



Máster Interuniversitario en Economía

Universidade da Coruña

Universidade de Santiago de Compostela

Universidade de Vigo

Traballo Fin de máster

La ecuación salarial de la Nueva Geografía Económica en el ámbito europeo: Un estudio con datos de panel de 1980 a 2015.

Edward Joseph Velo Fuentes

Dirixido por:
Fernando Bruna

Xuño 2021

Traballo de Fin de Máster presentado na Facultade de Economía e Empresa da Universidade de Coruña para a obtención do Máster en Economía

Resumo

Na literatura empírica da Nova Xeografía Económica (NXE) non se fixera antes un estudio da teoría para rexións europeas utilizando técnicas de datos de panel en conxunto con modelos espaciais, as cales poden ofrecer estimacións que non estean sesgadas debido a correlacións con variables omitidas de carácter fixo o con residuos espacialmente autocorrelacionados. Neste traballo facemos un primer acercamento a la teoría da NXE, para logo estimar con técnicas econométricas de datos de panel e componentes espaciais si existe una influencia espacial significativa nas variacións da actividade económica nos países europeos. Acorde coa literatura empírica realizada con modelos de sección cruzada hasta a fecha, os resultados seguen demostrando a existencia dunha influencia espacial positiva e significativa, aínda que cunha magnitude diferente nesta relación.

Palabras clave:

NXE, Potencial de Mercado, Ecuación Salarial, Datos de Panel, Econometría Espacial

Resumen

En la literatura empírica de la Nueva Geografía Económica (NGE) no se ha hecho antes un estudio de la teoría para regiones europeas utilizando técnicas de datos de panel en combinación con modelos espaciales, las cuales pueden ofrecer estimaciones que no estén sesgadas debido a correlaciones con variables omitidas de carácter fijo o con residuos espacialmente autocorrelacionados. En este trabajo hacemos un primer acercamiento a la teoría de la NGE, para luego estimar con técnicas econométricas de datos de panel y componentes espaciales si existe una influencia espacial significativa en las variaciones de la actividad económica en los países europeos. Acorde con la literatura empírica realizada con modelos de sección cruzada hasta la fecha, los resultados siguen demostrando la existencia de una influencia espacial positiva y significativa, aunque con una magnitud diferente en esta relación.

Palabras clave:

NEG, Potencial de Mercado, Ecuación Salarial, Datos de Panel, Econometría Espacial

Abstract

In the New Economic Geography (NEG) literature to date there hasn't been made a study of the theory for european regions using panel data techniques in addition to spatial models, which can offer estimates that are not correlated with fixed omitted variables and with spatial autocorrelated residuals. In this paper we make a first approach to the NGE theory, then estimate with panel data techniques and spatial components if there is a significant spatial influence in variations of the economic activity inside european countries. In line with related empirical research done using cross-section models, results still show evidence of a significant and positive spatial influence, although with a different magnitude in this relation.

Keywords:

NEG, Market Potential, Wage Equation, Panel Data, Spatial Econometrics

Índice de contenidos

1. Introducción	5
2. Marco Teórico	7
2.1. Un acercamiento a la Nueva Geografía Económica	7
2.2. El modelo básico de la NGE	10
2.2.1. Estructura del modelo	10
2.2.2. La demanda	11
2.2.3. La oferta	13
2.3. Equilibrio y dinámicas del modelo NGE	15
2.3.1. El equilibrio a corto plazo	15
2.3.2. El equilibrio a largo plazo	17
2.3.3. Dinámicas de localización	18
2.4. La ecuación salarial	19
3. Marco Econométrico	21
3.1. El modelo de panel con efectos individuales y componente espacial	21
3.1.1. El modelo de panel con efectos fijos	21
3.1.2. Los modelos SAR y SEM y su extensión a modelos de panel	23
3.1.3. La matriz de pesos espaciales: W	25
3.2. Revisión de la literatura empírica de la NGE	28
4. Metodología	30
4.1. Especificaciones econométricas	30
4.2. Tests	32
4.3. El problema de la endogeneidad en la estimación	33
5. Datos	35
6. Resultados	39
7. Conclusiones	43
Bibliografía	48

Anexo	49
A. Metodología de los tests empleados	49
B. Estimación MCO por variables instrumentales	52
C. Resultados del análisis de robustez	53

Índice de figuras

3.1. Mapa ilustrativo de la matriz W	26
5.1. Mapa del VAB per cápita y Potencial Mercado en Europa en el año 2018	37

Índice de cuadros

3.1. Ejemplo ilustrativo de la matriz W	27
3.2. Resumen de artículos donde se utiliza un modelo de datos de panel para estimar el modelo NGE	29
5.1. Análisis descriptivo de las variables	35
5.2. Correlaciones de las variables estudiadas	36
6.1. Resultados de las estimaciones MCO	40
6.2. Diagnóstico del test Baltagi-Song-Koh	40
6.3. Test del Multiplicador de Lagrange	40
6.4. Estimaciones SAR y SEM de panel con efectos fijos	41
6.5. Impactos de las estimaciones SAR	42

Capítulo 1

Introducción

Un fenómeno comúnmente asociado a la industrialización de una economía es la concentración espacial de la actividad económica, que normalmente se asocia al aprovechamiento de las economías de escala internas y externas, causando procesos como la movilización del factor trabajo desde lo rural hacia las ciudades, entre otros. En una parte de la literatura económica se ha buscado una explicación espacial a dicha concentración: esto es, que la cercanía entre los agentes económicos puede ser un factor relevante a la hora de explicar las variaciones en la actividad económica de las zonas (sean ciudades, regiones, países, etc...) en las que se encuentran. La localización de un mercado es relevante para la renta no sólo por las características físicas de la zona, como por ejemplo el clima o la orografía (Roback, 1982), sino también por la proximidad a otros mercados (Harris, 1954), idea que se traduce en que la actividad económica de una zona concreta tiende a correlacionarse positivamente con la misma actividad de otras zonas cercanas, siendo esta correlación menor con lugares más distantes, a priori por el aumento de los costes de transporte en los que se incurren para el comercio (Krugman, 1980). Esta última relación parece evidente, por ejemplo, en los Estados Unidos, donde el Producto Interior Bruto es más elevado en zonas próximas a la ciudad de Nueva York, mientras que el mismo indicador es cada vez más bajo a medida que nos fijamos en regiones más alejadas, lo que se suele denominar como un patrón *Centro-Periferia* (Prebisch, 1959). Uno de los campos de estudio que se han ocupado de esta cuestión, que además será el marco en el que este trabajo se centrará, es la *Nueva Geografía Económica* (o NGE), que sintetiza elementos tanto de la Economía Regional como de la Teoría del Comercio Internacional para establecer un modelo general que permite analizar la influencia espacial en la actividad económica tanto a nivel interno de la empresa como entre distintos mercados (Krugman, 1991). La NGE se ha utilizado en el ámbito empírico para analizar la influencia espacial de los mercados sobre la actividad económica, a mediante la estimación de una *ecuación salarial*, que explica las variaciones del salario de distintas regiones en base a el poder de compra de un conjunto de las mismas (Hanson, 2005). Este estudio se ha llevado a cabo en distintos ámbitos geográficos, como es a nivel mundial (Redding y Venables, 2004), a nivel de Estados Unidos (Hanson, 2005) o en la Unión Europea (Breinlich, 2006; Bruna et al., 2016), en

general encontrando evidencias significativas de la existencia de un componente espacial derivado de la cercanía de los mercados a la hora de explicar las variaciones en la actividad económica.

Gran parte de la literatura empírica de la NGE se ha llevado a cabo utilizando modelos diseñados para datos de sección cruzada y sin componentes espaciales. En términos econométricos, este hecho implica que la violación de la hipótesis de Mínimos Cuadrados Ordinarios puede proceder de tres vías distintas: la primera, a través de las características fijas de las regiones de estudio, que pueden generar estimaciones sesgadas y endogeneidad si las mismas están correlacionadas con la variable explicativa (Wooldridge, 2010), la segunda, en base a una correlación de la variable dependiente entre regiones cercanas, produciendo inconsistencia en la estimación del modelo, y la tercera, la existencia de una correlación espacial entre variables omitidas, que puede derivar en estimaciones ineficientes (Baltagi, 2005). La primera vía puede ser tratada mediante la estimación de un modelo de panel con efectos fijos, que elimina dichos efectos de la estimación, mientras que los problemas de la segunda y tercera vía pueden ser atenuados con la estimación de los modelos *Spatial Auto-Regressive (SAR)* y *Spatial Error Model (SEM)*, respectivamente. Adicionalmente, el uso de un SAR para estimar la ecuación salarial permitiría no sólo capturar el patrón Centro-Periferia de la actividad económica en la Unión Europea, sino también correlaciones de la variable dependiente entre regiones a una distancia más cercana (Bruna et al., 2015). Hasta la fecha, no se ha realizado una investigación de la ecuación salarial en el ámbito geográfico europeo que utilice técnicas de datos de panel en combinación con un componente espacial. Uno de los objetivos de este trabajo es, dentro del marco de la NGE, estudiar la influencia espacial en la actividad económica europea con el uso de un modelo de datos de panel por efectos fijos en combinación con dos modelos espaciales, SAR y SEM, utilizando un conjunto de datos de panel de 30 años. Nuestra hipótesis inicial no sólo es que el coeficiente estimado de la influencia espacial sobre las variaciones de la renta sea positiva y significativa, sino que la magnitud estimada del efecto será diferente de la obtenida por otros trabajos también realizados en el ámbito geográfico europeo (Breinlich, 2006; Bruna et al., 2016), debido a las razones econométricas aquí expuestas.

Nuestro trabajo se estructura como sigue: empezaremos con una introducción a la teoría NGE, donde explicaremos sus precedentes históricos, la estructura del modelo y de su ecuación salarial, para a continuación estimar dicha ecuación en un conjunto de regiones NUTS2 dentro de Europa en el periodo temporal 1980-2015, con la finalidad de comprobar si la cercanía entre sí de los distintos mercados que conforman la Unión Europea es importante para la actividad económica de las mismas.

Capítulo 2

Marco Teórico

2.1. Un acercamiento a la Nueva Geografía Económica

El campo denominado como *Nueva Geografía Económica* (NGE) hace referencia a una de varias ramas de la economía que busca comprender por qué la actividad económica tiende a concentrarse en determinados puntos en el mundo. En contraposición a otras teorías, como la Teoría del Comercio Internacional, donde se estudia la interacción entre agentes económicos sin recoger de forma alguna en su marco teórico la dimensión espacial donde surgen dichas interacciones, la NGE presenta la localización de los mercados y los agentes económicos como un factor importante a la hora de explicar las variaciones en la actividad económica. Un ejemplo de ello es el siguiente: En el año 2005, Alemania destinó la mayor parte de sus exportaciones a Francia, un 14% más que a Estados Unidos, a pesar de que este último generó un PIB diez veces mayor que el país vecino (Brakman, 2009). Este ejemplo pone de manifiesto que la distancia entre los distintos mercados importa a la hora de realizar transacciones, distancias que en la literatura han sido aproximadas mediante los costes de transporte, como veremos en este apartado más adelante.

Aunque se empezó a utilizar el término *Nueva Geografía Económica* para nombrar este campo, especialmente después de la publicación del trabajo de Krugman (1991), la NGE no es exactamente ni una teoría "nueva", ni Krugman fue su pionero. La NGE bebe en realidad de distintas ramas de la economía que también introdujeron el factor espacial en sus desarrollos teóricos, siendo las ramas más importantes la *Economía Regional* (a la cual pertenece la *Teoría de la Localización*, de la cual hablaremos a continuación) y la *Teoría del Comercio Internacional*. La introducción de ambas teorías permitirá al lector entender la importancia de la Nueva Geografía Económica en el ámbito de la economía espacial.

La primera de estas teorías, la Teoría de la Localización, data del siglo XIX, desde el trabajo de Von Thünen (1826), donde propuso un conocido modelo para explicar el uso de la tierra para la agricultura, de la cual se puede extraer la idea de unos costes de transporte que son más crecientes cuanto más lejos esté la producción de un cultivo del centro urbano donde se pretende vender. Dos hechos estilizados se pueden extraer a partir del uso de este modelo: primero, la densidad poblacional decrece a medida que nos alejamos de un centro urbano, y segundo, las principales capitales económicas occidentales han descentralizado su producción en el siglo XX, probablemente debido a una reducción de los costes de transporte (Anas et al., 1998). A pesar de estos resultados, el modelo de Von Thünen contenía algunas limitaciones, como el hecho de que no contabiliza rendimientos a escala crecientes (tanto a nivel de empresa como a nivel de industria) o de que se apoya en la existencia de un único centro urbano, sin que se generen interacciones de este con otros centros similares. Estos problemas son tratados por el modelo de Henderson (1974), donde introduce interacciones entre distintos centros y rendimientos a escala crecientes en su teoría, además que en su trabajo realiza una clara distinción entre dos tipos de fuerzas: Las fuerzas de aglomeración y las fuerzas de dispersión. Un ejemplo de la primera serían las economías de escala externas (también denominadas *derrames*) específicas de un sector industrial, en la que todas las empresa de un mismo sector industrial se benefician mutuamente de varios desarrollos (tecnológicos, educación, etc...), mientras que como ejemplos de fuerza de dispersión se nombran los costes de congestión, que implican costes tales como las rentas de la tierra o los costes de desplazamiento inter-urbano, que incentivarían a las empresas a deslocalizarse en otros sitios. En general, los principales aportes de la Teoría de la Localización al análisis económico espacial pueden resumirse en los siguientes puntos según Ottaviano (2004):

- El espacio económico es el resultado de un *trade-off* entre varias formas de rendimientos crecientes y diferentes tipos de costes de movilidad.
- La competición de precios, elevados costes de transporte y el uso de la tierra promueve la dispersión de la producción y el consumo; por lo tanto, las empresas son más propensas a aglomerarse dentro de áreas metropolitanas grandes cuando venden productos diferenciados y los costes de transporte son bajos.
- Las ciudades proveen un amplio abanico de bienes y un mercado de mano de obra especializada que hacen a las propias ciudades más atractivas de cara a consumidores y trabajadores.
- Las aglomeraciones son el resultado de procesos acumulativos que conciernen tanto el lado de la oferta como de la demanda.

Como consecuencia, el espacio económico ha de ser entendido como el resultado de la interacción entre las fuerzas de aglomeración y dispersión. Todos los puntos aquí resumidos se encuentran comprendidos en el campo de la entonces denominada *Geografía Económica*¹, aunque aún faltaba

¹ Los autores de la NGE normalmente utilizan Nueva Geografía Económica o Geografía Económica indistintamente, refiriéndose al mismo campo. Para simplificar, en este trabajo nos referiremos a ella como Nueva Geografía Económica o NGE.

por desarrollar una herramienta que permitiese estudiar de manera empírica todas las relaciones expuestas en esta rama de conocimiento: es aquí cuando Krugman (1991) presenta un modelo de equilibrio general, que recoge elementos desarrollados en la Geografía Económica hasta la fecha entonces y, en base a su trabajo previo realizado sobre Teoría del Comercio Internacional (Krugman, 1979; Krugman, 1980)², aumentando así la relevancia de la Geografía Económica en la economía *mainstream* y concediéndole el título de “Nueva” para así convertirse en la NGE.

La segunda teoría importante de la que bebe la NGE, la Teoría del Comercio Internacional, fue fundamental para resolver dos cuestiones que son cruciales para entender el modelo base: La cuestión de los rendimientos crecientes a escala y los costes de transporte. Establecer un modelo donde existen rendimientos crecientes a escala es incompatible con establecer un mercado de competencia perfecta, pues lo primero implica una situación en la que los ingresos marginales son superiores a los costes medios, por lo que se necesita plantear un tipo de competencia imperfecta. Para este fin, en la NGE se utiliza el modelo de competencia imperfecta Dixit-Stiglitz (Dixit y Stiglitz, 1977), que fue muy popular en diferentes ámbitos de la economía debido a su capacidad de combinar tanto economías de escala internas a nivel de empresa en conjunto con una estructura de mercado de competencia monopolística, puesto que este tipo de estructura era muy frecuente en el ámbito empírico y permitía analizar esta realidad de manera más fiel que utilizando herramientas de análisis neoclásicas.

