

Capítulo 1

Introducción

La sección 1.1 de este capítulo introduce el concepto de *dependencia espacial* que suele surgir en las muestras de datos espaciales transversales. Las muestras de datos espaciales representan observaciones asociadas a puntos o regiones, por ejemplo, hogares, condados, estados o secciones censales. Se ofrecen dos ejemplos de motivación para la dependencia espacial, uno basado en los efectos indirectos espaciales derivados de los efectos de la congestión y un segundo que se basa en las variables explicativas omitidas. En la sección 1.2 se exponen los procesos de generación de datos *autorregrésivos espaciales* para datos de muestra dependientes del espacio, junto con las *matrices de pesos espaciales* que desempeñan un papel importante en la descripción de la estructura de estos procesos. En el [capítulo 2](#) se analizan con más detalle los procesos generadores de datos espaciales y los modelos econométricos espaciales asociados, y en el [capítulo 4](#) las *matrices de pesos espaciales*. Nuestro objetivo aquí es ofrecer una introducción a los procesos autorregresivos espaciales y a los modelos de regresión espacial que se basan en este tipo de procesos. En la sección 1.3 se presenta un ejemplo sencillo de cómo los efectos de la congestión provocan desbordamientos espaciales que afectan a las regiones vecinas utilizando los tiempos de viaje a la región del distrito central de negocios (CBD) de un área metropolitana. En la sección 1.4 se describen varios escenarios en los que se pueden utilizar modelos econométricos espaciales para analizar los efectos de desbordamiento espacial. La sección final del capítulo expone el plan de este texto. Se ofrece una breve enumeración de los temas tratados en cada capítulo.

1.1 Dependencia espacial

Consideremos un vector de variables transversales que representa observaciones coloreadas con referencia a puntos o regiones en el espacio. Las observaciones puntuales podrían incluir los precios de venta de las viviendas, el empleo en varios establecimientos o la matriculación en escuelas individuales. Los sistemas de información geográfica suelen ofrecer *geocodificación* o *cotejo de direcciones*, lo que permite convertir automáticamente las direcciones en coordenadas de localización. La capacidad

de geocodificación ha dado lugar a grandes cantidades de datos con referencias espaciales. Las observaciones pueden incluir una variable como la población o el tiempo medio de desplazamiento de los residentes en regiones como las secciones censales, los condados o las áreas estadísticas metropolitanas (MSA). A diferencia de las observaciones puntuales, para una región nos basamos en las coordenadas de un punto interior que representa el centro (el *centroide*). Un punto importante es que en

En los modelos de regresión espacial, cada observación corresponde a un lugar o región. El *proceso de generación de datos* (PGD) para una muestra transversal no espacial convencional de n observaciones *independientes* $y_i, i = 1, \dots, n$ que se relacionan linealmente con las variables explicativas de una matriz X adopta la forma de (1.1), donde hemos suprimido el término de intercepción. n que se relacionan linealmente con las variables explicativas de una matriz X tiene la forma de (1.1), donde hemos suprimido el término de intercepción, que podría incluirse en la matriz X .

$$y_i = X\beta_i + \varepsilon_i \quad (1.1)$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1, \dots, n \quad (1.2)$$

En (1.2), utilizamos $N(a, b)$ para denotar una distribución normal univariante con media a y varianza b . En (1.1), X_i representa un vector $1 \times k$ de covariables o variables explicativas, con parámetros asociados β contenidos en un vector $1 \times k$. Este tipo de proceso de generación de datos se asume normalmente para los modelos de regresión lineal. Cada observación tiene una media subyacente de $X\beta_i$ y un componente aleatorio ε_i . Una implicación de esto para situaciones en las que las observaciones i representan regiones o puntos en el espacio es que los valores observados en un lugar (o región) son independientes de las observaciones realizadas en otros lugares (o regiones). Las observaciones independientes o *estadísticamente independientes* implican que $E(\varepsilon_i \varepsilon_j) = E(\varepsilon_i)E(\varepsilon_j) = 0$. La suposición de independencia simplifica enormemente los modelos, pero en contextos espaciales esta simplificación parece forzada.

Por el contrario, la *dependencia espacial* se refiere a una situación en la que los valores observados en un lugar o región, digamos la observación i , dependen de los valores de las observaciones *vecinas* en lugares cercanos. Supongamos que las observaciones $i = 1$ y $j = 2$ representan a los vecinos (tal vez regiones con fronteras que se tocan), entonces un proceso de generación de datos podría tener la forma mostrada en (1.3).

$$y_i = \alpha y_j + X\beta_i + \varepsilon_i \quad (1.3)$$

$$y_j = \alpha y_i + X\beta_j + \varepsilon_j$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1$$

$$\varepsilon_j \sim N(0, \sigma^2) \quad j = 2$$

Esta situación sugiere un proceso simultáneo de generación de datos, en el que el valor que toma y_i depende del de y_j y viceversa. Como ejemplo concreto, consideremos el conjunto de siete regiones mostradas en la [figura 1.1](#), que representan tres regiones al oeste y tres al este de un distrito comercial central (CBD). A efectos de este ejemplo, consideraremos que estas siete regiones constituyen una única área metropolitana, siendo la región R_4 el distrito comercial central. Dado que toda la región contiene una sola

carretera, todos los viajeros

compartir esta ruta hacia y desde el CBD.

Podríamos observar el siguiente conjunto de datos muestrales para estas regiones que relacionan los tiempos de viaje con el CBD (en minutos) contenidos en la variante dependiente

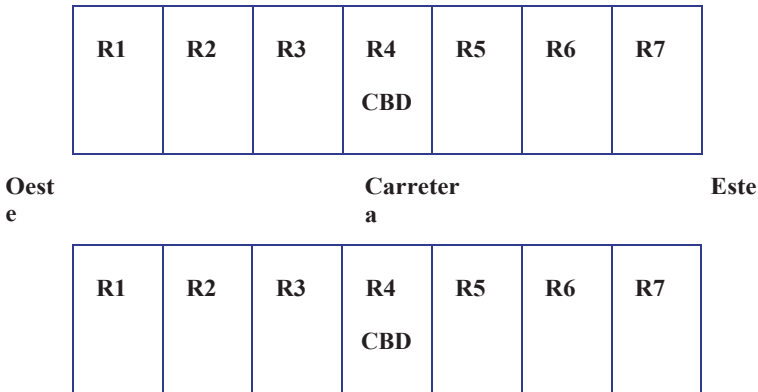
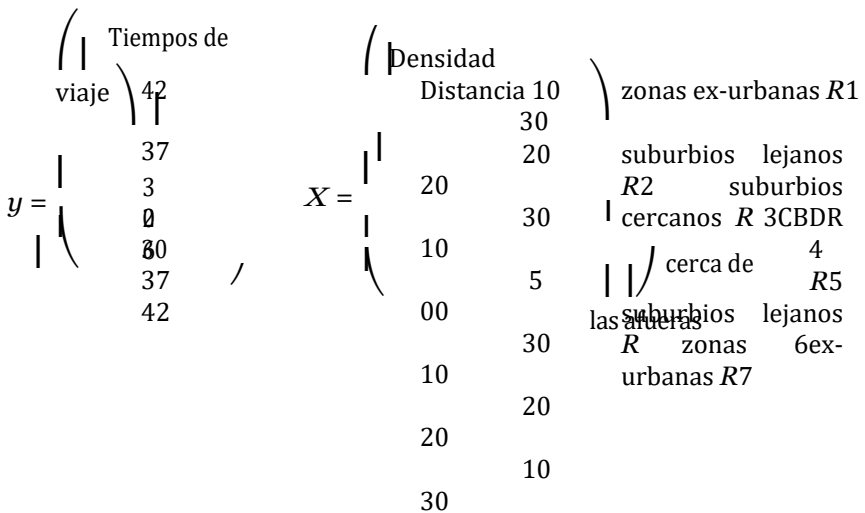


FIGURA1.1 : Regiones al este y al oeste del Distrito Comercial Central

y a la distancia (en millas) y a la densidad de población (población por manzana) de las regiones en las dos columnas de la matriz X .



El patrón de tiempos de viaje más largos para las regiones más distantes R_1 y R_7 frente a las regiones más cercanas R_3 y R_5 encontrado en el vector y parece violar claramente la independencia, ya que los tiempos de viaje parecen similares para las regiones vecinas. Sin embargo, podríamos suponer que este patrón se explica por las variables del modelo *Distancia* y *Densidad* asociadas a cada región, ya que éstas también parecen similares para las

regiones vecinas. Incluso para los individuos que se encuentran en el CBD, se necesita tiempo para ir a otro lugar del CBD. Por lo tanto, el tiempo de viaje dentro del CBD es de minutos a 26 a pesar de tener una distancia de kilómetros⁰.

Ahora bien, consideremos que nuestro conjunto de tiempos de viaje observados representa las mediciones realizadas en un día concreto, por lo que tenemos los tiempos de viaje al CBD promediados a lo largo de un periodo de una hora²⁴. En este caso, parte del patrón observado podría explicarse

por los efectos de congestión que surgen de la autopista compartida. Parece plausible que los tiempos de viaje más largos en una región conduzcan a tiempos de viaje más largos en las regiones vecinas en un día determinado. Esto se debe a que los viajeros pasan de una región a otra cuando viajan por la autopista hacia el CBD. Los tiempos más lentos en R_3 en un día concreto deberían producir tiempos más lentos para ese día en las regiones R_2 y R_1 . Los efectos de congestión representan un tipo de desbordamiento espacial, que no se produce simultáneamente, sino que requiere cierto tiempo para que surja el retraso de tráfico. Desde el punto de vista de la modelización, los efectos de congestión como éstos no se explican por las variables del modelo *Distancia y Densidad*. Se trata de efectos dinámicos de retroalimentación del tiempo de viaje en un día concreto que afectan a los tiempos de viaje de las regiones vecinas en el breve intervalo de tiempo necesario para que se produzca el retraso de tráfico. Como la variable explicativa distancia no cambiaría de un día a otro, y la densidad de población cambiaría muy lentamente en una escala temporal diaria, estas variables no serían capaces de explicar los fenómenos de retraso diario. La variación diaria observada en los tiempos de viaje se explicaría mejor si se basara en los tiempos de viaje de las regiones vecinas en ese día. Esta es la situación representada en (1.3), donde nos basamos en el tiempo de viaje de una observación vecina y_j como variable explicativa del tiempo de viaje en la región i , y_i . De forma similar, utilizamos y_j para explicar el tiempo de viaje de la región j , y_j .

Dado que nuestras observaciones se midieron utilizando los tiempos medios de un día, la escala de tiempo de medición no es lo suficientemente fina como para captar el aspecto dinámico del tiempo de intervalo corto de la demora de tráfico. Esto daría lugar a tiempos de viaje diarios observados en el vector y que parecen estar determinados simultáneamente. Este es un ejemplo de por qué la dependencia espacial medida puede variar con la escala temporal de recogida de datos.

Otro ejemplo en el que la dependencia espacial observada puede surgir de variables omitidas sería el caso de un modelo de precios hedónicos con los precios de venta de las viviendas como el vector y y las características de las viviendas como variables explicativas en la matriz X . Si tenemos una muestra transversal de los precios de venta en un vecindario recogida durante un período de un año, la variación en las características de las viviendas debería explicar parte de la variación en los precios de venta observados. Consideremos una situación en la que una sola vivienda se vende por un precio mucho mayor del que cabría esperar basándose únicamente en sus características. Supongamos que esta venta tuvo lugar a mediados de nuestro período de 12 observación de un mes, poco después de que se publicara un informe positivo sobre la calidad de los colegios cercanos. Dado que la calidad de la escuela no era una variable incluida en el conjunto de variables explicativas que representan las características de la vivienda, el precio de venta más alto de lo esperado podría reflejar una nueva prima por la calidad de la escuela. Esto podría indicar a otros vendedores de viviendas atendidas por la misma escuela que pidan precios más altos, o que acepten ofertas mucho más cercanas a sus precios de venta durante los últimos seis meses de nuestro período de observación. Esto llevaría a una situación en la que el uso

de los precios de venta de las viviendas vecinas produjera un mayor poder explicativo de las viviendas atendidas por la escuela de alta calidad durante los últimos seis meses de nuestra muestra. Otras variables omitidas podrían ser la accesibilidad al transporte, los servicios cercanos como tiendas o parques, etc. Si éstas se omitieran en

el conjunto de variables explicativas consistente únicamente en las características de las viviendas, encontraríamos que los precios de venta de las viviendas vecinas son útiles para la predicción. Un ejemplo de que los modelos de regresión no espaciales ignorarán la dependencia espacial en la variable dependiente lo proporciona un mapa de los residuos de mínimos cuadrados ordinarios de una regresión de la función de producción: $\ln(Q) = \alpha_0 + \beta_1 \ln(K) + \gamma \ln(L) + \varepsilon$, estimada utilizando los estados contiguos⁴⁸ de EE.UU. más el Distrito de Columbia. El producto estatal bruto del año 2001¹ utilizó como Q , siendo la mano de obra L el empleo total 2001¹ no agrícola de cada estado. Las estimaciones de capital K para los estados proceden de Garofalo y Yamarik (2002). Estos ~~residuos~~ suelen denominarse *residuos de Solow* si se imponen rendimientos constantes a escala, de modo que $\beta = \varphi$, $\gamma = (1 - \varphi)$. En el contexto de un modelo de crecimiento de Solow, se interpretan como reflejos del crecimiento económico por encima de la tasa de crecimiento del capital, o de la no explicada por el crecimiento de los factores de producción. En el caso de nuestro modelo de función de producción, se interpretarían como la productividad total de los factores, por lo que reflejan la producción atribuible a la variación regional del eficiencia tecnológica con la que se utilizan estos factores.

La figura 1.2 muestra un *mapa coroplético* de la productividad total de los factores (los residuos de nuestra regresión de la función de producción). Un mapa coroplético se basa en áreas sombreadas o con patrones para reflejar los valores medidos de la variable que se muestra en el mapa. Proporciona una representación visual de cómo los valores de una variable varían en el espacio. La figura 1.3 muestra una leyenda asociada al mapa en forma de histograma que muestra la distribución de la frecuencia de los estados según la magnitud de sus residuos. Vemos residuos negativos para los estados¹², incluyendo el grupo de estados vecinos⁷, Texas, Oklahoma, Luisiana, Mississippi, Tennessee Arkansas y Alabama. Un residuo negativo indicaría que la producción observada Q fue inferior a la predicha por la regresión basada en la mano de obra y el capital disponible en estos estados. En la leyenda de la Figura 1.3 vemos que los estados azules, verdes y morados representan residuos positivos. De los estados verdes¹¹ vemos un grupo de estos estados en el noreste, lo que indica que la producción observada para estos estados fue superior a la predicha por nuestro modelo de regresión, reflejando una productividad total de los factores superior a la esperada.

