Capítulo 3

Estimación por máxima verosimilitud

Como se muestra en el capítulo 2, la estimación de modelos espaciales mediante mínimos cuadrados puede llevar a estimaciones inconsistentes de los parámetros de regresión para modelos con variables dependientes espacialmente retrasadas, a una estimación inconsistente de los parámetros espaciales y a una estimación inconsistente de los errores estándar. En cambio, la máxima probabilidad es consistente para estos modelos (Lee, 2004). En consecuencia, este capítulo se centra en la estimación de máxima verosimilitud de los modelos de regresión espacial. Históricamente, gran parte de la literatura sobre econometría espacial se ha centrado en formas de evitar la estimación de máxima verosimilitud debido a las dificultades computacionales percibidas. Desde el texto de Anselin de 1988 se han producido muchas mejoras en los métodos computacionales para la estimación de máxima verosimilitud de los modelos de regresión espacial. Estas mejoras permiten que los modelos que implican muestras que contienen más de 60.000 observaciones de tramos censales de EE.UU. se estimen en sólo unos segundos en ordenadores de sobremesa y portátiles.

El apartado 3.1 aborda la estimación por máxima verosimilitud de los modelos SAR, SDM, SEM y otros. La sección 3.1 proporciona una serie de técnicas que reducen en gran medida las anteriores dificultades computacionales que surgían en la estimación de estos modelos. La sección 3.2 se centra en la estimación por máxima verosimilitud de las estimaciones de varianza-covarianza de la dispersión para los parámetros del modelo necesarios para la inferencia. Proporcionamos un nuevo enfoque que puede utilizarse para reducir las tareas computacionales necesarias para construir las estimaciones de máxima verosimilitud de la dispersión necesarias para la inferencia.

Como ya se ha motivado, las variables omitidas son un problema probable en los trabajos aplicados con datos económicos regionales, y la sección 3.3 explora más a fondo el impacto empírico de la dependencia espacial en el sesgo de las variables omitidas en los modelos de regresión espacial. Presentamos expresiones teóricas para el sesgo junto con pruebas estadísticas y especificaciones del modelo que mitigan los problemas que plantean las variables omitidas que están correlacionadas con las variables explicativas incluidas.

El capítulo concluye con una aplicación en la sección 3.4 que ilustra muchas de las cuestiones tratadas en el capítulo. Nos basamos en un modelo sencillo que contiene una única variable explicativa utilizada para explicar las diferencias de productividad de los factores entre las regiones de la Unión Europea.

3.1 Estimación del modelo

El2.6. modelo espacial de Durbin (SDM) proporciona un punto de partida general para la discusión de la estimación del modelo de regresión espacial, ya que este modelo incluye el modelo de error espacial (SEM) y el modelo autorregresivo espacial (SAR).

En la sección 3.1.1 se analiza la estimación por máxima verosimilitud de los modelos SAR y SDM, cuyas funciones de verosimilitud coinciden. En la sección 3.1.2 nos centramos en la función de verosimilitud y el procedimiento de estimación del modelo SEM, y en la sección 3.1.3 se analizan los modelos que implican múltiples matrices de pesos. 3.1.3.

3.1.1 Estimación de modelos SAR y SDM

El modelo SDM se muestra en (3.1) junto con su *proceso de generación de datos* asociado en (3.2),

$$y = \rho Wy + \alpha\iota_n + X\beta + WX\theta + \varepsilon \quad (3.1)$$
$$y = (I_n - \rho W)^{-1}(\alpha\iota_n + X\beta + WX\theta + \varepsilon) \quad (3.2)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

where 0 represents an $n \times 1$ vector of zeros and ι_n represents an n 1 vector of ones associated with the constant term parameter α. This model can be written as a SAR model by defining: $Z = \iota_n\ X\ WX$ and $\delta = \alpha\ \beta\ \theta$, leading to (3.3). This means that the likelihood function for SAR and SDM models can be written in the same form where: = for the SAR model and $Z = \iota_n\ X\ WX$ for the SDM model. $\iota\ X$ n

$$y = \rho Wy + Z\delta + \varepsilon \quad (3.3)$$
$$y = (I_n - \rho W)^{-1}Z\delta + (I_n - \rho W)\varepsilon^{-1} \quad (3.4)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

A partir del enunciado del modelo (3.3), *si* se supiera que el verdadero valor del parámetro ρ es, digamos, ρ*, podríamos reordenar el enunciado del modelo en (3.3) como se muestra en (3.5).

$$y - \rho W y^* = Z\delta + \varepsilon \quad (3.5)$$

Esto sugiere una estimación para δ de $\hat{\delta} = (ZZ)^{-Z1}(I_n - \rho W^*)y$. En este caso también podríamos finir una estimación para el parámetro de varianza del ruido $\hat{\sigma}^2 =$
$^{-1}ne(\rho^*)\ e(\rho^*)$, donde $e(\rho^*) = y - \rho W y^* - Z\hat{\delta}$.

Estas ideas motivan que podamos concentrar la verosimilitud completa (log) con respecto a los parámetros β, σ^2 y reducir la máxima verosimilitud a un problema de optimización univariante en el parámetro ρ.

La maximización de la log-verosimilitud completa para el caso del modelo SAR implicaría establecer las primeras derivadas con respecto a los parámetros β, σ ^{2}y ρ iguales a cero y resolver simultáneamente estas condiciones de primer orden para todos los parámetros.

Por el contrario, se podrían encontrar estimaciones equivalentes de máxima verosimilitud utilizando la función log-verosimilitud concentrada con respecto a los parámetros β y
σ^2. Esto implica sustituir las *soluciones de forma cerrada de las* condiciones de primer orden para los parámetros β y σ^2 para obtener una probabilidad logarítmica que se dice que está *concentrada con* respecto a estos parámetros. Etiquetamos estas expresiones $\hat{\beta}(\rho)$, $\hat{\sigma}^2(\rho)$, y observamos que dependen de los datos de la muestra más el pa-
En el caso del modelo SAR, esto nos deja con una log-verosimilitud *concentrada* que sólo depende del único parámetro escalar ρ. La optimización de la función log-verosimilitud *concentrada* con respecto a ρ, para finir la estimación de máxima verosimilitud $\hat{\rho}$ nos permite utilizar esta estimación en la forma cerrada
expresiones para $\hat{\beta}(\hat{\rho})$ y $\hat{\sigma}^2(\hat{\rho})$ para producir estimaciones de máxima verosimilitud para
estos parámetros.

Trabajando con la log-verosimilitud concentrada se obtienen exactamente las mismas estimaciones de máxima verosimilitud $\hat{\beta}$, $\hat{\sigma}$ y $\hat{\rho}$ que surgirían de maximizar la
la log-verosimilitud completa (Davidson y MacKinnon, p1993,. 267-269). El motivo para optimizar la log-verosimilitud concentrada es que esto simplifica el problema de optimización al reducir un problema de optimización multivariante a un problema univariante. Otra ventaja de utilizar la log-verosimilitud concentrada es que se pueden utilizar ajustes sencillos a la salida del problema de optimización (que describimos más adelante) para producir una matriz de varianza-covarianza computacionalmente efficiente que utilizamos para inferir los parámetros. Estas inferencias son idénticas a las que se obtendrían al resolver el problema de optimización más engorroso que implica la log-verosimilitud completa. La función de log-verosimilitud para los modelos SDM (y SAR) tiene la forma de (3.6) (Anselin, p1988,. 63), donde ω es el n $\times$ 1 vector de valores propios de
la matriz W.

$$\ln L = -(n/2)\ln(\pi\sigma^2) + \ln|I_n - \rho W| - \sigma \frac{ee}{2^2} \tag{3.6}$$

$$e = y - \rho Wy - Z\delta$$

$$\rho \in (\min(\omega)^{-1}, \max(\omega)^{-1})$$

Si ω contiene sólo valores propios reales, se asegura una ma- triz de varianza-covarianza positiva definita por la condición: ρ $\in$ (min(ω$)^{-1}$, max(ω$)^{-1}$), como se

muestra en Ord (1975). La matriz W siempre puede construirse para tener un valor propio máximo de Por1. ejemplo, escalando la matriz de pesos por su valor propio máximo
como señalan Barry y Pace (1999); Kelejian y Prucha (2007). En este caso,

el intervalo para ρ se convierte en $(\min(\omega)^{-1}, 1)$ y un subconjunto de este ampliamente empleado en la práctica es $\rho \in [0, 1)$. En el capítulo 4 se ofrecen más detalles sobre los valores admisibles de ρ. Los valores admisibles pueden ser más complicados para las matrices de peso no simétricas W, ya que éstas pueden tener valores propios complejos.

Como se ha señalado, la log-verosimilitud puede concentrarse con respecto al vector de coefi- cientes δ y al parámetro de varianza del ruido σ^2. Pace y Barry (1997) sugirieron un enfoque conveniente para concentrar los parámetros δ y σ^2, que se muestra en (3.7). El término κ es una constante que no depende del parámetro ρ, e $|I_n - \rho W|$ es el determinante de esta matriz $n \times n$. Utilizamos la notación $e(\rho)$ para indicar que este vector depende de los valores tomados por el parámetro ρ, al igual que el valor escalar concentrado de la función de probabilidad logarítmica $\ln L(\rho)$.

$$\ln L(\rho) = \kappa + \ln |I_n - \rho W| - (n/2)\ln(S(\rho)) \quad (3.7)$$
$$S(\rho) = e(\rho)' e(\rho) = e_o'e_o - 2\rho e_o'e_d + \rho^2 e_d'e_d$$
$$e(\rho) = e_o - \rho e_d$$
$$e_o = y - Z\delta_o$$
$$e_d = Wy - Z\delta_d$$
$$\delta_o = (Z'Z)^{-1}Z'y$$
$$\delta_d = (Z'Z)^{-1}Z'Wy$$

Para simplificar la optimización de la log-verosimilitud con respecto al parámetro escalar ρ, Pace y Barry (1997) propusieron evaluar la log-verosimilitud utilizando un vector $q \times 1$ de valores para ρ en el intervalo $[\rho_{\min}, \rho_{\max}]$, etiquetado como $\rho_1, \ldots, \rho_q$ en (3.8).

$$\begin{pmatrix} \ln L(\rho_1) \\ \ln L(\rho_2) \\ \vdots \\ \ln L(\rho_q) \end{pmatrix} = \kappa + \begin{pmatrix} \ln |I_n - \rho_1 W| \\ \ln |I_n - \rho_2 W| \\ \vdots \\ \ln |I_n - \rho_q W| \end{pmatrix} - (n/2) \begin{pmatrix} \ln(S(\rho_1)) \\ \ln(S(\rho_2)) \\ \vdots \\ \ln(S(\rho_q)) \end{pmatrix} \quad (3.8)$$

En el capítulo 4 se discuten una serie de enfoques para calcular efficientemente el término $\ln |I_n - \rho W_i|$ sobre un vector de valores para el parámetro ρ. En nuestra discusión aquí, simplemente asumimos que estos valores están disponibles durante la optimización de la log-verosimilitud. Dada una cuadrícula sufficientemente fine de *q valores* para la log-verosimilitud, la interpolación puede suministrar puntos intermedios con cualquier precisión deseada (lo que se deduce de la suavidad de la función log-verosimilitud). Obsérvese que los momentos escalares $e_o'e_o$, $e_d'e_o$ y $e_d'e_d$ y los vectores $k \times 1$ δ_o, δ_d se calculan antes de la optimización, por lo que dado un valor de ρ, el cálculo de $S(\rho)$ simplemente requiere ponderar tres números. Dado el valor óptimo de ρ, éste se convierte en la estimación de máxima verosimilitud de ρ denotada como $\hat{\rho}$. Por lo tanto, se requiere muy poco cálculo para llegar al vector de valores concentrados de log-verosimilitud.