Aunque el significado de los costes de transporte en un entorno de economía espacial es bien sabido desde el modelo de Von Thünen, existe una incógnita que no se ha tratado hasta ahora en la teoría que hemos expuesto: el transporte es un sector económico por su propia cuenta, lo cual significa que este *coste* es en realidad un beneficio para otro agente (en concreto, los que trabajan en el sector del transporte), por lo que un modelo teórico que intente explicar algún aspecto de la economía espacial debería introducir de algún modo el sector del transporte y sus agentes implicados. Para evitar este tipo de complicación, Samuelson (1952) ideó los denominados costes de transporte tipo *iceberg*, que introduce la idea de que un bien transportado de una región a otra se *derrite* en una proporción durante todo el trayecto, siendo esta parte del producto que se pierde el coste de transporte en cuestión. Esencialmente, este tipo de coste traduce los costes de transporte en unidades adicionales de un mismo producto, de tal forma que si un agente desea vender una unidad de un producto en otro mercado, deberá producir más de esa unidad para cubrir la proporción que se perderá en su venta (Por ejemplo, si al transportar un producto de una región A a una región B se espera que un 12% del producto se echará a perder, el productor deberá en realidad enviar 1.12 unidades del producto, para que así llegue una unidad entera a su destino).

En resumen, las claves importantes que la NGE ofrece al análisis económico espacial pueden resumirse en los siguientes puntos, expuestos por Head y Mayer (2004):

- Rendimientos a escala crecientes e internos a nivel de empresa

² Estos dos trabajos, junto con Krugman (1991), le valieron a Paul Krugman para conseguir el Premio Nobel de Economía en el año 2008.

-
- Competencia imperfecta
 - Costes de transporte positivos
 - Localización endógena de empresas
 - Localización endógena de la demanda, a través o bien de la movilidad de los trabajadores o de las empresas utilizando el output de su sector como inputs intermedios.

Los cuatro primeros puntos se han desarrollado desde la perspectiva de la Nueva Teoría Comercial, mientras que es el quinto punto la novedad que aporta la Nueva Geografía Económica desde Krugman (1991). Sin este último punto, el establecimiento de condiciones simétricas para las dos regiones en el modelo de la Teoría Económica Internacional de Krugman (1980) lleva a resultados simétricos, no dando lugar en ningún caso a la posibilidad que la actividad económica se aglomere en una única región. Este último punto es el que da inicio a un proceso denominado como *causalidad circular*, y que explicaremos más adelante.

2.2. El modelo básico de la NGE

La finalidad de este apartado es explicar el modelo de equilibrio general de Krugman (1991) de la NGE. Este modelo nos permitirá entender los intercambios comerciales entre sectores económicos de dos regiones distintas y, a su vez, los posibles movimientos de sus agentes económicos entre ambas regiones según distintas fuerzas. En primer lugar, detallamos la estructura básica del modelo con un ejemplo utilizado por Hoang (2019) y Brakman (2009). En segundo lugar, entramos más en detalle, describiendo las fórmulas que definen el lado de la oferta y de la demanda en el modelo. En tercer lugar, explicamos los equilibrios que se dan tanto a corto como a largo plazo. Finalmente, detallamos qué clase de fuerzas juegan en el movimiento de los agentes económicos entre las regiones, que es el valor añadido característico de la NGE al campo de la economía espacial.

2.2.1. Estructura del modelo

Asumimos que existen dos regiones simétricas, Norte y Sur, con dos agentes económicos distintos (empresas y trabajadores), donde sólo existen dos sectores económicos (uno manufacturero y otro del sector primario, la alimentación) y el único factor de producción utilizado por las empresas es el trabajo (l). Ambos sectores económicos están regidos por reglas distintas: mientras el sector de la alimentación está configurado como un mercado donde existen rendimientos a escala internos constantes (por lo que el mercado es de competencia perfecta), las empresas manufactureras

contienen rendimientos a escala internos crecientes, derivando en un mercado de competencia imperfecta, en concreto del tipo monopolístico Dixit-Stiglitz. A su vez, al ser monopolios, las empresas manufactureras producen N variedades de productos diferenciados, pero sólo una variedad por empresa. Adicionalmente, existen costes de transporte cuando únicamente se venden manufacturas de una región a otra, y no cuando la manufactura es vendida localmente, lo cual resulta en que dichos productos sean más caros en la región ajena.

En cuanto a la movilidad, sólo los trabajadores y las empresas del sector manufacturero pueden moverse entre las regiones Norte y Sur: en otras palabras, las empresas y trabajadores de la alimentación son inmóviles, y además existe siempre la misma cantidad tanto de trabajadores como de empresas de este sector en ambas regiones. La configuración de la movilidad permite que se den situaciones de equilibrio donde todas las empresas manufactureras se concentren en una única región, como parte del proceso de causación circular.

Los consumidores, que son a su vez trabajadores, utilizan todo su presupuesto tanto en manufacturas como en comida, y como este último proviene de un sector en competencia perfecta y sin costes de transporte asociados a su venta, le es indiferente comprarles en su región local o en la ajena. Los consumidores además tienen preferencia por la variedad, por lo que consumirán siempre al menos un producto de cada empresa manufacturera, sea en su propia región o importado a partir de la ajena.

En torno a esta configuración, lo que el modelo NGE explica es el movimiento de las empresas y trabajadores entre las regiones mencionadas según distintas variables, tales como el índice de precios (I), los costes de transporte (T), el salario (W) y la renta (Y), derivando en distintos equilibrios posibles tanto a corto como a largo plazo, que detallaremos después de describir el lado de la demanda y de la oferta.

2.2.2. La demanda

Los consumidores de ambas regiones tienen determinadas preferencias entre dos tipos de productos: alimentos y manufacturas. Entonces, su función de utilidad es:

$$U = F^{1-\delta} M^\delta \quad (2.1)$$

donde F es el consumo de alimentos, M el de manufacturas, y $0 < \delta < 1$ es la parte del presupuesto del consumidor destinada al consumo de manufacturas. Adicionalmente, el consumidor se rige por la siguiente restricción presupuestaria:

$$Y = F + I \times M \quad (2.2)$$

donde Y es el total de renta de los consumidores e I es el índice de precios de las manufacturas. Caben hacer dos aclaraciones respecto de este resultado: primero, como todos los trabajadores de ambas regiones son también la totalidad de los consumidores, la renta total (Y) de una región viene directamente determinada por el salario (W) de los consumidores de la misma región, y segundo, los precios de los dos tipos de bienes vienen determinados en número de alimentos, lo cual explica por qué tenemos un índice de precios para las manufacturas y no para los alimentos en la ecuación (2.2).

Dada la función de utilidad (2.1) y el presupuesto (2.2), el consumidor optimiza su cesta de la siguiente manera:

$$\begin{aligned} F &= (1 - \delta)Y \\ I \times M &= \delta Y \end{aligned} \quad (2.3)$$

por lo que el optimizar su utilidad implicará repartir el presupuesto entre los dos tipos distintos de bienes.

Al comenzar este apartado, señalamos que el sector manufacturero es de competencia monopolística, y de que cada empresa produce una variedad de producto distinta, para un total de N variedades (por lo que hay tantas empresas manufactureras como variedades). De la parte del presupuesto destinada a manufacturas, cabe aclarar en qué influye la existencia de estas variedades sobre la decisión de compra del consumidor. Definimos c_i como el nivel de consumo de un tipo de manufactura i , y establecemos el consumo agregado de las manufacturas M como sigue:

$$M = \left(\sum_{i=1}^N c_i^\rho \right)^{\frac{1}{\rho}} ; 0 < \rho < 1 \quad (2.4)$$

donde el parámetro ρ es el nivel de preferencia de los consumidores por la variedad de manufacturas. Al ser ρ menor que uno, se asegura que las distintas variedades sean sustitutos imperfectos las unas de otras, mientras que también necesitamos que ρ sea mayor que cero para que las variedades sean substitutas entre sí y no complementos según el modelo de monopolio propuesto. Por otro lado, suponiendo que $\rho = 1$, la ecuación (2.4) se simplificaría a $M = \sum_i c_i$ y para los consumidores no habría diferencia entre una variedad y otra (es decir, se vuelven sustitutos perfectos). Fijándonos de nuevo en la ecuación (2.4), se puede comprobar cómo el hecho de que ρ sea menor que uno, aumenta el consumo agregado de manufacturas M , *ceteris paribus*, por lo que los consumidores en

este modelo valoran positivamente la existencia de variedades.

Con el consumo agregado de manufacturas M , definimos ahora la restricción presupuestaria para esta ecuación. Siendo p_i el precio de una variedad i , para $i = 1, \dots, N$ variedades:

$$\sum_{i=1}^N p_i c_i = \delta Y \quad (2.5)$$

A partir de las ecuaciones (2.4) y (2.5), podemos optimizar el problema y obtener la demanda agregada de una variedad j :

$$c_j = p_j^{-\epsilon} I^{\epsilon-1} \delta Y \quad (2.6)$$

donde $I = \left(\sum_{i=1}^N p_i^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}}$ y $\epsilon = \frac{1}{1-\rho}$. A partir de la ecuación (2.6), podemos encontrar las siguientes relaciones: el consumo agregado de la variedad c_j se incrementa al aumentar los ingresos de la región (Y) y al aumentar el índice de precios de las restantes variedades (I), pero se reduce al aumentar el precio de la propia variedad j (p_j). Cabe mencionar la variable ϵ , interpretada como la elasticidad de la demanda por productos manufacturados, que varía de manera inversamente proporcional a ρ , por lo que cuanto mayor sea la preferencia por la variedad de los consumidores en una región, menor será el valor de ϵ .

2.2.3. La oferta

La descripción del lado de la oferta requiere que tratemos la estructura productiva del sector de la alimentación y del sector manufacturero. En el sector de la alimentación, como antes señalamos, las empresas trabajan con rendimientos a escala constantes y los trabajadores de este sector se asumen inmóviles (no pueden cambiar de región). De la fuerza total de trabajo entre ambos sectores económicos, L , una proporción $(1 - \gamma)$ trabaja en la alimentación, por lo que la producción de alimentos es equivalente a:

$$F = (1 - \gamma)L \quad (2.7)$$
$$0 < \gamma < 1$$

Para el sector manufacturero, donde existe competencia monopolística con rendimientos a escala crecientes, cada una de las empresas produce una única variedad. La fuerza de trabajo necesaria para producir una variedad i se define como:

$$l_i = a + \beta x_i \quad (2.8)$$

donde a y β son los costes fijos y variables, respectivamente, de la producción del producto manufacturado x_i . Debido a las economías de escala, cuanto mayor sea la producción de una variedad, menor será el coste medio por unidad, puesto que la misma cantidad de costes fijos (a) se reparten entre cada vez más unidades. Teniendo en cuenta la ecuación del input de trabajo (2.8), los beneficios de las empresas manufactureras vienen determinados por:

$$\pi = px - W(a + \beta x) \quad (2.9)$$

Respecto a la decisión de precios que las empresas manufactureras toman, el modelo Dixit-Stiglitz asume dos supuestos: primero, cada empresa conoce el comportamiento respecto a precios de las restantes empresas del sector (si una empresa cambia el precio de su producto, asume que las demás no harán ningún cambio en ese momento), y segundo, las empresas ignoran el efecto que tendrá el cambio del precio de su producto sobre el índice de precios I . La maximización de los beneficios (2.9) en competencia monopolística da lugar a que las empresas determinen el precio según un margen comercial sobre el coste del factor trabajo:

$$p = \left(\frac{\epsilon}{\epsilon - 1} \right) \beta W \quad (2.10)$$

La razón de la existencia de este margen sobre el coste del trabajo es porque la empresa decide aplicar un precio necesariamente mayor al coste marginal de la última unidad producida, siendo el margen mayor o menor dependiendo de la elasticidad de la demanda (ϵ). Este margen da lugar a la posibilidad de obtener beneficios positivos en el mercado del sector manufacturero. En base a esto, otras empresas externas manufactureras podrían moverse a una región con beneficios positivos, al menos hasta que los beneficios sean iguales a cero (recordemos por el anterior sub-apartado que los consumidores tienen un presupuesto limitado y las variedades son sustitutos imperfectos, por lo que los beneficios empresariales se anularán si los consumidores no pueden adquirir una variedad de manufactura adicional), lo cual implica que se da en el mercado una situación de equilibrio, donde ya no existen incentivos para que se desplacen más empresas a la región. Esta situación de equilibrio determina nuevos niveles para algunas de las variables que antes hemos mencionado:

$$l_i = a\epsilon \quad (2.11)$$

$$N = \gamma \frac{L}{l_i} = \gamma \frac{L}{a\epsilon} \quad (2.12)$$

$$x = \frac{a(\epsilon - 1)}{\beta} \quad (2.13)$$

donde, en equilibrio, (2.11) determina la cantidad de trabajo para producir una variedad i , (2.12) es el número de variedades producidas y 2.13 es el total de variedades manufacturadas por empresa. Fijándonos en las ecuaciones (2.12) y (2.13), vemos cómo la preferencia de los consumidores por la variedad (ϵ) influye de manera positiva en el número de variedades producidas y NGEativamente en la cantidad producida por empresa. Otro resultado interesante a partir de la ecuación (2.12) es el hecho de que el número de variedades (N) es directamente proporcional al número de trabajadores (L), resultado que también implica que el número de empresas manufactureras es proporcional a la cantidad de trabajadores, al producir cada empresa una variedad única.

2.3. Equilibrio y dinámicas del modelo NGE

En el anterior apartado (2.2.3) indicamos que en el modelo NGE se da una situación de equilibrio si en el sector manufacturero de una región entran tantas empresas como para que el beneficio de todas ellas sean cero. Una de las características más importantes del modelo NGE frente a otras teorías del ámbito de la economía espacial es la determinación de los movimientos de distintos factores según varios equilibrios. Los equilibrios espaciales pueden ser tanto a corto como a largo plazo. En este apartado, describiremos ambos tipos de equilibrio, empezando primero por las ecuaciones de los equilibrios a corto plazo. Seguiremos utilizando el ejemplo de las dos regiones que hemos empleado en los apartados 2.2.2 y 2.2.3.

2.3.1. El equilibrio a corto plazo

En el apartado 2.2.2 dijimos que toda la renta (Y) de una región viene determinada directamente por el salario de los trabajadores. Recogiendo la ecuación (2.2), podemos traducirla de tal forma que nos permita entender cómo se determina la renta en una situación de equilibrio. Esto es, para una región i , determinamos:

$$Y_i = \lambda_i W_i \gamma L + \phi_i (1 - \gamma) L \quad (2.14)$$

donde λ_i es el porcentaje de trabajadores que hay en el sector manufacturero de la región i , y ϕ_i es, para la misma región, los trabajadores del sector alimentario. En general, el primer término de la ecuación ($\lambda_i W_i \gamma L$) representa la renta para los trabajadores manufactureros, mientras que el segundo término es la renta para los trabajadores del sector alimentario. Nótese que sólo hay un salario para manufacturas (W), pues el salario de los trabajadores de alimentos es igual a uno, como antes establecimos en el apartado 2.2.2.

