KR	-0,061	-503,5	0,000	-0,061	-0,061
MX	0,040	138,2	0,000	0,039	0,041
PE	-0,018	-12,8	0,000	-0,021	-0,015
CH	-0,027	-41,8	0,000	-0,028	-0,026
US	-0,009	-1,7	0,174	-0,019	0,001
<b>Variables explicativas</b>		<b>t</b>			
IE	0,409	149,3	0,000	0,403	0,414
IEV	0,053	145,7	0,000	0,053	0,054
ln(IIB)	0,247	52,7	0,000	0,238	0,257
AT	-0,241	-76,8	0,000	-0,247	-0,235
BR	-0,240	4,4	0,000	-0,245	-0,234
CA	-0,248	-20,4	0,000	-0,254	-0,242
CN	-0,226	23,1	0,000	-0,232	-0,220
CO	-0,241	0,4	0,000	-0,247	-0,235
IN	-0,235	9,6	0,000	-0,240	-0,230
IT	-0,226	42,3	0,000	-0,232	-0,219
KR	-0,253	-30,4	0,000	-0,259	-0,248
MX	-0,237	10,8	0,000	-0,243	-0,231
PE	-0,248	-13,2	0,000	-0,254	-0,243
CH	-0,242	-1,9	0,000	-0,248	-0,236
US	-0,260	-36,8	0,000	-0,265	-0,254
R Cuadrado = 99,99 %					
R Cuadrado Ajustado = 99,99 %					
F = 112896		Valor - p = 0,000			

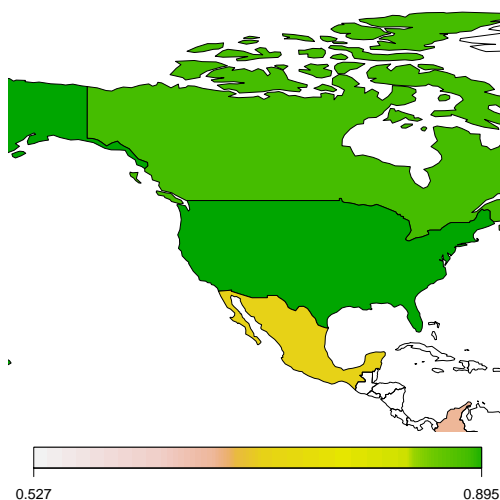


**Figura 4-3.:** Dispersión de IDH vs.  $\ln(\text{IIB})$  (Dimensión de estándar de vida) - Coeficiente único.

Por otro lado, llama la atención la ausencia de importancia asociada a los parámetros espaciales de Canadá, Estados Unidos y Colombia, lo que podría traducirse en ausencia de impactos por condiciones internas desde México y Colombia hacia sus vecinos.

Respecto a Canadá y Estados Unidos, los resultados son reflejo del hecho que aunque México se encuentra en esta vecindad, la reducida distancia entre este y sus vecinos del norte no implica que su índice de calidad de vida mejore (ver el mapa en 4-4), lo que no extraña dado que a pesar de los fuertes lazos económicos que comparten, México aún sufre de mayor debilidad institucional, problemas de violencia y narcotráfico, que incluso se alimentan de los consumidores en las economías desarrolladas con las que comparte vecindad, en tanto no existe un uso eficiente de los recursos para generar ingreso y reducir la inequidad, ver Ayvar-Campos, Navarro-Chávez & Giménez-García (2017).

Para el caso de Colombia, los resultados sugieren que las problemáticas internas que afectan las mediciones del país en relación con las dimensiones que componen el IDH no terminan por afectar a nuestros vecinos (ver mapa en 4-5), no obstante, la salida obtenida para Brasil y Perú podría ser un reflejo de los efectos que cada país tiene en la vecindad, ya que mientras el IDH superior de Brasil respecto a sus vecinos, en combinación con el tamaño de su economía implicaría un efecto positivo en LATAM (el PIB nominal en USD de Brasil para 2018 fue



**Figura 4-4.:** IDH América del Norte.

de más de USD 1'800.000 millones, mientras el de Colombia de menos de USD 300.000 millones), Perú terminaría por afectar en una menor proporción (con el IDH más bajo de la vecindad) debido a que es menos abierto y masivo en términos de producto que Colombia y Brasil (el PIB nominal del Perú en 2018 fue menor a los USD 230.000 millones<sup>4</sup>).

Por el lado de Europa, la relativa homogeneidad que existe en los IDH de los países de esa región explica los resultados significativos que tuvieron los coeficientes espaciales específicos para esas unidades transversales y que podrían terminar retroalimentándose entre sí (ver mapa en 4-9).

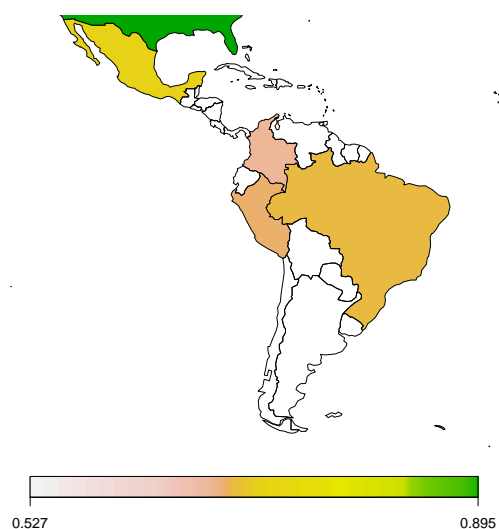
Finalmente, la distribución espacial (IDH mejorando de occidente a oriente) y el tamaño de la economía de las unidades asiáticas ayuda a explicar la significancia de estos resultados, la economía china (con un PIB nominal anual mayor a los 13'400.000 millones en 2018) tiene un IDH intermedio que impulsaría el resultado de India, mientras que el IDH superior de Korea del Sur no logra empujar a sus vecinos y se compensa con un coeficiente negativo (ver mapa en 4-10).

En relación a la calidad de las estimaciones, una vista detallada de las gráficas permite observar que el ajuste del modelo con coeficientes espaciales específicos se desempeña mejor en términos generales, pero las bondades en este caso tienden a verse con claridad en las

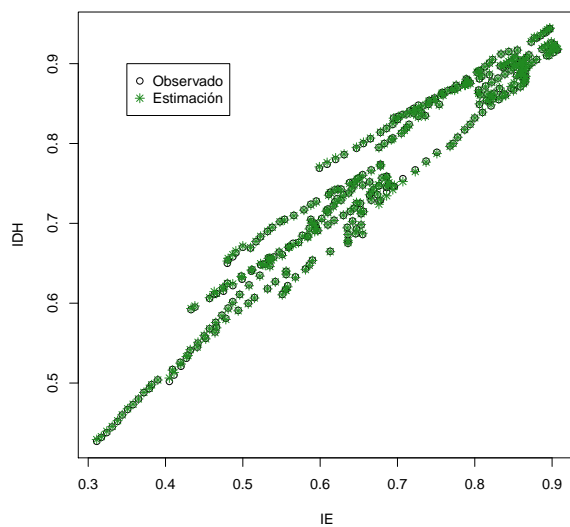
---

<sup>4</sup>Datos del FMI

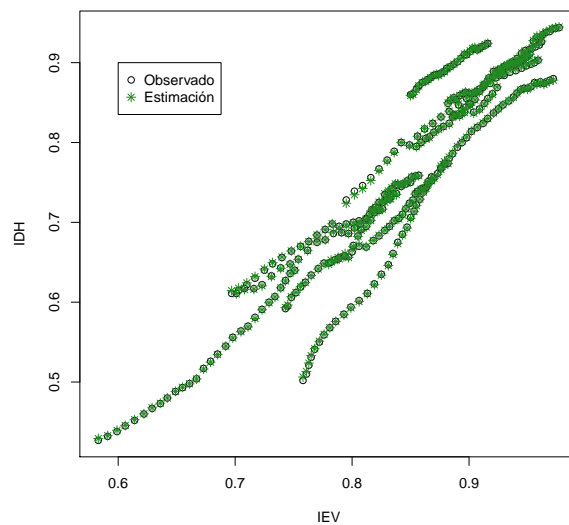




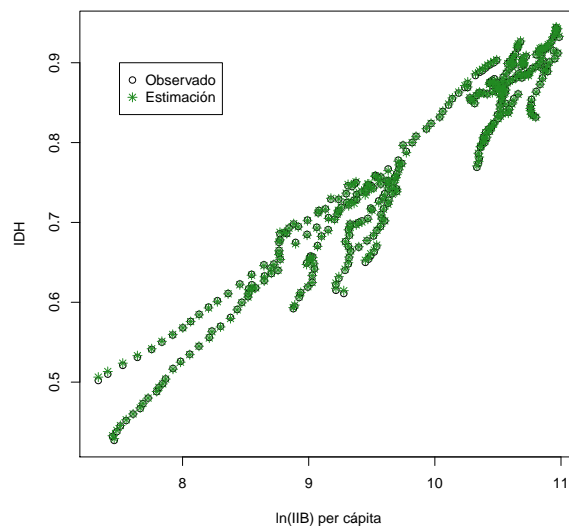
**Figura 4-5.:** IDH LATAM.



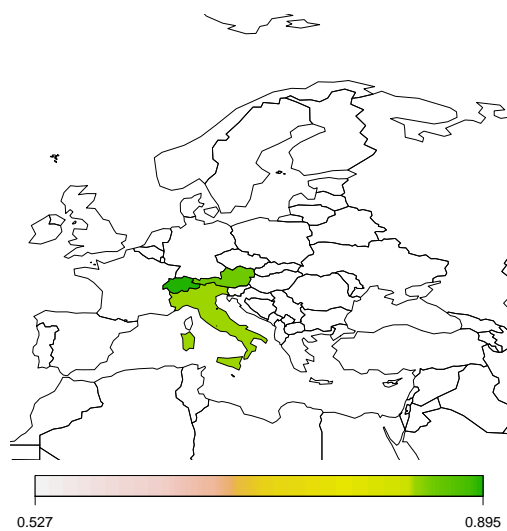
**Figura 4-6.:** Dispersión de IDH vs. IE (Dimensión de conocimiento) - Coeficientes específicos.



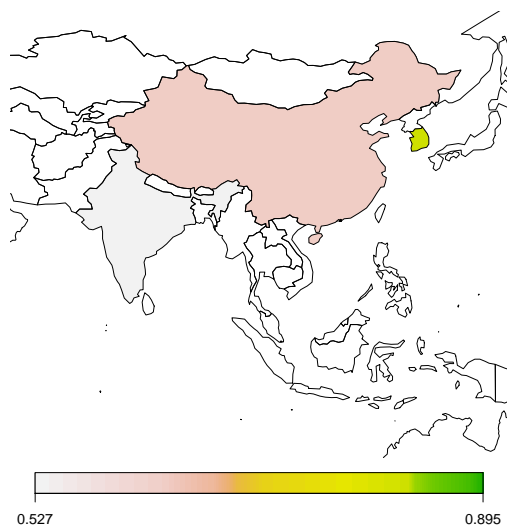
**Figura 4-7.:** Dispersión de IDH vs. IEV (Dimensión de salud y longevidad) - Coeficientes específicos.