Dada la estimación de máxima verosimilitud $\hat{\rho}$, (3. 9), (3.10) y (3.11) muestran las estimaciones de máxima verosimilitud para los coefficientes $\hat{\delta}$, el parámetro de varianza del ruido $\hat{\sigma}^2$ y la matriz de varianza-covarianza asociada para las perturbaciones.

$$\hat{\delta} = \delta_o - \hat{\rho}\delta_d \quad (3.9)$$

$$\hat{\sigma} = nS^{-1}(\hat{\rho}) \quad (3.10)$$

$$\hat{\Omega} = \hat{\sigma}^2 [(I_n - \hat{\rho}W)(I_n - \hat{\rho}W)]^{-1} \quad (3.11)$$

Aunque el enfoque vectorial funciona bien, en el capítulo 4 se discute una técnica alternativa de solución de forma cerrada para ρ. Sin embargo, preferimos discutir el enfoque vectorial aquí debido a su simplicidad.

La función de verosimilitud combina un término de suma de errores al cuadrado transformado con el término de determinante logarítmico que actúa como una función de penalización que impide que la estimación de máxima verosimilitud de ρ sea igual a una estimación basada únicamente en la suma de errores al cuadrado minimizada (transformada), $S(\rho)$. El enfoque vectorial ofrece la ventaja adicional de garantizar un óptimo global en lugar de local.

La estimación de máxima verosimilitud puede realizarse mediante diversas técnicas de optimización univariante. Éstas podrían incluir el enfoque vectorizado que acabamos de discutir, basado en una cuadrícula fine de valores de ρ (q grande), métodos de búsqueda no derivados, como el esquema de búsqueda simplex o bisección de Nelder-Mead, o aplicando una técnica de optimización basada en derivadas (Press et al., 1996). Alguna forma del método de Newton con derivadas numéricas tiene la ventaja de proporcionar el óptimo, así como la segunda derivada de la logverosimilitud concentrada en el óptimo $\hat{\rho}$. Esta estimación numérica de la segunda derivada, junto con otra información, puede ser útil para producir una estimación numérica de la matriz de varianza-covarianza de los parámetros. Discutimos este tema con más detalle en la sección 3.2.

Como se ha mostrado anteriormente, una barrera aparente para implementar estos modelos para n grandes es la matriz Wn. Si W contiene todos los elementos distintos de cero, se necesitarían enormes cantidades de memoria para almacenar esta matriz para problemas que implican grandes muestras, como los tramos del censo de EE.UU., donde n >60 , Afortunadamente000., W suele ser *escasa*, lo que significa que contiene una gran proporción de ceros. Por ejemplo, si se confía en las regiones contiguas o en algún número m de regiones vecinas más cercanas para formar W, la matriz de pesos espaciales sólo contendrá mn no ceros, frente a los n^2 no ceros de una matriz *densa*. La proporción de no ceros
se convierte en m/n que disminuye con n. Las matrices de pesos de contigüidad tienen una media
de seis vecinos por fila (para conjuntos de puntos espacialmente aleatorios en un plano). A modo de ejemplo, utilizando los 3.111 condados de EE.UU. que representan los 48 estados más el distrito de Columbia, hay elementos9,678,321 en el 3, 1113, 111
matriz W, pero sólo 3, 6 111= 18, 666 serían distintos de cero, es decir, el 0,1929 por ciento de las entradas. Además, el cálculo de productos matriciales-

vectoriales como Wy y WX requiere mucho menos tiempo para las matrices dispersas. En ambos casos, las matrices dispersas requieren un tiempo lineal en n operaciones ($O(n)$) mientras que una W densa requeriría un tiempo cuadrático en n operaciones ($O(n^2)$). Como se muestra en el capítulo 4, las técnicas de matrices dispersas aceleran en gran medida el cálculo del logaritmo-determinante y otras cantidades de interés.

En resumen, varias técnicas facilitan el cálculo de las estimaciones de máxima verosimilitud para los modelos SDM y SAR. Estas técnicas incluyen la concentración de la log-verosimilitud, el cálculo previo de una tabla de log-determinantes
así como momentos como *el eoe*$_d$, y el uso de W disperso . En conjunto, estas técnicas reducen en gran medida el número de operaciones y la memoria del ordenador necesarios para resolver problemas que implican grandes muestras de datos. El capítulo 4 ofrece más detalles sobre estas y otras técnicas que pueden ayudar a calcular
estimaciones de máxima verosimilitud.

3.1.2 Estimación del modelo SEM

El enunciado del modelo para un modelo que contiene dependencia espacial en las perturbaciones que etiquetamos como SEM se muestra en (3.12), con la DGP para este modelo en (3.13), donde definimos que X es la matriz de $n \times k$ variables explicativas que puede o no incluir un término constante, y β el vector1 asociado de $k \times$ parámetros.

$$y = X\beta + u \quad (3.12)$$
$$u = \lambda W u + \varepsilon$$
$$y = X\beta + (I_n - \lambda W)\varepsilon^{-1} (3.13)$$
$$\varepsilon \sim N(0, \sigma I^2_n)$$ La

log-verosimilitud completa tiene la forma de

(3.14).

$$\ln L = -(n/2)\ln(\pi\sigma^2) + \ln|I_n - \lambda W| - \sigma \frac{ee}{2^2} \quad (3.14)$$
$$e = (I_n - \lambda W)(y - X\beta)$$

Para una λ dada, la optimización de la función log-verosimilitud muestra (Ord, 1975; Anselin, 1988) que β (λ) = (X(λ) X(λ)$)^{-X1}$(λ) y(λ), donde X(λ) = (X
$\lambda W X$), y(λ) = ($y\lambda Wy$), y σ^2(λ) = e(λ) e(λ)/n donde e(λ) = y(λ) – X(λ)β(λ). Por tanto, podemos concentrar la log-verosimilitud con respecto a β y σ^2 para obtener la log-verosimilitud concentrada en función de λ en (3.15).

$$\ln L(\lambda) = \kappa + \ln|I_n - \lambda W| - (n/2)\ln(S(\lambda)) \quad (3.15)$$
$$S(\lambda) = e(\lambda)\, e(\lambda) \quad (3.16)$$

A diferencia del caso SAR o SDM, S(λ) no es una simple cuadrática en el parámetro espacial. Como se indica actualmente en (3.16), la evaluación de la log-verosimilitud concentrada para cualquier valor dado de λ requiere la manipulación de $n \times$ y1 $n \times$

k para cada elección de λ. Esto se vuelve tedioso para grandes conjuntos de datos, técnicas de optimización que requieren muchos valores de prueba de λ, y en simulaciones. Sin embargo, las variables que requieren $O(n)$ cálculos pueden ser precalculadas de manera que el cálculo de $S(\lambda)$ durante la optimización sólo requiere trabajar con matrices de momentos de dimensión k por k o más pequeñas. Estas matrices de momentos implican las variables independientes y dependientes en función de λ.[1]

$$A_{XX}(\lambda) = XX - \lambda XWX - \lambda XWX + \lambda XWWX^2 \quad A_{Xy}(\lambda)$$
$$= Xy - \lambda XWy - \lambda XWy + \lambda XW^2 \quad Wy \quad A_{yy}(\lambda) = yy -$$
$$\lambda yWy - \lambda yWy + \lambda yWWy^2$$
$$\beta(\lambda) = A_{XX}(\lambda)^{-1} A_{Xy}(\lambda)$$
$$S(\lambda) = A_{yy}(\lambda) - \beta(\lambda) A_{XX}(\lambda)\beta(\lambda)$$

Con estos momentos y una rejilla precalculada de determinantes logarítmicos (junto con una rutina de interpolación) la actualización de la log-verosimilitud concentrada en (3.15) para un nuevo valor de λ es casi instantánea. La aplicación de una técnica de optimización univariante como el método de Newton a (3.15) para finir $\hat{\lambda}$ y subtitularlo en $\sigma^2(\lambda)$, $\beta(\lambda)$ y $\Omega(\lambda)$ conduce a las estimaciones de máxima verosimilitud (3.17) a (3.19).

$$\hat{\beta} = \beta(\hat{\lambda}) \tag{3.17}$$
$$\hat{\sigma}^2 = n S^{-1}(\hat{\lambda}) \tag{3.18}$$
$$\hat{\Omega} = \hat{\sigma}^2 (I_n - \hat{\lambda} W)(I_n - \hat{\lambda} W)^{1-1} \tag{3.19}$$

Como se indicó en la sección 3.1.1, la aplicación del método de Newton con derivadas numéricas para encontrar el óptimo produce una estimación numérica de la segunda derivada.
tiva de la log-verosimilitud concentrada en el óptimo $\hat{\lambda}$. Esta estimación numérica de la segunda derivada puede utilizarse junto con otros datos.
mación para producir una estimación de la matriz de varianza-covarianza.

Obsérvese que el modelo SDM anida el modelo SEM como un caso especial. Para ver esto, considere el enunciado alternativo del modelo SEM en (3.20). Para evitar problemas de colinealidad para *la* fila-estocástica W , suponemos que la matriz X no contiene un término constante y lo especificamos por separado. Esto es necesario para evitar la creación de un vector columna $W\iota_n = \iota_n$ en WX que duplicaría el término de intercepción.

[1] El uso de las matrices de momentos requiere que evitemos conjuntos de variables explicativas que estén mal escalados o mal condicionados. En la práctica, esto puede no ser una restricción tremenda, ya que incluso las técnicas computacionales numéricamente robustas pueden verse afectadas por conjuntos de datos mal condicionados. Además, los conjuntos de variables explicativas mal escalados suelen dar lugar a estimaciones de parámetros difíciles de interpretar.

$$y = \alpha\iota_n + X\beta + (I_n - \lambda W)\varepsilon^{-1}$$
$$(I_n - \lambda W)y = \alpha(I_n - \lambda W)\iota_n + (I_n - \lambda W)X\beta + \varepsilon$$
$$y = \lambda W y + \alpha(I_n - \lambda W)\iota_n + X\beta + WX(-\beta\lambda) + \varepsilon \quad (3.20)$$

El modelo de (3.20) representa un modelo SDM en el que el parámetro del retardo espacial de las variables explicativas (WX) se ha restringido para que sea igual a $-\beta\lambda$. La estimación del modelo SDM más general ($y = \lambda W y + X\beta + WX\theta + \varepsilon$) y la comprobación de la restricción $\theta = -\beta\lambda$ podrían conducir al rechazo del SEM relativo al SDM.