Cada región tiene su propio índice de precios (I), que entre otras variables, viene determinado por el poder adquisitivo de los trabajadores de todas las regiones y de los costes de transporte necesarios para vender el producto en una región ajena. Tenemos entonces que, en equilibrio, la región i tiene el índice de precios siguiente:

$$I_i = \left(\frac{\beta}{\rho} \right) \left(\frac{\gamma L}{a\epsilon} \right)^{\frac{1}{1-\epsilon}} [\lambda_i W_i^{1-\epsilon} + \lambda_j T^{1-\epsilon} W_j^{1-\epsilon}]^{\frac{1}{1-\epsilon}} \quad (2.15)$$

donde es posible ver que el índice de precios en la región i depende tanto del poder adquisitivo de los trabajadores en la propia región como en la región ajena, solo que en esta última influyen los costes de transporte (T) necesarios para vender el producto en aquella región. Tanto las ecuaciones (2.14) como (2.15) contienen el salario manufacturero W , pero no sabemos aún cómo se determina el mismo. Por ejemplo, para saber el salario de la región i , necesitamos conocer la suma de las demandas de las manufacturas en i desde todas las regiones. Una vez obtenida la suma de dichas demandas e igualándolas a la oferta de la región, podemos determinar el salario en equilibrio. Utilizando la ecuación del consumo agregado (2.6) para cada región, obtenemos la demanda para la manufactura x_i , que es como sigue:

$$x_i = (\delta \beta^{-\epsilon} \rho^\epsilon) (Y_i W_i^{-\epsilon} I_i^{\epsilon-1} + Y_j W_j^{-\epsilon} I_j^{\epsilon-1}) \quad (2.16)$$

donde es posible ver cómo la demanda de un bien manufacturado depende en la renta de ambas regiones, los índices de precio de cada región y de los costes de transporte. Una vez obtenida la demanda, al igualarla a la ecuación de oferta que hemos presentado en (2.13) y despejando por el salario, obtenemos lo siguiente para la región i :

$$W_i = \rho \beta^{-\rho} \left(\frac{\delta}{(\epsilon-1)a} \right)^{\frac{1}{\epsilon}} [Y_i I_i^{\epsilon-1} + Y_j T^{1-\epsilon} I_j^{\epsilon-1}]^{\frac{1}{\epsilon}} \quad (2.17)$$

donde se hace manifiesto que el salario de la región i (W_i) depende de la capacidad de compra de todas las regiones, ponderada para la región j según sus costes de transporte. El salario es

determinante para el atractivo de la región de cara a potenciales trabajadores que vienen de otras regiones. Esta ecuación tiene ciertas similitudes con la *ecuación salarial* basada en el *potencial de mercado*, que se ha utilizado en las aplicaciones empíricas de la NGE y de la que hablaremos en un apartado posterior.

2.3.2. El equilibrio a largo plazo

El equilibrio a largo plazo del modelo NGE es esencialmente el mismo que el de corto plazo, con una diferencia: los salarios *reales* de ambas regiones son iguales, entiendo el salario real como el salario nominal deflactado según el índice de precios. Por ejemplo, el salario real de la región i es:

$$w_i = W_i I_i^{-\delta} \quad (2.18)$$

por lo que basta que $w_i = w_j$ para que el equilibrio a largo plazo se produzca. Si los salarios reales de las dos regiones son idénticos, se desincentiva cualquier movilidad del factor trabajo entre regiones. La situación donde los salarios reales de las regiones se igualan no sucede en un único escenario, sino en tres distintos: primero, que todas las empresas manufactureras se encuentren en la región i (proceso denominado *aglomeración completa*), segundo, lo mismo pero para la región j , y tercero, que haya una división equitativa de las empresas manufactureras entre ambas regiones. Para el caso de las aglomeraciones completas, esto sucede cuando partimos de tener a todos los trabajadores de la manufactura en una única región. En cambio, para el caso del reparto equitativo de la actividad manufacturera, esta se da si partimos de un escenario donde existen el mismo número de trabajadores en ambas regiones (esto es, $\lambda_i = \lambda_j = 1/2$), lo cual implica que tanto la renta como los salarios sean también iguales en ambos lugares.

Los tres escenarios que acabamos de exponer pueden ayudarnos a introducir un concepto importante en la NGE, conocido como *causalidad circular* y del que hemos hablado brevemente en el apartado (2.1). Supongamos el siguiente caso: en la región i se encuentran el 60% de los trabajadores manufactureros ($\delta_i = 0,6$), mientras que en la región j están el 40% restantes. Al haber más trabajadores en la primera región, los salarios manufactureros (y por ende, la renta de la región) serán mayores, incentivando la llegada tanto de empresas como de trabajadores de este sector que provengan de afuera, hasta que se alcance la aglomeración completa en el lugar. Esta situación donde las empresas tienen más incentivos en relocalizarse en la región *donde más empresas se encuentran*, se conoce en la literatura de la NGE como *causalidad circular* o *causalidad acumulada*.

2.3.3. Dinámicas de localización

En el anterior apartado, introdujimos los detalles principales del modelo NGE, el cual incluye competencia imperfecta, rendimientos a escala crecientes, localización, costes de transporte y movilidad para el factor trabajo, el cual es el único factor de producción. Después, en el apartado del equilibrio espacial, describimos tres escenarios posibles según las dotaciones iniciales del factor de producción entre las distintas regiones. La determinación del equilibrio puede verse influenciada debido a la interacción de distintas fuerzas económicas, que afectan a la decisión de los agentes económicos a la hora de cambiar de localización. Estas fuerzas son: El efecto del índice de precios, el efecto *home market* y el efecto *extent-of-competition*. Las dos primeras fuerzas son *fuerzas de aglomeración* (incentivan a los agentes a concentrarse en un mismo lugar), mientras que la última es una *fuerza de dispersión* (incentiva a los agentes a abandonar su actual localización y desplazar su actividad a otro lugar). A continuación detallamos los tres efectos:

- **Efecto del índice de precios:** Cuantas más empresas manufactureras tiene una región, mayor es el número de variedades ofrecidas. Esto hace que la región en cuestión sea más atractiva para otras empresas, puesto que un gran número de variedades pueden ser compradas sin incurrir en costes de transporte si las empresas y trabajadores se desplazan aquí.
- **Efecto *home market*:** Introducido por Krugman (1980). El número de variedades de manufacturas producidas en una región tiene relación directa con la cantidad de factor trabajo en el sector de la manufactura. Concretamente, la región con un mercado más grande obtendrá una ganancia en variedades más que proporcional ante un aumento unitario del factor trabajo en las manufacturas, lo cual hace que esta región sea más atractiva que otras regiones para la atracción de nuevos agentes. Eventualmente, esto convierte a la región en un exportador neto de manufacturas hacia regiones más pequeñas que apenas tienen un sector manufacturero desarrollado.
- **Efecto *extent-of-competition*:** El índice de precios tiende a reducirse a medida que el mercado donde se encuentra se hace cada vez más grande. Esto tiene como consecuencia inmediata que la demanda de los bienes bajo este índice se reduce (ver ecuación 2.16), por lo que la posición competitiva de una empresa manufacturera en la región se ve menguada, incentivándola a salir del mercado.

En conclusión, la concentración de la actividad económica en el espacio es el resultado de la interacción entre estas fuerzas, que a su vez están determinadas por los costes de transporte. En la actualidad, dentro del contexto de la globalización, donde los costes de transporte son cada vez más reducidos, la diferencia en la velocidad de deterioro de las fuerzas de aglomeración frente a las de dispersión han dado lugar, en la práctica, a la existencia de múltiples núcleos de actividad económica frente a la existencia de un único punto de aglomeración (Baldwin et al., 2001; Puga, 1999), en contra de lo que puede inferirse a partir de los equilibrios con aglomeración completa expuestos en

este capítulo. No obstante, el modelo NGE es útil para entender este tipo de concentraciones de la actividad económica.

2.4. La ecuación salarial

En el apartado del desarrollo del modelo NGE, hemos comprobado cómo la demanda de una manufactura en una región está influenciada por el nivel de renta no sólo de la propia región sino de la región adyacente, ponderado por el índice de precios y los costes de transporte, que en el ejemplo que utilizamos era la región Sur frente a la región Norte. Aunque este es un resultado que puede deducirse del modelo de Krugman (1991), esta idea realmente data de décadas atrás, concretamente del trabajo de Harris (1954), donde el autor sostuvo que la demanda de los bienes producidos en un lugar es la suma de las capacidades de compra en otras localizaciones, ponderadas por los costes de transporte. Este concepto lo expresó el autor mediante una función denominada función de *Acceso de Mercado* (o también llamada *Potencial de Mercado* en la literatura). Lo que teóricamente resultó ser una función que permite explicar relaciones entre la actividad económica de un lugar y el nivel de renta de lugares cercanos, fue traducida para el uso empírico en Fujita (2001), que demostró que puede derivarse a partir de modelos espaciales econométricos. En este apartado introducimos la ecuación salarial y el núcleo de su ecuación, que es el Potencial de Mercado.

Es posible derivar la ecuación salarial en base a los resultados que hemos expuesto en el apartado (2.2). Para un conjunto de J regiones que se consideran vecinas o adyacentes entre sí³ y en situación de equilibrio, si igualamos la oferta (2.13) y la demanda (2.6) de una variedad j de un bien manufacturado, obtenemos:

$$\frac{a(\epsilon - 1)}{\beta} = p_j^{-\epsilon} T^{\epsilon-1} \delta Y \quad (2.19)$$

donde si además sustituimos p_j por el precio monopolístico (2.10), y despejamos de la igualdad el salario W_j , el resultado que obtenemos es:

$$W_j = \theta \left(\sum_{j=1}^J Y_j T_{ij}^{1-\epsilon} T_j^{\epsilon-1} \right)^{\frac{1}{\epsilon}} \quad (2.20)$$

donde θ engloba distintas variables como la elasticidad de la demanda. El salario de la región j

³ Sobre cómo se determinan qué regiones son vecinas o no, y qué significa dicha vecindad, se entrara más a fondo en el apartado 3.1.3

está, en términos de Harris (1954), determinado por la capacidad adquisitiva de los consumidores de un conjunto de J regiones en total, entre las que se encuentra la región j . Lo contenido dentro del paréntesis es el *Potencial de Mercado*, que antes definimos. Lo que la ecuación (2.20) nos indica es que el salario en una región j , *ceteris paribus*, aumenta en la misma dirección que el nivel de ingresos en la propia región (Y_j) y el nivel de precios de productos manufacturados (I_j), pero se reduce si se produce un aumento de los costes de transporte a la región j ($T_{ij}^{1-\sigma}$). Lo que esta ecuación explica es que cuanto mayor sea el Potencial de Mercado en una región, más atractiva será la región en cuestión para empresas extranjeras, incentivándolas a desplazar su producción a j , lo cual conllevará a un aumento de los salarios a corto-medio plazo debido al incremento de la demanda en dicha región (en otras palabras, la estimación de θ se espera positiva en la práctica).

La ecuación (2.20) suscita una cuestión importante desde el punto de vista del NGE, que es la influencia espacial de la variable salarial. El salario en una región puede influenciar los salarios de otras regiones cercanas, lo cual significa que la variable produciría un *efecto derrame* en múltiples regiones (puesto que, a su vez, los salarios de otras regiones también afectan a los salarios de regiones ajenas, y así sucesivamente). Esta última característica suele traducirse a la hora de estimar la ecuación en la aparición de una autocorrelación espacial en los residuos, y puede recogerse con un modelo econométrico espacial, que veremos más adelante.

Capítulo 3

Marco Econométrico

3.1. El modelo de panel con efectos individuales y componente espacial

En la econometría existen un conjunto de modelos que permiten realizar estimaciones aprovechando los denominados datos de panel, obteniendo así varias ventajas que con datos de sección cruzada no serían posibles conseguir. Algunas de estas ventajas fueron resumidas en el apartado (3.2). En este apartado presentamos, en primer lugar, el modelo de panel con efectos fijos, y a continuación una adaptación para dos modelos del ámbito de la econometría espacial.

3.1.1. El modelo de panel con efectos fijos

La principal diferencia de un modelo econométrico de panel de uno clásico, como el modelo de regresión lineal simple para sección cruzada, es el uso de dos subíndices para distinguir las observaciones de panel. Un ejemplo de modelo de panel simple sería:

$$\begin{aligned}y_{it} &= \alpha + X'_{it}\beta + u_{it} \\i &= 1, \dots, N; \\t &= 1, \dots, T\end{aligned}\tag{3.1}$$

donde y es la variable dependiente, X es el conjunto de regresores, N es el número de secciones cruzadas de nuestra muestra, T el número de periodos temporales, α es el término independiente,

β es el coeficiente del regresor y u_{it} es el término del error. De especial mención en la literatura de modelos de panel es el término de error u_{it} , puesto que una de las principales hipótesis detrás del uso de datos de panel es el hecho de que estos modelos pueden controlar por *efectos individuales*, que de otra forma no podrían ser recogidos por un modelo de sección cruzada y podría derivarse en una violación de la hipótesis MCO (por ejemplo, porque el regresor del modelo esté correlacionado con el término de error). Por lo tanto, el término de error de la ecuación (3.1) se suele modelar en los modelos de panel como:

$$u_{it} = \mu_i + v_{it} \quad (3.2)$$

donde μ_i es un efecto individual no observado y v_{it} es un término de perturbación. Nótese que el efecto individual no tiene subíndice temporal, por lo que se considera que es una característica de parte del individuo en la muestra que no varía en el tiempo y que no está observada (es decir, no forma parte de la fórmula propuesta inicialmente). Un ejemplo de este tipo de efectos suele darse en la literatura empírica de la NGE: se asume que las distintas regiones bajo estudio contienen ciertas características invariantes en el tiempo y no observadas (por ejemplo, la orografía). El problema con este tipo de efectos individuales es que pueden estar correlacionadas con los regresores del modelo propuesto, lo cual puede derivar en estimadores sesgados si utilizásemos el método de Mínimos Cuadrados Ordinarios. Una forma de tratar con este tipo de problemas es con el modelo de panel con efectos fijos.

El método econométrico de panel con efectos fijos utiliza una versión centralizada de las variables aleatorias empleadas para controlar por aquellos efectos individuales no observados que pueden estar correlacionados con los regresores del modelo. Pongamos por ejemplo una regresión simple con datos de panel, donde sustituimos (3.2) en (3.1):

$$y_{it} = \alpha + \beta x_{it} + \mu_i + v_{it} \quad (3.3)$$

y obtengamos la misma fórmula pero con la media de todas las variables empleadas:

$$\bar{y}_i = \alpha + \beta \bar{x}_i + \mu_i + \bar{v}_i. \quad (3.4)$$

donde μ_i , al ser una constante (efecto invariante en el tiempo), es equivalente a su propia media aritmética. Entonces, de la ecuación (3.3) extraemos (3.4) y obtenemos:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (v_{it} - \bar{v}_i) \quad (3.5)$$

que formalmente podríamos expresarlo en la siguiente forma reducida:

$$y_{it}^* = \beta x_{it}^* + v_{it}^* \quad (3.6)$$

donde todas las variables con asterisco son las variables aleatorias de (3.3) pero centralizadas mediante la sustracción de sus respectivas medias aritméticas según individuo (en otras palabras, cada media se hace sobre todos los periodos temporales para cada sección cruzada distinta del conjunto de datos de panel). La idea detrás de sustraer las medias aritméticas a las variables es que, si a un efecto fijo que no varía en el tiempo se le resta su media, el efecto individual se eliminará de hecho en la estimación, pues la media de una constante es igual a la propia constante. En esencia, la estimación del modelo de panel por efectos fijos es equivalente a realizar un Mínimos Cuadrados Ordinarios sobre las variables centralizadas.

Si el modelo (3.6) es el modelo verdadero que intentamos estimar, entonces sus coeficientes serán insesgados, siempre y cuando v_{it}^* siga una distribución Normal con media 0 y matriz de varianzas-covarianzas $\sigma_v^2 I_{NT}$. Adicionalmente, si se usa un periodo temporal lo suficientemente grande ($T \rightarrow \infty$), el estimador por efectos fijos de todas las variables, tanto de los regresores como de los efectos, será consistente.