**Figura 4-8.:** Dispersión de IDH vs.  $\ln(\text{IIB})$  (Dimensión de estándar de vida) - Coeficientes específicos.



**Figura 4-9.: IDH Europa.**



**Figura 4-10.: IDH Asia.**

observaciones más alejadas del promedio, sobretodo donde hay grandes agrupaciones de datos, lo que ocurre en la dispersión del centro y la esquina superior derecha de las gráficas (figuras 4-1, 4-2, 4-3, 4-6, 4-7 y 4-8), donde se concentran las observaciones de la mitad y parte superior de los rangos de los índices.

### 4.3. Consideraciones acerca de los resultados

Con el fin de dotar al lector con una herramienta de decisión entre ambas estimaciones más allá de la bondad de ajuste (ya que los  $R^2$  son muy altos debido a la composición de las variables, el Anexo D incluye una tabla con criterios de información comparativos y los resultados de las pruebas realizadas para validar los supuestos realizados sobre  $\epsilon$ ). En la sección 3.2.3. se presenta un contraste de hipótesis que permite validar si los  $\rho$  de la diagonal de  $\mathbf{P}$  deberían ser el mismo o específicos para así escoger entre ambas estimaciones.

En el caso en que la hipótesis nula de igualdad de coeficientes espaciales para las unidades transversales no pueda ser rechazada, por parsimonia, flexibilidad y limitaciones computacionales, el SARP tradicional debería ser seleccionado frente a la metodología acá propuesta. En caso contrario bajo resultados y criterios coherentes, el SARP de coeficientes específicos sería la opción a elegir.

Para la aplicación acá expuesta, la prueba de hipótesis evidenció rechazo para  $H_0$  (ver Anexo B), con lo cual se puede afirmar que el SARP tradicional sobre-dimensiona el componente de riqueza en la estimación del IDH, y que el modelo que incluye coeficientes espaciales específicos debería escogerse sobre el de coeficiente único.

Lo anterior es congruente con la disminución en la correlación que arroja la estimación de los parámetros indicativos por país y los promedios del IDH (pasa de -91 % en el modelo con coeficiente espacial único a -46 % con el modelo de coeficientes específicos), dando a entender que los parámetros espaciales explican ahora parte de la variabilidad que antes estaba inmersa en las dummies.

#### 4.3.1. Tiempos computacionales y tamaño de la muestra en $N$

Aunque los resultados anteriores<sup>5</sup> sugieren el uso del modelo de coeficientes espaciales específicos sobre el de coeficiente único, es importante advertir al lector que la experiencia realizando experimentos para esta aplicación evidenció que los costos (en tiempos computacionales) son considerablemente más altos cuando se pasa de una estructura a otra, y pueden

---

<sup>5</sup>El código en R empleado para las estimaciones presentadas puede consultarse en el Anexo E

resultar altamente variables en función del punto de partida que sea provisto al algoritmo de optimización.

Por otro lado, la razón por la cual no se trabajó con un mayor tamaño en términos de cantidad de países, fue el importante crecimiento en tiempos de optimización que implica aumentar  $N$ . Esto debido a que el algoritmo que sea seleccionado tiene que optimizar en  $N$  dimensiones, lo que es claramente más costoso y aumenta la probabilidad de dar con máximos locales.

La selección del tamaño de  $N$  no resulta trivial y finalmente dependerá de los recursos que el investigador disponga para los experimentos (una computadora más o menos potente, tiempos, entre otros).

## 5. Conclusiones y recomendaciones

### 5.1. Conclusiones

En este trabajo se presentó una metodología para el proceso de estimación de parámetros en un modelo espacial auto-regresivo para datos panel, con coeficientes espaciales específicos y variables explicativas, así como la base teórica necesaria para realizar inferencia estadística sobre tales estimaciones.

Se hace uso del método de máxima verosimilitud para encontrar las estimaciones de la estructura propuesta en el Capítulo 3, de manera que se obtienen formas cerradas para el cálculo de los parámetros asociados a las variables explicativas y la varianza del proceso, mientras que se establece una función que permite estimar los coeficientes espaciales específicos mediante el uso de métodos numéricos flexibles a optimización multivariada.

También, se realiza una aplicación del estimador sobre un panel enfocado en explicar el desarrollo humano a través del IDH y sus componentes, en una muestra intercontinental y su respectivo contraste con la propuesta previa (el SARP con coeficiente único), obteniendo una mejora en el ajuste sobre las respuestas estimadas, y revelando desventajas propias de la construcción del IDH destacando la ausencia de uniformidad en la ponderación de sus componentes, lo que se refleja en la baja importancia de la dimensión de salud y longevidad y una necesidad de evaluar la forma en que se pondera o mide la dimensión de estandar de vida. Los resultados arrojan evidencia de la presencia de efectos espaciales específicos de un país sobre los demás, es decir, que estar a menor/mayor distancia de un país con cierta medida de desarrollo humano genera alguna influencia sobre el índice de los vecinos.

Finalmente se exponen los altos costos en tiempos computacionales que resultan de optar por la propuesta acá desarrollada y su aumento en función del crecimiento asociado al tamaño de la muestra (el número de individuos incluidos en el panel), los cuales serían compensados solo en caso de suponerse evidencia de efectos específicos significativamente diferentes al coeficiente único.

## 5.2. Recomendaciones

La propuesta acá desarrollada es la base para trabajos posteriores que incluyan coeficientes específicos en los modelos espaciales, aprovechando la ampliación en grados de libertad proveniente de las estructuras de datos panel. Un contraste con avances paralelos, abre la puerta a pensar en modelos de coeficientes específicos que disminuyan los supuestos acá implícitos. Por ejemplo, la inclusión de múltiples matrices de pesos espaciales, errores con estructuras espaciales auto-regresivas, o el modelamiento de la dimensión temporal de forma estocástica, son solo algunas de las opciones que surgen.

Por último, es importante recordar que las estructuras asociadas a la especificidad en los coeficientes espaciales tienden a generar dependencia sobre potentes métodos de optimización a la hora de encontrar las estimaciones, de forma que dar atención a este aspecto también se debería convertir en foco de la discusión.

# A. Anexo: Detalles matemáticos (estimadores SARP coeficientes específicos)

## A.1. Proceso Generador de Datos

El proceso generador de datos de este modelo puede escribirse como:

$$\mathbf{Y} = [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \quad (\text{A-1})$$

Prueba:

$$\mathbf{Y} = (\mathbf{I}_T \otimes \mathbf{WP})\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{donde} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT}) \quad (\text{A-2})$$

$$\mathbf{Y} - (\mathbf{I}_T \otimes \mathbf{WP})\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Con lo que invirtiendo la matriz que premultiplica a  $\mathbf{Y}$  se llega al resultado en la ecuación (A-1).

## A.2. Función de Log-Verosimilitud

$$\begin{aligned} \ln(\mathcal{L}) = & -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln |\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})| \\ & - \frac{1}{2\sigma^2} \{([\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t ([\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\} \end{aligned} \quad (\text{A-3})$$

Prueba:

Primero encontramos la media,



$$\begin{aligned} E(\mathbf{Y}) &= E\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\} \\ E(\mathbf{Y}) &= [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1}\mathbf{X}\boldsymbol{\beta} \end{aligned} \quad (\text{A-4})$$

y luego la varianza,

$$\begin{aligned} Var(\mathbf{Y}) &= Var\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon})\} \\ &= Var\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1}\boldsymbol{\epsilon}\} \end{aligned} \quad (\text{A-5})$$

Sea  $\mathbf{B} = [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]$ , entonces  $Var(\mathbf{Y}) = \mathbf{B}^{-1}Var(\boldsymbol{\epsilon})(\mathbf{B}^{-1})^t = \sigma^2\mathbf{B}^{-1}(\mathbf{B}^{-1})^t = \boldsymbol{\Sigma}$

Teniendo en cuenta que  $\mathbf{Y} \sim N([\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1}\mathbf{X}\boldsymbol{\beta}; \boldsymbol{\Sigma})$ ,

$$\mathcal{L}(\boldsymbol{\beta}, \mathbf{P} \mid \mathbf{Y}) = \frac{1}{(2\pi)^{\frac{NT}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \quad (\text{A-6})$$

Usando las propiedades descritas en Magnus & Neudecker (1988, pág, 10):

$$\text{Sea } \mathbf{A}_{n \times n} \text{ entonces } |c\mathbf{A}| = c^n |\mathbf{A}|$$

$$\text{Sea } \mathbf{A}_{m \times m} \text{ y } \mathbf{C}_{m \times m} \text{ entonces } |\mathbf{AC}| = |\mathbf{A}||\mathbf{C}|$$

$$|\mathbf{A}^t| = |\mathbf{A}|$$

$$|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$$

El denominador de la ecuación (A-6) se puede expresar como:

$$\begin{aligned} (2\pi)^{\frac{NT}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}} &= (2\pi)^{\frac{NT}{2}} |\sigma^2\mathbf{B}^{-1}(\mathbf{B}^{-1})^t|^{\frac{1}{2}} \\ &= (2\pi)^{\frac{NT}{2}} (\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}(\mathbf{B}^{-1})^t|^{\frac{1}{2}} \\ &= (2\pi)^{\frac{NT}{2}} (\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|^{\frac{1}{2}} |(\mathbf{B}^{-1})^t|^{\frac{1}{2}} \\ &= (2\pi)^{\frac{NT}{2}} (\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}| \\ &= (2\pi\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}| \end{aligned} \quad (\text{A-7})$$

Insertándolo en la verosimilitud (A-6), se obtiene:

$$\mathcal{L}(\boldsymbol{\beta}, \mathbf{P} \mid \mathbf{Y}) = \frac{1}{(2\pi\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \quad (\text{A-8})$$

Aplicando logaritmo natural a (A-8) para trabajar con la log-verosimilitud

$$\begin{aligned}
\ln(\mathcal{L}) &= \ln \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \right\} \\
&= \ln(1) - \ln \left\{ (2\pi\sigma^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}| \right\} + \ln \left\{ \exp \left[ -\frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \right] \right\} \\
&= -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln |\mathbf{B}| - \frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln |\mathbf{B}| - \frac{1}{2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t [\sigma^2 \mathbf{B}^{-1}(\mathbf{B}^{-1})^t]^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= -\frac{NT}{2} \ln\{2\pi\sigma^2\} + \ln |\mathbf{B}| - \frac{1}{2\sigma^2}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t [\mathbf{B}^{-1}(\mathbf{B}^{-1})^t]^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})
\end{aligned} \tag{A-9}$$