Si los residuos estuvieran distribuidos aleatoriamente con respecto a la ubicación, no veríamos grupos de estados rojos y verdes que son indicativos de residuos negativos y positivos asociados a estados vecinos. Este tipo de agrupación representa una representación visual de la dependencia espacial en los residuos o la productividad de los factores del modelo de regresión no espacial.

Se plantea una pregunta: ¿qué lleva a la dependencia espacial observada en la productividad total de los factores? Los métodos de modelización econométrica espacial tienen un papel que desempeñar para responder a esta pregunta. Como veremos, diferentes especificaciones de modelos sugieren justificaciones teóricas distintas, y viceversa. En la econometría tradicional hay tres usos de los modelos empíricos: 1) estimación e in-ferencia respecto a los

parámetros, 2) predicción o pronóstico fuera de muestra y
3) comparación de modelos de especificaciones alternativas.

Podemos utilizar los modelos econométricos espaciales de las mismas tres maneras para responder a la

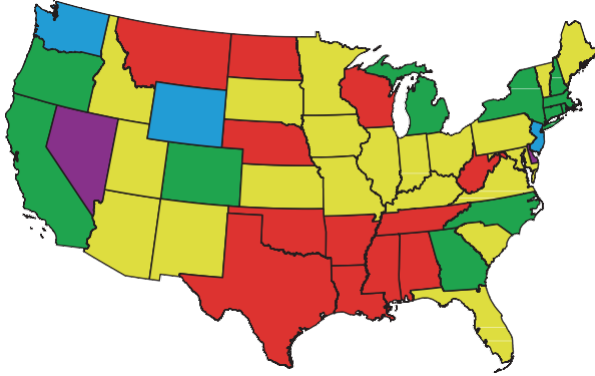


FIGURA1.2 : Residuos de Solow, 2001estados de EE.UU. (véase el gráfico en color en el inserto que sigue a la página 24)

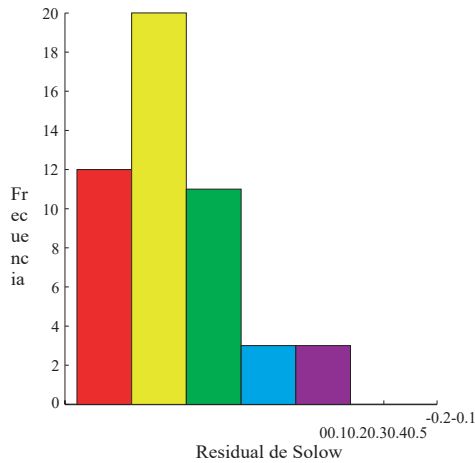


FIGURA1.3 : Leyenda del mapa de residuos de Solow (véase el inserto en color)

pregunta sobre la dependencia espacial observada en las variables dependientes de nuestros modelos, así como en los residuos. Por ejemplo, ha habido algunos trabajos teóricos sobre la ampliación de los modelos de crecimiento neoclásico para proporcionar una justificación de una variable dependiente *espacialmente retrasada* (López-Bazo et al., 2004; Ertur y Koch, 2007) en nuestro modelo de función de producción. Un rezago espacial de la variable dependiente es un vector de variables explicativas construido con un promedio de valores

de las regiones vecinas. Estos modelos teóricos postulan externalidades físicas y de capital humano, así como la interdependencia tecnológica entre regiones, lo que conduce a una regresión de forma reducida que incluye un retardo espacial de la variable dependiente.

Los métodos de comparación de modelos econométricos espaciales podrían utilizarse para poner a prueba estas teorías comparando los modelos que incluyen un retardo espacial de la variable dependiente con otras especificaciones de modelos que no lo hacen. Las predicciones o los pronósticos fuera de la muestra de los modelos que incluyen una variable dependiente con rezago espacial podrían compararse con los modelos que no incluyen estos términos para proporcionar evidencia a favor de estas teorías. Por último, las estimaciones e inferencias relativas a la significancia del parámetro asociado a la variable espacialmente retardada podrían utilizarse para mostrar la coherencia de estas teorías con los datos de la muestra.

Hay otras explicaciones posibles para el patrón de dependencia espacial observado. Dado que estamos mapeando residuos que reflexionan la producción total de los factores, éstos son condicionales a los insumos de capital y trabajo. Existe una gran cantidad de literatura que examina la producción regional desde el punto de vista de la nueva geografía económica (Duranton y Puga, 2001; Autant-Bernard, 2001; Autant-Bernard, Mairesse y Massard, 2007; Parent y LeSage, 2008). Estos estudios apuntan a los desbordamientos espaciales que surgen de la innovación tecnológica, medidos utilizando las patentes regionales como indicador del stock de conocimiento disponible en una región. En [el capítulo 3](#) se ofrece una ilustración aplicada de esta relación de productividad total de los factores que se utiliza para cuantificar la magnitud de los desbordamientos espaciales derivados de las diferencias regionales en la innovación técnica.

En las series temporales, las variables dependientes rezagadas pueden justificarse mediante modelos teóricos que incluyen ajustes costosos u otras fricciones de comportamiento que dan lugar de forma natural a rezagos de la variable dependiente. Como vimos con el ejemplo del tiempo de viaje al CBD, se puede utilizar una motivación similar para los rezagos espaciales. Otra justificación utilizada a menudo en el caso de las series temporales es que la variable dependiente retardada da cuenta de la variación en la variable dependiente que surge de influencias no observadas o latentes. Como hemos visto en el caso de nuestro ejemplo del precio hedónico de venta de viviendas, se puede utilizar una justificación similar para un rezago espacial de la variable dependiente. Las influencias latentes no observables relacionadas con la cultura, la infraestructura o los servicios recreativos pueden afectar la variable dependiente, pero pueden no aparecer como variables explicativas en

el modelo. El uso de un modelo de regresión espacial que incluya un desfase espacial del vector de la variable dependiente puede captar algunas de estas influencias.

1.2 El proceso autorregresivo espacial

Podríamos continuar de la forma de (1.3) para generar un conjunto mayor de observaciones como se muestra en (1.4).

$$\begin{aligned}
 y_i &= \alpha y_{i,jj} + \alpha y_{i,kk} + X\beta_i + \varepsilon_i \\
 y_j &= \alpha y_{j,ii} + \alpha y_{j,kk} + X\beta_j + \varepsilon_j \\
 y_k &= \alpha y_{k,ii} + \alpha y_{k,jj} + X\beta_k + \varepsilon_k \\
 \varepsilon_i &\sim N(0, \sigma^2) \quad i = 1 \\
 \varepsilon_j &\sim N(0, \sigma^2) \quad j = 2 \\
 \varepsilon_k &\sim N(0, \sigma^2) \quad k = 3
 \end{aligned} \tag{1.4}$$

Es fácil ver que esto tendría poca utilidad práctica, ya que daría lugar a un sistema con muchos más parámetros que observaciones.

Intuitivamente, una vez que permitimos las relaciones de dependencia entre un conjunto de n observaciones/ubicaciones, hay potencialmente n^2 relaciones que podrían surgir. Restamos n de las n^2 relaciones de dependencia potenciales porque descartamos la dependencia de una observación sobre sí misma.

La solución al problema de la sobreparametrización que surge cuando cada relación de dependencia tiene parámetros específicos de la relación es imponer una estructura a las relaciones de dependencia espacial. Ord (1975) propuso una parametrización parsimoniosa para las relaciones de dependencia (que se basó en un trabajo anterior de Whittle (1954)). Esta estructura da lugar a un proceso de generación de datos conocido como *proceso autorregresivo espacial*. Aplicado a las relaciones de dependencia entre las observaciones de la variable y , tenemos la expresión (1.5).

$$\begin{aligned}
 y_i &= \rho \sum_{j=1}^n W_{ij} y_j + \varepsilon_i \\
 \varepsilon_i &\sim N(0, \sigma^2) \quad i = 1, \dots, n
 \end{aligned} \tag{1.5}$$

Where we eliminate an intercept term by assuming that the vector of observations on the variable y is in deviations from means form. The term:

$\sum_{j=1}^n W_{ij} y_j$ se denomina *desfase espacial*, ya que representa una combinación lineal de

valores de la variable y contruidos a partir de las observaciones/regiones vecinas

observación i . Esto se consigue colocando elementos W_{ij} en el $n \times n$

matriz de pesos W , tal que $\sum_{j=1}^n W_{ij} y_j$ resulta en un escalar que representa una combinación lineal de los valores tomados por las observaciones vecinas.

Como ejemplo, consideremos las siete regiones mostradas en la [Figura 1.1](#). La única *vecina* de primer orden de la región $R1$ es la región $R2$, ya que ésta es la única región que tiene bordes que tocan la región $R1$.

Podemos escribir una versión matricial del proceso autorregresivo espacial como en (1.6), donde utilizamos $N(0, \sigma I_n^2)$ para denotar un proceso de perturbación de media cero que exhibe una varianza constante σ^2 , y una covarianza cero entre las observaciones. Esto da lugar a la matriz de varianza-covarianza diagonal σI_n^2 , donde I_n representa una matriz de identidad n -dimensional. La expresión (1.6) deja claro que estamos describiendo una relación entre el vector y y el vector Wy que representa una combinación lineal de valores vecinos a cada observación.

$$y = \rho W y + \varepsilon (1.6)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Para ilustrar esto, formamos una matriz de pesos espaciales W de 7×7 utilizando las relaciones de contigüidad de primer orden para las siete regiones mostradas en la figura 1.1. Esto implica asociar las filas de la matriz con el índice de observación i , y las columnas con el índice j que representa las observaciones/regiones vecinas a la región i . Comenzamos formando una matriz de contigüidad de primer orden C que se muestra en (1.7). Para la fila 1 colocamos un valor de 1 en la columna 2, reflejando el hecho de que la región R_2 es contigua de primer orden a la región R_1 . Todos los demás elementos de la fila 1 reciben valores de cero. Del mismo modo, para cada fila colocamos un 1 en las columnas asociadas a los vecinos contiguos de primer orden, lo que da lugar a la matriz C que se muestra en (1.7).

$$\begin{array}{r}
 \begin{array}{c} \mathbf{C} \\ = \end{array}
 \begin{array}{l}
 \begin{array}{c} \text{1R2 R3 R4 R 56R R 7R} \\ \text{R 1} \end{array} \\
 \begin{array}{c} \text{0100000} \\ \text{R 2} \end{array} \\
 \begin{array}{c} \text{1010000} \\ \text{R 3} \end{array} \\
 \begin{array}{c} \text{0101000} \\ \text{R 3} \end{array} \\
 \begin{array}{c} \text{0010100} \\ \text{R 50} \end{array} \begin{array}{c} \text{0 0 1 0 1} \end{array} \\
 \begin{array}{c} \text{0} \\ \text{R 6} \end{array} \\
 \begin{array}{c} \text{0000101} \\ \text{R 7} \end{array}
 \end{array}
 \end{array}$$

Observamos que los elementos diagonales de la matriz C son cero, por lo que las regiones no se consideran vecinas de sí mismas. Con el fin de formar un desfase espacial o una combinación lineal de valores de observaciones vecinas, podemos normalizar la matriz C para que tenga sumas de filas de la unidad. Esta *fila-estocástica*

La matriz que etiquetamos como W se muestra en (1.8), donde el término fila-estática se refiere a una matriz no negativa que tiene sumas de fila normalizadas para que sean iguales a uno.

$$W = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (1.8)$$

La matriz W puede multiplicarse con un vector y de valores tomados por cada región para producir un vector de *retraso espacial* de la variable dependiente vector que adopta la forma Wy . El producto matricial Wy funciona para producir un vector que representa el valor del vector de retraso espacial para cada observación i , $i = 1, \dots, 7$. En el capítulo 4 proporcionaremos detalles sobre varios enfoques para formular matrices de pesos espaciales, que implican formas alternativas de definir y ponderar las *observaciones vecinas*. Por ahora, observamos que el uso de la matriz W , que pondera por igual cada observación vecina, hará que el vector de retraso espacial sea un simple promedio de los valores de las observaciones vecinas (contiguas de primer orden) de cada región. El proceso de multiplicación de la matriz se muestra en (1.9), junto con el vector de retardo espacial resultante Wy .

$$Wy = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{pmatrix} = \begin{pmatrix} 0 \\ (y_1 + y_3)/2 \\ (y_2 + y_4)/2 \\ (y_3 + y_5)/2 \\ (y_4 + y_6)/2 \\ (y_5 + y_7)/2 \\ y_6 \end{pmatrix} \quad (1.9)$$

El parámetro escalar ρ en (1.6) describe la fuerza de la dependencia espacial en la muestra de observaciones. El uso de un único parámetro para reflejar un nivel medio de dependencia sobre todas las relaciones de dependencia que surgen de las observaciones $i = 1, \dots, n$, es una forma de *lograr la*

parsimonia mediante la autoreferencia espacial. , n , es una forma de lograr la parsimonia mediante la estructura autorregresiva espacial. Esto contrasta fuertemente con nuestro punto de partida en (1.3) y (1.4), donde permitimos que cada dependencia tuviera su propio parámetro.

Podemos examinar gráficamente un gráfico de dispersión de la relación entre las observaciones del vector y (en forma de desviación de la media) y los valores medios de las observaciones vecinas del vector Wy utilizando un *gráfico de dispersión de Moran*. Un ejemplo se muestra en [la Figura 1.4](#), en la que se traza la producción total de factores de los estados, construida utilizando los residuos de nuestra regresión de la función de producción de 2001 en el eje horizontal, y los valores de retardo espacial en el eje vertical. En virtud de la transformación a la desviación de la media, tenemos cuatro cuadrantes cartesianos en el gráfico de dispersión centrados en valores cero para los ejes horizontal y vertical. Estos cuatro cuadrantes se reflejan:

Cuadrante I (puntos rojos) estados que tienen una productividad de los factores (residuos) superior a la media, donde la media de la productividad de los factores de los estados vecinos también es superior a la media,

Cuadrante II (puntos verdes) Estados que presentan una productividad de los factores inferior a la media, pero la media de la productividad de los factores de los Estados vecinos es superior a la media,

Cuadrante III (puntos azules) Estados con una productividad de los factores inferior a la media, y la media de la productividad de los factores de los Estados vecinos también es inferior a la media,

Cuadrante IV (puntos morados) estados que tienen una productividad de los factores superior a la media, y la media de la productividad de los estados vecinos es inferior a la media.

