

ECONOMETRÍA ESPACIAL USANDO STATA. TEORÍA Y APLICACIONES*

MARCOS HERRERA

CONICET-IELDE

Universidad Nacional de Salta (Argentina)

(e-mail: mherreragomez@gmail.com)

Fecha de revisión: Octubre, 2016[†]

Comentarios Bienvenidos

*ESTE MANUSCRITO PUEDE SER IMPRESO Y REPRODUCIDO PARA USO PERSONAL O COMO MATERIAL INSTRUCTIVO, PERO NO PUEDE SER IMPRESO PARA FINES COMERCIALES.

[†]Previas versiones de este documento han sido presentadas en XL Reunión de Estudios Regionales, AEER (Zaragoza, España, 2014), en la 24th PRSCO & 7th SOCHER Annual Conference (Santiago y Viña del Mar, Chile, 2015) y en el 5th SAER Annual Conference (San Miguel de Tucumán, Argentina, 2015). Además, este material forma parte de la bibliografía del curso “Econometría Espacial” de la Maestría en Economía (Universidad de San Andrés, Argentina).

«... las cosas cercanas están más
relacionadas que las cosas
distantes».
(Tobler, 1970)

«... Mas vale amigo cercano que
pariente distante».
(Prov. 27:10b, La Biblia)

Índice general

Índice general	i
Prefacio	iii
1 Introducción	1
2 Análisis exploratorio de datos espaciales	4
2.1 Convirtiendo archivos shapes a Stata	4
2.2 Manipulación y visualización de datos espaciales	8
2.3 Creación de matrices de pesos espaciales	14
2.4 Estadísticos globales de autocorrelación espacial	19
3 Econometría espacial bajo corte transversal	26
3.1 Resumen del modelo teórico de Ertur y Koch (2007)	26
3.2 Detección de estructura espacial bajo MCO	28
3.3 Estimación por máxima verosimilitud	36
3.4 Estimación por método de momentos generalizados	47
4 Análisis espacial local	59
4.1 Estadísticos locales	59
Estadístico Local I de Moran	59
Estadísticos de Getis y Ord	63
4.2 Método de expansión espacial	65
4.3 Regresiones geográficamente ponderadas	65
5 Efectos espaciales en datos de panel	69
5.1 Introducción	69
5.2 Detección de dependencia espacial en datos de panel	71
5.3 Estimación de modelos espaciales estáticos	76
Modelo de Rezago Espacial	76
Modelo de Error Espacial	79
Modelo <i>SARAR</i>	81
Modelo espacial de Durbin	82
Modelo de Error de Durbin	83
Modelo de Cliff Ord	84

5.4	Interpretación de los modelos con rezago espacial endógeno	85
5.5	Estimación de modelos espaciales dinámicos	89
	Modelos dinámicos de rezago espacial	92
	Modelos dinámicos espaciales de Durbin	93
6	Econometría espacial para redes sociales	96
7	Modelos jerárquicos-espaciales	97
7.1	Modelo nulo	99
7.2	Modelo condicional de intercepto aleatorio	100
7.3	Modelo condicional de intercepto y pendiente aleatoria	101
8	Comentarios finales	102
	Bibliografía	103

Prefacio

Este documento pretende ser una guía para los economistas aplicados que desean trabajar con datos georreferenciados bajo el entorno de Stata. El documento está desarrollado para un nivel de posgrado tipo maestría o de un primer año doctoral.

El desarrollo del manuscrito está diseñado pensando en alumnos que tienen un entendimiento de análisis multivariado, álgebra lineal y matemática estadística. Un curso previo de econometría puede ser útil para una mejor comprensión del material pero no es completamente fundamental. Tres libros de textos que pueden servir de consulta son [Greene \(2003\)](#), [Davidson & MacKinnon \(2004\)](#) y [Wooldridge \(2009\)](#).

En principio, no es obvio que los datos georreferenciados requieran un tratamiento especial o técnicas econométricas diferentes a las estándares. Sin embargo, la econometría tradicional frecuentemente impone un conjunto de supuestos o restricciones sobre los datos analizados que no se cumplen para datos espaciales. Por ejemplo, gran parte de la teoría econométrica descansa en que la muestra de datos proviene de un muestreo aleatorio, implicando que las observaciones son independientes e idénticamente distribuidas (i.i.d.). Este requisito es casi siempre violado cuando se tienen datos espaciales debido a su propia naturaleza, generando la invalidez de la gran parte del análisis econométrico estándar. Esta limitación ha generado el surgimiento de diversas áreas de investigación específicas y, en particular, de la econometría espacial.

Debido a que este documento está en progreso, algunas secciones se encuentran incompletas y también hay varios tópicos que no se encuentran totalmente desarrollados.

Capítulo 1

Introducción

El área de la econometría espacial se ocupa del uso de técnicas estadísticas y econométricas que permiten el tratamiento explícito del espacio en modelos económicos. Este nombre fue dado por Jean Paelinck, quien utilizó por primera vez el término para destacar la necesidad de desarrollar una nueva rama de la econometría que sirviese de fundamentación metodológica a los modelos regionales y urbanos (*Annual Meeting of the Dutch Statistical Association*, Tilburg, 1974). [Anselin \(2010\)](#) ubica el inicio del área en el año 1979 debido a un conjunto de publicaciones que sentaron las primeras bases metodológicas distinguibles: *Spatial econometrics* ([Paelinck & Klaassen, 1979](#)), *Exploratory and explanatory statistical analysis of spatial data* ([Bartels & Ketellapper, 1979](#)), *Spatial time series* ([Bennett, 1979](#)) y *Problems in estimating econometric relations in space* ([Hordijk, 1979](#)). Una década más tarde, con la publicación del libro de texto: *Spatial econometrics: Methods and models* ([Anselin, 1988](#)), se consolidan las líneas metodológicas fundamentales que han marcado la agenda de investigación en los posteriores años.

Un punto crucial sobre el desarrollo y difusión de la econometría espacial ha sido el relacionado a los paquetes econométricos. En los inicios de la especialidad, década del 70' y 80', la falta de un programa apropiado para el análisis espacial fue uno de los principales impedimentos en la realización de investigaciones empíricas. Los investigadores utilizaban códigos de Fortran o de algún otro tipo de lenguaje de programación ya que los programas disponibles, tales como SAS y SPSS, solo proveían herramientas estadísticas estándares no considerando la presencia de dependencia espacial o directamente evitándola. La autocorrelación espacial era vista más bien como un problema a ser removido o evitado ([Goodchild et al., 1992](#)). En la década del 90', estas dificultades comienzan a sortearse debido al avance tecnológico de los sistemas de información geográfica (SIG) y al desarrollo de paquetes estadísticos específicos para manipular y modelizar estos datos como SpaceStat, primera vez lanzado en 1991 ([Anselin, 1992](#)).

En la actualidad, existen varias alternativas para realizar un análisis econométrico espacial. Sin ser exhaustivos, pueden mencionarse entre los más importantes: Econometrics Toolbox desarrollado en MATLAB ([LeSage, 1999](#)); en R puede utilizarse el paquete spdep desarrollado en el año 2001 ([Bivand, 2002](#)) y splm desarrollado por [Millo & Piras \(2012\)](#); el programa GeoDa, lanzado en el año 2002 ([Anselin et al., 2006](#)), con gran impacto a nivel usuario; y GeoDaSpace de reciente desarrollo basado en el código PySAL ([Rey & Anselin, 2010](#)). Además, numerosos especialistas en el área aportan códigos propios escritos en diferentes lenguajes que complementan estos programas.

Claramente Stata no ha sido un precursor del análisis espacial, aunque diferentes usuarios han aportado herramientas espaciales muy útiles bajo su entorno. El primer aporte importante data del año 2001, *Tools for spatial data analysis* (Pisati, 2001). Desde entonces, se han desarrollado gradualmente diferentes comandos que permiten alcanzar un nivel de análisis adecuado para realizar investigaciones en econometría espacial. En el presente documento, se presentarán los códigos disponibles más relevantes hasta la fecha.

En orden de facilitar la reproducibilidad de los resultados aquí presentados, se acompaña al texto con un conjunto de archivos, comprimidos en taller_2015.rar que permiten seguir el paso a paso de cada capítulo. El principal archivo es denominado taller_2015.do que incluye las instrucciones y comandos de ejecución.

Bases de datos a utilizar

A lo largo de la guía se utilizarán diferentes conjunto de datos. La primera parte del documento utilizará una base de datos de Europa. Estos datos corresponden a 164 regiones NUTS2 de la primera formación UE15, para el periodo 2001-2012, con únicamente dos variables regionales: la tasa de desempleo y la tasa de migración neta. Ambas variables han sido construidas con información proveniente de Eurostat.

La relación entre desempleo y migración neta sirve como ejemplo de estudio exploratorio, además de ser una relación de especial interés en el ámbito regional. Desde la teoría económica, una mayor inmigración neta puede afectar al desempleo mediante dos efectos en el mercado laboral: uno directo debido al incremento de la oferta laboral y otro indirecto producto de la mayor demanda de bienes de consumo que se traduce en un incremento de la demanda laboral. El impacto de la inmigración neta sobre el desempleo dependerá de cuál de los efectos es el que prevalece. Para la escuela neoclásica, el efecto directo es el principal y generará un mayor desempleo. Para la escuela de la Nueva Geografía Económica, el efecto indirecto es el que prevalece y la inmigración causará un menor desempleo en la región receptora. Ambas teorías dan pie a postular una relación inicial en donde la variable dependiente es la tasa de desempleo y la variable explicativa es la tasa de migración neta. Sin embargo, el signo y la magnitud del impacto de la migración neta sobre el desempleo es una cuestión empírica. Se explorará dicha relación mediante las técnicas espaciales para corte transversal y para datos de panel.

Otra base de datos a utilizar es la provista por Ertur & Koch (2007) en su trabajo sobre externalidades espaciales e interdependencia tecnológica entre países. Siguiendo la literatura empírica sobre crecimiento económico, los autores extraen una muestra de países de las Penn World Tables (PWT version 6.1) que contiene información sobre ingreso real, inversión y población, entre otras variables, para un gran número de países. La muestra utilizada contiene 91 países para el periodo 1960-1995, los mismos países que fueron utilizados en el conocido trabajo de Mankiw et al. (1992). Esta base de datos será utilizada en el capítulo de econometría espacial para corte transversal.

En el capítulo dedicado a los efectos espaciales en datos de panel, además de los datos de desempleo y migración neta en Europa, se utilizarán los datos de consumo de tabaco de un trabajo realizado por Baltagi & Li (2004) en el que muestran el comportamiento predictivo de diferentes modelos espaciales. La base de datos será utilizada para explicar la demanda de cigarrillos en 46

estados norteamericanos en un periodo de 30 años, desde 1963 a 1992. La variable dependiente es la venta per cápita de cigarrillos en términos reales, medida en packs por personas de 14 o más años. Las variables explicativas son: el precio de venta promedio del pack de cigarrillos y el ingreso disponible per cápita en términos reales. La base de datos se encuentra disponible en diferentes sitios web, en este documento se utilizará una versión adaptada que incluye la matriz de contactos espaciales¹.

El documento se encuentra estructurado de la siguiente manera. En el capítulo 2 se realiza un análisis exploratorio espacial de los datos. Debido a la particularidad de los mismos, se utilizan herramientas puntuales como son los diferentes tipos de mapas, diagramas de dispersión de Moran y mapas de clusters espaciales. El capítulo 3 presenta los modelos habituales de econometría espacial para datos de corte transversal, incluyendo las diferentes alternativas de estimación de máxima verosimilitud, variables instrumentales y método de momentos generalizados. En el último capítulo, capítulo 4, se exponen algunas herramientas disponibles para un análisis espacial local. En el capítulo 5 se presenta una serie de modelos espaciales para datos de panel de naturaleza estática y dinámica, incluyendo la estimación por efectos fijos o aleatorios. En el capítulo 6 se bosqueja un conjunto de comandos que pueden ser utilizados para redes sociales, destacando la particularidad y diferencia de este tipo de datos (INCOMPLETO). El capítulo 7 incorpora la posibilidad de utilizar modelos multinivel para datos espaciales (INCOMPLETO). En todos los capítulos se combina la presentación metodológica de cada modelo y/o tests con los comandos disponibles en Stata para su estimación.

¹Los datos se encuentran disponibles en el sitio web personal del Prof. Elhorst <http://www.regrooningen.nl/elhorst/software.shtml>.

Capítulo 2

Análisis exploratorio de datos espaciales

Este capítulo muestra las herramientas disponibles para realizar un análisis exploratorio de datos espaciales (AEDE). El AEDE es un conjunto de técnicas que permite detectar asociaciones espaciales, concentraciones locales y detección de datos anómalos (outliers). La idea detrás del AEDE es que los datos hablen por si mismos, sin imponer una estructura conceptual previa, similar al concepto que surge del Análisis Exploratorio de Datos (AED) (Anselin, 1999). Sin embargo, las herramientas usuales del AED no son válidas para datos espaciales ya que por naturaleza los mismos muestran autocorrelación y es por ello que se necesitan desarrollos específicos como los que se presentan a continuación.

El capítulo mostrará el potencial de los mapas de coropletas y algunos indicadores simples que permiten detectar patrones espaciales y sugerir hipótesis de trabajo.

2.1 Convirtiendo archivos shapes a Stata

Para iniciar el análisis espacial es necesario tener información georreferenciada. Un problema habitual es la obtención de dicha información. Sin embargo, esta restricción se ha vuelto menos importante a medida que diferentes sitios webs han ido ofreciendo archivos georreferenciados de manera gratuita. Entre los sitios gratuitos puede mencionarse Global Administrative Areas (GADM), que provee archivos de mapas base para casi todos los países¹; y el sitio DIVA-GIS².

El tipo de archivo estándar en el que se almacena la información geográfica es denominado shape (shapefile), originalmente desarrollado por la empresa ESRI para su programa ArcView GIS. Si bien existen otros formatos alternativos, el formato shape se ha convertido en el tipo de archivo más utilizado para el intercambio de información geográfica.

Un archivo shape es un formato vectorial que almacena la localización de elementos geográficos y sus atributos. La diferencia con otro tipo de archivo es que es un formato multiarchivo: conjunto relacionado de ficheros informáticos. El número mínimo y habitual es de tres ficheros con las siguientes extensiones:

.shp es el archivo que almacena las entidades geométricas de los objetos.

.shx es el archivo que almacena el índice de las entidades geométricas.

¹ <http://www.gadm.org/about>.

² <http://www.diva-gis.org/gdata>.

.dbf es la base de datos, en formato dBASE, donde se almacena la información de los atributos de los objetos.

Hay herramientas escritas por usuarios para convertir los shapes en archivos .dta de Stata. El comando **shp2dta** (Crow, 2013) permite crear dos conjuntos de datos: uno contendrá la información de los atributos del archivo .dbf, y otro contendrá la información geográfica.

El primer paso es leer la información georreferenciada mediante el siguiente comando:

```
. shp2dta using nuts2_164, database(datos_shp) coordinates(coord) genid(id) genc(c)
```

Este comando lee el archivo shape “nuts2_164”, generando dos archivos: “datos_shp.dta” y “coord.dta”. En el archivo de datos, “datos_shp”, se genera adicionalmente una variable identificadora de cada unidad geográfica, **id**, y los centroides, puntos que definen el centro geométrico de los polígonos, identificados por la latitud (**y_c**) y la longitud (**x_c**).

```
. use datos_shp, clear
. describe
```

Contains data from datos_shp.dta

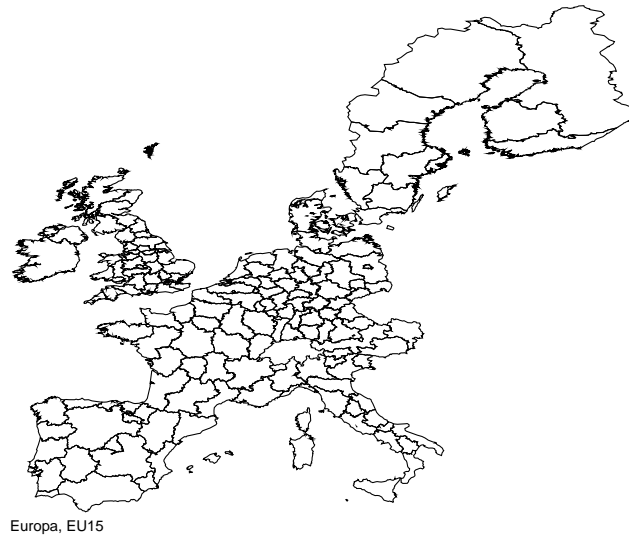
```
obs:          164
vars:          9                    5 Jul 2015 09:14
size:          5,904
```

variable name	storage type	display format	value label	variable label
id	int	%12.0g		Area ID
x_c	float	%9.0g		x-coordinate of area centroid
y_c	float	%9.0g		y-coordinate of area centroid
POLY_ID	int	%10.0g		POLY_ID
OBJECTID	int	%10.0g		OBJECTID
NUM	int	%10.0g		NUM
NUTS_ID	str4	%9s		NUTS_ID
X_COORD	double	%10.0g		X_COORD
Y_COORD	double	%10.0g		Y_COORD

El archivo datos_shp.dta contiene la información sobre los atributos del archivo dbf. El archivo de atributos tiene únicamente las variables que identifican a cada NUTS2, el nombre del objeto, y las coordenadas geográficas.

Pisati (2008) provee el comando **spmap** que permite obtener diferentes gráficos de mapas. Por ejemplo, puede obtenerse un mapa con los límites de cada área comúnmente denominado mapa base.

```
. spmap using datos_shp, id(id)
```



Puede suceder que el shape contenga polígonos que no son de interés o de los que no se tenga información de las variables. En este caso puede utilizarse el comando **drop** para eliminar las regiones sin información. Por ejemplo, si no hay información sobre Finlandia y Suecia:

```
. drop if id==3|id==5|id==6|id==164|id==7|id==8|id==12|id==4|id==2|id==1|id==11|id==12
. spmap using coord, id(id)
```



Estos mapas base solo contienen información geográfica, no hay variables que puedan ser analizadas. Tal como se comentó en la introducción, en este capítulo se analizarán dos variables que han sido construidas con información proveniente de la Oficina Europea de Estadísticas, más conocida como Eurostat.

La primera variable es la tasa de desempleo, definida por Eurostat como el porcentaje de personas desempleadas respecto a la fuerza laboral regional, esto es, población en edad de trabajar entre 15-74 años.

La segunda variable es la tasa de migración neta que se define como la razón de la migración neta durante el año (incluyendo el ajuste estadístico) respecto a la población promedio en ese año, expresada por cada mil habitantes. La migración neta es estimada como la diferencia entre el número de personas que ingresan y el número de personas que egresan de una región por lo que la variable puede tener valores positivos o negativos. Un valor positivo implica un exceso de personas ingresantes y es referida como inmigración neta. Un valor negativo implica un exceso de personas que egresan de la región y es referido como emigración neta.

Para incorporar la información descargada de la página de Eurostat el primer paso fue guardar el archivo .dbf en formato excel para utilizar la identificación proveniente del shape de cada polígono. Luego, se fueron incorporando los datos de las variables desempleo y migración neta al NUTS2 correspondiente. La variable que identifica a cada polígono servirá de vínculo entre el archivo excel y el shape.

Desde Stata, se importa el archivo excel generado y se guarda en formato .dta:

```
. import excel "C:\.\.\.\nuts2_164.xls", firstrow
. save "C:\.\.\.\migr_unemp.dta"
. use migr_unemp
. describe
```

Contains data from migr_unemp.dta

```
obs:          164
vars:          26                    5 Jul 2015 09:30
size:         32,472
```

	storage	display	value	
variable name	type	format	label	variable label

POLY_ID	int	%10.0g		POLY_ID
NUTS_ID	str4	%9s		NUTS_ID
NM2001	double	%10.0g		NM2001
NM2002	double	%10.0g		NM2002
NM2003	double	%10.0g		NM2003
NM2004	double	%10.0g		NM2004
.	.			.
.	.			.
.	.			.
U2008	double	%10.0g		U2008
U2009	double	%10.0g		U2009
U2010	double	%10.0g		U2010
U2011	double	%10.0g		U2011
U2012	double	%10.0g		U2012

La información ha sido guardada en formato wide: el número de filas se corresponde al número de NUTS2 y, para cada año, existe una variable de migración neta y desempleo. Con la información

guardada en archivos Stata, solo queda unir ambas bases de datos utilizando la variable link: `POLY_ID`.

```
. use datos_shp, clear
. merge 1:1 POLY_ID using migr_unemp, gen(union) force
```

```
Result                                # of obs.
-----
not matched                           0
matched                             164  (union==3)
-----

. assert union==3
. save migr_unemp_shp.dta
file migr_unemp_shp.dta saved
```

Mediante el comando `merge` la información de ambas bases es unida y la función `assert union==3` verifica que la condición sea verdadera. Si esta condición se cumple, entonces Stata no devolverá ninguna línea adicional. A pesar de estos recaudos, siempre es recomendable realizar un chequeo adicional en este paso comprobando visualmente que los polígonos se correspondan con la información de cada variable.

2.2 Manipulación y visualización de datos espaciales

Una vez unidas las bases de datos, es posible visualizar las variables mediante diferentes mapas de coropletas. Este tipo de mapa es un mapa temático en el que las áreas se colorean con una intensidad proporcional al valor que toma una variable cuantitativa (Slocum et al., 2009). Para ello utilizaremos el comando antes presentado, `spmap`.

En `spmap`, un mapa de coropletas es un mapa base cuyos polígonos estarán coloreados de acuerdo al valor tomado de un atributo (variable) *numérico*. La versión actual de `spmap` ofrece seis tipos de coropletas:

- Cuantiles: las clases corresponden a los cuantiles de la distribución de la variable, cada clase tendrá aproximadamente el mismo número de polígonos.
- Intervalos iguales: las clases se dividen según el valor que divide a la distribución de la variable en k intervalos iguales. El valor predefinido de k es 4.
- Diagrama de caja: las clases son seis y el número de polígonos en cada clase depende de la distribución de la variable elegida. Las clases son definidas de la siguiente manera: $[min, p25 - 1.5 \cdot ri]$, $(p25 - 1.5 \cdot ri, p25]$, $(p25, p50]$, $(p50, p75]$, $(p75, p75 + 1.5 \cdot ri]$, $(p75 + 1.5 \cdot ri, max]$, donde p es el percentil y ri es el rango intercuartílico.
- Desviación estándar: la distribución de la variable es dividida en k clases ($2 \leq k \leq 9$) cuya amplitud es definida como una fracción p de su desvío estándar, $s.d.$. El número de clases predefinido es $k = 4$: $[min, media - s.d.]$, $(media - s.d., media]$, $(media, media + s.d.]$, $(media + s.d., max.]$.

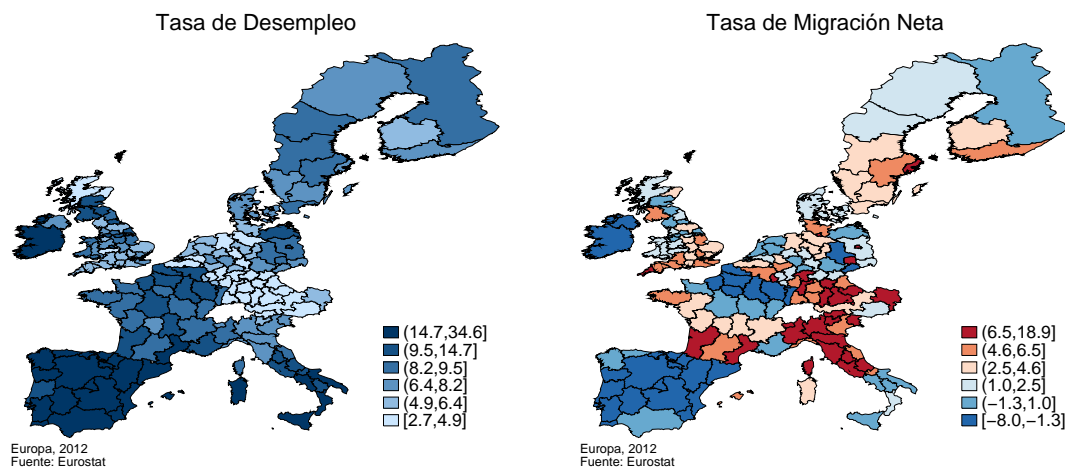
- k-medias: la distribución de la variable es dividida en k clases usando el análisis de partición de cluster de k-medias. Este procedimiento es iterativo finalizando cuando se alcanza la máxima bondad de ajuste según la varianza.
- Custom: los intervalos de clase son especificados por el usuario.

Las anteriores opciones solo permiten trabajar con variables cuantitativas. En el caso de las variables categóricas puede usarse la opción **unique**, que permite graficar variables asignando un color diferente a cada una de las categorías.

A continuación se presentan los mapas de coropletas aplicados a los datos de EU15 utilizando la información del último año disponible.

```
. format U2012 %12.1f
. spmap U2012 using coord, id(id) clmethod(q) cln(6) title("Tasa de Desempleo") ///
  legend(size(medium) position(5)) fcolor(Blues2) note("Europa, 2012" "Fuente: Eurostat")
. format NM2012 %12.1f
. spmap NM2012 using coord, id(id) clmethod(q) cln(6) title("Tasa de Migración Neta") ///
  legend(size(medium) position(5)) fcolor(BuRd) note("Europa, 2012" "Fuente: Eurostat")
```

Figura 2.1: Mapa de cuantiles.



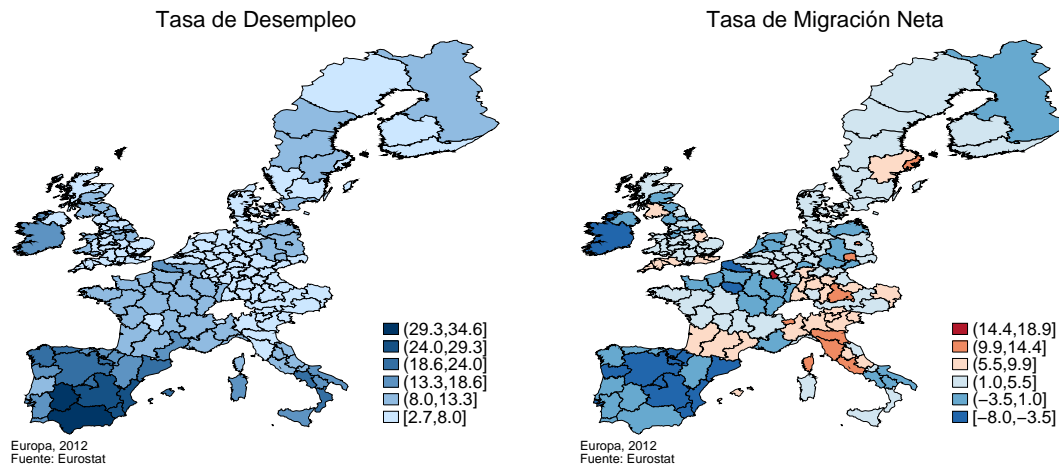
El mapa de cuantiles se genera utilizando el subcomando **clmethod(q)**. Además puede modificarse el número de clases usando **cln(#)**, así como los colores, la posición de la legenda, títulos, notas. La primera línea previa al comando del gráfico, **format**, permite controlar el formato numérico de la legenda.

Cada tipo de mapa destaca características diferentes de los datos. A continuación se presentan los mapas de intervalos iguales. Es notorio el cambio en comparación a los mapas de cuantiles. Cuando las variables tienen distribuciones de sus valores igualmente espaciados, o equidistantes, ambos mapas tenderán a asemejarse, pero no es este el caso.

```
. spmap U2012 using coord, id(id) clmethod(e) cln(6) title("Tasa de Desempleo") ///
  legend(size(medium) position(5)) fcolor(Blues2) note("Europa, 2012" "Fuente: Eurostat")
```

```
. spmap NM2012 using coord, id(id) clmethod(e) cln(6) title("Tasa de Migración Neta") ///
legend(size(medium) position(5)) fcolor(BuRd) note("Europa, 2012" "Fuente: Eurostat")
```

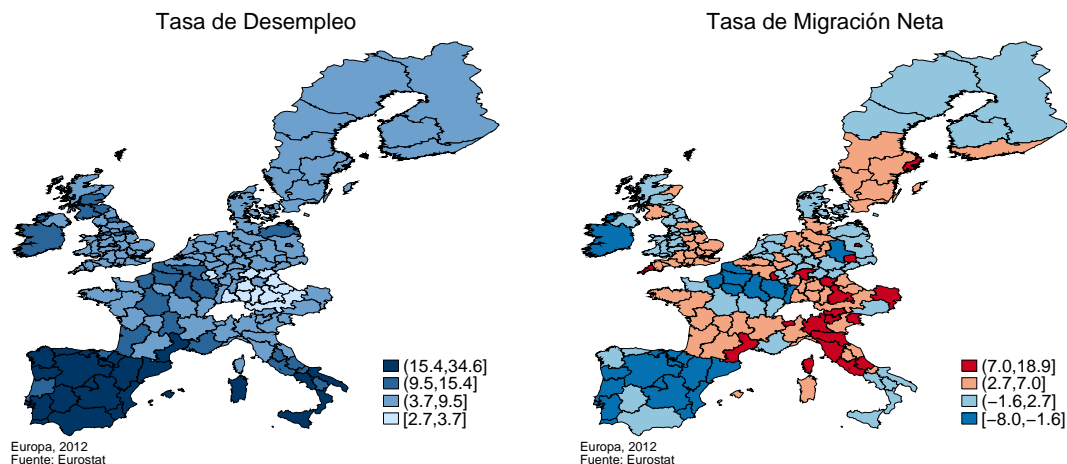
Figura 2.2: Mapa de intervalos iguales.



Otro tipo de mapa es el mapa de desvíos estándares respecto a la media muestral:

```
. spmap U2012 using coord, id(id) clmethod(s) title("Tasa de Desempleo") ///
legend(size(medium) position(5)) fcolor(Blues2) note("Europa, 2012" "Fuente: Eurostat")
. spmap NM2012 using coord, id(id) clmethod(s) title("Tasa de Migración Neta") ///
legend(size(medium) position(5)) fcolor(BuRd) note("Europa, 2012" "Fuente: Eurostat")
```

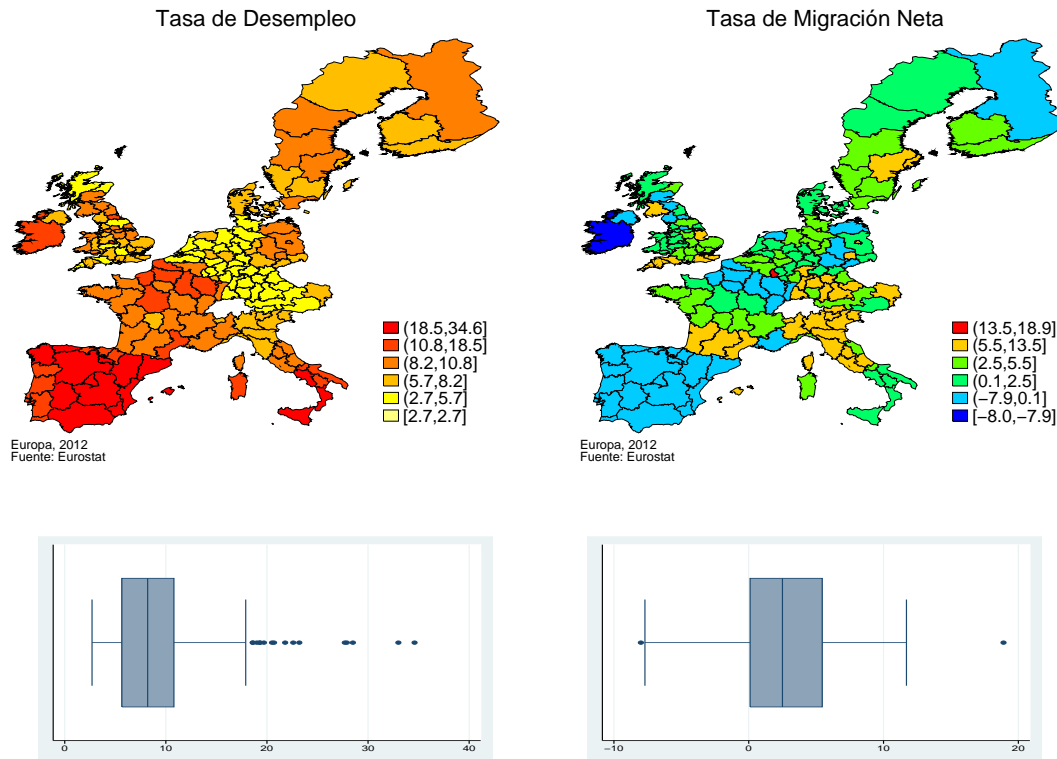
Figura 2.3: Mapa de desviaciones.



El mapa de diagrama de caja brinda una información diferente, permitiendo detectar valores atípicos. Este mapa puede presentarse conjuntamente con el diagrama de caja simple para poder vincular las observaciones:

```
. spmap U2012 using coord, id(id) clmethod(boxplot) title("Tasa de Desempleo") ///
  legend(size(medium) position(5)) fcolor(Heat) note("Europa, 2012" "Fuente: Eurostat")
. spmap NM2012 using coord, id(id) clmethod(boxplot) title("Tasa de Migración Neta") ///
  legend(size(medium) position(5)) fcolor(Rainbow) note("Europa, 2012" "Fuente: Eurostat")
```

Figura 2.4: Mapas de diagrama de caja.