Usando la siguiente propiedad se desarrolla el numerador del último sumando:

$$(\mathbf{AC})^{-1} = \mathbf{C}^{-1}\mathbf{A}^{-1}$$

y se aplica,

$$\begin{aligned}
&= (\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t [\mathbf{B}^{-1}(\mathbf{B}^{-1})^t]^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= (\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t [(\mathbf{B}^{-1})^t]^{-1}(\mathbf{B}^{-1})^{-1}(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= (\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta})^t (\mathbf{B}^t \mathbf{B})(\mathbf{Y} - \mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}) \\
&= [\mathbf{Y}^t \mathbf{B}^t - \boldsymbol{\beta}^t \mathbf{X}^t (\mathbf{B}^{-1})^t \mathbf{B}^t] [\mathbf{B}\mathbf{Y} - \mathbf{B}\mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta}] \\
&= [\mathbf{Y}^t \mathbf{B}^t - (\mathbf{X}\boldsymbol{\beta})^t (\mathbf{B}^{-1})^t \mathbf{B}^t] [\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}] \\
&= (\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t (\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})
\end{aligned} \tag{A-10}$$

Insertando en la log-verosimilitud (A-9) se obtiene el resultado de (A-3):

$$\ln(\mathcal{L}) = -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln |\mathbf{B}| - \frac{1}{2\sigma^2}(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t (\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{A-11}$$

### A.3. Maximización de la verosimilitud:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta} &= \frac{\partial \left( -\frac{NT}{2} \ln(2\pi\sigma^2) + \ln(|\mathbf{B}|) - \frac{1}{2\sigma^2} [\mathbf{BY} - \mathbf{X}\beta]^t [\mathbf{BY} - \mathbf{X}\beta] \right)}{\partial \beta} \\
&= -\frac{1}{2\sigma^2} \frac{\partial (\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]\mathbf{Y} - \mathbf{X}\beta\}^t \{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]\mathbf{Y} - \mathbf{X}\beta\})}{\partial \beta} \\
&= -\frac{1}{2\sigma^2} \frac{\partial (\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y} - \mathbf{Y}^t \mathbf{B}^t \mathbf{X} \beta - \beta^t \mathbf{X}^t \mathbf{B} \mathbf{Y} + \beta^t \mathbf{X}^t \mathbf{X} \beta)}{\partial \beta} \\
&= -\frac{1}{2\sigma^2} \left( \frac{\partial (\beta^t \mathbf{X}^t \mathbf{X} \beta)}{\partial \beta} - \frac{\partial (2\beta^t \mathbf{X}^t \mathbf{B} \mathbf{Y})}{\partial \beta} \right)
\end{aligned} \tag{A-12}$$

Usando las propiedades de las derivadas parciales descritas en Lax & Lax (2007):

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta} &= -\frac{1}{2\sigma^2} (2\mathbf{X}^t \mathbf{X} \beta - 2\mathbf{X}^t \mathbf{B} \mathbf{Y}) \\
&= \frac{1}{\sigma^2} (\mathbf{X}^t \mathbf{X} \beta - \mathbf{X}^t \mathbf{B} \mathbf{Y})
\end{aligned} \tag{A-13}$$

Se iguala a cero para maximizar:

$$\begin{aligned}
0 &= \frac{1}{\sigma^2} (\mathbf{X}^t \mathbf{X} \beta - \mathbf{X}^t \mathbf{B} \mathbf{Y}) \\
\mathbf{X}^t \mathbf{X} \beta &= \mathbf{X}^t \mathbf{B} \mathbf{Y}
\end{aligned}$$

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} \mathbf{Y} \\
\hat{\beta} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})] \mathbf{Y}
\end{aligned} \tag{A-14}$$

#### A.3.1. Insesgadez de $\hat{\beta}$

$$\begin{aligned}
E(\hat{\beta}) &= E\{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})] \mathbf{Y}\} \\
&= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})] E\{\mathbf{Y}\} \\
&= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})] E\{[\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1} (\mathbf{X}\beta + \epsilon)\} \\
&= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})] [\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{WP})]^{-1} \mathbf{X}\beta \\
&= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{X} \beta = \beta
\end{aligned} \tag{A-15}$$

### A.3.2. Estimador Máximo Verosímil de $\sigma^2$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \sigma^2} &= \frac{\partial[-\frac{NT}{2} \ln(2\pi\sigma^2)]}{\partial \sigma^2} + \frac{\partial[\ln |\mathbf{I}_{NT} - (\mathbf{I}_T \otimes \mathbf{W}\mathbf{P})|]}{\partial \sigma^2} + \frac{\partial[\frac{1}{2\sigma^2}(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})]}{\partial \sigma^2} \\
&= \frac{\partial[-\frac{NT}{2}(\ln(2\pi) + \ln(\sigma^2))]}{\partial \sigma^2} + \frac{\partial[-\frac{1}{2}(\sigma^2)^{-1}(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})]}{\partial \sigma^2} \\
&= \frac{\partial[-\frac{NT}{2} \ln(\sigma^2)]}{\partial \sigma^2} - \frac{1}{2} \frac{\partial[(\sigma^2)^{-1}[(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})]]}{\partial \sigma^2} \\
&= -\frac{NT}{2} \frac{1}{\sigma^2} + \frac{1}{2} \frac{1}{(\sigma^2)^2} [(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{B}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})] \\
\hat{\sigma}_{MV}^2 &= \frac{(\mathbf{B}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^t(\mathbf{B}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{NT}
\end{aligned} \tag{A-16}$$

### A.3.3. Sesgo y corrección $\hat{\sigma}_{MV}^2$

Primero se encuentra una forma idempotente para escribir el numerador de la expresión de  $\hat{\sigma}_{MV}^2$ . Se inicia reescribiendo el numerador del estimador máximo-verosímil:

$$l_1 = (\mathbf{B}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^t(\mathbf{B}\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Y}^t\mathbf{B}^t\mathbf{B}\mathbf{Y} - \mathbf{Y}^t\mathbf{B}^t\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{B}\mathbf{Y} + \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{X}\hat{\boldsymbol{\beta}}$$

Desarrollando el último sumando:

$$\hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{B}\mathbf{Y} = \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{B}\mathbf{Y}$$

A continuación vemos como esta expresión también puede ser escrita en términos de  $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ ,

$$\begin{aligned}
l_1 &= \mathbf{Y}^t\mathbf{B}^t\mathbf{B}\mathbf{Y} - \mathbf{Y}^t\mathbf{B}^t\mathbf{X}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{B}\mathbf{Y} + \hat{\boldsymbol{\beta}}^t\mathbf{X}^t\mathbf{B}\mathbf{Y} \\
&= \mathbf{Y}^t\mathbf{B}^t\mathbf{B}\mathbf{Y} - \mathbf{Y}^t\mathbf{B}^t\mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{Y}^t\mathbf{B}^t\mathbf{B}\mathbf{Y} - \mathbf{Y}^t\mathbf{B}^t\mathbf{H}\mathbf{B}\mathbf{Y}
\end{aligned} \tag{A-17}$$

Y re-expresarse como:

$$l_1 = (\mathbf{B}\mathbf{Y})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{B}\mathbf{Y}$$

Ahora como  $(\mathbf{I}_{NT} - \mathbf{H})$  es idempotente entonces la forma cuadrática:

$$\frac{(\mathbf{B}\mathbf{Y})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{B}\mathbf{Y}}{\sigma^2} \sim \chi_{(NT-k)}^2 \tag{A-18}$$

Prueba:

$$E(\mathbf{BY}) = \mathbf{B}E(\mathbf{Y}) = \mathbf{B}\mathbf{B}^{-1}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}$$

$$\text{Var}(\mathbf{BY}) = \mathbf{B}\text{Var}(\mathbf{Y})\mathbf{B}^t - \mathbf{B}\sigma^2\mathbf{B}^{-1}(\mathbf{B}^{-1})^t\mathbf{B}^t = \sigma^2\mathbf{I}_{NT}$$

$$\mathbf{BY} \sim \mathbf{N}_{NT}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_{NT})$$

Ahora, resulta idempotente el producto:

$$\left[ \frac{(\mathbf{I}_{NT} - \mathbf{H})\sigma^2\mathbf{I}_{NT}}{\sigma^2} \right] \left[ \frac{(\mathbf{I}_{NT} - \mathbf{H})\sigma^2\mathbf{I}_{NT}}{\sigma^2} \right] = (\mathbf{I}_{NT} - \mathbf{H})(\mathbf{I}_{NT} - \mathbf{H}) = (\mathbf{I}_{NT} - \mathbf{H})$$

Por ende:

$$\frac{(\mathbf{BY})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{BY}}{\sigma^2} \sim \chi^2_{[NT-k, \frac{1}{2}(\mathbf{X}\boldsymbol{\beta})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{X}\boldsymbol{\beta}]}$$

Si desarrollamos el parámetro de no centralidad:

$$\frac{1}{2}(\mathbf{X}\boldsymbol{\beta})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{X}\boldsymbol{\beta} = (\mathbf{X}\boldsymbol{\beta})^t[\mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\boldsymbol{\beta}] = (\mathbf{X}\boldsymbol{\beta})^t(\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

Lo que demuestra que (A-18):

$$\frac{(\mathbf{BY})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{BY}}{\sigma^2} \sim \chi^2_{(NT-k)}$$

Usando tal resultado:

$$E(\hat{\sigma}^2) = \frac{1}{NT}E\left[\frac{(\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})^t(\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})\sigma^2}{\sigma^2}\right] = \frac{\sigma^2}{NT}E\left[\frac{(\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})^t(\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})}{\sigma^2}\right]$$

Como el valor esperado de una variable aleatoria Chi-cuadrado son sus grados de libertad, y:

$E(\hat{\sigma}^2) = \frac{1}{NT}\sigma^2(NT - k)$ , así se verifica que  $\hat{\sigma}_{MV}^2$  es un estimador sesgado de la varianza, y su versión insesgada está dada por:

$$\hat{\sigma}_{ins}^2 = \frac{1}{NT - k}(\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})^t(\mathbf{BY} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

### A.3.4. Distribución de $\hat{\boldsymbol{\beta}}$

Ya se demostró que  $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ , y para la varianza:

$$\begin{aligned}
Var(\hat{\beta}) &= Var[(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} \mathbf{Y}] \\
&= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} Var(\mathbf{Y}) [(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B}]^t \\
&= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} (\sigma^2 \mathbf{I}) \mathbf{B}^t \mathbf{X} [(\mathbf{X}^t \mathbf{X})^{-1}]^t \\
&= \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} \mathbf{B}^t \mathbf{X} [(\mathbf{X}^t \mathbf{X})^{-1}]^t \\
&= \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} \mathbf{B}^t \mathbf{X} [(\mathbf{X}^t \mathbf{X})^t]^{-1} \\
&= \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} \mathbf{B}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1}
\end{aligned} \tag{A-19}$$

De lo que se puede concluir:

$$\hat{\beta} \sim N(\beta, \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{B} \mathbf{B}^t \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1})$$

### Pruebas $t$ -student para evaluar la significancia de las variables explicativas

$H_0 : \beta_j = 0$ , i.e. la  $j$ -ésima variable explicativa no influye significativamente sobre  $\mathbf{Y}$

$H_1 : \beta_j \neq 0$ , i.e. la  $j$ -ésima variable explicativa si influye significativamente sobre  $\mathbf{Y}$

$$t_j = \frac{\hat{\beta}_j}{std(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \sim t_{(NT-K-1)}, \text{ para } j = 1, \dots, K$$

Valor  $p < \alpha \%$  : se rechaza  $H_0 : \beta_j = 0$  y se concluye que la variable  $j$ -ésima es significativa.

