

**INFRAESTRUCTURAS, EXTERNALIDADES  
Y CRECIMIENTO REGIONAL:  
ALGUNAS APORTACIONES PARA EL CASO ESPAÑOL**

**Rosina Moreno Serrano**

Tesis dirigida por el Dr. Manuel  
Artís Ortuño en el marco  
del programa de doctorado  
“Economía i Territori” de la  
Universidad de Barcelona.

Departamento de Econometría,  
Estadística y Economía Española.

Barcelona, Septiembre de 1998.

B.U.B. Secció d'Econòmiques  
Diagonal, 690, 08034 Barcelona  
Tel. 402 19 66

# **PARTE I**

## **CAPÍTULO 2**

# **INFRAESTRUCTURAS Y CRECIMIENTO ECONÓMICO: ESTADO DE LA CUESTIÓN**

**“The complexity of infrastructure systems  
arises partly from the interface between a relatively  
static arena which supports and constrains  
a relatively dynamic traffic of activities”**

**Johnson (1995)**

## **2.1 INTRODUCCIÓN**

Dado que desde finales de los años ochenta la literatura económica ha visto aparecer gran número de trabajos que han tratado de estudiar y cuantificar el impacto de las infraestructuras, el objetivo del presente capítulo es el de ofrecer una visión general de las que han sido las aportaciones más interesantes al respecto.

En una primera sección se enmarca el estudio del impacto de las infraestructuras en las teorías del crecimiento endógeno, presentando las principales metodologías usadas así como las conclusiones más interesantes. Concretamente, se analizan los estudios que han tratado de cuantificar el efecto directo de las infraestructuras sobre la productividad mediante la estimación de funciones de producción agregadas con el factor público así como las principales críticas que dicha metodología ha recibido y las soluciones adoptadas.

Por otra parte, si bien el grueso de la literatura sobre las infraestructuras se ha centrado en la utilización de funciones de producción que intentan cuantificar el impacto que las mismas han supuesto en el crecimiento económico, existen investigaciones económicas que sin entrar en cuantificaciones de tal tipo analizan, bien teórica bien empíricamente, otras cuestiones relativas al impacto del capital público. A pesar de que el marco analítico utilizado en ellos es muy variado, las conclusiones que se ofrecen en dichos trabajos pueden resultar de interés, por lo que en la sección 2.3 se revisan los mismos.

## **2.2 LITERATURA SOBRE EL IMPACTO DE LAS INFRAESTRUCTURAS EN EL MARCO DE LAS TEORÍAS DEL CRECIMIENTO ENDÓGENO**

### **2.2.1 Las infraestructuras y las teorías de crecimiento endógeno. Relevancia de su análisis**

La literatura que estudia el efecto de las infraestructuras se enmarca dentro de las teorías del crecimiento endógeno que surgieron a mediados de los ochenta. Anteriormente, la teoría tradicional del crecimiento económico se basaba en el modelo neoclásico (Solow,

1957) caracterizado por una función de producción con rendimientos constantes de escala (hipótesis compatible con un marco de competencia perfecta) en la que el progreso técnico viene representado por una tendencia. De este modo, dados unos factores privados, capital y trabajo, con rendimientos marginales decrecientes, en un modelo neoclásico no se puede conseguir el crecimiento a largo plazo si no es gracias a un factor exógeno, la tecnología. Sin embargo, la aparición de las teorías del crecimiento endógeno supusieron, por una parte, el reexamen de las verdaderas fuentes del crecimiento de la productividad por la endogeneización del cambio técnico y sus determinantes y, por otra, la reconciliación de los rendimientos crecientes y la competencia perfecta.

La superación del enfoque neoclásico supone la endogeneización del cambio técnico, dado que los agentes económicos estarían dispuestos a dedicar parte de sus recursos a gastos en I+D o tiempo de formación. De este modo, resulta posible la aparición de un mecanismo que impide la anulación de la productividad marginal del capital con el consiguiente crecimiento perpetuo. Para que tenga lugar tal proceso, se concilian los conceptos de rendimientos a escala crecientes y existencia de competencia gracias a las externalidades. Las mismas, de tradición marshalliana, se basan en la idea de que la concentración de la industria en un área determinada permite que las empresas puedan beneficiarse de la existencia de un conjunto de trabajadores cualificados, de una oferta de los factores concretos necesarios para el proceso productivo y de la mayor rapidez en la difusión de la información y los avances tecnológicos. De este modo, en los distritos industriales marshallianos<sup>1</sup> las empresas son precio-aceptantes, considerándose la idea de que existen rendimientos crecientes externos que se manifiestan en la industria pero que a nivel de empresa pueden no serlo. Así, Marshall (1920) da un paso adelante en la consideración del progreso tecnológico al alejarse de las hipótesis de convexidad en el crecimiento que limitaban el entendimiento de dicho proceso.

Posteriormente, Meade (1952) formalizó el concepto de externalidad y Chipman (1970) demostró la existencia de un equilibrio consistente y competitivo con externalidades, de

---

<sup>1</sup> Un distrito industrial, en el sentido marshalliano, integra todas las actividades desde una perspectiva vertical (Becattini, 1979).

forma que el concepto propuesto por Marshall pudo verse justificado rigurosamente (véase Caballero y Lyons, 1989). Hoy en día, tras la aparición de las obras de Romer (1986, 1990) y Lucas (1988), entre otros, se encuentra comúnmente aceptada la idea de que la inclusión, por ejemplo, del conocimiento y del capital humano como un input de la producción supone aumentos en los rendimientos a escala.

Se ha propuesto una clasificación de los modelos de crecimiento endógeno según el factor acumulado que da origen al crecimiento: capital físico, tecnología a través de inversiones en I+D, capital humano y bienes e infraestructuras públicas. Dado que los tres primeros modelos son tratados en detalle en el Capítulo 3 del presente trabajo, nos centramos ahora en el último.

Según las teorías del crecimiento endógeno, el capital público permite aumentar, directa o indirectamente, la productividad de los factores privados mediante una mejor accesibilidad a través de la reducción de costes, permitiendo el aumento de la variedad de productos, etc. De hecho, los rendimientos a escala de las empresas serían decrecientes pero los factores que proporciona el Estado harían aumentar dichos rendimientos a través de un verdadero proceso de crecimiento endógeno.<sup>2</sup>

Según Biehl (1986), la infraestructura se define como aquella parte del capital global de las economías que, debido a su carácter público, normalmente no es suministrada por el mercado o que éste sólo la suministra de manera ineficiente, por lo que su provisión queda fundamentalmente confiada a las decisiones políticas. Los dos criterios que se suelen utilizar para distinguir el capital "público" del "privado", pueden resumirse en las dos propiedades musgravianas de no rivalidad y no exclusión, dos propiedades que ocasionan fallos de mercado.<sup>3</sup> Son categorías de infraestructura las redes de transporte y

---

<sup>2</sup> Si bien la mayoría de los modelos que endogeneizan el crecimiento se limitan a constatar que el crecimiento supone un aumento de la productividad de los factores, en algunos se introducen nuevas nociones como el incremento del número de bienes diferentes disponibles y la mejora de la calidad de los mismos, con repercusiones en la utilidad y bienestar de los consumidores en ambos casos.

<sup>3</sup> No obstante, ciertas infraestructuras se encuentran frecuentemente congestionadas, y aunque el consumo de un usuario puede tener un impacto mínimo en el resto, no resulta ser siempre así. Del mismo modo, no es difícil excluir a usuarios individuales de consumir ciertas formas de infraestructuras de transporte y comunicaciones. El concepto de exclusión se entiende en términos de que la provisión de un bien público

comunicaciones, las redes de distribución de la oferta energética, el sistema de abastecimiento de agua, los equipamientos docentes, sanitarios, asistenciales, instalaciones culturales, centros de control medio-ambiental, estructuras urbanas, etc.. En conjunto, y desde su dimensión material y cuantitativa, las infraestructuras comprenden los equipamientos, estructuras y servicios de soporte requeridos para el desarrollo económico de un área determinada.

Actualmente, se defiende, no obstante, una concepción más amplia de las infraestructuras, que combina la dimensión material (infraestructura física) y de gestión de las mismas (infraestructura lógica). Según Draper y Herce (1994), un ejemplo de esto último lo constituiría el sistema de gestión de tráfico que debería acompañar a una red de carreteras congestionada. Asimismo, Stern (1991) sugiere que la falta de un adecuado sistema de derechos de propiedad, una burocracia obstructiva o un sistema en el que los individuos se comporten de manera deshonesto pueden verse como una carencia de infraestructuras institucionales tan necesarias para el desarrollo económico como las infraestructuras físicas. Jochimsen (1966) ya daba una de las definiciones más amplias de las infraestructuras, que incluye casi todos los tipos de servicio público y hasta la administración general como infraestructura institucional. En este caso, los recursos necesarios para solucionar las deficiencias de ciertas infraestructuras físicas y la consecuente distorsión de incentivos pueden constituir serios impedimentos al crecimiento económico. Desde esta nueva perspectiva, se consideran las infraestructuras como un sistema de gestión de la dotación física que realza su importancia potencial en la economía, siendo el buen aprovechamiento de la misma tanto o más importante que su construcción física. Debido a multitud de factores (espaciales, medioambientales o regulatorios), la utilización eficiente de la infraestructura física es cada vez más necesaria.<sup>4</sup>

---

supone que todos deben efectivamente consumirlo (por ejemplo, la defensa nacional cubre a toda la población de un país tanto si se quiere como si no). Resulta difícil pensar en casos donde esta idea se pudiera aplicar para las infraestructuras, especialmente las de transporte. Por estas razones, la definición exacta de Musgrave (1959) "bienes públicos son aquellos bienes cuyas características inherentes requieren la provisión pública" debe ser matizada.

<sup>4</sup> No obstante, dado que resulta muy difícil obtener datos sobre el stock de infraestructuras en su concepción amplia, la mayoría de los trabajos aplicados (a los que se suma el presente) utilizan la definición estricta de infraestructuras físicas.

Hasta este punto se han utilizado indistintamente los términos infraestructura y capital público. Sin embargo, existen ciertas diferencias entre los dos. El término capital público debería reservarse para aquellas inversiones que son propiedad estricta del sector público, de forma que ciertas infraestructuras son capital público y otras no lo son. Las autopistas en España son un claro ejemplo. Así, si bien es parte de la infraestructura de la nación, en la mayoría de los casos son propiedad privada por lo que no entraría en la categoría estricta de capital público. Estas diferencias son importantes a la hora de hacer comparaciones internacionales dado que ciertas infraestructuras, principalmente las de telecomunicaciones, son provistas públicamente en ciertos países y por empresas privadas en otros. En el presente trabajo se van a utilizar indistintamente ambos términos, si bien el concepto final sería el equivalente a la totalidad de la infraestructura, independientemente del carácter de su provisión.

Según las teorías del crecimiento endógeno, por tanto, el capital público permite aumentar la productividad de los factores de producción privados, incrementando los rendimientos a escala de la economía agregada. En este sentido, se podría argumentar que el capital público es una verdadera externalidad asociable a las clásicas marshallianas. Así lo han considerado autores como Johansson y Wigren (1996) cuando hablan de las infraestructuras como parte del *production milieu* de una región. Formalmente, el *production milieu* comprende todos aquellos atributos de localización propios de una región que son durables y sobre los cuales (i) las empresas no tienen control, (ii) para los cuales no existen precios de mercado ni cargas directas y (iii) que influyen en las posibilidades de producción de las empresas. Parte de dichos atributos son dados por la naturaleza y otros son creados a través de inversiones de distintos tipos (incluyendo la experiencia, el conocimiento y la competencia de la fuerza laboral). Las infraestructuras se encontrarían en este segundo grupo, si bien son las consecuencias de las infraestructuras y no el tamaño o el valor del stock de capital las que realmente forman parte del *production milieu* de la zona, que afectaría a las posibilidades de producción así como a la dinámica de localización de la actividad.

Para poder estudiar y cuantificar la supuesta incidencia de las infraestructuras sobre el crecimiento, un gran número de estudios se centran en una función de producción

agregada<sup>5</sup> (tecnología de producción agregada) en la que el capital público y el privado aparecen como argumentos separados:

$$Y_t = A_t f(L_t, Kp_t, Kg_t) \quad (2.1)$$

donde

$Y_t$	producción agregada privada de bienes y servicios
$A_t$	medida del progreso tecnológico
$L_t$	empleo agregado del sector privado
$Kp_t$	stock de capital productivo privado
$Kg_t$	stock de infraestructuras públicas

En concreto, la mayor parte de los trabajos empíricos suponen que la función de producción es del tipo Cobb-Douglas:<sup>6</sup>

$$Y_t = A_t L_t^\alpha Kp_t^\beta Kg_t^\gamma \quad (2.2)$$

Dado que el exponente para cada input es la elasticidad del output respecto al input,  $\alpha$ ,  $\beta$  y  $\gamma$  representan la contribución relativa de cada input al output. En otras palabras, los exponentes indican el porcentaje de cambio en el output, dado un porcentaje de cambio en el input. El término  $A$  es el factor de cambio que tiene en cuenta los efectos específicos del tiempo en la productividad total.

En este modelo se está suponiendo que el Estado provee los servicios directamente a los productores privados, en el sentido de que no emplea impuestos, por lo que será un factor

---

<sup>5</sup> En la teoría de la producción, la función de producción juega, sin duda, un papel crucial. Esencialmente, se puede decir que se trata de una noción que revela el máximo producto que puede obtenerse en un período de tiempo, dadas unas cantidades de factores productivos, entre los que se consideran, generalmente, el capital privado y el empleo.

<sup>6</sup> La tecnología Cobb-Douglas es una forma funcional que implica elasticidades de sustitución entre los inputs constantes e igual a la unidad.

cuya productividad marginal no está remunerada sino que se traslada al resto de factores productivos. De hecho, este argumento en favor de posibles economías de escala gracias a la provisión pública sugiere que una especificación razonable de la tecnología privada supondría que dicha función presenta rendimientos constantes a escala sobre los inputs privados ( $L$ ,  $Kp$ ) pero rendimientos crecientes sobre el conjunto de todos los inputs, incluyendo el capital público ( $Kg$ ). En este caso, los factores privados son pagados según su productividad marginal y el output privado se distribuirá completamente.

Así, la suma de los exponentes en la ecuación (2.2) ( $\alpha+\beta+\gamma$ ) nos da el grado de las economías de escala. Esto es así ya que las medidas del output y de los inputs son el resultado de agregar todas las empresas en la economía. Dada la posibilidad de existencia de externalidades entre las empresas (economías de aglomeración o congestión), las posibilidades de producción nacional o regional pueden diferir (ser mayores o menores) de las posibilidades de producción derivadas de funciones de producción a nivel de empresas individuales. Según el teorema de Hicks,<sup>7</sup> si la suma de los exponentes es superior a 1, bien la mayoría de empresas están experimentando rendimientos a escalas crecientes, bien existen economías de aglomeración o ambas cosas al mismo tiempo. De esta manera, se podrá dar algo más de luz a la cuestión, no resuelta, de si las infraestructuras son una verdadera externalidad a la producción que operan de forma complementaria a los tres tipos de externalidades convencionales.

De esta forma, el tratamiento de las infraestructuras que realiza el presente trabajo se enmarca en el grupo de los modelos que hacen endógeno el crecimiento de la productividad, a través de aumentos en los rendimientos a escala. Las infraestructuras y servicios públicos son considerados factores acumulados que puede dar origen a dicho crecimiento. La noción de infraestructuras a la que nos vamos a referir se restringe al concepto estricto de las infraestructuras físicas, que son las únicas cuantificadas en términos monetarios para el caso español, dejando de lado servicios públicos como la

---

<sup>7</sup> El teorema de agregación de Hicks implica que el grado de las economías de escala derivadas de la función de producción será un promedio ponderado de los rendimientos a escala en la empresa corregido por las economías externas o inter-empresas que éstas se confieren unas a otras. Para mayor explicación véase Hicks (1946).

garantía de los derechos de propiedad, la justicia, la seguridad interior y exterior, entre otros.

La ventaja principal de la aparición de las teorías del crecimiento endógeno, entre las que se incluyen los trabajos que analizan el impacto de las infraestructuras, es, en opinión de Romer, la posibilidad de hacerse un grupo de nuevas preguntas que apenas han empezado a analizarse, y un conjunto de nuevas técnicas que permitan su exploración. Sin embargo, el estudio de todos estos modelos no resulta fácil y las conclusiones serán algo más confusas que para el caso del modelo neoclásico, pero se habrá ganado en que podrán tratarse muchos más problemas sobre la teoría del crecimiento agregado.

En la última década ha existido un interés importante en el estudio del papel que las infraestructuras juegan en el crecimiento económico. Ciertos autores piensan, no sin justificación, que dicho interés se debe a un modelo cíclico de cambio en los intereses académicos (Button, 1996), de forma que la importancia que se le ha dado al estudio del impacto de las infraestructuras se ha sobredimensionado en estos últimos años. Así, mirando retrospectivamente y en opinión de Gramlich (1994), existen numerosas situaciones en las que un tema razonablemente interesante es casi totalmente ignorado durante un largo tiempo y, repentinamente se convierte en el centro de atención de un número de trabajos fuera de toda proporción. El estudio de la inversión en infraestructuras sería claramente uno de ellos.

Son varias las razones a la hora de argumentar la importancia dada a los estudios sobre el impacto de las infraestructuras.

En primer lugar, no es nueva la concepción de que las infraestructuras pueden tener un efecto importante en el crecimiento. Así, ya Smith (1937) se refería al gasto en infraestructuras como la tercera razón de ser del estado, detrás de la provisión de defensa y justicia, mientras que Meade (1952) sugería la posibilidad de que determinadas formas de capital público pudieran considerarse como inputs relevantes en el proceso de producción de la industria privada, contribuyendo de forma importante en la generación de output. Asimismo, Hirshman (1958) asignaba al capital social en infraestructuras un importante

papel en *The Strategy for Economic Development*. Lo que esto y otros teóricos no llevaron a cabo es una evaluación cuantitativa sistemática de la contribución de las infraestructuras al crecimiento de la productividad, por lo que resulta lógico que una serie de estudios empíricos surgieran para avalar la teoría económica.

En segundo término, el tamaño del stock de capital público es suficientemente importante como para investigar su papel en el crecimiento de la economía. Así, en 1988 el stock neto del capital fijo del sector público representaba una tercera parte del valor del stock de capital fijo privado en los Estados Unidos (Lynde, 1992). De modo que analizar como los cambios en la provisión de estos servicios afectan al crecimiento puede resultar interesante en el estudio de toda política regional.

En tercer lugar, muchos estudios sobre infraestructuras surgieron en busca de una explicación a la reducción generalizada del crecimiento de la productividad en la mayoría de países occidentales desde inicios de los años setenta. Varios autores, entre ellos Aschauer (1989a), señalaron la elevada correlación entre la caída de la productividad y el descenso en la formación de capital público, levantando la sospecha de una posible relación causa-efecto entre las dos variables. De este modo, con los trabajos de Aschauer (1989a, 1989b, 1989c), se mostraba la evidencia empírica de tal relación, dando una explicación válida al problema de la caída de la productividad en varios países avanzados. A raíz de dichos resultados, son numerosos los trabajos que se sucedieron en un intento de cuantificar el impacto del capital público.

En cuarto lugar y a nivel europeo, se ha de señalar que una de las herramientas que más ha sido utilizada para aumentar la competitividad internacional de un país en el camino hacia una verdadera integración, ha sido la provisión de un buen sistema de infraestructuras. Por esta razón, parece razonable estudiar cuales han sido los efectos macroeconómicos que cabe esperar de este esfuerzo inversor.

Finalmente, y centrándonos en el caso español, a pesar de que la inversión pública no es la partida con más peso en el total de los gastos de las administraciones públicas (en 1991 la

inversión representó un 12.4% del total del gasto público consolidado en España),<sup>8</sup> los estudios sobre el impacto de las infraestructuras han sido posibles en estos últimos años gracias a las cuantificaciones monetarias de los stocks de capital público para las regiones españolas publicados a mediados de los noventa (Fundación BBV, 1995). De este modo, si bien algunos estudios sobre los efectos de las infraestructuras en el crecimiento utilizaban estimaciones aproximadas de dicho stock para el período 1980-1990 (Mas et al., 1994), (García-Fontes y Serra, 1993a), actualmente se pueden llevar a cabo investigaciones, como la presente, que consideran un período más dilatado en el tiempo (1964-1991).