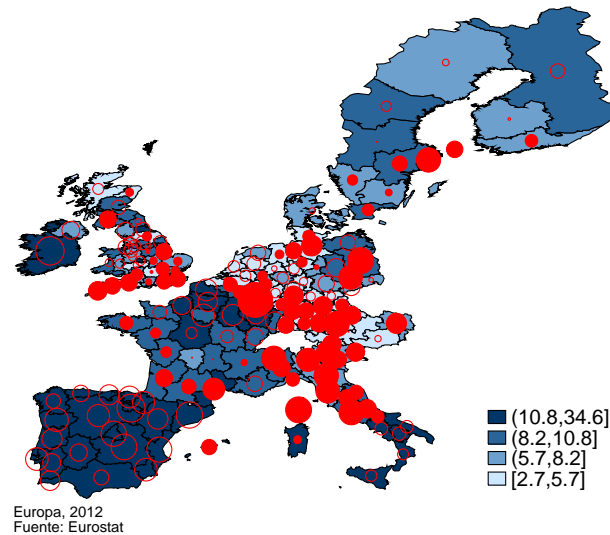


En el mapa de la tasa de desempleo pueden distinguirse los valores extremos altos, por encima del 75% de la distribución muestral. Estos valores se encuentran clasificados por la clase de color rojo intenso y contiene a la mayoría de las regiones españolas y al sur de Italia. En el caso de la migración neta hay dos observaciones atípicas, una en cada extremo. Se puede distinguir una región con emigración neta extrema (mayor valor negativo) identificada en Irlanda y otra región que experimenta una alta inmigración neta que representa a Luxemburgo.

Existen opciones adicionales tal como la posibilidad de combinar información de áreas con puntos. Por ejemplo, si se desea visualizar en el mapa la relación entre la distribución de cuantiles del desempleo y las desviaciones respecto a la media de la tasa de migración neta, puede aplicarse el siguiente comando:

```
. spmap U2012 using coord, id(id) fcolor(Blues2) point(data(migr_unemp_shp) xcoord(x_c)
  ycoord(y_c) deviation(NM2012) fcolor(red) size(*0.6)) legend(size(medium) position(5))
  note("Europa, 2012" "Fuente: Eurostat")
```


Figura 2.5: Mapa de cuantiles del desempleo y de desviaciones de migración neta.



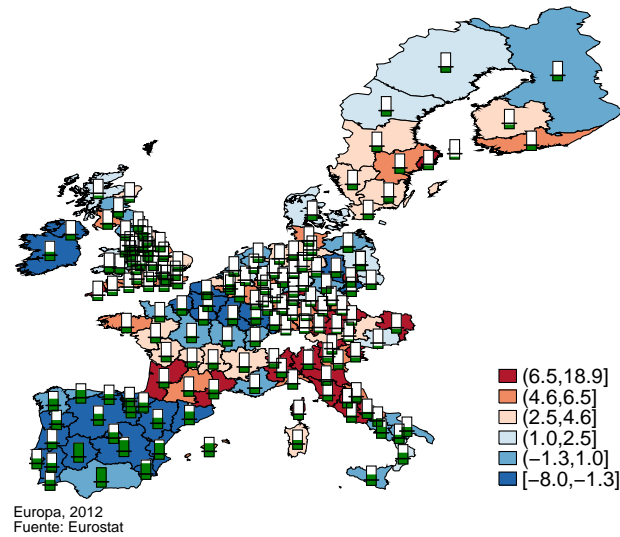
Los círculos huecos representan las desviaciones por debajo de la media de la migración neta y los círculos rellenos representan desviaciones por encima de la media. A mayor tamaño del círculo, más alejado se encuentra el valor de la región respecto a la media europea. Es fácil observar que los círculos huecos y de gran tamaño, que identifican a la emigración neta, se ubican en regiones de alto desempleo. Por lo contrario, los círculos que identifican altos valores de inmigración neta se ubican en áreas de bajo desempleo.

En general puede observarse que ambas variables muestran un patrón de agrupamiento espacial. Para el caso de la tasa de desempleo, los altos valores se encuentran al sur de Europa, en especial en las regiones de España, Portugal y parte de Italia. La distribución de los valores de migración neta refleja que hay zonas de alta inmigración como el norte de Italia, sur de Alemania, sur de Francia y sur del Reino Unido, principalmente. Este análisis visual permite anticipar que el comportamiento de las regiones NUTS2 no es independiente del espacio.

Otra combinación surge de reemplazar los puntos por diagramas de caja. Por ejemplo, combinar la distribución de la migración neta por cuantiles y el diagrama de caja del desempleo. En la Figura 2.6, los valores del desempleo se presentan mediante el color verde en el diagrama de caja. La altura del diagrama se refiere al máximo valor observado del desempleo y la línea horizontal en cada diagrama muestra el valor promedio del desempleo para todas las regiones. Tanto la altura de la caja como el valor representativo de la línea pueden modificarse mediante las opciones: `fweight()` y `refval()`.

```
. spmap NM2012 using coord, id(id) fcolor(BuRd) cln(6) diagram(var(U2012) xcoord(x_c)
ycoord(y_c) fcolor(green) size(1.5)) legend(size(medium) position(5)) ///
note("Europa, 2012" "Fuente: Eurostat")
```

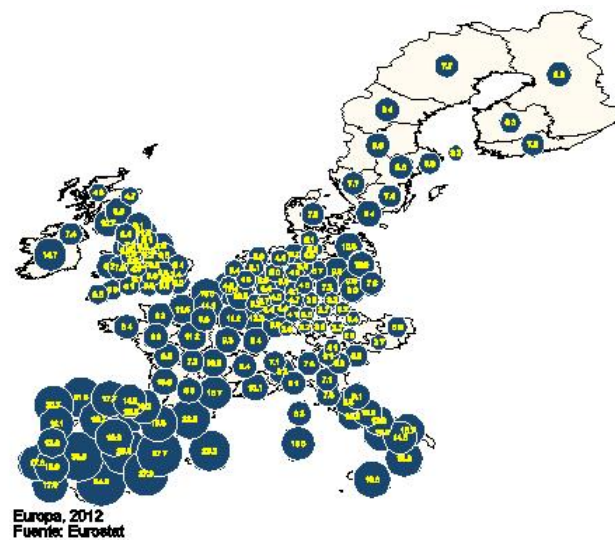
Figura 2.6: Mapa de cuantiles de migración neta y diagrama de caja de desempleo.



También puede combinarse el mapa base junto a círculos que representen proporcionalmente el valor de la variable, por ejemplo la tasa de desempleo.

```
. spmap using coord, id(id) fcolor(eggshell) point(xcoord(x_c) ycoord(y_c)    ///
proportional(U2012) fcolor(navy) ocolor(white) size(2.5)) legend(size(medium)  ///
position(5)) label(xcoord(x_c) ycoord(y_c) label(U2012) color(yellow) size(*0.5))///
note("Europa, 2012" "Fuente: Eurostat")
```

Figura 2.7: Mapa de proporcional de desempleo.



2.3 Creación de matrices de pesos espaciales

La sección previa ha brindado los primeros indicios de dependencia o autocorrelación³ espacial, sin embargo, no hay una cuantificación de esta autocorrelación y no se conoce si la misma es relevante. Es entonces necesario el uso de herramientas inferenciales para poder obtener una cuantificación de la misma. En este punto surge el problema de la matriz de pesos espaciales.

Siguiendo la información obtenida por el AEDE, se tiene cierta evidencia de que la distribución de la tasa de desempleo y la migración neta depende en cierta medida del valor observado de la variable en las regiones vecinas. Para simplificar la idea, supóngase que la variable de estudio es y , distribuida en solo tres regiones y que se ha detectado dependencia espacial. La formalización de esta dependencia puede realizarse de la siguiente manera:

$$\begin{aligned} y_i &= \alpha_{ij}y_j + \alpha_{ik}y_k + u_i, \\ y_j &= \alpha_{ji}y_i + \alpha_{jk}y_k + u_j, \\ y_k &= \alpha_{ki}y_i + \alpha_{kj}y_j + u_k, \\ u_i; u_j; u_k &\sim i.i.d. (0; \sigma^2), \end{aligned} \tag{2.3.1}$$

es decir, el valor de y que se observa en la región i depende del valor de y en la región j y k , más un término aleatorio que se distribuye de manera idéntica e independiente entre las localizaciones. Lo mismo puede decirse de las otras regiones, j y k . En términos matriciales, la expresión (2.3.1) se reduce a:

$$\begin{aligned} \begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix} &= \begin{bmatrix} 0 & \alpha_{ij} & \alpha_{ik} \\ \alpha_{ji} & 0 & \alpha_{jk} \\ \alpha_{ki} & \alpha_{kj} & 0 \end{bmatrix} \begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix} + \begin{bmatrix} u_i \\ u_j \\ u_k \end{bmatrix}, \\ y &= Ay + u, \end{aligned} \tag{2.3.2}$$

donde

$$A = \begin{bmatrix} 0 & \alpha_{ij} & \alpha_{ik} \\ \alpha_{ji} & 0 & \alpha_{jk} \\ \alpha_{ki} & \alpha_{kj} & 0 \end{bmatrix}.$$

El problema con el sistema (2.3.2) es que, bajo un corte transversal, hay más parámetros que observaciones (3 observaciones y 6 parámetros). Puede imponerse simetría tal que $\alpha_{ij} = \alpha_{ji}$, $\alpha_{ik} = \alpha_{ki}$ y $\alpha_{jk} = \alpha_{kj}$ reduciendo el número de parámetros pero, de todas formas, no podrá estimarse (LeSage & Pace, 2009, pp. 8). Por lo tanto, es necesario buscar alguna solución que permita capturar la dependencia subyacente en los datos.

La solución más ampliamente utilizada en econometría espacial es imponer un conjunto de restricciones sobre las relaciones de dependencia. De esta forma, la estructura de la matriz A es reparametrizada de la siguiente manera:

$$A = \rho W,$$

³En un sentido estricto, dependencia espacial y autocorrelación espacial no son sinónimos. Sin embargo, la literatura ha tendido a utilizar ambos términos de manera intercambiable. Véase Anselin & Bera (1998) para mayor detalle.

donde ρ es un parámetro a estimar que captura el efecto espacial promedio de los vecinos y W es una matriz de pesos espaciales (también denominada matriz de contigüidades, ponderaciones, distancias o interacciones espaciales). La W se re-especifica como:

$$W = \begin{bmatrix} 0 & w_{ij} & w_{ik} \\ w_{ji} & 0 & w_{jk} \\ w_{ki} & w_{kj} & 0 \end{bmatrix},$$

en la cual se ha intercambiado a los parámetros $\alpha's$ (parámetros originales del modelo) por $w's$ (coeficientes exógenos al modelo).

Generalizando para una muestra de tamaño n , la matriz W será de orden $n \times n$. Cada elemento de W , denominado peso espacial (w_{ij}) captura la vecindad, siendo diferente de cero cuando las regiones i y j son consideradas vecinas. Por convención, ninguna región puede ser vecina de si misma dando como resultado una diagonal principal de W con todos sus elementos iguales a cero, $w_{ii} = 0$ ($i = 1, \dots, n$).

La matriz de pesos espaciales ocupa una posición central en econometría espacial ya que define el conjunto de vecinos para cada localización y su elección condicionará al resto del análisis. El problema central radica en cómo se construyen los pesos de la matriz W .

En la práctica econométrica, la matriz de contactos es construida mediante diferentes criterios. Estos criterios van desde el uso de la posición geográfica hasta el uso de flujos que capturan interacciones sociales y utilizan otras fuentes de información socio-económica.

En este punto se procederá de la forma tradicional utilizando criterios geográficos, siguiendo la idea de la primera ley de la geografía (Tobler, 1970): “Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes.” Entre los criterios geográficos, puede definirse vecindad por contigüidad, por alguna función de distancia, por k vecinos más cercanos o alguna combinación de estas opciones.

El criterio de contigüidad considera como vecinos a los polígonos que comparten límites. Hay varias opciones dentro de este criterio: tipo torre (vecinos al norte-sur y este-oeste), tipo reina (todos los polígonos que comparten límite). Estos nombres hacen alusión a los movimientos del ajedrez, bajo un mapa con cuadrícula regular.

El criterio de distancia utiliza como punto de referencia al centroide de cada polígono, identificado por la latitud y la longitud. Es posible construir diversas medidas, por ejemplo: (1) todas las unidades son vecinas mediante una función de distancia, tal como la inversa; o (2) vecinos son aquellas unidades que se encuentran a una distancia inferior a un determinado umbral, definido por un radio de distancia d .

El problema con los criterios de contigüidad y de distancia por umbral es que pueden dar lugar a regiones aisladas que no tienen vecinos. Esto puede suceder debido a que la densidad de los centroides no es regular en el mapa o, bajo contigüidad, cuando se trabaja con mapas que contienen islas.

En el caso del criterio de contigüidad, el problema de regiones sin vecinos puede salvarse asignando manualmente vecinos a estas unidades. Bajo el criterio de distancia por umbral, puede determinarse un umbral d cuya amplitud asegure que cada observación contenga al menos un vecino. Este criterio, conocido como max-min, determina el umbral como aquella máxima distancia

del vecino más cercano de todas las unidades. Un umbral demasiado elevado puede ocasionar que existan unidades con una excesiva cantidad de vecinos.

Otro criterio de uso habitual es el de k – vecinos más cercanos. Utilizando la información de los centroides, se va eligiendo como vecino al centroide más cercano hasta obtener el número de vecinos establecidos, k . Bajo este criterio todas las regiones poseerán la misma cantidad de vecinos evitando el problema de unidades aisladas o unidades con excesiva cantidad de vecinos.

Una vez definido el criterio de vecindad, se puede representar al conjunto de vecinos mediante una elección binaria, con $w_{i,j} = 1$ cuando i y j son vecinos, y $w_{i,j} = 0$ cuando no lo son. Si el criterio utilizado es el de distancia los pesos binarios pueden reemplazarse por alguna función de distancia entre unidades espaciales (puede combinarse la distancia, el perímetro u otras características geográficas).

Generalmente, luego de la elección de los pesos espaciales la matriz es normalizada por filas o columnas debido a que mejora las propiedades estadísticas de los estimadores y los tests. La transformación más utilizada es la normalización por fila, en donde los nuevos pesos se obtienen como $w_{ij}^* = w_{ij} / \sum_j w_{ij}$, tal que la suma de cada fila de W será sea igual a la unidad: $\sum_j w_{ij}^* = 1$.

Además, es posible considerar diferentes órdenes de vecindad. Por ejemplo, puede seleccionarse para cada región a los vecinos y a los vecinos de los vecinos. En el caso del criterio por umbral, los vecinos de los vecinos serán las regiones que se encuentran a una distancia mayor al umbral pero inferior al doble del umbral. No es de uso común utilizar órdenes superiores a 1, pero si fuese el caso, se suele distinguir mediante un supra-índice en la matriz: $W^{(j)}$, $\forall j \geq 2$, siendo j el orden de vecindad.

La elección del criterio de vecindad es habitualmente un *a priori* del investigador, existiendo cierta libertad en la elección del criterio adecuado. Sin embargo, es importante destacar que el tratamiento econométrico bajo las derivaciones de los diferentes estadísticos y los métodos de estimación suponen que W contiene elementos exógenos. Esto implica que la elección de los vecinos no debería realizarse mediante variables consideradas en el modelo analizado, ya sean endógenas o explicativas. La importancia de este punto es tal que afecta no solo a la estimación sino además a la interpretación de los efectos marginales de los modelos espaciales. Como se verá en la sección 5.3 del capítulo 5, si el peso espacial es función de la variable dependiente o de alguna variable explicativa, la derivación de los efectos marginales deberá incluir el cambio marginal de los pesos espaciales.

Griffith (1996) presenta un conjunto de lineamientos útiles para la elección de la matriz: (i) es mejor utilizar una especificación razonable de la matriz de pesos geográficos que considerar todas las conexiones nulas; (ii) modelos con bajos órdenes deben ser preferidos por sobre los modelos de altos órdenes; (iii) en general, es mejor emplear una matriz de pesos subidentificada que una sobreidentificada. El último punto se relaciona al sesgo descendente que generan las matrices altamente conectadas o densas sobre las estimaciones de los parámetros de los modelos espaciales (véase Mizruchi & Neuman, 2008, y Farber et al., 2009, para mayor detalle).

En Stata es posible generar matrices espaciales mediante diferentes comandos. Uno de ellos es `spmat` (Drukker et al., 2013b) que permite crear, importar, manipular y guardar matrices W . Las matrices son almacenadas como objetos `spmat`, estructura requerida para el uso de otros comandos como `spivreg` (Drukker et al., 2013c) y `spreg` (Drukker et al., 2013d).

El comando **spatwmat** (Pisati, 2001) está integrado a las herramientas de análisis exploratorio que incluye los tests de dependencia espacial global y, además, el análisis de dependencia local. Si la matriz ha sido generada por **spmat** no podrá ser usada directamente para los comandos del análisis exploratorio de Pisati.

Una tercera opción para matrices espaciales es **spwmatrix** de Jeanty (2014). Este comando permite generar matrices similares a **spmat** pero además genera matrices de k vecinos más cercanos y matrices con criterios socio-económicos. Tanto **spwmatrix** como **spmat** permiten importar matrices creadas por GeoDa, software que genera de forma muy simple las matrices de criterios geográficos. Además, las matrices generadas por **spwmatrix** pueden ser directamente utilizadas en los comandos de Pisati.

El comando **spmat** ofrece varias ventajas. Entre ellas, permite la visualización de las matrices así como alternativas avanzadas de especificación. Los detalles de este comando no se verán aquí, pero se encuentran debidamente explicados por Drukker et al. (2013b).

Veamos cómo trabajan estos comandos bajo Stata. Por ejemplo, puede generarse una matriz contigüidad tipo reina con el siguiente comando:

```
. spmat contiguity Wcont using "coord.dta", id(id)
warning: spatial-weighting matrix contains 5 islands
```

Tal como se comentó, el mapa de Europa contiene 5 islas que no tendrán vecinos usando este criterio. A primera vista, no parece ser un problema que algunas observaciones no contengan vecinos. Sin embargo, una primera crítica parte desde lo conceptual ya que la falta de vecinos implica suponer que estas unidades se comportan de forma independiente a las demás, algo que los mapas parecen contradecir. Otras críticas son más técnicas: (i) las observaciones sin vecinos no serán incluidas en el cómputo de los tests de autocorrelación espacial (la sección 2.4 aborda este punto); (ii) la falta de vecinos genera que no puedan calcularse los autovalores correspondientes de la matriz W , necesarios para estimar los modelos espaciales (este punto se verá en el próximo capítulo).

Como criterio alternativo se empleará el de k vecinos más cercanos. Este criterio es aplicado por el comando **spwmatrix**:

```
. spwmatrix gecon y_c x_c, wn(W5st) knn(5) row

Nearest neighbor (knn = 5) spatial weights matrix (164 x 164) calculated successfully
and the following action(s) taken:
- Spatial weights matrix created as Stata object(s): W5st.
- Spatial weights matrix has been row-standardized.
```

Se ha generado una matriz de 5 vecinos más cercanos, estandarizada por fila. Dado que esta matriz es creada como un objeto de Stata, no podrá ser utilizada bajo el comando **spmat**. Para solucionar este problema, puede generarse una matriz no estandarizada y exportarla a un archivo .txt con la idea de manipular su estructura. La secuencia de comandos es la siguiente:

```
. spwmatrix gecon y_c x_c, wn(W5bin) knn(5) xport(W5bin,txt)
```

Nearest neighbor (knn = 5) spatial weights matrix (164 x 164) calculated successfully and the following action(s) taken:

- Spatial weights matrix created as Stata object(s): W5bin.
- Spatial weights matrix saved to .txt file, C:\..\W5bin.txt, for use with other Stata packages.

Ahora, ya generado el archivo .txt, puede leerse como una base de datos de Stata y modificar el archivo para que sea legible en `spmat`:

```
. insheet using "W5bin.txt", delim(" ") clear
(165 vars, 165 obs)
. drop in 1
(1 observation deleted)
. rename v1 id
. save "W5bin.dta"
file W5bin.dta saved
. spmat dta W5_st v*, id(id) norm(row)
```

La matriz binaria es estandarizada por fila y guardada como objeto `spmat`. Esto permite aprovechar algunos subcomandos como `summarize` y `graph`:

```
. spmat summarize W5_st, links

Summary of spatial-weighting object W5_st
-----
```

Matrix	Description
Dimensions	164 x 164
Stored as	164 x 164

```
-----
Links
```

total	820
min	5
mean	5
max	5

```
-----
```

El número total de links es de 820, cerca del 3.07% de todos los posibles links. Además, puede obtenerse un gráfico de la matriz mediante el comando:

```
. spmat graph W5_st
(gráfico omitido)
```

2.4 Estadísticos globales de autocorrelación espacial

Existen diferentes estadísticos univariantes que permiten detectar autocorrelación espacial. Los mismos pueden clasificarse como medidas de dependencia globales o locales.

Las medidas globales utilizan la información completa del conjunto de datos con el fin de obtener un valor promedio para todo el espacio geográfico. Al resumir en un único valor toda la información, no permiten detectar la variabilidad de la dependencia ni la localización de estos patrones. Por su parte, las medidas locales examinan la autocorrelación espacial en un subconjunto de datos. Para cada localización brindan un valor particular y este valor puede ser contrastado estadísticamente clasificándolo en diferentes grupos, denominados clusters espaciales.⁴

Posiblemente el test más ampliamente utilizado sea el estadístico I de [Moran \(1950\)](#):

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j (y_i - \bar{y}) w_{ji} (y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (2.4.1)$$

donde $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} = 1'W1$, siendo 1 un vector $(n \times 1)$ de unos, con \bar{y} siendo la media muestral.

Si la matriz W contiene islas, el elemento S_0 será incorrectamente calculado.

Los momentos del estadístico de Moran, asumiendo que y se distribuye normal, son:

$$E[I] = -\frac{1}{n-1}, \quad (2.4.2)$$

$$V[I] = \frac{(3S_0^2 + S_1n^2 - nS_2)}{S_0(n+1)(n-1)} - \frac{1}{(n-1)^2}, \quad (2.4.3)$$

con $S_1 = (1/2) \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$ y $S_2 = \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji} \right)^2$. Es decir, bajo normalidad, los momentos del estadístico están determinados únicamente por elementos de la matriz espacial.

Su distribución asintótica es normal:

$$\sqrt{n}[I - E(I)] \underset{as}{\sim} \mathcal{N}[0; V(I)]. \quad (2.4.4)$$

Cuando el test I de Moran toma valores positivos existe autocorrelación positiva, implicando que regiones con altos (bajos) valores poseen vecinos con altos (bajos) valores respecto a la media. Valores negativos de I implican autocorrelación negativa tal que, en general, regiones con altos (bajos) valores poseen vecinos con bajos (altos) valores.

Otro test de uso habitual es el estadístico c de [Geary \(1954\)](#):

$$c = \frac{n-1}{2S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.4.5)$$

donde S_0 y \bar{y} ya han sido definidos previamente.

⁴La definición de cluster espacial hace referencia a un grupo de observaciones geográficamente limitado, con un tamaño y concentración suficiente, tal que es poco probable que haya ocurrido aleatoriamente.

Los momentos del estadístico de Geary, bajo la hipótesis nula de aleatoriedad, son:

$$E[c] = 1, \quad (2.4.6)$$

$$V[c] = \frac{(2S_1 + S_2)(n-1) - 4S_0^2}{2(n+1)S_0^2}, \quad (2.4.7)$$

usando la notación definida para los momentos del test I de Moran. Su distribución asintótica es normal:

$$\sqrt{n}[c-1] \underset{as}{\sim} \mathcal{N}[0; V(c)]. \quad (2.4.8)$$

Para el estadístico c , un valor negativo (es decir, por debajo del valor esperado) implica que existe autocorrelación positiva y un valor de c positivo, por encima del valor esperado, implica autocorrelación negativa.

Por último, otro estadístico de utilidad es el G de [Getis & Ord \(1992\)](#):

$$G = \frac{\sum_i \sum_{j \neq i}^n w_{ij} y_i y_j}{\sum_i \sum_{j \neq i}^n y_i y_j}, \quad (2.4.9)$$

donde cada elemento ya ha sido definido previamente. Bajo permutación, el valor esperado de G es:

$$E[G] = \frac{\sum_i \sum_{j \neq i}^n w_{ij}}{n(n-1)}.$$

La particularidad de este test es que necesita que los pesos, w_{ij} , sean de tipo binario. [Getis & Ord \(1992\)](#) muestran que la distribución de G tiende a la normal a medida que el tamaño muestral crece. De esta forma, puede utilizarse la aproximación asintótica de la distribución:

$$z[G] = \frac{G - E(G)}{\sqrt{V(G)}} \underset{as.}{\sim} \mathcal{N}(0, 1). \quad (2.4.10)$$

En Stata, estos estadísticos pueden ser calculados por el comando `spatgsa` de [Pisati \(2001\)](#).

```
. spatgsa U2012, w(W5st) moran geary two

Measures of global spatial autocorrelation
Weights matrix
-----
Name: W5st
Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: Yes
-----

Moran's I
-----
Variables |      I      E(I)    sd(I)      z    p-value*
-----
```

```

                U2012 |  0.767  -0.006   0.045  17.084   0.000
-----
Geary's c
-----
      Variables |      c      E(c)   sd(c)      z    p-value*
-----
                U2012 |  0.228   1.000   0.054 -14.282   0.000
-----
*2-tail test

```

Ambos estadísticos, el I de Moran como el c de Geary, detectan autocorrelación espacial de la tasa de desempleo de 2012 usando la matriz de contactos de 5 vecinos más cercanos. Lo mismo puede decirse sobre la tasa de migración neta.

```

. spatgsa NM2012, w(W5st) moran geary two

Measures of global spatial autocorrelation
Weights matrix
-----
Name: W5st
Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: Yes
-----
Moran's I
-----
      Variables |      I      E(I)   sd(I)      z    p-value*
-----
                NM2012 |  0.431  -0.006   0.046   9.563   0.000
-----
Geary's c
-----
      Variables |      c      E(c)   sd(c)      z    p-value*
-----
                NM2012 |  0.559   1.000   0.050  -8.840   0.000
-----
*2-tail test

```

Para obtener los resultados del estadístico de Getis y Ord, primero se genera mediante `spwmatrix` una matriz binaria y posteriormente se aplica el comando para ambas variables.

```

. spwmatrix gecon y_c x_c, wn(W5bin) knn(5)

Nearest neighbor (knn = 5) spatial weights matrix (164 x 164) calculated successfully
and the following action(s) taken:
- Spatial weights matrix created as Stata object(s): W5bin.

```

```
. spatgsa U2012, w(W5bin) go two
```

Measures of global spatial autocorrelation
Weights matrix

Name: W5bin
Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: No

Getis & Ord's G

Variables	G	E(G)	sd(G)	z	p-value*
U2012	0.039	0.031	0.001	11.864	0.000

*2-tail test

```
. spatgsa NM2012, w(W5bin) go two
```

Measures of global spatial autocorrelation
Weights matrix

Name: W5bin
Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: No

Getis & Ord's G

Variables	G	E(G)	sd(G)	z	p-value*
NM2012	0.066	0.031	0.004	9.441	0.000

*2-tail test

Alternativamente, el estadístico de Moran puede calcularse por el comando `splagvar` ([Jeanty, 2012](#)). Este comando ofrece la versión aleatorizada como alternativa a la aproximación asintótica, versión similar a la ofrecida por GeoDa. Otra ventaja de este comando es la posibilidad de generar el diagrama de dispersión de Moran, habitual en investigaciones aplicadas.

El diagrama de dispersión de Moran muestra la relación entre la variable y su rezago espacial, Wy . En vez de utilizar la variable original, es preferible trabajar con desviaciones respecto a la media:

$$z_i = (y_i - \bar{y}),$$

donde y_i , en este caso es: `U2012` o `NM2012`.

La relación entre la variable estandarizada y el rezago espacial de dicha variable puede representarse mediante una regresión lineal simple, sin constante:

$$W \cdot z_i = \beta \cdot z_i + u_i,$$

donde u_i es un término de error que cumple con los supuestos habituales. El coeficiente estimado $\hat{\beta}$ es igual al valor I de Moran:

$$\hat{\beta} = (X'X)^{-1} X'y = \frac{z'W \cdot z}{z'z} = \hat{I}.$$

Los resultados para ambas variables se obtienen aplicando `splagvar`, y utilizando el subcomando `plot`:

```
. splagvar U2012, wname(W5st) wfrom(Stata) ind(U2012) order(1) plot(U2012) moran(U2012)
(permute U2012 : splagvar_randper)
```

Moran's I Statistics Under Normal Approximation and Randomization Assumptions

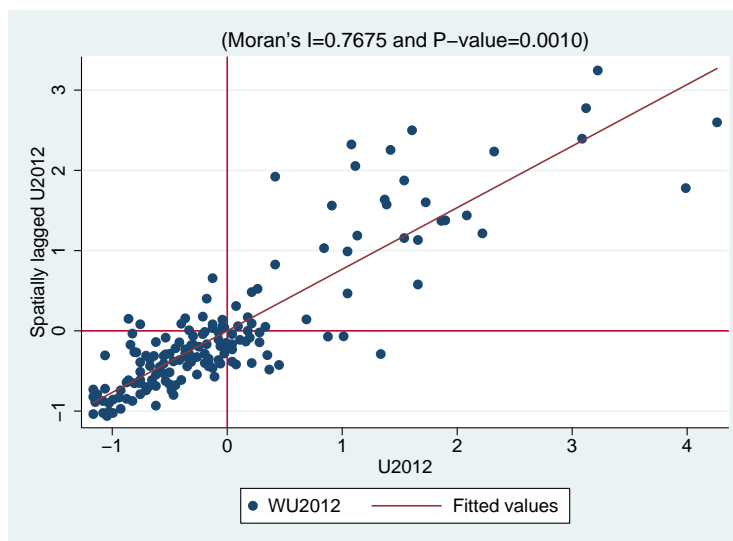
Statistics	Normal Approximation	Randomization
Moran's I	0.7675	0.7675
Mean	-0.0061	-0.0061
Std dev	0.0458	0.0453
Z-score	16.8866	17.0840
P-value*	0.0000	0.0000

*: Two-tailed test

Note: Under the random permutation procedure:

Mean = -0.0046 and Standard deviation = 0.0480

Spatially lagged variable(s) calculated successfully and/or all requests processed.



El diagrama de Moran divide al gráfico en cuatro cuadrantes respecto al valor promedio: valores altos de desempleo con valores alto de sus vecino (Alto-Alto), valores bajos de desempleo con valores bajos de sus vecinos (Bajo-Bajo), valores altos de desempleo con valores bajos de sus vecinos (Alto-Bajo), y valores bajos del desempleo con valores altos de los vecinos (Bajo-Alto).

Los resultados de la tasa migración neta se muestran a continuación:

```
. splotvar NM2012, wname(W5st) wfrom(Stata) ind(NM2012) order(1) plot(NM2012) moran(NM2012)
(permute NM2012 : splotvar_randper)
```

Moran's I Statistics Under Normal Approximation and Randomization Assumptions

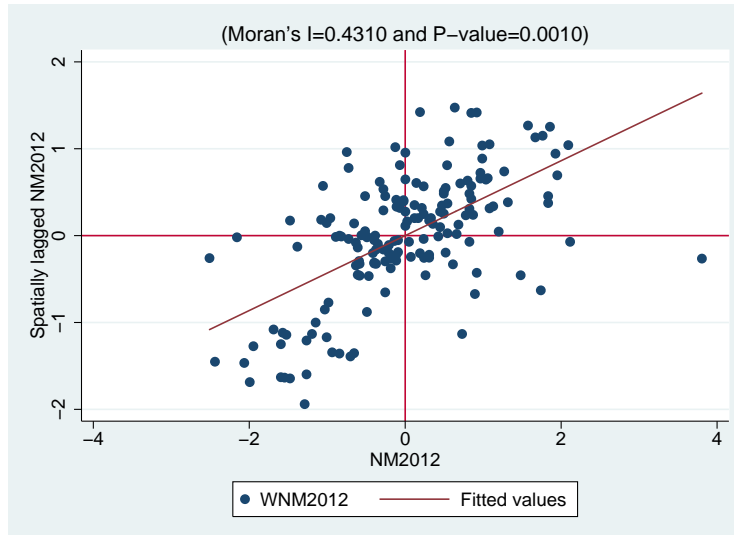
Statistics	Normal Approximation	Randomization
Moran's I	0.4310	0.4310
Mean	-0.0061	-0.0061
Std dev	0.0458	0.0457
Z-score	9.5417	9.5634
P-value*	0.0000	0.0000

*: Two-tailed test

Note: Under the random permutation procedure:

Mean = -0.0067 and Standard deviation = 0.0466

Spatially lagged variable(s) calculated successfully and/or all requests processed.



El estudio de la autocorrelación espacial puede ser extendido a la relación bivariada entre desempleo y el rezago espacial de la migración neta mediante el test I de Moran bivariado:

$$I_{yx} = \frac{n}{S_0} \frac{\sum_i \sum_j y_i w_{j \neq i} x_j}{\sqrt{\hat{\sigma}_y^2 \hat{\sigma}_x^2}}, \quad (2.4.11)$$

donde $\hat{\sigma}_{(\cdot)}^2$ representan las estimaciones de las varianzas de ambas variables. La hipótesis nula es aleatoriedad entre y y sus vecinos Wx .

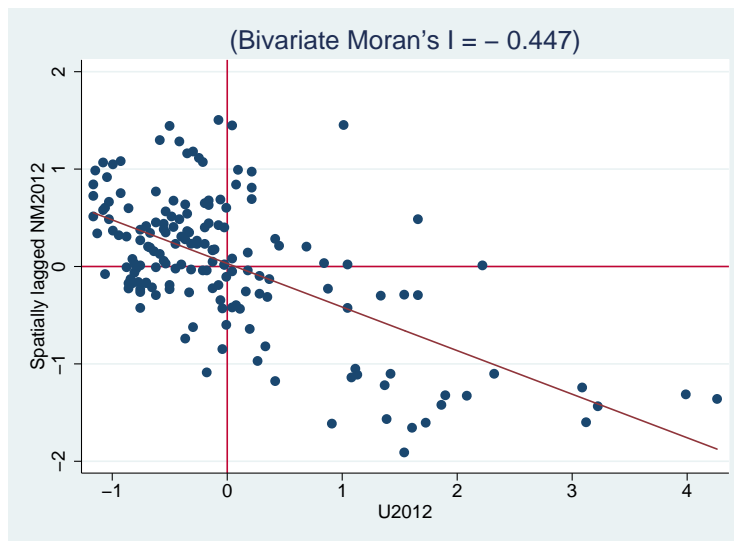
No hay un comando disponible en Stata para mostrar la significancia del test de Moran bivariado pero puede obtenerse el valor del estadístico bivariado mediante una regresión lineal. Como paso previo se deben generar las variables de la siguiente manera:

```
. egen z_U2012 = std(U2012)
. egen z_NM2012 = std(NM2012)
. splayvar z_NM2012, wname(W5st) wfrom(Stata) order(1)
```

La secuencia de comandos genera las variables estandarizadas para desempleo y migración neta. Luego, con el comando `splayvar` se genera el rezago espacial de migración neta.