### Razón de Verosimilitud para evaluar $H_0 : \beta = 0$

$$\lambda = \frac{\sup_{H_0: \beta=0} \mathcal{L}(\mathbf{Y}, \beta, \mathbf{P}, \sigma^2)}{\sup_{\Omega} \mathcal{L}(\mathbf{Y}, \beta, \mathbf{P}, \sigma^2)} \tag{A-20}$$

Con el estimador máximo-verosímil de  $\sigma^2$ :

$$\hat{\sigma}_{MV}^2 = \frac{(\mathbf{B}\mathbf{Y} - \mathbf{X}\hat{\beta})^t (\mathbf{B}\mathbf{Y} - \mathbf{X}\hat{\beta})}{NT}$$

y reemplazando la condición de  $\beta$ ,

$$\hat{\sigma}_{H_0}^2 = \frac{(\mathbf{B}\mathbf{Y})^t (\mathbf{B}\mathbf{Y})}{NT} = \frac{\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y}}{NT}$$

La función de verosimilitud bajo  $H_0 : \beta = 0$  se expresa como:

$$\mathcal{L}_{H_0} = \frac{1}{(2\pi\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left\{ -\frac{1}{2}(\mathbf{Y}^t \boldsymbol{\Sigma}^{-1} \mathbf{Y}) \right\}$$

y desarrollando el numerador,

$$(\mathbf{Y}^t \boldsymbol{\Sigma}^{-1} \mathbf{Y}) = \mathbf{Y}^t [\hat{\sigma}_{H_0}^2 \mathbf{B}^{-1} (\mathbf{B}^{-1})^t]^{-1} \mathbf{Y} = \frac{\mathbf{Y}^t [\mathbf{B}^{-1} (\mathbf{B}^{-1})^t]^{-1} \mathbf{Y}}{\hat{\sigma}_{H_0}^2} = \frac{\mathbf{Y}^t [(\mathbf{B}^{-1})^t] (\mathbf{B}^{-1})^{-1} \mathbf{Y}}{\hat{\sigma}_{H_0}^2} = \frac{\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y}}{\hat{\sigma}_{H_0}^2}$$

$$\mathcal{L}_{H_0} = \frac{1}{(2\pi\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left\{ -\frac{1}{2\hat{\sigma}_{H_0}^2} (\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y}) \right\}$$

Reemplazando  $\hat{\sigma}_{H_0}^2$  y  $\hat{\sigma}^2$  dentro del exponente, en A-20, se obtiene:

$$\begin{aligned} \lambda &= \frac{\frac{1}{(2\pi\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left\{ -\frac{1}{2\hat{\sigma}_{H_0}^2} (\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y}) \right\}}{\frac{1}{(2\pi\hat{\sigma}^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left\{ -\frac{1}{2\hat{\sigma}^2} (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})^t (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta}) \right\}} \\ \lambda &= \frac{\frac{1}{(2\pi\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left\{ \frac{NT(\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y})}{2(\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y})} \right\}}{\frac{1}{(2\pi\hat{\sigma}^2)^{\frac{NT}{2}} |\mathbf{B}^{-1}|} \exp \left\{ \frac{NT(\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})^t (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})}{2(\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})^t (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})} \right\}} \\ \lambda &= \frac{\frac{1}{(2\pi\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}}} \exp \left\{ \frac{NT}{2} \right\}}{\frac{1}{(2\pi\hat{\sigma}^2)^{\frac{NT}{2}}} \exp \left\{ \frac{NT}{2} \right\}} = \frac{(2\pi\hat{\sigma}^2)^{\frac{NT}{2}}}{(2\pi\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}}} = \frac{(\hat{\sigma}^2)^{\frac{NT}{2}}}{(\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}}} = \frac{(\hat{\sigma}^2)^{\frac{NT}{2}}}{(\hat{\sigma}_{H_0}^2)^{\frac{NT}{2}}} = \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_{H_0}^2} \right)^{\frac{NT}{2}} \\ \lambda_{\frac{2}{NT}} &= \frac{\hat{\sigma}^2}{\hat{\sigma}_{H_0}^2} = \frac{\frac{(\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})^t (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})}{NT}}{\frac{\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y}}{NT}} = \frac{(\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})^t (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})}{\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y}} \end{aligned}$$

Como ya se mostró  $(\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta})^t (\mathbf{B} \mathbf{Y} - \mathbf{X} \hat{\beta}) = (\mathbf{B} \mathbf{Y})^t (\mathbf{I}_{NT} - \mathbf{H}) \mathbf{B} \mathbf{Y}$ , además que:

$$\mathbf{Y}^t \mathbf{B}^t \mathbf{B} \mathbf{Y} = \mathbf{Y}^t \mathbf{B}^t \mathbf{H} \mathbf{B} \mathbf{Y} + \mathbf{Y}^t \mathbf{B}^t [\mathbf{I}_{NT} - \mathbf{H}] \mathbf{B} \mathbf{Y}$$

entonces:

$$\lambda_{\frac{2}{NT}} = \frac{(\mathbf{B} \mathbf{Y})^t (\mathbf{I}_{NT} \mathbf{H}) (\mathbf{B} \mathbf{Y})}{\mathbf{Y}^t \mathbf{B}^t \mathbf{H} \mathbf{B} \mathbf{Y} + \mathbf{Y}^t \mathbf{B}^t [\mathbf{I}_{NT} - \mathbf{H}] \mathbf{B} \mathbf{Y}}$$

re-expresando:

$$\lambda_{\frac{2}{NT}} = \frac{1}{1 + \frac{\mathbf{Y}^t \mathbf{B}^t \mathbf{H} \mathbf{B} \mathbf{Y}}{(\mathbf{B} \mathbf{Y})^t [\mathbf{I}_{NT} - \mathbf{H}] \mathbf{B} \mathbf{Y}}}$$

(A-21)

Si  $\lambda^{\frac{2}{NT}}$  es pequeña (cercana a cero), es porque  $H_0 : \boldsymbol{\beta} = \mathbf{0}$  debe ser rechazada, si por otro lado  $\lambda$  es cercana a 1,  $H_0$  no debe rechazarse.

De lo anterior y como ya se ha mostrado

$$\frac{(\mathbf{BY})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{BY}}{\sigma^2} \sim \chi_{(NT-K)}^2$$

y  $(\mathbf{BY})^t\mathbf{H}(\mathbf{BY})$  es una forma cuadrática donde  $\mathbf{H}$  es idempotente y de rango  $K$ . Luego,  $\frac{(\mathbf{BY})^t\mathbf{H}(\mathbf{BY})}{\sigma^2} \sim \chi_{(K)}^2$ .

Por lo tanto,

$$F = \frac{(NT - k)(\mathbf{BY})^t\mathbf{H}(\mathbf{BY})}{k(\mathbf{BY})^t(\mathbf{I}_{NT} - \mathbf{H})(\mathbf{BY})} \sim F_{(k, NT-K)}$$

De los resultados anteriores es importante notar que  $(\mathbf{BY})^t(\mathbf{BY}) = (\mathbf{BY})^t\mathbf{H}(\mathbf{BY}) + (\mathbf{BY})^t(\mathbf{I}_{NT} - \mathbf{H})\mathbf{BY}$ , lo que equivale a escribir  $SCT = SCM + SCE$ , en el mismo orden, donde  $SCT$  es la suma de cuadrados total, mientras que  $SCM$  y  $SCE$ , son las sumas de cuadrados del modelo y del error, respectivamente.

## Bondad de ajuste

$$R^2 = \frac{SCM_{CM}}{SCT_{CM}} = \frac{(\mathbf{BY})^t(\mathbf{H} - \frac{1}{NT}\mathbf{J}_{NT})(\mathbf{BY})}{(\mathbf{BY})^t(\mathbf{I}_{NT} - \frac{1}{NT}\mathbf{J}_{NT})(\mathbf{BY})}$$

$$R_{Adj}^2 = R^2 - \frac{k}{NT - (k + 1)}(1 - R^2)$$

Aquí  $\mathbf{J}$  es una matriz cuadrada de unos, de la dimensión indicada. Por otro lado,  $SCT_{CM}$  es la suma de cuadrados total corregida por la media, mientras que  $SCM_{CM}$  es la suma de cuadrados del modelo corregida por la media.



## B. Anexo: Índice de Desarrollo Humano, definiciones y cálculos

### B.1. Índice de Desarrollo Humano

Según lo definido en las Notas Técnicas del IDH<sup>1</sup>, esta es una medida de resumen de logros en tres dimensiones clave del desarrollo humano: una vida larga y saludable, acceso al conocimiento y un estándar decente de vida.

El cálculo no resulta en una medida lineal, ya que resulta ser la media geométrica de los índices que representan las dimensiones:

$$IDH = (I_{Salud} * I_{Educacion} * I_{Ingresos})^{\frac{1}{3}}$$

donde cada uno de los índices se construye como:

$$I_t = \frac{Indicador_t - \min(Indicador)}{\max(Indicador) - \min(Indicador)}$$

Los indicadores que se mencionan en la razón anterior, son la esperanza de vida para la dimensión de vida larga y saludable, la media aritmética entre la esperanza de años de escolaridad y la media de los años de escolaridad para la dimensión de acceso al conocimiento; y el Ingreso Interno Bruto (IIB) per cápita a precios de paridad adquisitiva para la dimensión de estándar decente de vida. Para este último índice es importante resaltar que se usa el logaritmo natural del IIB per cápita, puesto que suaviza la dimensión de la medida.

Ejemplos acerca del cálculo del índice y otros detalles pueden ser consultados en las notas técnicas del Programa de Desarrollo de las Naciones Unidas en su página web<sup>2</sup>.

---

<sup>1</sup>[http://hdr.undp.org/sites/default/files/hdr2018\\_technical\\_notes.pdf](http://hdr.undp.org/sites/default/files/hdr2018_technical_notes.pdf)

<sup>2</sup><http://hdr.undp.org/en/content/human-development-index-hdi>

## B.2. Prueba de Moran de correlación espacial al IDH

Los resultados de las pruebas de Moran para cada uno de los periodos del panel, con su respectivo estadístico y valores p se muestran a continuación (importante recordar que  $H_0$  : No hay auto-correlación espacial):

Para este caso, dado que  $N$  no cambia en el tiempo, se obtuvo que  $E(I) = -0,091$  y que  $Var(I) = 0,008$ .