En el gráfico de dispersión, vemos una asociación positiva entre las observaciones de la productividad de los factores y en el eje horizontal y las observaciones espacialmente retrasadas de Wy que se muestran en el eje vertical, lo que sugiere que el parámetro escalar ρ es mayor que cero. Otra forma de considerar la fuerza de la asociación positiva es observar que hay muy pocos puntos verdes y morados en el gráfico de dispersión. Los puntos verdes representan estados en los que la productividad de los factores está por debajo de la media y la de los estados vecinos Wy está por encima de la media. Lo contrario ocurre con los puntos morados, donde la productividad de los factores por encima de la media coincide con la productividad de los factores Wy de los estados vecinos por debajo de la media. Por el contrario, un gran número de puntos en los cuadrantes II y IV con pocos puntos en los cuadrantes I y III sugeriría una dependencia espacial negativa, de modo que $\rho < 0$.

Los puntos del gráfico de dispersión pueden situarse en un mapa utilizando el mismo esquema de codificación por colores, como en [la figura 1.5](#). Los estados rojos representan regiones con una productividad de los factores superior a la media (positiva) en las que la media de la productividad de los factores de los estados vecinos también está por encima de la media. El mapa muestra claramente la agrupación de los estados del noreste y del oeste con niveles de productividad de los factores superiores a la media en los que los

estados vecinos también tienen una productividad de los factores superior a la media. Del mismo modo, la agrupación de estados con niveles de productividad de los factores inferiores a la media y los estados circundantes que también están por debajo de la media es evidente en la parte central y meridional de EE.UU.

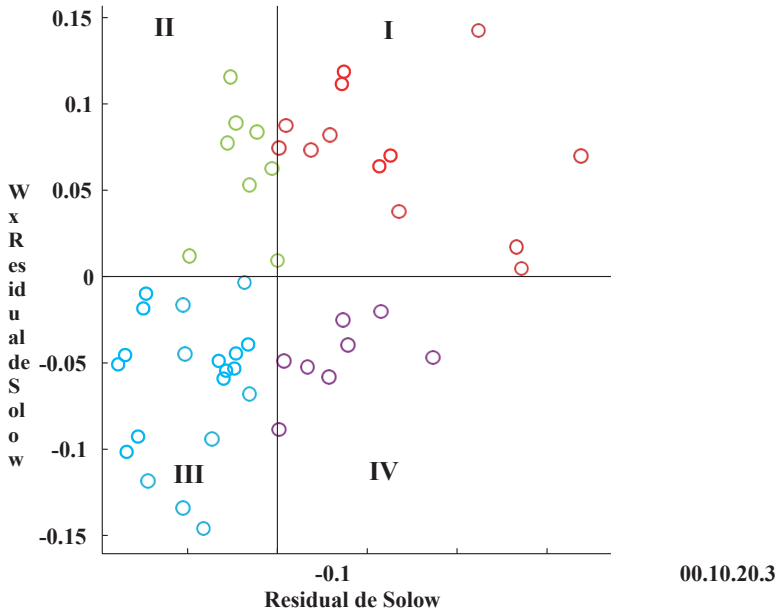


FIGURA1.4 : Gráfico de dispersión de Moran de la productividad de los factores de los 201Estados Unidos (véase el inserto en color)

Es tentador interpretar el parámetro escalar ρ en el proceso autorregresivo espacial como un coeficiente de correlación convencional entre el vector y y el vector de *retardo espacial* Wy . Esta tentación debe evitarse, ya que no es del todo precisa. Discutiremos este punto con más detalle en [el capítulo 2](#), pero tenga en cuenta que el rango de los coeficientes de correlación es $[-1, 1]$, mientras que ρ no puede ser igual a uno.

1.2.1 Proceso autorregresivo espacial de generación de datos

El proceso autorregresivo espacial se muestra en (1.10) utilizando la notación matricial, y el proceso generador de datos implícito para este tipo de proceso está en (1.11). Introducimos un vector de términos constantes de unos $\mathbf{1}_n$, y el parámetro asociado α para acomodar situaciones en las que el vector y no tiene un valor medio de cero.

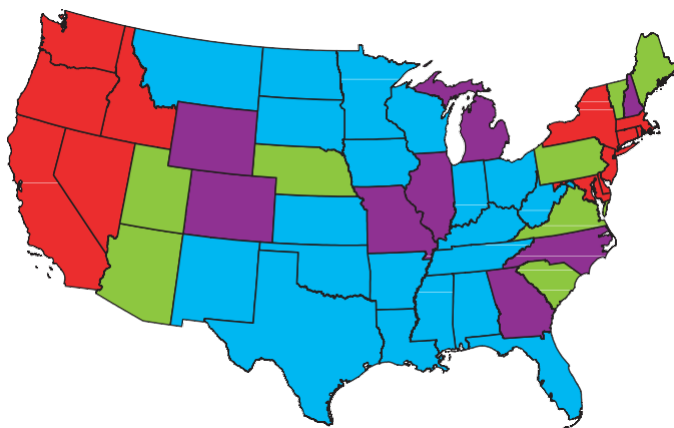


FIGURA1.5 : Mapa de Moran de la productividad de los factores de los estados de EE.UU. en 2001 (véase el inserto en color)

$$y = \alpha \mathbf{1}_n + \rho W y + \varepsilon \quad (1.10)$$

$$(\mathbf{I}_n - \rho W) y = \alpha \mathbf{1}_n + \varepsilon$$

$$y = (\mathbf{I}_n - \rho W)^{-1} \alpha \mathbf{1}_n + (\mathbf{I}_n - \rho W)^{-1} \varepsilon \quad (1.11)$$

$$\varepsilon \sim N(0, \sigma^2 \mathbf{I}_n)$$

El vector $n \times 1$ y contiene nuestra variable dependiente y ρ es un parámetro escalar, con W representando la matriz de pesos espaciales $n \times n$. Suponemos que ε sigue una distribución normal multivariante, con media cero y una matriz de varianza-covarianza diagonal constante $\sigma^2 \mathbf{I}_n$.

El enunciado del modelo en (1.10) puede interpretarse como que el valor esperado de cada observación y_i dependerá del valor medio α más

una combinación lineal de los valores tomados por las observaciones vecinas escaladas por el parámetro de dependencia ρ . El enunciado del *proceso generador de datos* en (1.11) expresa la naturaleza *simultánea* del proceso autorregresivo espacial. Para profundizar en su naturaleza, podemos utilizar la siguiente serie infinita para expresar la inversa:

$$(I_n - \rho W)^{-1} = I_n + \rho W + \rho W^2 + \rho W^3 + \dots \quad (1.12)$$

donde suponemos por el momento que $\text{abs}(\rho) < 1$. Esto conduce a un proceso autorregresivo espacial generador de datos para un vector variable y :

$$\begin{aligned} y &= (I_n - \rho W)^{-1} \alpha_n + (I_n - \rho W)^{-1} \varepsilon_n \\ y &= \alpha_n + \rho W \alpha_n + \rho W^2 \alpha_n + \dots \\ &\quad + \varepsilon_n + \rho W \varepsilon_n + \rho W^2 \varepsilon_n + \rho W^3 \varepsilon_n + \dots \end{aligned} \quad (1.13)$$

La expresión (1.13) puede simplificarse ya que la serie infinita: $\alpha_n + \rho W \alpha_n + \rho W^2 \alpha_n + \dots$ converge a $(1 - \rho)^{-1} \alpha_n$ ya que α es un escalar, el parámetro $\text{abs}(\rho) < 1$, y W es fila-estocástica. Por definición, $W_{ln} = \iota_n$ y por tanto $W(W_{ln})$ también es igual a $W_{ln} = \iota_n$. En consecuencia, $W^q \iota_n = \iota_n$ para $q \geq 1$ (recordemos que $W = \iota_n \iota_n'$). Esto nos permite escribir:

$$y = \frac{1}{(1 - \rho)} \alpha_n + \varepsilon_n + \rho W \varepsilon_n + \rho W^2 \varepsilon_n + \rho W^3 \varepsilon_n + \dots \quad (1.14)$$

Para explorar más a fondo la naturaleza de este proceso de generación de datos, consideramos las potencias de las matrices de pesos espaciales fila-estocásticas W, W^2, W^3, \dots que aparecen en (1.14). Supongamos que las filas de la matriz de pesos W se construyen para representar vecinos contiguos de primer orden. La matriz W^2 reflejará a los vecinos contiguos *de segundo orden*, aquellos que son vecinos de los vecinos de primer orden. Dado que el vecino del vecino (vecino de segundo orden) a una observación i incluye la propia observación i , W^2 tiene elementos positivos en la diagonal cuando cada observación tiene al menos un vecino. Es decir, los desfases espaciales de orden superior pueden conducir a una relación de conectividad para una observación i tal que W^2 extraerá las observaciones del vector ε que apuntan a la propia observación i . Esto contrasta fuertemente con nuestra relación de independencia inicial en (1.1), donde los supuestos de Gauss-Markov descartan la dependencia de ε_i de otras observaciones j , al suponer una covarianza cero entre las observaciones i y j en el proceso de generación de datos.

Para ilustrar este punto, mostramos W basándonos en la matriz de contigüidad de primer orden W de (1.8) en (1.15), donde los elementos positivos aparecen en la diagonal. Vemos que para la región R_1 , por ejemplo, los vecinos de segundo orden son las regiones R_1 y R_3 . Es decir, la región R_1 es vecina de segundo orden de sí misma y de la región R_3 , que es vecina de la región vecina R_2 .

$$W = \begin{pmatrix} 0.50 & 0 & 0 & 0.50 & & & & \\ & 0 & 0.75 & 0 & 0.25 & & & \\ & 0.25 & 0 & 0.50 & 0 & 0.25 & & \\ & & 0 & 0.25 & 0 & 0.50 & 0 & 0.25 \\ & & 0.00 & 0.25 & 0 & 0.50 & 0 & 0.25 \\ & & & 0.00 & 0.25 & 0 & 0.75 & 0 \\ & & & & 0.00 & 0.00 & 0.50 & 0.50 \end{pmatrix} \quad (1.15)$$

Dado que $\text{abs}(\rho) < 1$, el proceso de generación de datos asigna menos influencia a los términos de perturbación asociados con los vecinos de orden superior, con un decaimiento geométrico de la influencia a medida que aumenta el orden. Una dependencia espacial más fuerte reflexionada en valores mayores de ρ conduce a un papel más importante para los vecinos de orden superior.

La dependencia de cada observación y_i de las perturbaciones asociadas a las observaciones vecinas, así como a las vecinas de orden superior, sugiere una estructura de media y varianza-covarianza para las observaciones del vector y que dependen de forma complicada de otras observaciones. Es instructivo considerar la media de la variable y que surge del proceso autorregresivo espacial de generación de datos en (1.13). Obsérvese que suponemos que la matriz de pesos espaciales es exógena, o fija en el muestreo repetido, de modo que:

$$\begin{aligned} E(y) &= \frac{1}{(1 - \rho)} \alpha_n + E(\varepsilon) + \rho W E(\varepsilon) + \rho W^2 E(\varepsilon) + \dots \\ &= \frac{1}{(1 - \rho)} \alpha_n \end{aligned} \quad (1.16)$$

Es interesante observar que en las redes sociales (Katz, 1953; Bonacich, 1987) interpretan el vector $b = (I_n - \rho P)^{-1} 1_n$ como una medida de *centralidad* de los individuos en una red social, donde la matriz P es una matriz binaria de pares, por lo que el vector b refleja sumas de filas de la matriz inversa.¹ El vector b (denominado centralidad de Katz-Bonacich en las redes sociales) mide el número de conexiones directas e indirectas que tiene un individuo en una red social.

Por ejemplo, si la matriz P identifica a los amigos, entonces P apunta² a los amigos de los amigos, P^2 a los amigos de los amigos de los amigos, y así sucesivamente. En las redes sociales, se considera que los individuos están situados en nodos de una red, y el parámetro ρ refleja un factor de descuento que crea una decadencia de influencia para los amigos/pares que están situados en nodos más distantes. Estas observaciones no hacen más que señalar

que el proceso autorregresivo espacial ha desempeñado un papel importante en otras disciplinas además de la estadística espacial, y probablemente seguirá creciendo en uso e importancia.

La retroalimentación simultánea es útil en la modelización de las relaciones de dependencia espacial en las que deseamos acomodar los efectos de

retroalimentación espacial de las regiones vecinas a un lugar de origen i donde se produjo un impacto inicial. De hecho, estos

¹ La matriz binaria de pares se define como nuestra matriz de contigüidad C , teniendo valores para los 1pares y para los 0no pares.

nos permiten tratar todas las observaciones como orígenes potenciales de un impacto sin pérdida de generalidad. Se podría suponer que los efectos de retroalimentación llevarían tiempo, pero no hay un papel explícito para el paso del tiempo en una relación transversal. En su lugar, podemos ver las observaciones de los datos de la muestra transversal como reflexión de un resultado de equilibrio o estado estacionario del proceso espacial que estamos modelando. Desarrollamos esta idea en los capítulos 2 y 7. Esta es una interpretación que se utiliza a menudo en la modelización transversal y Sen y Smith (1995) proporcionan ejemplos de este tipo de situación para los modelos de interacción espacial convencionales utilizados en el análisis regional. El objetivo de los modelos de interacción espacial es analizar la variación de los flujos entre regiones que se producen a lo largo del tiempo utilizando una sección transversal de los flujos observados entre las regiones de origen y de destino que han tenido lugar durante un período de tiempo finito, pero medidos en un único momento. En el capítulo 8 se analizan los modelos econométricos espaciales para los flujos de origen-destino.

Esta situación de dependencia simultánea no se da en el análisis de series temporales, por lo que los procesos autorregresivos espaciales son distintos de los procesos autorregresivos de series temporales. En las series temporales, el operador de *desfase temporal* L es estrictamente triangular y contiene ceros en la diagonal. Las potencias de L también son estrictamente triangulares con ceros en la diagonal, de modo que L^2 especifica un desfase temporal de dos períodos mientras que L crea un desfase de un solo periodo. Nunca se da el caso de que L^2 produzca observaciones que apuntan hacia atrás para incluir el período de tiempo actual.