Una vez construidas las variables, se realiza la estimación de regresión lineal y se grafica la relación entre las variables mediante un diagrama de dispersión. Como ya se ha explicado, el coeficiente de la relación lineal es similar al valor estimado del test I_{yx} . En este caso es negativo y confirma la relación visualizada previamente mediante los diferentes mapas.

```
. reg wy_z_NM2012 z_U2012
. twoway (scatter wy_z_NM2012 z_U2012) (lfit wy_z_NM2012 z_U2012, lc(maroon)), xline(0)
  yline(0) xtitle("U2012") ytitle("Spatially lagged NM2012")
  title("(Bivariate Moran's I = - 0.447)") legend(off)
```



Capítulo 3

Econometría espacial bajo corte transversal

En este capítulo se desarrollan los principales modelos de regresión con efectos espaciales para datos de corte transversal. Una clara diferencia respecto a lo expuesto en el capítulo anterior radica en el tratamiento asimétrico de las variables bajo estudio. El modelo de regresión supone que existe una variable dependiente que es de principal interés y que hay un conjunto de variables explicativas que condicionan su comportamiento.

En el campo aplicado, la inclusión de los efectos espaciales en modelos de regresión se ha dividido en dos grandes estrategias. Por un lado, se encuentra la especificación *data-driven* en donde la inclusión de las interacciones espaciales se justifica por particularidades de los datos, sin una teoría económica que justifique cada nuevo elemento. Los efectos espaciales pueden provenir de hipótesis surgidas del AEDE, presentado en el capítulo anterior, o por contraste de hipótesis estadísticas sobre la presencia o no efectos espaciales. Por otro lado, las especificaciones *theory-driven* son típicamente motivadas por modelos económicos teóricos a los que se incorporan interacciones entre los agentes sociales. En este capítulo presentaremos una combinación de ambas estrategias utilizando los datos del trabajo de [Ertur & Koch \(2007\)](#).

3.1 Resumen del modelo teórico de Ertur y Koch (2007)

[Ertur & Koch \(2007\)](#) investigan la interdependencia tecnológica en el crecimiento económico entre países. Basados en un modelo neoclásico, los autores consideran una función de producción Cobb-Douglas que refleja rendimientos constantes a escala respecto al trabajo y capital físico para el país i en el periodo temporal t :

$$Y_i(t) = A_i(t) K_i^\alpha(t) L_i^{1-\alpha}(t), \quad (3.1.1)$$

donde $Y_i(t)$ es el producto, $K_i(t)$ el stock de capital físico reproducible, $L_i(t)$ el nivel de trabajo, y $A_i(t)$ el nivel tecnológico agregado:

$$A_i(t) = \Omega(t) k_i^\phi(t) \prod_{j \neq i}^n A_j^{\gamma_{ij}}(t). \quad (3.1.2)$$

El nivel tecnológico agregado para el país i depende de tres términos. Primero, una proporción exógena de progreso tecnológico idéntica para todos los países: $\Omega(t) = \Omega(0) e^{\mu t}$, donde μ es la tasa de crecimiento constante. Segundo, un nivel agregado de capital físico por trabajador $k_i(t) = K_i(t)/L_i(t)$ propio de cada país, donde el parámetro ϕ ($0 \leq \phi < 1$) describe la fortaleza de las externalidades internas en el país. Tercero, un término que refleja el efecto externo de la acumulación de capital generado por otros países, con una intensidad decreciente debido a diferencias socio-económicas e institucionales ($0 \leq \gamma < 1$). La interdependencia tecnológica implica que los países no pueden analizarse de forma aislada.

Manipulando la expresión (3.1.2) e introduciendo el resultado en la ecuación (3.1.1), el producto per cápita en el país i puede expresarse como:

$$y_i(t) = \Omega^{\frac{1}{1-\gamma}}(t) k_i^{u_{ii}}(t) \prod_{j \neq i}^n k_j^{u_{ij}}(t), \quad (3.1.3)$$

con $y_i(t) = Y_i(t)/L_i(t)$, $u_{ii} = \alpha + \phi \left(1 + \sum_{r=1}^{\infty} \gamma^r w_{ii}^{(r)}\right)$ y $u_{ij} = \phi \sum_{r=1}^{\infty} \gamma^r w_{ij}^{(r)}$. Esta función se simplifica a la habitual si $\phi = 0$.

Para completar el modelo, los autores suponen que una fracción constante del producto s_i es ahorrado para inversión y que la fuerza laboral crece exógenamente a una tasa l_i para el país i . Además, se supone que la tasa de depreciación anual, δ , es constante e idéntica para todos los países. Dados estos supuestos, la ecuación dinámica fundamental de Solow es igual a:

$$\dot{k}_i(t) = s_i y_i(t) - (l_i + \delta) k_i(t), \quad (3.1.4)$$

donde el punto sobre la variable representa su derivada respecto al tiempo. Esta ecuación implica que la razón capital-producto del país i es constante y converge a una tasa de crecimiento balanceado definida por $\dot{k}_i(t)/k_i(t) = g$, o $[k_i/y_i]^* = s_i/(l_i + g + \delta)$, en otros términos:

$$k_i^* = \Omega^{\frac{1}{(1-\gamma)(1-u_{ii})}}(t) \left(\frac{s_i}{l_i + g + \delta} \right)^{\frac{1}{(1-u_{ii})}} \prod_{j \neq i}^n k_j^{* \frac{u_{ij}}{(1-u_{ii})}}(t). \quad (3.1.5)$$

El capital per cápita en estado estacionario depende de los parámetros habituales así como del nivel del capital per cápita de los países vecinos.

Reescribiendo la función logarítmica del producto en forma matricial: $y = A + \alpha k$, y sustituyendo A por su expresión se obtiene:

$$y = \Omega + (\alpha + \phi) k - \alpha \gamma W k + \gamma W y. \quad (3.1.6)$$

Introduciendo en la última ecuación la razón capital-producto del estado estacionario en logaritmos, se obtiene la especificación de la renta real per cápita del país i a lo largo de la trayectoria de crecimiento balanceado, para un determinado periodo:

$$\begin{aligned} \ln y_i &= \frac{1}{1-\alpha-\phi} \ln \Omega + \frac{\alpha+\phi}{1-\alpha-\phi} \ln s_i - \frac{\alpha+\phi}{1-\alpha-\phi} \ln(l_i + g + \delta) \\ &\quad - \frac{\alpha\gamma}{1-\alpha-\phi} \sum_{j \neq i}^n w_{ij} \ln s_j + \frac{\alpha\gamma}{1-\alpha-\phi} \sum_{j \neq i}^n w_{ij} \ln(l_j + g + \delta) \\ &\quad + \frac{\gamma(1-\alpha)}{1-\alpha-\phi} \sum_{j \neq i}^n w_{ij} \ln y_j. \end{aligned} \quad (3.1.7)$$

De esta forma, la especificación final es un modelo conocido en la literatura como modelo espacial de Durbin (SDM):

$$y = X\beta + WX\theta + \lambda Wy + \varepsilon. \quad (3.1.8)$$

Los autores utilizan a la teoría como guía para especificar el modelo econométrico. Entre los principales resultados se encuentra contrastar las siguientes restricciones teóricas entre los coeficientes: $\beta_1 = -\beta_2 = \frac{\alpha+\phi}{1-\alpha-\phi}$ y $\theta_2 = -\theta_1 = \frac{\alpha\gamma}{1-\alpha-\phi}$ y recuperar los parámetros estructurales del modelo: α , ϕ y γ .

Datos a utilizar y matriz de pesos espaciales

Los autores utilizan datos de 91 países para el año 1995. Las variables utilizadas son:

- $\ln y_{95}$ es el logaritmo de la renta real per cápita computada como el PBI real dividido por la fuerza laboral.
- $\ln(l_i + 0.05)$ (l es expresado como n en el artículo original) es el logaritmo de la tasa de crecimiento medio de la población en edad de trabajo (de 15 a 64 años) más 0.05 que representan a la tasa de crecimiento balanceado y la tasa de depreciación ($g + \delta$), tal que como es habitual en la literatura.
- $\ln s$ es el logaritmo de la tasa de ahorro la cual es medida como la participación media de la inversión bruta en el PBI.

Los autores utilizan el criterio de distancia para construir dos matrices de pesos espaciales alternativas. Para simplificar el desarrollo del capítulo utilizaremos una sola matriz de las presentadas en el trabajo original. Esta matriz es construida considerando la distancia entre capitales, con pesos igual a

$$w_{ij} = \begin{cases} d_{ij}^{-2} & \text{si } i \neq j \\ 0 & \text{en otro caso} \end{cases}, \quad (3.1.9)$$

donde d_{ij} es la distancia del gran-círculo¹ entre capitales de países.

3.2 Detección de estructura espacial bajo MCO

A pesar del modelo teórico planteado en la sección previa. [Ertur & Koch \(2007\)](#) utilizan la estrategia de estimar un modelo simple e ir incorporando términos espaciales según el rechazo de hipótesis nulas. Para realizar este procedimiento se comienza con el modelo estático más simple

¹La distancia del gran círculo es la distancia más corta entre dos puntos sobre la superficie de una esfera medida a lo largo de la superficie de la esfera. La fórmula habitual del gran-círculo es:

$$d_{ij} = \text{radio} \times \cos^{-1} [\cos |long_i - long_j| \cos lat_i \cos lat_j + \sin lat_i \sin lat_j],$$

donde radio es el radio terrestre, y lat y $long$ son respectivamente latitud y longitud para las unidades de observación i y j .

que considera una variable dependiente y un conjunto de variables explicativas. La ecuación de trabajo es la siguiente:

$$\begin{aligned} y &= X\beta + u, \\ u &\sim \mathcal{N}(0, \sigma^2 I_n), \end{aligned} \quad (3.2.1)$$

siendo la variable dependiente y un vector de dimensión $(n \times 1)$, X es una matriz de variables explicativas, incluyendo una constante, de orden $(n \times k)$, β es un vector de parámetros desconocidos de orden $(k \times 1)$ y u es el término de error de dimensión $(n \times 1)$. Se supone que el modelo cumple con los supuestos habituales de la parte sistemática: forma funcional correcta, variables explicativas exógenas, no multicolinealidad.

La ecuación (3.2.1) es la hipótesis nula de los estadísticos de dependencia espacial. Típicamente, el término de error es supuesto independiente e idénticamente distribuido, *i.i.d.*, pero claramente este no será el caso bajo la hipótesis alternativa. El objetivo es encontrar evidencia tal que permita mantener la hipótesis nula de no autocorrelación espacial. En caso contrario, se incorporarán elementos espaciales al modelo.

La presencia de estructura espacial en el modelo (3.2.1) puede contrastarse en base a estadísticos simples que utilizan los resultados de la estimación por mínimo cuadrados ordinarios (*MCO*). El primer test que puede utilizarse es el test I de Moran aplicado sobre los residuos *MCO* del modelo (3.1), sugerido por [Cliff & Ord \(1972\)](#):

$$I = \frac{n}{S_0} \frac{\hat{u}' W \hat{u}}{\hat{u}' \hat{u}}, \quad (3.2.2)$$

donde \hat{u} es el vector de residuos *MCO*, n es el número de observaciones y S_0 es la suma de todos los elementos de W , tal como se ha definido en el capítulo previo. Bajo una matriz estandarizada por fila, el término $n/S_0 = 1$ y puede ser ignorado.

La distribución probabilística del estadístico I de Moran es desconocida para muestras finitas por lo que comúnmente se utiliza una aproximación empírica por permutación. Otra opción es plantear la inferencia estadística bajo el supuesto de normalidad asintótica.

Recuérdese que la hipótesis nula del test es no autocorrelación espacial. El problema con el test I de Moran es que el rechazo de la hipótesis nula no brinda información sobre el posible modelo a especificar. La hipótesis alternativa es general y no da una guía clara sobre el tipo de estructura espacial que se encuentra en el proceso generador de datos.

Como alternativa al test I de Moran, existe un conjunto de tests de Multiplicadores de Lagrange, *LM*, que resultan de la aplicación del principio de máxima verosimilitud. Estos tests tienen la ventaja de que la hipótesis alternativa se encuentra bien definida o restringida.

Una primera hipótesis alternativa proviene de plantear la presencia de autocorrelación espacial en el término de error. Esta estructura de dependencia puede deberse a un proceso autoregresivo o a uno de medias móviles. Sin pérdida de generalidad, se concentrará el análisis en la estructura autoregresiva dado que ante ambas alternativas el estadístico es el mismo ([Burridge, 1980](#)). Es así que un modelo de error espacial autoregresivo supone que el término de error aleatorio de (3.1) obedece a un proceso como:

$$u = \rho W u + \varepsilon, \quad (3.2.3)$$

donde ρ es el parámetro espacial autoregresivo, W es una matriz de pesos espaciales no estocástica de orden $(n \times n)$ y ε es un vector de innovaciones con media nula y varianza constante $(0, \sigma^2 I)$.

El test LM para detectar autocorrelación espacial en el error establece las siguientes hipótesis nula y alternativa:

$$\begin{aligned} H_0 : & \quad \rho = 0, \\ H_1 : & \quad \rho \neq 0, \end{aligned}$$

siendo su fórmula:

$$LM_{ERROR} = \frac{1}{T_1} \left(\frac{\hat{u}' W \hat{u}}{\hat{\sigma}^2} \right)^2 \underset{as}{\sim} \chi_{(1)}^2, \quad (3.2.4)$$

donde T_1 es igual a $\text{tr}[(W' + W)W]$ y \hat{u} son los residuos MCO y $\hat{\sigma}^2 = \hat{u}'\hat{u}/n$.

Una segunda hipótesis alternativa proviene de plantear un modelo con estructura espacial sustantiva, modelo de rezago espacial (SLM , spatial lag model). Este modelo incorpora un rezago espacial de la variable dependiente, Wy , como explicativa:

$$y = \lambda Wy + X\beta + u, \quad (3.2.5)$$

donde λ es un parámetro espacial autoregresivo, u es un vector de errores de dimensión $(n \times 1)$ y $u \sim (0, \sigma^2 I_n)$, con los demás elementos ya definidos.

El test para la detección de dependencia sustantiva plantea como hipótesis nula y alternativa:

$$\begin{aligned} H_0 : & \quad \lambda = 0, \\ H_1 : & \quad \lambda \neq 0, \end{aligned}$$

tal que el test tiene la siguiente estructura:

$$LM_{LAG} = \frac{\left(\frac{\hat{u}' Wy}{\hat{\sigma}^2} \right)^2}{n \hat{J}_{\lambda\beta}} \underset{as}{\sim} \chi_{(1)}^2, \quad (3.2.6)$$

con $\hat{J}_{\lambda\beta} = 1/n\hat{\sigma}^2 \left[(WX\hat{\beta})' M (WX\hat{\beta}) + T_1 \hat{\sigma}^2 \right]$, con $M = I - X(X'X)^{-1}X'$.

Estos tests LM presentan como inconveniente que son sensibles a diferentes tipos de errores de especificación. Por ejemplo, el LM_{ERROR} detecta autocorrelación espacial debido a la presencia de un rezago espacial de la variable endógena (Wy), y lo mismo puede decirse del LM_{LAG} , que da falsos positivos cuando el término de error contiene un rezago espacial (Wu).

En este contexto, [Anselin et al. \(1996\)](#) proponen dos nuevos multiplicadores de Lagrange diseñados para que su comportamiento sea robusto a dichos errores de especificación. El LM_{ERROR}^* analiza la falta de correlación en los residuos, siendo robusto a la omisión del término Wy :

$$LM_{ERROR}^* = \frac{\left[\left(\frac{\hat{u}' W \hat{u}}{\hat{\sigma}^2} \right) - T_1 \left(n \hat{J}_{\lambda\beta} \right)^{-1} \left(\frac{\hat{u}' W \hat{u}}{\hat{\sigma}^2} \right) \right]^2}{\left[T_1 - T_1^2 \left(n \hat{J}_{\lambda\beta} \right)^{-1} \right]} \underset{as}{\sim} \chi_{(1)}^2. \quad (3.2.7)$$

De igual manera, la versión robusta del test LM_{LAG} , LM_{LAG}^* , permite detectar la autocorrelación espacial sustantiva en presencia de estructura espacial en el término de error:

$$LM_{LAG}^* = \frac{\left[\left(\frac{\hat{u}' Wy}{\hat{\sigma}^2} \right) - \left(\frac{\hat{u}' W \hat{u}}{\hat{\sigma}^2} \right) \right]^2}{n \hat{J}_{\lambda\beta} - T_1} \underset{as}{\sim} \chi_{(1)}^2. \quad (3.2.8)$$

Para mayor detalle de la derivación de estos tests puede consultarse a [Anselin & Florax \(1995, pp. 23-26\)](#). El conjunto de tests habituales para la detección de autocorrelación espacial se resume en el Cuadro 3.1.

Estos tests permiten incorporar elementos espaciales de acuerdo al rechazo o no de cada una de las hipótesis nulas. Tal es así que puede establecerse una estrategia de especificación:

- Si I de Moran rechaza $H_0 \rightarrow$ Evidencia a favor de inclusión de elementos espaciales.
- Si LM_{ERROR} y LM_{ERROR}^* rechazan $H_0 \rightarrow$ Evidencia a favor de un modelo de error espacial (SEM).
- Si LM_{LAG} y LM_{LAG}^* rechazan $H_0 \rightarrow$ Evidencia a favor de un modelo de rezago espacial (SLM).
- Si no se rechaza H_0 bajo ninguno de los tests \rightarrow Evidencia a favor del modelo lineal general no espacial.
- Si ambos tests robustos, LM_{ERROR}^* y LM_{LAG}^* , rechazan H_0 entonces se deberán incorporar elementos espaciales en la parte sistemática y aleatoria.

Estos resultados son direcciones claras que ofrece la lectura conjunta de los tests.

Cuadro 3.1: Estadísticos espaciales

Hipótesis Nula, H_0	Parámetros en H_1		Estadístico
	Rezago espacial	Error espacial	
$\rho = 0$	-	si	LM_{ERROR}
	si	si	LM_{ERROR}^*
$\lambda = 0$	si	-	LM_{LAG}
	si	si	LM_{LAG}^*
No autocorrelación espacial			I de Moran

Ahora, se verá cómo obtener estos resultados en Stata. El modelo (3.2.1) se puede estimar por el comando **reg** que implementa la estimación por mínimos cuadrados ordinarios. Usando los datos de [Ertur & Koch \(2007\)](#), el modelo de Solow básico no considera la posible dependencia espacial tal que la especificación inicial es:

$$\ln y_{95i} = \beta_0 + \beta_1 \ln s_i + \beta_2 \ln(l_i + 0.05) + \varepsilon_i. \quad (3.2.9)$$

En Stata, la estimación se realiza de la siguiente manera:

```
. use ertur_koch_data, clear
. reg lny95 lns lnlgd
```

Source	SS	df	MS	Number of obs =	91
-----+-----				F(2, 88) =	74.02
Model	84.6761929	2	42.3380964	Prob > F =	0.0000
Residual	50.3310525	88	.571943778	R-squared =	0.6272

-----+-----					Adj R-squared = 0.6187	
Total		135.007245	90	1.5000805	Root MSE	= .75627
-----+-----						
lny95		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
lns		1.275585	.1251472	10.19	0.000	1.026881 1.524288
lnlgd		-2.708721	.6423168	-4.22	0.000	-3.98519 -1.432251
_cons		4.651014	1.772881	2.62	0.010	1.127786 8.174242
-----+-----						

Los resultados son los esperados de acuerdo a la teoría y equivalentes a los presentados por [Ertur & Koch \(2007, Tabla 1, pp. 1047\)](#).

Una vez estimado el modelo, se procede al diagnóstico de los residuos mediante el comando `spatdiag` de [Pisati \(2001\)](#). La implementación del comando requiere una matriz de contactos. Siguiendo con lo planteado en la especificación teórica, se utilizará la matriz de pesos usando la función inversa de la distancia al cuadrado, estandarizada por filas:

```
. spwmatrix gecon ylat xlong, wname(invdist2_st) wtype(inv) alpha(2) row

Inverse distance (alpha = 2) spatial weights matrix (91 x 91) calculated successfully
and the following action(s) taken:
- Spatial weights matrix created as Stata object(s): invdist2_st.
- Spatial weights matrix has been row-standardized.
```

Usando la matriz de contactos creada, los tests espaciales se obtienen con la siguiente secuencia:

```
. spatdiag, weights(invdist2_st)

Diagnostic tests for spatial dependence in OLS regression
Fitted model
-----
lny95 = lns + lnlgd
-----

Weights matrix
-----
Name: invdist2_st Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: Yes
-----

Diagnostics
-----
Test | Statistic df p-value
-----+-----
Spatial error: |
Moran's I | 6.860 1 0.000
Lagrange multiplier | 38.873 1 0.000
```

Robust Lagrange multiplier		2.512	1	0.113
Spatial lag:				
Lagrange multiplier		52.406	1	0.000
Robust Lagrange multiplier		16.044	1	0.000

Los resultados reportados muestran que existe autocorrelación espacial en los residuos según el test I de Moran, LM_{ERROR} y el LM_{LAG} . Ambas versiones del test de dependencia sustantiva detectan estructura espacial. No sucede lo mismo con los LM del error, ya que el LM_{ERROR}^* no rechaza la hipótesis de $\rho = 0$. La lectura conjunta de estos resultados implica que debe estimarse un modelo con autorrelación espacial sustantiva, modelo SLM .

Obsérvese que el comando de Stata ubica al test I de Moran como un estadístico de error espacial. Esto no implica que un valor significativo del estadístico brinde evidencia de estructura espacial en el error. Tal como ya se ha mencionado, la hipótesis alternativa del I de Moran no se encuentra restringida a un modelo específico y un valor significativo del test puede sugerir una o ambas estructuras espaciales.

Desde la teoría, puede establecer la restricción: $\beta_1 = -\beta_2$ dando lugar al modelo restringido:

$$\ln y_{95i} = \beta_0 + \beta_1 [\ln s_i - \ln (l_i + 0.05)] + \varepsilon_i. \quad (3.2.10)$$

En Stata, puede estimarse de la siguiente manera:

```
. gen lnslgd = lns - lnlgd
. reg lny95 lnslgd
```

Source	SS	df	MS	Number of obs = 91		
-----+-----				F(1, 89) = 138.30		
Model	82.1440506	1	82.1440506	Prob > F = 0.0000		
Residual	52.8631948	89	.59396848	R-squared = 0.6084		
-----+-----				Adj R-squared = 0.6040		
Total	135.007245	90	1.5000805	Root MSE = .77069		
-----+-----						
lny95	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
lnslgd	1.379085	.1172694	11.76	0.000	1.146073	1.612097
_cons	8.374839	.1066275	78.54	0.000	8.162973	8.586706

Del modelo restringido se obtiene una estimación de la participación del capital la cual es cercana al reportado por diferentes estudios, en orden del 0.6:

$$\hat{\beta}_1 = \alpha/(1-\alpha) \Rightarrow \hat{\alpha} = \hat{\beta}_1/(1+\hat{\beta}_1) \cong 0.58.$$

```
. display _b[lnslgd]/(1+_b[lnslgd])
.57967039
```

Los resultados obtenidos de los tests de autocorrelación espacial, tal como sucedía en el modelo no-restringido, muestran evidencia de autocorrelación espacial sustantiva.

```
. spatdiag, weights(invdist2_st)
```

Diagnostic tests for spatial dependence in OLS regression
Fitted model

lny95 = lns + lnlgd

Weights matrix

Name: invdist2_st Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: Yes

Diagnostics

Test	Statistic	df	p-value
Spatial error:			
Moran's I	6.498	1	0.000
Lagrange multiplier	35.979	1	0.000
Robust Lagrange multiplier	1.387	1	0.239
Spatial lag:			
Lagrange multiplier	54.281	1	0.000
Robust Lagrange multiplier	19.688	1	0.000

La detección de efectos espaciales lleva a desestimar el uso de MCO a favor de otros métodos más complejos de estimación. Sin embargo, un reciente trabajo de [Halleck Vega & Elhorst \(2015\)](#) aconsejan no abandonar tan rápidamente la estimación por *MCO*. Un modelo espacial que puede estimarse por la metodología tradicional es el llamado *SLX*, que incorpora efectos espaciales en las variables explicativas:

$$\begin{aligned} y &= X\beta + WX\theta + u, \\ u &\sim \mathcal{N}(0, \sigma^2 I_n), \end{aligned} \quad (3.2.11)$$

donde θ captura el efecto espacial de las variables explicativas de los vecinos a cada observación. Estos efectos son conocidos como contextuales en la literatura de modelos sociales y tienen un impacto local.

En Stata, para estimar este modelo se necesita calcular las variables WX . Existen varias alternativas para realizar este cálculo, uno de los más directos es utilizar el comando `splagvar`:

```
. splagvar, wname(invdist2_st) ind(lns lnlgd) wfrom(Stata) order(1)
```

Spatially lagged variable(s) calculated successfully and/or all requests processed.

Usando estas nuevas variables, simplemente se repite la secuencia de estimación antes utilizada:

```
. reg lny95 lns lnlgd wx_lns wx_lnlgd
```

Source	SS	df	MS	Number of obs = 91		
-----+-----				F(4, 86)	=	49.18
Model	93.939918	4	23.4849795	Prob > F	=	0.0000
Residual	41.0673274	86	.477527062	R-squared	=	0.6958
-----+-----				Adj R-squared	=	0.6817
Total	135.007245	90	1.5000805	Root MSE	=	.69103

lny95	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
lns	.9452114	.1369812	6.90	0.000	.6729017	1.217521
lnngd	-1.55606	.7785809	-2.00	0.049	-3.103828	-.0082928
wx_lns	.9206197	.220137	4.18	0.000	.4830018	1.358238
wx_lnngd	-.7802644	1.139519	-0.68	0.495	-3.045552	1.485023
_cons	6.850748	2.594039	2.64	0.010	1.693969	12.00753

```
. spatdiag, weights(invdist2_st)
```

Diagnostic tests for spatial dependence in OLS regression

Fitted model

```
lny95 = lns + lnlgd + wx_lns + wx_lnlgd
```

Weights matrix

Name: invdist2_st Type: Distance-based (binary)

Distance band: c1.c2 < d <= c3.c4

Row-standardized: Yes

Diagnostics

Test	Statistic	df	p-value
-----+-----			
Spatial error:			
Moran's I	5.937	1	0.000
Lagrange multiplier	27.688	1	0.000
Robust Lagrange multiplier	0.793	1	0.373
Spatial lag:			
Lagrange multiplier	36.299	1	0.000
Robust Lagrange multiplier	9.403	1	0.002

Los resultados corroboran la necesidad de incorporar como variable explicativa un rezago espacial endógena alcanzando la especificación sugerida por la teoría.

En trabajos aplicados como el de [Ertur & Koch \(2007\)](#), lo habitual será estimar el modelo indicado por los tests, es decir, un modelo *SLX* con un rezago espacial endógeno dando lugar a un modelo espacial derivado por la teoría, *SDM*. Sin embargo, debido a que esta guía pretende cubrir las opciones de especificación más habituales en econometría espacial, se estimará todo el conjunto de modelos espaciales para que el lector pueda conocer los comandos disponibles en Stata.

Una vez detectada la posible estructural espacial, surge el problema de cómo estimar este tipo de modelos. Lamentablemente, la estimación de los modelos que incluyen como variables Wy o Wu no pueden realizarse por *MCO* debido a que las estimaciones serán inconsistentes y/o ineficientes, dependiendo del caso ([Anselin, 1988](#), pp. 57-59).

Las alternativas de estimación han sido tradicionalmente de dos tipos. Una alternativa es por máxima verosimilitud, *MV*, suponiendo distribución normal de la perturbación aleatoria. La otra alternativa es por variables instrumentales o, su versión generalizada, método de momentos generalizado, *GMM*, que recurre a la teoría asintótica sin necesidad del supuesto de normalidad.

3.3 Estimación por máxima verosimilitud

La estimación por máxima verosimilitud ha sido extensamente tratada. Aquí solo se presentarán los elementos más importantes de dicho procedimiento para un conjunto de modelos espaciales. Para una mayor revisión de este método puede consultarse a [LeSage & Pace \(2009\)](#), capítulo 3).

Modelo de rezago espacial

El modelo de rezago espacial o *SLM* tiene la siguiente estructura:

$$\begin{aligned} y &= \lambda Wy + X\beta + u, \\ u &\sim \mathcal{N}(0, \sigma^2 I_n), \end{aligned} \quad (3.3.1)$$

donde el supuesto de normalidad del término de error permite plantear la habitual función de log-verosimilitud no-concentrada, esto es:

$$l(y|\varphi) = -\frac{n}{2} \ln(2\pi\sigma^2) + \ln|A| - \frac{1}{2\sigma^2} [(Ay - X\beta)'(Ay - X\beta)], \quad (3.3.2)$$

siendo $A = I - \lambda W$.

Existen importantes detalles a destacar en la maximización de la función (3.3.2). Uno de ellos es la necesidad de imponer alguna restricción sobre el rango de valores que puede alcanzar λ . Sin restricciones sobre este parámetro, el espacio paramétrico de la función será discontinuo e inestable. Para evitar este problema y también facilitar la estimación de λ , [Ord \(1981\)](#) propuso restringir los posibles valores del parámetro en un rango $1/\tau_{min} < \lambda < 1/\tau_{max}$, donde τ_{min} y τ_{max} son los autovalores, mínimo y máximo, de la matriz W . Bajo una matriz espacial estandarizada por filas automáticamente se impone una restricción sobre el máximo autovalor tal que λ se encontrará restringido entre los valores $1/\tau_{min} < \lambda < 1$. Algunos autores sugieren limitar el espacio paramétrico

de λ al intervalo $(-1, 1)$. Para mayor revisión puede consultarse a [Elhorst \(2014, pp. 13-17\)](#), [LeSage & Pace \(2009, capítulo 4\)](#) y [Kelejian & Prucha \(2010\)](#).

Otro detalle de importancia en la maximización de la función (3.3.2) involucra al determinante de una matriz de dimensión $n \times n$, $\ln |I - \lambda W|$. Este determinante debe ser evaluado repetidamente para cada valor estimado de λ y el procedimiento puede ser problemático para grandes bases de datos. Una solución propuesta por [Ord \(1975\)](#) es estimar dicho término mediante la siguiente transformación:

$$\ln |I - \lambda W| = \sum_{i=1}^n \ln (1 - \lambda \tau_i),$$

siendo τ_i los autovalores de la matriz W .

Para maximizar la función (3.3.2) se necesitan obtener las condiciones necesarias de primer orden. Estas condiciones conforman un sistema no lineal, imposibilitando la obtención de una solución analítica. La estimación de este sistema puede realizarse por medio de algoritmos numéricos o por el procedimiento basado en la log-verosimilitud concentrada. La aplicación de este último procedimiento es más simple, condicionando a los estimadores máximo-verosímiles respecto al parámetro de dependencia λ :

$$\begin{aligned} \beta_\lambda &= [X'X]^{-1} X' Ay, \\ &= [X'X]^{-1} X' y - \lambda [X'X]^{-1} X' W y = \beta_{mco} - \lambda \beta_{Wy, mco}, \end{aligned} \quad (3.3.3)$$

$$\sigma_\lambda^2 = \frac{\tilde{u}'\tilde{u}}{n} = \frac{1}{n} [u_{mco} - \lambda u_{Wy, mco}]' [u_{mco} - \lambda u_{Wy, mco}], \quad (3.3.4)$$

siendo $\beta_{mco} = [X'X]^{-1} X' y$, $\beta_{Wy, mco} = [X'X]^{-1} X' W y$, $u_{mco} = y - X\beta_{mco}$ y $u_{Wy, mco} = W y - X\beta_{Wy, mco}$.

Sustituyendo estos resultados en la función (3.3.2), se obtiene la función de log-verosimilitud concentrada que depende únicamente del parámetro λ . Esta función es optimizada mediante iteraciones hasta alcanzar convergencia. Posteriormente, el valor $\hat{\lambda}$ es utilizado para obtener estimaciones de las ecuaciones (3.3.3)-(3.3.4). Los errores estándares de los parámetros se pueden calcular a partir del Hessiano aproximado por una función de optimización numérica o por el cálculo analítico asintótico.

En Stata, el modelo *SLM* puede estimarse por el comando `spreg ml`, sub-opción `dlmat(Winv2_st)` ([Drukker et al., 2013c](#)). Para poder aplicar este comando debe realizarse un paso previo que es convertir la matriz de pesos espaciales en formato `.spmat`:

```
. spwmatrix gecon ylat xlong, wname(invdist2) wtype(inv) alpha(2) xport(Winv2,txt)
. insheet using "Winv2.txt", delim(" ") clear
. drop in 1
. rename v1 id
. save "Winv2.dta", replace
. spmat dta Winv2_st v*, id(id) norm(row)
```

Una vez generada la matriz, procedemos a aplicar el comando:

```
. spreg ml lny95 lns lnlgd, id(id) dlmat(Winv2_st)
```

```

Performing a grid search.... finished
Iteration 0:   log likelihood = -74.540178
Iteration 1:   log likelihood = -74.371247
Iteration 2:   log likelihood = -74.37095
Iteration 3:   log likelihood = -74.37095
Optimizing unconcentrated log likelihood
Iteration 0:   log likelihood = -74.37095

Spatial autoregressive model                                Number of obs   =       91
(Maximum likelihood estimates)                             Wald chi2(2)    =   66.7325
                                                            Prob > chi2     =    0.0000
-----+-----
              lny95 |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
lny95          |
              lns |   .7771338   .100483    7.73   0.000    .5801907    .9740768
              lnlgd | -1.208138   .4683733   -2.58   0.010   -2.126133   -2.2901429
              _cons |   1.722529   1.256213    1.37   0.170   -.739602    4.184661
-----+-----
lambda          |
              _cons |   .6403173   .0667183    9.60   0.000    .509552    .7710827
-----+-----
sigma2          |
              _cons |   .2702196   .0407237    6.64   0.000    .1904026    .3500366
-----+-----

```

A la hora de interpretar los resultados es importante tener presente la forma reducida del modelo *SLM*: $E(y) = (I - \lambda W)^{-1} X\beta$. Este valor esperado contiene como elemento diferencial respecto al modelo de regresión lineal al componente $(I - \lambda W)^{-1}$.