3.1.3 Estimaciones para modelos con dos matrices de pesos

La literatura espacial contiene una serie de modelos que implican dos o más matrices de pesos. El uso de múltiples matrices de pesos proporciona una generalización directa de los modelos SAR, SDM y SEM. Por ejemplo, Lacombe (2004) utiliza un modelo SAR de dos matrices de pesos similar al modelo SDM mostrado en (3.21).

$$y = \rho_1 W_1 y + \rho_2 W_2 y + X\beta + W_1 X\gamma_1 + W_2 X\theta_2 + \varepsilon \quad (3.21)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

La matriz de pesos W_1 se utilizó para captar el effecto de los condados vecinos dentro del estado, y W_2 capta el effecto de los condados vecinos en el estado fronterizo. Lacombe (2004) analizó políticas que variaban entre estados, lo que hace que este modelo sea atractivo. Para una muestra de condados situados en la frontera del estado, Lacombe (2004) analizó políticas que variaban entre los estados, lo que hace que este modelo sea atractivo.

La dependencia espacial se extiende tanto a los condados del estado como a los del otro lado de la frontera, en el estado vecino. Esta variante SDM del modelo de Lacombe permite que las influencias de la matriz de variables explicativas X *provengan por* separado de los vecinos del estado y de los del estado vecino.

La única desviación de nuestra discusión de la estimación de máxima verosimilitud para esta variante del modelo SDM implica un problema de optimización bivariante sobre el rango de valores factibles para ρ_1, ρ_2. La maximización de la log-verosimilitud (concentrada para $\beta, \gamma, \theta, \sigma^2$) para esta variante del modelo SDM requiere calcu-

lizando el término logarítmico determinante: $\ln|I_n - \rho_1 W_1 - \rho_2 W_2|$ sobre una retícula bivariada de valores para ambos ρ_1, ρ_2 en el rango factible. Estos valores escalares asociados

con la cuadrícula bivariada se almacenaría en una matriz en lugar de un vector.

La optimización de la función concentrada de verosimilitud logarítmica sobre los parámetros ρ_1, ρ_2 podría acceder repetidamente a esta matriz con un coste computacional muy pequeño. Como otro ejemplo de especificaciones que implican dos matrices de pesos, el modelo SAC contiene dependencia espacial tanto en la variable dependiente como en las perturbaciones, como se muestra en (3.22), junto con sus datos asociados que generan en (3.23). A diferencia del modelo de Lacombe, es posible implementar este modelo utilizando la misma matriz $W = W_1 = W_2$, pero tendremos más que decir sobre esto más adelante.

$$y = \rho W y_1 + X\beta + u$$
$$u = \lambda W u_2 + \varepsilon (3.22)$$
$$y = (I_n - \rho W_1)^{-1} X\beta + (I_n - \rho W_1)^{-1} (I_n - \lambda W_2) \varepsilon^{-1} (3.23)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

Las matrices W_1, W_2 pueden ser iguales o distintas. Obviamente, si el parámetro $\rho = 0$, este modelo colapsa al modelo SEM, y $\lambda = 0$ produce el modelo SAC. Normalmente, el SAC no contiene un término WX separado, por lo que el
El SAC no suele anidar el SDM. Sin embargo, se puede escribir un SDM extendido que anide el SAC, especificamente:

$$y = \rho W y + X\beta + WX\theta + u$$
$$u = \lambda W u + \varepsilon$$

La log-verosimilitud para el modelo SAC se muestra en (3.24) junto con las definiciones.

$$\ln L = -(n/2) \ln(\pi\sigma^2) + \ln|A| + \ln|B| - \frac{e'e}{2\sigma^2} \quad (3.24)$$
$$e = B(Ay - X\beta)$$
$$A = I_n - \rho W_1$$
$$B = I_n - \lambda W_2$$

La log-verosimilitud en (3.24) para el modelo SAC también puede concentrarse con respecto a los parámetros β, σ^2. La maximización de esta verosimilitud requiere el cálculo de dos log-determinantes para el caso en que $W_1 = W_2$, y la resolución de un problema de optimización bivariante en los dos parámetros ρ y λ.

Anselin (1988) planteó dudas sobre la identificación del modelo SAC en el caso de matrices idénticas W, pero Kelejian y Prucha (2007) proporcionan un argumento de que el modelo está identificado para este caso. Su argumento para la identificación requiere que $X\beta$ en la DGP haga una contribución material para explicar la variación en la variable dependiente y ($\beta = 0$). Para ver la importancia de esto, considere (3.25), y observe que en el caso en que $\beta = 0$, existe un problema de cambio de etiqueta ya que $AB = BA$ cuando A y B son funciones de la misma matriz de pesos W. Por tanto, los parámetros ρ y λ no están identificados.

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1} (I_n - \lambda W) \varepsilon^{-1} (3.25)$$

Aunque, 0en principio, β = identificará el modelo, a medida que la varianza del ruido de las perturbaciones aumenta, la importancia relativa de β disminuye. Esto se muestra en (3.26), donde las variables están todas escaladas por σ. Esto sugiere que en los problemas de baja relación señal-ruido (baja variación de los valores predichos en relación con la varianza del ruido), las estimaciones pueden mostrar síntomas de esta casi falta de identificación.

$$\sigma y^{-1} = ^{-1} AX(\sigma\beta^{-1}) + {}^{-1}AB\sigma\varepsilon^{-1-1} \quad (3.26)$$

También está el modelo SARMA mostrado en (3.27), con la correspondiente DGP en (3.28).

$$y = \rho W y_1 + X\beta + u$$
$$u = (I_n - \theta W_2)\varepsilon \quad (3.27)$$
$$y = (I_n - \rho W_1)^{-1}X\beta + (I_n - \rho W_1)^{-1}(I_n - \theta W_2)\varepsilon \quad (3.28)$$

Se requerirían cambios menores en la función de log-verosimilitud para este modelo, como se muestra en (3.29), donde hemos sustituido la definición $B = (I_n - \lambda W_2)$ del modelo SAC por $B = (I_n - \theta W_2)^{-1}$.

$$\ln L = \kappa + \ln|A| + \ln|B| - \frac{e'e}{2\sigma^2} \quad (3.29)$$
$$e = B(Ay - X\beta)$$
$$A = I_n - \rho W_1$$
$$B = (I_n - \theta W_2)^{-1}$$

Por último, en la literatura se han propuesto muchos otros modelos que implican múltiples matrices de pesos o combinaciones de potencias de matrices de pesos, como los modelos AR, MA y ARMA espaciales de orden superior (Huang y Anh, 1992). En el capítulo se 4discuten los enfoques para calcular los determinantes que surgen en dichos modelos.

3.2 Estimaciones de la dispersión de los parámetros

Hasta ahora, los procedimientos de estimación expuestos pueden utilizarse para producir estimaciones de los parámetros de dependencia espacial ρ y λ utilizando la maximización univariante o bivariante de la función de verosimilitud logarítmica concentrada con respecto a β y σ^2. Las estimaciones de máxima verosimilitud para los parámetros β y σ^2 pueden recuperarse utilizando las estimaciones de máxima verosimilitud para los parámetros de dependencia $\hat{\rho}$ y $\hat{\lambda}$.

Para muchos propósitos, existe la necesidad de realizar una inferencia. La inferencia de máxima verosimilitud a menudo se realiza mediante pruebas de *razón de verosimilitud* (LR), *multiplicador de Lagrange* (LM) o *Wald* (W). Asintóticamente, todos ellos deberían producir resultados similares, aunque pueden diferir para muestras finitas. A menudo, la elección de un método sobre el otro se reduce a la conveniencia computacional y otras preferencias.

Debido a la capacidad de calcular rápidamente las probabilidades, Pace y Barry (1997) proponen pruebas de razón de verosimilitud para hipótesis como la eliminación de una sola variable explicativa. Para poner estas pruebas de razón de verosimilitud en una forma similar a las pruebas *t*, Pace y LeSage (2003a) discuten el uso de estadísticas de desviación de raíz con signo.[2] La desviación de la raíz con signo aplica el signo de las estimaciones del coefficient β a la raíz cuadrada del estadístico de desviación (Chen y Jennrich, 1996). Estos estadísticos se comportan de forma similar a los coeficientes *t* cuando la muestra es grande, y pueden utilizarse en lugar de los estadísticos *t* para las pruebas de hipótesis.

La inferencia de Wald utiliza el hessiano (numérico o analítico) o la matriz de formación relacionada para proporcionar una matriz de varianza-covarianza para los parámetros estimados y, por tanto, la conocida *prueba t*. En este caso, el hessiano es simplemente la matriz de segundas derivadas de la log-verosimilitud con respecto a los parámetros. Los enfoques que utilizan el hessiano (Anselin, 1988, p. 76) o la matriz de información (Ord, 1975; Smirnov, 2005) han aparecido en la literatura de la econometría espacial.

Un problema de aplicación es la construcción de la matriz hessiana o de información. Utilizaremos el modelo SAR: $y = \rho Wy + X\beta + \varepsilon$ para simplificar nuestra discusión. La evaluación directa de la matriz analítica hessiana o de información implica el cálculo de un término de rastreo que contiene la matriz densa n n inversa $(I_n \ \rho W)^{-1}$. El capítulo proporciona4 medios para aproximar rápidamente los elementos
que surgen en la matriz hessiana o de información. En la siguiente discusión nos centramos en el hessiano.

Dada la capacidad de evaluar rápidamente la función log-verosimilitud, un enfoque puramente numérico podría parecer factible para calcular una estimación del hessiano. La aplicación de este enfoque en el software de uso general presenta algunos inconvenientes. En primer lugar, los profesionales suelen trabajar con datos de muestra mal escalados, lo que hace que las perturbaciones numéricas utilizadas para aproximar las derivadas que componen el hessiano sean difíciles. Un segundo punto es que la optimización univariante tiene lugar utilizando la verosimilitud concentrada con respecto a la
parámetros β y σ^2, por lo que una aproximación numérica al hessiano completo no surge de forma natural, como en los procedimientos típicos de estimación de máxima verosimilitud. Esto significa que hay que dedicar tiempo computacional después de la estimación de los parámetros para producir una estimación numérica separada del hessiano completo.

En la sección 3.2.1 se discuten las formas de combinar los resultados del hessiano analítico y del hessiano numérico para aprovechar los puntos fuertes de cada enfoque.

[2] La desviación es menos dos veces el logaritmo de la razón de verosimilitud para los modelos ajustados por máxima verosimilitud. La relación utilizada en estos cálculos es la que implica la probabilidad del modelo que excluye cada variable frente a la del modelo que contiene todas las variables.

En concreto, la mayoría de los elementos del hessiano analítico no requieren mucho tiempo de cálculo y son menos sensibles a los problemas de escala. Sin embargo, una aproximación numérica requiere menos tiempo y rinde bien para el único elemento difficultos del hessiano analítico.

En el capítulo se explican los métodos de estimación bayesiana5 de Markov Chain Monte Carlo (MCMC) para los modelos de regresión espacial, que pueden utilizarse para producir estimaciones de dispersión basadas en la muestra de extracciones realizadas por este enfoque de estimación basado en el muestreo. Siguiendo la teoría de regresión bayesiana estándar, el uso de un previo no informativo en estos modelos debería dar lugar a estimaciones posteriores e inferencias idénticas a las de máxima verosimilitud. Por lo tanto, estas estimaciones de la dispersión de los parámetros también proporcionan un medio válido, aunque poco ortodoxo, de llevar a cabo la inferencia de máxima verosimilitud.