1.3 Una ilustración de los efectos secundarios espaciales

La estructura autorregresiva espacial puede combinarse con un modelo de regresión convencional para producir una extensión espacial del modelo de regresión estándar mostrado en (1.17), con el proceso de generación de datos implícito en (1.18). Nos referiremos a esto simplemente como el *modelo autorregresivo espacial* (SAR) a lo largo del texto. Observamos que Anselin (1988) etiquetó este modelo como un modelo "mixto-regresivo, espacial-autorregresivo", donde la motivación de esta incómoda nomenclatura debería estar clara.

$$y = \rho W y + X\beta + \varepsilon \quad (1.17)$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1} \varepsilon \quad (1.18)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

En este modelo, los parámetros a estimar son los parámetros de regresión habituales β , σ y el parámetro adicional ρ . Cabe destacar que si el parámetro escalar ρ toma un valor de cero para que no haya dependencia espacial en el

vector de observaciones transversales y , se obtiene el modelo de regresión por mínimos cuadrados como un caso especial del modelo SAR.

Para ilustrar el modo en que el modelo de regresión espacial puede utilizarse para cuantificar los efectos indirectos espaciales, volvemos a utilizar el ejemplo anterior de los tiempos de viaje al CBD desde las siete regiones que se muestran en la Figura 1.1. Consideramos el impacto de un cambio en la densidad de población de una sola región sobre los tiempos de viaje al CBD de las siete regiones. Específicamente, duplicamos la densidad de población en la región R2 y hacemos una predicción del impacto en los tiempos de viaje al CBD para las siete regiones.

Utilizamos las estimaciones de los parámetros: $\hat{\beta} = 0.1350.561$ y $\hat{\rho} = 0.642$ para este ejemplo. El valor estimado de ρ indica una dependencia espacial positiva en los tiempos de desplazamiento. Las predicciones del modelo basadas en la matriz de variantes explicativas X tendrían la forma

$$\hat{y}^{(1)} = (I_n - \hat{\rho}W$$

)⁻¹ $X\hat{\beta}$ donde $\hat{\rho}, \hat{\beta}$ son estimaciones de máxima

verosimilitud.

La comparación de las predicciones $\hat{y}^{(1)}$ del modelo con variables explicativas de X y $\hat{y}^{(2)}$ del modelo basado en X que se muestra en (1.19) sirve para ilustrar cómo el modelo genera desbordamientos espaciales cuando cambia la densidad de población de una sola región. La matriz X refleja una duplicación de la densidad de población de la región R2.

$$\tilde{X} = \begin{pmatrix} 10 & 30 & 20 & 40 & 30 & 10 & 500 \\ 30 & 10 & 20 & 20 & 10 & 30 & 10 \end{pmatrix} \quad (1.19)$$

Los dos conjuntos de predicciones $\hat{y}^{(1)}, \hat{y}^{(2)}$ se muestran en la tabla 1.1, donde vemos que el cambio en la densidad de población de la región R2 tiene un efecto directo que aumenta los tiempos de desplazamiento de los residentes de la región R2 en 4 minutos. También tiene un efecto indirecto o de derrame que produce un aumento en los tiempos de desplazamiento para las otras seis regiones. El aumento de los tiempos de desplazamiento para las regiones vecinas al este y al oeste (regiones R1 y R3) es el mayor y estos efectos indirectos disminuyen a medida que nos desplazamos a las regiones de la muestra que están situadas más lejos de la región R2 donde se produjo el cambio en la densidad de población.

También es interesante saber que los impactos indirectos acumulados (spillovers) se pueden encontrar sumando el aumento de los tiempos de

desplazamiento en todas las demás regiones (excluyendo el cambio en el tiempo de desplazamiento de la propia región). Esto equivale a $2.57 + 1.45 + 0.53 + 0.20 + 0.07 + 0.05 = 4.87$ minutos, que es mayor que el impacto directo (de la propia región) de 4 minutos. El impacto total en todos los residentes del área metropolitana de las siete regiones por el cambio en la densidad de población de

TABLA 1.1: Efectos espaciales de los cambios en la densidad de población de la

Región R2			
Regiones / Escenario	$\hat{y}^{(1)}$	$\hat{y}^{(2)}$	$\hat{y}^{(2)} - \hat{y}^{(1)}$
R2 :	42.01	44.58	2.57
R3 :	37.06	41.06	4.00
R4 : CDB	29.94	31.39	1.45
R5 :	26.00	26.54	0.53
R6 :	29.94	30.14	0.20
R7 :	37.06	37.14	0.07
R7 :	42.01	42.06	0.05

La región R 2es la suma de los efectos directos e indirectos, o el aumento en minutos8.87 de los tiempos de viaje al CBD.²

El modelo sugiere literalmente que el cambio en la densidad de población de la región R2 provocaría inmediatamente un aumento de los tiempos de desplazamiento diario observados en todas las regiones. Una interpretación más aceptable sería que el cambio en la densidad de población conduciría con el tiempo a un nuevo estado estacionario de equilibrio en la relación entre los tiempos de desplazamiento diario y las variables de distancia y densidad. Las predicciones de los impactos directos derivados del cambio en densidad reflect $\partial y/\partial X_{ii}$, donde X_{ii} se refiere a la i-ésima observación de la segunda variable explicativa en el modelo. Las derivadas parciales cruzadas $\partial y/\partial X_{ji}$ representan los efectos indirectos asociados a este cambio.

Para explicar esto, observamos que la DGP para el modelo SAR puede escribirse como en (1.20), donde el subíndice r denota la variable explicativa r ,

$$y = S \sum_{r=1}^k r(W) X_r + (I_n - \rho W) \varepsilon^{-1} \quad (1.20)$$

$$E(y) = S \sum_{r=1}^k r(W) X_r \quad (1.21)$$

donde $S_r(W) = (I_n \rho W)^{-\beta_1}_r$ actúa como una matriz "multiplicadora" que aplica relaciones de vecindad de orden superior a X_r . Los modelos que contienen rezagos espaciales de la variable dependiente presentan una derivada complicada de y_i con respecto a X_{jr} , donde i, j denotan dos observaciones distintas. De (1.21) se deduce que:

$$\frac{\partial E(y_i)}{\partial X_{jr}} = S(W)_{rij} \quad (1.22)$$

donde $S_r(W)_{ij}$ representa el elemento ij de la matriz $S_r(W)$.

² A lo largo del texto utilizaremos los términos *impactos* y *efectos* indistintamente cuando nos refiramos a los efectos o impactos directos e indirectos.

Como indica la expresión (1.22), la interpretación de regresión estándar de Las estimaciones coeficientes como derivadas parciales: $\hat{\beta}_r = \partial y / \partial X_r$, ya no se cumplen.

Debido a la transformación de X_r por la matriz $n \times n S_r(W)$, cualquier cambio en una variable explicativa en una región determinada (observación) puede afectar a la variable dependiente en todas las regiones (observaciones) a través de la matriz inversa.

Dado que el impacto de los cambios en una variable explicativa differ a lo largo de todas las observaciones, parece deseable finir una medida resumen para la propia derivada

$\partial y / \partial X_{ir}$ en (1.22) que muestra el impacto que surge de un cambio en la i -ésima observación de la variable r . También sería interesante resumir la derivada cruzada $\partial y / \partial X_{ijr}$ ($i = j$) en (1.22) que mide el impacto en y_i de los cambios en la observación j de la variable r . Seguimos este tema en detalle en el capítulo 2, donde proporcionamos medidas resumidas e interpretaciones para los impactos que surgen de los cambios representados por las derivadas propias y cruzadas. A pesar de la simplicidad de este ejemplo, proporciona una ilustración de cómo los modelos de regresión espacial permiten los efectos indirectos de los cambios en las variables explicativas de una única región de la muestra. Este es un aspecto valioso de los modelos econométricos espaciales que los diferencia de la mayoría de los modelos estadísticos espaciales,

una cuestión que tratamos en la siguiente sección.

Un modelo de regresión ordinario predeciría que el cambio en la densidad de población en la región R2 sólo afecta al tiempo de desplazamiento de los residentes en la región R2, sin tener en cuenta los impactos espaciales. Para ver esto, podemos establecer el parámetro $\rho = 0$ en nuestro modelo, que produce el efecto no espacial modelo de regresión. En este caso $\hat{y}^{(1)} = X\hat{\beta}^{(0)}$ e $\hat{y}^{(2)} = X\hat{\beta}^{(0)}$ por lo que la diferencia

sería $X\hat{\beta}^{(0)} = (XX)^{-1}Xy$, donde los parámetros estimados $\hat{\beta}^{(0)}$ serían sean los de una regresión por mínimos cuadrados.

Si la DGP para nuestros tiempos de viaje diarios observados es la del modelo SAR, las estimaciones por mínimos cuadrados estarán sesgadas y serán inconsistentes, ya que ignoran el desfase espacial de la variable dependiente. Para ver esto, observe que las estimaciones para $\hat{\beta}$ del modelo de RAS toman la forma: $\hat{\beta} = (XX)^{-1}(I\rho^*W_n)y$, un sujeto que tratamos con más detalle en el capítulo 2. Para nuestra sencilla ilustración en la que todos los valores de y y X son positivos, y el parámetro de dependencia espacial también es positivo, esto sugiere un sesgo al alza en las estimaciones por mínimos cuadrados. Esto puede verse observando que:

$$\begin{aligned}\hat{\beta} &= (XX)^{-1}Xy - \rho^*(XX)^{-1}XW_ny \\ \hat{\beta} &= \hat{\beta}^{(0)} - \rho^*(XX)^{-1}XW_ny \\ \hat{\beta} &= \hat{\beta}^{(0)} + \rho^*(XX)^{-1}XW_ny\end{aligned}$$

Dado que todos los valores de y son positivos, el vector de desfase

espacial Wy contendrá los promedios de los valores vecinos que también serán positivos. Esto en conjunto con sólo elementos positivos en la matriz X así como ρ^{\wedge} positivos.

nos llevan a concluir que las estimaciones por mínimos cuadrados β^{\wedge} o estarán sesgadas al alza

en relación con las estimaciones insesgadas β^{\wedge} . Para nuestro ejemplo de siete regiones, las estimaciones por mínimos cuadrados fueron: $\beta^{\wedge}=0.551.25$ que muestran un sesgo al alza en relación con

las estimaciones del modelo autorregresivo espacial: $\hat{\beta} = 0.1350 . 561$. Intuitivamente, el modelo de mínimos cuadrados ordinarios intenta explicar la variación en los tiempos de viaje que surge de los efectos de la congestión mediante las variables de distancia y densidad de población. El resultado es una exageración de la verdadera influencia de estas variables en los tiempos de viaje.

Se presentan las predicciones por mínimos cuadrados basadas en las matrices X y \tilde{X} en el cuadro 1.2. Vemos que de este modelo no surgen desbordamientos espaciales, ya que sólo el tiempo de viaje al CBD para la región R2 está afectado por el cambio en la densidad de población de la región R2. También vemos el impacto del sesgo al alza en las estimaciones de mínimos cuadrados, que producen una predicción influida del cambio en el tiempo de viaje que surgiría del cambio en la densidad de población.

TABLA 1.2: Predicciones no espaciales de los cambios en la densidad de población de la Región R2

Regiones / Escenario	$\hat{y}^{(1)}$	$\hat{u}^{(2)}$	$\hat{u}^{(2)} - \hat{u}^{(1)}$
R1 :	42.98	42.98	0.00
R2 :	36.00	47.03	11.02
R3 :	29.02	29.02	0.00
R4 : CDB	27.56	27.56	0.00
R5 :	29.02	29.02	0.00
R6 :	36.00	36.00	0.00
R7 :	42.98	42.98	0.00

1.4 El papel de los modelos econométricos espaciales

Un tema de larga data en la economía es cómo la búsqueda del interés propio resulta en beneficios o costos que recaen en otros. Estos beneficios o costes se denominan externalidades. En situaciones en las que los efectos indirectos son de naturaleza espacial, los modelos económicos espaciales pueden cuantificar la magnitud de los mismos, como se ilustra en el ejemplo del tiempo de viaje al CBD.

Hay muchos otros ejemplos. La innovación tecnológica que surge como resultado de los desbordamientos de conocimiento espacial de las regiones cercanas es un ejemplo de externalidad o desbordamiento positivo. Se argumenta que una gran parte del conocimiento es tácito porque las ideas que conducen a la innovación técnica están plasmadas en las personas y

vinculadas a la experiencia del inventor. Este acervo de conocimientos aumenta en una región a medida que los inventores locales descubren nuevas ideas y disposiciones sobre todo a través de las interacciones cara a cara. Podemos considerar el conocimiento como un bien público local que beneficia a los investigadores de una región y a las regiones vecinas.

Esto motiva una especificación espacial para el conocimiento no observado que no se incluiría como variable explicativa del modelo. En general, se cree que el conocimiento tácito vinculado a la experiencia de los inventores e investigadores no "viaja bien", por lo que se piensa que los desbordamientos de conocimiento son de naturaleza local y recaen sólo en las regiones cercanas. Podemos utilizar modelos de regresión espacial para cuantificar el alcance espacial de los spillovers examinando los efectos indirectos mediante la expansión de la serie $I_n + \rho W + \rho W^2 + \dots$ que surge en la expresión de la derivada parcial

para estos efectos. El capítulo 3 explorará esta cuestión en una ilustración aplicada

que relaciona la productividad total de los factores regionales con los desbordamientos de conocimiento.

La contaminación es otro ejemplo, ya que es probable que estas externalidades negativas o desbordamientos sean de naturaleza espacial. La capacidad de cuantificar los efectos directos e indirectos de las fuentes de contaminación debería ser útil en el análisis empírico de las soluciones clásicas de impuestos y subsidios pigovianos para los fallos del mercado.

A menudo se piensa que los gobiernos regionales tienen en cuenta las acciones de los gobiernos vecinos a la hora de fijar los tipos impositivos (Wilson, 1986) y decidir sobre la prestación de servicios del gobierno local (Tiebout, 1956). Los modelos económicos espaciales pueden utilizarse para examinar empíricamente la magnitud y la significación estadística de la interacción de los gobiernos locales. El uso de las medidas derivadas parciales de los efectos directos e indirectos que surgen de los cambios en las variables explicativas debería ser especialmente útil desde la perspectiva de las políticas públicas. En un modelo de decisiones del gobierno de un condado, las estimaciones de los efectos directos se refieren a los impactos que serían de interés primordial para los funcionarios del gobierno de ese condado, mientras que los efectos indirectos y totales se refieren a la perspectiva más amplia de la sociedad en general. Gran parte de la literatura sobre la elección pública se centra en situaciones en las que los intereses privados y públicos, o de los gobiernos locales y nacionales, divergen. En el caso de los gobiernos locales y nacionales, la divergencia puede verse en términos de efectos de desbordamiento espacial. Una vez más, la capacidad de los modelos de regresión espacial para cuantificar la magnitud relativa de la divergencia debería ser útil para quienes estudian las cuestiones de elección pública.