### 2.2.2 Primer grupo de investigaciones para economías occidentales

A pesar de que antes de finales de los ochenta algunos autores<sup>9</sup> habían cuantificado en sus estudios la influencia del capital público en el crecimiento, el debate reciente sobre el tema de las infraestructuras tiene como componente destacado la aportación de Aschauer (1989a), en la que, incorporando de forma explícita el capital público en la función de producción tipo Cobb-Douglas, propone la relevancia que esa tipología de los bienes de capital podía suponer en el análisis de los determinantes del ritmo de avance de la productividad. En dicho estudio, Aschauer parte de una función de producción de tipo Cobb-Douglas en la que el stock de capital público se añade como un factor de producción adicional (ecuación 2.2). Dividiendo ambos lados de la ecuación por  $Kp_t$ , tomando logaritmos y asumiendo rendimientos constantes a escala en todos los inputs, se obtiene:

$$\ln \frac{Y_t}{Kp_t} = \ln A_t + \beta \ln \frac{L_t}{Kp_t} + \gamma \ln \frac{Kg_t}{Kp_t} \quad (2.3)$$

Aschauer (1989a) introduce una constante y una tendencia como una aproximación al factor tecnológico,  $A$ . También añade una tasa de utilización de la capacidad productiva

---

<sup>8</sup> No se incluyen en el cálculo las variaciones de activos y pasivos financieros.

<sup>9</sup> Mera (1973), Ratner (1983), Eberts (1986), Da Silva *et al.* (1987) y Holtz-Eakin (1988) son los trabajos que ofrecen las primeras estimaciones de una función de producción en la que se incluye el capital público como input.

para controlar la influencia del ciclo económico. Con esta especificación, los resultados de Aschauer indican que el capital público juega un papel importante en la evolución de la productividad total de los factores en la economía americana, resultado conocido como *efecto Aschauer*. Él mismo indica que "... un incremento de un 1% en el ratio entre el stock de capital público y el de capital privado aumenta la productividad del capital en un 0.39%". Estos resultados en un momento en que se buscaba una explicación a la caída de la tasa de crecimiento de los países industrializados, junto con las evidentes implicaciones de política económica que de él se derivan en forma de defensa del gasto público en inversión, hicieron que su trabajo tuviera una elevada repercusión en la literatura económica posterior.

Resultados parecidos fueron obtenidos por otros autores que utilizaban una metodología muy similar a la de Aschauer. Así, también a nivel nacional para la economía americana, Munnell (1990a) obtiene en diferentes especificaciones que la elasticidad del producto con respecto al capital público varía de 0.31 a 0.37, muy próximos a los valores estimados por Aschauer. Posteriormente, Munnell (1990b) y Garcia-Milà y McGuire (1992) confirmaban a nivel regional que el capital público tiene un impacto importante y significativo sobre el output, aunque la elasticidad estimada resulta más pequeña cuando se consideran los estados americanos en lugar del ámbito nacional en su conjunto. Así, Munnell obtiene una elasticidad del 0.15 mientras que Garcia-Milà y McGuire, centrándose en el stock de carreteras, obtienen un coeficiente para las mismas con un valor que oscila entre 0.05 y 0.17, según la estimación. El mayor número de observaciones que en ambos trabajos se utilizan respecto a los anteriores, así como la consideración de datos en diferentes años para los diferentes estados americanos, hizo que los resultados parecieran más fiables.

Una reducción similar de la elasticidad del producto respecto al capital público se obtiene cuando el ámbito territorial se refiere al área metropolitana. Deno y Eberts (1989) y Eberts (1990) estiman unos coeficientes para el capital público inferiores al 10 %, bastante lejos del 40 % estimado inicialmente por Aschauer para el conjunto norteamericano. Por tanto, parece ser que parte de las diferencias de las estimaciones están en relación con el nivel de desagregación adoptado, reduciéndose la magnitud de las elasticidades estimadas a medida que se desciende en el ámbito geográfico de referencia.

Asimismo, en muchos de estos trabajos se señala que la composición del capital público influye en los resultados. Desagregando las distintas formas de capital público, tanto Munnell (1990b) como Garcia-Milà y McGuire (1992) concluyen que son las infraestructuras básicas (las de transporte y comunicaciones, obras hidráulicas y estructuras urbanas, fundamentalmente) las que muestran una relación más estrecha con la productividad. En cambio, otras formas de capital público como el dedicado a sanidad y educación resultan menos relevantes.

También existen algunos trabajos que han estimado funciones de producción para distintos países y varios años. Así, Aschauer (1989b) trabaja con los países del Grupo de los Siete ofreciendo resultados muy similares a los obtenidos en su primer trabajo. También en el trabajo de Ford y Poret (1991) se estiman funciones de producción para once países de la OCDE en las que combinan el capital y el empleo en una sola variable imponiendo, por tanto, algunas restricciones que no son contrastadas. Con tal especificación, el efecto de las infraestructuras es siempre significativo en 5 casos (Estados Unidos, Alemania, Canadá, Bélgica y Suecia), nunca lo es en 3 casos (Gran Bretaña, Noruega y Australia) y tan sólo algunas veces en los tres restantes (Francia, Japón y Finlandia). En conjunto, sus estimaciones no son del todo favorables al efecto Aschauer. Evans y Karras (1994a) obtienen un coeficiente positivo y significativo del capital público, aunque algo menor que el estimado por Aschauer, para siete países de la OCDE. Hay que destacar, no obstante, que el coeficiente deja de ser significativo cuando se incluyen efectos específicos tanto temporales como nacionales.

En la Tabla 2.1 se encuentran los resultados obtenidos en diversos trabajos que confirman el efecto Aschauer a través de la utilización de funciones de producción. En dicha tabla se especifica el nivel geográfico considerado, así como el tipo de especificación utilizado.

Tabla 2.1 Estudios que confirman el efecto Aschauer

<b>ESTIMACIONES DE LA ELASTICIDAD DEL OUTPUT RESPECTO AL CAPITAL PÚBLICO</b>			
<i>TRABAJO</i>	<i>NIVEL DE AGREGACIÓN</i>	<i>ESPECIFICACIÓN</i>	<i>ELASTICIDAD</i>
Mera (1973)	Japón Regional 1954-1963	Cobb-Douglas	0.20
Ratner (1983)	EEUU Nacional 1949-1973	Cobb-Douglas	0.06
Eberts (1986)	EEUU Metropolitano 1958-1978	Translog	0.03 a 0.07
Da Silva <i>et al.</i> (1987)	EEUU Regional 1972	Translog	0.20
Holtz-Eakin (1988)	EEUU Nacional	Cobb-Douglas	0.39
Aschauer (1989a)	EEUU Nacional 1948-1985	Cobb-Douglas	0.39 a 0.56
Aschauer (1989b)	EEUU Regional 1966-1985	Cobb-Douglas	0.34 a 0.59
Ram y Ramsey (1989)	EEUU Nacional 1949-1985	Cobb-Douglas	0.24
Munnell (1990a)	EEUU Nacional 1949-1987	Cobb-Douglas	0.34
Munnell (1990b)	EEUU Regional 1970-1986	Cobb-Douglas	0.06 a 0.34
Aschauer (1990)	EEUU Regional 1965-1983	Cobb-Douglas	0.05 a 0.11
Munnell y Cook (1990)	EEUU Regional 1970-1986	Cobb-Douglas	0.15
Merriman (1990)	Japón Regional 1954-1963	Translog	0.43 a 0.58
Eisner (1991)	EEUU Regional 1970-1986	Cobb-Douglas	0.17
Ford y Poret (1991)	EEUU Nacional 1957-1989	Cobb-Douglas (en diferencias)	0.39 a 0.54
Hulten y Schwab (1991)	EEUU Nacional 1949-1985	Cobb-Douglas	0.21
Berndt y Hansson (1992)	Suecia Nacional 1964-1988	Cobb-Douglas	0.68
Garcia-Milà y McGuire (1992)	EEUU Regional 1970-1983	Cobb-Douglas	0.05 a 0.17
Holtz-Eakin (1992)	EEUU Regional 1969-1986	Cobb-Douglas	0 a 0.20
Munnell (1993)	EEUU Regional 1970-1986	Cobb-Douglas	0.14 a 0.17

Tabla 2.1. Continuación

Eisner (1994)	EEUU Nacional 1961-1991	Cobb-Douglas	0.27
Otto y Voss (1994)	Australia Nacional 1966-1990	Cobb-Douglas	0.38 a 0.45
Sturm y de Haan (1995)	EEUU Nacional 1949-1985	Cobb-Douglas	0.41
Dalamagas (1995)	Grecia Nacional 1950-1992	Translog	0.53
Holtz-Eakin (1992)	EEUU Regional 1969-1986	Cobb-Douglas	0 a 0.20

FUENTE: Elaboración propia

Como se puede observar, los coeficientes del capital público que se obtienen en todos estos estudios aportan evidencia a favor de la existencia de un efecto significativo de las infraestructuras sobre la productividad. Asimismo, en dicho cuadro se llega a la conclusión de que la función de producción Cobb-Douglas, tal y como ha sido especificada y estimada, no es capaz de recoger los efectos que las dotaciones de infraestructura de las áreas vecinas pueden suponer sobre el crecimiento económico de una zona concreta. Esto se pone de manifiesto por las menores elasticidades del output respecto al capital público que presentan los estudios en los que el ámbito de aplicación es regional y no nacional, agravándose el fenómeno cuando pasamos a áreas metropolitanas. En todos ellos se sigue obteniendo, no obstante, valores positivos y significativos de dicha elasticidad, confirmándose, de esta manera, el efecto Aschauer.

Dado que los coeficientes estimados para el capital público en las funciones de producción Cobb-Douglas dependen del nivel de desagregación geográfico que se adopte, parece deducirse que no deberíamos esperar captar todos los efectos de la inversión pública cuando la zona de referencia es pequeña. No obstante, esta conclusión debería interpretarse en palabras de Mas *et al.* (1993) "en el sentido de que existe mayor comparabilidad entre distintos resultados de la que parece", ya que los diferentes coeficientes que los diversos autores han obtenido para la economía norteamericana no son tan divergentes como en un

principio puede parecer, sino que son consecuencia del diferente ámbito de aplicación considerado.

### 2.2.3 Primer grupo de investigaciones para la economía española

A raíz de la extensa literatura sobre los efectos de las infraestructuras, y dados los diferentes resultados obtenidos principalmente para la economía de EEUU, ha surgido un amplio debate sobre el tema que ha traspasado las fronteras norteamericanas. En España, la citada controversia también ha dado lugar a la aparición de distintos trabajos de investigación que aportan evidencia empírica sobre el mismo. El hecho de que en España no se hubieran llevado a cabo estudios sobre los efectos de las infraestructuras hasta casi mediados de los noventa (el primer trabajo que conocemos es de 1993) puede ser debido a dos razones tal como señalan Mas *et al.* (1994): la falta de datos regionales y la defensa de intereses particulares.

En primer lugar, la falta de datos regionales sobre el stock de capital, tanto privado como público, no hacía posible la utilización de funciones de producción en los estudios de los efectos del capital público. Como muestra de la enorme complejidad del proceso de obtención de datos monetarios del stock del capital público, basta la frase de García-Fontes y Serra (1993a): "Los efectos de las inversiones y del stock de capital público existentes sobre el crecimiento regional han sido motivo de discusión reciente en la investigación aplicada sobre el crecimiento económico. El único punto sobre el cual los investigadores están de acuerdo reside en la dificultad de obtener datos para realizar dicho análisis". Tres son las razones de esa elevada complejidad: por una parte, las diferentes características asociadas a cada tipo de infraestructura hacen difícil la homogeneización de los datos a fin de clasificarlas por categorías; segundo, la falta de datos en las estadísticas oficiales, que en la mayoría de casos reflejan los flujos pero no las existencias, hace necesario bien un plan de compilación de datos bien un grupo de estudio dedicado a la obtención de series de inversiones monetarias transformables, mediante formulación matemática, en series de stock; por último, el hecho de que muchas fuentes de información no ofrezcan cifras desglosadas anualmente hace necesario un enorme trabajo para conseguir series temporales

sobre los equipamientos de infraestructura, por lo que algunos estudios sólo recogen datos transversales.<sup>10</sup>

La segunda posible causa de la escasa profundidad con que han sido abordados los trabajos españoles es, en opinión de Mas *et al.* (1994), la defensa de intereses particulares. En primer lugar, los empresarios, que opinaban que el gasto público debía reorientarse con el objetivo de atender a las necesidades del aparato productivo en lugar de destinarse a invertir en capital público. El segundo tipo de intereses eran consecuencia de la entrada en vigor del marco institucional de las autonomías, en el que aparece una fuerte rivalidad a la hora de reclamar mayores inversiones del Gobierno Central para su territorio, por lo que se tenía miedo a que los resultados de los estudios sobre los efectos que una infraestructura pública puede suponer sobre el crecimiento y desarrollo general perjudicasen a su región (principalmente las que ocupan una mejor posición relativa). Tanto en un caso como en otro, prevalecían los intereses particulares por lo que no había empuje a nivel privado de potenciar estudios en este sentido.

A pesar de las dificultades, en los últimos años se han realizado varios estudios que han superado el problema de los datos para España. En un primer momento, se publican trabajos que utilizan funciones de producción basadas en series de capital público y privado

---

<sup>10</sup> Para llevar a cabo los trabajos que abordan el efecto de las infraestructuras sobre el crecimiento económico en base a la teoría neoclásica, se necesita conocer las dotaciones de capital público, en términos monetarios, que es como se expresan el resto de factores productivos. En ese sentido existen datos de flujos de inversión como son la FBCF de las Administraciones Públicas, o más concretamente, el gasto público en infraestructuras, en diferentes cuentas de las AAPP. La variable utilizada en la mayoría de estudios es, sin embargo, el stock de capital público, cuya medición plantea grandes dificultades que, no obstante, han sido resueltas con mayor o menor éxito a través de varios métodos de estimación, siendo el más utilizado el Método del Inventario Permanente (MIP). Según dicho método, la dotación de capital bruto de un año se obtiene por acumulación de los flujos de inversiones anteriores y por minoración del valor acumulado de la inversión retirada, utilizando para ello un determinado esquema de retiro y valorando todo a precios de un año base.

Denominando  $KB_t^i$  al stock bruto de capital del activo  $i$  en el período  $t$ ,  $IB_t^i$  la inversión bruta en el activo  $i$  en el período  $t$ ,  $R_t^i$  el retiro del activo  $i$  producido en el período  $t$ ,  $M_i$  la duración máxima de vida del activo  $i$ ,  $r_j^i$  la tasa de retiro del activo  $i$  después de  $j-1$  periodos realizada la inversión, la obtención del stock bruto de capital se puede formular de la siguiente manera:

$$KB_t^i = KB_{t-1}^i + IB_t^i - R_t^i \quad \text{siendo} \quad R_t^i = \sum_{j=0}^{M_i} r_j^i IB_{t-j}^i$$

a nivel nacional como Argimón *et al.* (1993, 1994a) con datos obtenidos en Argimón y Martín (1993) sobre infraestructuras de transporte y comunicaciones y el trabajo de Bajo y Sosvilla (1993) que toma la serie de stocks de la base de datos del modelo MOISEES (Modelo de Investigación y Simulación para la Economía Española). En estos trabajos se obtiene un efecto positivo del capital público sobre el output, con una elasticidad en la línea de otros estudios para la economía norteamericana a nivel nacional. Bajo y Sosvilla (1993) estiman una función de producción Cobb-Douglas con la productividad del capital privado como endógena y utilizando la capacidad productiva como regresor, obteniendo una elasticidad significativa a largo plazo de 0.25. Argimón *et al.* (1993, 1994a) no introducen variables que recojan ni el progreso tecnológico ni el ciclo económico, ofreciendo un valor de la elasticidad del output que oscila entre 0.21 y 0.60. Esta variabilidad se debe a que los autores consideran diferentes series de capital público (para el total de las Administraciones públicas y para el Estado).

Posteriormente a nivel regional, García-Fontes y Serra (1993a, 1993b) estiman una función de producción ampliada con capital público para las regiones españolas en el periodo 1980-1988 a partir de series de capital público y privado construidas por ellos mismos en base a los datos suministrados por Frutos (1991), la Contabilidad Regional de España y otras fuentes. En dichos trabajos realizan estimaciones de la función de producción con diferentes combinaciones de efectos fijos (temporales y regionales), tanto con series en niveles como en diferencias. La magnitud del coeficiente del capital público cuando la estimación se realiza en niveles se encuentra entre el 0.02 y el 0.06, resultando significativa sólo en el caso de incluir efectos fijos temporales y regionales. En el caso de la especificación en diferencias, los coeficientes aumentan pero pierden significatividad. Por otra parte, Mas *et al.* (1993, 1994) también estiman una función de producción con datos de panel para las regiones españolas en el periodo 1980-1989, considerando los stocks de red de transportes, infraestructuras hidráulicas y estructuras urbanas para el capital público construidas por IVIE (1993), y la serie de capital privado construida por Calabuig *et al.* (1993). De su estimación en niveles, con efectos aleatorios en su primer trabajo y efectos fijos en el segundo, dichos autores obtienen un coeficiente positivo y significativo para el capital público, con un tamaño superior en el segundo caso (0.191 frente a 0.091). Además,

en su segundo trabajo, se distingue entre capital público productivo y social, concluyendo que es el primero el que tiene un efecto importante en la productividad de la economía.

De la Fuente (1993) estima una función de producción que incluye como variables explicativas el nivel medio de formación de la mano de obra y dos variables referentes al stock de infraestructuras: el stock de capital público productivo por ocupado y por Km<sup>2</sup> de superficie. Utilizando datos para las regiones españolas en los años 1980, 1985 y 1990, estiman el modelo en niveles por mínimos cuadrados ordinarios, introduciendo dummies temporales y una variable ficticia regional. Sus resultados confirman la importancia de las variables de educación y capital público, con mayor peso de la primera.

Sin embargo, el paso más importante en materia de obtención de datos regionales de stocks de capital privado y público se ha dado con la publicación por la Fundación BBV (1995) de las series elaboradas por el Instituto Valenciano de Investigaciones Económicas (IVIE) para el periodo 1964-1991. De esta manera, las series de capital público tienen una mayor desagregación funcional: carreteras, autopistas, RENFE y FEVE, puertos, aeropuertos, infraestructura hidráulica, estructuras urbanas, sanidad y educación. Con estos datos, Mas *et al.* (1996a) y de la Fuente (1996) realizan estudios sobre la relación capital público-productividad a nivel regional español en el periodo 1964-1991 basándose en la metodología propuesta por Aschauer. El trabajo de Mas *et al.* considera un modelo de efectos fijos regionales con una tendencia y desagregando el capital público en las dos categorías comentadas anteriormente: capital público básico y social. La elasticidad de las infraestructuras básicas tiene un valor de 0.08 junto a un valor negativo pero no significativo de la componente social. El trabajo de de la Fuente (1996) replica inicialmente algunas de las especificaciones típicamente utilizadas en los trabajos sobre infraestructuras, obteniendo que cuando no se controla por peculiaridades regionales el coeficiente del capital público suele ser generalmente positivo, pequeño y significativo. Sin embargo, cuando introduce efectos específicos por regiones o emplea una especificación en diferencias, las infraestructuras pierden su significatividad e incluso llegan a aparecer con signo negativo (dichos problemas se tratan más en detalle en el apartado siguiente).<sup>11</sup>

---

<sup>11</sup> En el trabajo de de la Fuente (1996) se desarrolla también un modelo de crecimiento que combina una función de producción con una especificación de la tasa de progreso técnico de forma que dicha variable

En la Tabla 2.2 aparecen resumidos los diferentes trabajos realizados en España y que concluyen con un resultado favorable al efecto Aschauer, que son la mayoría.

**Tabla 2.2** Estudios que confirman el efecto Aschauer en España

<b>ESTIMACIONES DE LA ELASTICIDAD DEL OUTPUT RESPECTO AL CAPITAL PÚBLICO</b>		
<i>TRABAJO</i>	<i>NIVEL DE AGREGACIÓN</i>	<i>ELASTICIDAD</i>
Bajo y Sosvilla (1993)	Nacional 1964-1988	0.25
Argimón <i>et al.</i> (1993)	Nacional 1964-1989	0.21 a 0.60
García-Fontes y Serra (1993a)	Nacional 1969-1988	0.18 a 0.27
García-Fontes y Serra (1993b)	Regional 1980-1988	0.17 a 0.22
de la Fuente (1993)	Regional 1980, 1985 y 1990 (efectos temporales y var. ficticia reg.)	0.07 y 0.16
Mas <i>et al.</i> (1993)	Regional 1980-1989 (efectos regionales aleatorios)	0.09
Mas <i>et al.</i> (1994)	Regional 1980-1989 (efectos regionales fijos)	0.18 a 0.24
Mas <i>et al.</i> (1996)	Regional 1964-1991 (efectos regionales fijos)	0.06 a 0.08
de la Fuente. (1996)	Regional 1964-1991	0.06

FUENTE: Elaboración propia

NOTA: Todos los trabajos presentados en la tabla estiman una función de producción del tipo Cobb-Douglas

Como se puede observar, el valor de la elasticidad del capital público, en general, es algo inferior al obtenido en estudios para otros países. No obstante, sigue constatándose el efecto positivo de las infraestructuras así como la reducción de las elasticidades cuando se tratan espacios económicos no muy amplios, como es el caso de las regiones.

varía sistemáticamente en el tiempo y entre regiones. La estimación con un panel de datos ofrece resultados coherentes en lo que respecta a los coeficientes de los factores privados de producción, pero conclusiones negativas en lo que concierne al impacto de las infraestructuras. Sólo la especificación en niveles presenta un coeficiente significativo para el capital público en torno a un valor del 0.06, mientras que dicha significatividad desaparece en las estimaciones que bien introducen un término autorregresivo en los errores o bien se estiman en diferencias.