## B.3. Resultados contraste de hipótesis coeficiente espacial único vs. específicos

El uso de la prueba diseñada en la sección 3.2.3 para evaluar:

$$H_0 : \rho = \rho_i$$

para  $i = 1, \dots, N$ .

Dentro de la aplicación que se trabajó arroja los siguientes resultados:

Por lo cual se puede concluir que existe evidencia estadística suficiente para rechazar  $H_0$  y por tanto, al menos un coeficiente espacial específico es estadísticamente diferente de los demás, de esta forma, un análisis a los coeficientes asociados a las variables explicativas debería hacerse bajo los resultados arrojados por el SAR Panel con coeficientes espaciales específicos sobre el de coeficiente único.

**Tabla B-1.:** Resultados de prueba de auto-correlación de Moran.

<b>Año</b>	<b>I</b>	<b>z</b>	<b>Valor-p</b>
1990	0,127	26,364	0,000
1991	0,123	25,849	0,000
1992	0,121	25,629	0,000
1993	0,117	25,196	0,000
1994	0,113	24,638	0,000
1995	0,105	23,754	0,000
1996	0,102	23,361	0,000
1997	0,101	23,194	0,000
1998	0,112	24,567	0,000
1999	0,106	23,807	0,000
2000	0,106	23,825	0,000
2001	0,107	23,950	0,000
2002	0,095	22,531	0,000
2003	0,100	23,103	0,000
2004	0,100	23,063	0,000
2005	0,098	22,870	0,000
2006	0,104	23,555	0,000
2007	0,110	24,327	0,000
2008	0,104	23,545	0,000
2009	0,112	24,585	0,000
2010	0,106	23,859	0,000
2011	0,104	23,608	0,000
2012	0,107	23,989	0,000
2013	0,099	22,978	0,000
2014	0,095	22,488	0,000
2015	0,099	23,025	0,000
2016	0,103	23,405	0,000
2017	0,102	23,401	0,000

**Tabla B-2.:** Resultados  $H_0 : \rho = \rho_i$ .

Componente	Resultado
$N$	12
$v$	35,51
$T^2$ de Hoteling	17,38
Estadística $F$	144
Valor-p	0,000

## C. Anexo: Criterios de información y resultados pruebas a los residuales

La tabla [C-1](#) muestra los resultados de los criterios de información que contrastan el modelo con y sin coeficientes específicos. La mejora es cercana en los tres casos al 3,7 %.

**Tabla C-1.:** Criterios de Información (CI) modelos.

CI	Coef. específicos	Coef. único
AIC	-14.268	-13.775
BIC	-14.420	-13.882
HQIC	-14.328	-13.818

Dado que la propuesta no considera el tiempo como una dimensión estocástica, se toma una prueba propuesta en Wooldridge (2010, pág. 176), con el fin de descartar que algún patrón o correlación proveniente del tiempo haya quedado omitida. Se toma el modelo:

$$\hat{\epsilon}_{it} = \phi \hat{\epsilon}_{i,t-1} + \mu_{it} \quad (\text{C-1})$$

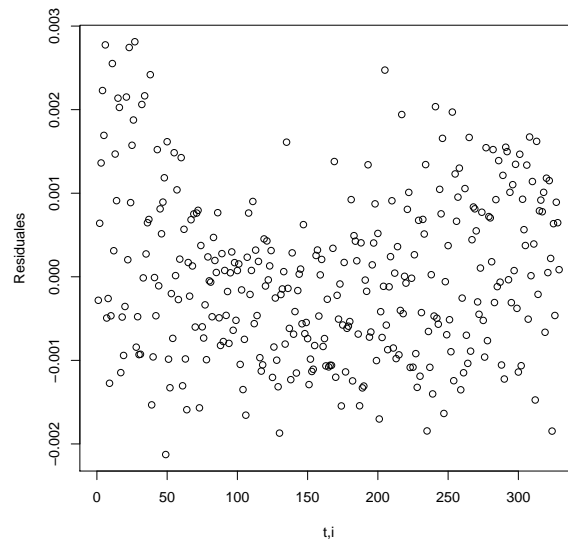
donde los  $\hat{\epsilon}_{it}$  corresponden al residual estimado del modelo espacial correspondiente. La prueba consiste en evaluar la significancia de  $\phi$  mediante una regresión ordinaria. Los resultados para los modelos de la aplicación se muestran en las tablas [C-2](#) y [C-3](#), e indican que no hay evidencia de alguna correlación temporal en los residuos.

**Tabla C-2.:** Prueba de auto-correlación temporal (Coef. específicos).

Parámetro ( $\phi$ )	t	Error Est.	Valor p
0,0714	0,054	1,314	0,19
$R^2 = 0,002$	$F = 1,725$		Valor p = 0,19

**Tabla C-3.:** Prueba de auto-correlación temporal (Coef. único)

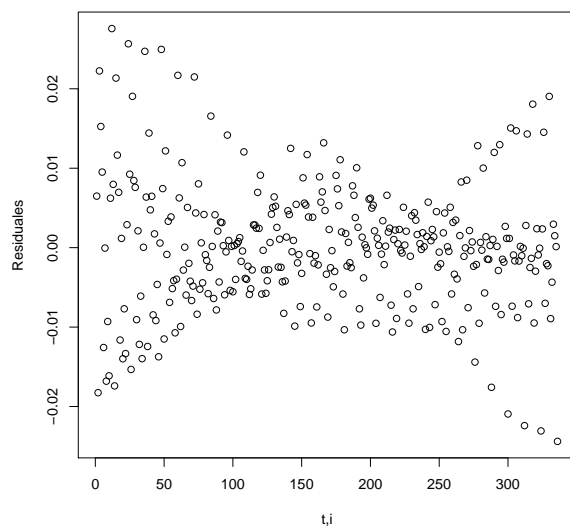
Parámetro ( $\phi$ )	t	Error Est.	Valor p
-0,020	-0,356	0,056	0.722
$R^2 = 0,000$	$F = 0,1266$	Valor p = 0,722	

**Figura C-1.:** Residuales especificación con coeficientes espaciales específicos.

Para evaluar presencia de heterocedasticidad, se hizo uso de la prueba planteada en Goldfeld & Quandt (1965), que consiste en ordenar los residuos de menor a mayor, para luego evaluar que la varianza de la primera mitad sea igual a la de la segunda haciendo uso de una prueba  $F$ .

Los resultados se pueden visualizar en las tablas [C-4](#) y [C-5](#), e indican a un 95 % de confianza homogeneidad en varianza del modelo de coeficientes específicos, lo que no ocurre en el de coeficiente único. Lo anterior también se puede intuir de las gráficas [C-1](#) y [C-2](#), donde se puede apreciar que los residuos del modelo de coeficientes únicos no presenta ningún patrón, mientras que el de coeficiente único, deja ver una forma de reloj de arena, una señal de presencia de heterocedasticidad.

Finalmente, con el fin de validar normalidad en los residuos, se realizan dos pruebas multi-



**Figura C-2.:** Residuales especificación con coeficiente espacial único.

**Tabla C-4.:** Prueba de Goldfeld Quandt para residuales del modelo con coeficientes específicos

Estadístico F	G. Lib. Num	G. Lib. Den	Valor p
0,800	167	167	0,07
Ratio estimado	IC 95 %		
0,800	0,58	1,03	

**Tabla C-5.:** Prueba de Goldfeld Quandt para residuales del modelo con coeficiente único

Estadístico F	G. Lib. Num	G. Lib. Den	Valor p
0,715	167	167	0,03
Ratio estimado	IC 95 %		
0,715	0,53	0,97	

variadas de normalidad para los  $N$  vectores aleatorios que se tienen (acordes a las unidades transversales). Las pruebas seleccionadas fueron las de Henze & Zirkler (1990) y la de Doornik & Hansen (2008), las cuales mejoran en términos de potencia frente a las propuestas más clásicas de Mardia (1970) y Baringhaus, Danschke & Henze (1989).

También se presentan los Q-Q plot para los residuos de cada unidad transversal de los dos modelos evaluados en las figuras [C-3](#) y [C-4](#).

Los resultados de las pruebas de Henze & Zirkler (1990) y Doornik & Hansen (2008) indicaron en ambos casos para el modelo de coeficientes específicos, normalidad en los residuales, mientras que sólo la de Doornik & Hansen (2008) evidencia normalidad en el caso del coeficiente único. Los resultados pueden observarse en la tabla [C-6](#) y [C-7](#).

**Tabla C-6.:** Pruebas de normalidad a los residuos, coeficientes específicos.

Prueba	Estadístico	valor p	Normalidad
<b>Henze-Zirkler</b>	0,981	0,303	<b>Si</b>
<b>Doornik-Hansen</b>	21,611	0,602	<b>Si</b>

**Tabla C-7.:** Pruebas de normalidad a los residuos, coeficiente único.

Prueba	Estadístico	valor p	Normalidad
<b>Henze-Zirkler</b>	1,012	0,000	<b>No</b>
<b>Doornik-Hansen</b>	17,500	0,826	<b>Si</b>