Además, para un n grande, a menudo es factible proporcionar una inferencia limitada. Por ejemplo, Pace y LeSage (2003a) introducen un límite inferior en la prueba de razón de verosimilitudes que permite una inferencia conservadora de máxima verosimilitud al tiempo que evita la tarea computacionalmente exigente de calcular incluso estimaciones puntuales de máxima verosimilitud exacta. Muestran que esta forma de *inferencia de dominio de la probabilidad* (Pollack y Wales, 1991) funcionó casi tan bien como la inferencia exacta de la probabilidad en los parámetros de un modelo de RAS que incluía 890.091 observaciones, donde el procedimiento tardó menos de un minuto en calcularse.

Un enfoque totalmente diferente al problema de la inferencia en los modelos de regresión espacial es confiar en un método de estimación que no esté basado en la verosimilitud. Algunos ejemplos son el enfoque de variables instrumentales de Anselin (1988, p. 81-90), el estimador de variables instrumentales/momentos generalizados de Kelejian y Prucha (1998, 1999) o la máxima entropía de Marsh y Mittelham- mer (2004). Gran parte de la motivación para utilizar estos métodos proviene de las dificultades percibidas para calcular las estimaciones de los métodos basados en la verosimilitud, un problema que se ha resuelto en gran medida. Una característica de los métodos basados en la verosimilitud es que el término determinante garantiza que las estimaciones de los parámetros de dependencia resultantes se encuentran en el intervalo definido por los valores propios máximos y mínimos de la matriz de pesos. Algunos de los métodos de estimación alternativos que evitan el uso del determinante logarítmico pueden no producir estimaciones de los parámetros de dependencia en este intervalo. Además, estos métodos pueden ser sensibles de forma no evidente a diversas cuestiones de aplicación, como la interacción entre la elección de los instrumentos y la especificación del modelo. Por estas razones, nos centramos en las técnicas basadas en la verosimilitud.

3.2.1 Un cálculo hessiano mixto analítico-numérico

Para el caso del modelo SAR, el hessiano con el que trabajaremos se organiza como en (3.30), que etiquetamos como H. Para el caso del modelo SEM, sustituiríamos el parámetro ρ por λ. Por supuesto, algunas de las expresiones de las derivadas también cambian.

$$H = \begin{bmatrix} \frac{\partial L^2}{\partial \rho^2} & \frac{\partial L^2}{\partial \rho \partial \beta} & \frac{\partial L^2}{\partial \rho \partial \sigma^2} \\ \frac{\partial L^2}{\partial \beta \partial \rho} & \frac{\partial L^2}{\partial \beta \partial \beta} & \frac{\partial L^2}{\partial \beta \partial \sigma^2} \\ \frac{\partial L^2}{\partial \sigma^2 \partial \rho} & \frac{\partial L^2}{\partial \sigma^2 \partial \beta} & \frac{\partial L^2}{\partial (\sigma^2)^2} \end{bmatrix} \tag{3.30}$$

El hessiano analítico que denominamos $H^{(a)}$ aparece en (3.31), donde empleamos las definiciones $A = (I_n - \rho W)^{-1}$, $B = y\,(W + W)\,y$, $C = yWWy$.

$$H^{(a)} = \begin{bmatrix} -\operatorname{tr}(WAWA) - \frac{C}{\sigma^2} & -\frac{yWX}{\sigma^2} & -\frac{2C - B + 2yWX\beta}{2\sigma^4} \\ & -\frac{X X}{\sigma^2} & 0 \\ & & -\frac{n}{2\sigma^4} \end{bmatrix} \tag{3.31}$$

Para los modelos que implican un gran número de observaciones n, la parte computacionalmente difficultad de evaluar el hessiano analítico en (3.31) implica el término: $\operatorname{tr}(WAWA) = \operatorname{tr}(W(I\rho W_n)^{-1} W (I\rho W_n)^{-1})$. Realizado de forma computacionalmente sencilla, esto requeriría calcular la matriz inversa $n \times n$, $A = (I\rho W_n)^{-1}$, así como multiplicaciones matriciales que implican la matriz de pesos espaciales n-dimensional W. Este enfoque requeriría operaciones $O(n^3)$, ya que A es densa para los problemas espacialmente conectados. Los términos restantes implican productos matriciales-vectoriales, y observamos que la matriz de pesos espaciales suele ser una matriz dispersa que contiene un número relativamente pequeño de elementos distintos de cero. Como ya se ha señalado, esto permite utilizar rutinas de matrices dispersas que pueden efficientemente realizar los productos matriz-vector.

Existen al menos tres formas de tratar el término $\operatorname{tr}(W(I\rho W_n)^{-1} W (I_n \rho W)^{-1})$. En primer lugar, se puede calcular exactamente como en Smirnov (2005). En segundo lugar, la estimación de esta traza lleva poco tiempo, y lo examinaremos en el capítulo 4. En tercer lugar, este término está subsumido en la segunda derivada de la concentración
La probabilidad logarítmica con respecto a ρ, una cantidad que surge como un subproducto de la optimización de la probabilidad logarítmica concentrada utilizando el método de Newton. Denominamos a esta última estrategia el *hessiano mixto analítico-numérico*. En esta sección, mostramos cómo funciona, y proporcionamos una ilustración aplicada que demuestra que este enfoque es computacionalmente fácil de implementar y preciso.

Para empezar, dado que nos basamos en la optimización univariante de la probabilidad logarítmica concentrada, esto no producirá un hessiano numérico completo, sino más bien un hessiano numérico que pertenece sólo al parámetro ρ (o λ) que surge de la probabilidad concentrada etiquetada L_p en (3.32).

$$\frac{\partial L^2_p}{\partial \rho^2} \tag{3.32}$$

Como señalan Davidson y MacKinnon (2004), podemos trabajar con la verosimilitud concen- trada L_p para producir valores correctos para el parámetro ρ (o λ), pero necesitamos el hessiano de la verosimilitud completa (L) H, que puede expresarse en términos del parámetro de dependencia espacial escalar ρ y un vector θ que contiene los restantes parámetros, θ = β σ^2.

$$H = \begin{bmatrix} \dfrac{\partial L^2}{\partial \rho^2} & \dfrac{\partial L^2}{\partial \rho \partial \theta} \\ \dfrac{\partial L^2}{\partial \theta \partial \rho} & \dfrac{\partial L^2}{\partial \theta \partial \theta} \end{bmatrix} \tag{3.33}$$

Es posible ajustar el hessiano de verosimilitud concentrada empírica para que produzca el elemento apropiado para el hessiano completo, como se ilustra en (3.34).

$$\frac{\partial L^2}{\partial \rho^2} = \frac{\partial Lp^2}{\partial \rho^2} + \frac{\partial L^2}{\partial \rho \partial \theta} \left(\frac{\partial L^2}{\partial \theta \partial \theta}\right)^{-1} \frac{\partial L^2}{\partial \theta \partial \rho} \tag{3.34}$$

Esta expresión (3.34), de fácil cálculo (los detalles se verán más adelante), se puede sustituir por el hessiano completo en (3.35).

$$H = \begin{bmatrix} \dfrac{\partial L^2_p}{\partial \rho^2} + \dfrac{\partial L^2}{\partial \rho \partial \theta} \left(\dfrac{\partial L^2}{\partial \theta \partial \theta}\right)^{-1} \dfrac{\partial L^2}{\partial \theta \partial \rho} & \dfrac{\partial L^2}{\partial \rho \partial \theta} \\ \dfrac{\partial L^2}{\partial \theta \partial \rho} & \dfrac{\partial L^2}{\partial \theta \partial \theta} \end{bmatrix} \tag{3.35}$$

Utilizando este enfoque, podemos sustituir el cálculo difficultoso que[11] implica $H^{(a)}$ por el hessiano empírico de probabilidad concentrada ajustado de (3.34). Un punto clave es que la estimación de máxima verosimilitud, tal como se expone en la sección 3.1, ya produce un vector de valores de log-verosimilitud concentrados en función del parámetro ρ. Dado este vector de log-verosimilitudes concentradas, $\partial L/\partial \rho^2_p{}^2$ no cuesta casi nada de calcular.

Para el caso del modelo SAR, esto da como resultado el hessiano numérico analítico mixto etiquetado $H^{(m)}$ en (3.36).

$$H^{(m)} = \begin{bmatrix} \frac{\partial L_p^2}{\partial \rho^2} + Q - \sigma^2 & \frac{yWX}{\sigma^2} & \frac{2C - B + 2yWX\beta}{2\sigma^4} \\ & \frac{X'X}{\sigma^2} & 0 \\ & & \frac{n}{2\sigma^4} \end{bmatrix} \quad (3.36)$$

$$Q = v' \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}^{-1} v \quad (3.37)$$

$$v = \begin{bmatrix} \frac{yWX}{\sigma^2} & \frac{2C - B + 2yWX\beta}{2\sigma^4} \end{bmatrix} \quad (3.38)$$

Observamos que $\partial L_p^2/\partial\rho^2$ representa la estimación de la segunda derivada de la logverosimilitud concentrada con respecto a ρ que surge como subproducto de la optimización. Añadimos este término a la forma cuadrática fácilmente calculada en Q que se muestra en (3.37) y (3.38). Esto da como resultado una simple mezcla numérica analítica
El hessiano que puede utilizarse para la inferencia de los parámetros del modelo.

Como es bien sabido, la matriz de varianza-covarianza pertinente para las estimaciones de los parámetros es igual a H^{-1}. Dada $H^{(m)}$, se pueden simular fácilmente las estimaciones de los parámetros utilizando desviaciones normales multivariantes. Esta capacidad de simular rápidamente la estimación de los parámetros facilita la identificación de la distribución de los impactos directos e indirectos que hemos analizado en el capítulo 2.

3.2.2 Una comparación de los cálculos hessianos

Para comparar los distintos enfoques del cálculo de *las estadísticas t* asociadas a los parámetros de regresión espacial, utilizamos un conjunto de datos de muestra de Pace y Barry (1997) que contiene información de 3.107 condados de EE.UU. sobre la participación de los votantes en las elecciones presidenciales1980. La variable dependiente representa la participación de los votantes, los que votan como *proporción (registrada)* de los que tienen derecho a votar. Las variables explicativas incluyen la población (registrada) mayor de 18 años que *vota*, la población (registrada) con título *universitario*, la población (registrada) que es propietaria *de* una *vivienda* y los *ingresos* medios de los hogares (registrados).