#### 2.2.4 Críticas y alternativas al primer grupo de investigaciones sobre los efectos de las infraestructuras

A pesar de ser mayoritaria la obtención de efectos positivos del capital público en el crecimiento económico, tanto a nivel internacional como para la economía española, no tardaron en surgir discusiones sobre las funciones estimadas y los resultados obtenidos, pues la cuantificación del efecto de las infraestructuras resultaba demasiado elevada para ser creíble. En cierto modo, no tiene mucho sentido que la inversión en capital público tenga un impacto sustancialmente mayor en el output del sector privado que la inversión en capital privado (conclusión a la que llegaba Aschauer), especialmente si se considera que la inversión pública también está destinada a mejorar el ambiente y a objetivos sociales que no pueden recogerse directamente en las medidas del producto nacional.<sup>12</sup> De este modo, han surgido una serie de trabajos que cuestionan la plausibilidad de los resultados de las investigaciones iniciales en base a problemas econométricos y de especificación en los modelos neoclásicos más comúnmente utilizados en el estudio de los efectos de las infraestructuras.

De este modo, las críticas a la utilización de las funciones de producción Cobb-Douglas pueden agruparse en dos grandes bloques: las debidas al uso de la función de producción y al posible problema de causación inversa.

Entre las críticas del primer grupo, podemos hacer las siguientes señalizaciones:<sup>13</sup>

- 1) Partiendo de la misma función de producción estimada por Aschauer, algunos autores opinan que puede haber un problema importante de sesgo derivado de la omisión de variables relevantes correlacionadas con el stock de capital público. Por

---

<sup>12</sup> Sin embargo, los factores destinados a objetivos sociales y a mejorar el medio ambiente son tenidos en cuenta de cara a la toma de decisiones de localización de las empresas, y por tanto, en un largo plazo aunque de forma indirecta, pueden acabar afectando al nivel de producto. No obstante, esta puntualización no hace cambiar la idea apuntada en esta afirmación por la que resulta ilógico que el papel del capital público en el producto privado sea mayor que el del capital privado.

<sup>13</sup> Para una explicación detallada de las críticas planteadas a las funciones de producción ampliadas con el capital público así como las posibles soluciones a las mismas, véase Moreno (1995).

ejemplo, autores como Tatom (1991) y Holtz-Eakin (1992) consideran que los precios de la energía debieran considerarse en una especificación como la utilizada por Aschauer dado que el aumento del precio del petróleo durante los setenta convirtió en obsoleto parte del capital privado, afectando negativamente a la productividad. Así, Holtz-Eakin comenta que la caída simultánea de la inversión pública y del crecimiento de la productividad puede ser tan solo una casualidad, y que la aparente significatividad del capital público se deba a que éste sirve de proxy para alguna otra variable omitida que sería la verdadera causante del problema de que el efecto del capital público se esté sobrecuantificando. De hecho, en el trabajo de Tatom, a partir de una especificación que considera los precios de la energía y la capacidad de utilización multiplicando tanto al capital privado como al público, observa que las infraestructuras pierden toda su significación.

Esta crítica se ve reforzada si comparamos los estudios de Aschauer (1989a) y Ratner (1983). Ambos trabajan con especificaciones y datos similares, con la única diferencia de que la muestra del segundo acaba en 1973. Los resultados, sin embargo, son muy diferentes pues el coeficiente para el capital público en el trabajo de Ratner es tan solo de 0.056. Parece deducirse, por tanto, que el resultado tan elevado de Aschauer pueda ser debido a la influencia de las últimas observaciones recogiendo la crisis de los setenta.

2) Una cuestión econométrica que ha suscitado la segunda crítica a los resultados de los trabajos iniciales del efecto de las infraestructuras, se basa en la idea de que la correlación entre productividad y capital público puede ser espúrea debido a la no estacionariedad de las variables. Es lo que se conoce como *la crítica de Jorgenson*. De ser así, se estaría diciendo que, por ejemplo, posibles factores demográficos o cambios en la demanda podrían haber determinado la evolución del gasto en capital público y su coincidencia con la productividad, haciendo de la relación entre ambas variables una relación espúrea, no determinada por la causalidad sino por la casualidad. La no estacionariedad de las series invalida el análisis macroeconómico entre ambas, de forma que parte de los resultados obtenidos no serían válidos. Por tanto, partiendo de la especificación inicial resultaría necesario eliminar esta tendencia común y poder

analizar así la verdadera relación entre las variables. Esto supondría especificar la relación en términos de primeras diferencias. Cuando se toman primeras diferencias en la función de producción agregada con capital público, se obtienen resultados en los que el efecto del capital público es bastante pequeño, incluso negativo en algunos casos y generalmente no significativo estadísticamente, tal como sucede en García-Fontes y Serra (1993a, 1993b), Holtz-Eakin (1994), Garcia-Milà *et al.* (1996) y de la Fuente (1996).

La especificación en primeras diferencias tiene, sin embargo, sus propios problemas. En concreto, es difícil esperar que el crecimiento del stock de capital en un año, tanto público como privado, esté correlacionado con el crecimiento del output en ese mismo año (Munnell, 1992). Por tanto, la principal crítica que ha recibido la diferenciación se ha centrado en el hecho de que, con este proceder, se elimina toda la información sobre las posibles relaciones de largo plazo, recogidas en los niveles de las variables. En consecuencia, se plantea como una solución excesivamente drástica a la prevención de regresiones espúreas. De hecho, las ecuaciones estimadas de esta manera ofrecen resultados negativos para el capital público pero también en muchos casos obtienen valores nada plausibles para el empleo y el capital privado, sin que eso les supusiera concluir que el capital privado y el empleo no contribuyen al output del sector privado. En consecuencia, a nuestro parecer, la diferenciación estaría destruyendo justamente lo que se está intentando estimar: la relación a largo plazo entre capital público y productividad.

Sin embargo, en econometría se ha sugerido recientemente que los resultados de la estimación en niveles pueden ser bastante más fiables de lo que se creyó en un principio. En concreto, la estimación en niveles de una relación entre series no estacionarias puede ser consistente cuando las variables están cointegradas, es decir, cuando existe una combinación de éstas que sí es estacionaria. Por tanto, en los estudios con series temporales se debería examinar no solamente si las variables crecen con el tiempo, es decir, ver si son estacionarias o no, sino analizar si crecen conjuntamente en el tiempo y convergen a su relación de equilibrio a largo plazo, que es justamente la situación en la que están cointegradas. En caso de estar ante una

relación de cointegración (lo que se puede comprobar a través de alguno de los contrastes de cointegración)<sup>14</sup>, los estimadores mínimos cuadrados ordinarios serán mejores de lo habitual en muestras grandes, aunque sus distribuciones no serán las estándar, por lo que los procedimientos habituales de inferencia pierden su validez. Analizando los trabajos que han contrastado la posible cointegración entre el capital público y el output del sector privado, nos encontramos con resultados varios. Así, mientras que Sturm y de Haan (1995) concluyen que tanto en EEUU como en Holanda ambas variables no están cointegradas, Bajo y Sosvilla (1993), Argimón *et al.* (1993) y Mas *et al.* (1993) encuentran que ambas variables están cointegradas en el caso español. Recientemente, Garcia-Milà *et al.* (1996) contrastan de forma sistemática la presencia de diversos problemas de especificación, entre ellos, la no-estacionariedad de los datos en el caso de una función de producción con datos de panel para los estados americanos. A pesar de que con datos de panel parece que dicho problema debiera mitigarse por la consideración de datos no sólo temporales sino también transversales, dichos autores obtienen que las variables son no estacionarias y no están cointegradas, por lo que se inclinan por un modelo en primeras diferencias, que no ofrece resultados significativos para el efecto del capital público. De este modo, dado que en el caso español las variables parecen estar cointegradas, se mantienen las conclusiones obtenidas en la mayoría de estudios en los que la variable capital público tiene un efecto significativo en la productividad de una economía. Por contra, en la economía americana y en la holandesa, tras la consideración de las variables en primeras diferencias, dicha relación deja de ser significativa. No obstante, la gran diversidad de coeficientes estimados en los diferentes estudios no permiten dar por zanjada la cuestión.

3) Una tercera crítica surge cuando se trabaja con datos de panel, es decir, cuando se consideran observaciones en el tiempo y en el espacio, y se utilizan los métodos de

---

<sup>14</sup> Entre los contrastes más utilizados en la comprobación de la presencia de cointegración se encuentran el de Durbin y Watson de la regresión de cointegración, el test de Dickey y Fuller de los residuos de cointegración, el procedimiento de Phillips y Ouliaris y el de Máximo Verosímil de Johansen. Para más detalle consultar Suriñach *et al.* (1995).

estimación que la literatura econométrica ofrece para estos casos.<sup>15</sup> Así, en los trabajos de Evans y Karras (1994b), Holtz-Eakin (1994) y Garcia-Milà *et al.* (1996) se analizan varias especificaciones de la función de producción agregada con capital público para los estados americanos, de forma que si se estima en niveles se reproducen los resultados obtenidos por Aschauer. Sin embargo, la elasticidad del producto respecto al capital público se reduce de manera importante y deja de ser significativa cuando se introducen efectos específicos por estados. La existencia de dichos efectos es, por otra parte, más que probable tal como demuestran los resultados del test de Hausman, por lo que la fragilidad de los resultados a su inclusión debieran hacernos dudar sobre la magnitud del efecto del capital público sobre la productividad. También en esta misma línea aunque a nivel de países en lugar de estados, Evans y Karras (1994a) utilizando un panel de datos para siete países observan que la inclusión de efectos específicos temporales, o temporales y nacionales simultáneamente, hace perder la significatividad del parámetro que acompaña al capital público.

En el caso de la economía española, se han realizado estimaciones con datos de panel para las regiones en los trabajos de García-Fontes y Serra (1993a, 1993b), Mas *et al.* (1994, 1996a) y de la Fuente (1996). En la economía regional española parece ser que el resultado no es tan unánime. Así, mientras en los trabajos de García-Fontes y Serra y Mas *et al.* es precisamente la inclusión de efectos específicos, principalmente regionales, los que tienden a mejorar los resultados relativos a la elasticidad del output respecto al capital público, en el trabajo de de la Fuente las infraestructuras pierden su significatividad cuando se controla por peculiaridades regionales.

Por tanto, parece concluirse que cuando se controlan las especificaciones para tener en cuenta las posibles diferencias tanto a lo largo del tiempo como entre unidades transversales, el efecto del capital público queda puesto en entredicho.

De todas maneras, es necesario advertir de un problema potencial inherente a la estimación de funciones que incluyen el capital público. Tal y como demuestran

---

<sup>15</sup> En el Capítulo 4 se ofrece una explicación detallada de los distintos métodos de estimación disponibles para datos de panel, juntos con las referencias más interesantes.

Chunrong y Cassou (1997), la correlación existente entre el stock de capital público de las distintas economías y la variables ficticias incluidas con el fin de considerar las diferencias regionales en el modelo de efectos fijos es muy elevada. Ello puede causar estimaciones muy inestables, y con ello grandes cambios en los resultados ante pequeñas alteraciones en la muestra. No obstante, con las técnicas actuales, dicho problema parece de difícil solución.

4) En cuarto lugar, el diferente nivel de agregación de los datos se ha considerado como una posible causa de las grandes diferencias existentes en los coeficientes estimados. La justificación de dicha dispersión de resultados se ha centrado en los efectos difusión o *spillovers* entre áreas, que no quedan recogidos cuando se trabaja con áreas pequeñas. Esa excesiva dependencia de la función de producción respecto del ámbito de referencia ha hecho que muchos la consideren no apta en el estudio de los efectos de las dotaciones infraestructurales, al menos no en la manera en que comúnmente se ha venido haciendo (Munnell, 1992). A fin de contrastar dicho efecto *spillover*, se han realizado estudios en los que en lugar del capital público de cada región o estado, se utiliza como variable explicativa el stock de capital público de la región propia más el de las regiones geográficamente adyacentes (Mas *et al.*, 1994, 1996a; Holtz-Eakin y Schwartz, 1995b) o el de las regiones con las que mantienen mayores relaciones comerciales (Corugedo *et al.*, 1995). En los trabajos realizados para la economía española (Mas *et al.* y Corugedo *et al.*) se obtiene que los efectos del capital público en la productividad dependen no sólo de la dotación infraestructural de la región misma sino también de la dotación global en todo el país y especialmente en las regiones contiguas físicamente. Sin embargo, cuando Holtz-Eakin y Schwartz (1995b) examinan si el stock de autopistas en un estado americano puede suponer beneficios en la productividad más allá de sus fronteras obtienen nula evidencia en favor de dicho efecto. Recientemente, Kelejian y Robinson (1997) han considerado el efecto que el capital público de los estados vecinos tienen sobre la productividad de un estado introduciendo una variable que recoge las infraestructuras de las regiones ponderadas por la densidad de población, a fin de tener en cuenta la población que va a hacer uso de dicha infraestructura. Su resultado, sin embargo, es contrario a la hipótesis de Munnell (1992), ya que obtienen un valor significativo pero negativo para

dicho efecto *spillover*. Los autores argumentan tal resultado indicando que unas buenas infraestructuras en las áreas vecinas pueden ser un aliciente para la población a desplazarse a vivir allí, por lo que dicho trasvase de fuerza laboral supondría un efecto negativo en el estado bajo consideración. Dado este resultado, cabría pensar que la reducción en el valor regional respecto al nacional, argumentado por Munnell en base a la presencia de efectos difusión, no parece tan evidente ya que si dichos *spillovers* existen pero son negativos, el valor del parámetro del capital público a nivel nacional debiera ser menor que a nivel regional, al contrario de lo que sucede.

5) Finalmente, se ha criticado la excesiva rigidez de la función de producción. Por una parte, la función de producción se ha considerado no adecuada por imponer demasiadas restricciones sobre la tecnología y el comportamiento de las empresas, y por omitir el precio de los factores que afectan a la intensidad con que son utilizados y que no son tenidos en cuenta (Berndt y Hansson, 1992). Para superar dicha rigidez se ha sugerido el seguimiento de los desarrollos de la teoría de la dualidad y especificar funciones de costes (Berndt y Hansson, 1992; Nadiri y Mamuneas, 1991; Morrison y Schwartz, 1992; Conrad y Seitz, 1994; Seitz, 1995, entre otros) y beneficios (Duffy-Deno, 1988; Lynde, 1992; Lynde y Richmond, 1993a; Dalamagas, 1995) que reflejan el comportamiento optimizador de las empresas. En estas especificaciones, minimizando costes o maximizando beneficios es posible diferenciar el impacto directo y el indirecto del capital público así como determinar si la provisión de capital público alcanza su nivel óptimo.<sup>16</sup> En términos globales, hay que decir que la mayoría de trabajos que utilizan la teoría de la dualidad llegan a la conclusión de que la dotación de infraestructuras tiene un efecto importante en la reducción de los costes de las empresas.

Por otra parte, existen modelos que adoptan una perspectiva más global, ya que consideran que la relación entre capital público y crecimiento es demasiado compleja para ser resumida en una función de producción neoclásica. Entre ellos se encuentran

---

<sup>16</sup> Dado que la teoría de la dualidad se utiliza en la Parte III del presente trabajo, véase la misma para una explicación detallada de dicha teoría.

los modelos de equilibrio general, en los que el Estado determina el capital público sin tener en cuenta los agentes privados (García-Milà, 1990) y la teoría integral basada en el enfoque del potencial del desarrollo regional (EPDR) que parte de la idea de que la infraestructura es tan sólo uno de los determinantes potenciales del desarrollo económico (Biehl, 1986; Cutanda y Paricio, 1992a). De este modo, según el EPDR, factores como la situación geográfica, las ventajas de aglomeración y la estructura sectorial son tenidas en cuenta junto con el capital público en una función de cuasi producción. Utilizando datos para las regiones de la CEE, Biehl obtiene que el indicador de infraestructuras es significativo en la mayoría de los casos.<sup>17</sup> En el trabajo de Cutanda y Paricio se utilizan varios indicadores de stock de capital público, obteniendo que el capital público económico (el utilizado en actividades productivas) tiene un impacto positivo y significativo en la renta de las regiones españolas, mientras que el social (educación y sanidad) no.

El segundo gran bloque de críticas se centra en el conocido problema de la causación inversa, por el que se argumenta que la dirección de causalidad puede ser la contraria a la que la mayoría de autores sugieren. Para argumentar esta discusión, existen dos teorías del gasto público. Por una parte, y defendiendo que los mayores niveles de output son los que requieren mayores dotaciones de capital público, se encuentra la Ley de Wagner por la que el gasto del Estado es un bien superior, por lo que a mayor crecimiento económico corresponde un mayor gasto público, adaptándose el sector público a la evolución de la economía. Por el contrario, la teoría keynesiana del activismo fiscal nos dice que el tamaño del sector público es el que sirve como estabilizador del crecimiento. Así, la mayoría de estudios siguen esta segunda corriente teórica por lo que se trata al capital público como variación exógena en lugar de como potencialmente endógena, pudiendo estar incorporando un sesgo en las estimaciones. Aschauer (1989a) ya considera la posibilidad de la existencia de una doble causalidad y ofrece evidencia en su contra. De este modo, observa que sus resultados apenas cambian cuando se utilizan valores retardados del capital público en una estimación ordinaria en dos etapas. Asimismo, Munnell (1992) analiza tal

---

<sup>17</sup> Se habla del mayor número de casos porque en el trabajo de Biehl se calcula un gran número de combinaciones posibles de variables exógenas y endógenas.

problemática reestimando algunas de sus ecuaciones con datos regionales incluyendo el valor del capital público al principio del periodo considerado, lo que excluye la posibilidad de un efecto *feedback* del crecimiento del producto en la inversión en capital público. A pesar de ello, la variable capital público continúa presentando un efecto positivo importante y significativo. En otros trabajos, como Argimón *et al.*, (1993), Bajo y Sosvilla (1993) y Mas *et al.*, (1994) se realizan contrastes de exogeneidad de Hausman para comprobar la posible simultaneidad entre ambas variables, no pudiendo rechazar en ningún caso la hipótesis de exogeneidad. A la misma conclusión llegan Bajo y Sosvilla mediante el contraste de causalidad de Granger, ya que el capital público causa en el sentido de Granger al capital privado, pero no al contrario. Finalmente, existen trabajos que recogen la posible doble relación de causalidad entre el stock de infraestructuras, por un lado, y la renta per cápita, por otro, a través de modelos de ecuaciones simultáneas que permiten considerar los efectos de retroalimentación (Flores *et al.*, 1994). En dicho trabajo se obtiene que no sólo el capital público determina el comportamiento de las variables del sector privado, sino que también retardos de las variables privadas determinan el nivel del stock de capital público. En este caso, la inversión pública se comportaría como un bien normal, por lo que introducen una ecuación de su demanda en función de la renta de forma que, de esta manera, se tiene un contexto multiecuacional dinámico.

Como conclusión, y a pesar de que los distintos procedimientos sugeridos para afrontar todos estos problemas no se pueden considerar todavía resueltos, no se puede desechar de forma radical la función de producción como enfoque metodológico para estimar los efectos de las infraestructuras públicas. Ciertamente, las críticas tienen cuerpo propio en el sentido de que los resultados obtenidos de los estudios con series temporales agregadas no son del todo creíbles y que se requiere una mayor evidencia en el tema de la causación inversa. No por ello, sin embargo, se ha de concluir que el resultado obtenido en un principio, favorable al efecto positivo del stock en infraestructuras sobre el crecimiento económico, sea falso, sino que resulta necesario buscar las posibles alternativas que solucionen las limitaciones que realmente posee la función de producción. De hecho, no se puede dar por zanjada la cuestión ya que los resultados varían mucho de un estudio a otro y no siempre resultan razonables.