3.1.2. Los modelos SAR y SEM y su extensión a modelos de panel

Antes indicamos que la violación de la hipótesis MCO puede provenir de una correlación de los regresores con una característica fija de los individuos (regiones, en nuestro trabajo) de una muestra de panel. Otra posible vía por la que la misma hipótesis puede no cumplirse es a través de la existencia de una correlación espacial entre las variables utilizadas para una ecuación a estimar. En el caso de la ecuación salarial de la NGE, por un lado, si consideramos que entre los salarios (la variable dependiente) de distintas regiones existe una correlación espacial, la omisión de este elemento en el modelo podría derivar en una estimación inconsistente de los coeficientes. Por el otro, si omitimos algún regresor que contenga también una autocorrelación espacial entre regiones, podrían generarse estimaciones ineficientes y sesgadas (LeSage y Pace, 2009). Existen dos modelos que recogen estos efectos de manera separada: estos son el SAR (*Spatial Auto-Regressive*) y el SEM (*Spatial Error Model*).¹ En este subapartado introduciremos ambos modelos para el uso de secciones cruzadas, y luego presentaremos una extensión de los mismos para su uso con modelos de panel con

¹ Existe un modelo espacial que combina los componentes SAR y SEM, denominado *Spatial Auto-regressive Combined* (SAC). Aunque este modelo presenta propiedades interesantes, decidimos descartarlo por varios motivos: en primer lugar, si en el modelo SAC las perturbaciones no están especificadas correctamente, la estimación del componente SEM puede distorsionar el cálculo de impactos del componente SAR (LeSage, 2014), y en segundo lugar, el Potencial de Mercado es de por sí como un lag espacial de la variable salario, por lo que introducir un componente tipo WX podría generar problemas de multicolinealidad y endogeneidad

efectos fijos.

El modelo SAR para sección cruzada lo definimos como sigue:

$$\begin{aligned} Y &= \lambda WY + X\beta + u \\ u &\sim N(0, \sigma^2 I) \end{aligned} \quad (3.7)$$

donde λ es el coeficiente de auto-regresión espacial, Y es la variable dependiente, X las variables independientes, W es una matriz de ponderaciones espaciales de dimensiones $N \times N$ (siendo N el número total de regiones en nuestro trabajo) y cuyos elementos diagonales son iguales a cero, λ es un parámetro que recoge la autocorrelación espacial de la variable dependiente, u es el término de error que recoge las desviaciones del modelo e I una matriz identidad de tamaño $N \times N$. En cambio, el modelo SEM se presenta de la siguiente manera:

$$\begin{aligned} Y &= X\beta + u \\ u &= \rho Wu + \epsilon \\ \epsilon &\sim N(0, \sigma^2 I) \end{aligned} \quad (3.8)$$

donde ρ es un término que recoge la autocorrelación espacial de las variables relevantes omitidas. De todos los componentes de los modelos descritos, probablemente W sea el más importante dentro de la econometría espacial, principalmente por la sensibilidad de los resultados de la estimación relacionada con la correcta elección de esta matriz, por lo que dedicaremos un apartado propio más adelante (véase sección 3.1.3).

Ahora que hemos introducido los modelos SAR y SEM para secciones cruzadas, podemos presentar dos modelos distintos de panel que recojan ambos componentes, en base a la ecuación (3.6). Para un modelo de panel con un componente SAR, tendríamos:

$$y_{it}^* = \lambda w_i y_{it}^* + \beta x_{it}^* + v_{it}^* \quad (3.9)$$

mientras que para un modelo con un componente SEM, obtendríamos:

$$\begin{aligned} y_{it}^* &= \beta x_{it}^* + v_{it}^* \\ v_{it}^* &= \rho w_i v_{it}^* + \epsilon_{it} \\ \epsilon &\sim N(0, \sigma^2 I) \end{aligned} \quad (3.10)$$

Los modelos SAR y SEM no se interpretan de la misma forma. Si bien la interpretación de los co-

eficientes del SEM es similar a la de un modelo de Mínimos Cuadrados Ordinarios (MCO), para el SAR es diferente: una variación en X tiene un *efecto directo* en la región i y un *efecto indirecto* en las restantes regiones j . Por lo tanto, para el caso del SAR sería erróneo interpretar la influencia de un regresor sobre la variable dependiente únicamente en base a la estimación $\hat{\beta}$. Estos efectos en su conjunto se denominan *impactos*. Los impactos en un modelo SAR con k coeficientes se representan por la matriz $(I - \lambda W)^{-1} \beta_k$, donde los efectos directos se encuentran en la diagonal de la matriz y los efectos indirectos en los elementos fuera de la diagonal. La utilidad que tiene el conocer los impactos de un regresor es la de comprender hasta qué punto las variaciones de dicho regresor en una región concreta afectan tanto a la propia región como a otras regiones consideradas vecinas. El concepto de vecindad o adyacencia compete al uso de una matriz de pesos espaciales, y se explicará a continuación en el siguiente apartado.

3.1.3. La matriz de pesos espaciales: W

Cuando una variable aleatoria recogida en un punto geográfico influye y es influida por la misma variable en al menos una localización distinta, decimos que existe una *autocorrelación espacial* en esta variable. En caso de existir este componente en un conjunto de variables, el uso de modelos tradicionales como el MCO puede dar lugar a estimaciones inconsistentes y/o ineficientes, por lo que se suele recurrir a un subconjunto de modelos pertenecientes a la denominada econometría espacial, que recogen este tipo de autocorrelaciones en caso de existir una. Una herramienta crucial para estos modelos es la matriz de pesos espaciales, comúnmente representada por la letra W , y cuya finalidad es reflejar la interdependencia espacial entre una o más variables, dependencia que a diferencia del caso de la autocorrelación temporal (la cual está siempre definida en una dirección), es una dependencia multi-direccional, puesto que es recurrente el hecho de que una variable de la región i afecte a la misma en j pero también en el sentido contrario.

Existen en general dos tipos de matrices W en la econometría espacial: de contigüidad, utilizada con muestras que contienen información de polígonos o regiones cuya relación espacial viene determinada por adyacencia a través de lados y/o vértices, y de distancia, que determina la interdependencia espacial entre dos regiones según una medida de distancia entre dos puntos. En el caso de la contigüidad, la matriz W busca, para una región dada, identificar qué otras regiones son vecinas o no de i , normalmente representado por una codificación binaria, y a partir de la definición de la matriz asignar unas ponderaciones a la variable aleatoria pertinente. En cambio, para el caso de la matriz W de distancias normalmente se determina que i está influida por todas las regiones de una muestra, pero dicha influencia es menor cuanto más alejados estén los puntos de dos regiones distintas. Los puntos en una matriz de distancia pueden ser de diferentes tipos: ciudades, capitales, parques, etc... En el caso de este trabajo, utilizaremos *centroides*, que es el punto central de una región calculado a partir de una serie de coordenadas.

Aunque hemos introducido dos tipos formales de matrices W espaciales, esta clasificación no es

exhaustiva, existiendo también otros tipos en la práctica. De hecho, en este trabajo utilizamos una matriz que combina ambos criterios, que es una matriz W de vecinos más próximos, identificando a varias regiones j como vecinas de una región i (criterio de contigüidad), pero no mediante el contacto directo entre lados de una región o polígono, sino mediante los k centroides ajenos más cercanos (criterio de distancia). En el Cuadro 3.1 puede verse un ejemplo de la selección de esta matriz para el caso de la comunidad autónoma de Castilla-La Mancha.

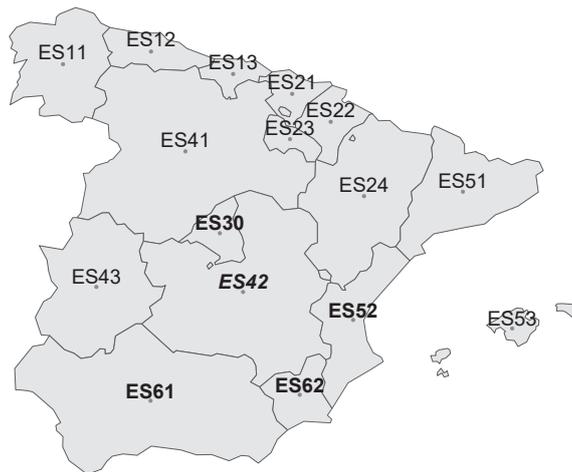


Figura 3.1: Ejemplo ilustrativo del método de ponderaciones mediante k vecinos próximos. Con la identificación ES42 en NGErita y cursiva se encuentra la región correspondiente a la Comunidad Autónoma de Castilla-La Mancha. Al crear una matriz W mediante 4 vecinos más cercanos, fueron seleccionadas como vecinos las 4 comunidades con el título en NGErita debido a la proximidad de sus centroides, con respecto de ES42, a pesar de estar Castilla-La Mancha rodeada por 7 regiones diferentes.

Respecto a la estructura de la matriz W , cabe señalar que sus elementos han de ser finitos y positivos. Es, además, simétrica, de tal forma que si por ejemplo trabajamos con N regiones, la matriz W tendrá dimensiones $N \times N$. En el caso de las matrices de contigüidad, los elementos de la misma suelen ser binarios, identificando con un 1 si una región j se considera vecina de otra región i , y 0 en caso contrario. Sin embargo, se suele normalizar en la práctica la matriz de tal forma que todos sus pesos sumen uno, pudiendo así interpretar el lag espacial de la variable dependiente como una media ponderada de la misma variable en otras regiones². Un ejemplo de una matriz de pesos espacial normalizada puede verse en el Cuadro 3.1. Dado una variable aleatoria Y que se encuentra registrada en n regiones diferentes, por lo que definimos un vector $y = (y_1, y_2, \dots, y_n)$, y para un subconjunto de j regiones vecinas, con la matriz W podemos crear los retardos espaciales de Y :

² Existen otras razones matemáticas detrás de la normalización de W , aunque no las comentamos puesto que escapan al enfoque de este trabajo

	ES11	ES12	ES13	ES21	ES22	ES23	ES24	ES30	ES41	ES42	ES43	ES51	ES52	ES53	ES61	ES62
ES13	0	0.25	0.00	0.25	0.00	0.25	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0	0.00	0.00
ES21	0	0.00	0.25	0.00	0.25	0.25	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0	0.00	0.00
ES22	0	0.00	0.25	0.25	0.00	0.25	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0	0.00	0.00
ES23	0	0.00	0.25	0.25	0.25	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0	0.00	0.00
ES24	0	0.00	0.00	0.25	0.25	0.25	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0	0.00	0.00
ES30	0	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.25	0.25	0.25	0.00	0.00	0	0.00	0.00
ES41	0	0.25	0.25	0.00	0.00	0.25	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0	0.00	0.00
ES42	0	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.25	0	0.25	0.25
ES52	0	0.00	0.00	0.00	0.00	0.00	0.25	0.25	0.00	0.25	0.00	0.00	0.00	0	0.00	0.25
ES53	0	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.25	0.25	0	0.00	0.25
ES62	0	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.25	0.00	0.00	0.25	0	0.25	0.00

Cuadro 3.1: Subconjunto de la matriz W mediante 4 vecinos próximos. Se han seleccionado comunidades autónomas españolas cuyos cuatro vecinos se encuentran dentro de España. En NGERita están seleccionadas las ponderaciones de Castilla-La Mancha (ES42) en referencia al mapa de la Figura 3.1. Todas las filas suman 1 al haberse normalizado la matriz.

$$W_y = \begin{pmatrix} \sum_j w_{1j}y_j \\ \sum_j w_{2j}y_j \\ \vdots \\ \sum_j w_{nj}y_j \end{pmatrix} \quad (3.11)$$

donde cada elemento del vector W_y es la media ponderada de los valores de la variable y en el subgrupo de las observaciones contiguas o vecinas.

El uso de la matriz W es un punto controvertido dentro de la econometría espacial. Una crítica recurrente respecto al uso de esta matriz es que su elección suele hacerse de manera arbitraria y por razones de eficiencia en las estimaciones antes que por justificaciones teóricas (Arbia y Fingleton, 2008). Adicionalmente, se ha considerado que las estimaciones dentro de este tipo de modelos son muy sensibles a la correcta especificación de dicha matriz, si bien esta afirmación puede haberse consolidado en la literatura debido a errores en la interpretación de modelos que la utilizan (LeSage y Pace, 2014). Otro detalle de interés es que para matrices W definidas mediante vecinos próximos existe una elevada correlación en los resultados para números de vecinos muy similares (por ejemplo, si estimamos dos ecuaciones de un modelo espacial con la misma especificación pero con matrices W de 7 y 8 vecinos, los resultados no variarán significativamente) (LeSage, 2014).

Para definir la matriz W existen varias recomendaciones en la literatura (véase por ejemplo Griffith, 1996; Bavaud, 2010; Stakhovych y Bijmolt, 2009), aunque destacaremos la propuesta de LeSage (2014) de utilizar una matriz dispersa (en otras palabras, con gran número de ceros). La razón detrás de esta recomendación es que si utilizásemos una matriz que para una región i se recogiesen todas las relaciones con las j regiones restantes (es decir, ninguna ponderación sería igual a cero), sería difícil para el modelo distinguir correctamente los derrames espaciales entre las variables, incluso aunque las ponderaciones se reduzcan de manera proporcional a la distancia. Esta recomendación se hace efectiva al elegir el método de vecinos más próximos, pues se introduce un umbral implícitamente al indicar que todas las ponderaciones de las regiones que no sean las k vecinas de una región i serán iguales a cero. Finalmente, decidimos elegir 4 vecinos próximos para comparar

nuestros resultados con los obtenidos por Bruna (2016), donde también utiliza el mismo número de vecinos.

3.2. Revisión de la literatura empírica de la NGE

La aplicación empírica de la NGE es un tema que ha sido ampliamente estudiado en los últimos años con el uso de herramientas econométricas, a raíz de la demostración de su derivación a partir de modelos espaciales en Fujita (2001). Se han realizado varios trabajos buscando una relación significativa y positiva entre el Potencial de Mercado y la variación de la actividad económica de un lugar: se encontraron evidencias de esta relación a nivel internacional (Redding y Venables, 2004) y a nivel de la Unión Europea (Gorter et al., 2005; Breinlich, 2006; Head y Mayer, 2006; Bruna et al., 2016). También se han encontrado las mismas relaciones significativas a nivel de país (por ejemplo, en Estados Unidos (Hanson, 2005), en Italia (Mion, 2004), en Bélgica (De Bruyne, 2009) y en Alemania (Brakman et al., 2004), entre otros) e incluso a nivel regional (Fingleton, 2010).

Cabe señalar que los estudios empíricos realizados dentro de la NGE que acabamos de mostrar se han hecho utilizando estimaciones con secciones cruzadas. En algunos casos, como es el trabajo de Bruna et al. (2016), se han utilizado modelos de sección cruzada pero con un componente espacial, como es el caso del SAR y SEM. En cuanto al uso de secciones cruzadas, la existencia de efectos fijos (orografía, proximidad a una costa, clima, etc...) en las localizaciones estudiadas tiene como consecuencia que las estimaciones realizadas con sección cruzada podría generar resultados sesgados (Elhorst, 2013). Una solución a este problema podría ser el uso de modelos econométricos espaciales de panel con efectos fijos, que controla por este tipo de características. En la literatura NGE empírica donde se han experimentado con este tipo de modelos se han encontrado, en general, las siguientes ventajas: la disponibilidad de una mayor fuente de información (debido a que suelen contener numerosas secciones cruzadas) (Baltagi, 2005), y la reducción del sesgo de las variables omitidas de los modelos (Fingleton, 2009). Desde un punto de vista teórico, este tipo de estimación también puede considerarse como una aproximación a un análisis causal en un entorno donde es casi imposible realizar experimentos controlados (Fingleton, 2010), como suele ser la norma en cualquier investigación económica. De nuevo, también se ha estimado la ecuación salarial utilizando un modelo de panel y con un componente espacial: en el trabajo de Hoang et al.(2019), utiliza un modelo con un componente SEM a la hora de estimar la ecuación salarial entre distintas prefecturas de Japón, obteniendo una estimación más eficiente que con MCO. En el Cuadro 2.1 podemos encontrar una tabla que resume varios de los artículos que han estimado modelos NGE con el uso de algún método de estimación de datos de panel.