Los datos se ajustaron utilizando el modelo SDM, que incluye rezagos espaciales de las variables explicativas, etiquetados como *Lag Voting Pop*, *Lag Education*, etc. La tabla 3.1 presenta los *estadísticos t* resultantes calculados utilizando: desviaciones de raíz con signo (SRD), el hessiano analítico (Analítico), el Monte Carlo de cadenas de Markov bayesiano (MCMC), la mezcla del hessiano empírico y teórico (Mixto), y un cálculo del hessiano puramente numérico (Numérico). Los resultados

CUADRO 3.1: Comparación de *los estadísticos t* calculados con enfoques alternativos

Variables	SRD	Analítica	MCMC	Mixto	Numérico
Votos/educación popular	−29.401	−31.689	−31.486	−31.689	−38.643
	7.718	7.752	7.752	7.752	7.922
Propietarios viviendas	27.346	29.191	28.977	29.191	29.837
Ingresos	1.896	1.897	1.930	1.897	2.633
Lag Votes/Pop	12.549	12.904	12.961	12.907	13.190
Lag Educación	1.570	1.560	1.621	1.560	1.510
Lag Propietarios de viviendas	12.114	12.381	12.375	12.382	12.671
Ingresos por retraso	4.662	4.661	4.713	4.661	5.038
Interceptar	11.603	11.449	11.529	11.453	11.597
ρ	33.709	41.374	41.430	41.427	47.254

en la tabla demuestran *estadísticas t* muy similares de las técnicas Analítica, MCMC y Mixta. Las estimaciones numéricas del hessiano difieren sustancialmente de los otros resultados del hessiano para algunas variables, como los ingresos, a pesar de que los datos de la muestra estaban bien escalados en este ejemplo. Los resultados de SRD, que utilizan la inferencia de la razón de verosimilitud, coinciden con los de los Hesianos analítico y mixto para los parámetros de regresión, aunque *los estadísticos t de* los parámetros de regresión de SRD parecen ligeramente conservadores y el *estadístico t* de ρ es sustancialmente más conservador.

El tiempo de cálculo requerido fue de unos 0,6 segundos para calcular los términos analíticos del hessiano junto con los ajustes de (3.34).

3.3 Variables omitidas con dependencia espacial

La existencia de variables omitidas espacialmente dependientes parece un hecho probable en la práctica aplicada. Por ejemplo, consideremos la literatura de regresión del crecimiento espacial que analiza el crecimiento de la renta regional transversal en función de los niveles de renta del período inicial y otras variables explicativas que describen las características regionales que se cree que influyen en el crecimiento económico (Abreu, de Groot y Florax, 2004; Ertur y Koch, 2007; Ertur, LeGallo y LeSage, 2007; Fingleton, 2001; Fischer y Stirbock, 2006). Si bien es posible que exista información regional sobre variables explicativas como el capital humano, es probable que no se disponga fácilmente de información de datos muestrales que afecten al capital físico y a otros determinantes importantes del crecimiento económico regional. Dado que el capital físico está probablemente correlacionado con el capital humano, y también es probable que presente una dependencia espacial, las circunstancias de las variables omitidas descritas en la sección 2.2 parecen plausibles.

En la sección 3.3.1,presentamos una prueba estadística que compara las estimaciones de mínimos cuadrados ordinarios (MCO) y SEM que puede utilizarse para diagnosticar la especificación errónea en general, y la posible existencia de variables omitidas. La motivación para este tipo de comparación es que la teoría indica que las estimaciones OLS y SEM deberían ser las mismas si la verdadera DGP es OLS, SEM o cualquier otro modelo de error.

Varios autores (Brasington y Hite, 2005; Dubin, 1988; Cressie, 1993, p. 25) han sugerido que las variables omitidas afectan menos a los métodos de regresión espacial que los mínimos cuadrados. En la sección 3.3.2 exploramos esta cuestión derivando una expresión para el sesgo de las variables omitidas de MCO en una versión univariante del modelo. Demostramos que la dependencia espacial en la variable explicativa exacerba el sesgo habitual de las variables omitidas que se produce cuando se utiliza incorrectamente OLS para estimar un modelo SEM en presencia de una variable omitida espacialmente dependiente.

En la sección 3.3.3 exploramos la conjetura de que los métodos de regresión espacial sufren menos el sesgo de las variables omitidas. Se demuestra que la DGP asociada a las variables omitidas espacialmente dependientes coincide con la DGP del MDE. El uso de este modelo en presencia de variables omitidas reduce el sesgo en relación con las estimaciones OLS, lo que proporciona una fuerte motivación econométrica para el uso del modelo SDM en el trabajo aplicado. También existen buenas motivaciones teóricas para el modelo SDM (Ertur y Koch, 2007).

3.3.1 Prueba de Hausman para las estimaciones OLS y SEM

Como ya se ha señalado en la sección 2.2, las estimaciones OLS de los parámetros β serán insesgadas si la DGP subyacente representa el modelo SEM, pero *los estadísticos t* de los mínimos cuadrados están sesgados. Como se muestra en la sección 2.2, el error de especificación derivado de la presencia de variables omitidas correlacionadas con la variable explicativa y de la dependencia espacial en las perturbaciones conducirá a una DGP que reflecte el modelo SEM. Como se muestra en la sección 3.1.2, el modelo SDM anida el modelo SEM como un caso especial, proporcionando la intuición para este resultado.

Exploramos una prueba estadística formal para la igualdad de las estimaciones del coefficiente de OLS y SEM, ya que pasar esta prueba sería una buena indicación de que los problemas de especificación (como las variables omitidas correlacionadas con las variables explicativas) no estaban presentes en el modelo SEM.

Como motivación para la prueba, observamos que si la verdadera DGP es cualquier modelo de error, en un contexto de muestreo repetido la media de las estimaciones del parámetro del modelo de error para β debería ser igual. Esto es cierto incluso con variables omitidas, siempre que éstas sean independientes de X. Para ver esto, considere el modelo de error DGP en (3.39) donde F es alguna matriz desconocida, arbitraria y fija, y z es una variable omitida que es independiente de X. Considere el estimador de *mínimos cuadrados generalizados (*GLS) en (3.40) basado en alguna matriz de covarianza fija y arbitraria G que puede no tener relación con una función de F. Para cualquier elección de F y G, incluso en presencia de z, el valor esperado de las estimaciones es igual a β, como se muestra en (3.41).

$$y = X\beta + z + F\varepsilon \quad (3.39)$$

$$\beta G = (XGX^{-1})^{-1} XGy^{-1} \quad (3.40)$$

$$\beta G = (XGX^{-1})^{-1} XGX\beta^{-1} + (XGX^{-1})^{-1} XG^{-1}(z + F\varepsilon)$$

$$E(\hat{\beta}G) = \beta \quad (3.41)$$

Intuitivamente, las perturbaciones con una expectativa nula, ya sean derivadas de variables omitidas o de una mala especificación (siempre que sean ortogonales a las variables explicativas incluidas) no affectan las estimaciones de los parámetros asociados a las variables explicativas.

Estos resultados teóricos sugieren que una DGP de error espacial debería dar lugar a estimaciones de parámetros OLS y SEM que son (en promedio) iguales para los parámetros β, a pesar de la presencia de algunos tipos de especificación errónea del modelo. Sin embargo, la literatura contiene una serie de ejemplos en los que los investigadores presentan estimaciones de MCO y SEM que no parecen cercanas en magnitud.

Se puede utilizar una prueba de Hausman (Hausman, 1978) siempre que haya dos estimadores, uno de los cuales es ineficaz pero consistente (MCO en este caso bajo la hipótesis mantenida de la DGP de SEM), mientras que el otro es eficaz (SEM en este caso). Exponemos una prueba de Hausman para detectar differencias estadísticamente significativas entre las estimaciones OLS y SEM. Argumentamos que esta prueba puede ser útil para diagnosticar la presencia de variables omitidas que están correlacionadas con las variables incluidas en el modelo. Dado que este escenario conduce a una especificación del modelo que debería incluir un retardo espacial de la variable dependiente, esperaríamos ver estimaciones OLS y SEM que son significativamente differentes.

Si dejamos que $\gamma = \hat{\beta}OLS\ \hat{\beta}SEM$ represente la differencia entre las estimaciones OLS y SEM, el estadístico T de la prueba de Hausman (bajo la hipótesis mantenida de la DGP SEM) tiene la forma simple en (3.42), donde $\hat{\Omega}O$ representa una estimación consistente de la matriz de varianza-covarianza asociada a $\hat{\beta}OLS$ (dado un modelo de error espacial DGP). La hipótesis nula es que las estimaciones SEM y OLS no son significativamente differentes. La hipótesis alternativa es que un
diflerencia significativa entre los dos conjuntos de estimaciones.

$$T = \gamma\ (\hat{\Omega}O - \hat{\Omega}S)\gamma^{-1} \quad (3.42)$$

La expresión (3.43) implica (3.44), y la expectativa del producto exterior de (3.44) se muestra en (3.45). Aunque la matriz de varianza-covarianza estimada por MCO_o habitual $\sigma^2(XX)^{-1}$ es inconsistente para la DGP SEM, Cordy y Griffith (1993) muestran que (3.45) es un estimador consistente. Bajo la hipótesis principal de la DGP SEM, las estimaciones SEM de máxima verosimilitud de $\hat{\sigma}^2$ y $\hat{\lambda}$ proporcionan estimaciones consistentes que pueden usarse para reemplazar σ^2 y λ en (3.45) dando como resultado (3.46).

$$\hat{\beta}_O = \beta + H(I_n - \lambda W)\varepsilon^{-1} \quad (3.43)$$

$$\hat{\beta}_O - E(\hat{\beta}_O) = H(I_n - \lambda W)\varepsilon^{-1} \quad (3.44)$$

$$H = (X'X)^{-1}X'$$

$$\Omega_O = \sigma^2 H(I_n - \lambda W)^{-1}(I_n - \lambda W')^{-1}H' \quad (3.45)$$

$$\hat{\Omega}_O = \hat{\sigma}^2 H(I_n - \hat{\lambda} W)^{-1}(I_n - \hat{\lambda} W')^{-1}H' \quad (3.46)$$

En (3.42), $\hat{\Omega}_S$ representa una estimación consistente para la varianza-covarianza asociada a $\hat{\beta}_{SEM}$, de nuevo bajo la hipótesis mantenida del proceso de error espacial, donde $\hat{\Omega}_S$ se muestra en (3. 47).

$$\hat{\Omega}_S = \hat{\sigma}^2 (X'(I_n - \hat{\lambda} W')(I_n - \hat{\lambda} W)X)^{-1} \quad (3.47)$$

Observamos que aunque las estimaciones SEM para β son insesgadas, las de la matriz de varianza-covarianza sólo son consistentes debido a la dependencia del parámetro estimado λ. Véase Lee (2004) sobre la consistencia de las estimaciones de regresión espacial y Davidson y MacKinnon (2004, p. 341-342) para una excelente discusión de las pruebas de Hausman.

El estadístico T sigue una distribución chi-cuadrado con grados de libertad iguales al número de parámetros de regresión probados. A modo de resumen, las estimaciones de máxima verosimilitud para $\hat{\beta}_{SEM}$, $\hat{\lambda}$, $\hat{\sigma}^2$ junto con $\hat{\beta}_{OLS}$ pueden utilizarse junto con las estimaciones consistentes para $\hat{\Omega}_O$ de (3.46) y $\hat{\Omega}_S$ en (3.47) para calcular el estadístico de prueba T. Esto nos permite comprobar si hay diferencias significativas entre las estimaciones del coeficiente SEM y OLS.

Si no podemos rechazar la hipótesis nula de igualdad, esto sería un indicio de que las variables omitidas no representan un problema grave o no están correlacionadas con las variables explicativas. Si el MEB tiene una probabilidad significativamente mayor que el MCO, pero la prueba de Hausman no encuentra una diferencia significativa entre las estimaciones del MEB y del MEB, esto indica que el término de error espacial en el MEB está capturando el efecto de las variables omitidas, pero éstas no están correlacionadas con las variables incluidas.