Por tanto, a pesar de que no en su totalidad, la mayor parte de los primeros estudios sobre los efectos de las infraestructuras, tanto los que utilizan la teoría neoclásica de producción como los que han intentado solucionar las críticas atribuidas a las mismas mediante la utilización bien de funciones de costes bien basándose en el enfoque del potencial de desarrollo regional, obtienen un efecto positivo entre infraestructuras y productividad tanto a nivel agregado como por regiones. Igualmente, en los pocos casos en que se han estimado modelos multiecuacionales, que permiten establecer relaciones simultáneas dinámicas entre el crecimiento económico y el capital público, se han obtenido resultados también favorables aunque no siempre contundentes. Sin embargo, recientemente se han realizado estimaciones del efecto del capital en infraestructuras utilizando la teoría neoclásica de producción a nivel regional y controlando al mismo tiempo las posibles diferencias entre regiones. En la mayoría de estudios para la economía americana, el coeficiente del capital público deja de ser significativo cuando se controlan dichos efectos, mientras que para la economía española el resultado no es contundente. Por tanto, si bien en un principio parecía existir unanimidad en el efecto positivo de las infraestructuras públicas en el crecimiento regional, los resultados posteriores no son tan optimistas. En opinión de algunos autores (Sturm y de Haan, 1995; de la Fuente, 1996) la conclusión más válida ante esta gran disparidad de resultados es pensar que la forma más sencilla de abordar la medición del impacto de las infraestructuras no ha funcionado bien, dada la enorme complejidad del tema.

### **2.3 OTRAS CUESTIONES RELATIVAS AL IMPACTO DEL CAPITAL PÚBLICO**

En un primer apartado de esta sección se ofrece una discusión de los que pueden ser considerados los principales efectos de las infraestructuras en el sentido más amplio posible. Dicha presentación nos permitirá considerar la complejidad del tema que se está tratando dados los múltiples efectos que las infraestructuras públicas suponen, siendo difícil el tratamiento conjunto de todos ellos. No obstante, la explicitación de los mismos resulta interesante en cualquier trabajo que se centre en el papel del capital público.

En segundo término, teniendo en cuenta que los resultados obtenidos en algunos estudios empíricos no son excesivamente favorables a una relación directa y positiva entre capital público y crecimiento de la productividad, es importante analizar las conclusiones a que se llega cuando se analizan canales indirectos en el impacto de las infraestructuras en la productividad industrial, principalmente. Entre dichos canales, nos centraremos en el efecto que el capital público puede suponer en la localización de la actividad económica en general, así como en las cantidades y productividades de los factores privados utilizados en el proceso productivo.

Posteriormente, se analizan las aportaciones más recientes sobre el papel que el capital público desempeña en el proceso de convergencia de las economías regionales. Esta cuestión resulta tanto más importante en el ámbito europeo en el que los Fondos de Desarrollo Regional dedican gran parte de sus recursos a la provisión de las infraestructuras necesarias para no limitar el crecimiento de las regiones pobres avanzando, de este modo, en el proceso de convergencia, uno de los objetivos claves del proyecto europeo actual.

Asimismo, parece difícil que puedan darse conclusiones válidas sobre el papel del capital público sin tener en cuenta factores como el tamaño ya existente del stock de capital público en el área en consideración, el grado de utilización de la red ya existente, el nivel de desarrollo económico de la zona, la etapa del ciclo económico en que se encuentra la economía, la existencia de economías de aglomeración, las características del tejido industrial como, por ejemplo, el tamaño de las empresas, entre otros. Estos condicionantes del papel de las infraestructuras han sido analizados de forma específica en diversas investigaciones, resumiendo aquí las principales conclusiones.

Finalmente, en estos últimos años se ha generalizado entre los economistas teóricos la idea de que, si bien el capital público en infraestructuras es un requisito para el desarrollo económico, de poco sirve si no se encuentra acompañado de otros factores que conjuntamente atraen y mantienen la actividad privada. De esta manera, deja de ser considerado como la panacea para conseguir una senda de crecimiento sostenido, convirtiéndose en una *condición necesaria pero no suficiente* para el crecimiento de la producción. Dicha idea es también analizada.

### 2.3.1 Vías de influencia teórica del capital público

Para entender los efectos del capital público en infraestructuras sobre el crecimiento y desarrollo económico se ha de comprender el proceso de desarrollo regional. Dado que sobre dicho tema se ha escrito y discutido ampliamente, tan solo vamos a realizar una descripción general que nos permita centrar los efectos de las infraestructuras. Así, siguiendo la pauta de Eberts (1991) se puede decir que una región crece de dos maneras. En primer lugar, un aumento de los inputs supondría un incremento del output, manteniendo todo lo demás constante. En segundo lugar, a pesar de que los inputs se mantuvieran invariables, pudiera suceder que una misma cantidad de inputs sirviera para producir más, gracias a cuestiones varias como los avances tecnológicos, la mejor preparación de los trabajadores, los incrementos de las economías de escala a nivel de la empresa o de las economías de aglomeración del área. Generalmente se consideran dos inputs: el capital privado y el empleo.

Dentro de la primera causa del crecimiento, se puede decir que los inputs pueden incrementar tanto por factores internos a la región como externos. Entre los factores internos que afectarían al trabajo, por ejemplo, podríamos considerar todo tipo de características demográficas como incrementos en la tasa de nacimiento o la tasa de participación en el trabajo. Por lo que respecta al capital, aparecen factores como una elevada tradición empresarial, la disponibilidad de buenas universidades o centros de investigación en el área o simplemente factores naturales o históricos. Por otra parte, las fuentes externas de crecimiento se relacionan con el hecho de que los factores de producción son móviles y fluyen entre las regiones en busca de la tasa de rendimiento más elevada. Así por ejemplo, los trabajadores migran a aquellas regiones en las que pueden obtener un salario más elevado, suponiendo que el resto de características regionales son iguales. Sin embargo, dichas características no son iguales en todas las regiones, inclinando la decisión de traslado de un trabajador y su familia de un área a otra. Así, por ejemplo, un factor importante en esta elección será la disponibilidad de instalaciones o servicios en la zona, tales como las oportunidades de ocio, atracciones culturales, disponibilidad de buenas escuelas, un buen transporte público y buenas comunicaciones con otras zonas. De la misma manera, las empresas se localizarán (y por tanto, el capital privado) en aquellas

regiones que pueden proporcionar una tasa de rendimiento más elevada, lo cual está directamente relacionado con el hecho de tener los costes de producción más bajos. Así, un área puede tener una ventaja en costes bien porque sus salarios u otros factores tienen precios más bajos o son de mayor calidad o bien porque los costes de transporte son más bajos. Se observa, por consiguiente, como una buena dotación de infraestructuras en un área influirá tanto en la atracción de mano de obra como de capital privado, con la posible repercusión en el crecimiento y desarrollo de la zona.

En segundo lugar, una región puede crecer a pesar de mantener constante la dotación de inputs gracias a que dichos factores pueden ser utilizados de una manera más eficiente. Entre las razones que pueden generar dicha mayor eficiencia, se puede pensar en la aparición de avances tecnológicos, tanto a través de la introducción de nuevo capital físico como de un mayor nivel de preparación de los trabajadores, por el aumento de tamaño de las empresas con la consiguiente reducción en los costes unitarios de producción (economías de escala) o el tamaño del área metropolitana con las consiguientes economías de aglomeración. El hecho de que en una zona se concentre la producción comporta el beneficio de disponer de bienes y servicios con menores costes de transporte y de tiempo. Asimismo, una empresa podrá decidir localizarse en un área no metropolitana si las comunicaciones con las áreas más cercanas son buenas y no suponen un incremento considerable en los costes. De esta manera, la empresa podrá beneficiarse de unos menores costes de instalación y de iniciación del negocio, junto con la disponibilidad de estar en fácil y barato contacto con zonas metropolitanas.

A través de estos dos canales de crecimiento regional, se observa como las infraestructuras tienen un papel importante en el mismo. De esta manera, pasamos a analizar, de forma detallada, las vías a través de las cuales las inversiones en infraestructuras afectan al desarrollo económico. A grandes rasgos se puede diferenciar entre los efectos de demanda y los de oferta.

Por el lado de la demanda, siguiendo a Herce y Sosvilla (1995), la misma realización de gasto en capital público supone fuertes efectos sobre la economía en su conjunto, a raíz de la capacidad de generación de demanda a otros sectores productivos (el caso típico de la

construcción). En consecuencia, se produce una reactivación sobre el empleo, la renta y los precios y salarios en el corto plazo pero sin afectar al nivel de equilibrio. De hecho, se puede pensar en las infraestructuras como un instrumento contracíclico (Kessides, 1996) en momentos de escasa demanda, de forma que el efecto multiplicador de las infraestructuras se realiza casi en su totalidad en el corto plazo, es decir, en el momento en el que se realizan las inversiones (Duffy-Deno y Eberts, 1991). En este sentido, Bean (1996) estimando una función de producción ampliada con capital público a través de vectores autorregresivos, obtiene un elevado efecto positivo del mismo sobre el capital privado y sobre el producto en el corto plazo, que desaparece completamente en la solución de largo plazo.

Desde el punto de vista de la oferta, las infraestructuras permiten reducir los costes de transporte y de otros servicios de comunicación que soportan las empresas, mejorando la competitividad de las mismas y aumentando, de este modo, la eficiencia del sistema productivo, efecto que se mantendrá en el largo plazo. Asimismo, este proceso estimula la inversión privada, nacional y extranjera, ya que la mayor eficiencia del sistema y la posibilidad de ser competitivo en el mercado empuja a las empresas a aumentar sus operaciones y a atraer a otras nuevas, con el consiguiente aumento en la demanda de empleo y los salarios del sector industrial. A través de un proceso multiplicador, este mayor salario del sector industrial conducirá a un aumento de la renta personal, sentando la base para un futuro crecimiento económico. Del mismo modo, los efectos de una nueva infraestructura, gracias a la generación de mayor actividad que suponen, comportan un incremento de los ingresos fiscales para el Estado (IVA, impuesto de sociedades, impuesto sobre la renta) y las corporaciones locales (IAE, IBI, licencias sobre construcciones, plusvalías, etc.).<sup>18</sup>

En último lugar, la inversión pública puede tener un efecto en el desarrollo económico a más largo plazo induciendo cambios estructurales en el área bajo consideración. Así, las infraestructuras afectan de forma importante a la habilidad de respuesta de la actividad privada a cambios en los precios de sus factores. Por ejemplo, la expansión de los sectores

---

<sup>18</sup> Se están exponiendo los beneficios de las inversiones públicas, si bien es evidente que a este beneficio bruto se le han de restar los costes que han supuesto.

de servicios, tecnológicos y financieros relacionados con la producción industrial necesitan cada vez más de las telecomunicaciones, que pueden considerarse como las nuevas infraestructuras, siendo menos dependientes de las infraestructuras de transporte. Asimismo, las infraestructuras permiten cambios importantes en economías rurales, permitiendo su acceso a inputs más baratos y la introducción de prácticas rurales que incorporan fuertes avances tecnológicos. Estos procesos permiten cambios en la estructura de producción y de consumo así como la diversificación de la economía.

Finalmente, las infraestructuras suponen efectos a nivel microeconómico. La población se beneficia de unas mejores dotaciones de infraestructuras de transporte, por ejemplo, a través de ahorros de tiempo, mejora de la seguridad, mayor comodidad, calidad del servicio, disminución del ruido y de la contaminación atmosférica en las zonas urbanas, lo que incluso puede suponer cambios en el comportamiento de los usuarios.

Todo este entramado de relaciones, directas e indirectas, entre las infraestructuras (principalmente las de transporte), y el desarrollo económico regional<sup>19</sup> ha sido presentado de forma esquemática por Bruinsma *et al.* (1995) (Figura 2.1).

Entre las *relaciones directas* cabe considerar que, como consecuencia de la construcción de las infraestructuras, se ven reducidos los costes de transporte que deben soportar las empresas gracias a la reducción de distancias y al aumento de la velocidad promedio (relación 1). Esto supondrá cambios en la elección del medio de transporte, de la ruta, el tiempo de trayecto (tanto mayor en caso de tener redes congestionadas) y la generación o atracción de nuevos movimientos en el área (relación 2).

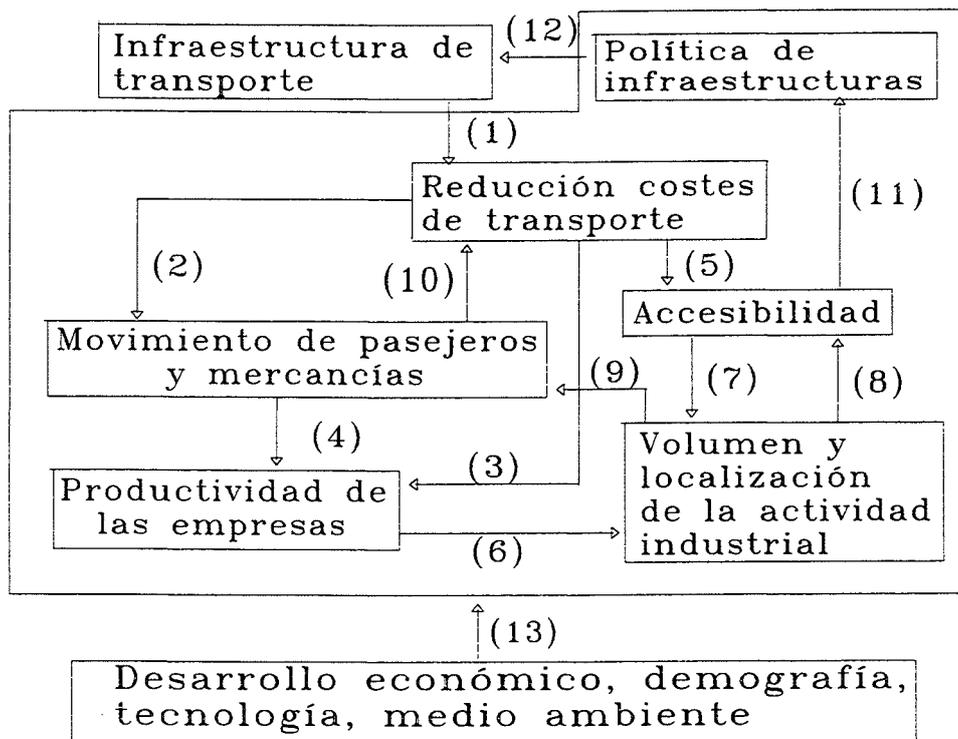
La reducción de los costes de transporte junto con los cambios en los modelos de traslado de las familias y de las empresas supondrán un aumento de la productividad en las zonas

---

<sup>19</sup> Parte de la evidencia que se muestra en este apartado se refiere a las infraestructuras de transporte. La razón se debe a que son éstas, probablemente, las que más afectan en las decisiones de localización de la actividad económica y, por tanto, en el crecimiento regional. Asimismo, dado que el mayor porcentaje del capital público lo representan los transportes (autopistas, carreteras, puertos, aeropuertos y ferrocarriles), la mayoría de las conclusiones que aquí se obtienen son generalizables al estudio de los efectos de las infraestructuras en general.

afectadas (relaciones 3 y 4). Otra consecuencia es el aumento de la accesibilidad de la zona (relación 5). Esta mejora de la productividad y accesibilidad se traducirá en una expansión de las actividades económicas y/o población en la zona (relaciones 6 y 7).

**Figura 2.1** Modelo conceptual sobre la relación entre infraestructuras de transporte y la distribución espacial de la actividad económica



Fuente: Bruinsma *et al.* (1995)

Sin embargo, los vínculos entre infraestructuras y crecimiento económico no siempre son tan directos, sino que pueden afectar más sutilmente al sistema económico. Así, un primer *ligazón indirecto* es el impacto que el capital público tiene sobre la relocalización de la actividad industrial. En consecuencia, unas determinadas áreas se afianzan como aglomeración de actividad económica que se benefician de la generación de economías externas. Esto puede tener, asimismo, efecto en la accesibilidad de las zonas. No obstante, se ha de tener en cuenta que el efecto final dependerá también de la relación con otros factores que influyen, también de forma importante, sobre las decisiones de localización (por ejemplo, formación y cualificación de la mano de obra, adecuadas condiciones

institucionales, estabilidad política y económica, etc.), que hacen aun más atractivas las economías de aglomeración.

Los cambios en la localización de la actividad económica influyen en el número de movimientos de pasajeros y de mercancías (relación 9), que en caso de congestión puede suponer cambios en los costes de transporte (relación 10).

Según estos autores, la provisión de infraestructuras no debe verse como un factor exógeno que está desarrollado por el Estado. La política de infraestructuras debe tener como principal objetivo el asegurar un determinado nivel de accesibilidad para cada región (relaciones 11 y 12). Esta accesibilidad les servirá, por una parte, a las regiones pobres para librarlas del posible aislamiento que les impediría relanzar cierta actividad económica y, por otra parte, para evitar problemas de congestión que pueden afectar a regiones con un desarrollo económico importante. Un último aspecto a señalar es que, evidentemente, no son sólo las infraestructuras las que tienen un importante efecto en la distribución espacial de la actividad económica. Aspectos como la tecnología, demografía, estructura sectorial, política pública y medioambiental y la conexión con centros dinámicos son importantes en la localización de la actividad (relación 13).

Principalmente, estos factores indirectos (en su mayoría efectos *feedback*) son los que conforman el contexto en el que se basan algunos de los trabajos que relacionan las infraestructuras de transporte y la distribución de las actividades económicas en el espacio. En los siguientes apartados se tratan algunos de estos vínculos.

### **2.3.2 Efectos de las infraestructuras en la localización de la actividad económica**

Las infraestructuras pueden afectar las decisiones de localización de las empresas de diferentes maneras (Eberts, 1991). En primer lugar, el hecho de que un área esté dotada de unas buenas infraestructuras hará que la zona sea más atractiva para las empresas por los menores costes que la producción y el traslado de los productos les supone. Incluso se puede pensar que el capital público hará reducir los costes del trabajo ya que el hecho de

poder disfrutar de unas adecuadas infraestructuras para las familias hará que éstas estén dispuestas a aceptar unos menores salarios en contrapartida, con la consiguiente reducción de costes en las empresas. De hecho, algunos estudios como el de Beeson y Eberts (1987) muestran que los salarios son menores en aquellas áreas que tienen una buena dotación de infraestructuras. En segundo término, algunas infraestructuras como las de transporte, pueden considerarse como un input directo en el proceso productivo, de forma que las empresas se trasladan allí donde pueden conseguir dicho input con una mayor calidad. De este modo, cualquier empresa que entrara en una región con buenas infraestructuras obtendría un beneficio de acuerdo al valor de la contribución del capital público en la producción. Finalmente, cabe pensar que las infraestructuras pueden aumentar la productividad de los otros inputs privados. Así, por ejemplo, en caso de que el capital público fuera un factor sustitutivo del empleo o del capital privado, las empresas considerarían que los incrementos en la cantidad de capital público proveen los mismos servicios que dicho input privado.

A pesar de que a nivel teórico parece bastante lógico que las infraestructuras pueden afectar las decisiones de localización de las empresas, los trabajos empíricos que han analizado el papel de ciertas medidas de actuación pública, entre ellas la provisión de infraestructuras, sobre la localización de la actividad no resultan tan contundentes. Así, Fox y Murray (1990) consideran el efecto que un conjunto de herramientas públicas, entre ellas diferentes formas de imposición o de transferencia de renta y gastos en inversión en transporte y educación, pueden tener sobre la tasa de entrada de empresas en el estado de Tennessee en el periodo 1980-86. En general, obtienen que la tasa de entrada de empresas responde muy poco a cambios en dichas variables, si bien hay que señalar que los resultados son muy sensibles a aspectos tales como el tamaño de las empresas. Sin embargo, como excepción a esta falta de efectividad de las políticas públicas, dichos autores encuentran que la entrada de empresas responde significativamente tanto a índices físicos del stock de infraestructura de transporte local como al nivel de educación de los trabajadores. En el caso concreto de las infraestructuras, obtienen que la presencia de una autopista interestatal o una línea de ferrocarril aumenta la tasa de entrada de nuevas empresas.

En un trabajo similar, Eberts (1991) confirma los resultados de Fox y Murray (1990), si bien en este estudio se utilizan medidas monetarias del stock de infraestructuras en lugar de índices, incluyendo autopistas e infraestructuras hidráulicas en 40 áreas metropolitanas de EEUU durante 1976 y 1978. Analizando la apertura de empresas a través de una ecuación de localización, obtienen que dicha tasa aumenta significativamente cuanto mayor sea la dotación de infraestructuras, aunque tal vínculo es sólo cierto para las empresas pequeñas. Un resultado adicional se centra en el hecho de que simples aumentos en la inversión en infraestructura pública no son suficientes para inducir la apertura de empresas, sino que la inversión pública debe suponer un incremento neto del stock de capital para que se obtenga tal resultado.

El trabajo de Fox y Murray (1991) en la misma línea que los anteriores, considera el efecto que diferentes tipos de políticas públicas pueden tener en la entrada de empresas mediante la estimación de un modelo TOBIT. Si bien consideran actuaciones públicas tanto en el corto plazo (principalmente actuaciones de política fiscal) como en el largo plazo (dotación de infraestructuras y educación), los autores concluyen que la influencia de una única política en la creación o localización de las empresas es muy pequeña, de forma que sólo afectaría en el largo plazo y siempre que se aplicara de forma conjunta a otras medidas. En opinión de los autores, esto se debe a que la mayoría de las políticas públicas influyen en la localización hasta el punto que dichas medidas de actuación difieren entre áreas (dichos autores están considerando, concretamente, los condados americanos), ya que de este modo el efecto consistirá en redistribuir la actividad. Así, concluyen que las políticas de dotación de infraestructuras en el largo plazo no suelen aumentar la actividad económica pero sí que la redistribuyen en el espacio.