Autor/es	Propósito	Variables	¿Por qué modelo de panel?
(Baltagi et al., 2014)	Testing de modelos de panel con lags espaciales y dinámicos con aplicación en la NGE. Uso de GMM con componente SAR. Como variable dependiente, utiliza VAB por trabajador.	Como potencial de mercado, utiliza VAB ponderado por distancias de gran círculo. Como instrumentos, utiliza variables tanto espaciales como no espaciales, que son lags espaciales de los regresores.	Una estimación MCO que no tenga en cuenta características fijas de cada región resulta en una estimación sesgada.
(Fingleton, 2009)	Con datos de UK, compara un modelo de la NGE con otro de Economía Urbana para relacionar salarios con Potencial de Mercado y densidad de empleo, respectivamente. Utiliza tres estimadores distintos de datos de panel, que son: Efectos Fijos, Efectos Aleatorios y Efectos Aleatorios con errores espacialmente correlacionados.	Para la variable dependiente se utiliza salario semanal bruto para trabajadores a tiempo completo en el periodo 1999-2003, mientras que para medir Potencial de Mercado utilizan VAB, costes de transporte e índices de precios. Para los instrumentos, utiliza un lag temporal de las variables regresoras de los modelos NGE y de Economía Urbana.	Por ventajas técnicas, tales como más información y posibilidad de modelar heterogeneidad entre individuos, eliminando sesgos de variables omitidas.
(Fingleton, 2010)	Se hace una revisión del uso del modelo de la NGE utilizando áreas pequeñas, comparando los resultados previos de distintos artículos en este ámbito y con artículos que utilizan modelos de Economía Urbana.	No se entra en detalle sobre de donde sacan los artículos las variable utilizadas para la estimación.	Ventajas teóricas y técnicas: Por un lado, a falta de poder hacer experimentos controlados para acercarnos a un análisis causal, el controlar por heterogeneidad con datos de panel facilita este objetivo. Por el otro, se pueden separar las heterogeneidades espaciales de los territorios del efecto del potencial de mercado con efectos fijos, además de controlar por una parte de la endogeneidad.
(Hoang et al., 2019)	Con datos de Vietnam por provincias, estimación de un modelo NGE con el uso del método GMM.	Para el Potencial de Mercado, utiliza la producción provincial per cápita (GPPC). Para la variable salario por provincia, utiliza el salario total partido por el número de trabajadores (AW). Como variable instrumental utiliza un lag temporal del Potencial de Mercado.	Por ventajas técnicas. El uso del método GMM permite controlar por endogeneidad.
(Wang & Haining, 2017)	Estimación del modelo NGE para una serie de prefecturas de Japón, utilizando un modelo de panel con efectos fijos mediante una herramienta denominada KKP, que contabiliza correlación temporal.	Para la variable dependiente, utiliza salarios nominales medios por prefectura extraídos directamente de una encuesta hecha por el Ministerio de Salud de Japón. Para el Potencial de Mercado, utiliza datos de ingresos totales por región, costes de transporte que aproxima por distancias y diferencial de precios al consumo entre prefecturas.	Ventajas técnicas. A diferencia de sección cruzada, los modelos de panel controlan por heterogeneidad y dependencia espacial. El estimador por efectos fijos estima efectos individuales no observados.

Cuadro 3.2: Resumen de artículos donde se utiliza un modelo de datos de panel para estimar el modelo NGE

Capítulo 4

Metodología

Como hemos señalado en el apartado (3.2), es posible estimar la ecuación salarial de la NGE utilizando métodos econométricos, derivados a partir de modelos espaciales. En este apartado presentaremos el procedimiento a seguir para las distintas estimaciones de la ecuación salarial. Todas las estimaciones se harán utilizando el software estadístico R (R Core Team, 2021), junto con un paquete especializado para la estimación de modelos de panel espaciales, desarrollado por Millo y Piras (2012). El paquete no sólo contiene formas para estimar los modelos, sino también incluye varios tests, de los cuales algunos son de nuestro interés y que explicaremos brevemente.

4.1. Especificaciones econométricas

Realizaremos primero la estimación MCO, que es como sigue:

$$\ln w_{it} = \beta_0 + \beta_1 \ln MP_{it} + \beta_2 \ln X_{it} + u_{it} \quad (4.1)$$

donde $u_{it} \sim N(0, \sigma^2 I)$, w_{it}^* es el salario nominal en una región i y en un periodo t en base al Valor Añadido Bruto per cápita de la región, MP_{it}^* es el *potencial de mercado*, X_{it}^* es una variable de control y u_{it} es un término residual. Específicamente, el potencial de mercado lo construimos de la siguiente manera:

$$MP_{it} = IMP_{it} + EMP_{it} = \frac{M_{it}}{d_{ii}} + \sum_{j \neq i}^{J-1} \frac{M_{jt}}{d_{ij}} \quad (4.2)$$

donde M_{it} y M_{jt} son la suma de los *mercados accesibles* para las regiones i e j , respectivamente, en el momento t , IMP_{it} es el *Potencial de Mercado Interno*, EMP_{it} es el *Potencial de Mercado Externo*, d_{ii} son las distancias intra-región en i y d_{ij} representa las distancias desde la región i con las regiones j .

Para estimar un modelo de panel con efectos fijos, seguimos a (Redding y Venables, 2004; Breinlich, 2006) para definir dos ecuaciones empíricas: una con un componente SAR y otro con un componente SEM. En el caso del SAR, tenemos:

$$\begin{aligned} \ln w_{it}^* &= \lambda W_i \ln w_{it}^* + \beta_1 \ln MP_{it}^* + \beta_2 \ln X_{it}^* + u_{it}^* \\ u_{it} &\sim N(0, \sigma^2 I) \end{aligned} \quad (4.3)$$

mientras que la especificación para el modelo SEM la presentamos como:

$$\begin{aligned} \ln w_{it}^* &= \beta_1 \ln MP_{it}^* + \beta_2 \ln X_{it}^* + u_{it}^* \\ u_{it}^* &= \rho W_i u_{it}^* + e_{it} \\ e_{it} &\sim N(0, \sigma^2 I) \end{aligned} \quad (4.4)$$

En este trabajo estimaremos cuatro especificaciones de las ecuaciones (4.1), (3.9) y (3.10). Las cuatro especificaciones se distinguen entre sí por dos clasificaciones: si se utiliza como regresor principal el Potencial de Mercado (MP) o el Potencial de Mercado Externo (EMP), y si se ha incluido (o no) la variable de control HDD. La primera clasificación es una forma de comprobar la robustez del modelo frente a la endogeneidad del *Potencial de Mercado Interno* (IMP), mientras que la segunda clasificación nos ayudará a reducir el sesgo en la estimación de la relación entre el Potencial de Mercado y el VAB per cápita.

La concentración espacial de ingresos puede deberse a factores ajenos al potencial de mercado, como por ejemplo las características climáticas locales de una región (Roback, 1982) o las diferencias entre regiones en cuanto a estructura económica (Rosés et al., 2010). Este hecho implica la posibilidad de que la estimación del Potencial de Mercado pueda sobrestimar o infraestimar la verdadera influencia de esta variable sobre el salario. Como tal, la inclusión de las denominadas *variables de control*, si bien no son relevantes para la ecuación teórica que estamos estimando, pueden ayudar a reducir este sesgo. Algunos ejemplos de variables de control utilizadas en la literatura aplicada de la NGE son el capital físico (Hoang et al., 2019), el capital humano (Bruna et al., 2016), determinantes de la eficiencia técnica (Redding y Venables, 2004), variables climatológicas (Roback, 1982), variables determinantes de la estructura económica de una región (Martinez-Galarraga et al., 2015), entre otras que son comúnmente utilizadas como determinantes fundamentales del nivel de renta entre regiones (Acemoglu et al., 2001; Hall y Jones, 1999). Una variable de control a destacar y utilizada por (Roback, 1982; Hanson, 2005) es la variable *Heating Degree Days* (HDD), que es un

indicador climático medido que calcula la necesidad de calor en los edificios de una región, teniendo en cuenta la temperatura ambiental y la temperatura media de las habitaciones de un edificio ¹. Esta variable entra dentro de un conjunto de *amenidades locales*, que son un conjunto de características físicas atractivas en una región a la hora de atraer agentes económicos para realizar sus actividades o vivir allí, y suelen ser un factor influyente a la hora de explicar diferencias salariales entre regiones (Roback, 1982). Por razones de disponibilidad muestral, utilizaremos únicamente HDD como variable de control.

A efectos de disponer de un *benchmark* que ilustre la importancia de emplear un modelo que recoja interdependencias espaciales entre las variables comentadas, comenzaremos en primer lugar con una estimación de *Mínimos Cuadrados Ordinarios* (MCO), y utilizaremos el resultado para comparar con lo que obtengamos con la estimación por Máxima Verosimilitud de los modelos SAR y SEM. Partimos con dos expectativas: primero, el Potencial de Mercado debería resultar en un coeficiente positivo y significativo, lo cual confirmaría la influencia que tiene el acceso a los mercados en el nivel de ingresos per cápita en cada una de las regiones, y segundo, que los modelos por Máxima Verosimilitud deberían poder recoger tanto una dependencia espacial local (que estaría representada por un coeficiente λ y ρ significativos) como de una tendencia global (a través de un efecto significativo del Potencial de Mercado), siguiendo a Bruna et al., (2016). En todo caso, los resultados irán acompañados de varios tests para detectar inconsistencia en las estimaciones y la existencia de componentes espaciales.

Por último, la parte empírica de este trabajo incluirá también un análisis de robustez de las estimaciones SAR y SEM, que incluirá estimaciones con las siguientes variables: uso de una matriz W de pesos normalizada de manera global (cada peso es $1 / N$ donde N es el número de regiones total), uso de la matriz W definida con 3 y 8 vecinos próximos, y la utilización de medias de 5 años en el conjunto de datos.

4.2. Tests

Las estimaciones que haremos de los modelos no estarán del todo completas si no las acompañamos con distintos tests que aumenten la validez de los resultados. Para este trabajo utilizaremos tres tests pensados para técnicas de datos de panel: el test *Baltagi-Song-Koh* (Baltagi et al., 2003), que permite detectar la existencia de autocorrelación espacial en el error, el test del Multiplicador de Lagrange (o test ML) para datos de panel (Anselin et al., 2008; Elhorst, 2013), que puede detectar tanto autocorrelación espacial en el error como en la variable dependiente, con formas tanto estándar (sólo detectan una u otra autocorrelación independientemente de la otra) como robustas (tratan de detectar una asumiendo la posibilidad de que exista la otra), por lo que su utilidad es la de ayudarnos

¹ Más información en <https://www.degreedays.net>

a justificar si existe un componente SAR o SEM en las estimaciones MCO, y el test de Hausman espacial de panel (Muhl y Pfaffermayr, 2011), que contrasta la consistencia de los estimadores bajo efectos fijos.²

4.3. El problema de la endogeneidad en la estimación

A la hora de estimar la ecuación (2.20), si la variable regresora X tiene correlación con el término de error u , decimos que estamos ante un problema de endogeneidad, lo cual puede producir estimaciones inconsistentes en el modelo. Este problema puede provenir de diferentes vías, de las cuales detallaremos algunas.

Una de las fuentes de endogeneidad puede ser la omisión de variables relevantes en el modelo. Si estamos interesados en estimar una regresión $E(y|x, q)$, donde y es la variable dependiente, x son variables observadas que incluimos y q son variables no observadas (de las cuales no disponemos de datos), no tendríamos más remedio que aproximar dicha regresión por $E(y|x)$. El problema es que si x y q están altamente correlacionadas, esto implicaría que habría una correlación entre x y u en la estimación de (2.20), generando así endogeneidad. Otra vía de endogeneidad puede ser la simultaneidad de las variables, que implica el hecho de que si bien y puede ser una función de x , también x puede llegar a ser una función de y (o al menos estar determinada parcialmente por y).

Entre las soluciones para la endogeneidad podemos encontrar el aplicar un modelo con *variables instrumentales*. Si por ejemplo un regresor x_k produce endogeneidad ($Cov(x_k|u) \neq 0$), se ha de estimar \hat{x}_k utilizando una variable exógena z que esté correlacionada parcialmente con x_k , pero no así con el error del modelo original ($Cov(z|u) = 0$). Esta variable se denomina *instrumento*³. Métodos econométricos como el *Mínimos Cuadrados de Dos Etapas* (2SLS) utilizan este tipo de variables para paliar la endogeneidad.

En el modelo propuesto (4.1) el problema de la correlación con el regresor puede ser atenuado aplicando efectos fijos a nuestro modelo, mientras que el problema de simultaneidad implica lo siguiente: un mayor potencial de mercado puede explicar un aumento de los ingresos salariales en una región, pero igualmente puede suceder que un mayor salario explique un mayor potencial de mercado, como explicamos en el capítulo 2. Para tratar este tipo de endogeneidad, podemos instrumentar el potencial de mercado con un lag temporal de la misma variable, tal y como han hecho autores como Hoang (2019) para otros contextos, asumiendo que el salario de una región en el momento t no tiene por

² Para más información acerca de estos tests, puede consultarse el apéndice (7)

³ Aunque es intuitivo considerar z como un instrumento, también suele denominarse como instrumentos a otros regresores del mismo modelo que sean exógenos (es decir, que no estén correlacionados con u , y que por ende pueden ser sus *propios* instrumentos en la estimación), por lo que es común que un paquete estadístico no solicite únicamente z como instrumento cuando se recurre a este método con más de un regresor disponible.

qué ser explicado directamente por el Potencial de Mercado en un periodo anterior. Si bien existen otros conocidos ejemplos de instrumentación del Potencial de Mercado en la literatura de la NGE, como la distancia a los tres mercados principales mundiales de Redding (2004), en su mayoría son instrumentos que dependen de distancias geográficas de puntos concretos y, por ende, no varían en el tiempo, lo cual hace imposible que los incluyamos de alguna manera en un modelo de panel con efectos fijos.

Finalmente, para el tratamiento de la endogeneidad en la estimación decidimos utilizar un lag temporal de un periodo sobre el Potencial de Mercado, que controlará por el problema de endogeneidad por simultaneidad, mientras que el caso de la endogeneidad por correlación con el término del error es tratado de manera implícita mediante el uso de efectos fijos en la estimación de modelos de panel, aunque este último no podremos utilizarlo en el caso de la estimación del modelo MCO.

Capítulo 5

Datos

Con la finalidad de estimar la ecuación salarial dentro del marco de la NGE, este trabajo presenta distintas muestras a nivel regional NUTS2 con periodo anual y entre los años 1980 y 2015, para un total de 189 regiones y 13 países, principalmente pertenecientes a Europa Occidental y del Norte. Por un lado, disponemos de las variables *Valor Añadido Bruto* (GVA) y el *Total de Población por región*, obtenidas de la base de datos de ARDECO ¹. Por el otro, utilizamos también la variable *HDD*, con el mismo periodo temporal que en las variables mencionadas anteriormente, descargada de la web Eurostat ². Una tabla descriptiva de todas estas variables puede verse en el Cuadro 5.1, mientras que en el Cuadro 5.2 tenemos las correlaciones. Por último, para el cálculo de la matriz de pesos W y del Potencial de Mercado externo, se han utilizado datos geométricos de las regiones europeas desde la base de datos GISCO y en formato NUTS 2016.

Para estimar la ecuación salarial, necesitamos antes construir sus variables. En primer lugar, la variable salario (w_{it}) se puede interpretar como *costes marginales* (Combes et al., 2008; Bruna

Estadístico	VAB	Población	VAB per capita	Potencial de Mercado	HDD
Mínimo	1.619e+08	22760	1523.295	1.244e+09	443.620
1er cuartil	1.288e+10	868147	12058.576	7.864e+09	2208.373
Mediana	2.648e+10	1462892	18907.151	1.274e+10	2775.505
Media	4.05e+10	2007167	19817.533	1.478e+10	2814.510
3er cuartil	4.928e+10	2507124	25759.393	2.007e+10	3296.760
Máximo	6.509e+11	12253906	59563.353	6.618e+10	7471.280

Cuadro 5.1: Estadísticos descriptivos de los datos de panel para el periodo 1980-2015

¹ ARDECO es una base de datos regional del área de Asuntos Económicos y Financieros de la Comisión Europea. Más información en <https://knowledge4policy.ec.europa.eu/territorial/ardeco-database.en>.