Hay una diferencia fundamental entre los modelos que contienen rezagos espaciales de la variable dependiente y los que modelan la dependencia espacial en las perturbaciones. Exploramos esto utilizando el modelo de error general en (1.23), donde $F(W)$ en (1.24) representa una función matricial no sinular que implica una matriz de pesos espaciales W .

$$\begin{aligned} y &= X\beta + \\ &= F(W)\varepsilon \end{aligned} \quad (1.23) \quad (1.24)$$

La expectativa de y para estos modelos de error aparece en (1.25).

$$E(y) = X\beta \quad (1,25)$$

Esto significa que todos los distintos tipos de modelos de error tienen la misma expectativa que el modelo no espacial. Muestras de tamaño suficientemente grande y consistente.

En el caso de muestras pequeñas, las estimaciones podrían variar, y el uso de modelos que difieren de la DGP podría dar lugar a estimaciones incoherentes de la dispersión de los parámetros del modelo. La interpretación de los parámetros β de este tipo de modelo es la misma que para un modelo de regresión lineal no espacial.

Anselin (1988) ofrece un argumento convincente de que la econometría espacial debería centrarse en la medición de los efectos indirectos. En este texto prestamos una atención limitada a los modelos de error porque estos modelos eliminan los efectos indirectos por construcción. Estos podrían añadirse haciendo que X sea más complejo espacialmente, pero hay alternativas más atractivas que exploraremos aquí.

1.5 El plan del texto

Este capítulo introductorio se ha centrado en una breve introducción a la dependencia espacial y a los procesos autorregresivos espaciales, así como a las matrices de pesos espaciales utilizadas en estos procesos. Estos procesos se pueden utilizar para producir una serie de modelos econométricos espaciales que tienen en cuenta la dependencia espacial de diversas formas.

En el capítulo 2 se ofrecen motivaciones más detalladas sobre la dependencia espacial y el uso de modelos de regresión espacial. Explicamos la idea de que las variables omitidas o excluidas en nuestros modelos que presentan dependencia espacial pueden conducir a modelos de regresión espacial que contengan rezagos espaciales de la variable dependiente. Los modelos de regresión espacial simultánea transversal también están motivados como un resultado de estado estacionario a largo plazo de situaciones de dependencia no simultánea. Consideramos situaciones en las que los agentes económicos pueden observar las acciones pasadas de los agentes vecinos, por ejemplo, los oficiales del gobierno del condado deben conocer los tipos impositivos del gobierno vecino o los niveles de prestación de servicios del gobierno en el período anterior. Este tipo de relación dinámica espacio-temporal no simultánea es consistente con una relación de regresión espacial simultánea transversal que representa el resultado de estado estacionario a largo plazo de la relación dinámica espacio-temporal. También proporcionamos detalles sobre la interpretación de las estimaciones de estos modelos. Se ofrece una explicación sobre los efectos directos e indirectos asociados a los cambios en las variables explicativas que se introdujeron en el ejemplo del tiempo de viaje de este capítulo.

El capítulo 3 se centra en una familia de modelos de regresión espacial popularizados por Anselin (1988) en su influyente texto sobre econometría espacial. Se discuten las implicaciones para las estimaciones e inferencias basadas en los mínimos cuadrados de la regresión no espacial en presencia de dependencia espacial. Este capítulo también proporciona detalles sobre los aspectos computacionales de la estimación de máxima verosimilitud para la

familia de modelos de regresión espacial. Los métodos computacionales han avanzado considerablemente desde el 1988,año del texto de Anselin.

El capítulo 4 aborda diversos aspectos computacionales y teóricos de los modelos econométricos espaciales. Los temas incluyen el cálculo de las matrices de pesos espaciales, los determinantes logarítmicos (incluyendo numerosos casos especiales como la matriz exponencial, los sistemas de ecuaciones, las matrices de pesos múltiples y las matrices de flujo), las derivadas de los determinantes logarítmicos, las diagonales de la matriz de varianza-covarianza y las soluciones de forma cerrada para una serie de modelos espaciales de un solo parámetro.

Los métodos bayesianos convencionales para el análisis de modelos econométricos espaciales (Anselin, 1988; Hepple, 1995a,b), así como los métodos *bayesianos de Markov Chain Monte Carlo* (MCMC) más recientes (LeSage, 1997) para la estimación de modelos de regresión espacial son el tema del capítulo 5. El enfoque expuesto en LeSage (1997) permite el tratamiento formal de la *heterogeneidad espacial* que se motiva en el capítulo 2. Demostramos que muchos de los avances computacionales descritos para la estimación de máxima verosimilitud en el capítulo 3 también sirven para simular la estimación bayesiana de estos modelos.

La especificación y comparación de modelos es el tema del capítulo 6. Las cuestiones de especificación consideradas incluyen la forma de la matriz de pesos, la preocupación habitual por las variables explicativas apropiadas y las cuestiones relativas a cuál de los miembros alternativos de la familia de modelos de regresión espacial introducidos en el capítulo debería emplearse. Mostramos cómo los métodos formales de comparación de modelos bayesianos propuestos por LeSage y Parent (2007) pueden utilizarse para responder a las preguntas relativas a las variables explicativas adecuadas para la familia de modelos del capítulo 3. Los métodos bayesianos de comparación de modelos también pueden utilizarse para discriminar entre los modelos basados en matrices de pesos espaciales alternativas, como señalan LeSage y Pace (2004a), y las distintas especificaciones que surgen de la familia de modelos de regresión espacial del capítulo (3Hepple, 2004).

El capítulo se 7 diferencia de otros capítulos del texto porque es más teórico y se centra en los fundamentos espacio-temporales de la dependencia espacial transversal observada. Partiendo del supuesto de que las regiones están influenciadas únicamente por los valores de *su propio* período y de *otras regiones* en el pasado, desarrollamos una motivación espacio-temporal para la dependencia espacial simultánea implícita en el proceso autorregresivo espacial. Desarrollamos la discusión en el capítulo mostrando cómo la dependencia temporal de las decisiones pasadas de los agentes económicos vecinos conducirá a especificaciones de regresión espacial simultánea. Demostramos que un marco espacio-temporal estricto consistente con un mecanismo de ajuste parcial espacial puede dar lugar a un equilibrio a largo plazo caracterizado por la dependencia espacial simultánea.

Las extensiones econométricas espaciales de los modelos convencionales de gravedad por mínimos cuadrados o de interacción espacial descritos en Sen y Smith (1995) son el tema del capítulo 8. Presentamos modelos de regresión espacial similares a los del capítulo presentado por LeSage y Pace (2008) que pueden aplicarse a modelos que intentan explicar la variación de los flujos entre orígenes y destinos. Si se tiene en cuenta la dependencia espacial en los

orígenes, los destinos y entre los orígenes y los destinos, se llega a una situación en la que los cambios en el origen o el destino darán lugar a fuerzas que pondrán en marcha una serie de acontecimientos. Exploramos la noción avanzada por Behrens, Ertur y Koch (2007) de que la dependencia espacial

La dependencia sugiere un mundo multilateral en el que las interacciones indirectas vinculan a todas las regiones. Esto contrasta con el énfasis convencional en los flujos bilaterales de las regiones de origen a las de destino.

El capítulo 9 presenta un enfoque alternativo para la modelización econométrica espacial que sustituye el proceso autorregresivo espacial por un enfoque exponencial matricial para especificar las estructuras de dependencia espacial (LeSage y Pace, 2007, 2004b). Esto tiene ventajas tanto computacionales como teóricas sobre el proceso autorregresivo espacial más convencional. Analizamos los enfoques de máxima verosimilitud y bayesiano para estimar los modelos basados en esta nueva especificación del proceso espacial.

El capítulo 10 aborda el tema de las regresiones espaciales con variables dependientes binarias, de recuento o truncadas. Esto se basa en el trabajo relativo a las variables dependientes binarias en el contexto de la familia de modelos del capítulo descrito en LeSage (2000) y analizado por Flemming (2004). También se discute el uso de procesos autorregresivos espaciales como una prioridad bayesiana para los efectos estructurados espacialmente introducidos en Smith y LeSage (2004) para el caso de los modelos probit y en LeSage, Fischer y Scherngell (2007) para los modelos de datos de conteo de Poisson. Este enfoque para estructurar los parámetros de los efectos individuales puede utilizarse para superar los problemas que suelen surgir al estimar los efectos individuales (Christensen, Roberts y Skold, 2006; Gelfand, Sahu y Carlin, 1995).

Capítulo 2

Motivación e interpretación de los modelos econométricos espaciales

En las primeras cinco secciones de este capítulo, proporcionamos motivaciones separadas para los modelos de regresión que incluyen procesos autorregresivos espaciales. Estas motivaciones se exploran con más detalle en capítulos posteriores del texto, siendo la presentación aquí menos formal. La sección 2.1 muestra cómo las relaciones de los modelos transversales que implican rezagos espaciales de la variable dependiente (el modelo SAR) provienen de los agentes económicos que consideran el comportamiento del período pasado de los agentes vecinos. La sección 2.2 presenta una segunda situación en la que las variables omitidas que presentan dependencia espacial conducen a un modelo que incluye rezagos espaciales tanto de la variable dependiente como de la independiente. En las secciones siguientes se 2.3ofrecen2.5 motivaciones adicionales basadas en la heterogeneidad espacial, las externalidades y la incertidumbre del modelo. En conjunto, las motivaciones de las secciones 2.1 a 2.5 muestran cómo surgen una serie de estructuras de regresión espacial alternativas cuando la dependencia entra en una combinación de las variables explicativas, las variables dependientes o las perturbaciones.

La sección 2.6 introduce brevemente una familia de modelos de regresión espacial convencionales que han aparecido en la literatura empírica. La sección 2.7 está dedicada a una discusión sobre la interpretación de las estimaciones de los parámetros de los modelos de regresión espacial. Esta cuestión ha sido especialmente malinterpretada en los estudios aplicados que se han basado en modelos de regresión espacial. Introducimos algunos procedimientos relativamente sencillos que simplifican el análisis de los impactos resultantes de los cambios en las variables explicativas de estos modelos.

2.1 Una motivación de dependencia temporal

Los agentes económicos a menudo toman decisiones en el período actual que están influenciadas por el comportamiento de otros agentes en períodos anteriores. Por ejemplo, los gobiernos locales pueden fijar los tipos impositivos tras observar los tipos fijados por las regiones vecinas en períodos anteriores. Aunque los tipos impositivos fueron fijados a lo largo del tiempo

por la sección transversal de las regiones que representan nuestra muestra, los tipos impositivos transversales observados mostrarían un patrón de dependencia espacial.

Para ilustrar esto, consideremos una relación en la que el vector de la variable dependiente en

tiempo t , denotado y_t , se determina utilizando un esquema autorregresivo espacial que depende de los valores rezagados en el espacio-tiempo de la variable dependiente de las observaciones vecinas. Esto llevaría a un desfase temporal de los valores vecinos medios de la variable dependiente observados durante el periodo anterior, Wy_{t-1} . También podemos incluir en nuestro modelo las características de la propia región del período actual, X_t .

En el caso de que las características de las regiones permanezcan relativamente fijas a lo largo del tiempo, podemos escribir $X_t = X$ e ignorar el subíndice de tiempo para esta matriz de características regionales. Como ejemplo concreto de este tipo de situación, consideremos un modelo que incluya los precios de venta de las viviendas como variable dependiente y_t , que dependen de los precios de venta de las viviendas vecinas en el período anterior, Wy_{t-1} .

Las características de las viviendas, como el número de dormitorios o los baños, cambian

muy lentamente en el tiempo. Esto sugiere la siguiente relación como representación del proceso autorregresivo con retardo en el tiempo:

$$y_t = \rho Wy_{t-1} + X\beta + \varepsilon_t \quad (2.1)$$

Obsérvese que podemos sustituir y_{t-1} en el lado derecho anterior por: $y_{t-1} =$

$\rho Wy_{t-2} + X\beta + \varepsilon_{t-1}$ produciendo:

$$y_t = X\beta + \rho W(X\beta + \rho Wy_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \quad (2.2)$$

$$y_t = X\beta + \rho WX\beta + \rho W^2 y_{t-2} + \varepsilon_t + \rho W\varepsilon_{t-1} \quad (2.3)$$

La sustitución recursiva de los valores pasados del vector y_{t-r} en el lado derecho de (2.3) a lo largo de q períodos conduce a (2.4) y (2.5).

$$y_t = I_n + \rho W + \rho^2 W^2 + \dots + \rho^{q-1} W^{q-1} X\beta + \rho^q W^q y_{t-q} + u \quad (2.4)$$

$$u = \varepsilon_t + \rho W\varepsilon_{t-1} + \rho W^2 \varepsilon_{t-2} + \dots + \rho W^{q-1} \varepsilon_{t-(q-1)} \quad (2.5)$$

Estas expresiones pueden simplificarse observando que $E(\varepsilon_{t-r}) = 0, r = 0, \dots, q-1$, implica que $E(u) = 0, q-1$, implica que $E(u) = 0$. Además, la magnitud de $\rho W^q y_{t-q}$ se hace pequeña para grandes q , bajo el supuesto habitual de que $\rho < 1$ y suponiendo

que W es una fila estocástica, por lo que la matriz W tiene un valor propio principal de 1. En consecuencia, podemos interpretar la relación transversal observada como el resultado o la expectativa de un equilibrio a largo plazo o estado estacionario mostrado en (2.6).

$$\lim_{q \rightarrow \infty} E(y_t) = (I - \rho W)^{-1} X\beta \quad (2.6)$$

Obsérvese que esto proporciona una motivación dinámica para el proceso de generación de datos del modelo SAR transversal que sirve de caballo de

batalla del modelado de regresión espacial. Es decir, una relación del modelo SAR transversal puede surgir de la dependencia temporal de las decisiones de los agentes económicos situados en varios puntos del espacio cuando las decisiones dependen de las de los vecinos.

2.2 Una motivación de variables omitidas

Las variables omitidas pueden surgir fácilmente en la modelización espacial porque los factores no observables, como las comodidades de la ubicación, la accesibilidad a las carreteras o el prestigio del vecindario, pueden ejercer una influencia en la variable dependiente. Es poco probable que las variables explicativas estén fácilmente disponibles para captar este tipo de influencias latentes. Exploramos esta situación utilizando un escenario muy simple que implica una variable dependiente y que se explica completamente por dos variables explicativas x y z con parámetros escalares asociados β y θ . Para simplificar, suponemos que los n vectores x y z se distribuyen $N(0, I_n)$, y suponemos que x y z son independientes.

$$y = x\beta + z\theta \quad (2.7)$$

Dados tanto x como z , la solución del sistema lineal produciría una β y θ exactas. La ausencia de un término de perturbación simplifica el descubrimiento de los parámetros. Consideremos el caso en el que el vector z no se observa. Dado que la variable no observada z no está correlacionada con el vector observado x , todavía podemos descubrir β . En este caso, el vector $z\theta$ actúa como el término de perturbación, que etiquetamos como ε en la relación mostrada en (2.8).

$$y = x\beta + \varepsilon \quad (2.8)$$

La expresión (2.8) representa un modelo lineal normal con perturbaciones independientes e idénticamente distribuidas (*iid*), donde se sabe que el estimador ordinario de mínimos cuadrados $\hat{\beta} = (x'x)^{-1}x'y$ es el mejor estimador lineal insesgado.