Finalmente, existen varios trabajos que se basan en los resultados obtenidos en diferentes encuestas a industrias para analizar cuales son los factores que más influyen en las decisiones de localización de sus fábricas. Calzonetti y Walker (1991) ofrecen una discusión sobre los factores de localización que pueden verse influidos por la intervención pública basándose en una muestra nacional de plantas manufactureras que iniciaron sus operaciones después de 1978 en EEUU. Entre sus resultados obtienen que las autopistas y las obras hidráulicas juegan un papel importante para las empresas a nivel local. Moore et

al. (1991) utilizan los resultados de una encuesta de casi 1400 compañías industriales y de servicios en la UE, a través de la cual se perseguía el objetivo de analizar cuales son los factores que más influyen en las decisiones de localización de las PYMEs en áreas asistidas por la UE. El factor más importante resulta ser la disponibilidad de asistencia al desarrollo regional, seguido del acceso a los clientes y la calidad y el tamaño de la oferta de empleo (incluyendo los niveles de salarios) y la disponibilidad de buenas condiciones para una futura expansión. Por el contrario, la disponibilidad de infraestructuras resulta ser de poca importancia como determinante locacional. Sin embargo, ahondando más en las respuestas relativas a las infraestructuras, se observa que muchas empresas no están satisfechas con la calidad de algunas de sus infraestructuras locales, lo que refuerza la idea de que existe un cierto nivel de infraestructuras necesario en una zona para que la empresa decida localizarse en la misma, pero una vez se ha sobrepasado dicho umbral las empresas se preocupan más por otros factores de localización. No obstante, esto no impide que las empresas una vez instaladas en un área se den cuenta de su inadecuación, por lo que se concluye que existen ciertas áreas de la UE donde se deberían hacer esfuerzos en mejorar la provisión de infraestructuras, principalmente de transporte. Button *et al.* (1995) analizan los resultados de una encuesta a la región Strathclyde de Escocia. En el análisis que ofrecen, las comunicaciones por carretera aparecen como el factor de localización que se considera más importante, seguidos de un conjunto de factores de accesibilidad tales como los mercados, acceso a proveedores y servicios de soporte. A continuación, se encuentran otros factores específicos de transporte como la comunicación por autobús o por aire, situando al ferrocarril como el medio de transporte menos significativo. Otros factores como los costes de alquiler y las posibilidades de aparcamiento aparecen a continuación, encontrándose en último término razones geográficas tales como la imagen del área y la disponibilidad de asistencia del Estado. En lo que a infraestructuras se refiere, se ha de remarcar que el 81 % de las empresas catalogaron la influencia de las comunicaciones por carretera como importantes o muy importantes en la elección de su actual localización.

Esta elevada significación de las infraestructuras de transporte, especialmente por carreteras, ha sido relativizada en algunos trabajos en los siguientes términos.

En primer lugar, tal y como concluyen Button *et al.* (1995), las infraestructuras de transporte juegan un papel relativamente pequeño como causante de la relocalización de la actividad industrial. Es decir, si bien unos buenos medios de comunicación por carretera parecen resultar relevantes en la atracción de las empresas a su localización presente, generalmente no es un factor determinante en hacer que se trasladen una vez ya establecidas. Por tanto, unas pobres o inadecuadas infraestructuras de transporte no parecen estimular la migración de las empresas, pero una buena calidad de las mismas es un elemento importante en los criterios de selección, una vez que se ha tomado la decisión de cambiar la ubicación de la empresa.

En segundo lugar, si bien para las empresas es imprescindible que exista un cierto nivel de infraestructuras en el lugar que elijan como ubicación de su actividad, adiciones a ese nivel base de infraestructuras son de escasa o nula importancia para las empresas. De este modo, algunos autores como Simon (1978) consideran que existe un concepto de satisfacción y no de maximización, en el sentido de que las empresas, dentro de una escala de costes y un abanico de posibilidades de localización, siguen una lógica intuitiva en la noción de una racionalidad limitada y el encuentro de unos objetivos umbral, antes que la búsqueda de las localizaciones minimizadoras de costes de transporte. Por otra parte, según Duffy-Deno (1991) cabe pensar que las ventajas de localización debidas a las infraestructuras pueden llegar a anularse. Para entender esto, se ha de partir de la idea de que una región atrae a las empresas si el beneficio neto que pueden extraer del capital público, es decir, la diferencia entre los servicios que provee dicho capital y lo que se paga por él en forma de impuestos, es positivo. Pero en la medida en que nuevas firmas se localizan en una región, este beneficio neto irá reduciéndose debido al mayor valor que supondrá la tierra o por un menor beneficio obtenido de las infraestructuras, como consecuencia de la congestión en el uso de las mismas. Por tanto, las empresas pagan adicionalmente por este capital público vía reducciones en el beneficio neto que obtienen. De cualquier modo, en el equilibrio a largo plazo los beneficios se igualarán a través de las regiones y en el margen, las ventajas de localización debido al capital público tenderán a diluirse o incluso eliminarse.

Finalmente, algunos inversores están interesados no únicamente en la modernidad y efectividad general de las infraestructuras en la región sino en el grado en que ellos pueden

ejercer control sobre el desarrollo presente y futuro de las infraestructuras. Según Peck (1996) las infraestructuras son, de este modo, algo más que un instrumento físico construido por la inversión pública para obtener objetivos de desarrollo regional; se convierten en una “arena” en la cual el poder y el control puede ejercitarse en la definición del uso del espacio. Es por ello que según Peck las infraestructuras pueden tener un papel como determinante de la distribución regional de la nueva inversión en la zona.

A modo de conclusión, se ha de decir que a pesar de que la evidencia demuestra que medidas de política pública como la provisión de infraestructuras parecen influir en las decisiones de localización de la actividad económica, es importante hacer énfasis en el hecho de que las respuestas estimadas son pequeñas y dependen principalmente de factores como el tamaño de las empresas, siendo las pequeñas las que parecen verse más beneficiadas. Por otra parte, estos estudios buscan una respuesta a la pregunta de si las infraestructuras pueden afectar en las decisiones de localización, pero no nos ayudan a entender el proceso por el que las infraestructuras influyen en dichas decisiones de ubicación, aspecto que es tratado en algunos trabajos de modelización económica, como los de Holtz-Eakin y Lovely (1996) y Martin y Rogers (1995), que relacionan las infraestructuras con la localización de la actividad industrial, obteniendo resultados interesantes.

Holtz-Eakin y Lovely (1996), mediante un estudio de la productividad del capital público en un contexto de equilibrio general, encuentran evidencia de la ausencia de efectos directos sobre el crecimiento del producto, así como de la existencia de un impacto positivo sobre la variedad de productos industriales medida por el número de establecimientos. El modelo que proponen trata de explicar los mecanismos por los que las infraestructuras afectan a las empresas, los mercados y la producción de equilibrio de cada sector en la economía. Para ello parten de una economía con dos sectores, el industrial y el no industrial, y dos bienes, uno agrario y otro final industrial, mientras que tanto el capital privado como el empleo son móviles entre los dos sectores y se encuentran distribuidos en mercados de factores competitivos en cada sector. El producto industrial final se fabrica a partir de los componentes industriales intermedios (el producto agrario es utilizado con una tecnología de rendimientos constantes para fabricar productos que serán utilizados en la

producción de bienes intermedios) según la función de producción de elasticidad de sustitución constante (CES):

$$M = n^\alpha \left[ \sum_{i=1}^n \frac{x_i^\beta}{n} \right]^{\frac{1}{\beta}} \quad (2.4)$$

donde  $x_i$  es el input  $i$  en la producción del producto industrial  $M$ ,  $n$  es la variedad de componentes que se producirán,  $\alpha$  es una medida de las economías de escala con respecto al número de productos intermedios ( $\alpha > 1$  indica rendimientos crecientes a la variedad) y  $\beta$  una medida del grado de diferenciación entre cada par de inputs, siendo la elasticidad de sustitución entre cada par  $1/(1-\beta)$  (mayor  $\beta$ , menor diferenciación entre componentes).

De este modo, si un mayor número de variedades de componentes suponen un entorno con un mayor número de empresas, se obtendrá un aumento de la productividad de los factores. Es en este punto que la cuestión de cómo afectan las infraestructuras presenta un mayor interés. Así, dado que se asume que todas las componentes tienen la misma tecnología, en el equilibrio se produce una cantidad igual para cada una,  $x_0$ , con lo que se puede transformar la función de producción en:

$$M = n^\alpha x_0 \quad (2.5)$$

Por tanto, el producto final es homogéneo en  $x_0$  pero homogéneo de grado  $\alpha$  en  $n$ , obteniendo, en este caso, economías externas al sector industrial final. Esto nos llevaría a la posibilidad de que incluso en el caso de que el capital público no alterara  $x_0$ , sí puede elevar  $n$  y, por tanto, obtener mejoras en la productividad.

En este modelo, ¿cuál sería el efecto de un incremento de la dotación de infraestructuras? En el caso del sector industrial, la provisión de capital público altera la estructura de costes de cada productor de componentes y, por tanto, sus niveles deseados de producto:

$$\hat{x}_0 = \frac{-G}{b-G} \hat{G} + \frac{s}{a-s} \hat{s} \equiv -\delta_G \hat{G} + \delta_s \hat{s} \quad (2.6)$$

siendo  $\hat{\cdot}$  la tasa de variación,  $G$  una medida de la reducción en los costes fijos provocada por el incremento en capital público,  $s$  un subsidio implícito en los costes variables, consecuencia asimismo, de un incremento de capital público, y  $a, b > 0$  indican, en la forma funcional  $ax + b$ , que existen rendimientos a escala a nivel de empresa.

Según la ecuación (2.6) el efecto sobre el nivel de output de las empresas de productos intermedios será distinta si el capital público afecta a los costes fijos o a los variables, incentivando, en el primer caso, la entrada de nuevos productores y la consiguiente reducción del output de los demás.

A partir de varias ecuaciones del modelo les es posible estudiar el efecto que la provisión de capital público supone sobre el precio de los componentes y su variedad y la proporción de recursos destinada al bien industrial. Así, por ejemplo, la provisión de capital público supone un traspaso de factores primarios a la producción de componentes, aumentando, por consiguiente, el coste marginal de los factores intermedios que los productores tratarán de trasladar en el precio con un determinado *markup*. Sin embargo, este incremento en el precio de los componentes debe estar acompañado de un aumento en el número de variedades si la economía quiere seguir siendo competitiva en producción industrial final. Asimismo, dado que las empresas disfrutaban de un determinado *markup*, el sector de componentes tendrá beneficios con el número inicial de empresas, induciendo la entrada de otras nuevas. En conclusión, un incremento de la dotación de capital público incrementa el número de productores de componentes y favorece las economías externas del sector industrial final. Además, si las firmas tienen suficiente poder de mercado para forzar mayores *markups*, habrá una contracción en la producción de componentes y del producto final. Sin embargo, se dará un impacto distinto según el poder de mercado que las empresas posean.

Tal y como el modelo preveía, en la parte empírica de su artículo, Holtz-Eakin y Lovely (1996) obtienen que el impacto de las infraestructuras sobre el output por empresa es pequeño e impreciso, implicando que los efectos de la reducción de costes fijos compensan los de la reducción de los variables. Sin embargo, sí que comprueban que existe una relación positiva entre el número de establecimientos y el capital público y que éste

contribuye significativamente al crecimiento del número de los mismos. De este modo, en la medida en que la variedad de productos suponga rendimientos a escala, las infraestructuras pueden aumentar la productividad del sector de inputs intermedios manufactureros. Todos estos resultados, no obstante, se encuentran limitados por una de sus hipótesis iniciales, que es la consideración de un entorno de competencia monopolística, ya que en este caso la reducción de costes que experimentan las empresas no se trasladan a los consumidores.

Estos resultados están relacionados con las conclusiones presentadas por Martin y Rogers (1995), que obtienen que las infraestructuras<sup>20</sup> favorecen el comercio de la región atrayendo nuevas empresas. Martin y Rogers utilizan un modelo propuesto por Krugman (1993) del tipo de economía abierta para poder introducir los procesos de integración. Con el objetivo de estudiar los efectos del capital público en el comercio, la localización y el bienestar, diferencian entre dos tipos de infraestructuras: las *domésticas* son aquéllas que favorecen los intercambios en el interior de un territorio mientras que las *internacionales*, situadas dentro o fuera de un territorio, favorecen el intercambio entre regiones. La hipótesis básica que utilizan es que la dotación de capital público afecta a los costes de comercio, tanto en el interior como entre regiones. Es precisamente en este punto en el que cabe recordar la advertencia de Krugman sobre el hecho de que un incremento del nivel de infraestructuras en una región pobre, al disminuir los costes de transporte, podría hacer que las empresas abandonaran la región pobre y se situaran en una rica donde poder aprovecharse de las economías de aglomeración y las economías externas de los distritos industriales, y desde donde ahora se puede suministrar al mercado pobre de forma barata.

En su modelo, los autores consideran dos países con dos productos, uno básico ( $Y$ ) y otro compuesto de un número determinado de productos diferenciados ( $D$ ). De este modo, el consumidor elige entre ambos productos dando lugar a las siguientes demandas de los bienes producidos en el país doméstico ( $D_i$ ) y en el país foráneo ( $D_j$ ):

---

<sup>20</sup> En el modelo de dichos autores las infraestructuras son entendidas en sentido amplio, incluyendo además de las típicas, la legislación, las condiciones sociales, políticas, etc. Esto es así por la forma en que introducen el capital público en el modelo, cuya función básica es la de disminuir el coste de los intercambios comerciales tanto internos como externos de una región.

$$D_i = \frac{\sigma - 1}{\beta \sigma} \frac{\rho_D \alpha I}{\tau_D (n \rho_D + n^* \rho_I \rho_I^*)} \quad (2.7)$$

$$D_j = \frac{\sigma - 1}{\beta \sigma} \frac{\rho_I \rho_I^* \alpha I}{\tau_I \tau_I^* (n \rho_D + n^* \rho_I \rho_I^*)} \quad (2.8)$$

donde  $\beta$  es el coste laboral unitario incorporado en cada bien;  $\sigma$  es una medida de la sustituibilidad entre bienes que, como en los modelos de Krugman (1992), se puede interpretar como una medida de las economías de escala;  $I$  es el ingreso individual, de forma que  $I = 1 + r K/L$ , siendo  $K$  las unidades de capital privado en la economía del país doméstico (una unidad de  $K$  produce una variedad de un bien, por lo que el número total de variedades iguala la dotación total de  $K$ ). De esta manera, los consumidores poseen y ofertan 1 unidad de trabajo inelásticamente y  $K/L$  unidades de  $K$ , retribuidas según su rendimiento,  $r$ , en el caso del  $K$ , cuando  $L$  es la población total del país doméstico;  $n$  y  $n^*$  es el número total de bienes en el país propio y en el foráneo, respectivamente, que se asume igual al de empresas dado que cada una de éstas produce un bien diferenciado. Los costes de transporte los introducen a través de la forma de *iceberg* típica del modelo de Samuelson:

- una cantidad  $1/\tau_D < 1$  de los bienes domésticos adquiridos están disponibles finalmente para el consumo
- una cantidad  $1/\tau_I \tau_I^* < 1$  de los bienes foráneos están disponibles para consumir

Por tanto, dado que  $\tau_D$  son los costes de transporte que sufren los bienes del propio país (internos), se asume que sólo un porcentaje del coste total del producto es susceptible de consumo dado que la otra parte se destina a coste de transporte. Es decir, es como si una parte del bien se "consumiese" en el transporte. Por su parte,  $\tau_I$  son los costes de un bien adquirido en el país foráneo por el hecho de colocarlo en la frontera de ese país, mientras que  $\tau_I^*$  son los costes del mismo bien, por transportarlo desde la frontera hasta el lugar de consumo, de modo que los bienes importados tendrían dos costes de transporte. Para el país foráneo las cosas serían simétricas desde su óptica. De este manera, a través de los costes de transporte se puede introducir el impacto de las infraestructuras.

Finalmente,  $\rho_D = \tau_D^{1-\sigma}$ , siendo  $0 < \rho_D < 1$  dado que  $\sigma > 1$  y  $\tau_D > 1$  (lo mismo para los otros tipos de costes), de forma que al aumentar  $\rho_D$  mejoran las infraestructuras domésticas en el país doméstico; finalmente,  $\alpha$  es el parámetro de la función de utilidad:

$$U = \frac{1}{\alpha^\alpha (1 - \alpha)^{1-\alpha}} D^\alpha y^{1-\alpha} \quad (2.9)$$

De las ecuaciones que determinan la demanda en el país doméstico de los bienes del propio país y de los del foráneo, (2.7 y 2.8), respectivamente, se puede deducir el efecto que tendrán las infraestructuras. Una mejora del capital público doméstico aumentará  $\rho_D$  y con ello la demanda en el país de sus propios productos (que serán más atractivos al ser más baratos), al contrario de lo que sucederá con los foráneos (serán menos demandados dado que en esa ecuación  $\rho_D$  está dividiendo). A su vez cuanto peor sea la dotación de infraestructuras internacionales, ese efecto se verá reforzado ( $\rho_I \rho_I^*$  está dividiendo en la primera ecuación, mientras que en la segunda está en el numerador). Es decir, se trata de que disminuya el coste de los productos propios en relación a los foráneos. Así, en este tipo de modelo, las infraestructuras intervienen en la relación entre empresas y consumidores, de forma que su impacto en la actividad económica se produce a través de la demanda, asumiendo entonces que el capital público no afecta a la función de producción directamente. Como es evidente, las infraestructuras domésticas, al disminuir los costes de transporte en el interior de la región, facilitan el comercio doméstico mientras que las internacionales favorecen las importaciones/exportaciones.

Un resultado adicional del modelo de Martin y Rogers es que, en el equilibrio, el país con un mayor número de empresas será aquél que tenga una mayor demanda y una mejor dotación de infraestructuras domésticas. Este resultado se obtiene de las expresiones que obtienen los autores, correspondientes al número de empresas en cada país:

$$n = \frac{\sigma - \alpha}{\sigma} \frac{K + K^*}{L + L^*} \left[ \frac{L I \rho_D^*}{\rho_D^* - \rho_I \rho_I^*} - \frac{L^* I^* \rho_I \rho_I^*}{\rho_D - \rho_I \rho_I^*} \right] \quad (2.10)$$

$$n^* = \frac{\sigma - \alpha}{\sigma} \frac{K + K^*}{L + L^*} \left[ \frac{L^* I^* \rho_D}{\rho_D - \rho_I \rho_I^*} - \frac{L I \rho_I \rho_I^*}{\rho_D^* - \rho_I \rho_I^*} \right] \quad (2.11)$$

donde

$$I = I + \frac{\alpha}{\sigma - \alpha} \left( \frac{L + L^*}{K + K^*} \right) \frac{K}{L} \quad I^* = I + \frac{\alpha}{\sigma - \alpha} \left( \frac{L + L^*}{K + K^*} \right) \frac{K^*}{L^*} \quad (2.12)$$

De estas ecuaciones se desprende, como se ha comentado anteriormente, que el país que acapare el mayor número de empresas será aquél que tenga un mayores ingresos y mayor número de trabajadores (potenciales consumidores), en conclusión, un mercado grande y rico que, además, esté bien dotado de infraestructuras domésticas. Es decir, tener unas buenas infraestructuras no nos asegura un desarrollo de la región si no van acompañadas de una demanda fuerte. Por tanto, los movimientos de las empresas vendrán determinados por esos factores en un entorno, por ejemplo, como el de los últimos años en la UE, marcado por una mayor integración y el consiguiente aumento de los intercambios comerciales. Es lo que predicen los modelos de *tipo Krugman*: en una primera etapa, el desmantelamiento de las barreras puede llevar a incrementar las diferencias entre regiones. De todas formas, en esos modelos parece ser que la peor situación es la de desmantelamiento intermedio, ya que cuando los costes de transporte tienden a cero (desaparición de barreras), tan sólo pequeñas diferencias en los precios relativos de los factores inducirían a la relocalización de la actividad económica (si las barreras fueran absolutas, la industria tendría que estar repartida en las distintas regiones para poder abastecer a todos los mercados).