El rendimiento de esta prueba de Hausman espacial se examinó en Pace y LeSage (2008) en condiciones controladas utilizando una DGP SEM simulada basada en 3, observaciones000 y niveles variables de dependencia espacial asignados al parámetro λ. Muestran que los tamaños estimados para esta prueba se ajustan estrechamente a los tamaños teóricos.

3.3.2 Sesgo de las variables omitidas en los mínimos cuadrados

A menudo, las variables explicativas utilizadas en los modelos de regresión espacial muestran depen- dencia, ya que reflectan las características regionales. Por ejemplo, en un modelo de precios hedónicos de la vivienda, variables como los niveles de renta, el nivel educativo y los tiempos de desplazamiento al trabajo suelen mostrar similitudes a lo largo del espacio, o dependencia espacial. Además, los precios de la vivienda se ven afectados por influencias latentes no observables, como la calidad arquitectónica, la atención a la jardinería en un barrio, el acceso conveniente a restaurantes populares, la accesibilidad a pie, el ruido, así como otros factores. Estas variables latentes también pueden presentar similitudes a lo largo del espacio. Debido a las limitaciones de los datos, es probable que estas variables latentes se omitan en los modelos. Discutimos una expresión para el sesgo de la variable omitida que surge cuando se utilizan estimaciones OLS en circunstancias en las que las variables explicativas incluidas y omitidas presentan dependencia espacial y el proceso de perturbación es espacialmente dependiente como en el modelo SEM. La expresión muestra que la dependencia espacial en una única variable explicativa incluida exacerba el sesgo habitual que se produce al utilizar MCO para estimar un modelo SEM en presencia de una variable omitida espacialmente dependiente que está correlacionada con la variable explicativa incluida.

Derivamos una expresión para el sesgo que surgiría al utilizar es- timas OLS en presencia de dependencia espacial en las perturbaciones, las variables explicativas incluidas y las omitidas. Trabajamos con un vector x que representa una única variable explicativa (no constante) con una media de cero y que sigue una distribución normal *iid* y dejamos que y sea la variable dependiente. Añadimos una variable omitida al modelo SEM y permitimos que un proceso de dependencia espacial gobierne esta variable así como la variable explicativa incluida, dando lugar al modelo de (3.48) a (3.51). Los vectores ε, y ν representan $n \times 1$ vectores de perturbación, y suponemos que ε se distribuye $N(0, \sigma_\varepsilon^2 I_n)$, ν se distribuye $N(0, \sigma_v^2 I_n)$, y ε es independiente de ν.

$$y = x\beta + u \tag{3.48}$$

$$u = \lambda W u + \eta \tag{3.49}$$

$$\eta = x\gamma + \varepsilon \tag{3.50}$$

$$x = \varphi W x + \nu \tag{3.51}$$

Los parámetros escalares del modelo son: β, λ, φ y γ, y W es un $n \times n$ matriz de pesos espaciales simétrica no negativa con ceros en la diagonal.

Las expresiones (3.48) y (3.49) son los enunciados habituales del modelo SEM y (3.50) añade una variable omitida, donde la fuerza de la dependencia (correlación) entre el vector de la variable incluida x y el vector de la variable omitida η se controla mediante el parámetro γ. Por último, (3.51) especifica una autorregulación espacial.

Nos centramos en la dependencia espacial no negativa, asumiendo que $\lambda, \varphi \in [0, 1)$.

Pace y LeSage (2009b) derivan expresiones teóricas para el sesgo asociado con el uso de estimaciones OLS en estas circunstancias, como se muestra en

(3.52) a (3.54).

$$\text{plim}_{n\to\infty}\ \hat{\beta}_o = \beta + T_\gamma(\varphi, \lambda)\gamma \tag{3.52}$$

$$T_\gamma(\varphi, \lambda) = \frac{\text{tr}[H(\varphi)^2 G(\lambda)]}{\text{tr}[H(\varphi)^2]} \tag{3.53}$$

$$G(\lambda) = (I_n - \lambda W)^{-1},\ H(\varphi) = (I_n - \phi W)^{-1} \tag{3.54}$$

A medida que el factor $T_\gamma(\varphi, \lambda)$ adquiere valores superiores a la unidad, aumenta el sesgo de las estimaciones OLS para este modelo. La magnitud del sesgo depende del parámetro φ que representa la fuerza de la dependencia espacial en la variable explicativa, el parámetro λ que refleja la dependencia del error y el parámetro γ que gobierna la correlación entre la variable incluida y la omitida.

Pace y LeSage (2009b) muestran que $T_\gamma(\varphi, \lambda) > 1$para $\lambda > 0$y la dependencia espacial en el regresor, $\varphi > 0$, amplía estos factores. Este modelo en- abarca el modelo SEM como un caso especial. A continuación se enumeran los sesgos asintóticos que surgen al utilizar las estimaciones por mínimos cuadrados en circunstancias alternativas como la presencia/ausencia de variables omitidas y la presencia/ausencia de dependencia espacial en las variables independientes y las perturbaciones.

1. *Dependencia espacial en las perturbaciones y el regresor* : ($\gamma = \lambda 0,, \varphi > 0$), conduce a $\text{plim}_{n\to\infty}\ \hat{\beta}_o = \beta$, y no hay sesgo asintótico.
2. *Dependencia espacial en el regresor en presencia de una variante omitida de la misma manera*: ($\lambda = 0$), mientras que $\hat{\beta}_o = \beta + \gamma$, representa- ($\gamma = 0$), resulta en plimn el sesgo de variable omitida estándar.
3. *Existe una variable omitida en presencia de la dependencia espacial en el regresores y perturbaciones*: ($\gamma \neq 0$, $\varphi, \lambda > 0$) entonces $\text{plim}_{n\to\infty}\ \hat{\beta}_o = \beta + T_\gamma(\varphi, \lambda)\gamma$, y OLS tiene un sesgo de variables omitidas ampliado por la dependencia espacial en las perturbaciones y en el regresor.

El primer resultado es bien conocido, y el segundo es una extensión menor del caso convencional de variables omitidas para los mínimos cuadrados. El tercer resultado muestra que la dependencia espacial en las perturbaciones (y/o en el regresor) en presencia de variables omitidas conduce a una magnificación del sesgo convencional de las variables omitidas. Este tercer resultado difiere del hallazgo habitual de que la dependencia espacial de las perturbaciones no conduce al sesgo.

Para dar una idea de la magnitud de estos sesgos, presentamos los resultados de un pequeño experimento de Monte Carlo en la Tabla 3.2. Simulamos un conjunto espacialmente aleatorio de 1.000 lugares y los utilizamos para construir una matriz basada en la contigüidad W. La1 matriz de pesos espaciales 000simétrica000 Wresultante1 se estandarizó para que fuera doblemente estocástica (tener sumas de unidad tanto en las filas como en las columnas). La variable independiente *x se estableció* como un vector normal unitario *iid* con media cero. Fijamos β = 0,75 y γ = 0,25 para todos los ensayos. Dado Wy un valor para λ y φ, utilizamos la DGP para simular1 , muestras 000de y, y para

cada muestra se calculó la estimación OLS y se registró la media de las estimaciones (etiquetada como media β̂o en la Tabla 3. 2). Se utilizó un conjunto de nueve combinaciones de λ y φ, y la expectativa teórica (etiquetada como E(β̂o) en la Tabla 3. 2) se calculó para cada una de ellas utilizando la expresión (3.52).

TABLA 3.2: Sesgo de las variables omitidas en función de la dependencia espacial

Experimento	φ	λ	media β̂o	E(β̂o)
1	0.0	0.0	0.9990	1.0000
2	0.5	0.0	1.0015	1.0000
3	0.9	0.0	1.0003	1.0000
4	0.0	0.5	1.0173	1.0159
5	0.5	0.5	1.0615	1.0624
6	0.9	0.5	1.1641	1.1639
7	0.0	0.9	1.1088	1.1093
8	0.5	0.9	1.3099	1.3152
9	0.9	0.9	1.9956	2.0035

La tabla muestra la media empírica de las estimaciones y los valores esperados para las nueve combinaciones de λ y φ. Los resultados teóricos y empíricos muestran una estrecha concordancia, y la tabla documenta que puede producirse un grave sesgo cuando las variables omitidas se combinan con la dependencia espacial en el proceso de perturbación, especialmente en presencia de dependencia espacial en el regresor. Por ejemplo, las estimaciones OLS arrojan una media empírica de 1.9956 (que se acerca al valor teórico de 2,0035) cuando λ y φ son iguales a 0,9, aunque β = 0,75 y γ = 0. 25.En este caso, $T_\gamma(\varphi, \lambda)$ es aproximadamente igual a 5. Si β = 1 y γ = 0,*2*, un $T_\gamma(\varphi, \lambda)$ de 5 significaría que una regresión OLS produciría una estimación cercana a 0. Por lo tanto, la inflación del sesgo habitual de las variables omitidas podría dar lugar a que no se perciba ninguna relación entre y y x. *A fortiori*, la estimación del parámetro de MCO sería igual a 1 cuando el parámetro verdadero fuera igual a 1, cuando γ = 0.4. Por lo tanto, la inflación del sesgo de la variable omitida en presencia de la dependencia espacial puede tener graves consecuencias inferen- ciales al utilizar MCO.

Además, Pace y LeSage (2009b) estudian un modelo más general que incluye la dependencia espacial en y, *así como* las perturbaciones y las variables explicativas. Naturalmente, la dependencia espacial en y aumenta aún más el sesgo de OLS.

3.3.3 Sesgo de las variables omitidas en las regresiones espaciales

Consideramos la conjetura hecha por varios autores (Brasington y Hite, 2005; Dubin, 1988; Cressie, 1993, p. 25) de que las variables omitidas affectan los métodos de regresión espacial menos que los mínimos cuadrados ordinarios.

Comenzamos examinando la DGP implícita para el caso de dependencia espacial en las variables omitidas y las perturbaciones para una x dada. Al manipular estas ecuaciones se obtiene una ecuación que se muestra en (3.55) en términos de rezagos espaciales de las variables dependientes e independientes.

$$y = \lambda W\ y + x(\beta + \gamma) + \ Wx(-\lambda\beta) + \varepsilon(3.55)$$
$$y = \lambda W\ y + x\beta + Wx\psi + \varepsilon(3.56)$$

Podemos utilizar el modelo SDM en (3.56) para producir estimaciones consistentes para los parámetros λ y ψ, ya que este modelo coincide con la DGP en las circunstancias de variables omitidas expuestas. Estas estimaciones consistentes serían iguales a las
parámetros estructurales subyacentes del modelo en muestras grandes.[3] En otras palabras, para n sufficientemente grande la estimación (3.56) produciría $E(\hat\beta) = \beta + \gamma$, $E(\hat\psi) = \lambda\beta$, y $E(\hat\lambda) = \lambda$. No hay sesgo asintótico en la estimación de λ para el modelo SDM en (3.56) a pesar de la presencia de variables omitidas.