Como escenario alternativo, consideremos una situación en la que el vector explicativo z presenta una covarianza nula con el vector x , pero sigue el proceso autorregresivo espacial mostrado en (2.9).

$$z = \rho Wz + r \quad (2.9)$$

$$z = (I_n - \rho W)^{-1}r \quad (2.10)$$

En (2.9), ρ es un parámetro escalar real, r es un vector de perturbaciones distribuido $N(0, \sigma^2 I_n)$, y W es una matriz de pesos espaciales $n \times n$ con $W_{ij} > 0$ cuando la observación j es vecina de la observación i , y $W_{ij} = 0$ en caso contrario. También establecemos $W_{ii} = 0$, y suponemos que W tiene sumas de filas de la unidad y que $(I_n - \rho W)^{-1}$ existe. A partir de nuestra discusión de los procesos autorregresivos espaciales, cada elemento de Wz representaría una combinación lineal de elementos del vector z asociados con las ubicaciones vecinas. Cuando se trabaja con procesos espaciales de datos, parece intuitivamente plausible que factores latentes no observados, como las comodidades de la ubicación, la accesibilidad a las carreteras o el

prestigio del vecindario, presenten una dependencia espacial del tipo asignado al vector z .

Sustituyendo (2.10) en (2.7) se obtiene (2.11), que reflexiona el modelo lineal normalizado generalizado que contiene perturbaciones no esféricas. El efecto de θ es aumentar la varianza de r y en (2.12) define $u = \theta r$. Como es bien sabido, las estimaciones por mínimos cuadrados del parámetro β en (2.11) siguen siendo insesgadas, pero no efficientes.

$$y = x\beta + (I_n - \rho W)^{-1}(\theta r) \quad (2.11)$$

$$y = x\beta + (I_n - \rho W)^{-1}u \quad (2.12)$$

$$E(y) = x\beta \quad (2.13)$$

Dada la prevalencia de las variables omitidas en la práctica de la econometría espacial, parece poco probable que x y u no estén correlacionadas. Un enfoque sencillo para representar esta correlación es especificar que u depende linealmente de x , más un término de perturbación v que es independiente de x como en (2.14), donde el parámetro escalar γ y la varianza del término de perturbación v (σ^2) determinan la fuerza de la relación entre x y $z = (I_n - \rho W)^{-1}u$.

$$u = x\gamma + v \quad (2.14)$$

$$v \sim N(0, \sigma^2 I_n)$$

En este escenario, la DGP más complicada se muestra en (2.16).

$$y = x\beta + (I_n - \rho W)^{-1}(x\gamma + v) \quad (2.15)$$

$$y = x\beta + (I_n - \rho W)^{-1}x\gamma + (I_n - \rho W)^{-1}v \quad (2.16)$$

Ya no es el caso de que la estimación por mínimos cuadrados $\hat{\beta}$ sea insesgada. Si transformamos la expresión (2.16) para que tenga errores *iid*, vemos que esta situación da lugar a un modelo mostrado en (2.18) que Anselin (1988) etiquetó como modelo espacial de Durbin (MDE). Este modelo incluye un retardo espacial de la variable dependiente Wy , así como el vector de variables explicativas x , y un retardo espacial de la variable explicativa Wx .

$$(I_n - \rho W)y = (I_n - \rho W)x\beta + x\gamma + v \quad (2.17)$$

$$y = \rho Wy + x(\beta + \gamma) + Wx(-\rho\beta) + v \quad (2.18)$$

En el capítulo 3 se estudiará la relación entre las variables omitidas que presentan dependencia espacial y los modelos de regresión espacial implícitos resultantes. La magnitud del sesgo que surge en estos casos también se explorará con más detalle.

2.3 Una motivación de heterogeneidad espacial

La especificación de los modelos para que tengan un efecto individual, normalmente modelado como un intercepto de tipo distinto para cada individuo o unidad, se ha hecho más popular con la prevalencia de grandes conjuntos de *datos de panel*. Para darle alguna forma, dejemos que el vector $n \times 1$ a en (2.19) represente los interceptos individuales.

$$y = a + X\beta \quad (2.19)$$

Normalmente, los conjuntos de datos de panel incluyen múltiples observaciones para cada unidad, por lo que la estimación de un vector de parámetros como a es factible. En un contexto espacial en el que sólo tenemos una única observación para cada región, podemos tratar el vector a como un vector de efectos aleatorios estructurado espacialmente. Una forma de modelar la heterogeneidad espacial consiste en suponer que las unidades de observación cercanas deben presentar niveles de efectos similares a los de las unidades vecinas. Esto puede implementarse asignando el proceso espacial autoregresivo mostrado en (2.20) y (2.21) para gobernar el vector de intercepciones a . Por el momento, asumimos que a es independiente de X .

$$a = \rho W a + \varepsilon \quad (2.20)$$

$$a = (I_n - \rho W) \varepsilon \quad (2.21)$$

Dado que hemos introducido el parámetro escalar ρ y un parámetro escalar de varianza del ruido σ^2 junto con la información exógena de conectividad de la muestra contenida en la matriz W , podemos estimar de forma factible el vector $n \times 1$

de los parámetros a . Combinando (2.19) y (2.21) se obtiene la DGP del modelo de error espacial (SEM).

$$y = X\beta + (I_n - \rho W) \varepsilon \quad (2.22)$$

En consecuencia, la heterogeneidad espacial proporciona otra forma de motivar la dependencia espacial. En este caso, la dependencia puede considerarse como una dependencia del error.

¿Qué ocurre si a no es independiente de X ? Supongamos que ε en (2.21) se sustituye por $X\gamma + E$ para modelar las perturbaciones, donde hay una parte que está correlacionada con las variables explicativas y una parte que es ruido independiente. En este caso, a tiene la forma de (2.23), que conduce a las formas reducidas de a e y en (2.24) y (2.25) y al modelo empírico con perturbaciones *iid* en (2.26).

$$a = \rho W a + X\gamma + E \quad (2.23)$$

$$a = (I_n - \rho W)^{-1} X\gamma + (I_n - \rho W)^{-1} E \quad (2.24)$$

$$y = X\beta + (I_n - \rho W)^{-1} (X\gamma + E) \quad (2.25)$$

$$y = \rho W y + X(\beta + \gamma) + WX(-\rho\beta) + E \quad (2.26)$$

El modelo de (2.26) adopta la forma del MDS. Los modelos que implican parámetros effects espacialmente estructurados se analizan en el capítulo 8 en el contexto de los flujos de origen-destino y en el capítulo 10 el caso de los modelos de variable dependiente limitada.

2.4 Una motivación basada en las externalidades

En un contexto espacial, las externalidades (tanto positivas como negativas) derivadas de las características del vecindario suelen tener un impacto sensorial directo. Por ejemplo, los solares con basura proporcionan un hábitat para ratas y serpientes que pueden visitar los patios contiguos y reducir el valor de sus propiedades. Por otro lado, las viviendas rodeadas por aquellas con patios bellamente ajardinados que contienen plantas aromáticas tendrían un efecto positivo en los valores de la vivienda. En términos de modelización, la media espacial de las características de las viviendas vecinas (WX) podría desempeñar un papel directo en la determinación de los precios de la vivienda contenidos en el vector y , como se muestra en (2.27).

$$y = \alpha \mathbf{1}_n + X\beta_1 + WX\beta_2 + \varepsilon \quad (2.27)$$

Lo denominamos modelo de rezago espacial de X o SLX , ya que el modelo contiene rezagos espaciales (WX) de las características de las viviendas vecinas como variables explicativas.

2.5 Un modelo de motivación de la incertidumbre

En la práctica aplicada, a menudo nos enfrentamos a la incertidumbre sobre el tipo de modelo a emplear, así como a la incertidumbre sobre los parámetros convencionales y a la incertidumbre sobre la especificación de las variables explicativas adecuadas. Como ejemplo, supongamos que existe incertidumbre respecto al uso de la especificación del modelo autorregresivo (SAR) $y = \rho W y + X\beta + \varepsilon$. En particular, introducimos una especificación de modelo competitiva que implica la dependencia espacial en las distancias del modelo, $y = X\beta + u$, $u = \rho W u + \varepsilon$, a la que nos referimos como la

modelo de error espacial (SEM). Las DGP respectivas de estos dos modelos se muestran en (2.28) y (2.29). Se trata de DGPs y no de modelos de estimación, por lo que podríamos tener vectores de parámetros β idénticos y valores iguales para ρ en cada DGP.

$$y_a = (I_n - \rho W)^{-1} X \beta + (I_n - \rho W) \varepsilon^{-1} \quad (2.28)$$

$$y_b = X \beta + (I_n - \rho W) \varepsilon^{-1} \quad (2.29)$$

En el [capítulo 6](#) discutiremos los métodos bayesianos de comparación de modelos que pueden utilizarse para producir probabilidades de modelos posteriores. Supongamos que π_a y π_b representan las ponderaciones o probabilidades asociadas a los modelos autorregresivo y de error, y supongamos además que estos dos modelos representan los únicos modelos considerados de forma que: $\pi_a + \pi_b = 1$. Una solución bayesiana a la incertidumbre del modelo es confiar en *el promedio del modelo*, que implica hacer inferencias a partir de una combinación lineal de modelos. Las probabilidades posteriores del modelo se utilizan como pesos para producir estimaciones e inferencias basadas en los parámetros del modelo combinado o promediado.

Es interesante considerar la DGP asociada a una combinación lineal de los modelos SAR y SEM, mostrada en (2.30). Como muestran las manipulaciones, esto conduce al modelo SDM en (2.31).

$$y_c = \pi y_{aa} + \pi y_{bb} \quad (2.30)$$

$$y_c = R X^{-1} (\pi \beta_a) + X (\pi \beta_b) + (\pi_a + \pi_b) R \varepsilon^{-1} y_c$$

$$= R X^{-1} (\pi \beta_a) + X (\pi \beta_b) + R \varepsilon^{-1}$$

$$R y_c = X (\pi \beta_a) + R X (\pi \beta_b) + \varepsilon$$

$$R y_c = X \beta + W X (-\rho \pi \beta_b) + \varepsilon$$

$$R y_c = X \beta_1 + W X \beta_2 + \varepsilon$$

$$y_c = \rho W y_c + X \beta_1 + W X \beta_2 + \varepsilon \quad (2.31)$$

$$R = I_n - \rho W$$

Las combinaciones de otros modelos que introducimos en la siguiente sección pueden utilizarse para producir versiones más elaboradas del modelo SDM que contienen rezagos espaciales de orden superior que implican términos como $W^2 X$.

Esto sugiere que la incertidumbre respecto al carácter específico de la dependencia espacial en la DGP subyacente proporciona otra motivación para los modelos que implican rezagos espaciales de las variables dependientes y explicativas. En este ejemplo, tenemos incertidumbre respecto a la presencia de dependencia espacial en la variable dependiente frente a las

perturbaciones. Abordaremos el promediado bayesiano de modelos como solución a la incertidumbre del modelo con más detalle en el capítulo 6.

2.6 Modelos de regresión autorregresiva espacial

Como se señaló en el [capítulo 1](#), la estructura autorregresiva espacial puede combinarse con un modelo de regresión convencional para producir una extensión espacial del modelo de regresión lineal que hemos denominado modelo SAR. Este modelo se muestra en (2.32), con el *proceso de generación de datos* implícito en (2.33).

$$y = \rho W y + \alpha \iota_n + X\beta + \varepsilon \quad (2.32)$$

$$y = (I_n - \rho W)^{-1} (\alpha \iota_n + X\beta) + (I_n - \rho W)^{-1} \varepsilon \quad (2.33)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

En este modelo, los parámetros que hay que estimar son los parámetros de regresión habituales α , β , σ y el parámetro adicional ρ . Los rezagos espaciales son una característica de los modelos de regresión espacial, y pueden utilizarse para proporcionar versiones ampliadas del modelo SAR. Ya hemos visto una de estas extensiones, el Modelo Durbin espacial (MDE), que surgió de nuestra motivación de las variables omitidas. Este modelo incluye rezagos espaciales de las variables explicativas así como de la variable dependiente. Este modelo se muestra en (2.34) junto con su *proceso de generación de datos* asociado en (2.35).

$$y = \rho W y + \alpha \iota_n + X\beta + WX\gamma + \varepsilon \quad (2.34)$$

$$y = (I_n - \rho W)^{-1} (\alpha \iota_n + X\beta + WX\gamma + \varepsilon) \quad (2.35)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

También podemos utilizar rezagos espaciales para reflejar la dependencia en el proceso de perturbación, lo que conduce al modelo de error espacial (SEM), mostrado en (2.36).

$$y = \alpha \iota_n + X\beta + u \quad (2.36)$$

$$u = \rho W u + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Otro miembro de la familia de modelos de regresión espacial es el que denominamos SAC, que tiene la forma de (2.37), donde la matriz W puede ser igual a W_2 . Este modelo contiene dependencia espacial tanto en la variable dependiente como en las perturbaciones.

$$y = \alpha \iota_n + \rho W y_1 + X\beta + u \quad (2.37)$$

$$u = \theta W u_2 + \varepsilon$$

$$\sim N(0, \sigma^2 I_n)$$

$$y = (I_n - \rho W_1)^{-1} (X\beta + \alpha \iota_n) + (I_n - \rho W_1)^{-1} (I_n - \theta W_2)^{-1} \varepsilon \quad (2.38)$$

Observamos que se han propuesto modelos de regresión espacial que utilizan un proceso de media móvil en lugar del proceso autorregresivo espacial. Por ejemplo, se podría utilizar $u = (I_n - \theta W) \varepsilon$ para modelar las perturbaciones. Este tipo de proceso proporciona un método para capturar *los efectos locales* que surgen de los vecinos inmediatos, a diferencia del proceso autorregresivo que modela *los efectos globales* (Anselin, 2003).