La idea de que el tamaño de las empresas está limitado por el alcance del mercado ya fue propugnada por Stigler (1951),<sup>21</sup> y más tarde defendida por Krugman (1991) a través de la utilización del concepto de *market size effect*. Por tanto, este resultado enlaza con la idea de que las infraestructuras son una condición necesaria pero no suficiente para impulsar el desarrollo económico, ya que necesitan interaccionar con otros determinantes de la

---

<sup>21</sup> Según dicho autor, la cuestión básica para que una estructura productiva pueda generar y aprovechar las economías externas es que exista un mercado local lo suficientemente grande y articulado para que las empresas puedan realizar funciones de producción especializadas y complementarias (relaciones inter e intrasectoriales).

localización industrial (como los estudiados en Krugman y Venables 1990, Krugman 1991 y Bertolá 1992).

Así, a partir de las ecuaciones anteriores y bajo el supuesto de que  $\rho_D < \rho_D^*$ , es decir, bajo la hipótesis de que existe una menor dotación de infraestructuras en el país propio, la diferencia en el número de empresas entre países vendrá dada por:

$$n^* - n = 2(K - n) = 2K \frac{\rho_I \rho_I^* (\rho_D^* - \rho_D)}{(\rho_D - \rho_I \rho_I^*) (\rho_D^* - \rho_I \rho_I^*)} \quad (2.13)$$

Una disminución de  $\rho_D$  supone un aumento de los costes de transporte doméstico (por tanto, de los bienes intermedios en la producción),  $\tau_D$ , que provocan un aumento del precio de los bienes producidos en el país doméstico para los consumidores del mismo, con la consiguiente disminución de la demanda de bienes domésticos y aumento de la de los foráneos. Por tanto, al eliminar las barreras, las empresas tenderán a localizarse en las regiones con mejores infraestructuras domésticas. Asimismo, se observa que las diferencias en las dotaciones de capital público internacional, por sí solas, no inducen a la relocalización industrial sino que, por el contrario, aumentan la sensibilidad de la misma.

De este modo, teniendo en cuenta la clasificación que hacen los autores sobre las infraestructuras, los resultados que obtienen sostienen dos ideas principales:

- Con libre comercio, como es el caso de los procesos de integración comercial, las empresas del sector con rendimientos a escala se localizan en los países (regiones) con mejores dotaciones de capital público doméstico para así aprovechar las economías de escala generadas y producir a un coste más bajo. Una mejora de las infraestructuras domésticas aumentará la demanda, en el país, de sus propios productos ya que serán más atractivos al ser más baratos, mientras que los productos del país foráneo serán menos demandados. A su vez, este efecto se verá reforzado cuanto mayor sea la dotación de capital público internacional. Por tanto,

las diferencias en la dotación de capital público doméstico afectan a la dirección de la relocalización de la actividad económica

- Por el contrario, las diferencias en el capital público internacional afectan a la sensibilidad del proceso de relocalización industrial, es decir, una elevada dotación de infraestructuras internacionales y elevados rendimientos a escala fortalecen el proceso de relocalización que surge como consecuencia de las diferentes dotaciones de capital público doméstico, pero por sí solas no lo inducen.

Una cuestión que los autores no comentan es qué pasaría si se parte de una situación en la que en un país no hay industria, estando ésta concentrada en el país foráneo. En esta situación, una mejora de la dotación de infraestructuras domésticas presionará sobre la demanda de productos domésticos, haciendo atractiva la localización de las empresas en ese país. En ese momento es probable que alguna empresa ya haya saltado a ese país y que entonces el impacto ahora sea proporcional al número de empresas ya instaladas. Por el contrario, si mejoran las infraestructuras internacionales, la demanda de productos del país disminuirá tanto más cuanto mayor sea el número de empresas localizadas en el país foráneo.

Asimismo, en el modelo de Martin y Rogers se comprueba que si bien el tamaño de las empresas está inversamente relacionado con el ratio capital privado/empleo, no depende del capital público. Este resultado es común con el del modelo de Holtz-Eakin y Lovely, en el que obtenían que la cantidad de output producida por cada empresa no depende del capital público, sino que éste sólo afecta al número de empresas que se localizan en una zona.

### **2.3.3 Influencia del capital público en los inputs de producción privados**

Se puede considerar que las infraestructuras públicas influirán no sólo directamente en el crecimiento económico sino que puede existir un efecto indirecto a través de cambios en los factores de producción privada. Dos son los posibles efectos que el capital público puede tener sobre los inputs privados: directamente en la cantidad que se necesita de los

mismos y en la efectividad con que son utilizados en el proceso productivo. De hecho, la dificultad que se ha encontrado en asignar un papel cuantitativo al capital público en los estudios que tratan sus efectos en el crecimiento económico, ha sido argumentado por algunos (Mullen *et al.*, 1996) como un reflejo del énfasis que en dichos estudios se hace sobre la aproximación de input directo de las infraestructuras. En opinión de dichos autores, no es que las infraestructuras públicas no tengan un papel directo importante en el crecimiento de la productividad o del output, sino que también se puede pensar en un papel del input público como factor indirecto que refuerza la productividad de los inputs tradicionales.

Ya Negishi (1973) sugería que las regiones con mayores stocks de capital público suelen caracterizarse por sectores manufactureros con ratios capital privado-empleo mayores. De su argumentación se desprende la idea de que la función de producción del sector industrial presenta rendimientos constantes a escala en el capital privado, empleo y capital público, por lo que las firmas conseguirán rentas según el valor del producto marginal del capital público. Además, esta renta probablemente se asignará a aquellos que asuman mayores riesgos, en otras palabras, a los dueños del capital privado. Por tanto, el capital privado será atraído a aquellas regiones que ofrecen unas mayores ganancias para el mismo: aquellas que relativamente poseen unas mejores dotaciones de capital público. Pero además de estos *efectos escala*, pueden aparecer *efectos composición*, por los que las firmas localizadas en regiones en las que la dotación de capital público se está expandiendo pueden responder a través de cambios en su composición factorial, haciéndose más capital intensivas.

De hecho, existen más trabajos que analizan, tanto teórica como empíricamente, la influencia de las infraestructuras sobre el capital privado que no sobre el empleo. Buitter (1977) discute a nivel teórico la idea de que el capital en infraestructuras tiene un impacto positivo en la inversión privada. Dado que las industrias combinan su stock de capital privado con el stock de capital público, los efectos que supone dicha combinación permiten incrementar el valor de las empresas. Si una parte de la idea de que las firmas juzgan la rentabilidad de los proyectos de inversión en base a la tasa de rendimiento interna, a medida que el valor esperado de las firmas aumenta, la inversión en capital también aumentará. Si los costes de instalación de una empresa son más pequeños cuando se puede

disfrutar de una buena provisión de infraestructuras y/o si los costes de los materiales son menores (debido, por ejemplo a una mejora en los sistemas de transporte) o si los beneficios son mayores (debido, por ejemplo a una mayor renta futura generada de una mayor cantidad de infraestructuras), entonces la tasa de rendimiento será mayor. Es decir, a mayor dotación de infraestructuras, mayor será la inversión privada.

No obstante, cabe pensar que las infraestructuras pueden tener un efecto *crowding-out* con la consiguiente influencia negativa sobre las inversiones privadas (teoría neoclásica o monetarista). *Ceteris paribus*, cabe pensar que el sector público y el privado están compitiendo por unos mismos recursos para poder llevar a cabo unos proyectos de inversión (públicos/privados), de forma que el gasto en inversión pública puede suponer una expulsión de gastos en inversión privada. Asimismo, la financiación de la inversión en infraestructuras puede suponer unos mayores tipos de interés, de forma que se genera un desincentivo para los inversores privados. Por otra parte, Eberts (1993) entiende dicha relación de sustituibilidad de forma diferente. Según dicho autor, las empresas consideran que ciertos servicios que provee el capital privado pueden ser sustituidos por los que proporcionarían ciertos componentes del capital público, de forma que se demanda menos capital privado. Así, por ejemplo, las instalaciones de agua que son provistas por los gobiernos locales hacen que las empresas privadas no tengan necesidad de construir sus propias instalaciones de forma que se observará menos inversión privada en el área.

Al igual que existen teorías con diferentes conclusiones sobre el efecto del capital público sobre el privado, la evidencia empírica se encuentra también dividida.

Por una parte, existen varios estudios que a través de la utilización de formas funcionales flexibles, tanto en la teoría de la producción como en la teoría de la dualidad (funciones de costes y de beneficios), han analizado la influencia del capital público en el resto de factores de producción. El uso de formas funcionales flexibles nos permite dividir el efecto total del capital público en el sector privado en sus efectos sobre los diferentes factores productivos. No obstante, los resultados no apuntan hacia una conclusión clara. Así, para la economía americana, mientras Nadiri y Mamuneas (1991), Dalenberg y Eberts (1992), Morrison y Schwartz (1992), Lynde y Richmond (1992), Andrews y Swanson (1995)

indican que el capital público es complementario del capital privado y sustitutivo del empleo, Eberts (1986), Dalenberg (1987), Da Silva *et al.* (1987), Munnell (1990b) y Pinnoi (1994) obtienen que el capital público y el privado son sustitutivos mientras que el capital público y el empleo son complementarios. Para la economía alemana, Conrad and Seitz (1992, 1994) y Seitz (1993, 1994) encuentran que el capital público y el privado se comportan de forma complementaria mientras que la relación entre el empleo y el capital público no resulta clara ya que Conrad y Seitz (1994) y Seitz (1993, 1994) obtienen una relación de sustitución, mientras que Conrad y Seitz (1992) ofrecen una relación de complementariedad. Shah (1992) ofrece resultados para la economía mexicana, con una relación positiva tanto entre capital público y privado como entre capital público y empleo.

En la Figura 2.2 se presentan las implicaciones económicas de una relación de complementariedad entre el capital público y el capital privado y de sustituibilidad con el empleo, según Seitz (1994). Con un capital público igual a  $Kg_0$  las empresas minimizarían sus costes para una cierta producción  $Y_0$ , siendo  $L_0$  y  $Kp_0$  respectivamente las cantidades de empleo y de capital privado que les permitirían minimizar sus costes productivos. Un aumento del capital público hasta  $Kg_1$  ( $Kg_1 > Kg_0$ ) moverá la isocuanta de  $L-Kp$ ,  $Y_0(Kg_0)$  hacia el origen suponiendo que el capital público tiene un producto marginal positivo. La nueva isocuanta,  $Y_0(Kg_1)$  muestra que el output  $Y_0$  puede producirse a un coste menor,  $C(Kg_1) < C(Kg_0)$ . Y si el capital público es sustitutivo del empleo y complementario del capital privado, la isocuanta  $L-Kp$  también cambiará su forma como resultado del incremento del capital público. Si bien el cambio de forma no se aprecia muy bien en la figura, sí se puede analizar observando que si  $Y = f(L, Kp, Kg)$  es la función de producción, la pendiente de la isocuanta  $L-Kp$  viene dada por:

$$\frac{dL}{dKp} = -\frac{f_{Kp}}{f_L} < 0 \quad (2.14)$$

donde los subíndices indican derivadas parciales. Los cambios en  $Kg$  afectaran la curva de la isocuanta de la siguiente manera:

$$\frac{\partial (dL / dKp)}{\partial Kg} = - \frac{f_{KpKg} f_L - f_{Kp} f_{LKg}}{f_L^2} < 0 \tag{2.15}$$

Por tanto, un incremento en  $Kg$  supondrá un aumento de la pendiente de la isocuanta, dado que  $f_{KpKg} > 0$  ( $Kp$  y  $Kg$  son complementarios) y  $f_{LKg} < 0$  ( $L$  y  $Kg$  son sustitutos).

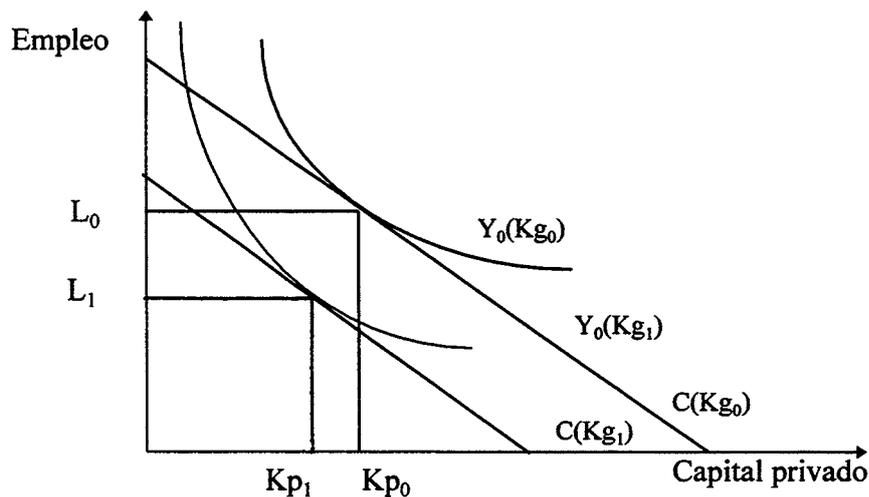


Figura 2.2 Respuesta de la demanda del empleo y el capital privado respecto a aumentos del capital público

Por tanto, el efecto de ahorro de empleo que supone la dotación de capital público es de una naturaleza indirecta antes que directa: el capital en infraestructuras lo que haría sería aumentar la productividad del capital privado que hace que la producción privada sea, *ceteris paribus*, más intensiva en capital privado. De esta manera, el empleo es sustituido por el capital privado y la nueva elección de inputs privados es  $L_1$  y  $Kp_1$ .

Se observa, por tanto, una escasa unanimidad en lo que se refiere a los resultados sobre el tipo de relación entre las infraestructuras y los inputs de producción privada a través de la estimación de formas funcionales flexibles. No obstante, el problema que suponen estas aproximaciones flexibles, es que sí permiten observar el signo de la relación entre ambas variables (de complementariedad o de sustituibilidad) pero no nos permiten determinar la dirección de la causalidad. A fin de mostrar evidencia respecto a este punto, dos tipos de

análisis se han llevado a cabo: por una parte, la utilización de modelos que tratan de analizar los determinantes de los factores privados, y por otra, los análisis de causalidad entre el input público y los privados. Tanto en un caso como en otro la mayor parte de la literatura se centra en el efecto del capital público sobre el capital privado.

Dentro del primer grupo de trabajos, analizando en concreto el papel de las infraestructuras públicas sobre el capital privado, Aschauer (1989c) y Erenburg (1993) ofrecen resultados empíricos en favor de una influencia positiva de la provisión de infraestructuras públicas sobre la inversión privada. En el trabajo de Aschauer (1989c) sobre el efecto nacional del capital público, se obtiene que el efecto *crowding-out* de la inversión pública es sobrepasado por el efecto *crowding-in* asociado con el papel del capital público como input productivo y su complementariedad con el capital privado. Entre sus resultados, obtienen que la inversión pública en carreteras y educación complementa el gasto del sector privado. Por otra parte, los resultados de Erenburg (1993) indican que la inversión en equipamiento del sector privado está inversamente relacionada con el gasto en inversión pública y directamente relacionado con el stock existente de capital público. En dicho estudio, que estima un macromodelo con expectativas racionales, se obtiene que la inversión en equipamiento privado es mucho más sensible a la dotación de infraestructuras públicas que a la inversión en estructuras o medidas de inversión totales. Concretamente, llega a la conclusión de que si la tasa de crecimiento del stock de capital público hubiera continuado de 1966 hasta 1987 a la tasa de crecimiento de 1947-1965, en lugar de disminuir, como fue el caso, la tasa de crecimiento de la inversión en equipamiento del sector privado hubiera estado entre 4 y 6 puntos porcentuales por encima de la tasa de crecimiento en el momento de realización del estudio. En Duffy-Deno (1991), a través de una estimación de los determinantes de la composición del sector industrial medido por el ratio  $Kp/L$  para 36 SMSAs, también se confirma la hipótesis de que un aumento del stock de capital público puede incrementar el capital privado de las empresas. Los resultados empíricos de Argimón *et al.* (1994b) están, asimismo, a favor de la existencia de un efecto *crowding-in* de la inversión privada por la inversión pública en 14 países industrializados en el periodo 1979-1987, a través del impacto positivo de las infraestructuras sobre la productividad de la inversión privada.

Sin embargo, los resultados de Bairam y Ward (1993) apoyan la hipótesis de *crowding-out*. En su trabajo estiman una función de la inversión privada en la que no se impone *a priori* la forma funcional sino que se contrasta a través de la transformación Box-Cox para 25 países de la OCDE con datos anuales para el periodo 1950-88. Los resultados avalan de forma rotunda en 19 países que aumentos en los gastos del Estado expulsan de forma significativa la inversión privada, mientras que en otros 5 también se obtiene un signo negativo aunque no significativo.

En otro orden, se han utilizado análisis de causalidad en los que se determina no sólo el signo de la relación entre inversión privada y pública sino también su secuencia en el tiempo: ¿Precede la inversión en capital público a la inversión privada o al revés? En el primer caso, se podría concluir que las infraestructuras públicas son usadas activamente como instrumentos de desarrollo local directo. Por el contrario, en el segundo caso se concluiría que la construcción de infraestructuras públicas meramente responde a las iniciativas de inversión privada.

Así, Eberts y Fogarty (1987) analizan la relación causal entre la inversión pública y la privada para un grupo de ciudades en los EEUU desde 1904 hasta 1978. En dicho análisis encuentran una relación causal significativa entre las inversiones públicas y la inversión privada para 33 de las 40 áreas metropolitanas que examinan, si bien la relación de causación va en las dos direcciones. La inversión pública influye mayormente en la inversión privada en ciudades que no están en el sur del país y en ciudades que experimentaron el mayor crecimiento de población antes de 1950. Por otra parte, la inversión privada afecta más a la inversión pública en las ciudades del sur y las que crecieron sustancialmente después de 1950. Así, los autores concluyen que si bien la secuencia de la inversión pública influyendo en la privada no es cierta en toda la muestra, sigue manteniéndose la idea de que la inversión pública es necesaria para el desarrollo económico local.

En el trabajo de Erenburg y Wohar (1995) se utilizan los tests de causalidad de Granger en un modelo formado por una ecuación de la inversión privada y otra de la inversión pública con valores retardados de las variables explicativas. Sus resultados sugieren la ausencia de

causalidad unidireccional en el sentido de Granger tanto de la inversión en capital privado a la de capital público como al revés, pero concluyen en favor de la existencia de efectos *feedback* entre ambas variables. En la estimación de la ecuación de la inversión privada, el efecto de la inversión pública resulta tener un efecto conjunto significativo, pero la dirección del impacto es ambigua. Así, se obtiene un efecto estadístico negativo de los tres primeros retardos, pero el cuarto es significativo y positivo. Estos resultados estarían indicando, en opinión de los autores, que la inversión del sector público puede originar un efecto *crowding-out* en el sector financiero. Sin embargo, cabría argumentar que dada la existencia de un tiempo intermedio entre el momento en el que el sector público pone el dinero del proyecto y el momento en que dicho proyecto se finaliza, las empresas que se benefician del capital público pueden posponer los planes de inversión en equipamiento hasta que el capital público se encuentre realmente disponible. Asimismo, en dicha ecuación de inversión en capital privado, la variable explicativa que recoge el stock de capital público no tiene ningún tipo de efecto. Este resultado no es interpretado, sin embargo, como una falta de influencia del stock de infraestructuras públicas sobre la inversión privada, sino que se considera que dicha variable puede no estar ejerciendo una influencia independiente en este contexto, dado que el efecto de la inversión del sector público ya se habría considerado de forma retardada, de manera que ya se estaría teniendo en cuenta el posible efecto del stock de capital público.

Por tanto, a tenor de los resultados de estos trabajos, no podemos concluir en favor de una relación clara entre capital público y privado, dado que los resultados dependen, en gran manera, del nivel de actividad económica y el grado de desarrollo de la zona que se esté considerando.

#### **2.3.4 Capital en infraestructuras y convergencia regional**

Tal y como se ha visto hasta este punto, el impacto del capital público en el crecimiento, en las decisiones de localización de la actividad económica y en la productividad de los inputs privados no ha alcanzado aún el consenso. De la misma manera, el efecto que las infraestructuras poseen en la reducción de las disparidades regionales no es una excepción.

Si bien en los medios políticos parece considerarse que la dotación de unas buenas infraestructuras es necesario para crear las condiciones que permitan alcanzar la convergencia en los procesos de integración, desde el mundo académico no se tiene tan claro.

Desde el punto de vista político nos vamos a centrar en el proceso de integración europea y las diversas actitudes que se han adoptado para asegurar la convergencia regional en la UE. En este sentido, como consecuencia de las disparidades en los niveles de renta y bienestar que se observaron en la Comunidad Económica Europea (CEE) tras la entrada de España y Portugal en 1986, y temiéndose que el paso siguiente hacia un Mercado Único Europeo (MUE) agravará tales disparidades, el Consejo Europeo articuló una serie de medidas para conseguir una mayor cohesión entre los países y regiones europeas. En un primer momento, año 1987, se reformaron los Fondos Estructurales reformulando y expandiendo un grupo variado de políticas con objetivos de reforma estructural y de crecimiento económico, bajo el nombre de Marco de Apoyo Comunitario (MAC). Desde su creación, estos fondos están dirigidos fundamentalmente a la promoción de un crecimiento autosostenido en los países y regiones menos desarrolladas de la UE, dedicando un considerable volumen de recursos a tres grandes áreas de actuación:<sup>22</sup>

- 1) inversión pública en infraestructuras (37.3%)
- 2) inversión en recursos humanos (17.2%)
- 3) ayudas a la inversión productiva (45%).