Sin embargo, existe un sesgo asintótico de variable omitida en las estimaciones de este modelo para β, ya que $E(\hat\beta)\ \beta = \gamma$. A diferencia de los resultados para OLS presentados en (3. 52), este sesgo no depende de x, eliminando la influencia del parámetro
φ que reflecta la fuerza de la dependencia espacial en la variable incluida x. Además, el sesgo no depende de la dependencia espacial en las perturbaciones especificadas por el parámetro λ. En cambio, el sesgo de la variable omitida es constante y depende sólo de la fuerza de la relación entre la variable explicativa incluida y la omitida reflectada por el parámetro γ. Esto es similar al resultado del sesgo de la variable omitida del modelo de regresión convencional.

Estos resultados coinciden con la observación anterior de que las variables omitidas afectan menos a los métodos de regresión espacial que los mínimos cuadrados ordinarios. Esta protección contra el sesgo de las variables omitidas está sujeta a algunas advertencias, ya que debemos pro- ducir estimaciones utilizando un modelo que coincida con la DGP implícita del modelo después de tener en cuenta la presencia de variables omitidas (y la presencia de dependencia espacial en éstas y en las variables explicativas, así como las perturbaciones). Como se ha demostrado, el uso de la regresión SEM no contendrá el desfase espacial de las variables dependientes y explicativas que implica la presencia de variables omitidas. Recordemos que, según la teoría básica de la regresión, la inclusión de variables explicativas que no figuran en la DGP no da lugar a un sesgo en las estimaciones. Sin embargo,

[3] Véanse Kelejian y Prucha (1998), Lee (2004) y Mardia y Marshall (1984) sobre la coherencia de las estimaciones de los modelos de regresión espacial.

El sesgo de las variables omitidas surge cuando se excluyen del modelo las variables que intervienen en la DGP.

Por otro lado, el modelo SDM sí coincide con la DGP implícita que surge de la presencia de variables omitidas y de variables explicativas dependientes del espacio. Consideremos el caso inverso en el que aplicamos el modelo SDM para producir estimaciones cuando la verdadera DGP es la del SEM y no hay variables omitidas. Las estimaciones del SDM deberían seguir siendo consistentes, pero no efficientes.

Como enfoque algo más general, Pace y LeSage (2009b) utilizan la DGP del SAC y examinan los efectos de una variable omitida que está correlacionada con la variable incluida, x. La presencia de una variable omitida también conduce a un modelo SDM ampliado que incluye un retardo espacial de las variables explicativas, Wx, y que subsume el SAC. Consideremos el caso de que no haya variables omitidas, donde la verdadera DGP es el modelo SAC. El uso del modelo SDM ampliado para producir estimaciones en estas circunstancias (donde la verdadera DGP es el modelo SAC) da lugar a es- tados inefficientes, pero consistentes, del modelo SDM ampliado para la variable explicativa. Obsérvese que la efficiencia de las estimaciones no suele ser la principal preocupación para las muestras espaciales grandes. Consideremos ahora el caso inverso en el que la verdadera DGP es el modelo SDM ampliado, pero estimamos el modelo SAC. Las estimaciones de los coefficientes de las variables explicativas estarán sesgadas debido a una exclusión incorrecta de las variables explicativas espacialmente retardadas (WX) del modelo. En otras palabras, cuando la verdadera DGP está asociada al modelo SDM ampliado en el que las variables explicativas de las re- giones vecinas son importantes, el uso del modelo SAC producirá estimaciones sesgadas que sufferán de los problemas de variables omitidas del tipo que hemos considerado.

A modo de conclusión, examinamos el impacto de las variables omitidas en los mínimos cuadrados y en varias estimaciones de modelos de regresión espacial cuando la DGP reflecta la dependencia espacial en: la variable dependiente, la variable independiente y las perturbaciones. Encontramos que el sesgo convencional de las variables omitidas se am- plía cuando se utilizan procedimientos de estimación OLS para estos modelos. El uso de ciertos modelos de regresión espacial, como el SDM, junto con estimadores consistentes, producirá estimaciones que no sufren el sesgo ampliado. Estos resultados proporcionan una fuerte motivación para el uso de la especificación del modelo SDM en trabajos aplicados donde los problemas de variables omitidas parecen probables.

3.4 Un ejemplo aplicado

Para ilustrarlo de forma sencilla, nos basamos en una relación entre la *productividad total de los factores* regionales (pft) como variable dependiente *y las* existencias regionales de conocimientos como única variable explicativa. Como se ilustra en el capítulo 1, la variable dependiente tfp puede construirse utilizando los residuos de un modelo log-lineal

Regresión de la función de producción Cobb-Douglas con imposición de rendimientos constantes a escala. La variable dependiente utilizada aquí se construyó utilizando una estimación empírica de los porcentajes relativos del trabajo y el supuesto de rendimientos constantes a escala.

La variable dependiente (la productividad total de los factores) representa lo que a veces se denomina el residuo de Solow, como se explica en el capítulo 1. Desde este punto de vista, es plausible que nos basemos en un único vector de variables explicativas A que represente el stock regional de conocimientos, lo que da lugar al modelo de (3.57), donde utilizamos a en (3.58) para representar lnA.

$$y = \alpha\iota_n + \beta \ln A + \varepsilon \quad (3.57)$$
$$y = \alpha\iota_n + \beta a + \varepsilon \quad (3.58)$$

La variable A se construyó utilizando el stock de patentes regionales descontado de forma adecuada como proxy del stock regional de conocimiento. LeSage, Fischer y Scherngell (2007) ofrecen una descripción detallada de los datos de la muestra, que abarca 198 regiones de la Unión Europea de los 15 Estados miembros anteriores a 2004. El modelo relaciona las existencias de conocimientos regionales con la productividad total de los factores regionales para explorar si las existencias de conocimientos influyen en la efficiencia con la que las regiones utilizan sus factores físicos de producción.

Aunque utilicemos el stock regional de patentes como indicador empírico de la tecnología, es poco probable que refleje la verdadera tecnología disponible en las regiones. Esto se debe a que los conocimientos producidos por las empresas innovadoras sólo se apropian parcialmente debido a la naturaleza de bien público del conocimiento que se extiende a otras empresas dentro de la región y en las regiones cercanas. Podríamos plantear la existencia de un conocimiento no medido a^* que está excluido del modelo pero correlacionado con la variable incluida a. Es bien sabido que las patentes regionales exhiben dependencia espacial (Parent y LeSage, 2008; Autant-Bernard, 2001), por lo que, como ya se ha motivado, esto llevaría a un modelo SDM:

$$y = \alpha\iota_{0n} + \rho Wy + \alpha a_1 + \alpha W a_2 + \varepsilon \quad (3.59)$$

El modelo SDM en (3.59) subsume el modelo de error espacial SEM como un caso especial cuando la restricción del parámetro: $\alpha_2 = \rho\alpha_1$. El modelo SEM surgiría si no hubiera correlación entre los stocks de conocimiento medidos y no medidos, a y a^*, y cuando la restricción $\alpha_2 = \rho\alpha_1$ es verdadera.[4] En el capítulo 6 aplicamos una sencilla prueba de razón de verosimilitud del modelo SEM frente al SDM para probar la restricción $\alpha_2 = -\rho\alpha_1$ para este modelo y los datos de la muestra.

3.4.1 Estimaciones coefficientes

Recordemos que en la sección 2.2 mostramos cómo las variables omitidas espacialmente dependientes conducirán a la presencia de rezagos espaciales de las variables explicativas.

[4] Anselin (1988) lo denomina "restricción del factor común".

Las estimaciones del MEB y del MDE, junto con los estadísticos *t, se presentan en* el cuadro 3.3.

Es frecuente que los estudios aplicados comparen estimaciones como las de un modelo SEM con las de modelos que contienen un retardo espacial de la variable dependiente, como el SAR o el SDM. Esta comparación no es válida, ya que el MEB no contempla los efectos indirectos. En el cuadro 3.4 se presentan las estimaciones resumidas del impacto del SDM basadas en las derivadas parciales, que se analizarán en breve.

TABLA 3.3: Estimaciones del modelo SEM y SDM

	ParámetrosEstimacionesdel modelo MECEstimaciones del modelo MDM*Estadística t* centrada Coefficientt Estadística Coefficientt				
α_0	2.5068	17.28	0.5684	3.10	
α_1	0.1238	6.02	0.1112	5.33	
α_2			0.0160	−0.48	
ρ			− 0.64508.970.6469		9.11

Muchos estudios interpretan erróneamente el coefficiente α_2 sobre el retardo espacial de la variable de capital de conocimiento $(W a)$ como una prueba de la existencia de desbordamientos espaciales. Dado que este coefficiente no es significativamente distinto de cero, errarían.
nalmente, concluyen que no existen desbordamientos espaciales asociados al capital de conocimiento.

3.4.2 Estimaciones de effectos acumulados

La inferencia relativa a los impactos directos e indirectos (spillover) del modelo SDM se basaría en las medidas resumidas de los impactos directos e indirectos para el modelo SDM. La expresión matricial que refleja las derivadas propias y cruzadas para este modelo tiene la forma:

$$S_r(W) = V(W)(I\alpha_{n1} + W\alpha_2)$$
$$V(W) = (I_n - \rho W)^{-1} = I_n + \rho W + \rho W^2 +^2 \rho W^3 + {}^3 \ldots$$

La tabla muestra las estimaciones de los 3.4efectos que se produjeron mediante la simulación de los parámetros utilizando la distribución de parámetros normal multivariante de máxima probabilidad y el hessiano analítico mixto descrito en la sección 3.2.1. Se utilizó una serie de 2.000 extracciones simuladas. Las medias, las desviaciones estándar y *los estadísticos t* se construyeron a partir de los resultados de la simulación.

Si consideramos los impactos directos, vemos que éstos se acercan a las estimaciones del coefficiente del modelo SDM asociado a la variable *a* reportada en Ta- ble La 3.3.differencia entre la estimación del coefficiente de0 .1112 y el

CUADRO 3.4: Estimaciones resumidas de los effectos acumulados

	Efectos medios	Desviación estándar	*Estadística t*
effectodirecto	0.1201	0.0243	4.95
effecto indirecto	0.1718	0.0806	2.13
total effecto	0.2919	0.1117	2.61

La estimación del effecto directo de 0,1201 igual a 0,0089 representa los effectos de retroalimentación que surgen como resultado de los impactos que pasan por las regiones vecinas y vuelven a la propia región. La discrepancia es positiva, ya que la estimación del impacto supera la estimación del coefficiente, lo que refleja cierta retroalimentación positiva. Dado que la diferencia entre el coeflciente del MDS y la estimación directa del impacto es muy pequeña, concluiríamos que los efectos de la retroalimentación son pequeños y no es probable que tengan significación económica.

En contraste con la similitud de las estimaciones de impacto directo y el coefficiente α del SDM$_1$, existen grandes discrepancias entre el coefficiente α_2 de retardo espacial del modelo SDM y las estimaciones de impacto indirecto. Por ejemplo, el impacto indirecto es de 0,1718, y significativamente differente de cero utilizando el estadístico *t*-. La estimación del coefficiente del SDM asociado a la variable de retardo espacial Wa que se presenta en el cuadro es de 3.3-0,0160, y no es significante según el estadístico *t*-.
estadística. Si consideramos incorrectamente el coefficient α_2 del MDE sobre el reflejo espacial de los stocks de conocimiento (Wa) como reflexión del impacto indirecto, esto llevaría a inferir que la variable de capital de conocimiento Wa ejerce un impacto indirecto negativo e insignificante sobre la productividad total de los factores. Sin embargo, el verdadero
La estimación del impacto apunta a un impacto indirecto positivo y significativo (spillover) derivado de los cambios en la variable a.