² Eurostat es una institución de la Comisión Europea cuyas principales labores son proporcionar información estadística a otras instituciones de la Unión Europea y armonizar métodos estadísticos entre los países miembros.

	VAB	Poblacion	VAB per cápita	Potencial de mercado	HDD
VAB	1.0000	0.8105	0.4562	0.4622	-0.1419
Poblacion		1.0000	0.0414	0.1343	-0.2869
VAB per cápita			1.0000	0.7516	0.1819
Potencial de mercado				1.0000	-0.0712
HDD				-	1.0000

Cuadro 5.2: Correlaciones entre las variables utilizadas en este trabajo para el periodo 1980-2015

et al., 2016), utilizando para su aproximación datos del *Valor Añadido Bruto per cápita* (VAB_{pc}), que conseguimos calcular dividiendo el Valor Añadido Bruto entre la población total de cada región. En segundo lugar, para estimar el *Potencial de Mercado*, definido por la ecuación (4.2), estimamos la suma de los mercados M utilizando el *Valor Añadido Bruto* (VAB) de las regiones pertinentes (esto es, dada una región i y las restantes regiones j , VAB_i se incluye en el componente interno y los VAB_j en el componente externo), y las distancias d_{ii} y d_{ij} se calculan de distintas maneras, que explicamos a continuación.

La distancia d_{ii} tiene como finalidad ponderar los costes de transporte en los mercados intra-región en la región i . Para su cálculo, el método estándar consiste en asumir que la región es circular y que d_{ii} sea un porcentaje del radio de dicho círculo. Esto es:

$$d_{ii} = z \times r_i \quad (5.1)$$

donde z es un porcentaje elegido por el usuario y r_i es el radio del círculo intra-región. Dado un r_i concreto, cuanto mayor sea z , menor será el peso del potencial de mercado interno en el cálculo de MP_{it} . Optamos por la elección de Keeble (1982) de $z = \frac{1}{3}$ para contabilizar la concentración de la actividad económica alrededor de los centroides en el entorno europeo. Conociendo la fórmula para el cálculo del área de un círculo, $a_i = \pi r_i^2$, y sustituyendo en (5.1), obtenemos:

$$d_{ii} = \frac{1}{3} \times r_i = \frac{1}{3} \times \sqrt{\frac{a_i}{\pi}} = 0,188 \times \sqrt{a_i} \quad (5.2)$$

por lo que d_{ii} se vuelve una función del área de una región i . Por otro lado, las distancias entre regiones, (d_{ij}), se calculan mediante *distancias de gran círculo*, que no es más que la distancia en kilómetros entre dos puntos distintos sobre una esfera (que en nuestro caso, serían las distancias entre los diferentes centroides en el mapa). Esta distancia tiene como finalidad ponderar la influencia de la accesibilidad a los mercados de las regiones j sobre la región i . Al estar incluida en un denominador de la ecuación (4.2), cuanto mayor sea la distancia entre las regiones i y j , menor va a ser la influencia de un mercado j sobre i , en coherencia con la teoría expuesta de la NGE.

Con la muestra recogida acerca del VAB, la población y los datos geométricos de las regiones, hemos

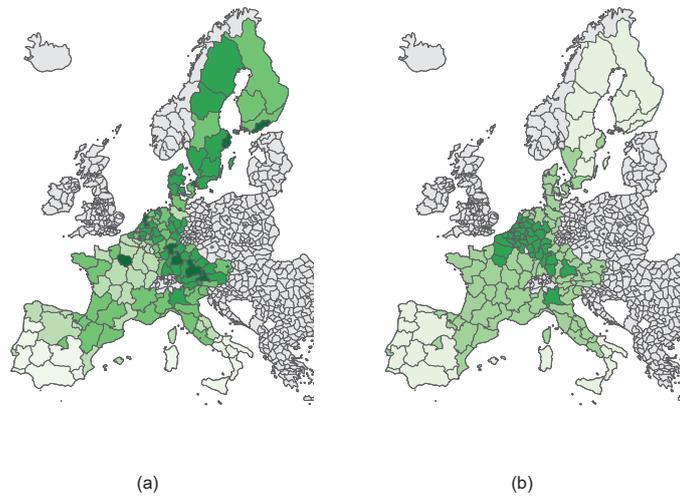


Figura 5.1: Representación de las variables VAB per cápita (a) y Potencial de Mercado (b) en logaritmos dentro de Europa en el año 2018. Las tonalidades más oscuras representan valores más elevados y viceversa para las más claras. Las regiones delimitadas son aquellas según el nivel 2 de la clasificación NUTS2016. Fuente: elaboración propia con datos de ARDECO.

podido calcular tanto el *VAB per cápita* como el potencial de mercado para cada una de las regiones NUTS2 en el periodo 1980-2015. En la Figura 5.1 puede observarse la distribución espacial del VAB per cápita y del Potencial de Mercado, respectivamente, para las regiones europeas del año 2018. En este último mapa se puede distinguir la denominada *banana azul* en Europa, que está conformada por las regiones con mayor actividad económica del continente, atravesando países como Alemania e Italia, y que dejan en manifiesto la existencia de la distribución Centro-Periferia de la actividad económica europea que antes señalábamos.

A efectos de estimar una ecuación con un modelo de panel de efectos fijos que explique la relación entre VAB per cápita y las variables regresoras a medio-largo plazo, se han realizado medias de todas estas variables en grupos de 3 años, de tal forma que dentro del periodo antes señalado de 1980-2015, este cálculo resulta en un total de 12 grupos. La decisión acerca del por qué realizar medias y de cuántos años se determina principalmente por dos motivos: primero, estimar por efectos fijos utilizando años de manera individual resultaría en estimaciones de las variaciones a corto plazo, por lo que se perdería la oportunidad de aprovechar el amplio periodo temporal del que disponemos para obtener una visual acerca de la relación definida por la ecuación salarial a un plazo más largo. Segundo, utilizar demasiados años para cada media calculada podría suponer un problema, pues un número reducido de periodos temporales puede derivar en estimaciones inconsistentes para los modelos de panel por efectos fijos (Elhorst, 2013).

A la hora de reunir un tamaño muestral suficiente para las variables de control, nos hemos encontrado con varias limitaciones. Aunque la estimación de panel por efectos fijos permite el uso de datos de panel no equilibrados (i.e. no tiene por qué haber el mismo número de observaciones muestrales para cada año estudiado), el método econométrico que vamos a utilizar de Millo y Piras (2012) sólo

funciona con paneles equilibrados, por lo que necesitaremos, para cada periodo, el mismo número de regiones a analizar. Dentro de la literatura empírica de la NGE es frecuente utilizar como control una variable que sea *proxy* del capital humano, puesto que las ganancias de productividad derivadas de este tipo de característica podrían explicar aumentos en la renta de un país. No obstante, para una estimación de panel por efectos fijos no hemos podido encontrar información suficiente de manera equilibrada y homogeneizada entre todas las regiones NUTS2 estudiadas. Si para las principales variables de la ecuación salarial, que son el VAB per cápita y el Potencial de Mercado, disponemos de información completa para 189 regiones NUTS2 diferentes, para la variable de control HDD únicamente dispondremos de 144 regiones, debido a que no se pudo obtener información de esta variable para las regiones de Reino Unido y Noruega. En base a esto, en la sección de estimación indicaremos esta diferencia en el número de observaciones para cada especificación estimada.

Finalmente, para el caso del uso de instrumentos, gran parte de la instrumentación utilizada en la literatura empírica de la NGE es de naturaleza invariante en el tiempo, como por ejemplo la distancia en kilómetros de una región a una capital económica como instrumento del Potencial de Mercado (Hanson, 2005), por lo que no es posible utilizarla para una estimación de panel por efectos fijos. Aunque finalmente utilizamos un instrumento mediante un lag temporal del Potencial de Mercado, tener más variedad de instrumentación para tratar la endogeneidad podría resultar en mejores estimaciones.

Capítulo 6

Resultados

En el Cuadro 6.1 pueden observarse las estimaciones MCO de panel de las cuatro especificaciones junto con algunas medidas de bondad del ajuste. Nótese que esta estimación se hace sobre la muestra con datos de panel, por lo que el modelo no hace distinciones entre distintas secciones cruzadas y/o periodos temporales. Los cuatro últimos modelos es la estimación de las especificaciones pero utilizando como instrumento del regresor principal (MP o EMP) un lag temporal del propio regresor en logaritmo. Tanto el Potencial de Mercado como el Potencial de Mercado externo se muestran significativos y con signo positivo, lo cual encaja con la teoría presentada en este trabajo. No obstante, puede verse como el test de Hausman de panel (*PHtest* en los cuadros) rechaza la consistencia de los estimadores en todas las especificaciones. Este problema de inconsistencia puede deberse no sólo a la endogeneidad generada por el Potencial de Mercado y sus componentes (Interno y Externo), sino también por un componente espacial en los residuos no recogidos por el modelo, como así parece indicar el test Baltagi-Song-Koh (Cuadro 6.2), que rechaza en todos los modelos la no existencia de residuos auto correlacionados en el espacio.

Utilizando un test del Multiplicador de Lagrange localmente robusto podemos determinar si la autocorrelación espacial que hemos detectado anteriormente proviene de la propia variable dependiente o del término de error. Los resultados pueden observarse en el Cuadro 5.3. Para las cuatro especificaciones mostradas en el Cuadro 6.1, utilizando un nivel de significación del 5%, todos los tests detectan lags espaciales tanto en la variable dependiente como en el componente del error, sea de manera dependiente (LME y LML) como introduciendo la posibilidad de que también exista el otro componente que no es el objeto del test (RLME y RLML).

En base a el diagnóstico LM, decidimos estimar un SAR y un SEM para cada una de las cuatro especificaciones (es decir, ocho modelos en total) por máxima verosimilitud. El resultado de las estimaciones puede observarse en el Cuadro 6.4. Para un nivel de significación del 5%, el test de Hausman de Panel espacial no rechaza la consistencia en la estimación de coeficientes de todas las

	(1)	(2)	(3)	(4)
(Intercept)	-4,60*** (0,26)	-4,40*** (0,29)	-7,47*** (0,32)	-6,53*** (0,35)
log(MP)	0,62*** (0,01)		0,64*** (0,01)	
log(EMP)		0,61*** (0,01)		0,61*** (0,01)
log(HDD)			0,30*** (0,02)	0,28*** (0,02)
R ²	0,57	0,51	0,64	0,57
Adj. R ²	0,57	0,51	0,64	0,56
Num. obs.	2268	2268	1728	1728
PWald-stat	2958,276	2339,375	3063,095	2244,08
PWald-pvalue	0,00	0,00	0,00	0,00
PHtest-stat	26,811	24,551	25,759	22,134
PHtest-pvalue	0,000	0,000	0,000	0,000
Variable de control	No	No	Sí	Sí

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 6.1: Estimaciones MCO. La variable dependiente de cada modelo es el Valor Añadido Bruto per cápita en logaritmo

Modelo	HDD	Estadístico	Valor P
Potencial de Mercado	No	55.01	0.00
Potencial de Mercado Externo		45.46	0.00
Potencial de Mercado	Si	49.40	0.00
Potencial de Mercado Externo		40.12	0.00

Cuadro 6.2: Diagnóstico Baltagi-Song-Koh para la detección de autocorrelación espacial.

Test	Modelo	Con HDD	Estadístico	P Valor
LME	Potencial de Mercado Externo	No	2066.589	0.000
	Potencial de Mercado	No	3026.008	0.000
	Potencial de Mercado Externo	Si	1609.300	0.000
	Potencial de Mercado	Si	2440.416	0.000
LML	Potencial de Mercado Externo	No	2046.217	0.000
	Potencial de Mercado	No	2102.599	0.000
	Potencial de Mercado Externo	Si	1770.804	0.000
	Potencial de Mercado	Si	1904.439	0.000
RLME	Potencial de Mercado Externo	No	37.734	0.000
	Potencial de Mercado	No	966.185	0.000
	Potencial de Mercado Externo	Si	9.201	0.002
	Potencial de Mercado	Si	541.234	0.000
RLML	Potencial de Mercado Externo	No	17.362	0.000
	Potencial de Mercado	No	42.777	0.000
	Potencial de Mercado Externo	Si	170.705	0.000
	Potencial de Mercado	Si	5.257	0.022

Cuadro 6.3: Diagnósticos del Multiplicador de Lagrange para la detección de lag espacial en la variable dependiente o autocorrelación espacial en el error

	<i>Spatial Auto-Regression (SAR)</i>				<i>Spatial Error Model (SEM)</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lambda	0,55*** (0,02)	0,65*** (0,02)	0,61*** (0,02)	0,69*** (0,02)				
rho					0,66*** (0,02)	0,62*** (0,02)	0,70*** (0,02)	0,67*** (0,02)
log(MP)	0,44*** (0,02)		0,39*** (0,02)		1,00*** (0,01)		1,02*** (0,01)	
log(EMP)		0,35*** (0,02)		0,31*** (0,02)		0,98*** (0,01)		0,99*** (0,01)
log(HDD)			0,09* (0,03)	0,06 (0,03)			0,23*** (0,06)	0,08 (0,07)
Num. obs.	2268	2268	1728	1728	2268	2268	1728	1728
SPHausman-stat	2.79	1.47	1.81	1.44	0.16	2.62	1.36	0.50
SPHausman-pvalue	0.10	0.22	0.40	0.49	0.69	0.11	0.51	0.78

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 6.4: Estimaciones SAR y SEM con efectos fijos. La variable dependiente es el Valor Añadido Bruto per cápita en logaritmo)

especificaciones, tanto las que incluyen la variable de control HDD como las que no. Tanto las variables *MP* como *EMP* se muestran significativas para un nivel de significación del 5 %, confirmándose así la teoría acerca de la influencia positiva del Potencial de Mercado sobre los ingresos per cápita de distintas regiones, además de que la estimación de la primera variable parece mostrarse robusta a la endogeneidad del Potencial de Mercado interno. Para las especificaciones SAR, el parámetro lambda (λ) es significativo en todos los modelos, lo cual confirmaría la existencia de un componente espacial autoregresivo en estas especificaciones, y también implica que las estimaciones MCO que antes hemos realizado son insesgadas e inconsistentes al no tener en cuenta esta característica. Por otro lado, en las estimaciones SEM el parámetro ρ es significativo, mostrando así indicios de interdependencia entre variables no observadas en las especificaciones, e implica que las estimaciones MCO que antes hemos realizado, aunque sean insesgadas, no son eficientes. En las estimaciones SEM (columnas (5) a (8)), los coeficientes tanto de $\log(MP)$ como de $\log(EMP)$ son cercanos a uno, lo cual implica que, teniendo en cuenta que todas las variables están en logaritmos, una variación en un uno por ciento de ambos regresores aumentaría, en término medio, otro uno por ciento los ingresos per cápita de la misma región. Finalmente, el uso de la variable de control *HDD* (especificaciones (3), (4), (7) y (8)) aumenta las magnitudes de los regresores del SAR y reduce las de los SEM entre una y dos décimas, probablemente debido a que la estimación sin control está sesgada.