Se observa, por tanto, que las inversiones en capital público son tenidas en cuenta por los gestores de política económica a fin de promover la cohesión (es decir, la convergencia real) de las regiones europeas. Algo más de la mitad de dichos fondos están destinados a infraestructuras de transporte, un cuarto a temas medioambientales asociados al ciclo del agua y a la conservación de los recursos del medio ambiente y el resto se distribuye por igual entre infraestructuras energéticas y de comunicaciones, equipamientos sanitarios y

---

<sup>22</sup> En el caso español, el programa de infraestructuras del MAC constituyó un 0.74 % del PIB en 1994, con la intención de incrementarlo hasta el 0.97 en 1999.

educativos (Herce y Sosvilla, 1995). Entre los objetivos fijados por los Fondos se encuentran la articulación del territorio, la mayor accesibilidad de áreas rurales aisladas, la mejora del transporte urbano y la coherencia de los distintos modos de transporte.

En concreto, se observa que son las infraestructuras de transporte las que centran la atención de la CEE. Así, en el Acta Única Europea de 1985 se incorporaba la reflexión sobre redes transeuropeas que hicieran posible el flujo de personas, servicios y mercancías que suponía la creación de un área sin fronteras internas en 1992, así como la utilización de las mismas como instrumento de cohesión económica y social. De este modo, se estaba fraguando lo que en el Tratado de la Unión de Maastricht se conoce como política de redes transeuropeas, con el doble objetivo de constituir un único mercado interior y alcanzar la cohesión económica y social. Respecto a este último objetivo, el que centra nuestro análisis, y según reza en el propio Tratado "esta política tiene una indudable relación con el territorio y con la necesidad de conectar las regiones aisladas, enclavadas o periféricas" (González-Finat, 1995). De hecho, en el quinto informe periódico sobre la situación y la evolución socioeconómica de las regiones de la CEE (Comisión, 1994) se sostiene que la menor renta per cápita de las regiones de Grecia, España, Irlanda y Portugal con relación al resto de la UE está estrechamente relacionada con la insuficiencia de carreteras, autopistas, líneas ferroviarias, líneas telefónicas, redes de energía, suministro de agua y tratamiento de aguas residuales en estas regiones.

También la política económica española mantiene la idea de que resulta necesario invertir en infraestructuras para conseguir la convergencia con los países más desarrollados de la UE. Esto queda patente en el compromiso que ha adquirido el gobierno español en el Programa de Convergencia de mantener las inversiones públicas en el 5 % del PIB durante el quinquenio 1996-2000 (De Rus *et al.*, 1995). De este modo, queda patente la importancia del capital público en el caso español si se tiene en cuenta la perifericidad de la economía española que dificulta la convergencia con los países más desarrollados situados en el centro de Europa, que concentran los núcleos de actividad más importantes.

Si bien estas actuaciones reflejan la idea de los agentes políticos europeos de defender la política de inversión en infraestructuras como medio para avanzar en el proceso de

cohesión económica y social o convergencia real (en términos de renta per cápita), la visión académica sobre el tema es algo distinta.

Si nos basamos en que los resultados obtenidos por Martin y Rogers (1995) (comentados en el apartado 2.3.2) son válidos, la provisión de capital público en infraestructuras no favorecería la convergencia. Por el contrario, parece que, si bien una buena dotación de infraestructuras domésticas es necesaria para atraer nuevas empresas, cuando se crean unas buenas comunicaciones internacionales, las empresas de las regiones pobres se trasladan a los centros de actividad más próximos donde pueden aprovechar las economías de aglomeración y desde donde ahora resulta más barato suministrar al mercado de la región pobre. De esta manera, en una primera etapa de los procesos de integración comercial, el desmantelamiento de las barreras puede llevar a incrementar las diferencias entre regiones. La política regional, de este modo, debería facilitar las infraestructuras domésticas más que favorecer el comercio internacional si se desea favorecer la convergencia.

No obstante, esta argumentación podría no ser aplicable directamente al caso español dado que la importancia del peso de las pequeñas y medianas empresas (PYMEs) en el conjunto del tejido industrial unido a la posibilidad de una menor tendencia al cambio de localización del negocio por parte del pequeño empresario nos llevaría al resultado de una menor sensibilidad a todo tipo de incentivos locacionales al contrario de lo que parece apreciarse para la economía americana, en la cual la tendencia a la migración del negocio parece ser bastante mayor. Por esta razón, a pesar de que el establecimiento de una buena red de comunicaciones internacionales parece que debiera favorecer el movimiento de las empresas, podría estar pesando más el efecto de la mejora en la accesibilidad a mercados exteriores (resultando las exportaciones más competitivas merced a la obtención de ventajas competitivas en su producción por menores costes de transporte y menor tiempo de abastecimiento principalmente) que el efecto contrario, tal y como propone el modelo de Martin y Rogers. En este sentido se puede recurrir al ejemplo de parte de las PYMEs de regiones como Cataluña, País Vasco y Valencia, orientadas a la exportación (sin haberse tenido que localizar en regiones extranjeras) y que, por tanto, podrán obtener un fuerte rendimiento de las infraestructuras, tanto locales como internacionales, al ser más competitivas en los mercados externos. Sin embargo, no es menos cierto que también

parece estar detectándose un proceso mediante el cual algunas empresas españolas están empezando a establecer sucursales en países como Bélgica, Holanda o Alemania para abastecer más fácilmente al mercado del norte de Europa, por lo que si bien se puede concluir que los modelos analizados no se adaptan de manera taxativa al caso español, no deben descartarse en su totalidad. De igual forma, las empresas orientadas al mercado interno también podrán competir en mejores condiciones frente a los productos foráneos si la dotación de infraestructuras domésticas hace que sus productos sean más competitivos en el mercado local. Por tanto, parece deducirse que la primera decisión a adoptar es dotar de capital público doméstico a aquellas zonas que se encuentran desprovistas del mismo y, a medida que dicha dotación va siendo adecuada a las necesidades de la zona, dotar de capital público internacional para permitir que aumenten sus relaciones con otras regiones y países.

Siguiendo con el análisis de Martin y Rogers (1995) consideramos que existe un cierto dilema en el mismo en lo referente a la convergencia. Según las conclusiones de los autores, la política de provisión de capital público doméstico en las regiones pobres será más efectiva cuanto mejor sea la dotación de capital público internacional. Sin embargo, la mejora de este último tipo de infraestructuras en una región con baja dotación de capital público doméstico puede hacer que las empresas abandonen dicha región y el comercio se vea favorecido, tanto más cuanto mayores sean los rendimientos a escala. Por tanto, si el capital público internacional acentúa tanto positiva como negativamente los efectos del capital público doméstico, entonces no resulta claro qué tipo de capital hay que priorizar. Si el capital internacional fuese muy bajo, entonces los efectos de la dotación de capital doméstico serían más pequeños que si el internacional fuese elevado. Pero, en contrapartida, si se mejora el capital internacional entonces también se acentúan las fuerzas de localización que conducen a la aglomeración.

Por tanto, las tendencias actuales en transportes y comunicaciones podrían llegar a reforzar la divergencia. Para sustentar esta afirmación, De Rus *et al.* (1995) ofrecen varios ejemplos: las redes de alta velocidad y las autopistas aumentan la accesibilidad y la comunicación entre regiones especialmente centrales; Vickerman (1987) consideró que el túnel del Canal de la Mancha no desarrollaría las regiones adyacentes, beneficiándose, en

todo caso, las situadas a 100 o 150 kms.; las políticas de desregulación del transporte aéreo han reforzado la primacía de los principales aeropuertos de Europa, etc.

En lo que se refiere al caso español, Mas *et al.* (1995c) analizan el proceso de convergencia entre las regiones españolas y estiman la influencia de la dotación de capital público en dicho proceso de convergencia. Tras un análisis de los datos (1955-1991) observan que existe una correlación positiva entre la dotación de infraestructuras relativizada al valor añadido bruto y la tasa de crecimiento del valor añadido bruto per cápita. No obstante, concluyen que la provisión de capital público sólo supuso un impacto positivo sobre el crecimiento hasta mediados de los años sesenta, debido tanto a la mayor homogeneización de la dotación infraestructural a partir de entonces como a la posible existencia de rendimientos decrecientes en los efectos del capital público. Por tanto, parece deducirse que teniendo en cuenta el acortamiento de distancias entre las regiones españolas en cuanto a dotación infraestructural se refiere, la efectividad de una política pública de inversión en infraestructuras en las regiones menos desarrolladas será limitada y posiblemente suponga un coste de oportunidad elevado (*trade-off* entre equidad y eficiencia).

De hecho, Mas *et al.* (1993) opinan que las dotaciones relativas de infraestructuras en España con respecto al PIB son similares a los casos, por ejemplo, de Alemania y Reino Unido, mientras que estamos muy por debajo de las dotaciones de capital privado. Por tanto, si bien las políticas públicas no deben reducir el gasto en inversión pública, sobre todo de transportes y comunicaciones, la localización periférica de las regiones españolas con respecto al centro europeo y la insuficiencia de otros factores (capital privado, capital humano, cierto nivel tecnológico, entre otros) serían las que realmente afectan a las posibilidades de crecimiento.

Sin embargo, en de la Fuente y Vives (1995) se obtiene como conclusión principal que las diferencias en capital humano y público explican un tercio de la desigualdad regional en España. En opinión de dichos autores, las políticas de oferta ofrecen un papel importante en la búsqueda de la convergencia regional, argumentando que la poca efectividad de la política regional activa en la reducción de las disparidades en España no se debe tanto a la

ineficiencia de la misma como instrumento específico de intervención sino a la insuficiencia de los fondos destinados a la misma.

El problema no es de fácil solución. Así, si las regiones pobres tienen prioridad en la política de inversión pública llevada por el Estado, la eficiencia de todo el sistema económico puede ponerse en cuestión imponiendo un coste elevado a las regiones con condiciones más favorables para el crecimiento económico. Lo que debiera propugnarse es no incrementar el nivel de infraestructuras con independencia del nivel de utilización, ya que sino se produciría sobrecapacidad y despilfarro de recursos, sino invertir a diferentes niveles según el estado de desarrollo. De este modo, siguiendo las ideas extraídas de los modelos que hasta este punto se han analizado, en las regiones más desarrolladas las infraestructuras necesarias serían aquéllas que permitieran descongestionar aquellos servicios que se encuentran ya saturados por la elevada actividad económica que tiene lugar en las mismas, prestando especial atención a las infraestructuras internacionales. Por el contrario, las regiones menos privilegiadas deberían conseguir el capital público mínimo para evitar que su ausencia suponga un impedimento a la implantación de nueva actividad económica, consolidando básicamente la dotación de infraestructura doméstica. Es decir, la convergencia parece exigir que las regiones periféricas modernicen y amplíen su red con el objetivo de que su perifericidad no comprometa su futuro por falta de infraestructuras que aseguren una cierta accesibilidad y unos costes medios de acceso razonables (Delgado, 1995; De Rus *et al.* 1995).

### 2.3.5 Condicionantes en el impacto de las infraestructuras

A la vista de los resultados anteriores cabe pensar que cuantificar el impacto de las infraestructuras sobre el crecimiento económico o analizar la importancia que tienen las mismas en las decisiones de localización de la actividad económica no es una empresa fácil. No es que dichos objetivos no sean factibles de estudiar, sino que la dificultad radica en que los resultados obtenidos variarán mucho de un estudio a otro, siendo difícil obtener resultados concretos que puedan generalizarse como ayudas en la toma de decisiones políticas. Las razones de esta variedad de conclusiones se encuentra en el hecho de que el

impacto de las infraestructuras es un tema complejo que depende de un gran número de factores, tantos como la realidad económica puede llevar emparejados en sí misma: el momento de desarrollo económico de la zona en la que se planea instalar la infraestructura, la etapa del ciclo económico en que se encuentra la economía, la existencia o no de economías de aglomeración, el nivel de utilización de las infraestructuras ya existentes, o el tamaño de las empresas de la zona son algunos de los condicionantes del impacto de las infraestructuras. Sin ánimo de ser exhaustivos, se ofrecen a continuación las principales ideas que se han extraído de estudios en los que estos factores han sido tenidos en cuenta.

### **2.3.5.1 El nivel existente de infraestructuras y la congestión en su uso**

La idea sobre el diferente impacto que pueden tener las infraestructuras según el nivel existente de las mismas puede resumirse en tres puntos.

En primer lugar, como consecuencia de la característica tipo red que tienen gran parte de las infraestructuras, principalmente las de transportes, se requiere una capacidad mínima de las mismas para que puedan suponer un impacto diferenciable en el desarrollo económico de un área: parece lógico pensar que construir una autopista en medio de un desierto tendrá un escaso efecto en la zona. Es decir, la productividad de una parte de la red depende del tamaño y el estado de la red entera. Nijkamp (1986) habla de la existencia de un tamaño umbral o una mínima capacidad del capital público, especialmente en lo que se refiere a transportes, de forma que una infraestructura aislada no tendrá mucho efecto en el crecimiento pero sí la red entera. Wheat (1969) y Isserman *et al.* (1989) concluyen que inversiones en infraestructuras previas a la instalación de una autopista, influyen positivamente en el impacto de las mismas. Asimismo, la existencia de otros tipos de infraestructuras, tales como los aeropuertos, pueden contribuir a la efectividad de las inversiones en autopistas.

En segundo lugar, la naturaleza de red de la mayor parte de infraestructuras implica un mayor efecto en la fase inicial de construcción e instalación de las mismas que en fases posteriores de ampliación. Una mejora en una conexión o red ya establecida reducirá

costes, pero tendrá un impacto menor que cuando se construyó por primera vez. Incluso puede existir un determinado tamaño de la red de infraestructuras, más allá del cual el capital público no tiene un efecto importante sobre la producción, mientras que el impacto de la inversión pública podría ser muy alto cuando existen barreras que limitan el desarrollo. Es decir, parece existir un nivel umbral de las infraestructuras, hasta el cual las mismas tienen un efecto importante en el crecimiento económico ya que son tomadas como condición indispensable para establecer la actividad económica, pero pasado dicho umbral, no resultan un factor determinante ni en el crecimiento ni en el establecimiento de la actividad. Así lo argumentan Moore *et al.* (1991) cuando, al analizar los resultados de las encuestas a empresas de los países de la UE, obtienen que los empresarios no suelen tener en cuenta las infraestructuras como factor determinante en su decisión de localización porque dan por sentado que existen unas infraestructuras básicas que van a poder utilizar en su proceso productivo.

Da Silva *et al* (1987) encuentran empíricamente, para el caso de los estados americanos a través de una función de producción translogarítmica, que los impactos diferenciales de las infraestructuras en el crecimiento regional están condicionados por los niveles existentes de infraestructura: a mayor stock de capital público per cápita, menor será el impacto del stock de capital público en la producción. Es decir, consiguientes adiciones de stock en infraestructuras suponen rendimientos marginales decrecientes.

Y por último, aun habiendo sobrepasado dicho umbral parece sensato pensar que incrementos del stock de un cierto tipo de capital público tendrán un efecto positivo en la zona siempre y cuando el stock existente esté congestionado. Así, añadir capacidad a una red no congestionada no mejorará la productividad privada, mientras que los beneficios que se pueden obtener de un aumento en la cantidad de capital público son muy altos cuando la congestión es alta. Esta idea se encuentra relacionada con el concepto de infrutilización o sobreutilización de las infraestructuras, de modo que si bien en el primer caso no resulta necesario realizar más inversiones en el área considerada, en el segundo, la ampliación de la red ya existente permitirá aumentar la productividad de la zona gracias al mejor aprovechamiento de la misma.

En el trabajo de Boarnet (1997a) se estima una función de producción translog que incorpora no sólo el stock de carreteras y autopistas de los 58 condados de California en el periodo 1977-1988, sino también una medida de congestión de las mismas basada en la intensidad media diaria del tráfico y la adecuación de la capacidad de las infraestructuras mencionadas anteriormente. El resultado que obtiene confirma la teoría, ya que el efecto de incrementar el stock de las carreteras y las autopistas es mayor, *ceteris paribus*, en aquellos condados que tienen una mayor congestión.

Sin embargo, al hablar de la congestión de las infraestructuras, principalmente de transportes, se está haciendo referencia al hecho de que el capital público es productivo dado que ofrece un servicio útil en la producción de un output. Cuando en la mayoría de estudios se utiliza el stock de capital público como aproximación a los servicios ofrecidos por el capital público, no se está teniendo en cuenta si dicho stock está siendo utilizado de manera eficiente. Así, existe una amplia literatura sobre el precio de la congestión que ha considerado desde antiguo que las autopistas que no tienen precio resultan en un uso ineficiente de las mismas (ver por ejemplo, Mohring y Harwitz, 1962; Vickrey, 1963; Keeler y Small, 1977; Small, 1983). De este modo, una manera de incrementar el flujo de servicios provistos por una autopista congestionada consistiría, en opinión de autores como Winston (1990), Gramlich (1994), y Boarnet (1997a), en poner precio al stock ya existente antes que construir más.

### **2.3.5.2 El nivel de desarrollo económico, la etapa de crecimiento económico y la existencia de economías de aglomeración**

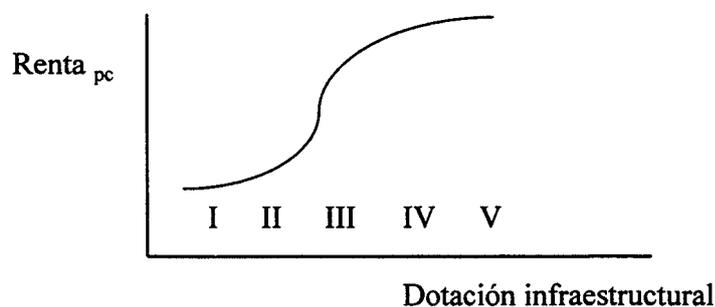
Algunos trabajos analizan la influencia del capital público en el crecimiento regional partiendo de hipótesis tales como que la misma es mayor en el caso de que la región en la que se instala la infraestructura esté pasando un momento de recesión, que la zona esté en vías de desarrollo o que existan economías de aglomeración de forma que las infraestructuras se vean como un input necesario para poder llevar a cabo el proceso productivo. En el caso contrario, se esperaría que el efecto de las infraestructuras fuera menor.

A nivel teórico, dichas hipótesis han sido tratadas desde antiguo. Hirschman (1958) y Hansen (1971) argumentan que la inversión en infraestructura pública sería más efectiva en regiones urbanizadas que presenten economías de aglomeración y un dinamismo previo que en regiones rurales y con menor desarrollo que padecen principalmente de falta de inversión privada. Wheat (1969) comenta también que los efectos de las autopistas, por ejemplo, son mayores en áreas geográficas con una elevada densidad de población. Así, según dichos autores, el capital público no serviría para atraer actividad privada en un inicio, sino que previamente debe existir un cierto nivel de demanda de la misma para que los efectos del capital público puedan realizarse. Isserman *et al.* (1989) sugieren que los condados rurales que contienen ciudades de un cierto tamaño (por encima de los 15000 habitantes) suelen ser estimulados más por la autopistas que los condados que carecen de las mismas. De esta manera, el hecho de que puedan servir a una mayor población resulta en reducciones de costes de transporte mayores que las autopistas localizadas en otras zonas. Siguiendo esta misma idea pero avanzando un poco más, Hansen (1965) considera que se debería concentrar el capital público básico (transportes y sistema hidráulico, principalmente) en las regiones de desarrollo intermedio, y el capital público social (educación y sanidad) en las regiones subdesarrolladas.

Varios estudios confirman las ideas expuestas anteriormente. Chernoff (1978) obtiene que las autopistas suponen mayores efectos socioeconómicos en áreas con mayores densidades de población. Martín (1979) muestra que los programas de obras públicas son menos efectivos en áreas de renta baja, de la misma manera que Looney y Frederiksen (1981) en un estudio para México demuestran que el transporte por carretera en general es efectivo en regiones intermedias pero no en regiones subdesarrolladas. Según obtienen Sasaki *et al.* (1995) a través de un análisis de simulación basado en un modelo econométrico regional, la inversión en carreteras en áreas urbanas es más efectiva que en áreas suburbanas, en el sentido de que provoca un aumento del producto interior que es tres veces mayor en el primer caso. No obstante, no es menos cierto que la construcción de carreteras supone un mayor coste en las zonas urbanas, por lo que la diferencia de la tasa de rendimiento de las inversiones en carreteras entre zonas urbanas y suburbanas no es tan grande como en un principio pudiera parecer. La única excepción que conocemos a este resultado general es el

trabajo de Duffy-Deno (1988) donde se obtiene que las autopistas tienen su mayor efecto en el output manufacturero en ciudades deprimidas económicamente.