También se da el caso de que tratar la suma de las estimaciones del coefficiente SDM de las variables a y Wa como estimaciones del impacto total llevaría a resultados erróneos. El impacto total de las reservas de conocimiento sobre la productividad total de los factores es un 0,2919 positivo que es significativo, mientras que el impacto total sugerido por la suma de los coefficientes del SDM equivaldría a menos de la mitad de esta magnitud. Estas differencias dependerán de la magnitud de los impactos indirectos que no pueden inferirse correctamente a partir de los coefficientes del MDF. En los casos en los que los impactos indirectos fueran nulos, y las estimaciones de los impactos directos se aproximaran a las estimaciones del SDM en las variables no retrasadas espacialmente, el impacto total podría inferirse correctamente. Por supuesto, no se sabría si los impactos indirectos eran pequeños o insignificantes sin calcular las medidas de impacto escalares resumidas que se presentan en el cuadro 3.4.

Podemos interpretar las estimaciones del impacto total como elasticidades, ya que el modelo se especifica utilizando niveles registrados de la productividad total de los factores y de las reservas de conocimientos. Basándonos en la estimación0.2919 positiva del impacto total de los stocks de conocimiento, concluiríamos que un aumento del 10 por ciento en el conocimiento regional daría lugar a un aumento del porcentaje2.9 en la productividad total de los factores. Alrededor de 2/5 de este impacto proviene de la magnitud directa effect de 0,1201, y 3/5 del impacto indirecto o de derrame espacial basado en su estimación de impacto escalar de 0.1718.

3.4.3 Distribución espacial de las estimaciones de impacto

Podemos dividir espacialmente estos impactos para ilustrar la naturaleza de su influencia a medida que pasamos de los vecinos inmediatos a los de orden superior. Esto podría ser interesante en aplicaciones en las que el alcance espacial de los efectos indirectos es un objeto de inferencia.

Estos se presentan para el modelo SDM en la Tabla que3.5, muestra la media, la desviación estándar y un *estadístico t* para los effectos *marginales* asociados
con matrices W de orden 0 a Los9. effectos directos de W serán ^{1}iguales a cero y los effectos indirectos de W serán iguales a ^{0}cero, como se ha comentado en el capítulo 2. Por supuesto, si acumuláramos los effectos marginales de la tabla sobre todos los órdenes de W hasta la convergencia empírica de las series infinitas, éstos serían iguales a los effectos acumulativos reportados en la Tabla3.4 .

TABLA 3.5: Distribución espacial marginal de los impactos

	Efectos directos	Desviación estándar	*Estadística t*
W^0	0.1113	0.0205	5.4191
W^1	0.0000	0.0000	-
W^2	0.0046	0.0013	3.5185
W^3	0.0016	0.0007	2.3739
W^4	0.0010	0.0005	2.0147
W^5	0.0006	0.0003	1.6643
W^6	0.0004	0.0002	1.4208
W^7	0.0002	0.0002	1.2285
W^8	0.0001	0.0001	1.0761
W^9	0.0001	0.0001	0.9516
	Efectos indirectos	Desviación estándar	*Estadística t*
W^0	0.0000	0.0000	-
W^1	0.0622	0.0188	3.3085
W^2	0.0353	0.0131	2.6985
W^3	0.0243	0.0105	2.3098
W^4	0.0160	0.0083	1.9220
W^5	0.0107	0.0066	1.6283
W^6	0.0072	0.0052	1.3978
W^7	0.0049	0.0040	1.2151
W^8	0.0033	0.0031	1.0672
W^9	0.0023	0.0024	0.9455

En la tabla vemos que los effectos directos e indirectos muestran la decadencia esperada con matrices W de orden superior. Si utilizamos un valor *estadístico t* de 2 como medida de cuándo los effectos dejan de ser estadísticamente diferentes de cero, vemos que la extensión espacial de los spillovers de las reservas regionales de conocimiento se sitúa en torno a W^4. Para nuestra matriz W basada en los 7vecinos más cercanos, la matriz W contiene elementos no ^{2}nulos18 que representan a los vecinos de segundo orden (para la región media de nuestra muestra). La matriz W contiene ^{3}(en promedio) 30,8 vecinos de tercer orden y W tiene vecinos de ^{4}cuarto45 orden. Esto sugiere que los efectos de desbordamiento espacial que emanan de una sola región ejercen un impacto en una gran proporción de las regiones198 de nuestra muestra. Sin embargo, observamos que el tamaño de los efectos indirectos no es probable que sea económicamente significativo para las regiones vecinas de orden superior.

Utilizando nuestra interpretación de la elasticidad, podemos deducir que un aumento relativamente grande del 10% en las reservas de conocimientos tendría efectos indirectos o espaciales correspondientes a un aumento del 0,6% en la productividad de los factores de la región vecina de primer orden, un aumento del porcentaje0.35 en la productividad de los factores de los vecinos de segundo orden, para0.24 los vecinos de tercer orden, y así sucesivamente.

La otra característica notable de la tabla 3.5 es la pequeña cantidad de effecto de retroalimentación que se muestra en los effectos directos marginales, y el decaimiento relativamente rápido con órdenes de W.

3.4.4 Una comparación de los impactos de los distintos modelos

Es interesante comparar las estimaciones del modelo SDM y el resumen escalar de los effectos con los de los modelos SAR y SAC. Las estimaciones coefficientes se presentan en el cuadro 3.6. Dada la falta de significancia de la variable de-retardo espacial Wa en el modelo SDM, cabría esperar que las estimaciones de los modelos SAR y SDM fueran bastante similares, como se muestra en la Tabla El3.6. modelo SAC dio lugar a una estimación insignificante para el parámetro de dependencia espacial λ asociado a las perturbaciones. Esto también produce estimaciones similares a las de los modelos SAR y SDM.

TABLA 3.6: Estimaciones de los modelos SAR y SAC

Parámetros	Estimaciones del modelo SAR	*Estadístico t*	Estimaciones del modelo SAC	Estadístico t
α_0	0.5649	3.10	0.5625	2.11
α_1	0.1057	5.93	0.1144	5.09
α_2				
ρ	0.6279	10.12	0.6289	6.27
λ			−0.0051	−0.02
σ^2	0.1479			
Log-Likelihood	−29.30		−30.65	

0.1509

Las estimaciones de Effects para el modelo SAR y SAC tienen la misma forma analítica ya que se basan en las expresiones matriciales de la Sección 2.7.

$$S_r(W) = V(W)I\beta_{nr}$$
$$V(W) = (I_n - \rho W)^{-1} = I_n + \rho W + \rho W^2 +^2 \rho W^3 + ^3 \ldots$$

La diferencia entre los impactos de estos dos modelos y los del modelo SDM es el término adicional $W\theta$ que aparece en el caso del modelo SDM. Dado que la distribución de θ está centrada cerca de cero según la estimación puntual y el *estadístico t* asociado, cabría esperar que las estimaciones de impacto de los modelos SAR, SDM y SAC fueran similares en esta ilustración concreta.

TABLA 3.7: Comparación de los impactos acumulativos de SAR, SAC y SDM

	SAR effects	Desviación	*t-estadística*
effecto directo		0.11450.02075.53	
effecto indirecto	0.1746	0.0620	2.81
total effecto	0.2891	0.0827	3.49
	SDM effects	Desviación típica	*Estadística t*
effecto directo	0.1201	0.0243	4.95
effecto indirecto	0.1718	0.0806	2.13
total effecto	0.2919	0.1117	2.61
	SAC effects	Desviación	*t-estadística*
effecto directo	0.1199	0.0241	4.98
effecto indirecto	0.1206	0.0741	1.62
total effecto	0.2405	0.0982	2.44

Observamos que las comparaciones no válidas de las estimaciones puntuales de diferentes especificaciones de modelos de regresión espacial han llevado a los profesionales a concluir que cambiar las especificaciones de los modelos conducirá a inferencias muy diferentes. Esto también puede haber llevado a que la literatura de la econometría espacial se centre excesivamente en los procedimientos para la prueba comparativa de especificaciones de modelos alternativos, un tema que abordamos en el capítulo 6. Sin embargo, el uso de la derivada parcial correcta de los parámetros de varios modelos da como resultado una menor divergencia en las inferencias de differentes especificaciones de modelos. Este resultado está relacionado con la interpretación en derivadas parciales del impacto de los cambios en las variables de las distintas especificaciones del modelo, que representa una base válida para estas comparaciones.

Esto no quiere decir que la especificación del modelo no sea importante. Por ejemplo, el uso de un modelo SEM conduciría a la omisión de los importantes efectos espaciales indirectos encontrados aquí. Además, las estimaciones de los effectos del SAC llevan a inferir que los impactos indirectos no son significativamente diferentes de cero sobre la base del *estadístico t* reportado en la Tabla 3.7.

3.5 Resumen del capítulo

En la sección 3.1 presentamos enfoques computacionalmente efficientes para la estimación de máxima verosimilitud de la familia básica de modelos de regresión espacial. La parte más difícil de la estimación de máxima verosimilitud es el cálculo del término determinante del logaritmo que aparece en la función de logaritmo-verosimilitud, y el capítulo 4 proporcionará detalles al respecto. Además de las estimaciones puntuales, también es necesario proporcionar una estimación de la matriz de varianza-covarianza que pueda utilizarse para la inferencia. En la sección 3.2 se analizaron varias estrategias y se expuso un enfoque mixto que utiliza los resultados numéricos del hessiano para modificar un único término computacionalmente difícil del hessiano analítico.

La caja de herramientas de econometría espacial de dominio público (LeSage, 2007) y la *caja de herramientas de estadística espacial* (Pace, 2007) proporcionan ejemplos de código escritos en el lenguaje MAT- LAB que implementan la mayoría de los métodos discutidos en este texto. Esto debería permitir al lector interesado examinar ejemplos detallados que implementan las ideas presentadas aquí.

La modelización de las relaciones espaciales suele dar lugar a la omisión de influencias latentes de naturaleza espacial. Por ejemplo, las regresiones hedónicas del precio de la vivienda suelen basarse en las características individuales de la casa, que pueden excluir importantes variables del barrio que afectan a la accesibilidad, la calidad de los colegios, los servicios, etc. En la sección 3.3, examinamos la naturaleza del sesgo que surgirá de las variables omitidas en las estimaciones de regresión espacial y de mínimos cuadrados. Una característica interesante de las variables omitidas en los modelos de regresión espacial es que conducirán a procesos de generación de datos que incluyen rezagos espaciales de las variables explicativas, lo que proporciona una poderosa motivación para el uso del modelo espacial de Durbin.

En la sección se ofrece una ilustración aplicada para 3.4reforzar las ideas expuestas en este capítulo. Se compararon las estimaciones de máxima verosimilitud de una familia de modelos de regresión espacial, junto con una interpretación de los parámetros. El modelo simple de una variable se basó en la variación regional de la productividad de los factores para una muestra de 198 regiones de la Unión Europea. Esta ilustración aplicada se centró en el papel de las reservas regionales de conocimiento para explicar la variación de la productividad total regional de los factores.