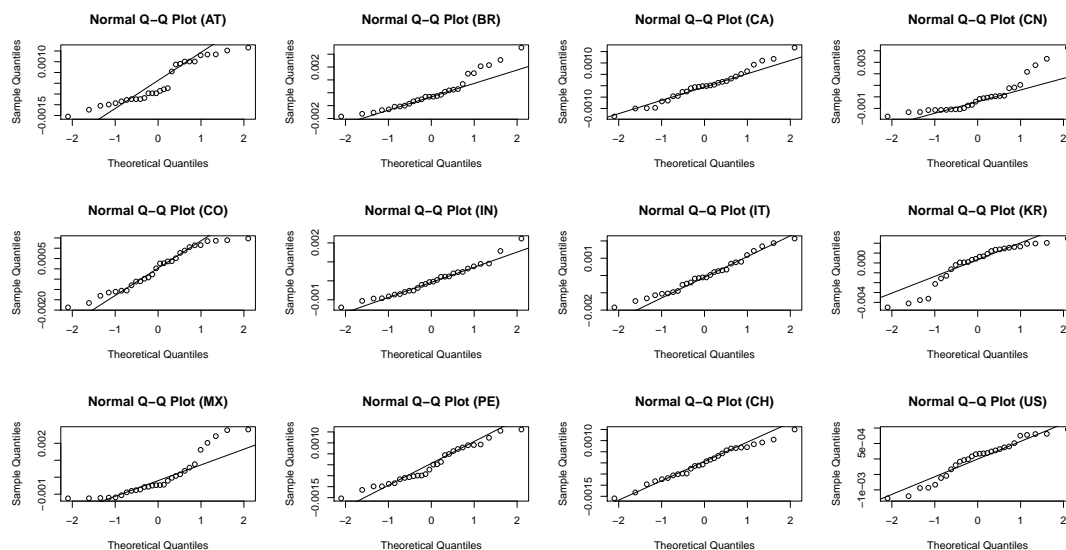


Figura C-3.: Q-Q Plot residuales modelo coeficientes espaciales específicos

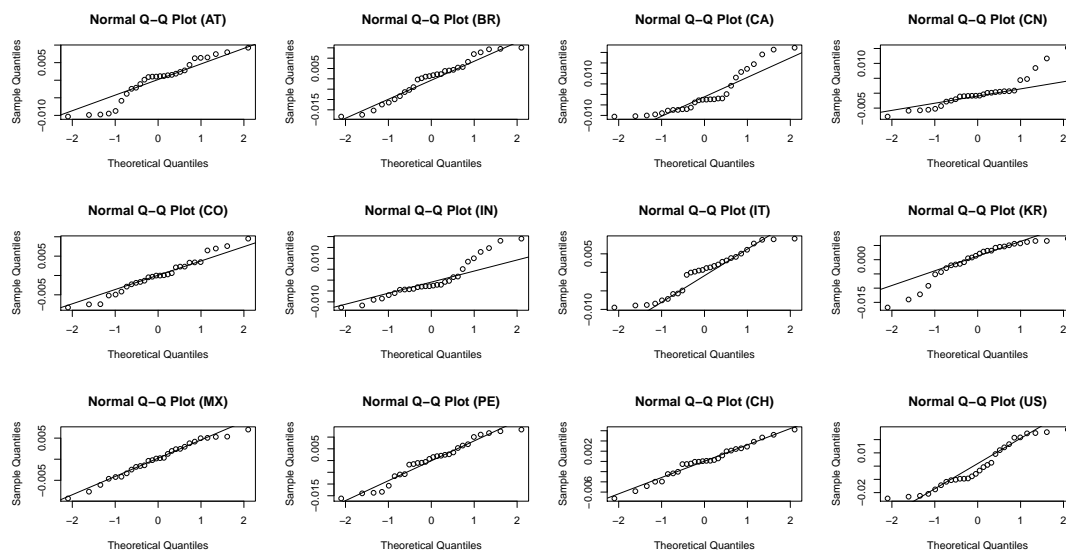


Figura C-4.: Q-Q Plot residuales modelo coeficiente espacial único

# D. Anexo: Matrices de pesos espaciales

La matriz de pesos espaciales inicial con las distancias entre los centroides de los países en kilómetros se presenta a continuación en la tabla [D-1](#).

Tabla D-1.: Matriz inicial de pesos espaciales en km												
	AT	BR	CA	CN	CO	IN	IT	KR	MX	PE	CH	US
AT	0	9498	7253	7137	9546	6430	647	8615	9920	10727	484	8344
BR	9498	0	9305	16633	3234	14775	9070	17605	6928	2574	9043	7317
CA	7253	9305	0	9386	6396	11469	7705	8579	3627	7816	7056	2262
CN	7137	16633	9386	0	15516	2984	7566	2118	12826	17048	7596	11647
CO	9546	3234	6396	15516	0	15972	9409	14961	3697	1532	9069	4222
IN	6430	14775	11469	2984	15972	0	6569	5017	15094	16951	6906	13577
IT	647	9070	7705	7566	9409	6569	0	9147	10148	10490	649	8619
KR	8615	17605	8579	2118	14961	5017	9147	0	11543	16255	9020	10743
MX	9920	6928	3627	12826	3697	15094	10148	11543	0	4720	9562	1633
PE	10727	2574	7816	17048	1532	16951	10490	16255	4720	0	10243	5580
CH	484	9043	7056	7596	9069	6906	649	9020	9562	10243	0	8011
US	8344	7317	2262	11647	4222	13577	8619	10743	1633	5580	8011	0

Luego de las transformaciones adecuadas, la matriz de pesos espaciales estandarizada que se usa en las estimaciones resulta en la tabla [D-2](#).

Finalmente, la tabla de homologación de los códigos asociados de los países se encuentra documentado en la tabla [D-3](#).

**Tabla D-2.:** Matriz de pesos espaciales transformada

	AT	BR	CA	CN	CO	IN	IT	KR	MX	PE	CH	US
AT	0,00	0,02	0,03	0,03	0,02	0,03	0,33	0,02	0,02	0,02	0,44	0,03
BR	0,07	0,00	0,07	0,04	0,19	0,04	0,07	0,04	0,09	0,24	0,07	0,09
CA	0,08	0,06	0,00	0,06	0,09	0,05	0,07	0,06	0,15	0,07	0,08	0,24
CN	0,08	0,04	0,06	0,00	0,04	0,20	0,08	0,28	0,05	0,04	0,08	0,05
CO	0,05	0,14	0,07	0,03	0,00	0,03	0,05	0,03	0,13	0,30	0,05	0,11
IN	0,11	0,05	0,06	0,24	0,04	0,00	0,11	0,14	0,05	0,04	0,10	0,05
IT	0,37	0,03	0,03	0,03	0,03	0,04	0,00	0,03	0,02	0,02	0,37	0,03
KR	0,08	0,04	0,08	0,32	0,04	0,13	0,07	0,00	0,06	0,04	0,07	0,06
MX	0,05	0,07	0,13	0,04	0,13	0,03	0,05	0,04	0,00	0,10	0,05	0,30
PE	0,05	0,19	0,06	0,03	0,32	0,03	0,05	0,03	0,10	0,00	0,05	0,09
CH	0,44	0,02	0,03	0,03	0,02	0,03	0,33	0,02	0,02	0,02	0,00	0,03
US	0,05	0,06	0,20	0,04	0,11	0,03	0,05	0,04	0,28	0,08	0,06	0,00

**Tabla D-3.:** Homologación código países

País	Código
Austria	AT
Brasil	BR
Canadá	CA
China	CN
Colombia	CO
India	IN
Italia	IT
Corea del Sur	KR
México	MX
Perú	PE
Suiza	CH
USA	US

## E. Anexo: Código empleado en la aplicación

```
#####  
### Aplicacion estimador SARP con coeficientes espaciales especificos ###  
#####
```

```
#####  
### David Fernando Delgado Contreras ###  
### Universidad Nacional de Colombia ###  
### Estudiante M. Sc. Estadística #####  
### Marzo 2019 #####  
#####
```

```
# Limpiar ambiente  
rm(list=ls())
```

```
# Establecer directorio de trabajo  
setwd("C:/Users/DavidD/Videos/Dropbox UNAL/Dropbox/  
Tesis Spatial Panel/Experiments/R")
```

```
# Cargar panel y matriz de pesos espaciales
```

```
Data_Panel = read.csv(file="Panel_HDI_EI_dummy.csv", header=TRUE, sep = ",")  
# Distance weights matrix  
Data_W = read.csv(file="W_HDI.csv", header=FALSE, sep = ",")
```

```
# Fijando parametros generales
```

```
Years <- as.matrix(Data_Panel[3])
Years

Panel <- as.matrix(Data_Panel[5:20])
Panel
T <- max(Years)
T

N <- dim(Panel)[1]/T
N

NT = dim(Panel)[1]

# Extraer matrices Y, X, y W, y fijando k

W <- as.matrix(Data_W)
W

X <- as.matrix(Data_Panel[6:20])
Y <- as.matrix(Data_Panel[5])

k <- dim(X)[2]
k

H <- X%*%solve(t(X)%*%X)%*%t(X)

# Calculando rango para rho

eigen_W <- eigen(W)
eigen_W

max_rho <- max(eigen_W[[1]])
max_rho
```

---

```

min_rho <- min(eigen_W[[1]])
min_rho

### Optimizacion bajo metodologia tradicional (rho unico)

# Establecer funcion objetivo para la optimizacion

# -f(P) (ya que el algoritmo minimiza)
neg_lklh_func <- function(rho) {
  Y=Y
  H=H
  W=W
  P_diag = rho*diag(N)
  LL = (NT/2)*log(t((diag(NT)-H)%*%(diag(NT)-(diag(T)%x%P_diag)
  %*%(diag(T)%x%W))%*%Y)%*%((diag(NT)-H)%*%(diag(NT)
  -(diag(T)%x%P_diag)*(diag(T)%x%W))%*%Y))
  - log(det((diag(NT)-(diag(T)%x%P_diag)%*%(diag(T)%x%W))))
}

# Optimizacion usando rutinas PORT

P_diag_nlminb = nlminb(rep(0, 1), neg_lklh_func)
rho_1_nlminb <- round(P_diag_nlminb[[1]], 3)
P_nlminb_diag <- round(P_diag_nlminb[[1]], 3)
P_nlminb <- P_nlminb_diag*diag(N)
P_nlminb

# Hesiano (y matriz de informacion observada de Fisher)

#install.packages('pracma')
library(pracma)

Hess <- hessian(neg_lklh_func, P_nlminb_diag)

```

---

```

Hess_1_rho <- Hess
Hess

# Matriz de covarianza de la optimizacion

Var_P <- solve(-Hess)
Var_1_rho <- Var_P
Var_P

# Betas

B <- diag(NT)-diag(T)%x%(W%*%P_nlminb)

beta <- solve(t(X)%*%X)%*%t(X)%*%B%*%Y
beta_1_rho <- beta
beta

beta_0 = solve(t(X)%*%X)%*%t(X)%*%Y
beta_d = solve(t(X)%*%X)%*%t(X)%*%(diag(T)%x%W)%*%Y

beta_0
beta_d

# Sigma_2

sigma2_MV <- t(B%*%Y-X%*%beta)%*%(B%*%Y-X%*%beta)/NT
sigma2_MV_1_rho <- sigma2_MV
sigma2_MV

sigma_2_unb <- t(B%*%Y-X%*%beta)%*%(B%*%Y-X%*%beta)/(NT-dim(beta)[1])
sigma_2_unb_1_rho <- sigma_2_unb
sigma_2_unb

# Varianza de los betas

```

---

```

Var_beta <- sigma2_MV[1]*solve(t(X)%*%X)
      %*%t(X)%*%B%*%t(B)%*%X%*%solve(t(X)%*%X)
Var_beta_1_rho <- Var_beta
Var_beta

# Valores p

vector_t_betas <- matrix(, nrow = 0, ncol = 1)

for (n in 1:k) {
  vector_t_betas <- rbind(vector_t_betas,
    beta[n,1]/sqrt(Var_beta[n,n]))
}

vector_t_betas

p_values_betas <- matrix(, nrow = 0, ncol = 1)

for (n in 1:k) {
  p_values_betas <- rbind(p_values_betas,
    2*dt(vector_t_betas[n,1], NT-k-1))
}

p_values_betas

# Valores p rho

vector_z_rhos <- matrix(, nrow = 0, ncol = 1)

for (n in 1:1) {
  vector_z_rhos <- rbind(vector_z_rhos,
    P_nlminb[n,n]/sqrt(Var_P[n,n]))
}