La media móvil espacial (local) puede combinarse con un proceso autorregresivo espacial (global) para producir un modelo que Anselin y Bera (1998) denominan modelo de media móvil autorregresivo espacial, SARMA. Este modelo adopta la forma de (2.39), con la DGP mostrada en (2.40), donde, como en el caso del modelo SAC, la matriz W_1 podría ser igual a W_2 .

$$y = \alpha_1 + \rho W y + X\beta + u$$

$$u = (I_n - \theta W_2) \varepsilon \quad (2.39)$$

$$\varepsilon \sim N(0, \sigma^2 I_n)$$

$$y = (I_n - \rho W_1)^{-1} (X\beta + \alpha_1) + (I_n - \rho W_1)^{-1} (I_n - \theta W_2) \varepsilon \quad (2.40)$$

La distinción entre el modelo SAC y el SARMA radica en las diferencias entre las perturbaciones en sus respectivas DGP (2.38) y (2.40). El SAC utiliza $(I_n - \rho W_1)^{-1} (I_n - \theta W_2) \varepsilon$ mientras que SARMA utiliza $(I_n - \rho W_1)^{-1} (I_n - \theta W_2) \varepsilon$. Dada la representación en serie de la inversa en términos de potencias de la matriz, debería estar claro que el SAC dará más peso a las potencias más altas de W que SARMA. Sin embargo, ambos modelos tienen: $E(y) = (I - \rho W_1)^{-1} (X\beta + \alpha_1)$, que es lo mismo que $E(y)$ para el modelo SAR. Por lo tanto, estos modelos se concentran en un modelo más elaborado para las desviaciones, mientras que el SDM elabora el modelo para los spillovers.

Además, existen muchos otros modelos, como el exponencial matricial, el differencing fraccional y otras variantes de las especificaciones ARMA. Nos ocuparemos de algunas de ellas en el capítulo 9.

2.7 Interpretación de las estimaciones de los parámetros

Los modelos de regresión espacial explotan la complicada estructura de dependencia entre las observaciones que representan países, regiones, condados, etc. Por ello, las estimaciones de los parámetros contienen una gran cantidad de información sobre las relaciones entre las observaciones o regiones. Un cambio en una sola observación (región) asociado a una variable explicativa determinada afectará a la propia región (un impacto directo) y potencialmente afectará a todas las demás regiones de forma indirecta (un impacto indirecto). De hecho, la capacidad de los modelos de regresión espacial para captar estas interacciones representa un aspecto importante de los modelos econométricos espaciales señalados en Behrens y Thisse (2007).

Una de las virtudes de la econometría espacial es la capacidad de dar cabida a estrategias de modelización ampliadas que describen interacciones multirregionales. Sin embargo, este rico conjunto de información también aumenta la dificultad de interpretar las estimaciones resultantes. En la sección 2.7.1 describimos la teoría que subyace al análisis del impacto del cambio de las variables explicativas en la variable dependiente del modelo. Los enfoques computacionales para calcular las medidas de resumen de estos impactos son el tema de la Sección 2.7.2, con medidas de dispersión para estas estadísticas de resumen discutidas en la Sección 2.7.3. En la sección 2.7.4 se expone una partición de las medidas resumidas de impacto que permite examinar la tasa de descomposición del impacto en el espacio, y en la sección 2.7.5 se analiza un modelo de error que contiene tanto X como WX y que denominamos (SDEM) como medio simplificado de estimar los impactos directos e indirectos.

2.7.1 Impactos directos e indirectos en teoría

Los parámetros de regresión lineal tienen una interpretación directa como la derivada parcial de la variable dependiente con respecto a la variable explicativa. Esto se debe a la linealidad y a la supuesta independencia de los observadores.

vaciones en el modelo: $y = \sum_{r=1}^K \beta_r x_{ir} + \varepsilon_i$. Las derivadas parciales de y_i con respecto a x_{ir} tienen una forma sencilla: $\partial y_i / \partial x_{ir} = \beta_r$ para todo i, r ; y $\partial y_i / \partial x_{ijr} = 0$, para $j \neq i$ y todas las variables r .

Una forma de pensar en esto es que el conjunto de información de una observación i en la regresión consiste únicamente en variables exógenas o predeterminadas asociadas

con la observación i . Así, una regresión lineal específica: $E(y_i) = \sum_{r=1}^K \beta_r x_{ir}$, y adopta una visión restringida del conjunto de información en virtud del supuesto de independencia.

En los modelos que contienen rezagos espaciales de las variables explicativas o dependientes, la interpretación de los parámetros se vuelve más rica y complicada. Varios investigadores han señalado que los modelos que contienen rezagos espaciales de la variable dependiente requieren una interpretación especial de los parámetros (Anselin y LeGallo, 2006; Kelejian, Tavlas y Hondronyiannis, 2006; Kim, Phipps y Anselin, 2003; LeGallo, Ertur y Baumont, 2003).

En esencia, los modelos de regresión espacial amplían el conjunto de información para incluir información de las regiones/observaciones vecinas. Para ver el efecto de esto, considere el modelo SDM que hemos reescrito en (2.41).

$$(I_n - \rho W)y = X\beta + WX\theta + \alpha_n + \varepsilon$$

\mathbb{E}

$$y = \sum_{r=1}^K \beta_r(W)x_r + V(W)\alpha_n + V(W)\varepsilon \quad (2.41)$$

$r=1$

$$S_r(W) = V(W)(I_{Bnr} + W\theta_r)$$

$$V(W) = (I_n - \rho W)^{-1} = I_n + \rho W + \rho W^2 + \rho W^3 + \dots$$

Para ilustrar el papel de $S_r(W)$, consideremos la expansión de los datos gener-

de (2.41) como se muestra en (2.42) (Kim, Phipps y Anselin, 2003, c.f. ecuación(4)).

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{r=1}^k L \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \dots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & \dots & S_r(W)_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_r(W)_{n1} & S_r(W)_{n2} & \dots & S_r(W)_{nn} \end{pmatrix} \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{pmatrix} + V(W)\alpha_n + V(W)\varepsilon \quad (2.42)$$

El caso de una única observación de la variable dependiente en (2.43) hace más transparente el papel de la matriz $S_r(W)$. Utilizamos $S_r(W)_{ij}$ en esta ecuación para denotar el i ,ésimo elemento de la matriz $S_r(W)$, y $V(W)_i$ para indicar la i ª fila de $V(W)$.

$$y_i = \sum_{r=1}^k [S_r(W)_{i1}x_{1r} + S_r(W)_{i2}x_{2r} + \dots + S_r(W)_{in}x_{nr}] + V(W)_i\alpha_n + V(W)_i\varepsilon \quad (2.43)$$

De (2.43) se deduce que, a diferencia del caso del modelo de datos independientes, la derivada de y_i con respecto a x_{jr} es potencialmente distinta de cero, tomando un valor determinado por el i ,ésimo elemento de la matriz $S_r(W)$. También ocurre que la derivada de y_i con respecto a x_{ir} suele ser igual a β_r como en los mínimos cuadrados.

$$\frac{\partial y_i}{\partial x_{jr}} = S(W)_{rij} \quad (2.44)$$

Una implicación de esto es que un cambio en la variable explicativa para una sola región (observación) puede potencialmente afectar a la variable dependiente en todas las demás observaciones (regiones). Esto es, por supuesto, una consecuencia lógica de nuestro modelo SDM, ya que el modelo tiene en cuenta las variables dependientes y explicativas de otras regiones mediante la introducción de Wy y WX . Para un modelo en el que el vector de la variable dependiente y refleja los niveles de renta per cápita regional, y las variables explicativas son las características regionales (por ejemplo, el capital humano y físico, la estructura industrial, la densidad de población, etc.), la variación regional de los niveles de renta se modela para que dependa de los niveles de renta de las regiones vecinas capturados por el vector de retardo espacial Wy , así como de las características de las regiones vecinas representadas por WX .

La derivada propia para la i -ésima región mostrada en (2.45) da lugar a una expresión $S_r(W)_{ii}$ que mide el impacto sobre la variable dependiente observación i de un cambio en x_{ir} . Este impacto incluye el efecto de los *bucles de retroalimentación* donde la observación i afecta a la observación j y

la observación j también afecta a la observación i

así como caminos más largos que pueden ir de la observación i a la j a la k y de vuelta a la i .

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \quad (2.45)$$

Consideremos el término escalar $S_r(W)_{ii}$ a la luz de la matriz, $S_r(W) = (I_n - \rho W)^{-1}(\beta_{nr} + W\theta_r)$. Centrándonos en el término inverso y en la expansión en serie de esta inversa a partir de la expresión 1 del capítulo (1.12), las influencias de las regiones vecinas surgen como resultado de los impactos que pasan por las regiones vecinas y vuelven a la propia región. Para ver esto, observe que la matriz W de la expresión 1 del capítulo (1.15) refleja a los vecinos de segundo orden y contiene elementos no nulos en la diagonal. Estos surgen porque la región i se considera vecina de su vecino, de modo que los impactos que pasen por las regiones vecinas ejercerán una influencia de retroalimentación en la propia región i . La magnitud de este tipo de retroalimentación dependerá de: (1) la posición de las regiones en el espacio, (2) el grado de conectividad entre las regiones que se rige por la matriz de pesos W en el modelo, (3) el parámetro ρ que mide la fuerza de la dependencia espacial, y (4) los parámetros β y θ . Los elementos diagonales de la $n \times n$ matriz $S_r(W)$ contienen los impactos directos, y los elementos off-diagonales representan los impactos indirectos.

Hay algunas situaciones en las que los profesionales están interesados en los impactos derivados de los cambios en una sola región o en el impacto de los cambios en una sola región, que se reflexionarían en una columna o fila de la matriz $S_r(W)$, como motivaremos en breve. Por ejemplo, Kelejian, Tavlás y Hondronyannis (2006) examinan el impacto del contagio financiero que surge de un solo país en otros países del modelo, Anselin y LeGallo (2006) examinan la difusión de la contaminación atmosférica de fuentes puntuales, y LeGallo, Ertur y Baumont (2003) y Dall'erna y LeGallo (2007) examinan los impactos de los cambios en las variables explicativas (como los fondos estructurales de la Unión Europea) en regiones estratégicas sobre el crecimiento económico global.

Sin embargo, en general, dado que el impacto de los cambios en una variable explicativa varía en todas las regiones/observaciones, Pace y LeSage (2006) sugieren una medida resumida deseable de estos impactos variables. Un resumen escalar natural se basaría en la suma de los impactos totales sobre las filas (o columnas) de la matriz $S_r(W)$, y luego se tomaría un promedio sobre todas las regiones. Los autores etiquetan la media de las sumas de las filas de esta matriz como *Impacto total medio en una observación*, y se refieren a la media de las sumas de las columnas como *Impacto total medio de una observación*. Un promedio de la diagonal de la matriz $S_r(W)$ proporciona una medida resumida del *Impacto Directo Medio*. Por último, un resumen escalar del impacto *indirecto medio* es, por definición, la diferencia entre el impacto total medio y el impacto directo medio. Formalmente, las definiciones de estas medidas resumidas de impacto son:

1. *Impacto directo medio*. El impacto de los cambios en la i -ésima observación de

x_r , que denotamos x_{ir} , sobre y_i podría resumirse midiendo el

promedio de $S_r(W)_{ii}$, que es igual a $n^{-1}\text{tr}(S_r(W))$. Obsérvese que promediar sobre el impacto directo asociado a todas las observaciones i es similar en espíritu a las típicas interpretaciones del coefficient de regresión que representan la respuesta media de la variable dependiente a la independiente sobre la muestra de observaciones.

2. *Impacto total medio en una observación.* La suma de la i -ésima fila de $S_r(W)$ representaría el impacto total en la observación individual y_i resultante de cambiar la r -ésima variable explicativa en la misma cantidad en todas las n observaciones (por ejemplo, $x_r + \delta_{in}$ donde δ es el cambio escalar). Hay n de estas sumas dadas por el vector de columnas $c_r = S_r(W)_{in}$, por lo que una media de estos impactos totales es $nc^{-1}r$.
3. *Impacto total medio de una observación.* La suma de la j ª columna de $S_r(W)$ daría el impacto total sobre todos los y_i de cambiar la r ª variable explicativa por una cantidad en la j ª observación (por ejemplo, $x_{jr} + \delta$). Hay n de estas sumas dadas por el vector de filas $r_r = nS_r^T(W)$, por lo que una media de estos impactos totales es $nr^{-1}rn$.

Es fácil ver que los valores numéricos de las medidas de resumen para las dos formas de impactos totales medios expuestas en 2) y 3) son iguales, ya que $nc^{-1}r = nS_r^T(W)_{in}$, al igual que $nr^{-1}rn = nS_r^T(W)_{in}$. Sin embargo, estas dos medidas permiten puntos de vista interpretativos diferentes, a pesar de su igualdad numérica.

El punto de vista *de una observación expresado en 3)* se refiere a cómo los cambios en una sola observación j influyen en todas las observaciones. Por el contrario, el punto de vista de una observación expresado en 2) considera cómo los cambios en todas las observaciones influyen en una sola observación i . El promedio de todos los n impactos totales, ya sea tomando los enfoques *de una observación* o de *una observación*, conduce al mismo resultado numérico. Por tanto, el impacto total medio es la media de todas las derivadas de y_i con respecto a x_{jr} para cualquier i, j . El impacto directo medio es la media de todas las derivadas propias. En consecuencia, la media de todas las derivadas (impacto total medio) menos la derivada propia media (impacto directo medio) es igual a la derivada cruzada media (impacto indirecto medio).

La aplicación de Kelejian, Tavlas y Hondronyiannis (2006) examina el impacto del contagio financiero que surge de un solo país en otros países del modelo, tomando el punto de vista de la *observación* expresado en 3). Por otro lado, Dall'erba y LeGallo (2007) examinan el impacto de la modificación de las variables explicativas (como los fondos estructurales de la Unión Europea) que se aplican a todas las regiones en el crecimiento económico propio y global, un ejemplo más coherente con el punto de vista de la *observación*. A lo largo del texto ofreceremos otros ejemplos en nuestras ilustraciones aplicadas.

Hay que tener en cuenta que las medidas escalares de impacto resumen la forma en que estos cambios funcionarían a través del sistema de dependencia

simultánea en el tiempo para culminar en un nuevo equilibrio de estado estacionario.