Considerando que $|\lambda| < 1$ y $w_{ij} < 1$ (matriz estandarizada), la matriz inversa puede expandirse en una serie infinita conocida como “multiplicador espacial”:

$$(I - \lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \dots,$$

que introducido en el valor esperado del *SLM* genera:

$$E(y) = (I - \lambda W)^{-1} X\beta = X\beta + \lambda W X\beta + \lambda^2 W^2 X\beta + \dots.$$

Este resultado permite destacar algunos puntos de interés:

- El resultado en una localización i no solo será afectado por las variables exógenas de i , sino que será retroalimentado por las otras localizaciones a través de la inversa de la transformación espacial $(I - \lambda W)^{-1}$. Este efecto irá decayendo debido a $|\lambda| < 1$.
- Un shock aleatorio en una localización i no solo afectará el resultado de i , además tendrá un impacto en todas las localizaciones.

En términos del modelo de crecimiento, este resultado implica: (1) en media condicional, la renta per cápita por trabajador en el país i no solo será afectada por la tasa de inversión y la tasa efectiva

de capital físico del propio país i , además será afectada por el valor de estas variables en todos los demás países a través del multiplicador espacial, denominado efecto de interacción global; (2) un shock aleatorio en un país i no solo afectará a la renta per cápita de ese país sino que se distribuirá a todo el conjunto de países debido a la misma transformación de la matriz inversa, esto es llamado proceso de difusión espacial.

La matriz de varianzas-covarianzas para y es fácilmente obtenida siendo igual a:

$$\begin{aligned} \text{Var}[y] &= E\{[y - Ey][y - Ey]'\}, \\ &= E[A^{-1}uu'A^{-1'}], \\ &= \sigma^2 A^{-1}A^{-1'}, \\ \text{Var}[y] &= \sigma^2 (I - \lambda W)^{-1} (I - \lambda W')^{-1}, \end{aligned}$$

dando lugar a que cada país esté correlacionado con los demás en el sistema, con una mayor correlación según la cercanía. Además, los elementos de la diagonal principal no sean constantes, es decir, el modelo muestra heteroscedasticidad aún cuando el proceso inicial no es heteroscedástico.

La presencia de la matriz λW permite obtener efectos directos e indirectos de interacción espacial. Los detalles sobre estos efectos y su obtención serán desarrollados en la sección ??.

Modelo de error espacial

Otro modelo con estructura espacial es el modelo de error espacial (*SEM*, spatial error model):

$$\begin{aligned} y &= X\beta + u, \\ u &= \rho Wu + \varepsilon, \Rightarrow u = B^{-1}\varepsilon, \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2 I_n), \end{aligned} \tag{3.3.5}$$

donde $B = [I - \rho W]$ y ε un vector de innovaciones.

Este modelo tiene una matriz de varianzas y covarianzas igual a:

$$\begin{aligned} E[uu'] &= E\{[B^{-1}\varepsilon][B^{-1}\varepsilon]'\}, \\ &= E[B^{-1}\varepsilon\varepsilon'B^{-1'}], \\ &= \sigma^2 B^{-1}B^{-1'}, \\ E[uu'] &= \sigma^2 \Omega(\rho)^{-1}, \end{aligned}$$

siendo $\Omega(\rho) = (I - \rho W)'(I - \rho W) = B'B$. Esta estructura genera que los errores se desborden afectando a todo el sistema.

Suponiendo distribución normal de las innovaciones, la función de log-verosimilitud del *SEM* será:

$$L = -\frac{n}{2} \ln(2\pi\sigma^2) + \ln|B| - \frac{1}{2\sigma^2} [(y - X\beta)' \Omega(\rho) (y - X\beta)]. \tag{3.3.6}$$

Tal como sucedía en el modelo *SLM*, las condiciones de primer orden son expresiones altamente no lineales y por lo tanto la log-verosimilitud no puede maximizarse directamente.

Para un ρ dado, la optimización de la función de log-verosimilitud (consúltese [LeSage & Pace, 2009](#)) muestra que:

$$\beta_\rho = [X' \Omega(\rho) X]^{-1} X' \Omega(\rho) y,$$

y

$$\sigma_\rho^2 = \frac{e' \Omega(\rho) e}{n},$$

donde $e = y - X\beta_\rho$ y $\Omega(\rho) = (I - \rho W)'(I - \rho W)$.

Ambos parámetros son funciones del ρ . Una log-verosimilitud concentrada puede ser obtenida como:

$$L_c = C + \ln |I - \rho W| - \frac{n}{2} \ln [e' \Omega(\rho) e].$$

Igual que el procedimiento implementado para el *SLM*, se maximiza la log-verosimilitud concentrada hasta lograr convergencia y posteriormente con el valor $\hat{\rho}$ se obtienen los demás parámetros del modelo. Una vez logrado la convergencia de L_c , puede sustituirse en las expresiones previas tal que:

$$\begin{aligned} \hat{\beta}_\rho &= [X' \Omega(\hat{\rho}) X]^{-1} X' \Omega(\hat{\rho}) y, \\ \hat{\sigma}_\rho^2 &= \frac{1}{n} (e' \Omega(\hat{\rho}) e). \end{aligned}$$

El modelo *SEM* por máxima verosimilitud puede estimarse por medio del comando `spreg ml`, sub-opción `elmat(Winv2_st)`:

```
. spreg ml lny95 lns lnlgd, id(id) elmat(Winv2_st)
```

```
Performing a grid search.... finished
```

```
Iteration 0: log likelihood = -85.194677
Iteration 1: log likelihood = -76.467742
Iteration 2: log likelihood = -76.225266
Iteration 3: log likelihood = -76.22308
Iteration 4: log likelihood = -76.223063
Iteration 5: log likelihood = -76.223063
Iteration 6: log likelihood = -76.223063
Iteration 7: log likelihood = -76.223063
Optimizing unconcentrated log likelihood
Iteration 0: log likelihood = -76.223063
```

```
Spatial autoregressive model          Number of obs   =          91
(Maximum likelihood estimates)        Wald chi2(2)     =       70.4829
                                      Prob > chi2       =       0.0000
```

```
-----+-----
      lny95 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      lny95 |
      lns |   .8239677   .1016769     8.10   0.000    .6246848    1.023251
      lnlgd | -1.694553   .5664868    -2.99   0.003   -2.804847   -1.5842598
      _cons |   6.461502   1.53853     4.20   0.000    3.446039    9.476965
```

```

-----+-----
rho      |
      _cons |   .8266491   .0631425   13.09   0.000   .702892   .9504061
-----+-----
sigma2    |
      _cons |   .2531681   .03965    6.39   0.000   .1754554   .3308807
-----+-----
. estimates store modelSEM

```

La interpretación de estos resultados no difiere del modelo de regresión lineal. El efecto espacial solo afecta a los errores estándares.

Modelo *SARAR*

El modelo *SARAR* es un modelo utilizado en diferentes trabajos teóricos en econometría espacial aunque de escasa difusión en aplicaciones empíricas. Este modelo incorpora, simultáneamente, estructura sustantiva y residual:

$$\begin{aligned} y &= \lambda W y + X\beta + u, \\ u &= \rho W u + \varepsilon, \end{aligned} \quad (3.3.7)$$

suponiendo $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$.

Una forma útil de presentar este modelo es la siguiente:

$$\begin{aligned} Ay &= X\beta + u, \\ Bu &= \varepsilon, \end{aligned}$$

donde $A = (I - \lambda W)$ y $B = (I - \rho W)$.

Bajo esta notación, suponiendo $|\lambda| < 1$ y $|\rho| < 1$, la función de log-verosimilitud es:

$$L = -\frac{n}{2} \ln(\sigma^2) + \ln|A| + \ln|B| - \frac{1}{2\sigma^2} v'v, \quad (3.3.8)$$

con $v'v = (Ay - X\beta)' \Omega(\rho) (Ay - X\beta)$ como la suma de cuadrados de los errores transformados, siendo $\Omega(\rho) = B'B$.

Puede concentrarse la log-verosimilitud primero maximizando la función 3.3.8 respecto a β y σ^2 , obteniendo:

$$\beta_{\lambda\rho} = [X'\Omega(\rho)X]^{-1} X'\Omega(\rho) Ay,$$

y

$$\sigma_\rho^2 = \frac{1}{n} [Ay - X\beta_{\lambda\rho}]' \Omega(\rho) [Ay - X\beta_{\lambda\rho}].$$

Reemplazando estas expresiones en la ecuación 3.3.8 se genera la log-verosimilitud concentrada que depende de λ y ρ . Una vez obtenidos los valores $\hat{\lambda}$ y $\hat{\rho}$ que maximizan la función pueden estimarse los valores de β y σ^2 como $\hat{\beta}_{\hat{\lambda}\hat{\rho}}$ y $\hat{\sigma}_{\hat{\lambda}\hat{\rho}}^2$.

En Stata este modelo se obtiene de la siguiente manera:

```
. spreg ml lny95 lns lnlgd, id(id) dmat(Winv2_st) elmat(Winv2_st)
```

Performing a grid search.... finished

Iteration 0: log likelihood = -73.318597

Iteration 1: log likelihood = -73.301799

Iteration 2: log likelihood = -73.301797

Optimizing unconcentrated log likelihood

Iteration 0: log likelihood = -73.301797

Spatial autoregressive model

(Maximum likelihood estimates)

Number of obs	=	91
Wald chi2(2)	=	68.2209
Prob > chi2	=	0.0000

	lny95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lny95						
lns		.8761535	.1076121	8.14	0.000	.6652376 1.087069
lnlgd		-1.612564	.5925813	-2.72	0.007	-2.774002 -.4511258
_cons		2.202529	1.5265	1.44	0.149	-.7893569 5.194414
lambda						
_cons		.4933581	.1430422	3.45	0.001	.2130006 .7737156
rho						
_cons		.3871749	.2484714	1.56	0.119	-.09982 .8741699
sigma2						
_cons		.2678415	.0403271	6.64	0.000	.188802 .3468811

El coeficiente espacial del error, $\hat{\rho}$, no es significativo al 5%, favoreciendo la estimación de un modelo sin estructura espacial en el término de error.

Modelo espacial de Durbin

El modelo espacial de Durbin, *SDM*, es el modelo teórico que deriban [Ertur & Koch \(2007\)](#) y posee la siguiente especificación:

$$y = \lambda W y + X\beta + WX\theta + u, \quad (3.3.9)$$

donde WX captura el efecto espacial local de las variables exógenas y anida al modelo *SLX*.

El procedimiento para estimar este modelo no difiere del presentado por el *SLM*, ya que solo se incorporan nuevas variables exógenas, considerando que W es por construcción no-estocástica.

En Stata, para estimar el modelo *SDM*, se necesita generar los rezagos espaciales de las variables explicativas. En la sección previa se obtuvieron mediante un comando, en esta sección se muestra otra alternativa mediante las opciones que brinda el comando **spmat**: el subcomando es **lag**, seguido por el nombre de la nueva variable, la matriz de pesos y la variable original de la que se desea obtener el rezago espacial:

```
. spmat lag wx_lns Winv2_st lns
. spmat lag wx_lnlgd Winv2_st lnlgd
```

Una vez generada la variable, se procede a estimar el modelo *SLM*, pero incorporando **wx_lns** y **wx_lnlgd** como variable explicativa:

```
. spreg ml lny95 lns lnlgd wx_lns wx_lnlgd, id(id) dlmat(Winv2_st)
Performing a grid search.... finished
Iteration 0:   log likelihood = -73.284801
Iteration 1:   log likelihood = -73.18522
Iteration 2:   log likelihood = -73.184547
Iteration 3:   log likelihood = -73.184547
Optimizing unconcentrated log likelihood
Iteration 0:   log likelihood = -73.184547
Spatial autoregressive model                               Number of obs   =       91
(Maximum likelihood estimates)                             Wald chi2(4)    =   75.6347
                                                            Prob > chi2     =    0.0000
```

	lny95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lny95						
lns		.8201799	.1003312	8.17	0.000	.6235343 1.016826
lnlgd		-1.494475	.5645324	-2.65	0.008	-2.600938 -.3880123
wx_lns		-.3138399	.2127504	-1.48	0.140	-.7308229 .1031431
wx_lnlgd		.5506592	.8399951	0.66	0.512	-1.095701 2.197019
_cons		.9501218	1.997277	0.48	0.634	-2.964469 4.864712
lambda						
_cons		.7397691	.0843045	8.77	0.000	.5745352 .9050029
sigma2						
_cons		.2510158	.0390827	6.42	0.000	.1744152 .3276164

```
. estimates store modelSDM
```

Es interesante mencionar que el modelo *SDM*, a pesar de contener una estructura más compleja que el *SLM*, puede reducirse a un modelo *SEM*. Esta relación se hace evidente mediante el test denominado *COMFAC* (de factores comunes).

Suponiendo que se ha estimado el modelo *SDM*:

$$y = \lambda W y + X \beta + W X \theta + u. \quad (3.3.10)$$

El estadístico de factores comunes plantea las siguientes hipótesis, nula y alternativa:

$$\begin{aligned} H_0 : & \theta + \lambda \beta = 0, \\ H_1 : & \theta + \lambda \beta \neq 0. \end{aligned}$$

Bajo H_0 , $\gamma = -\lambda \beta$, entonces reemplazando en el modelo (3.3.10):

$$\begin{aligned}
y &= \lambda W y + X\beta + WX(-\lambda\beta) + u, \\
&= \lambda W y + X\beta - \lambda WX\beta + u, \\
(I - \lambda W)y &= (I - \lambda W)X\beta + u.
\end{aligned}$$

La última expresión puede resumirse en el *SEM*:

$$y = X\beta + (I - \rho W)^{-1} \varepsilon,$$

donde se ha reemplazado a λ por ρ .

El modelo bajo hipótesis alternativa es el *SDM* y el modelo restringido, bajo hipótesis nula, es el *SEM*. Bajo estimación *MV*, puede plantearse el estadístico *LR*:

$$LR_{COMFAC} = 2 [l_{|H_1} - l_{|H_0}] \underset{as}{\sim} \chi_q^2 \quad (3.3.11)$$

siendo $l_{|H_1}$ la log-verosimilitud obtenida en la estimación del modelo amplio y $l_{|H_0}$ la correspondiente al modelo de la hipótesis nula, q es el número de restricciones e igual al número de parámetros incluidos en β , sin considerar la constante.

Para aplicar este test en Stata, previamente se han guardado las estimaciones del modelo *SDM* y *SEM*. Luego, se aplica el siguiente comando:

```
. lrtest modelSDM modelSEM
Likelihood-ratio test                LR chi2(2) =      6.08
(Assumption: modelSEM nested in modelSDM) Prob > chi2 =    0.0479
```

donde *modelSDM* y *modelSEM* son las estimaciones almacenadas en Stata del modelo *SDM* y *SEM*, respectivamente.

En conclusión, se rechaza la hipótesis nula, arrojando evidencia a favor de un modelo *SDM*. En términos del modelo teórico, la restricción de factores comunes implica ausencia de externalidades del capital físico representadas por ϕ . De hecho, si $\phi = 0$, las restricciones en la especificación *SDM* son: $\beta_1 = -\beta_2 = \frac{\alpha}{1-\alpha}$ y $\theta_2 = -\theta_1 = \frac{\alpha\gamma}{1-\alpha}$, implicando las siguientes restricciones no-lineales: $\theta_1 + \beta_1\gamma = 0$ y $\theta_2 + \beta_2\gamma = 0$. Sin embargo, estas restricciones son rechazadas en favor del modelo *SDM* y por consecuencia se encuentra evidencia de externalidades positivas del capital físico.

Otra restricción ya contrastada en el modelo de Solow básico han sido: $\beta_1 = -\beta_2$ y $\theta_2 = -\theta_1$. Ahora contrastarán estas restricciones bajo el modelo *SDM*:

```
. spmat lag wx_lnsldg Winv2_st lnsldg

. spreg ml lny95 lnsldg wx_lnsldg, id(id) dlmata(Winv2_st)
Performing a grid search.... finished
Iteration 0:  log likelihood = -74.065563
Iteration 1:  log likelihood = -73.987198
Iteration 2:  log likelihood = -73.986799
Iteration 3:  log likelihood = -73.986799
```

```

Optimizing unconcentrated log likelihood
Iteration 0:   log likelihood = -73.986799
Spatial autoregressive model
(Maximum likelihood estimates)
Number of obs   =      91
Wald chi2(2)    =    72.7351
Prob > chi2     =    0.0000

-----+-----
            lny95 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
lny95       |
      lnslgd |   .8378764   .1001751     8.36   0.000     .6415369    1.034216
      wx_lnsld | -.2650967   .2029543    -1.31   0.191    -.6628799    .1326865
      _cons   |   2.106019   .6944463     3.03   0.002     .7449292    3.467108
-----+-----
lambda       |
      _cons   |   .7356144   .0851873     8.64   0.000     .5686504    .9025784
-----+-----
sigma2       |
      _cons   |   .2560763   .0398541     6.43   0.000     .1779637    .3341889
-----+-----

. estimates store modelSDMr

```

El test de LR entre el modelo SDM y el SDM restringido no puede rechazar la hipótesis nula, por lo que el modelo SDM restringido es el modelo final.

```

. lrtest modelSDM modelSDMr
Likelihood-ratio test
(Assumption: modelSDMr nested in modelSDM)
LR chi2(2) =      1.61
Prob > chi2 =    0.4471

```

Utilizando las últimas estimaciones, pueden recuperarse los parámetros estructurales α , ϕ y γ :

```

. mata
----- mata (type end to exit) -----
: b = st_matrix("e(b)")
: b
            1            2            3            4            5
+-----+-----+-----+-----+-----+
1 |   .8378764298   -.2650967104   2.106018843   .7356144169   .2560762852 |
+-----+-----+-----+-----+-----+
: beta = b[1,1]
: theta = b[1,2]
: lambda = b[1,4]
: alpha = theta / (theta - lambda)
: phi = beta / (1 + beta) - alpha
: gamma = - (theta - lambda) / (1 + beta)
: alpha
      .2649083268
: phi

```

```
.1909854515
: gamma
.5444931504
: end
```

Los resultados de los parámetros estructurales son similares a los presentados por [Ertur & Koch \(2007, Tabla 1, pp. 1047\)](#).

Modelo de Cliff-Ord

El modelo más completo en términos espaciales es denominado modelo de Cliff-Ord. Este modelo incorpora dependencia sustantiva, dependencia residual y dependencia local debido a las variables exógenas, siendo su expresión:

$$\begin{aligned} y &= \lambda W y + X\beta + WX\theta + u, \\ u &= \rho W u + \varepsilon. \end{aligned} \quad (3.3.12)$$

Este modelo posee escasa relevancia en aplicaciones empíricas debido a que se encuentra débilmente identificado.

La virtud del modelo radica en que la imposición de diferentes restricciones en la ecuación (3.3.12) sobre γ , ρ y λ permite obtener los modelos presentados previamente:

- $\theta = 0, \rho = 0, \lambda \neq 0 \rightarrow SLM$.
- $\theta = 0, \rho \neq 0, \lambda = 0 \rightarrow SEM$.
- $\theta = 0, \rho \neq 0, \lambda \neq 0 \rightarrow SARAR$.
- $\theta \neq 0, \rho = 0, \lambda = 0 \rightarrow SLX$.
- $\theta \neq 0, \rho = 0, \lambda \neq 0 \rightarrow SDM$.

El método de estimación *MV* para el modelo de Cliff-Ord no difiere del presentado por el modelo *SARAR*, ya que solo incorpora un elemento adicional en las variables explicativas, WX .

La estimación en Stata se realiza con la siguiente secuencia:

```
. spreg ml lny95 lnslgd wx_lnslgd, id(id) dlmats(Winv2_st) elmat(Winv2_st)
```

```
Performing a grid search.... finished
Iteration 0:   log likelihood = -73.683684
Iteration 1:   log likelihood = -73.642565
Iteration 2:   log likelihood = -73.642192
Iteration 3:   log likelihood = -73.642192
Optimizing unconcentrated log likelihood
Iteration 0:   log likelihood = -73.642192
Spatial autoregressive model
```

```
Number of obs   =      91
```

(Maximum likelihood estimates)				Wald chi2(2)	=	68.8575
				Prob > chi2	=	0.0000

lny95		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

lny95						
lnslgd		.8770193	.1059499	8.28	0.000	.6693614 1.084677
wx_lnsld		-.238518	.2212245	-1.08	0.281	-.6721101 .1950741
_cons		2.684079	1.183426	2.27	0.023	.3646062 5.003552

lambda						
_cons		.6683512	.1404543	4.76	0.000	.3930658 .9436367

rho						
_cons		.2144963	.2634297	0.81	0.416	-.3018164 .730809

sigma2						
_cons		.2602409	.0405884	6.41	0.000	.180689 .3397927

Obsérvese que el parámetro estimado ρ es no significativo, reduciendo el modelo de Cliff-Ord a un modelo *SDM*.

Metodológicamente, puede comenzarse con una estimación *MCO* e incorporar elementos espaciales o, alternativamente, puede estimarse el modelo más complejo e ir eliminando los elementos no significativos. La primera estrategia es denominada “de lo particular a lo general” y la segunda “de lo general a lo particular”. Para una discusión sobre las ventajas de una u otra estrategia en econometría espacial puede consultarse el trabajo de [Mur & Angulo \(2009\)](#) y las referencias allí citadas.

3.4 Estimación por método de momentos generalizados

[Ertur & Koch \(2007\)](#) únicamente aplican la estimación por máxima verosimilitud. En esta sección se replicaron los principales resultados del modelo final, versión restringida, usando la estimación por el método de momentos generalizados. Los principales precursores de este método han sido los trabajos de [Kelejian & Prucha \(1998, 1999\)](#), con generalizaciones recientes realizadas por [Kelejian & Prucha \(2010\)](#); [Arraiz et al. \(2010\)](#); [Drukker et al. \(2013a\)](#). [Ertur & Koch \(2007\)](#) no exploran los resultados obtenidos de su modelo teórico bajo este método de estimación.

Modelo de rezago espacial

Para implementar la estrategia de estimación se reexpresará al modelo *SLM* como:

$$y = Z\delta + u, \quad (3.4.1)$$

donde $Z_{n \times (k+1)} = [X, Wy]$ y $\delta' = [\beta', \lambda]$, con $u \sim (0, \sigma_u^2 I_n)$.

La estimación de este modelo se basa en el enfoque de variables instrumentales (*IV*). La idea básica es simple, el modelo (3.4.1) no puede estimarse por los métodos habituales debido a que existe una variable endógena, Wy , que se correlaciona con el término de error. La alternativa es encontrar un instrumento, H , que se encuentre fuertemente correlacionado con la variable que genera endogeneidad Wy pero no con el error, asintóticamente hablando.

Es decir, el instrumento H debe cumplir con:

$$plim \left(\frac{1}{n} \right) H'Wy = M_{HWy},$$

y

$$plim \left(\frac{1}{n} \right) H'u = 0,$$

siendo M_{HWy} una matriz finita no-singular.

Recordando que la forma reducida del modelo *SLM* es: $E(y) = (I - \lambda W)^{-1}(X\beta + u)$, y considerando que $(I - \lambda W)^{-1}$ puede expresarse como una progresión: $(I - \lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 + \dots$. Entonces, el valor esperado de y puede expresarse como:

$$E(y) = X\beta + \lambda WX\beta + \lambda^2 W^2 X\beta + \lambda^3 W^3 X\beta + \dots,$$

y, además,

$$E(Wy) = WX\beta + W^2 X(\lambda\beta) + W^3 X(\lambda^2\beta) + \dots.$$

Luego, $E(Wy)$ se relaciona linealmente con WX , $W^2 X$, $W^3 X$, ..., $W^q X$, ..., siendo instrumentos directos de la variable endógena Wy . Diferentes experimentos Monte Carlo muestran un buen comportamiento para $q = 2$, por lo que habitualmente se utilizan WX y $W^2 X$ como instrumentos. El uso de ordenes mayores tiende a generar problemas de multicolinealidad entre los regresores.

Aplicando esta idea al modelo (3.4.1) se obtiene el estimador de variables instrumentales (*IV*):

$$\hat{\delta} = \left(\hat{Z}'Z \right)^{-1} \hat{Z}'y,$$

donde $\hat{Z} = P_H Z = [X, \widehat{Wy}]$, $\widehat{Wy} = P_H Wy$ y $P_H = H(H'H)^{-1}H'$, siendo los instrumentos elegidos $H = [X, WX, W^2 X]$.

[Kelejian & Prucha \(1998\)](#) demuestran que

$$n^{1/2} (\hat{\delta} - \delta) \xrightarrow{d} \mathcal{N} \left(0, \sigma^2 \lim_{n \rightarrow \infty} \left[\hat{Z}'Z \right]^{-1} \right).$$

La estimación *IV* de estos parámetros puede ser obtenida como un estimador de mínimos cuadrados en dos etapas (*2SLS*):

1. Estimar la relación entre los instrumentos $H = [X, WX]$ y $Z = [X, Wy]$:

$$Z = H\theta + \epsilon,$$

$$\hat{\theta} = (H'H)^{-1} H'Z,$$

tal que las predicciones del modelo son: $\widehat{Z} = H\widehat{\theta} = H(H'H)^{-1}H'Z = P_H Z$.

2. Estimar $y = Z\delta + \varepsilon$, usando los resultados del paso anterior:

$$\widehat{\delta} = \left(\widehat{Z}'\widehat{Z}\right)^{-1}\widehat{Z}'y,$$

donde $\widehat{Z} = P_H Z = [X, \widehat{W}y]$, donde $\widehat{W}y = P_H Wy$.

En Stata, el comando que permite realizar esta estimación es `spivreg` desarrollado por [Drukker et al. \(2013c\)](#):

```
. spivreg lny95 lnslgd, id(id) dl(Winv2_st)
Spatial autoregressive model          Number of obs   =          91
(GS2SLS estimates)

-----+-----
      lny95 |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
lny95      |
  lnslgd   |      .821611    .111586     7.36   0.000     .6029065    1.040316
    _cons  |      2.97183    .7597846    3.91   0.000     1.482679    4.46098
-----+-----
lambda     |
    _cons  |      .6253418   .0875356    7.14   0.000     .4537752    .7969083
-----+-----
```

La estimación realizada puede obtenerse sin necesidad de un comando especial mediante la generación previa de los instrumentos y utilizando el comando de regresión por variables instrumentales simple:

```
. spmat lag wx2_lnslgd Winv2_st wx_lnslgd
. spmat lag wy_lny95 Winv2_st lny95
. ivregress 2sls lny95 lnslgd (wy_lny95 = wx_lnslgd wx2_lnslgd)
Instrumental variables (2SLS) regression          Number of obs   =          91
                                                Wald chi2(2)    =    349.98
                                                Prob > chi2     =    0.0000
                                                R-squared       =    0.8148
                                                Root MSE       =    .52419

-----+-----
      lny95 |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
wy_lny95   |      .6253418   .0875356    7.14   0.000     .4537752    .7969084
  lnslgd    |      .8216111   .111586     7.36   0.000     .6029065    1.040316
    _cons   |      2.97183    .7597846    3.91   0.000     1.482679    4.46098
-----+-----

Instrumented:  wy_lny95 Instruments:   lnslgd wx_lnslgd wx2_lnslgd
```

Si se supone que el término de error es independiente pero heteroscedástico, entonces se modifica la matriz de varianzas-covarianzas asintótica de $\hat{\delta}$, adquiriendo la forma *sandwich*:

$$\left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}'\hat{\Sigma}\hat{Z}\left(\hat{Z}'\hat{Z}\right)^{-1},$$

donde $\hat{\Sigma}$ es una matriz diagonal cuyos elementos i -ésimos son \tilde{u}_i^2 con $\tilde{u}_i = y_i - Z_i\hat{\delta}$.

Para realizar esta corrección en Stata basta agregar al comando anterior la sub-opción **het**:

```
. spivreg lny95 lnslgd, id(id) dl(Winv2_st) het
Spatial autoregressive model          Number of obs   =          91
(GS2SLS estimates)

-----+-----
            lny95 |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
lny95          |
      lnslgd |      .821611   .1622652     5.06   0.000     .5035772     1.139645
        _cons |      2.97183   1.008858     2.95   0.003     .9945052     4.949154
-----+-----
lambda          |
        _cons |      .6253418   .116842     5.35   0.000     .3963356     .8543479
-----+-----
```

Comparando las estimaciones bajo homoscedasticidad y heteroscedasticidad, puede observarse que las correcciones afectan a los errores estándares de los coeficientes.

Modelo de error espacial

El modelo *SEM* puede expresarse como:

$$\begin{aligned} y &= Z\delta + u, \\ u &= \rho Wu + \varepsilon, \end{aligned} \tag{3.4.2}$$

donde $Z = X$, $\delta = \beta$, ρ es un parámetro autoregresivo espacial, siendo las innovaciones $\varepsilon \sim (0, \sigma^2 I)$, con I_n siendo una matriz identidad de orden $n \times n$.

El objetivo es obtener un estimador consistente de ρ . Este paso genera un estimador inicial *GMM* para ρ surge de las condiciones de momentos cuadráticos de las innovaciones:

$$E(\varepsilon' A_s \varepsilon) = 0, \quad s = 1, \dots, S,$$

donde las matrices A_s satisfacen $\text{tr}(A_s) = 0$.

El estimador inicial *GMM* viene dado por

$$\tilde{\rho} = \arg \min \left\{ \left[\tilde{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right]' \left[\tilde{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right] \right\},$$

donde

$$\tilde{\Gamma} = \frac{1}{n} \begin{bmatrix} \hat{u}'(A_1 + A'_1)\tilde{u} & -\hat{u}'A_1\hat{u} \\ \vdots & \vdots \\ \hat{u}'(A_S + A'_S)\tilde{u} & -\hat{u}'A_S\hat{u} \end{bmatrix}, \tilde{\omega} = \frac{1}{n} \begin{bmatrix} \hat{u}'A_1\hat{u} \\ \vdots \\ \hat{u}'A_S\hat{u} \end{bmatrix}, \text{ donde } \hat{u} = y - Z\hat{\beta} \text{ y } \tilde{u} = W\hat{u}.$$

Así planteado, el estimador *GMM* proviene de un estimador de mínimos cuadrados no-lineales. Generalmente, se utilizan dos ecuaciones ($S = 2$).

Bajo homoscedasticidad de las innovaciones:

$$A_1 = \left[1 + \left\{ \frac{1}{n} \text{tr}(W'W) \right\}^2 \right]^{-1} \left\{ W'W - \frac{1}{n} \text{tr}(W'W) I_n \right\},$$

y

$$A_2 = W.$$

Obtenida la estimación inicial de ρ , se procede a obtener los coeficientes del modelo 3.4.2. Para ello se utiliza la transformación Cochrane-Orcutt:

$$\begin{aligned} (I - \rho W)y &= (I - \rho W)Z\beta + \varepsilon, \\ y_\rho &= Z_\rho\beta + \varepsilon. \end{aligned} \tag{3.4.3}$$

Bajo esta especificación, se utiliza la estimación *GMM* $\tilde{\rho}$:

$$\hat{\delta} = \left(\hat{Z}_\rho' Z_\rho \right)^{-1} \hat{Z}_\rho' y_\rho,$$

donde $y_\rho = (I - \tilde{\rho}W)y$, $Z_\rho = (I - \tilde{\rho}W)Z$, y $\hat{Z}_\rho = P_H Z_\rho = X(X'X)^{-1}X'Z_\rho$.

El procedimiento completo es conocido como mínimos cuadrados espaciales en dos etapas generalizados (*GS2SLS*).