```



```
vector_z_rhos

p_values_rhos <- matrix(, nrow = 0, ncol = 1)

for (n in 1:1) {
  p_values_rhos <- rbind(p_values_rhos,
    2*dnorm(vector_z_rhos[n,1], mean = 0, sd = 1))
}

p_values_rhos

## Intervalos de confianza

## De los betas

t_statistic_betas <- qt(0.975, df = NT-k-1)
band_betas <- matrix(, nrow = 0, ncol = 1)

for (n in 1:k) {
  band_betas <- rbind(band_betas, t_statistic_betas*sqrt(Var_beta[n,n]))
}

band_betas

vector_IC_betas <- matrix(, nrow = 0, ncol = 3)

for (n in 1:k) {
  vector_IC_betas <- rbind(vector_IC_betas,
    c(beta[n,1] - band_betas[n,1], beta[n,1], beta[n,1] + band_betas[n,1]))
}

vector_IC_betas

## De rho

z_statistic_rhos <- qnorm(0.975, mean = 0, sd = 1)
```

---

```

band_rhos <- matrix(, nrow = 0, ncol = 1)

for (n in 1:1) {
  band_rhos <- rbind(band_rhos, z_statistic_rhos*sqrt(Var_P[n,n]))
}

band_rhos

vector_IC_rhos <- matrix(, nrow = 0, ncol = 3)

for (n in 1:1) {
  vector_IC_rhos <- rbind(vector_IC_rhos, c(P_nlminb[n,n]
    - band_rhos[n,1], P_nlminb[n,n], P_nlminb[n,n] + band_rhos[n,1]))
}

vector_IC_rhos

## Prueba F de significancia global

Fstat = ((NT-k)*(t(B%*%Y)%*%H%*%B%*%Y))/(k*t(B%*%Y)%*%
  (diag(NT)-H)%*%B%*%Y)
Fstat
p_val <- 1 - pf(Fstat, df1 = k, df2 = NT-k)
p_val

## Prediccion

Y_gorro <- (diag(T)%x%(W%*%P_nlminb))%*%Y + X%*%beta
Y_gorro_1_rho <- Y_gorro
Y_gorro

resid <- Y_gorro - Y
resid_1_rho <- resid
resid

```

```
Y_gorro2 <- solve(diag(NT)-(diag(T)%x%(W%*%P_nlminb)))%*%(X%*%beta)
Y_gorro2_1_rho <- Y_gorro2
Y_gorro2
```

```
resid2 <- Y_gorro2 - Y
resid2_1_rho <- resid2
resid2
```

```
plot(resid)
plot(resid2)
```

```
## Bondad de ajuste
```

```
Prom_Y <- as.numeric(t(ones(NT,1))%*%Y/NT[1])
Prom_Y
R2 <- 1-(t(resid2)%*%resid2)/(t(Y-Prom_Y))%*%(Y-Prom_Y)
R2
```

```
R2_Adj <- R2 - k/(NT-(k+1))*(1-R2)
R2_Adj
```

```
### Graficas
```

```
attach(Data_Panel)
```

```
pdf("Fit_HDI_dummies_1_rho_EI.pdf")
plot(EI,HDI,col="black", xlab="IE", ylab="IDH", main =
"Estimación usando coeficiente espacial único")
points(EI,Y_gorro2,col="blue")
legend(0.35, 0.9, c("Observado", "Estimación"),
col=c("black", "blue"), pch=c(1))
dev.off()
```

```
pdf("Fit_HDI_dummies_1_rho_LEI.pdf")
plot(LEI,HDI,col="black", xlab="IEV", ylab="IDH", main =
```

---

```

    "Estimación usando coeficiente espacial único")
    points(LEI,Y_gorro2,col="blue")
    legend(0.6, 0.9, c("Observado", "Estimación"),
           col=c("black", "blue"), pch=c(1))
dev.off()

pdf("Fit_HDI_dummies_1_rho_GNI.pdf")
    plot(GNI,HDI,col="black", xlab="ln(IIB) per cápita", ylab="IDH",
         main = "Estimación usando coeficiente espacial único")
    points(GNI,Y_gorro2,col="blue")
    legend(7.5, 0.9, c("Observado", "Estimación"),
           col=c("black", "blue"), pch=c(1))
dev.off()

### Proceso para SARP de coeficientes espaciales especificos

# Establecer funcion objetivo

neg_lklh_func <- function(P) {
  Y=Y
  H=H
  W=W
  P_diag = diag(P)
  LL = (NT/2)*log(t((diag(NT)-H)%*(diag(NT)-(diag(T)%x%P_diag)
%*(diag(T)%x%W))%*Y)%*((diag(NT)-H)%*(diag(NT)
-(diag(T)%x%P_diag)*(diag(T)%x%W))%*Y))
- log(det((diag(NT)-(diag(T)%x%P_diag)%*(diag(T)%x%W))))
}

# Optimizacion usando rutinas PORT

P_diag_nlmnb = nlmnb(rep(0, N), neg_lklh_func)
P_diag_nlmnb <- round(P_diag_nlmnb[[1]], 3)
P_nlmnb <- diag(P_diag_nlmnb)

```

```

P_nlminb

# Hesiano (y matriz de informacion observada de Fisher)

#install.packages('pracma')
library(pracma)

Hess <- hessian(neg_lklh_func, diag(P_nlminb))
Hess

# Matriz de covarianza de la optimizacion

Var_P <- solve(Hess)
Var_P

# Betas

B <- diag(NT)-diag(T)%x%(W%*%P_nlminb)

beta <- solve(t(X)%*%X)%*%t(X)%*%B%*%Y
beta

beta_0 = solve(t(X)%*%X)%*%t(X)%*%Y
beta_d = solve(t(X)%*%X)%*%t(X)%*%(diag(T)%x%W)%*%Y

beta_0
beta_d

# Sigma_2

sigma2_MV <- t(B%*%Y-X%*%beta)%*%(B%*%Y-X%*%beta)/NT
sigma2_MV

sigma_2_unb <- t(B%*%Y-X%*%beta)%*%(B%*%Y-X%*%beta)/(NT-dim(beta)[1])
sigma_2_unb

```

---

```

# Varianza de los betas

Var_beta <- sigma2_MV[1]*solve(t(X)%*%X)%*%
  t(X)%*%B%*%t(B)%*%X)%*%solve(t(X)%*%X)
Var_beta

# Valores p para los betas

vector_t_betas <- matrix(, nrow = 0, ncol = 1)

for (n in 1:k) {
  vector_t_betas <- rbind(vector_t_betas,
    beta[n,1]/sqrt(Var_beta[n,n]))
}

vector_t_betas

p_values_betas <- matrix(, nrow = 0, ncol = 1)

for (n in 1:k) {
  p_values_betas <- rbind(p_values_betas,
    2*dt(vector_t_betas[n,1], NT-k-1))
}

p_values_betas

# Valores p para los rhos especificos

vector_z_rhos <- matrix(, nrow = 0, ncol = 1)

for (n in 1:N) {
  vector_z_rhos <- rbind(vector_z_rhos, P_nlminb[n,n]/sqrt(Var_P[n,n]))
}

```

```
vector_z_rhos

p_values_rhos <- matrix(, nrow = 0, ncol = 1)

for (n in 1:N) {
  p_values_rhos <- rbind(p_values_rhos,
    2*dnorm(vector_z_rhos[n,1], mean = 0, sd = 1))
}

p_values_rhos

## Intervalos de confianza

## Para los betas

t_statistic_betas <- qt(0.975, df = NT-k-1)
band_betas <- matrix(, nrow = 0, ncol = 1)

for (n in 1:k) {
  band_betas <- rbind(band_betas, t_statistic_betas*sqrt(Var_beta[n,n]))
}

band_betas

vector_IC_betas <- matrix(, nrow = 0, ncol = 3)

for (n in 1:k) {
  vector_IC_betas <- rbind(vector_IC_betas, c(beta[n,1] -
    band_betas[n,1], beta[n,1], beta[n,1] + band_betas[n,1]))
}

vector_IC_betas

## Para los rhos
```

---

```

z_statistic_rhos <- qnorm(0.975, mean = 0, sd = 1)
band_rhos <- matrix(, nrow = 0, ncol = 1)

for (n in 1:N) {
  band_rhos <- rbind(band_rhos, z_statistic_rhos*sqrt(Var_P[n,n]))
}

band_rhos

vector_IC_rhos <- matrix(, nrow = 0, ncol = 3)

for (n in 1:N) {
  vector_IC_rhos <- rbind(vector_IC_rhos, c(P_nlminb[n,n] -
    band_rhos[n,1], P_nlminb[n,n], P_nlminb[n,n] + band_rhos[n,1]))
}

vector_IC_rhos

## Prueba F de significancia global del modelo

Fstat = ((NT-k)*(t(B%*%Y)%*%H%*%B%*%Y))/(k*t(B%*%Y)
  %*%(diag(NT)-H)%*%B%*%Y)
Fstat
p_val <- 1 - pf(Fstat, df1 = k, df2 = NT-k)
p_val

## Prediccion

Y_gorro <- (diag(T)%x%(W%*%P_nlminb))%*%Y + X%*%beta
Y_gorro

resid <- Y_gorro - Y
resid

```



```
Y_gorro2 <- solve(diag(NT)-(diag(T)%x%(W%%P_nlminb)))%%(X%%beta)
Y_gorro2
```

```
resid2 <- Y_gorro2 - Y
resid2
```

```
plot(resid)
plot(resid2)
```

```
## Bondad de ajuste
```

```
Prom_Y <- as.numeric(t(ones(NT,1))%*%Y/NT[1])
Prom_Y
R2 <- 1-(t(resid2)%*%resid2)/(t(Y-Prom_Y))%*%(Y-Prom_Y)
R2
```

```
R2_Adj <- R2 - k/(NT-(k+1))*(1-R2)
R2_Adj
```

```
### Graficas
```

```
attach(Data_Panel)
```

```
pdf("Fit_HDI_dummies_spec_rho_EI.pdf")
plot(EI,HDI,col="black", xlab="IE", ylab="IDH", main =
"Estimación usando coeficientes espaciales específicos")
points(EI,Y_gorro2,col="forestgreen")
legend(0.35, 0.9, c("Observado", "Estimación"),
col=c("black", "forestgreen"), pch=c(1))
dev.off()
```

```
pdf("Fit_HDI_dummies_spec_rho_LEI.pdf")
plot(LEI,HDI,col="black", xlab="IEV", ylab="IDH", main =
"Estimación usando coeficientes espaciales específicos")
```