Si bien los modelos SAR presentan significación tanto en *MP* como en *EMP*, la interpretación de la influencia del acceso de mercado sobre los ingresos per cápita ha de hacerse a través de los *impactos*, que a su vez están desglosados en un efecto directo y otro indirecto. En el Cuadro 6.5 podemos observar los impactos generados por las cuatro especificaciones SAR que hemos estimado, según si se ha empleado la variable de control HDD o no. En primer lugar, todos los efectos se muestran estadísticamente significativos para un nivel de significación del 5 %, utilizando una simulación Monte Carlo con 200 réplicas, y presentan signo positivo, en acorde con la teoría de la NGE. En segundo

lugar, los efectos indirectos en todas las especificaciones son mayores que los efectos directos, lo cual es intuitivo si tenemos en cuenta que los modelos SAR contabilizan el efecto de las variaciones del regresor en los ingresos per cápita de entre 144 y 188 regiones (dependiendo de si usamos HDD o no), frente al efecto directo que sólo contabiliza el efecto en la propia región. En tercer lugar, aquellos modelos que tienen como regresor el potencial de mercado externo muestran un mayor efecto indirecto que los modelos con el potencial de mercado completo, aunque los primeros generen un efecto total menor que los segundos, lo cual es intuitivo al haber prescindido del Potencial de Mercado interno en una de las especificaciones. Por último, el uso de la variable de control aumenta la magnitud de todos los efectos en general, pasando de un efecto total menor que la unidad en las especificaciones sin dicha variable, a superar la unidad una vez se incluye el control HDD, por lo que la influencia tanto directa como indirecta del potencial de mercado parece haberse infraestimado en las primeras especificaciones.

Las estimaciones SAR y SEM son sensibles al uso de una matriz W distinta. El uso de la matriz W normalizada globalmente vuelve inconsistentes, según el test de Hausman, las cuatro especificaciones SAR antes mencionadas y el modelo SEM con el Potencial de Mercado con el regresor, además de que todos los impactos dejan de ser significativos. En el segundo caso, en línea de lo que LeSage (2014) indicaba, el uso de 3 vecinos no varía la significación de los coeficientes y la consistencia de los modelos con respecto al uso de 4 vecinos, como igualmente sucede con el uso de 8 vecinos. Finalmente, el uso de medias de 5 años en los datos mantiene la significación de todos los coeficientes y resulta en consistencia en todas las especificaciones, excepto para el modelo SAR con el Potencial de Mercado como único regresor, problema que parece resolverse al incluir la variable HDD como control. Para más información acerca de estos resultados, pueden consultarse las tablas en el Anexo (apartado C).

Regresor	Con HDD	Tipo Impacto	Efecto	Valor P
Potencial de Mercado	No	Direct	0.469	0.000
		Indirect	0.515	0.000
		Total	0.985	0.000
Potencial de Mercado Externo		Direct	0.383	0.000
		Indirect	0.598	0.000
		Total	0.981	0.000
Potencial de Mercado	Sí	Direct	0.439	0.000
		Indirect	0.574	0.000
		Total	1.013	0.000
Potencial de Mercado Externo		Direct	0.363	0.000
		Indirect	0.640	0.000
		Total	1.004	0.000

Cuadro 6.5: Impactos generados por las cuatro especificaciones SAR. Los valores p generados en esta tabla se han calculado con una simulación Monte Carlo de 200 repeticiones.

Capítulo 7

Conclusiones

El objetivo de este trabajo es hacer una introducción a la teoría de la Nueva Geografía Económica, explicando su importancia en el ámbito general de la economía, para luego presentar un caso práctico donde estimamos el modelo base de la NGE con una muestra de datos de panel de un conjunto de regiones europeas, con la finalidad de comprobar si existe un componente espacial que explica parte de las variaciones de la renta en cada una de las regiones estudiadas. En la literatura empírica relacionada no se ha realizado hasta ahora una estimación de la ecuación salarial en el ámbito geográfico europeo utilizando un modelo de datos de panel, por lo que, en este aspecto, el valor añadido de este trabajo es ofrecer una estimación de la ecuación con esta técnica, contabilizando por efectos fijos, en combinación con dos modelos de la econometría espacial, SAR y SEM.

En coherencia con los resultados de la literatura empírica que han estimado la ecuación salarial con secciones cruzadas de Europa (Bruna et al., 2016, Breinlich, 2006), las estimaciones realizadas en el presente trabajo mediante la ecuación salarial confirman que existe una influencia espacial del acceso a los mercados en distintas regiones de Europa Occidental y Central sobre la actividad económica en las mismas, recogida a través de un coeficiente positivo y significativo del Potencial de Mercado. También hemos encontrado evidencias de la existencia de una interdependencia espacial entre distintas regiones para la renta per cápita, a través de los coeficientes significativos de los componentes SAR y SEM, resultado también alcanzado por Bruna et al. (2016), pero que en nuestro caso se confirma también para una muestra de datos de panel de 35 años. Tanto los impactos del SAR como las estimaciones del SEM resultaron en magnitudes positivas y cercanas a la unidad, lo cual es evidencia de que, teniendo en cuenta que hemos estimado por efectos fijos, las variaciones porcentuales del Potencial de Mercado son aproximadamente proporcionales a las variaciones porcentuales de los ingresos per cápita regionales. Adicionalmente, el análisis de robustez realizado demuestra que las estimaciones son sensibles a la especificación de la matriz de pesos W , donde las estimaciones se volvieron inconsistentes al utilizar una matriz globalmente normalizada, pero no así al número de vecinos empleados, donde al variar de 4 a 10 vecinos la significación de

las estimaciones no se han perdido, lo cual es coherente con lo expuesto por LeSage (2014).

El presente trabajo tiene dos limitaciones en cuanto a la estimación de la ecuación salarial: primero, en el uso de instrumentos para el tratamiento de la endogeneidad, y segundo, en la falta de más variables de control para la estimación. Debido a que gran parte de la literatura empírica de la NGE se ha realizado con el uso de secciones cruzadas, existen pocos precedentes a la hora de tomar como ejemplo para el uso de instrumentos y variables de control para la estimación de la ecuación salarial. Como consecuencia, no se han podido encontrar instrumentos o variables de control para las que pudiésemos disponer de una muestra de panel equilibrada y sin necesidad de excluir un número significativo de regiones.

Se plantean dos posibilidades para futuras investigaciones en esta línea: la primera, el uso de las mismas herramientas que en este trabajo pero con una base de datos más amplia, preferiblemente para obtener como variables de control *proxys* del capital físico y/o humano, que han sido frecuentemente utilizadas en la literatura NGE con secciones cruzadas, y la segunda, contemplar la estimación con el uso de un modelo espacial de panel dinámico. En el primer caso, se necesita más investigación en cuanto a la instrumentación del Potencial de Mercado en el uso de modelos de panel, concretamente el encontrar instrumentos que sean invariantes en el tiempo, y para las variables de control se necesita una base de datos homogénea para un gran número de regiones, como en Europa, y que ofrezcan suficiente información para estimar una variable que permita controlar por ganancias de productividad, como son el caso de las variables relacionadas con la eficiencia técnica. En referencia al uso de modelos espaciales de panel dinámicos, estos son útiles para controlar por posibles autocorrelaciones temporales, que pueden generar problemas de endogeneidad. Al depender estos modelos de lags temporales en todas sus variables, la existencia de información de varias décadas requerida para la estimación de la ecuación salarial tipo NGE en Europa con múltiples lags temporales es viable, por lo que cabría contemplar el uso de esta herramienta para el estudio de la influencia espacial de la actividad económica.

Bibliografía

- Acemoglu, D., Johnson, S. & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review*, 91(5), 1369-1401. <https://doi.org/10.1257/aer.91.5.1369>
- Anas, A., Arnott, R. & Small, K. A. (1998). Urban Spatial Structure. *Journal of Economic Literature*, 36(3), 1426-1464. <http://www.jstor.org/stable/2564805>
- Anselin, L., Gallo, J. L. & Jayet, H. (2008). Spatial Panel Econometrics, 625-660. https://doi.org/10.1007/978-3-540-75892-1_19
- Arbia, G. & Fingleton, B. (2008). New spatial econometric techniques and applications in regional science. *Papers in Regional Science*, 87(3), 311-317. <https://doi.org/10.1111/j.1435-5957.2008.00208.x>
- Baldwin, R. E., Forslid, R., Martin, P., Ottaviano, G. I. P. & Robert-Nicoud, F. (2001). The core-periphery model key features and effects. *The Monopolistic Competition Revolution in Retrospect* (pp. 213-235). Cambridge University Press. <https://doi.org/10.1017/CBO9780511492273.011>
- Baltagi, B. H. (2005). *Econometric analysis of panel data*. J. Wiley & Sons.
- Baltagi, B. H., Song, S. H. & Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117(1), 123-150. [https://doi.org/10.1016/S0304-4076\(03\)00120-9](https://doi.org/10.1016/S0304-4076(03)00120-9)
- Bavaud, F. (2010). Models for Spatial Weights: A Systematic Look. *Geographical Analysis*, 30(2), 153-171. <https://doi.org/10.1111/j.1538-4632.1998.tb00394.x>
- Brakman, S. (2009). *The new introduction to geographical economics*. Cambridge University Press.
- Brakman, S., Garretsen, H. & Schramm, M. (2004). The Spatial Distribution of Wages: Estimating the Helpman-Hanson Model for Germany. *Journal of Regional Science*, 44(3), 437-466. <https://doi.org/https://doi.org/10.1111/j.0022-4146.2004.00344.x>
- Breinlich, H. (2006). The spatial income structure in the European Union—what role for Economic Geography? *Journal of Economic Geography*, 6(5), 593-617. <https://doi.org/10.1093/jeg/ibl018>

-
- Bruna, F., da Coruña. E-mail, U. & f.bruna@udc.es. (2015). A GENERALIZED NEG WAGE-TYPE EQUATION.
- Bruna, F., Lopez-Rodriguez, J. & Faíña, A. (2016). Market Potential, Spatial Dependences and Spillovers in European Regions. *Regional Studies*, 50(9), 1551-1563. <https://doi.org/10.1080/00343404.2015.1048796>
- Combes, P.-P., Mayer, T. & Thisse, J.-F. (2008, 28 de septiembre). *Economic Geography: The Integration of Regions and Nations*. Princeton University Press. https://www.ebook.de/de/product/7491299/pierre_philippe_combes_thierry_mayer_jacques_francois_thisse_economic_geography.html
- De Bruyne, K. (2009). *Explaining the Location of Economic Activity. Is there a Spatial Employment Structure in Belgium?* (Working Papers N.º 2009/28). Hogeschool-Universiteit Brussel, Faculteit Economie en Management. <https://ideas.repec.org/p/hub/wpecon/200928.html>
- Dixit, A. K. & Stiglitz, J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review*, 67(3), 297-308. <http://www.jstor.org/stable/1831401>
- Elhorst, J. P. (2013, 30 de septiembre). *Spatial Econometrics*. Springer-Verlag GmbH. https://www.ebook.de/de/product/21989791/j_paul_elhorst_spatial_econometrics.html
- Fingleton, B. (2010). The empirical performance of the NEG with reference to small areas. *Journal of Economic Geography*, 11(2), 267-279. <https://doi.org/10.1093/jeg/lbq044>
- Fingleton, B. (2009). Testing The Neg Model : Further Evidence From Panel Data. *Region et Development*, 30, 141-158. <https://ideas.repec.org/a/tou/journal/v30y2009p141-158.html>
- Fujita, M., Krugman, P. & Venables, A. J. (2001, 27 de julio). *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press Ltd. https://www.ebook.de/de/product/3249077/masahisa_kyoto_univ_fujita_paul_cuny_krugman_anthony_j_venables_the_spatial_economy.html
- Gorter, J., van der Horst, A., Brakman, S., Garretsen, H. & Schram, M. (2005). *New economic geography, empirics, and regional policy* (CPB Special Publication 56.rdf). CPB Netherlands Bureau for Economic Policy Analysis. <https://ideas.repec.org/p/cpb/spcial/56.rdf.html>
- Griffith, A. (1996). Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. *Practical handbook of spatial statistics*.
- Hall, R. E. & Jones, C. I. (1999). *Fundamental Determinants of Output per Worker across Countries* (Working Papers N.º 97021). Stanford University, Department of Economics. <https://ideas.repec.org/p/wop/stanec/97021.html>
- Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics*, 67(1), 1-24. <https://doi.org/10.1016/j.jinteco.2004.09.008>
- Harris, C. D. (1954). The Market as a Factor in the Localization of Industry in the United States. *Annals of the Association of American Geographers*, 44(4), 315. <https://doi.org/10.2307/2561395>
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251. <https://doi.org/10.2307/1913827>

-
- Head, K. & Mayer, T. (2004). The empirics of agglomeration and trade. En J. V. Henderson & J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (pp. 2609-2669). Elsevier. <https://ideas.repec.org/h/eee/regchp/4-59.html>
- Head, K. & Mayer, T. (2006). Regional wage and employment responses to market potential in the EU. *Regional Science and Urban Economics*, 36(5), 573-594. <https://doi.org/10.1016/j.regsciurbeco.2006.06.002>
- Henderson, J. V. (1974). The Sizes and Types of Cities. *The American Economic Review*, 64(4), 640-656. <http://www.jstor.org/stable/1813316>
- Hoang, H. T., Huynh, L. T. D. & Chen, G. S. (2019). How New Economic Geography Explains Provincial Wage Disparities: Generalised Methods Of Moments Approach. *Economy of Region*, 15(1), 205-215. <https://doi.org/10.17059/2019-1-16>
- Keeble, D., Owens, P. L. & Thompson, C. (1982). Regional accessibility and economic potential in the European community. *Regional Studies*, 16(6), 419-432. <https://doi.org/10.1080/09595238200185421>
- Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *The American Economic Review*, 70(5), 950-959. <http://www.jstor.org/stable/1805774>
- Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3), 483-499. <https://doi.org/10.1086/261763>
- LeSage. (2014). What Regional Scientists Need to Know About Spatial Econometrics. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2420725>
- LeSage & Pace. (2009). *Introduction to spatial econometrics* [OCLC: 166358420]. CRC Press.
- LeSage & Pace. (2014). The Biggest Myth in Spatial Econometrics. *Econometrics*, 2(4), 217-249. <https://doi.org/10.3390/econometrics2040217>
- Martinez-Galarraga, J., Tirado, D. A. & González-Val, R. (2015). Market potential and regional economic growth in Spain (1860–1930). *European Review of Economic History*, 19(4), 335-358. <https://doi.org/10.1093/ereh/hev014>
- Millo, G. & Piras, G. (2012). splm : Spatial Panel Data Models in R. *Journal of Statistical Software*, 47(1). <https://doi.org/10.18637/jss.v047.i01>
- Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy. *Journal of Urban Economics*, 56(1), 97-118. <https://doi.org/10.1016/j.jue.2004.03.004>
- Moulton, B. & Randolph, W. C. (1989). Alternative Tests of the Error Components Model. *Econometrica*, 57(3), 685-93. <https://EconPapers.repec.org/RePEc:ecm:emetrp:v:57:y:1989:i:3:p:685-93>
- Mutl, J. & Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *The Econometrics Journal*, 14(1), 48-76. <https://doi.org/10.1111/j.1368-423X.2010.00325.x>

-
- Ottaviano, G. & Thisse, J.-F. (2004). Agglomeration and economic geography. En J. V. Henderson & J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics* (pp. 2563-2608). Elsevier. <https://ideas.repec.org/h/eee/regchp/4-58.html>
- Prebisch, R. (1959). Commercial Policy in the Underdeveloped Countries. *The American Economic Review*, 49(2), 251-273. <http://www.jstor.org/stable/1816120>
- Puga, D. (1999). The rise and fall of regional inequalities. *European Economic Review*, 43(2), 303-334. [https://doi.org/https://doi.org/10.1016/S0014-2921\(98\)00061-0](https://doi.org/https://doi.org/10.1016/S0014-2921(98)00061-0)
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Redding, S. & Venables, A. J. (2004). Economic geography and international inequality. *Journal of International Economics*, 62(1), 53-82. <https://doi.org/10.1016/j.jinteco.2003.07.001>
- Roback, J. (1982). Wages, Rents, and the Quality of Life. *The Journal of Political Economy*, 90(6), 1257-1278. <http://www.jstor.org/stable/1830947>
- Rosés, J. R., Martínez-Galarraga, J. & Tirado, D. A. (2010). The upswing of regional income inequality in Spain (1860–1930). *Explorations in Economic History*, 47(2), 244-257. <https://doi.org/10.1016/j.eeh.2010.01.002>
- Samuelson, P. A. (1952). The Transfer Problem and Transport Costs: The Terms of Trade When Impediments are Absent. *The Economic Journal*, 62(246), 278-304. <http://www.jstor.org/stable/2227005>
- Stakhovych, S. & Bijmolt, T. H. (2009). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, 88(2), 389-408. <https://doi.org/10.1111/j.1435-5957.2008.00213.x>
- von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. G. Fischer, <https://doi.org/10.5962/BHL.TITLE.28726>
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. https://www.ebook.de/de/product/6071525/jeffrey_m_wooldridge_econometric_analysis_of_cross_section_and_panel_data.html