Posterior a la estimación de los parámetros del modelo, puede darse paso adicional para obtener un estimador eficiente de ρ , para ello se utilizan los residuos del procedimiento *GS2SLS*:

$$\hat{\rho} = \arg \min \left\{ \left[\hat{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right]' \left\{ \hat{\Psi}_\rho^{\rho\rho} \right\}^{-1} \left[\hat{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right] \right\},$$

donde $\hat{\Psi}_\rho^{\rho\rho}$ es un estimador de la matriz de varianzas-covarianzas del vector de momentos muestrales (normalizados) basados en los residuos *GS2SLS*. Bajo homoscedasticidad los elementos r, s (con $r, s = 1, 2$) de $\hat{\Psi}_\rho^{\rho\rho}$ están dados por:

$$\begin{aligned} \hat{\Psi}_\rho^{\rho\rho} &= \left\{ \tilde{\sigma}_\rho^2 \right\}^2 \frac{1}{2n} \text{tr} \{ (A_r + A'_r)(A_s + A'_s) \} + \frac{1}{n} \tilde{\sigma}_\rho^2 \tilde{a}_{\rho r}' \tilde{a}_{\rho s} \\ &\quad + \frac{1}{n} \left[\tilde{\mu}_\rho^{(4)} - 3 \left\{ \tilde{\sigma}_\rho^2 \right\}^2 \right] \text{vec}_D(A_r)' \text{vec}_D(A_s) \\ &\quad + \frac{1}{n} \tilde{\mu}_\rho^{(3)} \left\{ \tilde{a}_{\rho r}' \text{vec}_D(A_s) + \tilde{a}_{\rho s}' \text{vec}_D(A_r) \right\} \end{aligned}$$

donde

$$\tilde{a}_{\rho r} = \hat{T}_\rho \tilde{\alpha}_{\rho r},$$

$$\begin{aligned}
\hat{T}_{\rho} &= H\hat{P}_{\rho}, \\
\hat{P}_{\rho} &= \hat{Q}_{HH}^{-1} \hat{Q}_{\rho HZ} \left\{ \hat{Q}'_{\rho HZ} \hat{Q}_{HH}^{-1} \hat{Q}'_{\rho HZ} \right\}^{-1}, \\
\hat{Q}_{HH} &= \left(\frac{1}{n} H' H \right), \\
\hat{Q}_{\rho HZ} &= \left(\frac{1}{n} H' Z_{\rho} \right), \\
Z_{\rho} &= (I - \hat{\rho} W) Z, \\
\tilde{a}_{\rho r} &= -\frac{1}{n} \left\{ Z'_{\rho} (A_r + A'_r) \hat{\varepsilon}_{\rho} \right\}, \\
\hat{\varepsilon}_{\rho} &= (I - \hat{\rho} W) \hat{u}, \\
\hat{\sigma}_{\rho}^2 &= \frac{1}{n} \hat{\varepsilon}'_{\rho} \hat{\varepsilon}_{\rho}, \\
\hat{\mu}_{\rho}^{(3)} &= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{i\rho}^3, \\
\hat{\mu}_{\rho}^{(4)} &= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{i\rho}^4.
\end{aligned}$$

En Stata, la estimación de este modelo se obtiene mediante:

```
. spivreg lny95 lnslgd, id(id) el(Winv2_st)
```

Estimating rho using 2SLS residuals

Iteration 0: GMM criterion = .00079191

Iteration 1: GMM criterion = .00077114

Iteration 2: GMM criterion = .00077112

Estimating rho using GS2SLS residuals

Iteration 0: GMM criterion = .06695338

Iteration 1: GMM criterion = .02455973

Iteration 2: GMM criterion = .02453116

Iteration 3: GMM criterion = .02453101

Iteration 4: GMM criterion = .02453101

Spatial autoregressive model

Number of obs = 91

(GS2SLS estimates)

	lny95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lny95							
lnslgd		.893372	.0968803	9.22	0.000	.70349	1.083254
_cons		8.705266	.6442985	13.51	0.000	7.442465	9.968068
rho							
_cons		.9191018	.1141702	8.05	0.000	.6953323	1.142871

La estimación obtenida supone innovaciones homoscedásticas. En el caso de heteroscedasticidad, se debe modificar en el estimador inicial el elemento A_1 : $A_1 = W'W - \text{diag}(W'W)$.

En la segunda etapa, que realiza una estimación eficiente de ρ usando los residuos *GS2SLS*, se modifica el elemento r, s de $\hat{\Psi}_{\rho}^{\rho\rho}$:

$$\hat{\Psi}_{\rho}^{\rho\rho} = \frac{1}{2n} \text{tr} \left\{ (A_r + A'_r) \hat{\Sigma}_{\rho} (A_s + A'_s) \right\} + \frac{1}{n} \tilde{\sigma}_{\rho}^2 \tilde{a}_{\rho r} \hat{\Sigma}_{\rho} \tilde{a}_{\rho s},$$

Paso 2: Obtener $\hat{u} = y - Z\hat{\delta}$, $\hat{\tilde{u}} = W\hat{u}$. Usando estos vectores residuales obtener el estimador inicial *GMM* minimizando:

$$\tilde{\rho} = \arg \min \left\{ \left[\tilde{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right]' \left[\tilde{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right] \right\},$$

Paso 3: Estimador *GS2SLS*. Se utiliza la estimación inicial *GMM*, $\tilde{\rho}$, en la transformación Cochrane-Orcutt para obtener

$$\hat{\delta} = \left(\hat{Z}'_{\tilde{\rho}} Z_{\tilde{\rho}} \right)^{-1} \hat{Z}'_{\tilde{\rho}} y_{\tilde{\rho}},$$

donde $y_{\tilde{\rho}} = (I - \tilde{\rho}W)y$, $Z_{\tilde{\rho}} = (I - \tilde{\rho}W)Z$, y $\hat{Z}_{\tilde{\rho}} = P_H Z_{\tilde{\rho}} = H(H'H)^{-1}H'Z_{\tilde{\rho}}$.

Paso 4: Estimador eficiente y consistente *GMM*. Utilizando los residuos del procedimiento *GS2SLS*, se estima:

$$\hat{\rho} = \arg \min \left\{ \left[\hat{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right]' \left\{ \hat{\Psi}_{\hat{\rho}}^{\rho\rho} \right\}^{-1} \left[\hat{\Gamma} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \tilde{\omega} \right] \right\},$$

Habiendo computado los estimadores $\hat{\delta} = (\hat{\beta}, \hat{\lambda})$ y $\hat{\rho}$, el próximo paso es computar un estimador consistente de la matriz de varianzas-covarianzas, Υ . El estimador viene dado por $n\hat{\Upsilon}$ donde:

$$\hat{\Upsilon} = \begin{pmatrix} \hat{\Upsilon}^{\delta\delta} & \hat{\Upsilon}^{\delta\rho} \\ \hat{\Upsilon}^{\delta\rho'} & \hat{\Upsilon}^{\rho\rho} \end{pmatrix}$$

donde

$$\hat{\Upsilon}^{\delta\delta} = \hat{P}'_{\hat{\rho}} \hat{\Psi}_{\hat{\rho}}^{\delta\delta} \hat{P}_{\hat{\rho}},$$

$$\hat{\Upsilon}^{\delta\rho} = \hat{P}'_{\hat{\rho}} \hat{\Psi}_{\hat{\rho}}^{\delta\rho} \hat{P}_{\hat{\rho}} \left\{ \hat{\Psi}_{\hat{\rho}}^{\rho\rho} \right\}^{-1} \hat{J} \left[\hat{J}' \left\{ \hat{\Psi}_{\hat{\rho}}^{\rho\rho} \right\}^{-1} \hat{J} \right]^{-1},$$

$$\hat{\Upsilon}^{\rho\rho} = \hat{J} \left[\hat{J}' \left\{ \hat{\Psi}_{\hat{\rho}}^{\rho\rho} \right\}^{-1} \hat{J} \right]^{-1},$$

$$\hat{J} = \hat{\Gamma} \begin{pmatrix} 1 \\ 2\hat{\rho} \end{pmatrix}.$$

Bajo el supuesto de homoscedasticidad de las innovaciones, se utilizan:

$$\hat{\Psi}_{\hat{\rho}}^{\delta\delta} = \hat{\sigma}_{\hat{\rho}}^2 \hat{Q}_{HH},$$

$$\hat{\Psi}_{\hat{\rho}}^{\delta\rho} = \frac{1}{n} \hat{\sigma}_{\hat{\rho}}^2 H' \left\{ a_{1\hat{\rho}}, a_{2\hat{\rho}} \right\} + \frac{1}{n} \mu_{\hat{\rho}}^3 H' \{ vec_D(A_1), vec_D(A_2) \}.$$

En Stata, el comando para estimar un *SARAR* homoscedástico es:

```
. spivreg lny95 lnslgd, id(id) el(Winv2_st) dl(Winv2_st)
```

```
Estimating rho using 2SLS residuals
Iteration 0:   GMM criterion =   .00010232
Iteration 1:   GMM criterion =   .00009871
Iteration 2:   GMM criterion =   .0000987
Estimating rho using GS2SLS residuals
```

```

Iteration 0:   GMM criterion = .00595159
Iteration 1:   GMM criterion = .00465292
Iteration 2:   GMM criterion = .0046526
Iteration 3:   GMM criterion = .0046526

```

```

Spatial autoregressive model                Number of obs   =       91
(GS2SLS estimates)

```

	lny95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lny95						
	lnslgd	.8594081	.1039971	8.26	0.000	.6555775 1.063239
	_cons	3.299241	.8426125	3.92	0.000	1.647751 4.950731
lambda						
	_cons	.5871404	.0947579	6.20	0.000	.4014183 .7728625
rho						
	_cons	.2781695	.2672189	1.04	0.298	-.24557 .8019089

Cuando se especifica heteroscedasticidad, la estructura de los siguientes elementos se modifica (además de las modificaciones presentadas en los pasos previos):

$$\hat{\Psi}_{\rho}^{\delta\delta} = \frac{1}{n} H' \hat{\Sigma}_{\rho} H,$$

$$\hat{\Psi}_{\rho}^{\delta\rho} = \frac{1}{n} H' \hat{\Sigma}_{\rho} \left\{ a_{1\rho}, a_{2\rho} \right\}.$$

Nuevamente, en Stata basta con incorporar al comando `spivreg`, la sub-opción `het`:

```
. spivreg lny95 lnslgd, id(id) el(Winv2_st) dl(Winv2_st) het
```

```
Estimating rho using 2SLS residuals
```

```

Iteration 0:   GMM criterion = .00004497
Iteration 1:   GMM criterion = .00004009
Iteration 2:   GMM criterion = .00004009

```

```
Estimating rho using GS2SLS residuals
```

```

Iteration 0:   GMM criterion = .01580705
Iteration 1:   GMM criterion = .01326804
Iteration 2:   GMM criterion = .01324619
Iteration 3:   GMM criterion = .01324598
Iteration 4:   GMM criterion = .01324598

```

```

Spatial autoregressive model                Number of obs   =       91
(GS2SLS estimates)

```

	lny95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lny95						
	lnslgd	.8578936	.1515402	5.66	0.000	.5608802 1.154907
	_cons	3.283914	1.066033	3.08	0.002	1.194528 5.3733

lambda							
_cons		.5889115	.1203226	4.89	0.000	.3530835	.8247394
rho							
_cons		.3361195	.3225898	1.04	0.297	-.296145	.968384

Modelo espacial de Durbin

El modelo *SDM* presenta una estructura de estimación similar al modelo *SLM*, la principal diferencia es la introducción de variables explicativas rezagadas espacialmente, WX .

La introducción de estas variables exógenas rezagadas tiende a generar problemas de multicolinealidad en la estimación *IV*. El motivo se debe a los instrumentos utilizados para la variable endógena Wy :

$$H = [X, WX, W^2X].$$

Al introducir como variables explicativas WX , el conjunto de instrumentos pasa a ser:

$$H = [X, WX, W^2X, WX, W(WX), W^2(WX)],$$

en donde puede observarse que WX y W^2X se encuentran repetidos en el conjunto de instrumentos. Luego, los instrumentos disponibles para la estimación del *SDM* son:

$$H = [X, WX, W^2X, W^3X]. \quad (3.4.5)$$

Siguiendo con el ejemplo aplicado en Stata:

```
. spivreg lny95 lnslgd wx_lnslgd, id(id) dl(Winv2_st)
note: 2 instruments omitted from the H matrix because of collinearity
see e(H_omitted) for names of the omitted instruments
Spatial autoregressive model          Number of obs   =          91
(GS2SLS estimates)
```

lny95							
lnslgd		.7095822	.1135207	6.25	0.000	.4870857	.9320788
wx_lnslgd		-1.539129	.5908767	-2.60	0.009	-2.697226	-.3810316
_cons		-4.056033	2.77482	-1.46	0.144	-9.49458	1.382515
lambda							
_cons		1.497175	.3427818	4.37	0.000	.8253347	2.169015

Esta estimación tiene dos puntos a destacar. El primero es referido a la omisión de dos instrumentos. Esta omisión es un problema habitual en la estimación por *IV*/*GMM* de modelos

espaciales ya que al incluir como variables explicativas los rezagos espaciales, entonces los mismos no pueden ser instrumentos para el rezago endógeno. Mediante el comando `ereturn list` puede visualizarse a cada instrumento omitido:

```
. ereturn list
e(H_omitted) : "wx_lnsldg lnsldg*W^2"
```

Otro punto a destacar es el valor que toma el coeficiente `lambda` que no se encuentra en el rango de estabilidad. Esto se debe a que la estimación IV/GMM no impone ninguna restricción sobre el espacio paramétrico siendo una de las debilidades del proceso de estimación.

Tal como sucedía en el modelo *SLM*, el modelo puede corregirse por heteroscedasticidad de las innovaciones (resultados que no se presentaron).

Modelo de Cliff-Ord

Sin mayor detalle que el expuesto en el modelo *SARAR*, la estimación del modelo de Cliff-Ord supone utilizar el conjunto de instrumentos del modelo *SDM*.

```
. spivreg lny95 lnsldg wx_lnsldg, id(id) dl(Winv2_st) el(Winv2_st)
note: 2 instruments omitted from the H matrix because of collinearity
see e(H_omitted) for names of the omitted instruments
```

Estimating rho using 2SLS residuals

```
Iteration 0: GMM criterion = .00011272
Iteration 1: GMM criterion = .00010807
Iteration 2: GMM criterion = .00010807
```

Estimating rho using GS2SLS residuals

```
Iteration 0: GMM criterion = .00664606
Iteration 1: GMM criterion = .00334232
Iteration 2: GMM criterion = .00333935
Iteration 3: GMM criterion = .00333935
```

```
Spatial autoregressive model          Number of obs   =       91
(GS2SLS estimates)
```

	lny95	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lny95							
lnsldg		.710113	.1148064	6.19	0.000	.4850966	.9351295
wx_lnsldg		-1.522226	.5834616	-2.61	0.009	-2.66579	-.3786627
_cons		-3.98205	2.855745	-1.39	0.163	-9.579208	1.615108
lambda							
_cons		1.487987	.3516368	4.23	0.000	.7987918	2.177183
rho							
_cons		.1632927	.2785443	0.59	0.558	-.382644	.7092294

Los comentarios sobre la estimación *SDM* son válidos para este modelo: problemas con algunos instrumentos y estimación del coeficiente espacial fuera del área de estabilidad.

Un último punto a considerar es la posibilidad de incorporar variables endógenas no espaciales, además del tratamiento por heteroscedasticidad. El comando **spivreg** permite trabajar simultáneamente con ambos tipos de endogeneidad.

Capítulo 4

Análisis espacial local

Los modelos presentados a lo largo de este libro tratan con dependencia global, significando que los resultados obtenidos son aplicables a toda la muestra de regiones. Sin embargo, si la relación examinada varía a lo largo de las regiones, los resultados globales tendrán una limitada aplicación si el investigador está interesado en una zona o región particular. En este capítulo exploraremos estadísticos y técnicas econométricas locales que permiten establecer el foco de atención en la presencia de diferencias a través del espacio.

Los estadísticos locales capturan una heterogeneidad no detectada por los tests globales y permiten explorar un camino alternativo de modelización espacial: la heterogeneidad espacial. La heterogeneidad espacial es vista como un síntoma de no estacionariedad espacial y forma parte del análisis espacial local que incluye métodos de regresiones geográficamente ponderadas, modelos de expansión de Casetti, entre otros.

4.1 Estadísticos locales

En primeros capítulos se ha considerado que la dependencia espacial puede ser resumida en un único valor surgido como una medida promedio de toda la muestra. Sin embargo, si la relación que está siendo investigada varía significativamente a través del espacio, puede ser conveniente avanzar en el estudio de dependencia local.

Una primera aproximación local puede realizarse con algún estadístico univariado. [Anselin \(1995\)](#) define un conjunto de indicadores locales de asociación espacial comúnmente denominados LISA. Los estadísticos LISA permiten la descomposición de indicadores globales, tal que cumplan con dos requisitos:

1. El valor LISA de cada observación indica el alcance de la significancia espacial alrededor de la observación.
2. La suma de todos los LISA es proporcional al indicador global de asociación espacial.

Estadístico Local I de Moran

Un estadístico que cumple con las condiciones LISA es el estadístico local I de Moran definido, para la observación i —ésima, como:

$$I_i = \frac{n(y_i - \bar{y})}{\sum_{j=1}^n (y_j - \bar{y})^2} \sum_{j=1}^n w_{ij} (y_j - \bar{y}), \quad (4.1.1)$$

donde w_{ij} es un elemento de la matriz de pesos espaciales W . En el caso de que W contenga islas, observaciones sin vecinos, el valor del estadístico local será siempre 0 para estas observaciones.

Bajo el supuesto de aleatorización completa, el valor esperado de I_i es:

$$E[I_i] = -\frac{1}{n-1} \sum_{j=1}^n w_{ij}. \quad (4.1.2)$$

La varianza de I_i , suponiendo completa aleatorización, es

$$\text{Var}[I_i] = \frac{w_{i(2)}(n - b_2)}{n - 1} + \frac{2w_{i(kh)}(2b_2 - n)}{(n - 1)(n - 2)} - \left(\frac{\sum_{j=1}^n w_{ij}}{n - 1} \right)^2, \quad (4.1.3)$$

donde $b_2 = m_4/m_2^2$ y $m_r = \sum_{i=1}^n z^r/n$, con $z = (y_i - \bar{y})$, $w_{i(2)} = \sum_{j=1, j \neq i}^n w_{ij}^2$ y $2w_{i(kh)} = \sum_{k=1, k \neq i}^n \sum_{h=1, h \neq i}^n w_{ik}w_{ih}$.

Mediante la esperanza y la varianza estimadas, la significancia del estadístico I_i puede ser contrastada sobre la base de la distribución normal donde el Moran Local es estandarizado como:

$$z[I_i] = \frac{[I_i - E[I_i]]}{\sqrt{\text{Var}[I_i]}}. \quad (4.1.4)$$

El análisis del test local de Moran permite obtener la clasificación de las observaciones *significativas* en 4 grupos: (1) Alto-Alto, valores altos de y_i y altos de Wy_i ; (2) Bajo-Bajo, valores bajos de y_i y bajos de Wy_i ; (3) Bajo-Alto, valores bajos de y_i y altos de Wy_i ; y (4) Alto-Bajo, valores altos de y_i y bajos de Wy_i . El diagrama de Moran ayuda a visualizar los patrones espaciales representados por los 4 grupos antes definidos.

Pisati (2001) ha generado el comando para aplicar el análisis LISA en Stata. La secuencia del comando es la siguiente:

```
. spatlsa U2012, w(W5st) id(id) moran sort

Measures of local spatial autocorrelation
Weights matrix
-----
Name: W5st
Type: Distance-based (binary)
Distance band: c1.c2 < d <= c3.c4
Row-standardized: Yes
-----

Moran's Ii (U2012)
-----
id | Ii E(Ii) sd(Ii) z p-value*
-----+-----
62 | -0.431 -0.006 0.434 -0.979 0.164
```

```

          90 | -0.207  -0.006   0.434  -0.464   0.321
(Resultados omitidos)
          142 | 10.418  -0.006   0.434  24.010   0.000
          150 | 11.001  -0.006   0.434  25.353   0.000

```

```
-----
*1-tail test
```

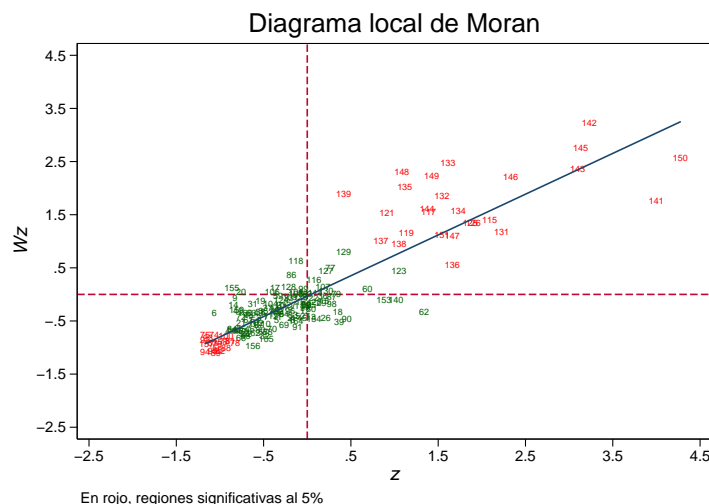
Este comando arroja los resultados para cada observación. Sin embargo, la forma más simple de visualizar los resultados es mediante el diagrama de dispersión de Moran, en donde se muestran las regiones que son significativas y mediante el mapa de clusters. [Pisati \(2012\)](#) desarrolló un comando simple, presentado aquí bajo el nombre `genmspi`, que permite realizar ambas tareas.

```

. genmspi U2012, w(W5st)

. graph twoway (scatter Wstd_U2012 std_U2012 if pval_U2012>=0.05, msymbol(i) mlabel (id) ///
mlabsize(*0.6) mlabpos(c)) (scatter Wstd_U2012 std_U2012 if pval_U2012<0.05, msymbol(i) ///
mlabel (id) mlabsize(*0.6) mlabpos(c) mlabcol(red)) (lfit Wstd_U2012 std_U2012), yline(0, ///
lpattern(--)) xline(0, lpattern(--)) xlabel(-2.5(1)4.5, labsize(*0.8)) xtitle("{it:z}") ///
ylabel(-2.5(1)4.5, angle(0) labsize(*0.8)) ytitle("{it:Wz}") legend(off) scheme(sicolor) ///
title("Diagrama local de Moran") note("En rojo, regiones significativas al 5%")

```



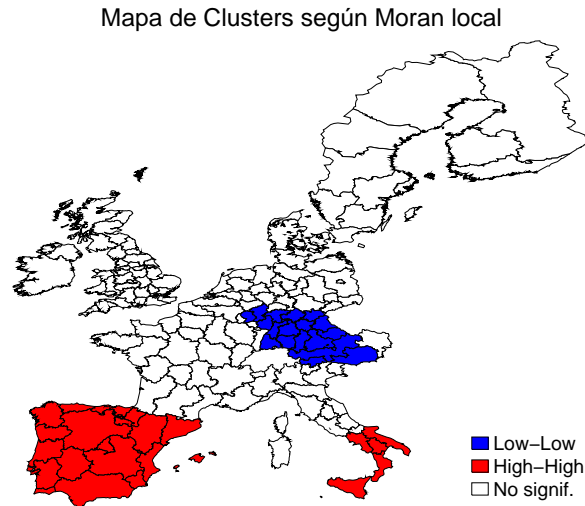
El diagrama de dispersión de Moran identifica a las regiones con valores significativamente diferentes de cero considerando un nivel de significancia del 5%.

En el mapa de clusters pueden visualizarse dos clusters Alto-Alto (hot spots) que se corresponden con el cuadrante superior del diagrama de Moran previo. También hay un cluster Bajo-Bajo (cold spot).

```

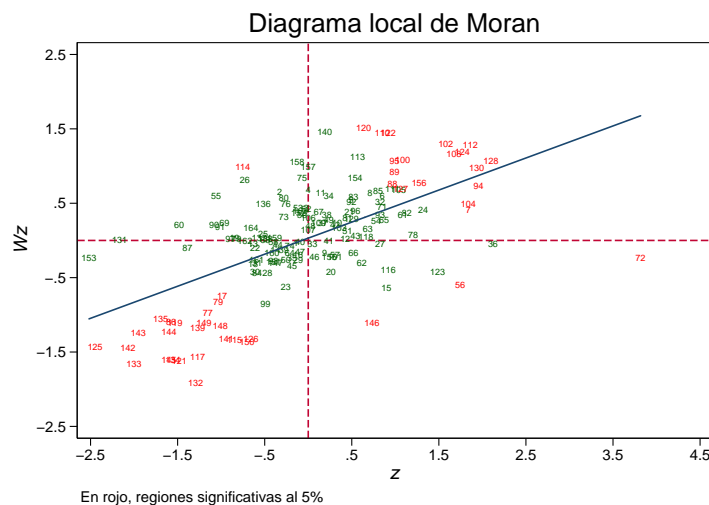
. spmap msp_U2012 using coord, id(id) clmethod(unique) title("Mapa de Clusters según Moran ///
local") legend(size(medium) position(4)) ndl("No signif.") fcolor(blue red)

```



Aplicando similares comandos para la migración neta, se obtiene una mayor cantidad de agrupamientos.

```
. genmspi NM2012, w(W5st)
. graph twoway (scatter Wstd_NM2012 std_NM2012 if pval_NM2012>=0.05, msymbol(i) mlabel (id) ///
mlabsize(*0.6) mlabpos(c)) (scatter Wstd_NM2012 std_NM2012 if pval_NM2012<0.05, msymbol(i) ///
mlabel (id) mlabsize(*0.6) mlabpos(c) mlabcol(red)) (lfit Wstd_NM2012 std_NM2012), yline(0, ///
lpattern(--)) xline(0, lpattern(--)) xlabel(-2.5(1)4.5, labsize(*0.8)) xtitle("{it:z}") ///
ylabel(-2.5(1)2.5, angle(0) labsize(*0.8)) ytitle("{it:Wz}") legend(off) scheme(sicolor) ///
title("Diagrama local de Moran") note("En rojo, regiones significativas al 5%")
```

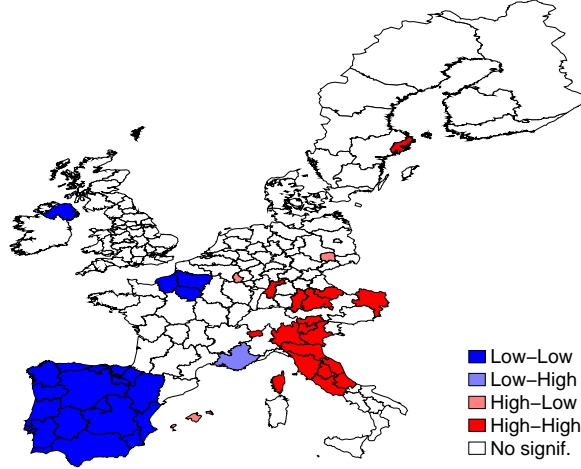


En el mapa de cluster de migración neta se visualiza un cluster Bajo-Bajo que coincide con el cluster Alto-Alto de desempleo. A su vez, hay observaciones que se identifican en los cuadrantes Bajo-Alto y Alto-Bajo que pueden considerarse en transición o como observaciones atípicas. La

distribución de los clusters Alto-Alto es más dispersa geográficamente que el observado para el desempleo.

```
. spmap msp_NM2012 using coord, id(id) clmethod(unique) title("Mapa de Clusters ///  
según Moran local") legend(size(medium) position(4)) ndl("No signif.") ///  
fcolor(blue blue*0.5 red*0.5 red)
```

Mapa de Clusters según Moran local



Estadísticos de Getis y Ord

Getis & Ord (1992) proponen dos estadísticos espaciales de dependencia local: el estadístico G_i , donde el valor de la localización i es excluido en el cómputo del estadístico, y el estadístico G_i^* donde se considera el valor de localización i —ésima. La idea de estos autores es que la asociación espacial puede ser localmente heterogénea aún si se observase autocorrelación espacial de forma global.

La expresión del estadístico G_i es igual a:

$$G_i = \frac{\sum_{j=1}^n w_{ij}(d) x_j}{\sum_{j=1}^n x_j}, \quad j \neq i, \quad (4.1.5)$$

donde el peso espacial está definido para una función de distancia con umbral d .

El valor esperado del test, bajo el supuesto de completa aleatoriedad, depende del número de vecinos locales:

$$E(G_i) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n w_{ij}(d). \quad (4.1.6)$$

De manera similar, el estadístico local G_i^* es computado como:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij}(d) x_j}{\sum_{j=1}^n x_i}, \quad (4.1.7)$$

donde el valor esperado, nuevamente bajo el supuesto de completa aleatoriedad, es:

$$E(G_i^*) = \frac{1}{n} \sum_{j=1}^n w_{ij}(d). \quad (4.1.8)$$

Getis & Ord (1992) y Ord & Getis (1995) proveen los valores esperados y varianzas de ambos estadísticos. Sus distribuciones se aproximan a la normal. No hay una guía teórica sobre cuál de los dos estadísticos debe usarse para determinadas situaciones y las diferencias entre ambos tiende a ser pequeña.

Como una forma equivalente al test de Moran local, la literatura ha adoptado para la detección de clusters utilizar el test G^* ya que brinda la usual definición de cluster entre la relación que existe de la observación y su vecindario local.

El estadístico G^* solo clasifica a las observaciones en dos grupos: clusters de altos valores (hotspots) y clusters de bajos valores (coldspots), difiriendo de la clasificación en cuatro grupos brindada por el estadístico local de Moran.

La estimación del test G^* ha sido recientemente implementada en Stata (Kondo, 2016). El comando es muy eficiente en el cómputo ya que no necesita generar una matriz de pesos espaciales de orden $(n \times n)$ y permite generar tres tipos de matrices de distancia.

El primer tipo es una matriz binaria como:

$$w_{ij}(d) = \begin{cases} 1, & \text{si } d_{ij} \leq d, \\ 0, & \text{en otro caso.} \end{cases}$$

Los elementos de la diagonal principal toman el valor 1.

Además, puede utilizarse una matriz espacial no binaria con función potencia:

$$w_{ij}(d) = \begin{cases} (a + d_{ij})^{-\delta}, & \text{si } d_{ij} \leq d, \quad \forall i, j, \quad \delta > 0, \\ 0, & \text{en otro caso,} \end{cases}$$

donde δ es un parámetro alisado de distancia y a una constante que evita pesos $w_{ii} = \infty$ cuando la distancia es nula.

Otra alternativa es el caso exponencial que tiene la siguiente forma:

$$w_{ij}(d) = \begin{cases} \exp(-\delta d_{ij}), & \text{si } d_{ij} \leq d, \quad \forall i, j, \quad \delta > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

El comando de Stata utiliza el test estandarizado

$$G_i^* - \text{estandar} = \frac{G_i^* - E(G_i^*)}{\sqrt{\text{Var}(G_i^*)}},$$

y genera dos nuevas variables, el $G_i^* - \text{estandar}$ para cada región y su correspondiente p -valor. Más detalle puede consultarse a Kondo (2016).

Aplicando el comando a las dos variables analizadas previamente pueden obtenerse los siguientes resultados y sus respectivos mapas de clusters.

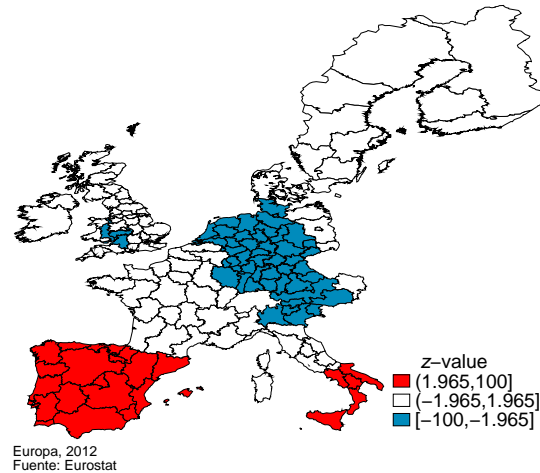
```
. getisord U2012, lat(y_c) lon(x_c) swm(bin) dist(300) dunit(km) detail approx
. spmap go_z_U2012_b using coord, id(id) clmethod(custom) ///
```

```

clb(-100 -2.576 -1.965 1.965 2.576 100) legtitle("{it: z}-value") ///
legend(size(medium) pos(5)) title("Mapa de Clusters según Getis-Ord") ///
legstyle(1) fcolor(ebblue eltblue white orange red) ///
note("Europa, 2012" "Fuente: Eurostat")

```

Mapa de Clusters según Getis-Ord

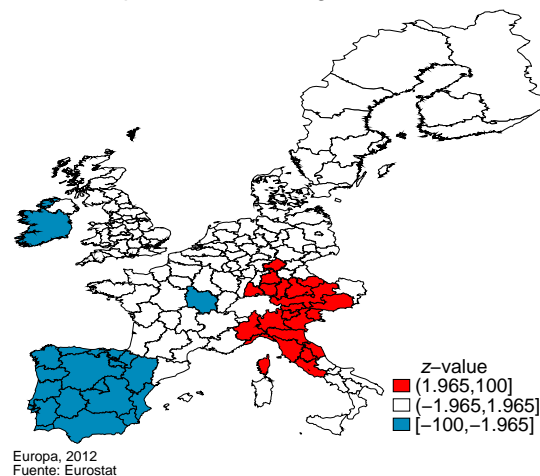


```

. getisord NM2012, lat(y_c) lon(x_c) swm(bin) dist(300) dunit(km) detail approx
. spmap go_z_NM2012_b using coord, id(id) clmethod(custom) ///
clb(-100 -2.576 -1.965 1.965 2.576 100) legtitle("{it: z}-value") ///
legend(size(medium) pos(5)) title("Mapa de Clusters según Getis-Ord") ///
legstyle(1) fcolor(ebblue eltblue white orange red) ///
note("Europa, 2012" "Fuente: Eurostat")

```

Mapa de Clusters según Getis-Ord



4.2 Método de expansión espacial

4.3 Regresiones geográficamente ponderadas

Una vez detectada la variación en intensidad en el espacio, el investigador enfrenta la necesidad de modelizar dicha característica. Una alternativa es utilizar una técnica de estimación mínimo

cuadrática adaptada al caso de datos georeferenciados comúnmente conocida como *regresiones geográficamente ponderadas* (geographically weighted regressions, GWR).

Los datos georeferenciados, debido a sus particularidades, limitan el uso de las técnicas econométricas habituales. En el capítulo 4 veremos los modelos de econometría espacial más utilizados, que utilizan el supuesto de estacionariedad del proceso espacial, es decir, plantean que es posible describir la relación espacial mediante la estimación de un único parámetro para todas unidades de estudio. Ejemplos de ello son el modelo de rezago espacial y el modelo de error espacial, por mencionar a los modelos más tradicionales. En contraposición a estos modelos, GWR destaca que el proceso espacial puede ser no estacionario de modo que la relación lineal investigada varíe localmente en su intensidad.

La técnica GWR, propuesta por Brunson et al. (1996), calibra un modelo que permite el análisis local de la regresión lineal. Entre las ventajas de este método se pueden mencionar: (i) identificación de asociaciones entre las variables dependiente e independientes que varían espacialmente; y (ii) obtención de estimaciones alternativas a modelos espaciales globales, los cuales pueden ser producto de una incorrecta especificación local. En particular, la idea central al utilizar GWR es que la conducta reproductiva, como la conducta humana en general, se encuentra influenciada por el contexto local y el mismo no es constante en todo el espacio geográfico. La detección de variabilidad espacial en la relación investigada brinda indicios de la importancia de las interacciones sociales a nivel local producto de diferentes comportamientos relacionados a la cultura, actitudes y conductas reproductivas a lo largo del territorio.