---

```

points(LEI,Y_gorro2,col="forestgreen")
  legend(0.6, 0.9, c("Observado", "Estimación"),
    col=c("black", "forestgreen"), pch=c(1))
dev.off()

pdf("Fit_HDI_dummies_spec_rho_GNI.pdf")
plot(GNI,HDI,col="black", xlab="ln(IIB) per cápita", ylab="IDH",
  main = "Estimación usando coeficientes espaciales específicos")
points(GNI,Y_gorro2,col="forestgreen")
  legend(7.5, 0.9, c("Observado", "Estimación"),
    col=c("black", "forestgreen"), pch=c(1))
dev.off()

## Prueba de contraste para la igualdad de los rho
# vs rho espaciales significativamente distintos
#####

p <- N
S_1 <- ones(N,N)
S_1 <- Var_1_rho*S_1
S_2 <- Var_P
S_0 <- (1/NT)*(S_1+S_2)

sum_tr1 <- (1/NT)*(sum(diag((1/NT)*(S_1%%solve(S_0))%%
  ((1/NT)*(S_1%%solve(S_0)))+sum(diag((1/NT)*(S_1%%solve(S_0))))^2)))
sum_tr2 <- (1/NT)*(sum(diag((1/NT)*(S_2%%solve(S_0))%%
  ((1/NT)*(S_2%%solve(S_0)))+sum(diag((1/NT)*(S_2%%solve(S_0))))^2)))

v <- (p+p^2)/(sum_tr1 + sum_tr2)
v

T2 <- v*p/(v-p+1)
T2

```

```
Fstat <- T2*(v-p+1)/v*p
Fstat
p_val <- 1 - pf(Fstat, df1 = p, df2 = v-p+1)
p_val

#####
## Normalidad ##
#####

# Clear enviroment
rm(list=ls())

# Set work directory
setwd("C:/Users/DavidD/Videos/Dropbox UNAL/Dropbox/Tesis Spatial Panel/Experiments/R")

# Name and direct work file

Resids_U = read.csv(file="Resid_Unique.csv", header=TRUE, sep = ",")
Resids_S = read.csv(file="Resid_Specific.csv", header=TRUE, sep = ",")

Resid_U <- Resids_U[2:13]
Resid_S <- Resids_S[2:13]

## Requerided package
#install.packages('MVN')
library(MVN)

## Unique rho
#####

# create univariate Q-Q plots
Q_Q_U <- mvn(data = Resid_U, mvnTest = "royston", univariatePlot = "qqplot")
# create univariate histograms
```

---

```
Histo_U <- mvn(data = Resid_U, mvnTest = "royston", univariatePlot = "histogram")

# Henze-Zirkler's MVN test

Henze_U <- mvn(data = Resid_U, mvnTest = "hz")
Henze_U

# Doornik-Hansen's MVN test
Doornik_U <- mvn(data = Resid_U, mvnTest = "dh")
Doornik_U

## Specific rho
#####

# create univariate Q-Q plots
Q_Q_S <- mvn(data = Resid_S, mvnTest = "royston", univariatePlot = "qqplot")
# create univariate histograms
Histo_S <- mvn(data = Resid_S, mvnTest = "royston", univariatePlot = "histogram")

# Henze-Zirkler's MVN test

Henze_S <- mvn(data = Resid_S, mvnTest = "hz")
Henze_S

# Doornik-Hansen's MVN test
Doornik_S <- mvn(data = Resid_S, mvnTest = "dh")
Doornik_S
```

# Bibliografía

- Anselin, L. (2001), 'Spatial econometrics', *A companion to theoretical econometrics* **310330**.
- Anselin, L. (2013), *Spatial Econometrics: Methods and Models*, Vol. 4, Springer Science & Business Media.
- Anselin, L. & Kelejian, H. H. (1997), 'Testing for spatial error autocorrelation in the presence of endogenous regressors', *International Regional Science Review* **20**(1-2), 153–182.
- Anselin, L., Le Gallo, J. & Jayet, H. (2008), Spatial panel econometrics, in 'The Econometrics of Panel Data', Springer, pp. 625–660.
- Arellano, M. (2003), *Panel data econometrics*, Oxford University Press.
- Ayvar-Campos, F. J., Navarro-Chávez, J. C. L. & Giménez-García, V. M. (2017), 'La eficiencia de la dimensión ingreso del IDH en México', *Ensayos. Revista de economía* **36**(2), 95–121.
- Baltagi, B. (2001), *Econometric Analysis of Panel Data*, John Wiley & Sons.
- Baltagi, B. H., Egger, P. & Pfaffermayr, M. (2013), 'A generalized spatial panel data model with random effects', *Econometric Reviews* **32**(5-6), 650–685.
- Baltagi, B. H. & Li, D. (2006), 'Prediction in the panel data model with spatial correlation: the case of liquor', *Spatial Economic Analysis* **1**(2), 175–185.
- Baltagi, B. H., Song, S. H. & Koh, W. (2003), 'Testing panel data regression models with spatial error correlation', *Journal of Econometrics* **117**(1), 123–150.
- Baringhaus, L., Danschke, R. & Henze, N. (1989), 'Recent and classical tests for normality-a comparative study', *Communications in Statistics-Simulation and Computation* **18**(1), 363–379.
- Barry, R. P. & Pace, R. K. (1999), 'Monte carlo estimates of the log determinant of large sparse matrices', *Linear Algebra and its Applications* **289**(1-3), 41–54.

- Bibby, J., Kent, J. & Mardia, K. (1979), 'Multivariate analysis'.
- Breusch, T. & Pagan, A. (1980), 'The lagrange multiplier test and its applications to model specification in econometrics', *The Review of Economic Studies* **47**(1), 239–253.
- Brueckner, J. K. (2006), 'Strategic interaction among governments', *A companion to urban economics* **26**, 332–347.
- Celemín, J. P. & Velázquez, G. Á. (2017), 'Spatial analysis of the relationship between a life quality index, HDI and poverty in the province of Buenos Aires and the Autonomous City of Buenos Aires, Argentina', *Social Indicators Research* **17**, 1–21.
- Chintagunta, P. K., Jain, D. C. & Vilcassim, N. J. (1991), 'Investigating heterogeneity in brand preferences in logit models for panel data', *Journal of Marketing Research* **1**, 417–428.
- Cliff, A. D. & Ord, J. K. (1973), 'Spatial autocorrelation, monographs in spatial environmental systems analysis', *London: Pion Limited* **1**.
- Doornik, J. A. & Hansen, H. (2008), 'An omnibus test for univariate and multivariate normality', *Oxford Bulletin of Economics and Statistics* **70**, 927–939.
- Druska, V. & Horrace, W. C. (2004), 'Generalized moments estimation for spatial panel data: Indonesian rice farming', *American Journal of Agricultural Economics* **86**(1), 185–198.
- Egger, P., Pfaffermayr, M. & Winner, H. (2005), 'An unbalanced spatial panel data approach to us state tax competition', *Economics Letters* **88**(3), 329–335.
- Elhorst, J. P. (2003), 'Specification and estimation of spatial panel data models', *International Regional Science Review* **26**(3), 244–268.
- Frazier, C. & Kockelman, K. (2005), 'Spatial econometric models for panel data: incorporating spatial and temporal data', *Transportation Research Record: Journal of the Transportation Research Board* (1902), 80–90.
- Goldfeld, S. M. & Quandt, R. E. (1965), 'Some tests for homoscedasticity', *Journal of the American Statistical Association* **60**(310), 539–547.
- Greene, W. (2000), *Econometric analysis, International edition*, Pearson US Imports & PHIPEs.
- Hausman, J. A. (1975), 'An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models', *Econometrica: Journal of the Econometric Society* **43**, 727–738.

- Hausman, J. A. (1978), ‘Specification tests in econometrics’, *Econometrica: Journal of the Econometric Society* **46**, 1251–1271.
- Henze, N. & Zirkler, B. (1990), ‘A class of invariant consistent tests for multivariate normality’, *Communications in Statistics-Theory and Methods* **19**(10), 3595–3617.
- Kao, C. (1999), ‘Spurious regression and residual-based tests for cointegration in panel data’, *Journal of Econometrics* **90**(1), 1–44.
- Kapoor, M., Kelejian, H. H. & Prucha, I. R. (2007), ‘Panel data models with spatially correlated error components’, *Journal of Econometrics* **140**(1), 97–130.
- Kmenta, J. & Rafailzadeh, B. (1997), *Elements of econometrics*, University of Michigan Press.
- Kumar, S. (2014), ‘Eigenvalue statistics for the sum of two complex wishart matrices’, *EPL (Europhysics Letters)* **107**(6), 60002.
- Lax, P. D. & Lax, P. (2007), *Linear Algebra and its Applications*, Wiley-Interscience.
- Lee, L.-F. (2004), ‘Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models’, *Econometrica* **72**(6), 1899–1925.
- Lee, L.-f. & Yu, J. (2010a), ‘Estimation of spatial autoregressive panel data models with fixed effects’, *Journal of Econometrics* **154**(2), 165–185.
- Lee, L.-f. & Yu, J. (2010b), ‘Some recent developments in spatial panel data models’, *Regional Science and Urban Economics* **40**(5), 255–271.
- LeSage, J. & Pace, R. K. (2009), *Introduction to spatial econometrics*, Chapman and Hall/CRC.
- Letskovičová, H. et al. (2014), ‘Spatial distribution and relationship between sustainable development measures in EU countries’, *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* **53**(3), 87–94.
- Maddala, G. S. (1987), ‘Limited dependent variable models using panel data’, *Journal of Human Resources* **22**.
- Maddala, G. S. & Wu, S. (1999), ‘A comparative study of unit root tests with panel data and a new simple test’, *Oxford Bulletin of Economics and Statistics* **61**, 631–652.
- Magnus, J. R. & Neudecker, H. (1988), ‘Matrix differential calculus with applications in statistics and econometrics’, *Wiley Series in Probability and Mathematical Statistics*.

- Mardia, K. V. (1970), 'Measures of multivariate skewness and kurtosis with applications', *Biometrika* **57**(3), 519–530.
- Moran, P. A. (1950), 'Notes on continuous stochastic phenomena', *Biometrika* **37**(1/2), 17–23.
- Mutl, J. & Pfaffermayr, M. (2008), The spatial random effects and the spatial fixed effects model: the hausman test in a cliff and ord panel model, Technical report, Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS).
- Ord, K. (1975), 'Estimation methods for models of spatial interaction', *Journal of the American Statistical Association* **70**(349), 120–126.
- Pace, R. K. & Barry, R. (1997), 'Quick computation of spatial autoregressive estimators', *Geographical analysis* **29**(3), 232–247.
- Pawitan, Y. (2001), *In all likelihood: statistical modelling and inference using likelihood*, Oxford University Press.
- Pesaran, M. H. & Smith, R. (1995), 'Estimating long-run relationships from dynamic heterogeneous panels', *Journal of Econometrics* **68**(1), 79–113.
- Qiu, Q., Sung, J., Davis, W. & Tchernis, R. (2018), 'Using spatial factor analysis to measure human development', *Journal of Development Economics* **132**, 130–149.
- Shaker, R. R. (2015), 'The spatial distribution of development in Europe and its underlying sustainability correlations', *Applied Geography* **63**, 304–314.
- Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.