Observamos que nuestras medidas de impacto para el modelo SAR pueden derivarse de (2.46).

$$\begin{aligned}(I_n - \rho W)y &= X\beta + \alpha_n + \varepsilon \\ y &= S^{-1}_r(W)x_r + V(W)\alpha_n + V(W)\varepsilon \quad (2.46) \\ S_r(W) &= V(W)IV_{nr} \\ V(W) &= (I_n - \rho W)^{-1} = I_n + \rho W + \rho W^2 + \rho W^3 + \dots\end{aligned}$$

La medida resumida de los impactos totales, $nunS^{-1}_r(W)\alpha_n$, para este modelo toma la forma simple de (2.47) para el W estocástico de filas.

$$\begin{aligned}nunS^{-1}_r(W)\alpha_n &= nun^{-1}(I_n - \rho W)\beta_{1-rn} \\ &= (1 - \rho)\beta_{1-r} \quad (2.47)\end{aligned}$$

En contraste con nuestro enfoque y nomenclatura, Abreu, de Groot y Florax (2004) consideran el modelo SAR (más simple) y la expresión en (2.48). Se refieren a β_r como un efecto directo, $W\rho\beta_r$ como un efecto indirecto, y el término entre paréntesis se etiqueta como efectos inducidos.

$$\frac{\partial y}{\partial x} = I\beta_{nr} + W\rho\beta_r + [W\rho\beta_r^2 + W\rho\beta_r^3 + \dots] \quad (2.48)$$

Bajo su etiquetado, como señalan correctamente, los efectos directos *no* corresponden a la derivada parcial de y_i con respecto a x_{ir} , y los efectos indirectos *no* corresponden a la derivada parcial de y_i con respecto a x_{jr} para $i = j$. En cambio, nuestras definiciones de efecto directo y efecto indirecto *sí corresponden* a las derivadas parciales propias y cruzadas, respectivamente. Un beneficio adicional de nuestro enfoque es que reducimos el número de etiquetas de tres (directa, indirecta e inducida) a sólo dos (directa e indirecta).

En relación con el modelo RAS, los impactos totales del modelo SDM derivados de cambios en X , presentan una gran heterogeneidad derivada de la presencia de la matriz adicional $W\theta_r$ en los efectos totales. En particular, esto permite que los efectos indirectos de un cambio en cada variable explicativa sean diferentes, a diferencia del caso del SAR, que tiene un multiplicador común y global para cada variable.

Para el caso del modelo SAC, los impactos totales adoptan la misma forma que en el modelo SAR, ya que el modelo autorregresivo espacial para las perturbaciones en este modelo no entran en juego al considerar la derivada parcial de y con respecto a los cambios en las variables explicativas X . Por supuesto, en la práctica aplicada, las estimaciones de impacto para el modelo RAS se basarían en las estimaciones del modelo RAS para los coeficientes ρ , β , mientras que las del SAC se basarían en las estimaciones del modelo SAC para estos parámetros, que probablemente serían diferentes. También se da el

caso de que el modelo SARMA introducido en la sección 2.6

tendría los mismos impactos totales que los modelos SAR y SAC, de nuevo con la advertencia de que las estimaciones basadas en el modelo SARMA para ρ, β se utilizarían para calcular estos impactos. En consecuencia, para conjuntos de datos más grandes, SAR, SAC y SARMA deberían producir impactos estimados muy similares (en ausencia de una especificación errónea que afecte a variables distintas de los términos de perturbación).

2.7.2 Cálculo de medidas de impacto resumidas

Formalmente define en (2.49) hasta (2.51), $\bar{N}(\bar{r})_{total}$, $\bar{N}(\bar{r})_{direct}$, y $\bar{N}(\bar{r})_{indirect}$, que representa el promedio de los impactos totales, el promedio de los impactos directos y el promedio de los impactos indirectos de los cambios en la variable del modelo X_r .

$$\bar{N}(\bar{r})_{direct} = n^{-1} \text{tr}(S_r(W)) \quad (2.49)$$

$$\bar{N}(\bar{r})_{total} = n n S^{-1} (W)_{ln} \quad (2.50)$$

$$\bar{N}(\bar{r})_{indirect} = \bar{N}(\bar{r})_{total} - \bar{N}(\bar{r})_{direct} \quad (2.51)$$

Es computacionalmente ineficiente calcular las estimaciones de impacto resumidas utilizando las definiciones anteriores, ya que esto requeriría la inversión de la matriz $n n (I_n \rho W)$ en $S_r(W)$. Proponemos una aproximación a la expansión infinita de $S_r(W)$ basada en las trazas de las potencias de W . Esto, por supuesto, requiere que la mayor potencia considerada en la aproximación sea lo suficientemente grande como para asegurar la convergencia aproximada. En el capítulo 4 se analiza una aproximación lineal en n de este tipo que ayuda a calcular las medidas escalares de resumen para los impactos directos, indirectos y totales.

2.7.3 Medidas de dispersión para las estimaciones de impacto

Para extraer conclusiones sobre la significación estadística de los impactos asociados a la modificación de las variables explicativas, necesitamos la distribución de nuestras medidas escalares de resumen para los distintos tipos de impacto. Se pueden utilizar enfoques de simulación computacionalmente eficientes para producir una distribución empírica de los parámetros $\alpha, \beta, \theta, \rho, \sigma$ que se necesitan para calcular las medidas escalares de resumen. Esta distribución puede construirse utilizando un gran número de parámetros simulados extraídos de la distribución normal multivariante de los parámetros implicados por las estimaciones de máxima verosimilitud.

Como alternativa, se pueden utilizar los métodos de estimación Bayesiana de Markov Chain Monte Carlo (MCMC) expuestos en LeSage (1997), discutidos en el capítulo, para producir estimaciones de dispersión para los impactos escalares. Dado que la estimación MCMC produce muestras (extracciones) de la distribución posterior de los parámetros del modelo, éstas pueden utilizarse en (2.49) y (2.50) para producir una distribución posterior de las medidas de impacto escalares. Como muestran Gelfand et al. (1990), MCMC puede producir una inferencia

válida sobre funciones no lineales de los parámetros, como los impactos directos e indirectos en (2.49) y (2.50). Todos los

lo que se requiere es la evaluación y el almacenamiento de las extracciones que reflejan las combinaciones no lineales de los parámetros. Las estimaciones posteriores de la dispersión se basan en simples cálculos de varianza aplicados a estas extracciones almacenadas, tema que se aborda en [el capítulo 5](#).

2.7.4 Repartición de los impactos por orden de vecindad

Debe quedar claro que los impactos derivados de un cambio en las variables explicativas influirán más en los vecinos de bajo orden que en los de orden superior. Se esperaría un patrón de disminución de la magnitud de los impactos a medida que se pasa de los vecinos de orden inferior a los de orden superior. En algunas aplicaciones, puede ser interesante el patrón particular de disminución de la influencia en los vecinos de distinto orden. Proporcionamos un ejemplo de esto en [el capítulo 3](#).

Dado que los impactos son una función de $S_r(W)$, éstos pueden expandirse como una combinación lineal de potencias de la matriz de pesos W utilizando la expansión en serie infinita de $(I_n - \rho W)^{-1}$. Aplicar esto a (2.49) y (2.50), donde utilizamos la definición de $S_r(W)$ para el modelo SAR, nos permite observar el impacto asociado a cada potencia de W . Estas potencias corresponden a las propias observaciones (de orden cero), a los vecinos inmediatos (de primer orden), a los vecinos de los vecinos (de segundo orden), etc.

$$S_r(W) \approx (I_n + \rho W + \rho W^2 + \rho W^3 + \dots + \rho W^{qq})\beta_r \quad (2.52)$$

A modo de ejemplo, [la tabla 2.1](#) muestra los impactos directos, indirectos y totales, tanto *acumulativos* como *marginales o divididos espacialmente*, asociados a los órdenes 0 a para9 el caso de un modelo de RAS en el que $\beta_r = 0.5$ y $\rho = 0.7$. En la tabla vemos un efecto directo acumulativo igual a 0,586, que dado el coeficiente de 0,5 indica que hay una retroalimentación igual a 0,086 que surge de cada región que impacta en los vecinos que a su vez impacta en los vecinos y así sucesivamente. En este caso, estos efectos de retroalimentación explican la diferencia entre el valor del coeficiente de $\beta_r = 0.5$ y el efecto directo acumulado de 0.586.

Los efectos acumulativos indirectos iguales a 1,0841 son casi el doble de la magnitud de los efectos acumulativos directos de 0.5860. Según *las estadísticas t* calculadas a partir de un conjunto de 5.000 valores de parámetros simulados, los tres efectos son significativamente diferentes de cero.

La partición espacial del efecto directo muestra que cuando llegamos a los vecinos de noveno orden hemos contabilizado 0,5834 del efecto directo acumulado de 0,5860. Cabe destacar el hecho de que para W no ⁰hay efecto indirecto, sino directo, y para W no ¹hay efecto directo, sino indirecto. Para ver esto, considera que cuando $q = W$ 0, ⁰ $= I_n$, y tenemos: $S_r(W) = I\beta_r = 0.5I_n$.

Cuando $q = 1$ tenemos un efecto indirecto, ya que hay elementos nulos en

la diagonal de la matriz W . Además, la naturaleza estocástica de filas de W conduce a una media de la suma de las filas que toma la forma: $\rho\beta_{\chi} = 0.70.5 = 0.35$, cuando $q = 1$.

TABLA 2.1: Distribución espacial de los impactos directos, indirectos y totales

Efectos acumulados			
MediaDev. t-estadística			
Directo efecto XO_r	.5860	0.014839.6106	
efecto indirecto $X1_r$.0841	0.058718.4745
Total effect $X1_r$.6700	0.073522.7302	
Efectos espacialmente divididos			
		W- orderTotalDirectIndirect	
W^0	0.5000	0.5000	0
W^1	0.3500	0	0.3500
W^2	0.2452	0.0407	0.2045
W^3	0.1718	0.0144	0.1574
W^4	0.1204	0.0114	0.1090
W^5	0.0844	0.0066	0.0778
W^6	0.0591	0.0044	0.0547
W^7	0.0415	0.0028	0.0386
W^8	0.0291	0.0019	0.0272
W^9	0.0204	0.0012	0.0191
${}^9_{q=0} W^q$	1.6220	0.5834	1.0386

Aunque los efectos indirectos acumulados que tienen magnitudes mayores que los efectos directos pueden parecer contraintuitivos, los impactos marginales o particionados dejan claro que los efectos indirectos individuales que recaen en regiones vecinas de primer orden, de segundo orden y de orden superior son menores que el efecto directo medio de 0.5 que recae en la "propia región". Sin embargo, la acumulación de estos efectos conduce a un efecto indirecto mayor que representa impactos más pequeños repartidos entre muchas regiones.

Vemos que los efectos directos desaparecen rápidamente a medida que nos desplazamos a vecinos de orden superior, mientras que los efectos indirectos o de desbordamiento espacial decaen más lentamente a medida que nos desplazamos a vecinos de orden superior.

2.7.5 Alternativas simplificadas a los cálculos de impacto

Los modelos de regresión espacial como el SEM que no implican rezagos espaciales de la variable dependiente producen estimaciones para los parámetros β que pueden interpretarse en el sentido habitual de la regresión como derivadas parciales: $\partial y / \partial x_{ir} = \beta_r$ para todas las i, r ; y $\partial y / \partial x_{ijr} = 0$, para $j = i$ y todas las variables r . Por supuesto, estos modelos no permiten que surjan impactos indirectos de los cambios en las variables explicativas, de forma similar a la situación de mínimos cuadrados en la que las observaciones de la variable dependiente se tratan como independientes.

Una alternativa al modelo SEM que denominamos *error de Durbin espacial*

(SDEM) incluye un retardo espacial de las variables explicativas WX , así como perturbaciones dependientes del espacio. Este modelo, que aumenta el modelo SEM con un retardo espacial de las variables explicativas se muestra en (2.53), con el modelo DGP en (2.54).

$$y = X\beta + WX\gamma + \iota\alpha_n + u \quad (2.53)$$

$$u = R\varepsilon^{-1}$$

$$y = X\beta + WX\gamma + \iota\alpha_n + R\varepsilon^{-1} \quad (2.54)$$

$$R = I_n - \rho W$$

$$\mathbf{I}_n$$

$$E(y) = S_r(W)x_r + \iota\alpha_n \quad (2.55)$$

$$r=1$$

$$S_r(W) = (I\beta_{nr} + W\gamma_r)$$

El modelo SDEM no permite una variable dependiente retardada separada effect, pero sí permite errores espacialmente dependientes y retardos espaciales de las variables explicativas. En relación con el MDE más general, simplifica la interpretación de los impactos, ya que los impactos directos están representados por los parámetros del modelo β y los impactos indirectos corresponden a γ . Esto también nos permite utilizar medidas de dispersión como la desviación estándar o *el estadístico t* para estos parámetros de regresión como base para la inferencia respecto a la significancia de los impactos directos e indirectos.

La versión del SDEM en (2.53) utiliza la misma matriz de pesos espaciales para los errores y las variables explicativas espacialmente retardadas, pero esto podría generalizarse para permitir pesos diferentes y no afectar la simplicidad de interpretar los impactos directos e indirectos como correspondientes a los parámetros del modelo.

El SDEM sustituye el multiplicador global del SDM por multiplicadores locales que simplifican la interpretación de las estimaciones del modelo. Sin embargo, observamos que el SDEM podría dar lugar a una subestimación de los impactos indirectos de orden superior (globales). El SDEM no anida el SDM y *viceversa*. Sin embargo, se puede concebir un MDE ampliado que anide el MDE.

2.8 Resumen del capítulo

Hemos proporcionado numerosas motivaciones de por qué pueden surgir relaciones de regresión espacial que incluyan rezagos espaciales del vector de la variable dependiente. Los modelos que contienen variables dependientes rezagadas espacialmente se han utilizado con mayor frecuencia en situaciones en las que existe una motivación intuitiva o teórica de que y dependerá de los valores vecinos de y . Por ejemplo, el precio hedónico de la vivienda

donde generalmente se piensa que los precios de las viviendas dependen de los precios de las viviendas vecinas vendidas recientemente. Esto se debe a que los tasadores/agentes inmobiliarios presumiblemente utilizan información sobre las viviendas vendidas recientemente para determinar el precio de venta. Otro ejemplo es la competencia entre gobiernos locales, donde parece intuitivo que los gobiernos locales puedan reaccionar a las acciones tomadas por los gobiernos locales cercanos.

Es posible que estos modelos tengan un papel más amplio de lo que se pensaba, ya que pudimos aportar tres motivaciones basadas en: variables omitidas, dependencia espacio-temporal e incertidumbre del modelo que dieron lugar a modelos que implican rezagos espaciales de la variable dependiente.

Este tipo de desarrollo tiene amplias implicaciones para la interfaz de la teoría económica y la econometría espacial. Sugiere que los modelos de econometría espacial pueden ser aplicables en muchas situaciones en las que no se han empleado anteriormente.