La estimación básica GWR (Brunson et al., 1996, 1998) supone que los coeficientes varían espacialmente de la siguiente forma:

$$y_i = \sum_{j=1}^k \beta_j(\text{lat}_i, \text{lon}_i) x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (4.3.1)$$

donde $(y_i; x_{i1}, \dots, x_{ik})$ son observaciones de la variable dependiente y explicativas para la localidad i —ésima, identificada geográficamente por su latitud y longitud $(\text{lat}_i, \text{lon}_i)$. Los coeficientes $\beta_j(\text{lat}_i, \text{lon}_i)$ son parámetros desconocidos que dependen de las coordenadas de localización, y ε_i es un término de error con media cero y varianza σ^2 .

La ecuación (4.3.1) representa una relación lineal que puede ser estimada por medio de mínimos cuadrados ponderados. Para una localización dada, representada por $(\text{lat}_0, \text{lon}_0)$, sea $d_{0,i}$ la distancia entre $(\text{lat}_0, \text{lon}_0)$ y la localización i —ésima $(\text{lat}_i, \text{lon}_i)$. Los coeficientes en el modelo (4.3.1) en $(\text{lat}_0, \text{lon}_0)$ son estimados minimizando la siguiente función de pérdida:

$$\sum_{i=1}^n \left[y_i - \sum_{j=1}^k \beta_j(\text{lat}_0, \text{lon}_0) x_{ij} \right]^2 w(d_{0i}), \quad (4.3.2)$$

con respecto a $\beta_j(\text{lat}_0, \text{lon}_0)$, $j = 1, 2, \dots, k$, donde $w(\cdot)$ es una ponderación basada en la proximidad geográfica. Resolviendo este problema de optimización, las estimaciones $\hat{\beta}_j(\text{lat}_0, \text{lon}_0)$ de $\beta_j(\text{lat}_0, \text{lon}_0)$, $j = 1, 2, \dots, k$, para $(\text{lat}_0, \text{lon}_0)$ pueden ser expresadas en notación matricial de la siguiente forma:

$$\hat{\beta}(\text{lat}_0, \text{lon}_0) = [X'W(\text{lat}_0, \text{lon}_0)X]^{-1} X'W(\text{lat}_0, \text{lon}_0)Y, \quad (4.3.3)$$

siendo $\widehat{\beta}(\text{lat}_0, \text{lon}_0) = \begin{bmatrix} \widehat{\beta}_1(\text{lat}_0, \text{lon}_0) & \widehat{\beta}_2(\text{lat}_0, \text{lon}_0) & \cdots & \widehat{\beta}_k(\text{lat}_0, \text{lon}_0) \end{bmatrix}'$,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad W(\text{lat}_0, \text{lon}_0) = \begin{bmatrix} w(d_{01}) & 0 & \cdots & 0 \\ 0 & w(d_{02}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w(d_{0n}) \end{bmatrix},$$

y con $Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}'$.

La ponderación $w(\cdot)$ es una función *kernel* que puede especificarse de diferentes maneras. Una elección simple puede ser definir una distancia d y que w_{ij} sea igual a 1 si la observación está a una distancia menor o igual a d y que sea igual a 0 en otro caso. Sin embargo, este tipo de función genera una discontinuidad en las ponderaciones w y puede generar inestabilidad en las estimaciones debido al tipo de función definida (Fotheringham et al., 2002). Una solución a la discontinuidad es utilizar una función de ponderación suavizada, como el *kernel* gaussiano:

$$w_{ij} = \exp \left[-\frac{1}{2} (d_{ij}/h)^2 \right], \quad (4.3.4)$$

donde h es referido como el ancho de banda (bandwidth). Si i y j coinciden, la función será igual a la unidad y a medida que la distancia entre las observaciones aumenta, la función irá decreciendo.

Alternativamente, puede definirse un *kernel* doblemente cuadrático:

$$w_{ij} = \begin{cases} (1 - d_{ij}^2/h^2)^2, & \text{si } |d_{ij}/h| \leq 1, \\ 0, & \text{en otro caso.} \end{cases} \quad (4.3.5)$$

que genera un tipo de suavizado similar al gaussiano.

Ambos *kernels* son denominados de tipo fijo en términos de la forma y magnitud que adquieren bajo el espacio. Ante una distribución espacial regular de los datos (o de alta densidad), este tipo de estrategia funcionará correctamente. Sin embargo, en estudios aplicados, la distribución espacial es habitualmente poco densa y los *kernels* de tipo fijo presentan un problema potencial de estimación. Por ejemplo, puede generarse una situación extrema de imposibilidad de estimación de los parámetros debido a insuficientes datos locales. Para evitar este tipo de problema, el *kernel* puede ser de tipo adaptativo cambiando la forma y magnitud de acuerdo a la densidad de los datos.

Para realizar la adaptación del *kernel* existen diferentes opciones, una de ellas es la selección de un número estable de k vecinos más cercanos ($k - nn$) asegurando un número suficiente de observaciones para estimar el *kernel*. Por ejemplo, el *kernel* gaussiano será definido como:

$$w_{ij} = \begin{cases} \exp \left[-\frac{1}{2} (d_{ij}/h_{i(k)})^2 \right], & d_{ij} < h_{i(k)} \\ 0, & \text{en otro caso.} \end{cases} \quad (4.3.6)$$

donde $h_{i(k)}$ captura el ancho de banda adaptativo definido por la distancia del $k - \text{ésimo}$ vecino. Los datos más cercanos al punto i tendrán un peso cercano a 1 e irá decreciendo según la distancia de cada vecino, hasta alcanzar al $k - \text{ésimo}$ vecino.

La estimación de los parámetros del modelo GWR dependerá de la función de ponderación o *kernel* elegido. En la ecuación (4.3.6), un valor de h muy grande tenderá a ponderar a los vecinos

con pesos próximos a 1. Inversamente, con valores bajos de ancho de banda, los pesos tenderán a 0. El problema es, entonces, cómo elegir el ancho de banda óptimo.

Uno de los procedimientos elegidos para optimizar el valor h es el procedimiento de validación cruzada (CV). Este procedimiento utiliza la expresión:

$$CV(h) = \sum_{i=1}^n [y_i - \hat{y}_{-i}(h)]^2, \quad (4.3.7)$$

como función objetivo, donde $\hat{y}_{-i}(h)$ es el valor estimado de y_i bajo un ancho de banda h con la observación de la localización $(\text{lat}_i, \text{lon}_i)$ omitida en el proceso de calibración. El valor óptimo h_0 es tal que:

$$CV(h_0) = \min_{h>0} CV(h). \quad (4.3.8)$$

Alternativamente, puede utilizarse un procedimiento generalizado de validación cruzada GCV o el criterio de información de Akaike, AIC . Una versión corregida del criterio Akaike, AIC_c , es computada mediante la siguiente fórmula:

$$AIC_c = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + \text{tr}(P)}{n - 2 - \text{tr}(P)} \right\}, \quad (4.3.9)$$

donde $\hat{\sigma}$ es la estimación máxima verosímil del parámetro de dispersión, $\text{tr}(P)$ es la traza de la matriz proyección P del modelo GWR. El mejor modelo será aquel que compute el menor valor AIC_c . Además de utilizarse para elegir el ancho de banda óptimo, el criterio AIC_c sirve como método de selección de modelos alternativos (por ejemplo, comparar la estimación MCO con la obtenida por GWR).

El supuesto inicial de GWR es que todos los coeficientes son localmente variables, pero puede suceder que algunos de ellos no varíen sustancialmente. En estos casos puede considerarse una extensión del modelo básico hacia uno de coeficientes mixtos, fijos y variables:

$$y_i = \sum_{l=1}^p \beta_l x_{il} + \sum_{j=1}^k \beta_j(\text{lat}_i, \text{lon}_i) x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (4.3.10)$$

donde β_l ($l = 1, \dots, p$) son coeficientes fijos o globales y los demás elementos varían según la localización geográfica, como en la ecuación (4.3.1).

La elección entre coeficientes fijos y variables puede realizarse por medio de la diferencia de AIC_c estimada por dos especificaciones alternativas. Por ejemplo, si el modelo a estimar contiene una constante y una pendiente, entonces, se estima un modelo GWR completo y un modelo mixto con constante fija y pendiente variable. El modelo que mejor ajuste según el AIC_c determinará si debe considerarse a la constante como fija o variable. Los mismo pasos se repiten considerando a un modelo con pendiente fija y a otro con pendiente variable. Si la diferencia entre los modelos, $AIC_c^{GWR} - AIC_c^{Mixto}$, es mayor o igual a 2 entonces el término variable debería considerarse como fijo.

Entre las diferentes aplicaciones empíricas de este método pueden mencionarse, por ejemplo, [Fotheringham et al. \(2001\)](#) en donde se analiza la variación espacial del rendimiento escolar; [Calvo & Escolar \(2003\)](#) aplican GWR para resolver problemas de sesgo de agregación espacial y autocorrelación espacial en las votaciones de la Ciudad Autónoma de Buenos Aires.

Capítulo 5

Efectos espaciales en datos de panel

5.1 Introducción

En este capítulo se realiza una breve revisión de la especificación y estimación de paneles espaciales por máxima verosimilitud. Los paneles espaciales se refieren a datos que contienen observaciones de series temporales de un número de unidades espaciales. Los datos de panel son habitualmente más informativos que datos de corte transversal, conteniendo mayor variación y menor colinealidad entre las variables. Además, este tipo de estructura de datos permite una especificación más compleja, incluyendo efectos que no son posibles de controlar bajo un corte transversal (véase [Hsiao, 2003](#), para mayor detalle).

En econometría espacial se trabaja habitualmente con paneles micro, es decir, con un número amplio de unidades de corte transversal y un número reducido de periodos temporales. Por otra parte, los paneles espaciales requieren que los datos del panel se encuentren balanceados en el sentido que las mismas unidades espaciales deben ser observadas en cada uno de los periodos. Esto plantea un problema adicional ya que los datos de panel tienden a sufrir el fenómeno de desgaste muestral, es decir, pérdida información entre diferentes periodos, ocasionando paneles desbalanceados.

El análisis comenzará por el modelo más simple de datos de panel, denominado *pooled*:

$$y_{it} = x_{it}\beta + u_{it}, \quad (5.1.1)$$

donde i es un índice de corte transversal (unidades espaciales), con $i = 1, \dots, n$, y t es un índice temporal (periodos temporales), con $t = 1, \dots, T$; y_{it} es una observación de la variable dependiente en la unidad i y tiempo t , x_{it} es un vector fila ($1 \times k$) de observaciones de variables independientes, y β es un vector de ($k \times 1$) parámetros desconocidos. Además, u_{it} es un término de errores aleatorios.

La estimación por *MCO* de un modelo como el (5.1.1) puede detectar dependencia espacial pero no genera resultados confiables ya que no considera que las observaciones del corte transversal se repiten en el tiempo, violando uno de los supuestos habituales del término de error que es la no-correlación de los errores. Básicamente, el término de error puede descomponerse en tres componentes:

$$u_{it} = \mu_i + \phi_t + \varepsilon_{it},$$

donde μ_i es un efecto específico, invariante temporalmente, para cada región y ϕ_t es un efecto temporal específico común para todas las regiones, invariante para el corte transversal. La inclusión de estos efectos específicos permite controlar por variables invariantes cuya omisión puede sesgar las estimaciones.

Los efectos específicos pueden ser tratados como efectos fijos (**fe**) o efectos aleatorios (**re**). Si los modelos de efectos aleatorios son una especificación apropiada en el contexto espacial es controversial y aún es un tema no cerrado. Cuando el modelo de efectos aleatorios es implementado, las unidades de observación deben ser representativas de una gran población, y el número de unidades debe potencialmente ser capaz de tender a infinito. Por su parte, cuando se implementan los efectos fijos la inferencia es condicional tal que los resultados son válidos para los individuos observados en el panel y no pueden generalizarse a toda la población.

Elhorst (2014, pp. 55) comenta que existen dos tipos de asintótica comúnmente utilizadas en el contexto espacial. Una de ellas es la estructura “infill”, donde las observaciones espaciales se encuentran delimitadas en el espacio geográfico a medida que $n \rightarrow \infty$. Otra estructura es “increasing domain”, donde los límites geográficos de las regiones se expanden cuando $n \rightarrow \infty$. Por otra parte, hay dos tipos de diseño muestral en el espacio: (a) diseño estocástico donde las unidades espaciales son aleatoriamente seleccionadas; y (b) diseño fijo donde las unidades espaciales se encuentran relacionadas a un espacio no aleatorio, posiblemente irregular.

La literatura de econometría espacial se enfoca principalmente en la estructura asintótica “increasing domain” bajo un diseño muestral fijo esto es debido a que mayoritariamente se utilizan data areales. Por lo tanto, una posición con mayor respaldo ha sido la estimación por efectos fijos, dado que el espacio es usualmente muestreado en forma exhaustiva y la inferencia realizada puede considerarse como poblacional. De todas formas, hay argumentos para realizar estimaciones por efectos aleatorios ya que no todos los estudios utilizan de forma exhaustiva el espacio geográfico y, por otra parte, los datos de panel para un intervalo temporal pueden considerarse como una muestra representativa de una superpoblación (concepto relacionado a un proceso generador de datos espacio-temporal, ver Graubard & Korn, 2002 y las referencias allí citadas para mayor detalle).

Llegado a este punto, a continuación se presentarán los modelos espaciales para datos de panel según efectos fijos y efectos aleatorios. Si los regresores están correlacionados con los parámetros μ_i s, el estimador **fe** será consistente y el estimador **re** será inconsistente. Pero, si los regresores no se correlacionan con μ_i s, el estimador **fe** seguirá siendo consistente pero ineficiente, mientras que el estimador **re** será consistente y eficiente.

Tal como en la metodología tradicional, la elección entre estos dos modelos puede realizarse por el test de Hausman. Este test plantea como hipótesis nula la ortogonalidad de las variables con respecto a los efectos aleatorios. Está basado en la comparación de dos matrices de covarianzas bajo la hipótesis que ambas son consistentes. El test de Hausman (1978) es computado como:

$$H = \left(\hat{\beta}_{fe} - \hat{\beta}_{re} \right)' \left(\hat{V}_{fe} - \hat{V}_{re} \right)^{-1} \left(\hat{\beta}_{fe} - \hat{\beta}_{re} \right),$$

donde $\hat{\beta}_{fe}$ es el vector de coeficientes del estimador consistente **fe**, $\hat{\beta}_{re}$ es el vector de coeficientes del estimador eficiente **re**, con \hat{V}_{fe} y \hat{V}_{re} como la matriz de varianzas-covarianzas del estimador **fe** y **re**, respectivamente. Este estadístico se distribuye como una χ_q^2 , siendo q el número de coeficientes comunes en ambos modelos.

El estadístico de Hausman puede considerarse como un estadístico de validez del estimador **re**, hipótesis nula. La lógica de Hausman está basada en que, bajo hipótesis nula, ambos estimadores son consistentes. Bajo hipótesis alternativa solo uno de ellos es consistente. El rechazo de la hipótesis nula, implica evidencia sobre no-exogeneidad de los regresores y la elección del modelo de efectos fijos ya que es, al menos, consistente. En definitiva, si no se ha adoptado una posición clara sobre la naturaleza de la inferencia, la elección entre **fe** o **re** dependerá de la evidencia práctica.

En todos los modelos que se presentan a continuación la estimación se realiza por máxima verosimilitud, necesitando del supuesto de normalidad del error y también de estacionariedad temporal (ausencia de raíces unitarias). En la práctica, puede ser necesaria la transformación de las series para que estos supuestos se cumplan.

En la sección 4.2 se comienza trabajando simultáneamente con dos bases de datos: desempleo-migración europea y demanda de tabaco norteamericana. Por simplicidad, la sección 4.3 solo muestra los resultados para la demanda de tabaco en Estados Unidos. El ejemplo de desempleo-migración se utiliza para los modelos dinámicos de la sección 4.4. Los resultados omitidos se encuentran a disposición en el archivo .do que acompaña al documento.

5.2 Detección de dependencia espacial en datos de panel

Tal como sucedía bajo un corte transversal, la interdependencia espacial entre las observaciones de un panel puede tener consecuencias severas en las propiedades de los estimadores, ya sea por **fe** o **re**, generando ineficiencia e inconsistencia. De aquí la necesidad de detectar y corregir por tal dependencia.

Comenzando con un modelo pooled como el presentado en la ecuación (5.1.1), la literatura ha extendido los tests LM , presentados en el capítulo anterior, para detectar la dependencia espacial.

Anselin et al. (2008) presenta las versiones simples de los tests para detección de estructura espacial residual y sustantiva. El test de autocorrelación residual en un *pooled* es:

$$LM_{ERROR} = \frac{\left[\hat{u}'(I_T \otimes W) \hat{u} / \hat{\sigma}^2 \right]^2}{T \times T_1} \underset{as}{\sim} \chi_{(1)}^2, \quad (5.2.1)$$

donde T_1 es igual a $\text{tr}[(W' + W)W]$ y \hat{u} son los residuos MCO del modelo *pooled* y $\hat{\sigma}^2 = \hat{u}'\hat{u}/(n \times T)$.

El test para detección de dependencia sustantiva en un *pooled* tiene la siguiente estructura:

$$LM_{LAG} = \frac{\left[\hat{u}'(I_T \otimes W)y / \hat{\sigma}^2 \right]^2}{\hat{J}} \underset{as}{\sim} \chi_{(1)}^2, \quad (5.2.2)$$

con $\hat{J} = 1/\hat{\sigma}^2 \left[\left((I_T \otimes W) X \hat{\beta} \right)' M_{Tn} \left((I_T \otimes W) X \hat{\beta} \right) + T \times T_1 \hat{\sigma}^2 \right]$, con $M_{Tn} = I_{Tn} - X(X'X)^{-1}X'$.

Las versiones robustas de estos tests pueden encontrarse en Elhorst (2010b):

$$LM_{ERROR}^* = \frac{\left[\left(\frac{\hat{u}'(I_T \otimes W)\hat{u}}{\hat{\sigma}^2} \right) - T \times T_1 \hat{J}^{-1} \times \left(\frac{\hat{u}'(I_T \otimes W)y}{\hat{\sigma}^2} \right) \right]^2}{T \times T_1 \left[1 - T \times T_1 \hat{J}^{-1} \right]} \underset{as}{\sim} \chi_{(1)}^2. \quad (5.2.3)$$

La versión robusta del test LM_{LAG} , LM_{LAG}^* , es igual a:

$$LM_{LAG}^* = \frac{\left[\left(\frac{\widehat{u}'(I_T \otimes W)y}{\widehat{\sigma}^2} \right) - \left(\frac{\widehat{u}'(I_T \otimes W)\widehat{u}}{\widehat{\sigma}^2} \right) \right]^2}{\widehat{J} - T \times T_1} \underset{as}{\sim} \chi_{(1)}^2. \quad (5.2.4)$$

En Stata pueden utilizarse los comandos disponibles de estos tests para corte transversal, además de la versión panel del test I de Moran, siempre que se brinde una estructura adecuada de los datos de panel.

Por ejemplo, si se utilizan los datos de desempleo y migración neta de Europa, los tests para un modelo de regresión *pooled* pueden obtenerse con la siguiente secuencia: (1) importar la matriz de 5 vecinos más cercanos, (2) dejar solo la información de vecindad, (3) construir la matriz espacio-temporal, $W_{Tn} = I_T \otimes W$, donde \otimes es el producto de Kronecker. En Stata, estos pasos son:

```
. insheet using "W5bin.txt", delim(" ") clear (165 vars, 165 obs)
. drop in 1 (1 observation deleted)
. drop v1
. mkmat v2-v165, mat(W5nn_bin)
. save W5nn_bin.dta, replace

. mat TMAT=I(4)
. mat W5xt_bin=TMAT#W5nn_bin
. drop v2-v165
. svmat W5xt_bin
. save W5xt_bin.dta, replace
```

Una vez generada la matriz, se cargarán los datos de desempleo y migración neta en el formato requerido por Stata, esto es *long*.

```
. use migr_unemp_shp.dta, clear
. reshape long NM U, i(id) j(year)
```

Dos modificaciones a la base de datos son necesarias. La primera es ordenar los datos por año y luego por NUT2. La segunda modificación es considerar solo los últimos 4 años para reducir los tiempos de estimación de los modelos.

```
. sort year id
. drop if year<2009
```

Ahora, simplemente se repite la secuencia del capítulo de corte transversal, importando la matriz recientemente generada:

```
. spwmatrix import using W5xt_bin.dta, wname(W5xt_st) row dta

. reg U NM
```

Source	SS	df	MS	Number of obs =	656
-----+-----				F(1, 654) =	59.06
Model	1219.25133	1	1219.25133	Prob > F =	0.0000

Residual		13501.0803	654	20.6438537		R-squared	=	0.0828
-----+-----								
Total		14720.3316	655	22.4737888		Adj R-squared	=	0.0814

						Root MSE	=	4.5436

U		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		
-----+-----								
NM		-.3606983	.0469346	-7.69	0.000	-.4528589		-.2685377
_cons		9.873942	.2137589	46.19	0.000	9.454205		10.29368

```
. spatdiag, weights(W5xt_st)
```

Diagnostic tests for spatial dependence in OLS regression

Fitted model

U = NM

Weights matrix

Name: W5xt_st

Type: Distance-based (inverse distance)

Distance band: c1.c2 < d <= c3.c4

Row-standardized: Yes

Diagnostics

Test		Statistic	df	p-value
-----+-----				
Spatial error:				
Moran's I		28.056	1	0.000
Lagrange multiplier		773.334	1	0.000
Robust Lagrange multiplier		0.087	1	0.769

Spatial lag:				
Lagrange multiplier		828.024	1	0.000
Robust Lagrange multiplier		54.777	1	0.000

El diagnóstico de los residuos muestra evidencia a favor de la incorporación de un rezago espacial sustantivo. El test I de Moran también es válido, detectando autocorrelación espacial.

Este ejercicio puede repetirse para el ejemplo del consumo de tabaco. Baltagi & Li (2004) explican la demanda de cigarrillos en 46 estados norteamericanos en un periodo de 30 años, desde 1963 a 1992. La variable dependiente, $\log c$, es el logaritmo de la venta per cápita de cigarrillos en términos reales, medida en packs por personas de 14 o más años. Las variables explicativas son: $\log p$, el logaritmo del precio de venta promedio del pack de cigarrillos y $\log y$, logaritmo del ingreso disponible per cápita. Los resultados presentados a continuación siguen las estimaciones

presentadas por [Elhorst \(2014\)](#) bajo el entorno de Matlab. La matriz de pesos espaciales es especificada como una matriz de contigüidad binaria, estandarizada por fila.

```
. spwmatrix import using W_xt_bin.dta, wname(Wxt_st) row dta
```

```
DTA file W_xt_bin.dta imported successfully as a (1380 x 1380)
```

```
SPW matrix and the following action(s) taken:
```

- Spatial weights matrix created as Stata object(s): Wxt_st.
- Spatial weights matrix has been row-standardized.

```
. sort year state
```

```
. reg logc logp logy
```

Source	SS	df	MS	Number of obs = 1380		
-----+-----				F(2, 1377) = 325.34		
Model	22.3226886	2	11.1613443	Prob > F = 0.0000		
Residual	47.2406342	1377	.034306924	R-squared = 0.3209		
-----+-----				Adj R-squared = 0.3199		
Total	69.5633228	1379	.050444759	Root MSE = .18522		
-----+-----						
logc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----+-----						
logp	-.8590232	.0341394	-25.16	0.000	-.925994	-.7920525
logy	.267733	.0246803	10.85	0.000	.2193179	.3161481
_cons	3.485067	.1133264	30.75	0.000	3.262756	3.707378

```
. spatdiag, weights(Wxt_st)
```

```
Diagnostic tests for spatial dependence in OLS regression
```

```
Fitted model
```

```
logc = logp + logy
```

```
Weights matrix
```

```
Name: Wxt_st
```

```
Type: Distance-based (inverse distance)
```

```
Distance band: c1.c2 < d <= c3.c4
```

```
Row-standardized: Yes
```

```
Diagnostics
```

Test	Statistic	df	p-value
-----+-----			
Spatial error:			
Moran's I	12.491	1	0.000

Lagrange multiplier		153.040	1	0.000
Robust Lagrange multiplier		144.838	1	0.000
Spatial lag:				
Lagrange multiplier		66.466	1	0.000
Robust Lagrange multiplier		58.264	1	0.000

Los resultados son similares a los obtenidos por el programa Matlab (véase [Elhorst, 2014](#), tabla 3.2, pp. 62), se rechazan las dos versiones robustas de los tests *LM* de dependencia espacial sustantiva y residual.

La estimación del modelo *pooled* puede extenderse incorporando efectos fijos individuales y temporales en ambos conjuntos de datos. Otra extensión puede ser la inclusión de efectos espaciales en las variables explicativas, convirtiendo al modelo inicial en un modelo *SLX*. El archivo de Stata .do contiene los pasos para realizar estas modificaciones propuestas.

La detección de dependencia espacial ha utilizado herramientas diseñadas para datos de corte transversal. Alternativamente, pueden utilizarse otros tests generales que permiten detectar la presencia de dependencia espacial. Recuérdese que, bajo un corte transversal, la dependencia espacial ha sido reducida a la estimación de un parámetro debido a la insuficiente información para estimar la correlación entre pares de observaciones. Con datos de panel, esta limitación puede ser sorteada y es posible obtener estimaciones de la correlación entre las observaciones del corte transversal.

Un estadístico que detecta dependencia de corte transversal es el CD de [Pesaran \(2004\)](#):

$$CD = \sqrt{\frac{2T}{n(n-1)}} \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{\rho}_{ij} \right),$$

donde $\hat{\rho}_{ij}$ es la estimación muestral de la correlación por pares de los residuos:

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^T \hat{u}_{it} \hat{u}_{jt}}{\left(\sum_{t=1}^T \hat{u}_{it}^2 \right)^{1/2} \left(\sum_{t=1}^T \hat{u}_{jt}^2 \right)^{1/2}},$$

siendo \hat{u}_{it} los residuos estimados ya sea por **fe** o **re**.

Bajo hipótesis nula de no-dependencia de corte transversal, el test $CD \xrightarrow{d} \mathcal{N}(0, 1)$ para $n \rightarrow \infty$ y T suficientemente grande. Este estadístico puede ser aplicado a una amplia variedad de paneles, incluyendo dinámicos y no-estacionarios.

En Stata, la implementación del *CD* se realiza por el comando **xtcsd** desarrollado por [De Hoyos & Sarafidis \(2006\)](#), posterior estimación del panel por **fe** o **re**:

```
. xtreg U NM, fe
    (omitted product)

. xtcsd, pes abs
```



```
Pesaran's test of cross sectional independence =    9.073, Pr = 0.0000
Average absolute value of the off-diagonal elements =    0.544
```

```
. xtreg U NM, re
  (omitted product)
```

```
. xtcsd, pes abs
```

```
Pesaran's test of cross sectional independence =    9.189, Pr = 0.0000
Average absolute value of the off-diagonal elements =    0.544
```

La opción **abs** muestra el coeficiente de correlación estimado en valor absoluto. El *CD* detecta dependencia espacial en el panel y esto lleva a buscar diferentes especificaciones incorporen explícitamente estructuras espaciales.

Para el modelo de consumo de tabaco, usando efectos fijos y aleatorios, los resultados son:

```
. xtreg logc logp logy, fe
  (omitted product)
```

```
. xtcsd, pes abs
```

```
Pesaran's test of cross sectional independence =   42.821, Pr = 0.0000
Average absolute value of the off-diagonal elements =    0.409
```

```
. xtreg logc logp logy, re
  (omitted product)
```

```
. xtcsd, pes abs
```

```
Pesaran's test of cross sectional independence =   43.171, Pr = 0.0000
Average absolute value of the off-diagonal elements =    0.411
```

5.3 Estimación de modelos espaciales estáticos

En esta sección se explorarán las diferentes alternativas de especificación comenzando desde las más simple hasta alcanzar las más complejas.

Modelo de Rezago Espacial

El primer modelo a presentar es por efectos fijos. En este caso, modelo *SLM* incluye una variable endógena rezagada como explicativa y una estructura de error con tal como:

$$y_t = \rho W y_t + X_t \beta + \mu + \varepsilon_t, \quad (5.3.1)$$

$$\varepsilon_t \sim \mathcal{N} [0, \sigma_\varepsilon^2 I_n],$$

donde

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{pmatrix}, \quad X_t = \begin{pmatrix} x_{11t} & x_{21t} & \cdots & x_{k1t} \\ x_{12t} & x_{22t} & \cdots & x_{k2t} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1nt} & x_{2nt} & \cdots & x_{knt} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}.$$

La ecuación (5.3.1) puede expresarse de una forma más compacta como:

$$Y = \rho (I_T \otimes W) Y + X \beta + (1_T \otimes I_n) \mu + \varepsilon,$$

$$Y = [I_{nT} - \rho (I_T \otimes W)]^{-1} [X \beta + (1_T \otimes I_n) \mu + \varepsilon],$$

$$\varepsilon \sim \mathcal{N} [0, \sigma_\varepsilon^2 I_{nT}],$$

donde \otimes es el producto de Kronecker.

La función de log-verosimilitud puede expresarse como:

$$L = -\frac{nT}{2} \ln \sigma_\varepsilon^2 + T \ln |A| - \frac{1}{2\sigma_\varepsilon^2} \{ [I_{nT} - \rho (I_T \otimes W)] Y - X \beta - (1_T \otimes I_n) \mu \}' \{ [I_{nT} - \rho (I_T \otimes W)] Y - X \beta - (1_T \otimes I_n) \mu \},$$

donde $A = I_n - \rho W$.

La estimación de este modelo puede resumirse en los siguientes pasos: Primero, para eliminar los efectos fijos se debe calcular los desvíos respecto a la media temporal de cada unidad de corte transversal: $y_{it}^* = y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it}$ y $x_{it}^* = x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it}$.

Segundo, denotando como b_0 y b_1 a los estimadores *MCO* obtenidos de las estimaciones de Y^* y $(I_T \otimes W) Y^*$ sobre X^* , donde el asterisco simboliza a las variables expresadas en desvíos, y e_0^* y e_1^* los residuos correspondientes. Luego, el estimador *MV* de ρ es obtenido maximizando la log-verosimilitud concentrada:

$$L_c = C - \frac{nT}{2} \ln [(e_0^* - \rho e_1^*)' (e_0^* - \rho e_1^*)] + T \ln |I_n - \rho W|,$$

donde C es una constante que no depende de ρ .

Tercero, los estimadores β y σ^2 son computados considerando la estimación de ρ :

$$\begin{aligned} \beta &= b_0 - \rho b_1 = (X^{*'} X^*)^{-1} X^{*'} [Y^* - \rho (I_T \otimes W) Y^*], \\ \sigma^2 &= \frac{1}{nT} (e_0^* - \rho e_1^*)' (e_0^* - \rho e_1^*). \end{aligned}$$

Finalmente, la matriz de varianzas-covarianzas asintótica es computada. Para mayor detalle véase [Anselin et al. \(2008\)](#) y [Elhorst \(2014\)](#).

La estimación por efectos aleatorios del *SLM* se modifica de la siguiente manera:

$$\begin{aligned} y_t &= \rho W y_t + X_t \beta + \underbrace{\mu + \varepsilon_t}_{u_t}, \\ \mu &\sim \mathcal{N}[0, \sigma_\mu^2 I_n], \quad \varepsilon_t \sim \mathcal{N}[0, \sigma_\varepsilon^2 I_n]. \end{aligned} \quad (5.3.2)$$

Re-expresando el modelo (5.3.2) de la siguiente manera:

$$\begin{aligned} Y &= \rho (I_T \otimes W) Y + X \beta + u, \\ u &= (1_T \otimes I_n) \mu + \varepsilon, \\ u &\sim \mathcal{N}[0, \Sigma_{nT}], \end{aligned}$$

siendo $\Sigma_{nT} = \sigma_\mu^2 [(11') \otimes I_n] + \sigma_\varepsilon^2 [I_{nT}]$, tal que la inversa de Σ_{nT} es:

$$\Sigma_{nT}^{-1} = \frac{1}{\sigma_\varepsilon^2} \left\{ \theta \left[\left(\frac{11'}{T} \right) \otimes I_n \right] + [Q_T \otimes I_n] \right\}$$

con $\theta = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T \sigma_\mu^2}$; $Q_T = I_T - \frac{11'}{T}$.

La función de log-verosimilitud del modelo es:

$$\begin{aligned} L &= -\frac{nT}{2} \ln \sigma_\varepsilon^2 + T \ln |A| + \frac{n}{2} \ln \theta \\ &\quad - \frac{1}{2} \{ [I_{nT} - \rho (I_T \otimes W)] Y - X \beta \}' \Sigma_{nT}^{-1} \{ [I_{nT} - \rho (I_T \otimes W)] Y - X \beta \}, \end{aligned} \quad (5.3.3)$$

Para estimar este modelo se deben transformar las variables de la siguiente manera: $y_{it}^\bullet = y_{it} - (1 - \theta) \frac{1}{T} \sum_{t=1}^T y_{it}$ y $x_{it}^\bullet = x_{it} - (1 - \theta) \frac{1}{T} \sum_{t=1}^T x_{it}$. Dado θ , la log-verosimilitud es similar a la presentada para el caso de efectos fijos. Esto implica que puede utilizarse el mismo procedimiento reemplazando al supra-índice * por \bullet . Obtenidas las estimaciones de β , ρ y σ^2 , es posible obtener una estimación de θ mediante la log-verosimilitud concentrada que dependerá de θ :

$$L_c = -\frac{nT}{2} \ln [e^\bullet(\theta)' e^\bullet(\theta)] + \frac{n}{2} \ln \theta^2.$$

En Stata, la estimación de este modelo ha sido implementado por [Belotti et al. \(2014\)](#) mediante el comando `xsmle`:

```
. xsmle logc logp logy, fe type(both) wmat(Wst) mod(sar) noeffects nolog hausman
... estimating random-effects model to perform Hausman test
```

SAR with spatial and time fixed-effects	Number of obs =	1380
Group variable: state	Number of groups =	46
Time variable: year	Panel length =	30
R-sq: within = 0.3795		
between = 0.3227		
overall = 0.2861		

Mean of fixed-effects = 1.6709							
Log-likelihood = 2470.2271							
-----+-----							
	logc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----							
Main							
	logp	-.9938784	.0399908	-24.85	0.000	-1.072259	-.915498
	logy	.4619618	.0454312	10.17	0.000	.3729184	.5510053
-----+-----							
Spatial							
	rho	.1911609	.0283223	6.75	0.000	.1356502	.2466716
-----+-----							
Variance							
	sigma2_e	.0050565	.0001931	26.19	0.000	.0046781	.0054349
-----+-----							
Ho: difference in coeffs not systematic chi2(3) = 77.10 Prob>=chi2 = 0.0000							
-----+-----							

Una virtud del comando de [Belotti et al. \(2014\)](#) es que incorpora como sub-opción el test de Hausman. El estadístico de Hausman rechaza la hipótesis nula implicando que debe estimarse por efectos fijos.