Anexo

A. Metodología de los tests empleados

A.1. Test Baltagi-Song-Koh

Para la detección de autocorrelación espacial en el error o efectos aleatorios, (Baltagi et al., 2003) realizaron una adaptación de los tests del Multiplicador de Lagrange en sus versiones conjunta, marginal y condicional, conocido como el test *Baltagi-Song-Koh*. Definiendo ρ como el parámetro que recoge la autocorrelación espacial, y μ como el parámetro que identifica los efectos aleatorios, las hipótesis que los distintos tests manejan son las siguientes:

- **Conjunta:** $H_0 : \lambda = \sigma_\mu^2 = 0$ y $H_1 : \lambda = \sigma_\mu^2 \neq 0$
- **Marginal en μ :** $H_0 : \sigma_\mu^2 = 0$ y $H_1 : \sigma_\mu^2 > 0$, asumiendo la inexistencia de correlación espacial ($\lambda = 0$)
- **Marginal en λ :** $H_0 : \lambda = 0$ y $H_1 : \lambda \neq 0$, asumiendo que no existan efectos aleatorios ($\sigma_\mu^2 = 0$)
- **Condicional en λ :** $H_0 : \lambda = 0$ y $H_1 : \lambda \neq 0$, asumiendo la posible existencia de efectos aleatorios ($\sigma_\mu^2 > 0$)
- **Condicional en μ :** $H_0 : \sigma_\mu^2 = 0$ y $H_1 : \sigma_\mu^2 > 0$, asumiendo la posible existencia de correlación espacial ($\lambda \neq 0$)

Para el contraste de la hipótesis conjunta, el estadístico a utilizar es el siguiente:

$$LM_j = \frac{NT}{2(T-1)}G^2 + \frac{N^2T}{b}H^2 \quad (1)$$

donde $G = \tilde{u}' (J_T \otimes I_N) \tilde{u} / \tilde{u}' \tilde{u} - 1$, $H = \tilde{u}' (I_T \otimes (W + W') / 2) \tilde{u} / \tilde{u}' \tilde{u}$, $b = \text{tr} (W + W')^2 / 2$, $J_T = \iota^T$, y \tilde{u} son los residuos generados por una estimación de Mínimos Cuadrados Ordinarios de la ecuación. La distribución de este estadístico para muestras grandes es similar a la seguida por el estadístico I de Moran (Kelejian & Prucha, 2001), que es un conocido estadístico para la detección de correlación espacial en los residuos para datos de sección cruzada.

Los estadísticos para los dos tests marginales se derivan del estadístico del contraste de significación conjunta, y son como siguen:

$$LM_1 = \sqrt{\frac{NT}{2(T-1)}} G \quad (2)$$

$$LM_2 = \sqrt{\frac{N^2 T}{b}} H \quad (3)$$

donde ambos términos son las raíces cuadradas del primer y segundo término de (1), respectivamente. Ambos estadísticos deberían estar asintóticamente distribuidos como una $N(0, 1)$ bajo la hipótesis nula, es decir, para $N \rightarrow \infty$. No obstante, si el número de regresores en el modelo econométrico a estimar es elevado, este comportamiento es difícilmente alcanzable (Moulton y Randolph, 1989). De particular interés para nuestro trabajo es el estadístico (3), puesto que asumimos la no existencia de efectos aleatorios. Utilizaremos este último estadístico para la detección de auto-correlación espacial en la variable dependiente.

A.2. Test Multiplicador de Lagrange para componentes espaciales

Otro test del Multiplicador de Lagrange es la adaptación creada por (Elhorst, 2013) para datos de panel de la técnica ideada en (Anselin et al., 2008), que nos permite no sólo detectar autocorrelación espacial en la variable dependiente sino también en el término del error. De manera similar al test BSK, existen tanto formas estándar como formas robustas. Las formas estándar se definen como:

$$LM_\lambda = \frac{[e^T (I_T \otimes W) Y / \hat{\sigma}^2]^2}{J} \quad (4)$$

$$LM_\rho = \frac{[e^T (I_T \otimes W) e / \hat{\sigma}^2]^2}{T \times T_W} \quad (5)$$

donde e es el vector de residuos de una regresión MCO, y tanto J como T_W se definen como:

$$J = \frac{1}{\hat{\sigma}^2} \left[\left((I_T \otimes W) X \hat{\beta} \right)^T \left(I_{NT} - X (X^T X)^{-1} X^T \right) (I_T \otimes W) X \hat{\beta} + TT_W \hat{\sigma}^2 \right] \quad (6)$$

$$T_W = \text{tr} (WW + W^T W) \quad (7)$$

Mientras que las versiones robustas son:

$$\text{robust LM}_\lambda = \frac{\left[e^T (I_T \otimes W) Y / \hat{\sigma}^2 - e^T (I_T \otimes W) e / \hat{\sigma}^2 \right]^2}{J - TT_W} \quad (8)$$

$$\text{robust LM}_\rho = \frac{\left[e^T (I_T \otimes W) e / \hat{\sigma}^2 - TT_W / J \times e^T (I_T \otimes W) Y / \hat{\sigma}^2 \right]^2}{TT_W [1 - TT_W / J]} \quad (9)$$

A.3. Test Hausman de panel espacial

Un test que utilizaremos para comprobar la consistencia de los estimadores es el test de Hausman (Hausman, 1978) en su versión extendida por (Mull y Pfaffermayr, 2011) para datos de panel espaciales. Estableciendo $\hat{\theta}_{FGLS}$ como la estimación de θ bajo Mínimos Cuadrados Generalizados y $\hat{\theta}_W$ como la estimación con efectos fijos, la *hipótesis nula* indica que ambos estimadores son consistentes, pero el estimador por efectos fijos es más eficiente, mientras que la *hipótesis alternativa* señala que $\hat{\theta}_{FGLS}$ es consistente, pero no así la estimación por efectos fijos.

El estadístico para esta variante del test de Hausman se define como sigue:

$$H = NT \left(\hat{\theta}_{FGLS} - \hat{\theta}_W \right)^T \left(\hat{\Sigma}_W - \hat{\Sigma}_{FGLS} \right)^{-1} \left(\hat{\theta}_{FGLS} - \hat{\theta}_W \right) \quad (10)$$

donde $\hat{\Sigma}_{FGLS}$ y $\hat{\Sigma}_W$ son las estimaciones de las matrices de covarianzas de $\hat{\theta}_{FGLS}$ y $\hat{\theta}_W$, respectivamente, y H está asintóticamente distribuida como una chi cuadrado (χ^2) con k grados de libertad, donde k es el número de regresores en el modelo.

B. Estimación MCO por variables instrumentales

	(1)	(2)	(3)	(4)
(Intercept)	-3,67*** (0,28)	-4,74*** (0,33)	-6,55*** (0,33)	-7,03*** (0,38)
log(MP)	0,58*** (0,01)		0,60*** (0,01)	
log(EMP)		0,63*** (0,01)		0,63*** (0,02)
log(HDD)			0,32*** (0,02)	0,28*** (0,02)
R ²	0,51	0,44	0,60	0,51
Adj. R ²	0,51	0,44	0,60	0,51
Num. obs.	2079	2079	1584	1584
PWald-stat	2245.884	1935.147	2472.607	1983.662
PWald-pvalue	0.00	0.00	0.00	0.00
PHtest-stat	25.479	17.559	24.152	14.026
PHtest-pvalue	0.000	0.000	0.000	0.001
Variable de control	No	No	Sí	Sí

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 1: Estimaciones MCO con instrumentos. La instrumentación utilizada es el lag temporal de un periodo del regresor principal. La variable dependiente de cada modelo es log(GVA per capita)

C. Resultados del análisis de robustez

	<i>Spatial Auto-Regression (SAR)</i>				<i>Spatial Error Model (SEM)</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lambda	104,70*** (3,77)	122,33*** (3,27)	88,20*** (3,04)	99,50*** (2,65)				
log(MP)	0,44*** (0,02)		0,39*** (0,02)		1,00*** (0,01)		1,02*** (0,01)	
log(EMP)		0,35*** (0,02)		0,31*** (0,02)		0,98*** (0,01)		0,99*** (0,01)
log(HDD)			0,09* (0,03)	0,06 (0,03)			0,23*** (0,06)	0,08 (0,07)
rho					124,43*** (3,20)	118,00*** (3,42)	101,15*** (2,58)	96,97*** (2,76)
Num. obs.	2268,00	2268,00	1728,00	1728,00	2268,00	2268,00	1728,00	1728,00
SPHausman-stat	750.17	1406.82	825.14	1437.74	7.15	0.03	4.54	0.97
SPHausman-value	0.00	0.00	0.00	0.00	0.01	0.87	0.10	0.61

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 2: Estimaciones SAR y SEM con efectos fijos utilizando la matriz W normalizada globalmente

Regresor	Variable_control	Tipo Impacto	Efecto	Pvalor
Potencial de Mercado	No	Direct	3.522138e+56	0.381
		Indirect	1.646278e+58	0.381
		Total	1.681499e+58	0.381
Potencial de Mercado Externo		Direct	2.526609e+58	0.211
		Indirect	1.180987e+60	0.211
		Total	1.206253e+60	0.211
Potencial de Mercado	Sí	Direct	2.857822e+54	0.379
		Indirect	1.010894e+56	0.379
		Total	1.039472e+56	0.379
Potencial de Mercado Externo		Direct	7.445511e+55	0.224
		Indirect	2.633751e+57	0.224
		Total	2.708206e+57	0.224

Cuadro 3: Impactos generados por las cuatro especificaciones SAR con la matriz W globalmente normalizada. Los p-valores generados en esta tabla se han calculado con una simulación Monte Carlo de 200 repeticiones.

	<i>Spatial Auto-Regression (SAR)</i>				<i>Spatial Error Model (SEM)</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lambda	0,54*** (0,02)	0,63*** (0,02)	0,58*** (0,02)	0,65*** (0,02)				
log(MP)	0,45*** (0,02)		0,43*** (0,02)		1,00*** (0,01)		1,02*** (0,01)	
log(EMP)		0,37*** (0,02)		0,35*** (0,02)		0,98*** (0,01)		0,99*** (0,01)
log(HDD)			0,09** (0,03)	0,06 (0,03)			0,21*** (0,06)	0,09 (0,07)
rho					0,63*** (0,02)	0,60*** (0,02)	0,66*** (0,02)	0,63*** (0,02)
Num. obs.	2268,00	2268,00	1728,00	1728,00	2268,00	2268,00	1728,00	1728,00
SPHausman-stat	1.96	1.01	1.95	1.56	0.30	2.65	0.70	1.80
SPhausman-pvalue	0.16	0.32	0.38	0.46	0.58	0.10	0.70	0.41

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 4: Estimaciones SAR y SEM con efectos fijos para 3 vecinos próximos

Regresor	Variable_control	Tipo Impacto	Efecto	Pvalor
Potencial de Mercado	No	Direct	0.479	0
		Indirect	0.508	0
		Total	0.987	0
Potencial de Mercado Externo		Direct	0.401	0
		Indirect	0.584	0
		Total	0.985	0
Potencial de Mercado	Sí	Direct	0.472	0
		Indirect	0.542	0
		Total	1.014	0
Potencial de Mercado Externo		Direct	0.403	0
		Indirect	0.604	0
		Total	1.007	0

Cuadro 5: Impactos generados por las cuatro especificaciones SAR para 3 vecinos próximos. P-valores generados por Monte Carlo con 200 repeticiones

	<i>Spatial Auto-Regression (SAR)</i>				<i>Spatial Error Model (SEM)</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lambda	0,64*** (0,02)	0,76*** (0,02)	0,73*** (0,02)	0,82*** (0,02)				
log(MP)	0,35*** (0,02)		0,28*** (0,02)		1,02*** (0,01)		1,05*** (0,02)	
log(EMP)		0,24*** (0,02)		0,19*** (0,02)		0,97*** (0,01)		0,98*** (0,02)
log(HDD)			0,07* (0,03)	0,04 (0,03)			0,18** (0,07)	0,02 (0,07)
rho					0,77*** (0,02)	0,74*** (0,02)	0,83*** (0,02)	0,80*** (0,02)
Num. obs.	2268,00	2268,00	1728,00	1728,00	2268,00	2268,00	1728,00	1728,00
SPHausman-stat	2.73	1.09	0.87	0.77	0.58	3.14	1.78	0.28
SPhausman-pvalue	0.10	0.30	0.65	0.68	0.45	0.08	0.41	0.87

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 6: Estimaciones SAR y SEM con efectos fijos para 8 vecinos próximos

Regresor	Variable_control	Tipo Impacto	Efecto	Pvalor
Potencial de Mercado	No	Direct	0.391	0
		Indirect	0.595	0
		Total	0.986	0
Potencial de Mercado Externo		Direct	0.280	0
		Indirect	0.700	0
		Total	0.979	0
Potencial de Mercado	Sí	Direct	0.335	0
		Indirect	0.688	0
		Total	1.023	0
Potencial de Mercado Externo		Direct	0.247	0
		Indirect	0.761	0
		Total	1.008	0

Cuadro 7: Impactos generados por las cuatro especificaciones SAR para 8 vecinos próximos. P-valores generados por Monte Carlo con 200 repeticiones

	<i>Spatial Auto-Regression (SAR)</i>				<i>Spatial Error Model (SEM)</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lambda	0,54*** (0,03)	0,64*** (0,02)	0,60*** (0,03)	0,68*** (0,02)				
log(MP)	0,45*** (0,03)		0,41*** (0,03)		1,00*** (0,01)		1,04*** (0,02)	
log(EMP)		0,36*** (0,02)		0,33*** (0,03)		0,98*** (0,01)		1,00*** (0,02)
log(HDD)			0,16** (0,06)	0,11 (0,06)			0,34*** (0,10)	0,14 (0,11)
rho					0,65*** (0,02)	0,61*** (0,02)	0,70*** (0,02)	0,66*** (0,03)
Num. obs.	1323,00	1323,00	1008,00	1008,00	1323,00	1323,00	1008,00	1008,00
SPHausman-stat	4.75	2.68	4.40	3.35	0.24	2.77	3.92	0.71
SPhausman-pvalue	0.03	0.10	0.11	0.19	0.62	0.10	0.14	0.70

*** $p < 0,001$; ** $p < 0,01$; * $p < 0,05$

Cuadro 8: Estimaciones SAR y SEM con efectos fijos para medias de 5 años

Regresor	Variable_control	Tipo Impacto	Efecto	Pvalor
Potencial de Mercado	No	Direct	0.480	0
		Indirect	0.507	0
		Total	0.988	0
Potencial de Mercado Externo		Direct	0.393	0
		Indirect	0.592	0
		Total	0.985	0
Potencial de Mercado	Sí	Direct	0.461	0
		Indirect	0.574	0
		Total	1.035	0
Potencial de Mercado Externo		Direct	0.381	0
		Indirect	0.644	0
		Total	1.025	0

Cuadro 9: Impactos generados por las cuatro especificaciones SAR para medias de 5 años. P-valores generados por Monte Carlo con 200 repeticiones