Modelo de Error Espacial

El modelo *SEM* de efectos fijos es:

$$\begin{aligned}
 y_t &= X_t \beta + \mu + \varepsilon_t \\
 \varepsilon_t &= \lambda W \varepsilon_t + \eta_t \\
 \eta_t &\sim \mathcal{N} [0, \sigma_\eta^2 I_n]
 \end{aligned}
 \tag{5.3.4}$$

Re-expresando de manera más compacta (5.3.4):

$$\begin{aligned}
 Y &= X\beta + (1 \otimes I_n) \mu + \varepsilon, \\
 \varepsilon &= (I_T \otimes B^{-1}) \eta, \quad B = [I_n - \lambda W], \\
 \varepsilon &\sim \mathcal{N} [0, \Sigma_{nT}], \quad \Sigma_{nT} = \sigma_\eta^2 \left[I_T \otimes (B' B)^{-1} \right].
 \end{aligned}$$

La función de log-verosimilitud de este modelo se expresa como:

$$\begin{aligned}
 \ell &= -\frac{1}{2} \ln |\Sigma_{nT}| - \frac{1}{2} \varepsilon' \Sigma_{nT}^{-1} \varepsilon, \\
 &= -\frac{nT}{2} \ln \sigma_\eta^2 + T \ln |B| - \frac{1}{2\sigma_\eta^2} \varepsilon' [I_T \otimes (B' B)] \varepsilon.
 \end{aligned}$$

Dado λ , los estimadores MV de β y σ^2 pueden ser resueltos de las condiciones de primer orden de la log-verosimilitud, tal que:

$$\begin{aligned}\beta &= \{[X^* - \lambda(I_T \otimes W)X^*]'[X^* - \lambda(I_T \otimes W)X^*]\}^{-1} \\ &\quad \times [X^* - \lambda(I_T \otimes W)X^*]'[Y^* - \lambda(I_T \otimes W)Y^*], \\ \sigma^2 &= \frac{1}{nT}e(\lambda)'e(\lambda),\end{aligned}$$

donde $e(\lambda) = Y^* - \lambda(I_T \otimes W)Y^* - [X^* - \lambda(I_T \otimes W)X^*]\beta$. La log-verosimilitud concentrada en función de λ toma la siguiente forma:

$$L_c = -\frac{nT}{2}\ln[e(\lambda)'e(\lambda)] + T\ln|I_n - \lambda W|.$$

Esta función es maximizada para obtener una estimación de λ mediante un procedimiento iterativo tal como se ha expuesto previamente.

En el caso de considerar un *SEM* con efecto aleatorios, la especificación vendrá dada por:

$$\begin{aligned}y_t &= X_t\beta + \underbrace{\mu + \varepsilon_t}_{u_t}, \\ \varepsilon_t &= \lambda W\varepsilon_t + \eta_t, \\ \eta_t &\sim \mathcal{N}[0, \sigma_\eta^2 I_n], \quad \mu \sim \mathcal{N}[0, \sigma_\mu^2 I_n],\end{aligned}$$

tal que su forma compacta podrá expresarse como:

$$\begin{aligned}Y &= X\beta + u, \\ u &= (1 \otimes I_n)\mu + (I_T \otimes B^{-1})\eta, \\ B &= [I_n - \lambda W], \quad u \sim \mathcal{N}[0, \Sigma_{nT}],\end{aligned}$$

siendo $\Sigma_{nT} = \sigma_\mu^2[(11') \otimes I_n] + \sigma_\eta^2[I_T \otimes (B'B)^{-1}]$ y su inversa:

$$\Sigma_{nT}^{-1} = \left(\frac{11'}{T}\right) \otimes [(B'B)^{-1} + T\varphi I_T]^{-1} + Q_T \otimes (B'B),$$

con $\varphi = \sigma_\eta^2/\sigma_\mu^2$, $Q_T = I_T - \frac{11'}{T}$.

La función de log-verosimilitud del *SEM* con efectos aleatorios es:

$$\begin{aligned}L &= -\frac{nT}{2}\ln\sigma_\eta^2 - (T-1)\ln|B| - \frac{1}{2}\ln|(B'B)^{-1} + T\varphi I_T| \\ &\quad - \frac{1}{2\sigma_\eta^2}\left\{u' \left(\frac{11'}{T}\right) \otimes [(B'B)^{-1} + T\theta I_T]^{-1} u + u'Q_T \otimes (B'B)u\right\}.\end{aligned}$$

La expresión $[(B'B)^{-1} + T\varphi I_T]$ complica el análisis y no existe una expresión simple a la que pueda reducirse. [Elhorst \(2014, pp. 51\)](#) realiza una revisión de las diferentes propuestas. En cualquier caso, se mantiene el resultado general (*MCGF*): $\tilde{\beta} = [X'\hat{\Sigma}_{nT}^{-1}X]^{-1}[X'\hat{\Sigma}_{nT}^{-1}Y]$.

En Stata, el comando `xsmle, fe ... mod(sem) hausman` permite estimar el *SEM* por efectos fijos y realizar el test de Hausman:

```
. xsmle logc logp logy, fe type(both) emat(Wst) mod(sem) nolog hausman
... estimating random-effects model to perform Hausman test
```

```
SEM with spatial and time fixed-effects          Number of obs =      1380
Group variable: state                            Number of groups =       46
Time variable: year                             Panel length =       30
R-sq:      within = 0.3508
           between = 0.3247
           overall = 0.2915
Mean of fixed-effects = 2.1691
Log-likelihood = 2470.2271
```

	logc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Main						
logp		-1.004296	.0401848	-24.99	0.000	-1.083057 - .9255358
logy		.5538486	.049324	11.23	0.000	.4571754 .6505218
Spatial						
lambda		.2400392	.0325699	7.37	0.000	.1762034 .303875
Variance						
sigma2_e		.0050002	.0001916	26.10	0.000	.0046247 .0053757
Ho: difference in coeffs not systematic chi2(3) = 78.85 Prob>=chi2 = 0.0000						

Los resultados reportados por el test de Hausman apoyan la estimación por efectos fijos.

Modelo *SARAR*

El modelo *SARAR* de efectos fijos es:

$$y_t = \rho W y_t + X_t \beta + \mu + \varepsilon_t$$

$$\varepsilon_t = \lambda W \varepsilon_t + \eta_t \quad (5.3.5)$$

$$\eta_t \sim \mathcal{N}[0, \sigma_\eta^2 I_n]$$

La estimación de este modelo combina los elementos del modelo *SLM* y *SEM*, por lo que no se presentan los detalles con el fin de evitar la duplicación de resultados.

El comando `xsmle` solo permite estimar este modelo por efectos fijos. A continuación se muestran los resultados de la estimación.

```
. xsmle logc logp logy, fe wmat(Wst) type(both) emat(Wst) mod(sac) noeffects nolog
SAC with spatial and time fixed-effects          Number of obs =      1380
Group variable: state                            Number of groups =       46
Time variable: year                             Panel length =       30
```

```

R-sq:    within = 0.3534
         between = 0.3251
         overall = 0.2913
Mean of fixed-effects = 2.1114
Log-likelihood = 1686.8515
-----
              logc |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
Main              |
      logp | -1.004669   .0402561   -24.96   0.000   -1.08357   - .9257687
      logy |  .5461385   .0561382    9.73   0.000   .4361096   .6561674
-----+-----
Spatial           |
      rho |  .0193429   .0746271    0.26   0.795   -.1269235   .1656093
      lambda | .2206412   .080468    2.74   0.006   .0629267   .3783556
-----+-----
Variance          |
      sigma2_e |  .005183   .000196   26.45   0.000   .0047989   .0055671
-----

```

La estimación del modelo *SARAR* por el comando `xsmle` solo puede realizarse por efectos fijos y por eso no puede incluirse la sob-opción `hausman`.

Modelo espacial de Durbin

La especificación del *SDM*:

$$y_t = \rho W y_t + X_t \beta + W X_t \theta + \varepsilon_t, \quad (5.3.6)$$

donde la estimación es similar a la presentada por el *SLM*, con la diferencia de que existirán efectos locales de las variables explicativas. En Stata el comando para este modelo es:

```

. xsmle logc logp logy, fe type(both) wmat(Wst) mod(sdm) hausman noeffects nolog
Warning: All regressors will be spatially lagged
convergence not achieved
... estimating random-effects model to perform Hausman test
SDM with spatial and time fixed-effects          Number of obs =      1380
Group variable: state                            Number of groups =       46
Time variable: year                             Panel length =       30
R-sq:    within = 0.4591
         between = 0.2912
         overall = 0.3110
Mean of fixed-effects = 2.2026
Log-likelihood = 1686.8515
-----
              logc |      Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
Main              |

```

	logp		-1.002454	.0401168	-24.99	0.000	-1.081081	-.9238265
	logy		.6006541	.0571792	10.50	0.000	.488585	.7127232
-----+-----								
Wx								
	logp		.0564066	.0828302	0.68	0.496	-.1059376	.2187508
	logy		-.2949799	.0770509	-3.83	0.000	-.4459968	-.143963
-----+-----								
Spatial								
	rho		.229154	.0327665	6.99	0.000	.1649328	.2933752
-----+-----								
Variance								
	sigma2_e		.0049782	.0001906	26.12	0.000	.0046046	.0053518
-----+-----								
Ho: difference in coeffs not systematic chi2(5) = 90.39 Prob>=chi2 = 0.0000								
-----+-----								

Dada la evidencia aportada por el estadístico de Hausman, no es necesario estimar el modelo *SDM* por efectos aleatorios. Una vez obtenida la estimación puede implementarse el test de factores comunes para decidir entre los modelos *SEM* y *SDM*. El test se implementa con un estadístico tipo-Wald:

```
. testnl ([Wx]logp = -[Spatial]rho*[Main]logp) ([Wx]logy = -[Spatial]rho*[Main]logy)
(1)  [Wx]logp = -[Spatial]rho*[Main]logp
(2)  [Wx]logy = -[Spatial]rho*[Main]logy
      chi2(2) =      8.62          Prob > chi2 =      0.0134
```

Rechazándose la hipótesis nula, modelo *SEM*, a favor de un modelo *SDM*.

El modelo *SDM* puede ser reducido a un *SLM* si los efectos locales no son significativos conjuntamente:

```
. test ([Wx]logp = 0) ([Wx]logy = 0)
(1)  [Wx]logp = 0
(2)  [Wx]logy = 0
      chi2(2) =     16.06          Prob > chi2 =      0.0003
```

Nuevamente, la hipótesis nula se rechaza en favor del modelo *SDM*.

Modelo de Error de Durbin

El modelo *SDEM* de efectos fijos es:

$$y_t = X_t\beta + WX_t\theta + \mu + \varepsilon_t, \quad (5.3.7)$$

$$\varepsilon_t = \lambda W\varepsilon_t + \eta_t.$$

En Stata:


```
. xsmle logc logp logy wx_logp wx_logy, fe type(both) emat(Wst) mod(sem) hausman nolog
... estimating random-effects model to perform Hausman test
SEM with spatial and time fixed-effects          Number of obs =      1380
Group variable: state                          Number of groups =       46
Time variable: year                            Panel length =       30
R-sq:      within = 0.4470
           between = 0.2989
           overall = 0.3114
Mean of fixed-effects =  2.7553
Log-likelihood = 2470.2271
```

	logc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Main						
logp		-1.01122	.0399806	-25.29	0.000	-1.089581 - .9328596
logy		.5885175	.0548176	10.74	0.000	.481077 .695958
wx_logp		-.1769101	.07763	-2.28	0.023	-.329062 -.0247582
wx_logy		-.168392	.0783297	-2.15	0.032	-.3219155 -.0148686
Spatial						
lambda		.2270169	.0327979	6.92	0.000	.1627341 .2912997
Variance						
sigma2_e		.0049798	.0001907	26.12	0.000	.0046061 .0053535

```
Ho: difference in coeffs not systematic chi2(5) = 99.14    Prob>=chi2 = 0.0000
```

Modelo de Cliff Ord

El modelo de Cliff-Ord de efectos fijos incorpora todos los efectos espaciales posibles:

$$y_t = \rho W y_t + X_t \beta + W X_t \theta + \mu + \varepsilon_t, \quad (5.3.8)$$

$$\varepsilon_t = \lambda W \varepsilon_t + \eta_t.$$

Similar a lo comentado para el modelo de corte transversal, la identificación del modelo es débil y no suele utilizarse en aplicaciones empíricas.

En Stata puede estimarse este modelo mediante la siguiente secuencia de comandos:

```
. xsmle logc logp logy wx_logp wx_logy t2-t30, fe wmat(Wst) ematrix(Wst) ///
type(ind) model(sac) noeffects nolog
SAC with spatial fixed-effects          Number of obs =      1380
Group variable: state                  Number of groups =       46
Time variable: year                    Panel length =       30
R-sq:      within = 0.6790
           between = 0.3029
```

overall = 0.4234						
Mean of fixed-effects = 3.9690						
Log-likelihood = 2470.2271						

	logc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

Main						
	logp	-1.017526	.0401194	-25.36	0.000	-1.096159 - .9388937
	logy	.5787523	.0518391	11.16	0.000	.4771496 .680355
	wx_logp	-.5603647	.1183861	-4.73	0.000	-.7923972 -.3283322
	wx_logy	.0306273	.0934856	0.33	0.743	-.1526011 .2138558
	t2	-.0467175	.0307498	-1.52	0.129	-.1069861 .013551
					
	t30	-.1624458	.0789715	-2.06	0.040	-.317227 -.0076646

Spatial						
	rho	-.396394	.0871629	-4.55	0.000	-.5672303 -.2255578
	lambda	.5386568	.0603142	8.93	0.000	.4204432 .6568704

Variance						
	sigma2_e	.0045799	.0002392	19.15	0.000	.0041111 .0050488

Los resultados del modelo Cliff-Ord dan evidencia de un resultado que se produce con frecuencia: los coeficientes espaciales toman un valor opuesto entre sí y de magnitud similar. Esto suele tomarse como evidencia de la no-identificación del modelo.

La virtud del modelo Cliff-Ord es que anida a todos los demás modelos y puede servir, ante determinados supuestos, para deducir el mejor modelo cuando se utiliza la estrategia de especificación “de lo general a lo particular”. La selección entre modelos puede ser realizada mediante el test LR, en donde el modelo restringido forma parte de la hipótesis nula.

5.4 Interpretación de los modelos con rezago espacial endógeno

Un punto de interés, también válido para los modelos de corte transversal, es la interpretación de los resultados obtenidos. Por ejemplo, en los inicios de la especialidad el modelo *SLM* fue adoptado como un modelo central debido a que brinda una estimación de la interdependencia espacial. La interpretación de los resultados obtenidos era directa, en el sentido que un valor significativo del parámetro ρ era evidencia de la importancia de los efectos desbordamiento o contagio. Sin embargo, [LeSage & Pace \(2009\)](#) destacan la errónea interpretación de estos resultados y muestran la necesidad de considerar derivadas parciales del impacto de un cambio en una unidad de las variables explicativas.

Para ver más claramente el punto destacado por [LeSage & Pace \(2009\)](#), debe tenerse presente que los efectos desbordamiento en un modelo *SDM*, por ejemplo, se deben a la presencia de Wy (efecto global) y WX (efecto local):

$$y = \rho W y + X\beta + WX\theta + u. \quad (5.4.1)$$

Supongamos que se desea analizar el impacto del incremento unitario de la k –ésima variable explicativa, este incremento es considerado igual para todas las localizaciones. Bajo un modelo de regresión no espacial, el efecto total sobre la variable dependiente será igual a la estimación del coeficiente β_k , ceteris paribus, sin importar la localización. Pero ante un modelo como el presentado en (5.4.1), el efecto total depende de las unidades vecinas de cada localización y de la magnitud de los coeficientes que acompañan a las variables espaciales.

El efecto total puede ser descompuesto en un efecto directo y otro indirecto. El efecto directo es el impacto del cambio de la variable explicativa sobre la variable dependiente en cada localidad, este efecto tenderá a ser similar al obtenido por un modelo de regresión no espacial si ρ y θ_k son cercanos a 0. El efecto indirecto se debe a la dinámica espacial generada por la presencia de ρ y θ que afectará a todas las unidades el modelo.

Bajo un modelo de Durbin pueden distinguirse dos tipos de efectos indirectos producto de la interdependencia entre las unidades. Uno es de tipo global, afectando a todas las unidades: y_i aumenta inicialmente β_{ik} unidades, y esto genera un nuevo incremento producto del cambio en las y_j 's de los J vecinos, capturado por el término endógeno $\rho W y$. Es decir, para cada i –ésima localidad existirá un nuevo impacto igual a $\rho \sum_{j \neq i}^J w_{ij} \beta_{jk}$ unidades, retroalimentando el cambio en la variable dependiente y que solo cesará al converger (suponiendo $|\rho| < 1$). Este efecto es identificado como un *efecto espacial global* para distinguirlo del *efecto espacial local* que surge del incremento unitario en la k –ésima variable explicativa de los J vecinos de la localidad i : $\sum_{j \neq i}^J w_{ij} \theta_{jk}$. Este efecto espacial refuerza al efecto inicial β_k y es de tipo local en el sentido que no posee un efecto dinamizador como el generado por la presencia de ρ .

En términos matriciales, y de forma más compacta, el *efecto marginal total* en el modelo *SDM* puede ser representado por la matriz de derivadas parciales de y respecto al cambio en una unidad de la k –ésima variable explicativa en X , dado un periodo t :

$$\begin{aligned} \left[\frac{\partial E(y)}{\partial x_{1k}} \quad \cdots \quad \frac{\partial E(y)}{\partial x_{nk}} \right]_t &= \left[\begin{array}{ccc} \frac{\partial E(y_1)}{\partial x_{1k}} & \cdots & \frac{\partial E(y_1)}{\partial x_{nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_{1k}} & \cdots & \frac{\partial E(y_n)}{\partial x_{nk}} \end{array} \right]_t, \\ &= (I_n - \rho W)_t^{-1} \left[\begin{array}{cccc} \beta_k & w_{12}\theta_k & \cdots & w_{1n}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2n}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}\theta_k & w_{n2}\theta_k & \cdots & \beta_k \end{array} \right]_t, \\ &= S_t [\beta_k I_n + \theta_k W]_t, \end{aligned} \quad (5.4.2)$$

donde w_{ij} es el elemento (i, j) –ésimo de W , β_k es el k –ésimo elemento del vector β , θ_k es el elemento k –ésimo del vector γ y S_t es igual a $(I_n - \rho W)_t^{-1}$. Obsérvese que para la estimación de los efectos marginales se utiliza el valor esperado de la forma reducida del modelo, $E(y_t) = (I - \rho W)^{-1} [X_t \beta + W X_t \theta]$.

En la expresión (5.4.2) puede observarse la presencia de un efecto directo y otro indirecto. El *efecto directo* es representado por los elementos de la diagonal principal de la ecuación (5.4.2).

Este cambio será diferente para cada región por lo que no existe un único efecto directo. Por su parte, el *efecto indirecto* (spatial spillover), proviene de los elementos fuera de la diagonal principal de (5.4.2) y, nuevamente, será diferente para cada región. En este caso se tiene una sumatoria de efectos para cada localidad.

En el caso de un *SLM*, $y_t = \rho W y_t + X_t \beta + \varepsilon_t$, los *efectos marginales totales* se simplifican tal que:

$$\begin{aligned} \left[\begin{array}{ccc} \frac{\partial E(y)}{\partial x_{1k}} & \cdots & \frac{\partial E(y)}{\partial x_{nk}} \end{array} \right]_t &= \left[\begin{array}{ccc} \frac{\partial E(y_1)}{\partial x_{1k}} & \cdots & \frac{\partial E(y_1)}{\partial x_{nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial x_{1k}} & \cdots & \frac{\partial E(y_n)}{\partial x_{nk}} \end{array} \right]_t, \\ &= (I_n - \rho W)_t^{-1} \left[\begin{array}{cccc} \beta_k & 0 & \cdots & 0 \\ 0 & \beta_k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_k \end{array} \right]_t, \\ &= S_t [\beta_k I_n], \end{aligned} \quad (5.4.3)$$

donde β_k es el k -ésimo elemento del vector β y S_t definido como en (5.4.2). El *efecto directo* será igual al que surge en el modelo *SDM*, pero el *efecto indirecto* se torna más simple siendo igual a la suma de los elementos fuera de la diagonal principal de $\beta_k S_t$, conteniendo solo efectos globales.

Ya sea que se estime un modelo *SLM* o *SDM*, siempre se tendrá efecto marginal que no es único para todas las localidades (en contraposición a lo que sucede bajo un modelo no-espacial). Debido a esta característica, LeSage & Pace (2009) proponen una medida resumen para cada uno de los efectos considerando valores promedio. Por ejemplo para un modelo *SLM*, para cada periodo temporal se estiman los siguientes efectos:

$$\begin{aligned} \overline{M}_{Total} &= \frac{\beta_k}{n} 1'_n S 1_n = \frac{\beta_k}{n} \sum_i \sum_j S_{i,j}, \\ \overline{M}_{Directo} &= \frac{\beta_k}{n} \text{tr}(S) = \frac{\beta_k}{n} \sum_i S_{i,i}, \\ \overline{M}_{Indirecto} &= \frac{\beta_k}{n} 1'_n S 1_n - \frac{\beta_k}{n} \text{tr}(S) = \frac{\beta_k}{n} \sum_i \sum_{j \neq i} S_{i,j}. \end{aligned}$$

La significancia de estos efectos pueden ser obtenidos mediante una simulación Monte Carlo por medio de shocks aleatorios en el término de error.

Es importante recordar que estos efectos deben ser cuidadosamente interpretados y no están libres de controversias ya que su obtención tiene un fuerte componente causal. Existe cierta discusión sobre la adecuada interpretación de los resultados y no es un tema cerrado, por ejemplo, algunos autores sugieren que el parámetro espacial del *SLM* (o del *SDM*) es un parámetro causal (Gibbons & Overman, 2012). Nuestra posición es que el parámetro ρ únicamente captura la dinámica espacial de la propia variable endógena y no debe interpretarse como causal. El proceso espacial guarda similitudes a lo que sucede en un proceso AR(1) en el tiempo, $y_t = \rho y_{t-1} + \varepsilon_t$, en donde ρ no se interpreta como un parámetro causal en series de tiempo. Para una discusión sobre este punto véase Herrera et al. (2014).

Una corrección adicional debe ser mencionada. Bajo la estimación de un modelo de efectos fijos *SLM*, *SDM* o incluso un *SEM*, la estimación del parámetro de dispersión σ^2 será sesgada. La magnitud del sesgo y su dirección dependerá de la especificación del modelo. Lee & Yu (2010) muestran que si los efectos fijos son individuales o temporales, solo se necesita corregir el parámetro de dispersión. Sin embargo, si los efectos fijos incluyen ambos efectos individuales y temporales, otros parámetros necesitan ser ajustados. El código `xsmle` permite realizar este tipo de corrección.

En Stata, los efectos marginales pueden obtenerse directamente bajo el comando `xsmle`:

```
. quietly xsmle logc logp logy t2-t30, fe type(ind) wmat(Wst) mod(sdm) nolog r nsim(999)
. estimate store SDMfe2
. quietly xsmle logc logp logy t2-t30, fe type(ind, leeyu) wmat(Wst) mod(sdm) nolog r ///
nsim(999)
. estimate store SDMleeyu
. estimates table SDMfe2 SDMleeyu, b(%7.3f) drop(t2-t30) star(0.1 0.05 0.01) stf(%9.0f)
```

Variable		SDMfe2	SDMleeyu
-----+-----			
Main			
logp		-1.002***	-1.002***
logy		0.601***	0.601***
-----+-----			
Wx			
logp		0.056	0.056
logy		-0.296***	-0.295***
-----+-----			
Spatial			
rho		0.229***	0.229***
-----+-----			
Variance			
sigma2_e		0.005***	0.005***
-----+-----			
Direct			
logp		-1.013***	-1.013***
logy		0.590***	0.590***
-----+-----			
Indirect			
logp		-0.215**	-0.211**
logy		-0.197**	-0.194**
-----+-----			
Total			
logp		-1.229***	-1.223***
logy		0.393***	0.396***
-----+-----			

legend: * p<.1; ** p<.05; *** p<.01

Obsérvese que la especificación del modelo puede ser refinada mediante la corrección del sesgo propuesta por [Lee & Yu \(2010\)](#) y mediante la opción robusta para el cálculo de los errores estándares. La estimación de ambas versiones del modelo *SDM* son prácticamente iguales, solo existe una modificación en el valor de los efectos indirectos. Adicionalmente, puede determinarse el número de réplicas de la simulación para mejorar la precisión de los efectos.

5.5 Estimación de modelos espaciales dinámicos

En los modelos estáticos se ha explorado la dependencia de las unidades en su dimensión espacial. En esta sección se explora la dependencia en la dimensión restante, la temporal. Ante presencia de correlación serial en el término de error de un modelo de panel, los errores estándares serán sesgados y causarán pérdida de eficiencia.

Un test que permite detectar correlación serial bajo **fe** o **re** ha sido desarrollado por [Wooldridge \(2002\)](#). El método utiliza los residuos de una regresión en primeras diferencias de la ecuación (5.1.1):

$$\begin{aligned} y_{it} - y_{it-1} &= (x_{it} - x_{it-1})\beta + u_{it} - u_{it-1}, \\ \Delta y_{it} &= \Delta x_{it}\beta + \Delta u_{it}, \end{aligned}$$

donde Δ es el operador de primeras diferencias.

El procedimiento de Wooldridge es obtener los residuos de la ecuación en primeras diferencias, \hat{u}_{it} , y bajo no-autocorrelación serial de u_{ij} , se espera $\text{corr}(\Delta u_{it}, \Delta u_{it-1}) = -0.5$. El método es robusto ante diferentes especificaciones y ha sido implementado en Stata por [Drukker \(2003\)](#) bajo el comando **xtserial**, mediante un estadístico F :

```
.xtserial logc logp logy
Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
F( 1,      45) =    112.351
Prob > F =      0.0000
```

Los resultados observados muestran que existe dependencia temporal y la misma debería considerarse en la especificación del modelo.

La incorporación de dinámica temporal genera un modelo básico, que en su forma vectorial puede escribirse, para un periodo t :

$$y_t = \tau y_{t-1} + X_t\beta + \mu + \varepsilon_t, \quad (5.5.1)$$

con $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$, con y_t e y_{t-1} de orden $n \times 1$ y efectos específicos $\mu' = [\mu_1, \mu_2, \dots, \mu_n]$.

El modelo (5.5.1) permite capturar la heterogeneidad entre las regiones, μ_i ($i = 1, 2, \dots, n$). Igual que en los modelos estáticos, los efectos específicos pueden ser tratados como efectos fijos o aleatorios. En el caso un modelo de efectos fijos, una variable ficticia es introducida por cada unidad espacial (sin considerar a la unidad base), mientras que bajo el modelo de efectos aleatorios, μ_i es considerada como una variable aleatoria *i.i.d.* con media cero y varianza σ_μ^2 .

La literatura de datos de panel ha discutido extensamente el modelo dinámico (5.5.1) en comparación a su versión estática, esto es cuando $\tau = 0$ (puede consultarse [Hsiao, 2003](#) o [Baltagi, 2008](#)). La estrategia de estimación de los efectos fijos proviene de eliminar el intercepto y las variables ficticias de la ecuación de regresión (procedimiento de desviaciones respecto a la media temporal). Los coeficientes de las pendientes, β (sin el intercepto), en esta ecuación en desviaciones pueden ser estimados por *MCO*, denominado estimador de mínimos cuadrados de variables ficticias o estimador intragrupos (*LSDV*). Sin embargo, cuando $\tau \neq 0$ se introduce una variable temporalmente rezagada al modelo, generando que el estimador *LSDV* sea inconsistente si T es fija, sin importar el tamaño de n . El problema con el estimador intragrupos es que genera una correlación de orden $(1/T)$, conocida como sesgo de Nickell, entre las variables y_{it-1} y el término de error, ambas expresadas en desvíos ([Nickell, 1981](#); [Hsiao, 2003](#)).

La literatura econométrica ha desarrollado un número de estimadores consistentes que utilizan el método de variables instrumentales ([Anderson & Hsiao, 1982](#)) y de momentos generalizados ([Arellano & Bond, 1991](#); [Arellano & Bover, 1995](#); [Blundell & Bond, 1998](#)). Otro procedimiento es el estimador *LSDV* corregido por sesgo, que obtiene el estimador intragrupo corregido, *CLSDV* ([Kiviet, 1995, 1999](#); [Bun & Kiviet, 2003](#)). [Elhorst \(2008\)](#) muestra que el estimador *CLSDV* de τ y β puede ser obtenido por medio de la log-máxima verosimilitud del modelo, condicionado por las observaciones iniciales, con respecto a τ , β y σ_ε^2 . Otro método alternativo aplica estimación por *MV* basado en la función de verosimilitud no condicionada del modelo ([Hsiao et al., 2002](#)).

Detectada la dependencia temporal, el modelo a estimar deberá contener elementos dinámicos. Bajo este tipo de paneles dinámicos deberá incorporarse la dependencia espacial ya detectada en la sección previa. Omitir la dependencia espacial en paneles dinámicos puede generar una importante pérdida de eficiencia tal que la comparación con un modelo *pooled*, que no distingue la estructura panel, puede ser mínima ([Phillips & Sul, 2003](#)).

Tal como sucedía en los modelos estáticos, los efectos espaciales pueden introducirse en el modelo (5.5.1) en diferentes formas. Una primera especificación puede hacerse mediante el denominado modelo dinámico de rezago espacial (*SLM*), al que se le incorpora un rezago temporal endógeno, denominado *SLM* simultáneo espacio-temporal ([Anselin et al., 2008](#)):

$$y_t = \tau y_{t-1} + \rho W y_t + X_t \beta + \mu + \varepsilon_t, \quad (5.5.2)$$

con $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_n)$. En este modelo, el parámetro ρ captura la dependencia espacial contemporánea de la variable endógena. Además, la inclusión de un rezago temporal endógeno, genera dinámica espacio temporal de corto y largo plazo.

[Anselin et al.](#) presenta otros modelos *SLM* espacio-temporales en donde incluye al modelo recursivo espacial puro:

$$y_t = \gamma W y_{t-1} + X_t \beta + \mu + \varepsilon_t, \quad (5.5.3)$$

donde γ es un parámetro autoregresivo espacio-temporal. Este modelo puede ser fácilmente extendido incluyendo X_{t-1} o $W X_t$ pero dado que $W y_{t-1}$ ya incluye al término $W X_{t-1}$ pueden generarse problemas de identificación. Este modelo puede ser muy adecuado para el análisis de fenómenos de difusión espacial.

Otro modelo sugerido es el *SLM* recursivo espacio-temporal:

$$y_t = \tau y_{t-1} + \gamma W y_{t-1} + X_t \beta + \mu + \varepsilon_t, \quad (5.5.4)$$

que incorpora dos parámetros autoregresivos temporales. Este modelo tiene aplicaciones particulares en cuestiones de predicciones espacio-temporales (por ejemplo, véase [Giacomini & Granger, 2004](#)).

El último modelo sugerido por [Anselin et al.](#) es una combinación de los anteriores denominado como modelo *SLM* dinámico espacio-temporal:

$$y_t = \tau y_{t-1} + \gamma W y_{t-1} + \rho W y_t + X_t \beta + \mu + \varepsilon_t, \quad (5.5.5)$$

donde se incorporan todas las alternativas de rezagos espacio-temporales de la variable dependiente.

Los modelos previamente mencionados son tipos de modelos espacio-temporales *SLM* ya que solo incorporan elementos dinámicos espacio-temporales endógenos. Esta clasificación puede ampliarse hacia modelos espacio-temporales de Durbin (*SDM*) como:

$$y_t = \tau y_{t-1} + \rho W y_t + X_t \beta + W X_t \theta + \mu + \varepsilon_t, \quad (5.5.6)$$

$$y_t = \gamma W y_{t-1} + X_t \beta + W X_t \theta + \mu + \varepsilon_t, \quad (5.5.7)$$

$$y_t = \tau y_{t-1} + \gamma W y_{t-1} + \rho W y_t + X_t \beta + W X_t \theta + \mu + \varepsilon_t, \quad (5.5.8)$$

en donde se han introducido efectos espaciales locales, $W X_t$, además de los efectos ya presentes en los modelos *dynSLM*.

Los efectos directos e indirectos desarrollados para paneles estáticos pueden extenderse hacia paneles dinámicos. La ventaja del modelo dinámico en tiempo es que pueden obtenerse efectos directos e indirectos de corto y largo plazo. Por ejemplo, bajo un modelo *dynSDM*, tipo dinámico espacio temporal (5.5.8), los efectos de corto plazo se obtienen ignorando τ y γ . La matriz de derivadas parciales de y con respecto a un cambio en una unidad de la variable explicativa k -ésima en las n regiones, para un punto temporal dado puede obtenerse como:

$$\left[\frac{\partial y}{\partial x_{1k}} \quad \cdots \quad \frac{\partial y}{\partial x_{nk}} \right]_t = (I_n - \rho W)^{-1} [\beta_k I_n + \theta_k W]. \quad (5.5.9)$$

Para obtener los efectos de largo plazo se puede suponer se alcanza un estado estacionario ($y_t = y_{t-1} = y^*$) tal que:

$$\left[\frac{\partial y}{\partial x_{1k}} \quad \cdots \quad \frac{\partial y}{\partial x_{nk}} \right]_t = [(1 - \tau) I_n - (\rho + \gamma) W]^{-1} [\beta_k I_n + \theta_k W]. \quad (5.5.10)$$

Estos resultados permiten determinar los efectos directos en el corto y largo plazo, y los efectos indirectos del corto y largo plazo (spatial spillovers). Desde estos modelos pueden obtenerse trayectorias secuenciales periodo a periodo como las presentadas [Debarsy et al. \(2012\)](#).

Respecto a la estimación de los paneles espaciales dinámicos (estacionarios), hay diferentes alternativas que extienden los procedimientos de máxima verosimilitud condicional y no-condicional. [Elhorst \(2010a\)](#) muestra diferentes procedimientos basados en *MV* que consideran las condiciones iniciales. [Yu et al. \(2008\)](#) desarrollan un estimador de cuasi-máxima verosimilitud

que muestra propiedades comparables a los estimadores *MV*. Este último estimador es el disponible en Stata bajo el comando desarrollado por [Belotti et al. \(2014\)](#).

La literatura ha avanzado en la estimación de modelos dinámicos más complejos. Una taxonomía de los mismos se presenta en [Anselin et al. \(2008\)](#). Sin embargo, en Stata, no hay comandos disponibles para modelos dinámicos con estructura espacial en el término de error. A continuación se muestran los modelos disponibles.

Modelos dinámicos de rezago espacial

La última actualización del comando `xsmle` permite la estimación de todos los tipos de modelos presentados previamente.

```
. quietly xsmle logc logp logy, dlag(1) fe wmat(Wst) type(ind) mod(sar) nolog effects nsim(999)
. estimate store SLM_simul
. quietly xsmle logc logp logy, dlag(2) fe wmat(Wst) type(ind) mod(sar) nolog effects nsim(999)
. estimate store SLM_recur
. quietly xsmle logc logp logy, dlag(3) fe wmat(Wst) type(ind) mod(sar) nolog effects nsim(999)
. estimate store SLM_dinam
```

Los resultados de estos modelos pueden presentarse conjuntamente para comparar particularmente los efectos de corto y largo plazo estiamdos.

```
. estimates table SLM_simul SLM_recur SLM_dinam, b(%7.3f) star(0.1 0.05 0.01) stats(ll aic bic) stf(%9.0f)
```

Variable		SLM_simul	SLM_recur	SLM_dinam
Main				
	logc			
	L1.	0.924***		0.931***
	logp	-0.049***	-0.482***	-0.086***
	logy	-0.032***	-0.008	-0.022***
	Wlogc			
	L1.		0.182***	-0.290***
Spatial				
	rho	0.096***	0.199***	0.296***
Variance				
	sigma2_e	0.002***	0.007***	0.002***

(continua en la siguiente página)

SR_Direct				
logp		-0.049***	-0.488***	-0.087***
logy		-0.031***	-0.006	-0.023***

SR_Indirect				
logp		-0.005***	-0.115***	-0.034***
logy		-0.003***	-0.001	-0.009***

SR_Total				
logp		-0.054***	-0.603***	-0.121***
logy		-0.034***	-0.007	-0.032***

LR_Direct				
logp		-2.016	-0.504***	-1.339**
logy		-1.175	-0.006	-0.351**

LR_Indirect				
logp		-31.506	-0.277***	-0.391
logy		-11.704	-0.003	-0.066

LR_Total				
logp		-33.522	-0.781***	-1.730
logy		-12.880	-0.009	-0.417

Statistics				
ll		2405	1459	2438
aic		-4783	-2894	-4848
bic		-4716	-2831	-4775

legend: * p<.1; ** p<.05; *** p<.01				

De los tres modelos estimados el que mejor ajusta es el modelo más completo que incluye todos los posibles parámetros espacio temporales dinámicos.

Modelos dinámicos espaciales de Durbin

En el caso de los modelos de Durbin, el comando `xsmle` también permite la estimación de los tres tipos de modelos. Sin embargo, para el caso analizado el algoritmo no logra convergencia evidenciando algunos problemas de identificación tal como [Anselin et al. \(2008\)](#) menciona. El único modelo que puede estimarse es el 5.5.7:

```
. xsmle logc logp logy, dlag(2) fe wmat(Wst) type(ind) mod(sdm) effects nsim(999)
```

```
Warning: All regressors will be spatially lagged
```

```
Iteration 0: Log-likelihood = 1567.5229
```

```
Iteration 1: Log-likelihood = 1576.8146
```

```
Iteration 2: Log-likelihood = 1576.8967
```

```
Iteration 3: Log-likelihood = 1576.8967
```

```
Computing marginal effects standard errors using MC simulation...
```

	logc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Main							
	Wlogc						
	L1.	.3506499	.0516162	6.79	0.000	.249484	.4518158
	logp	-.9015616	.0408432	-22.07	0.000	-.9816127	-.8215105
	logy	.5153353	.0602908	8.55	0.000	.3971674	.6335032
Wx							
	logp	.6489984	.0487541	13.31	0.000	.5534421	.7445546
	logy	-.5589645	.0613487	-9.11	0.000	-.6792058	-.4387233
Spatial							
	rho	.2739521	.0401772	6.82	0.000	.1952062	.352698
Variance							
	sigma2_e	.0055818	.0002115	26.39	0.000	.0051672	.0059963
SR_Direct							
	logp	-.8725182	.0372714	-23.41	0.000	-.9455687	-.7994676
	logy	.4891803	.054339	9.00	0.000	.3826778	.5956827
SR_Indirect							
	logp	.524279	.0529529	9.90	0.000	.4204933	.6280647
	logy	-.5490412	.055717	-9.85	0.000	-.6582446	-.4398378
SR_Total							
	logp	-.3482392	.0398476	-8.74	0.000	-.4263391	-.2701393
	logy	-.0598609	.0206368	-2.90	0.004	-.1003084	-.0194135
LR_Direct							
	logp	-.8829027	.0373096	-23.66	0.000	-.9560282	-.8097772
	logy	.4635183	.0527446	8.79	0.000	.3601407	.5668959
LR_Indirect							
	logp	.2041062	.0946115	2.16	0.031	.0186711	.3895413
	logy	-.5801251	.0620675	-9.35	0.000	-.7017752	-.458475
LR_Total							
	logp	-.6787965	.0982465	-6.91	0.000	-.8713561	-.486237
	logy	-.1166068	.0415472	-2.81	0.005	-.1980378	-.0351758

Como ya se ha mencionado, es importante destacar que los modelos aquí presentados no agotan las posibles especificaciones y esto principalmente notorio para el caso de paneles dinámicos espaciales.

Capítulo 6

Econometría espacial para redes sociales

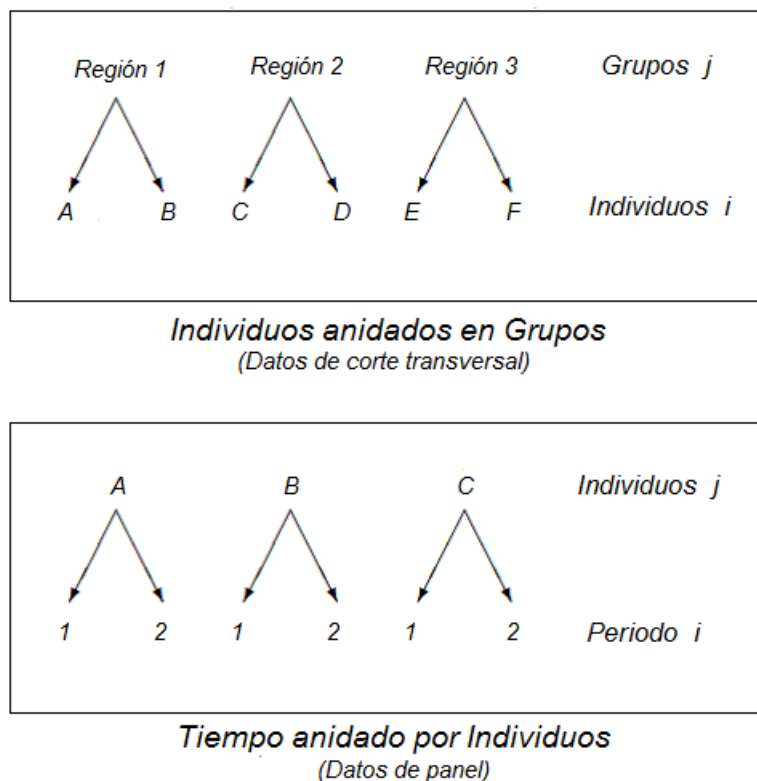
Capítulo 7

Modelos jerárquicos-espaciales

En una investigación crítica desarrollada por [Corrado & Fingleton \(2012\)](#), los autores mencionan a los modelos multinivel o jerárquicos como una alternativa viable para tratar datos georreferenciados. La metodología multinivel se enfoca en el análisis *simultáneo* del comportamiento de una variable en diferentes niveles de agregación. El caso más habitual es el representado por dos niveles, nivel micro y nivel macro, tal como se presenta en la Figura 7.1.

El primer caso de la Figura 7.1 es bajo un corte transversal, en donde los individuos pueden estar anidados en regiones. El nivel macro está representado por las regiones y, en cada una de estas regiones, existe un conjunto de individuos sobre los que se obtiene información estadística, conformando el nivel micro. El segundo caso hace referencia a un tipo de estructura jerárquica que involucra el tiempo, generalmente conocida como datos de panel: mediciones repetidas para un mismo individuo. En este caso, el nivel micro es representado por cada periodo temporal y el nivel macro por los individuos. Obsérvese que la notación destaca a j como nivel macro o de agrupamiento, siguiendo la estructura habitual del análisis multinivel. Es posible generalizar estas estructuras utilizando tres o más niveles, en donde los periodos temporales serán la medición micro, anidados en individuos y estos anidados en grupos representados por regiones. Asimismo, no es necesario que la estructura jerárquica sea balanceada de manera estricta pudiéndose trabajar con grupos de heterogéneos en su conformación.

Figura 7.1: Datos con Estructuras Jerárquicas



La estructura jerárquica genera que las unidades del nivel micro tiendan a asemejarse dentro de cada grupo, además tenderán a ser conjuntamente diferentes a unidades de otros grupos o clusters. Dicha dependencia intra-cluster viola uno de los supuestos habituales del modelo lineal general, unidades condicionalmente independientes (suponiendo un conjunto de variables explicativas).

Ante el problema mencionado, tradicionalmente, hay dos alternativas muy aplicadas. El primero de ellos es agregar la información micro al nivel superior. Es bien conocido que este tipo de análisis puede generar interpretaciones inválidas desde el nivel macro hacia el nivel micro, comúnmente llamada falacia ecológica (Robinson, 1950). Además, no es un método eficiente en el uso de la información perdiendo la variabilidad intra-grupo. El segundo método es analizar la variabilidad individual ignorando el contexto grupal, conocido como modelo combinado (*pooled model*). Este método tiene como consecuencia habitual que no permite distinguir efectos composicionales (debido a variables individuales) respecto a efectos contextuales (debido al efecto de pertenencia grupal).

Una solución para capturar la heterogeneidad a nivel macro puede ser la introducción, en el modelo combinado, de variables ficticias que identifican a los diferentes grupos, estrategia comúnmente denominada *de efectos fijos* en la literatura de datos de panel (Wooldridge, 2012). Esta estrategia puede ser válida para el caso de pocos agrupamientos, pero a medida que el número de grupos tiende a crecer, los grados de libertad se verán sensiblemente reducidos tornando dicha estrategia inviable (Goldstein, 1999).

El análisis multinivel utiliza modelos *de efectos aleatorios* estableciendo modelos más parsimoniosos, respecto a los modelos *de efectos fijos*, ya que solo requiere estimar unos pocos parámetros adicionales. Más importante aún, el modelo multinivel tiene la ventaja de permitir la distinción de procesos aleatorios a diferentes niveles de agregación. Dependiendo del ámbito en el que se desarrolle la investigación, este modelo ha recibido diferentes nombres como modelo jerárquico, de medidas repetidas, modelo de efectos mixtos, por citar los más recurrentes. Para una revisión de las diversas terminologías véase [De Leeuw & Meijer \(2008\)](#). A continuación presentamos los modelos relevantes para nuestra la aplicación empírica.

7.1 Modelo nulo

Como paso preliminar, utilizaremos un modelo sin variables explicativas o modelo nulo:

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}, \quad (7.1.1)$$

siendo y_{ij} la variable dependiente para el individuo i en el grupo j , donde β_{0j} es la media aleatoria grupal y ε_{ij} es el término de error que cumple con los supuestos: $\varepsilon_{ij} \sim i.i.d.\mathcal{N}(0, \sigma_\varepsilon^2)$. El intercepto es representado como una combinación lineal de un parámetro β_0 , que captura la media general, y una desviación aleatoria, δ_{0j} , que captura el efecto grupal específico. Es decir, para un j dado, la media aleatoria grupal se expresa como:

$$\beta_{0j} = \beta_0 + \delta_j.$$

El término δ_j corresponde al término de error aleatorio del nivel macro, y se supone que $\delta_j \sim i.i.d.\mathcal{N}(0, \sigma_\delta^2)$. De esta forma, se tienen dos componentes de error aleatorio, δ_j y ε_{ij} . Bajo independencia entre ambos errores, este modelo descompone la varianza total de la siguiente manera:

$$\text{var}(y_{ij}) = \text{var}(\beta_0 + \delta_{0j} + \varepsilon_{ij}) = \sigma_\delta^2 + \sigma_\varepsilon^2, \quad (7.1.2)$$

resultado equivalente a realizar un análisis de varianza (ANOVA) con efectos aleatorios. La significancia del componente σ_δ^2 es un indicativo de la importancia del efecto grupal.

La hipótesis nula de *no diferencia entre los grupos*, $H_0 : \sigma_\delta^2 = 0$, puede realizarse por medio del test $LR = 2(l_{H_1} - l_{H_0}) \sim \chi_{(k_1 - k_0)}^2$, donde l_{H_1} y l_{H_0} son las log-verosimilitudes del modelo bajo hipótesis alternativa y nula, respectivamente. Los grados de libertad de la distribución Chi-cuadrada, $k_1 - k_0$, son igual a la unidad en este caso.

Usualmente, es de interés realizar comparaciones entre grupos, para ello se necesitarán las estimaciones de δ_j . Sin embargo, los efectos δ_j son tratados como variables aleatorias y no como parámetros a estimar del modelo. Por lo que, de forma análoga a la obtención de los residuos en un modelo de regresión simple, puede aplicarse un mecanismo que brindará la realización de δ_j para grupo. Para ello, se utilizarán las estimaciones de los parámetros β_0 , σ_δ^2 y σ_ε^2 .

Los residuos totales del modelo pueden estimarse como: $e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\beta}_0$. Para obtener los residuos grupales se estima la media residual por grupo $\bar{e}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} e_{ij}$, donde n_j es la cantidad de observaciones en el grupo j . Luego, la predicción del residuo grupal j es:

$$\hat{\delta}_j = w_j \times \bar{e}_j = \frac{\hat{\sigma}_\delta^2}{\hat{\sigma}_\delta^2 + \frac{\hat{\sigma}_\varepsilon^2}{n_j}} \times \bar{e}_j. \quad (7.1.3)$$

El factor w recibe la denominación de contracción residual (shrinkage factor). Este factor es siempre menor a 1, por lo que los residuos grupales son menores o iguales a los residuos medios. Existirá una mayor contracción de los residuos si tenemos una varianza grupal pequeña (σ_δ^2), o una gran varianza micro (σ_ε^2) o un tamaño grupal pequeño (n_j). Para completar la estimación residual, pueden obtenerse los residuos del nivel 1, $\hat{\varepsilon}_{ij}$, mediante una simple operación: $\hat{\varepsilon}_{ij} = e_{ij} - \hat{\delta}_j$.

7.2 Modelo condicional de intercepto aleatorio

La variabilidad entre grupos del modelo nulo, σ_δ^2 , puede ser explicada por efectos composicionales, es decir, por las características de los individuos que componen a cada grupo. Esto lleva a la necesidad de plantear un modelo condicional de la forma:

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \cdots + \beta_k x_{kij} + \varepsilon_{ij}, \quad (7.2.1)$$

donde x_{1ij} hasta x_{kij} son variables explicativas a nivel individual. El término β_{0j} está compuesto $\beta_0 + \delta_j$, tal que habrá un intercepto por cada grupo, condicionado al conjunto de variables explicativas. Este modelo es conocido como *de intercepto aleatorio* y es el ejemplo más simple de los llamados modelos de efectos mixtos donde hay efectos fijos, parte sistemática, y efectos aleatorios.

La introducción de variables individuales puede explicar completamente la variabilidad entre grupos. Para contrastar esta hipótesis puede utilizarse el test LR , planteado en la sección 2.1, especificando bajo hipótesis nula al modelo de regresión lineal general y, bajo hipótesis alternativa, al modelo de intercepto aleatorio.

Asimismo, es posible introducir otro tipo de variables, las denominadas contextuales. Estas variables pueden reflejar características del nivel macro, obtenidas directamente de las unidades superiores, Z_j ; o pueden provenir de agregaciones del nivel micro como, por ejemplo, el promedio de individuos con NBI, $Z_{1j} = \bar{x}_{1j} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{1ij}$. La nueva especificación es:

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \cdots + \beta_k x_{kij} + \theta_1 Z_{1j} + \cdots + \theta_g Z_{gj} + \varepsilon_{ij}. \quad (7.2.2)$$

Ambos tipos de variables contextuales pueden incorporarse sin problemas al modelo de intercepto aleatorio, teniendo especificado un error aleatorio para cada nivel de agregación: ε_{ij} para las variables individuales y δ_j para las variables contextuales. Además, si es de interés, existe la posibilidad de incorporar términos de interacción entre variables de diferentes niveles. De esta forma, este modelo multinivel permite distinguir entre efectos composicionales, efectos contextuales y efectos de interacciones entre niveles.

La estimación de este tipo de modelo puede realizarse por diferentes métodos, uno de los más habituales es por máxima verosimilitud. La precisión de los coeficientes estimados dependerá en buena medida del tamaño muestral disponible. Una distinción de estos modelos es que hay dos tamaños muestrales, la cantidad de grupos y la cantidad de individuos por grupo. Diferentes investigaciones han propuesto que idealmente se necesita un mínimo de 30 unidades por cada nivel

y, si el modelo es complejo en su especificación, ampliar esta condición a un mínimo de 50 grupos con 20 unidades por grupo (Hox, 1998). Claramente, este problema es una cuestión de grado y variará dependiendo del campo de estudio, aunque en general un gran número de grupos parece ser más importante que una gran cantidad de individuos por grupo (Maas & Hox, 2005).

7.3 Modelo condicional de intercepto y pendiente aleatoria

La principal limitación del modelo de intercepto aleatorio radica en que el efecto de las variable explicativas individuales es constante entre los grupos. Para evitar esta restricción, puede modificarse el modelo tal que permita efectos aleatorios en el intercepto y la pendiente. Es decir, puede especificarse un modelo como:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \sum_{p=2}^k \beta_p x_{pij} + \sum_{l=1}^g \theta_l Z_{lij} + \varepsilon_{ij} \quad (7.3.1)$$

donde $\beta_{1j} = \beta_1 + \eta_j$, tal que $\eta_j \sim i.i.d.\mathcal{N}(0, \sigma_\eta^2)$. La pendiente de x_{1ij} ahora es aleatoria y puede ser descompuesta en un efecto global, común a todos los grupos, β_1 , y en un efecto aleatorio, η_j , que captura la desviación del grupo j respecto al efecto global. Por simplicidad se ha considerado un único efecto de pendiente aleatoria, en x_{1ij} , pudiéndose generalizar a todas las demás variables explicativas individuales.

Los efectos aleatorios, intercepto y pendiente, pueden estar relacionados tal que el modelo puede ampliarse para incluir un parámetro adicional. Este parámetro, $\sigma_{\eta\delta}$, es la covarianza que capturará la variación del intercepto y la pendiente entre-grupos. Un valor positivo de $\sigma_{\eta\delta}$ implica que grupos con altos valores del intercepto, δ_j , tienden a tener altos valores de la pendiente, η_j .

Los modelos de intercepto aleatorio, ya sea el nulo o el condicional, suponen que la varianza entre grupos es constante e igual a σ_δ^2 . Bajo un modelo de intercepto y pendiente aleatoria, la varianza entre grupos es función de las variables explicativas que acompañan al intercepto aleatorio. Por ejemplo, para el modelo presentado en la ecuación (6) se obtiene:

$$\text{var}(\delta_j + \eta_j x_{1ij}) = \sigma_\delta^2 + 2\sigma_{\eta\delta}x_{1ij} + \sigma_\eta^2 x_{1ij}^2, \quad (7.3.2)$$

es decir, la varianza entre grupos es una función cuadrática que depende de los valores de la variable a nivel micro. En el caso de que x_{1ij} sea dicotómica, la varianza solo tomará dos valores:

$$\begin{aligned} \text{var}(\delta_j + \eta_j x_{1ij}) &= \sigma_\delta^2, & x_{1ij} = 0, \\ \text{var}(\delta_j + \eta_j x_{1ij}) &= \sigma_\delta^2 + 2\sigma_{\eta\delta} + \sigma_\eta^2, & x_{1ij} = 1. \end{aligned}$$

Un fenómeno habitual en modelos de corte transversal es la presencia de heteroscedasticidad, esto es, que la varianza residual es diferentes para diferentes unidades. Más específicamente, usualmente significa que la varianza residual depende funcionalmente de las variables explicativas. Por lo tanto, podría esperarse que este fenómeno se encuentre presente en muchos modelos multinivel. De hecho, la heteroscedasticidad entre grupos es una parte explícita en los modelos multinivel tal como se ha expresado en la ecuación (7.3.2).

Capítulo 8

Comentarios finales

En los últimos años el avance de los paquetes informáticos ha posibilitado que las técnicas de econometría espacial se encuentren disponibles para los economistas aplicados.

Esta guía ha buscado utilizar el programa Stata íntegramente para realizar el análisis econométrico espacial tanto para datos de corte transversal como para panel de datos.

Los capítulos desarrollados han incluido el análisis espacial exploratorio como un primer paso para aproximarse a la modelización espacial. Este tipo de análisis brinda información importante para desarrollar posibles hipótesis y detectar datos anómalos. Otro capítulo incorpora la estimación de modelos espaciales bajo un corte transversal. La revisión muestra las principales especificaciones así como los métodos usuales de estimación. Finalmente, se presenta un último capítulo con los avances en modelos de datos de panel que incorporan efectos espaciales. Claramente, esta última área es de intensa investigación y los desarrollos son recientes.

Por otra parte, el documento no compara los resultados obtenidos por Stata respecto a otros paquetes estadísticos. Otras investigaciones han realizado dicha comparación respecto a Matlab, R y Pysal ([Bivand & Piras, 2013](#)). Algunos investigadores solo se han enfocado en un modelo en particular, *SEM*, explorando las alternativas de estimación de Stata respecto a GeoDaSpace, Pysal y R ([Anselin et al., 2012](#)). Consúltense estas referencias para conocer las diferencias entre los paquetes disponibles.

Bibliografía

- Anderson, T. W. & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of econometrics*, 18(1), 47–82.
- Anselin, L. (1988). *Spatial econometrics: Methods and models*, volume 4. Springer.
- Anselin, L. (1992). Spacestat tutorial: A workbook for using SpaceStat in the analysis of spatial data. *Urbana-Champaign: University of Illinois*, 263.
- Anselin, L. (1995). Local indicators of spatial association lisa. *Geographical analysis*, 27(2), 93–115.
- Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In P. Longley, M. Goodchild, D. Maguire, & D. Rhind (Eds.), *Geographical Information Systems: Principles, Techniques, Management and Applications*. Cambridge: Geoinformation Int.
- Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in regional science*, 89(1), 3–25.
- Anselin, L., Amaral, P. V., & Arribas-Bel, D. (2012). Technical aspects of implementing GMM estimation of the spatial error model in PySAL and GeoDaSpace. *GeoDa Center for Geospatial Analysis and Computation*.
- Anselin, L. & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs*, 155, 237–290.
- Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional science and urban economics*, 26(1), 77–104.
- Anselin, L. & Florax, R. J., Eds. (1995). *New directions in spatial econometrics*. Springer, New York.
- Anselin, L., Le Gallo, J., & Jayet, H. (2008). Spatial panel econometrics. In L. Mátyás & P. Sevestre (Eds.), *The econometrics of panel data: fundamentals and recent developments in theory and practice*, volume 46 (pp. 625–660). Springer.
- Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical analysis*, 38(1), 5–22.
- Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2), 277–297.
- Arellano, M. & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1), 29–51.

- Arraiz, I., Drukker, D. M., Kelejian, H. H., & Prucha, I. R. (2010). A spatial cliff-ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science*, 50(2), 592–614.
- Baltagi, B. (2008). *Econometric analysis of panel data*. John Wiley & Sons.
- Baltagi, B. H. & Li, D. (2004). Prediction in the panel data model with spatial correlation. In *Advances in spatial econometrics* (pp. 283–295). Springer.
- Bartels, C. & Ketellapper, R. (1979). *Exploratory and explanatory statistical analysis of spatial data*. Martinus Nijhoff, Boston, MA.
- Belotti, F., Hughes, G., & Mortari, A. P. (2014). Xsmle: Stata module for spatial panel data models estimation. *Statistical Software Components*.
- Bennett, R. (1979). *Spatial time series*. Pion, London.
- Bivand, R. (2002). Spatial econometrics functions in R: Classes and methods. *Journal of geographical systems*, 4(4), 405–421.
- Bivand, R. & Piras, G. (2013). *Comparing Implementations of Estimation Methods for Spatial Econometrics*. Technical report, Regional Research Institute, West Virginia University.
- Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1), 115–143.
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281–298.
- Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 431–443.
- Bun, M. J. & Kiviet, J. F. (2003). On the diminishing returns of higher-order terms in asymptotic expansions of bias. *economics Letters*, 79(2), 145–152.
- Burridge, P. (1980). On the cliff-ord test for spatial correlation. *Journal of the Royal Statistical Society. Series B (Methodological)*, (pp. 107–108).
- Calvo, E. & Escobar, M. (2003). The local voter: A geographically weighted approach to ecological inference. *American Journal of Political Science*, 47(1), 189–204.
- Cliff, A. & Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4(3), 267–284.
- Corrado, L. & Fingleton, B. (2012). Where is the economics in spatial econometrics? *Journal of Regional Science*, 52(2), 210–239.
- Crow, K. (2013). Shp2dta: Stata module to converts shape boundary files to stata datasets. *Statistical Software Components*.
- Davidson, R. & MacKinnon, J. G. (2004). *Econometric theory and methods*, volume 5. Oxford University Press New York.

- De Hoyos, R. E. & Sarafidis, V. (2006). Testing for cross-sectional dependence in panel-data models. *Stata Journal*, 6(4), 482.
- De Leeuw, J. & Meijer, E. (2008). Introduction to multilevel analysis. In *Handbook of multilevel analysis* (pp. 1–75). Springer.
- Debarsy, N., Ertur, C., & LeSage, J. P. (2012). Interpreting dynamic space–time panel data models. *Statistical Methodology*, 9(1), 158–171.
- Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *Stata Journal*, 3(2), 168–177.
- Drukker, D. M., Egger, P., & Prucha, I. R. (2013a). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews*, 32(5-6), 686–733.
- Drukker, D. M., Peng, H., Prucha, I. R., & Raciborski, R. (2013b). Creating and managing spatial-weighting matrices with the `spmat` command. *Stata Journal*, 13(2), 242–286.
- Drukker, D. M., Prucha, I. R., & Raciborski, R. (2013c). A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. *The Stata Journal*, 13(2), 287–301.
- Drukker, D. M., Prucha, I. R., & Raciborski, R. (2013d). Maximum-likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *Stata Journal*, 13(2), 221–241.
- Elhorst, J. (2008). Estimation of dynamic panels with endogenous interaction effects. In *II World Conference of the Spatial Econometrics Association*.
- Elhorst, J. P. (2010a). Dynamic panels with endogenous interaction effects when t is small. *Regional Science and Urban Economics*, 40(5), 272–282.
- Elhorst, J. P. (2010b). Spatial panel data models. In M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications* chapter 2, (pp. 377 – 407). Springer-Verlag Berlin Heidelberg.
- Elhorst, J. P. (2014). *Spatial Econometrics. From Cross-sectional data to Spatial Panels*. SpringerBriefs in Regional Science, Springer.
- Ertur, C. & Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of applied econometrics*, 22(6), 1033–1062.
- Farber, S., Páez, A., & Volz, E. (2009). Topology and dependency tests in spatial and network autoregressive models. *Geographical Analysis*, 41(2), 158–180.
- Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.

- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (2001). Spatial variations in school performance: a local analysis using geographically weighted regression. *Geographical and Environmental Modelling*, 5(1), 43–66.
- Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The incorporated statistician*, (pp. 115–146).
- Getis, A. & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical analysis*, 24(3), 189–206.
- Giacomini, R. & Granger, C. W. (2004). Aggregation of space-time processes. *Journal of econometrics*, 118(1), 7–26.
- Gibbons, S. & Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2), 172–191.
- Goldstein, H. (1999). *Multilevel Statistical Models*. Institute of Education, Multilevel Models Project.
- Goodchild, M., Haining, R., & Wise, S. (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographical Information Systems*, 6(5), 407–423.
- Graubard, B. I. & Korn, E. L. (2002). Inference for superpopulation parameters using sample surveys. *Statistical Science*, (pp. 73–96).
- Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.
- Griffith, D. A. (1996). Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. *Practical handbook of spatial statistics*, (pp. 65–82).
- Halleck Vega, S. & Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, 55(3), 339–363.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, (pp. 1251–1271).
- Herrera, M., Mur, J., & Ruiz, M. (2014). Detecting causal relationships between spatial processes. *Papers in Regional Science*.
- Hordijk, L. (1979). Problems in estimating econometric relations in space. *Papers in Regional Science*, 42(1), 99–115.
- Hox, J. (1998). Multilevel modeling: When and why. In *Classification, data analysis, and data highways* (pp. 147–154). Springer.
- Hsiao, C. (2003). *Analysis of panel data*, volume 34. Cambridge university press.
- Hsiao, C., Hashem Pesaran, M., & Kamil Tahmiscioglu, A. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of econometrics*, 109(1), 107–150.

- Jeanty, P. W. (2012). Splagvar: Stata module to generate spatially lagged variables, construct the moran scatter plot, and calculate moran's i statistics. *Statistical Software Components*.
- Jeanty, P. W. (2014). Spwmatrix: Stata module to generate, import, and export spatial weights. *Statistical Software Components*.
- Kelejian, H. H. & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99–121.
- Kelejian, H. H. & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2), 509–533.
- Kelejian, H. H. & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1), 53–67.
- Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics*, 68(1), 53–78.
- Kiviet, J. F. (1999). Expectation of expansions for estimators in a dynamic panel data model: Some results for weakly exogenous regressors. In C. Hsiao, K. Lahiri, L.-F. Lee, & M. Pesaran (Eds.), *Analysis of Panel Data and Limited Dependent Variables*. Cambridge university press.
- Kondo, K. (2016). Hot and cold spot analysis using stata. *The Stata journal*, 16(3), 613–631.
- Lee, L.-f. & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2), 165–185.
- LeSage, J. & Pace, R. (2009). *Introduction to spatial econometrics*. CRC press.
- LeSage, J. P. (1999). *Spatial econometrics*. Regional Research Institute, West Virginia University.
- Maas, C. J. & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 1(3), 86.
- Mankiw, G., Weil, D., & Romer, D. (1992). A contribution to the empirics of growth theory. *The Quarterly Journal of Economics*, 107, 407–37.
- Millo, G. & Piras, G. (2012). splm: Spatial panel data models in r. *Journal of Statistical Software*, 47(1), 1–38.
- Mizruchi, M. S. & Neuman, E. J. (2008). The effect of density on the level of bias in the network autocorrelation model. *Social Networks*, 30(3), 190–200.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, (pp. 17–23).
- Mur, J. & Angulo, A. (2009). Model selection strategies in a spatial setting: Some additional results. *Regional Science and Urban Economics*, 39(2), 200–213.
- Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, (pp. 1417–1426).

- Ord, J. (1981). Towards a theory of spatial statistics: a comment. *Geographical Analysis*, 13(1), 86–91.
- Ord, J. K. & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, 27(4), 286–306.
- Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349), 120–126.
- Paelinck, J. & Klaassen, L. (1979). *Spatial econometrics*. Saxon House, Farnborough.
- Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. *Cambridge Working Papers in Economics No. 0435*. University of Cambridge.
- Phillips, P. C. & Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence*. *The Econometrics Journal*, 6(1), 217–259.
- Pisati, M. (2001). Sg162: Tools for spatial data analysis. *Stata Technical Bulletin*, 60, 21–37.
- Pisati, M. (2008). Spmap: Stata module to visualize spatial data. *Statistical Software Components*.
- Pisati, M. (2012). Exploratory spatial data analysis using stata. In *Presentation at German Stata User Group meeting, Berlin*.
- Rey, S. J. & Anselin, L. (2010). Pysal: A Python library of spatial analytical methods. In F. M. & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 175–193). Springer.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351–357.
- Slocum, T., McMaster, R., Kessler, F., & Howard, H. (2009). *Thematic cartography and geovisualisation*. Prentice Hall, New Jersey, NJ.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, (pp. 234–240).
- Wooldridge, J. (2012). *Introductory econometrics: A modern approach*. Cengage Learning.
- Wooldridge, J. M. (2009). *Introductory Econometrics: A Modern Approach*. Southwestern, 4th edition edition.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, ma: mit Press.
- Yu, J., de Jong, R., & Lee, L.-f. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics*, 146(1), 118–134.