Nijkamp (1986) relaciona lo que denomina *problema umbral* (la existencia de una capacidad mínima de capital en infraestructura para que suponga un efecto en el crecimiento) con la noción de Rostow *de los estados de desarrollo regional*. Así, en un estudio de la Comunidad Europea en el que se recogen datos sobre infraestructuras (Biehl *et al.*, 1982) se indica la existencia de una curva logística en forma de S, que suele concebirse como la senda de crecimiento general en un sistema económico dinámico, tal como se ilustra en la Figura 2.3:



**Figura 2.3** Senda de crecimiento general en un sistema económico dinámico

Esta curva “S” puede dividirse en algunos estados de desarrollo económico de la región bajo consideración y relacionarlos con su dotación infraestructural. Cada estado corresponde a un tipo específico de región:

- I: regiones subdesarrolladas con una dotación infraestructural inadecuada, tales como áreas rurales o periféricas
- II: regiones que están alcanzando un estado de despegue económico, de forma que los requerimientos mínimos de infraestructuras para el crecimiento económico ya se han alcanzado
- III: regiones de rápido crecimiento que tienen una suficiente dotación infraestructural

- IV: regiones que se caracterizan por estar en el camino hacia la madurez y que muestran los primeros signos de externalidades negativas relacionadas con un rápido crecimiento
- V: regiones que han crecido rápidamente pero que no pueden crecer más debido a la existencia de cuellos de botella en su dotación infraestructural.

Según Nijkamp, la cuestión estriba en si se puede identificar la existencia de un cierto nivel específico de infraestructuras en cada estado de desarrollo regional. Así, por ejemplo, tal como se tratará posteriormente en este mismo apartado, se considera que las infraestructuras son una condición necesaria para el desarrollo regional, de forma que las infraestructuras de red son particularmente importantes durante las primeras etapas de desarrollo. En las siguientes etapas, según Nijkamp, las infraestructuras urbanas son las que resultan ser más importantes mientras que, finalmente, las infraestructuras sociales pueden jugar un papel relevante.

Con el objetivo de identificar cuáles son las regiones en las que utilizar el capital público como herramienta para estimular el crecimiento económico puede ser de utilidad, Nijkamp propone la utilización de un análisis shift-share que permite separar la contribución nacional al desarrollo regional de los impactos intrarregionales específicos de los cuales, según dicho autor, las infraestructuras forman parte. En el mismo, el componente estructural en el cambio regional estaría representando una medida de la calidad de la estructura económica regional, mientras que el componente regional reflejaría si la combinación de otros factores de potencialidad confieren unas ventajas locacionales relativas que explicaría la tendencia de las industrias a crecer más rápido en unas regiones que en otras. Como en todo análisis shift-share, se obtiene la siguiente tipología de las regiones:

		COMPONENTE REGIONAL	
		+	-
COMPONENTE	+	I	II
ESTRUCTURAL	-	III	IV

De este modo, las regiones del grupo I tienen una estructura industrial satisfactoria y unas condiciones de localización que favorecen la realización de las economías de aglomeración. En el grupo II, las regiones poseen unas condiciones favorables para atraer actividad industrial pero otras industrias pueden tender a estacionarse debido a la existencia de cuellos de botella principalmente por factores de potencialidad (accesibilidad, nivel salarial, etc.). En el grupo III, las regiones presentan una infrarrepresentación relativa del crecimiento industrial pero las condiciones de localización son favorables para poderse beneficiar de los efectos *spillover* de las industrias del grupo II; a estas regiones se las conoce como las regiones intermedias, localizadas en la frontera de las áreas centrales de cada país. Finalmente, las regiones del grupo IV presentan condiciones locacionales desfavorables de forma que ni el crecimiento de las industrias ni los efectos *spillover* pueden ser atraídos a dichas áreas, presentando un panorama desalentador. De esta forma, si se asume que las infraestructuras son uno de los elementos de la dimensión regional, la contribución de las mismas como instrumento de política para el desarrollo regional puede resultar importante en aquellas regiones en las que el componente regional presenta un valor negativo, celdas II y IV. En la celda II, se podría estar dando el caso de que una insuficiente dotación infraestructural esté frenando el crecimiento regional, mientras que en el caso de las regiones de la celda IV, es aconsejable combinar tanto políticas locacionales como sectoriales.

Por tanto, parece concluirse que las infraestructuras, sobre todo las de transporte, parecen tener un mayor efecto sobre la actividad económica en regiones que no están muy industrializadas pero que poseen un cierto potencial de desarrollo, como por ejemplo, regiones que están en un estado intermedio de desarrollo y quizás experimentando un bajo crecimiento. Asimismo parece ser que los efectos de las infraestructuras están correlacionados con los niveles de urbanización y la proximidad metropolitana. De hecho, según Justman (1995) se debe conseguir una interdependencia entre el establecimiento de una infraestructura y conseguir una masa crítica de demanda de sus servicios. Aunque la presencia de dicha infraestructura suele aumentar la actividad local, se requiere que exista un mínimo nivel de demanda de la misma para justificar su establecimiento. Posteriormente, cuanto mayor sea la actividad económica de una zona, lo cual está

directamente relacionado con el nivel de aglomeración de la actividad y urbanización, mayor será la demanda de infraestructuras y, por tanto, se asegurara la eficiencia en su uso.

Sin embargo, si bien es cierto que la eficiencia de las infraestructuras se consigue mediante la localización de las mismas en regiones con cierta actividad económica, se puede pensar en la utilización del capital público como política de convergencia, ubicándolo en zonas deprimidas a fin de disminuir las disparidades regionales. Es la relación de *trade-off* entre eficiencia y equidad en la localización del capital público, tal como se ha visto en el apartado anterior.

### 2.3.5.3 El tamaño de las empresas

El efecto de las infraestructuras estará condicionado a las características del tejido empresarial de la región, y en concreto, al tamaño de las empresas. Empresas de tamaño pequeño pueden ser más vulnerables a todo tipo de políticas públicas, principalmente a la hora de decidir la localización de su actividad. Varias pueden ser las razones de esta mayor vulnerabilidad. En primer lugar, empresarios con pequeños establecimientos pueden prestar más importancia, en el momento de decidir su localización, a la disponibilidad de unas buenas instalaciones en el área y las comunicaciones con sus clientes/proveedores, que no empresas grandes, con sucursales en distintas localizaciones, que basarán sus decisiones de localización en términos de rentabilidad. Otra razón puede explicarse por el hecho de que las empresas pequeñas probablemente intentarán reducir sus costes de búsqueda de la mejor ubicación considerando menores opciones de localización y, por tanto, confiando su decisión a un área geográfica o económica más pequeña, de manera que considerando menos opciones, cabría esperar una mayor sensibilidad a los atributos locales como las infraestructuras. Por el contrario, las empresas grandes estarán más influidas por variaciones en atributos regionales y no sólo locales.

A nivel empírico existen varios trabajos que han tenido en cuenta el diferente efecto de las infraestructuras en la actividad económica según el tamaño de las empresas. Analizando las respuestas de encuestas sobre cuáles son los factores más determinantes en las decisiones

de localización de las empresas, Moore *et al.* (1991) obtienen que efectivamente las empresas pequeñas (aquéllas con menos de 200 trabajadores) son las que parecen resultar más afectadas por todas las políticas públicas. Sin embargo, en lo que se refiere en concreto a las infraestructuras, concluyen que no existen diferencias significativas en su importancia según el tamaño de las empresas.

Por otra parte, varios trabajos que utilizan modelos de regresión condicionan los diferentes resultados de los efectos de las infraestructuras al tamaño de las empresas. Fox y Murray (1990) obtienen que la influencia de las infraestructuras de transporte en la entrada de nuevas empresas es homogénea en todos los tamaños considerados, pero la disponibilidad de infraestructuras hidráulicas sólo resulta una variable significativa en la entrada de empresas pequeñas. No obstante, los autores obtienen que la entrada de empresas está positiva y significativamente relacionada con los costes del agua. La explicación que dan los autores a tal resultado se basa en la idea de que las empresas ven unos costes superiores como un reflejo de la disponibilidad de una mayor calidad del servicio de la infraestructura. Por tanto, el mantenimiento de unos servicios a bajo coste a costa de reducir la calidad probablemente desanimaría a las empresas a instalarse en dicho área. Por tanto, su resultado positivo respecto a los costes sólo es, en parte, favorable a la hipótesis de la mayor vulnerabilidad de las empresas a menor tamaño.

En un trabajo similar al anterior, Eberts (1993) obtiene que el aumento de las infraestructuras públicas afecta positiva y significativamente al número de apertura de empresas pequeñas, mientras que el número de aperturas de empresas con elevado número de trabajadores no resulta significativamente influido por el crecimiento de las infraestructuras. Fox y Murray (1993) llegan a la misma conclusión de que las PYMEs son las que resultan más afectadas por las diferentes políticas públicas, entre las que consideran la provisión de infraestructuras.

Parece ser, por tanto, que las empresas de pequeño tamaño consideran menos opciones de localización pero resultan más afectadas que las grandes a dichas opciones, siendo las infraestructuras uno de sus determinantes más importantes.

### 2.3.6 Capital público en infraestructuras: Condición necesaria pero no suficiente

La consideración del capital público como *condición necesaria* para el crecimiento económico la han utilizado muchos autores con bastante frecuencia: "Como mínimo hay acuerdo en que las infraestructuras son un prerequisite para el desarrollo económico, influyendo positivamente sobre la producción, la inversión privada y el empleo" (De Rus, 1995); "Las infraestructuras públicas juegan, con seguridad, un importante papel en las economías metropolitanas. Sin embargo, su contribución marginal no es ni mayor ni menor que otras formas de inversión" (Crihfield y Panggabean, 1995). Del mismo modo, De Rus *et al.* (1995) consideran que no sería posible tener una economía con una elevada especialización del trabajo y con transmisiones de mercancías, personas e información, si no se contara con redes de transportes y comunicaciones, sistemas hidrológicos y de saneamiento, más o menos sofisticados. Este requisito para el desarrollo económico se explica por el hecho de que su carencia o insuficiencia origina cuellos de botella al normal funcionamiento del sistema económico bien mediante la elevación de los costes de producción bien por el empeoramiento de la calidad de los servicios prestados, resultando negativamente afectados tanto la competitividad de la economía como el nivel de vida de los ciudadanos.

Es en el sector industrial manufacturero y en el de las nuevas tecnologías (telecomunicaciones, *software*, biotecnología) donde se considera que determinadas categorías de infraestructuras tienen su principal aportación como condición necesaria. Así, por ejemplo, Martin y Rogers (1995) destacan que los procesos de integración comercial provocan que las empresas tiendan a localizarse en países o regiones que disponen de una buena dotación de capital público (proceso de relocalización de la actividad), dejando en evidencia la necesidad de las mismas. Asimismo y en concreto para el caso español, varias encuestas realizadas por el Centro de Investigaciones Sociológicas muestran que una de las principales cuestiones que todo empresario considera imprescindible para invertir en un área es la existencia de buenos transportes y comunicaciones, junto con la existencia de una mano de obra cualificada.<sup>23</sup> Solamente cuando se han cumplido estos dos requisitos, los

---

<sup>23</sup> Este tema fue tratado en el Curso Magistral que bajo el título "La Economía Española de Fines de Siglo: Convergencia y Reformas" impartió Julio Segura en la Universidad Internacional Menéndez Pelayo en

empresarios se preocupan por aspectos como la conflictividad laboral, el nivel de los salarios, etc. De la misma manera, en una encuesta de la Comisión de las Comunidades Europeas (Begg, 1995) se concluye que cuando se toma una decisión de invertir en el extranjero se tienen en cuenta diversos factores endógenos y que, sólo tras asegurarse de la presencia de los mismos, se analizan las diferencias en costes laborales. Por tanto, la actividad económica difícilmente podría llevarse a cabo en una zona desprovista de un cierto nivel de infraestructuras que aseguren la accesibilidad a los mercados de proveedores y consumidores.

Pero donde resulta más evidente la obstrucción que puede suponer en el desarrollo económico la falta de unas adecuadas infraestructuras es en las regiones económica y geográficamente periféricas (Peck, 1996). La incertidumbre o la falta de acceso a servicios provistos por las infraestructuras hace que los usuarios tengan que invertir en fuentes alternativas, con el consiguiente aumento en los costes de capital. Según Kessides (1996), un estudio basado en una encuesta a 179 establecimientos manufactureros en Nigeria pone de manifiesto que la desconfianza de las empresas en las infraestructuras, principalmente de utilización de la energía, ha hecho que el 92% de las empresas se proveyeran de generadores de electricidad propios y equipos de comunicaciones por radio, que constituyen el 15% de los costes de equipo y de maquinaria para las empresas de más de 50 trabajadores y un 25% para las empresas pequeñas. Con resultado parecido, se han realizado varios estudios en Latinoamérica. En Chile y Costa Rica, se ha concluido que por cada dólar que no se gasta en un necesario mantenimiento de las carreteras, los costes de funcionamiento de los vehículos llegan a incrementarse en tres dólares, y hace que la reconstrucción de las carreteras sea más costosa. En lo que se refiere al aprovisionamiento de agua, varios estudios muestran la importante inversión privada en que se ha incurrido para compensar una insuficiente oferta pública, con unos costes muy superiores.

En contrapartida, se apunta al concepto de *condición no suficiente* cuando se afirma que si bien nadie duda de la importancia de tener unas buenas dotaciones de infraestructuras, éstas no pueden ser consideradas como el único factor a tener en cuenta a la hora de explicar el

crecimiento de una región determinada. En esta línea apuntan autores como Johansson (1993) y Forslund y Johansson (1995) al considerar al capital público como un potencial para la región en que éste se encuentra ubicado<sup>24</sup> o al hablar de las autopistas como catalizadores del desarrollo económico (Gillis, 1989). De este modo, la cuestión más importante es saber si dicho potencial va a ser utilizado y, en tal caso, la velocidad con la que dicha economía se va a mover hacia el nivel de producción indicado por dicho potencial.

La idea de potencialidad enlaza con la visión apuntada por Biehl (1986) en la que se sugiere que un mejor equipamiento infraestructural redundará en un incremento de la productividad de la inversión privada con repercusiones en un PIB regional más alto y/o un mayor nivel de empleo, teniendo en cuenta que se refiere a los valores potenciales de los mismos. En este sentido, se ha de considerar que los valores reales pueden diferir de los potenciales por la existencia de otros factores que influyan en el índice de utilización de la capacidad total de una infraestructura, de manera que el valor real y potencial sólo coincidirán en el caso de una utilización óptima de dicha capacidad. El capital público en infraestructura representa, en conclusión, la capacidad productiva de una economía, si bien los factores privados, como el capital y el trabajo, son necesarios para explotar el potencial de desarrollo regional. Por tanto, existe un grupo de factores explicativos de la situación económica de las regiones, como son la situación geográfica, la aglomeración y la estructura sectorial, que resulta necesario introducir en toda teoría que quiera evaluar la contribución de la infraestructura al desarrollo regional. En caso contrario, se estaría sobrevalorando el papel del capital provisto públicamente.

En conclusión, tal y como señala Looney (1992), a pesar de que existe entre los economistas una amplia variedad de puntos de vista en lo que se refiere al papel de las infraestructuras, existe consenso en la necesidad de que se posea un cierto nivel de las

---

<sup>24</sup> Dichos autores se refieren, en concreto, a las infraestructuras de transportes y comunicaciones por lo que relacionan infraestructuras con accesibilidad. Según ellos, dado que el capital público consiste en recursos duraderos que pueden ser usados colectivamente por empresas y familias, las infraestructuras deben considerarse como un abanico de posibilidades de interacción, comunicación, contactos interempresariales y de mercado. Es en este sentido que se considera un potencial ya que proveen a los agentes económicos de una mayor accesibilidad a los recursos locales.

mismas para evitar que se convierta en un factor limitativo sin el cual no podría llevarse a cabo un proceso de desarrollo. Sin embargo, una vez pasado este punto resulta difícil la aproximación de las diferentes opiniones. Así, mientras unos consideran que las infraestructuras sólo sirven para aliviar las presiones generadas por los cuellos de botella, otros van mucho más lejos argumentando una verdadera influencia del factor público sobre la inversión y el crecimiento a través de un proceso de retroalimentación.<sup>25</sup> Sin embargo, la mayoría de economistas parecen adoptar una posición intermedia entre estos dos puntos de vista. Esta es la posición que Glomm y Ravikumar (1994) defienden cuando argumentan que no todo el capital público entra a formar parte de la función de producción de cada empresa, sino que sólo entraría el necesario para obtener un buen uso agregado de los factores de producción privados.<sup>26</sup> O la de Button *et al.* (1995) cuando comentan que "dadas las limitaciones de los estudios agregados de las infraestructuras (doble causalidad, flexibilidad del término infraestructura, etc.), la posición más objetiva es decir que el papel de los transportes y de otras infraestructuras es visto por muchos como una condición necesaria pero no suficiente para el desarrollo económico, en el sentido de que no ejercen un efecto primordial sino que refuerzan otros instrumentos ...". Asimismo, dichos autores consideran que cada vez más se debe analizar el papel del capital público como un problema de satisfacción y no tanto de maximización. Es decir, un cierto nivel de infraestructuras son necesarias para que un área sea considerada atractiva por parte de la iniciativa privada, pero añadir infraestructuras a dicho nivel puede tener poca importancia. De hecho, dados los costes de búsqueda de la mejor ubicación y el rango de opciones disponibles, existe una lógica en la noción de una racionalidad limitada y la búsqueda de unos objetivos umbral en lugar de buscar la localización que permite minimizar costes. O, en palabras de Wheeler y Mody (1992), aunque la provisión de infraestructuras claramente no es una condición suficiente para atraer la inversión privada a una determinada

---

<sup>25</sup> El trabajo de Flores *et al.* (1994) plantea el problema de la evaluación de los efectos a corto y largo plazo de la inversión en infraestructuras sobre el crecimiento del producto agregado en un contexto multiecuacional dinámico, de manera que se pueden tener en cuenta explícitamente la presencia de efectos de retroalimentación entre las variables consideradas (producto agregado, empleo y stock de capital en el sector privado). El resultado que obtienen muestra que el efecto a largo plazo de las infraestructuras en las variables privadas es positivo.

<sup>26</sup> De esta manera, la contribución de las infraestructuras a la productividad de los factores privados estaría sujeta a una cierta congestión.

localización, las diferencias en la calidad de las infraestructuras pueden ser un importante factor en el margen a la hora de determinar la elección entre potenciales ubicaciones.

Por tanto, la cuestión primordial no es analizar si la dotación de unas buenas infraestructuras es imprescindible en todo proceso de desarrollo económico, sino discutir y dotar de cuantificaciones económicas que nos permitan saber en qué grado la dotación de capital público explica el crecimiento económico y, en consecuencia, si se puede utilizar la inversión en infraestructuras para conseguir un mayor nivel de equilibrio interregional.

## **CAPÍTULO 3**

# **ECONOMÍA REGIONAL Y EXTERNALIDADES. ECONOMETRÍA ESPACIAL**

**“Everything is related to everything else,  
but near things are more related than distant things”**

**Tobler (1979)**

### 3.1 INTRODUCCIÓN

Los hechos económicos suelen ser el resultado de una multitud de factores que interaccionan en un marco caracterizado por múltiples condiciones. El análisis económico, por tanto, no debería aislar causas concretas e intentar explicar una realidad compleja de forma unilateral. Intentar analizar el efecto de las infraestructuras en el crecimiento económico implica el peligro de que toda una compleja red causa-efecto sea reducida a unas pocas relaciones consideradas como las más importantes. Sin embargo, difícilmente puede realizarse un estudio en el que todos los determinantes que influyen en la contribución de las infraestructuras, y que han sido expuestos en el capítulo anterior, puedan ser controlados. Lo que sí se puede hacer es incluir el estudio de los efectos de las infraestructuras en un marco de la teoría de crecimiento regional. En el presente trabajo se tienen en cuenta dos aspectos. En primer lugar, se considera la región como ámbito de estudio de forma que los resultados que se extraigan se enmarquen en el ámbito de la economía regional. En segundo lugar, se intenta evitar el traslado automático de conceptos macroeconómicos, otorgando una mayor relevancia al papel de la geografía y el espacio con la consideración explícita de *spillovers* o vínculos regionales.

Con estos dos objetivos, en este capítulo se realiza una revisión y discusión de la base teórica que es utilizada posteriormente a nivel empírico. Así, en un primer apartado se analizan las razones de la proliferación de estudios regionales en el ámbito de la economía en estos últimos años. Posteriormente se revisan las teorías del crecimiento endógeno, centrándonos en el papel de las externalidades como instrumento de transmisión del crecimiento. Finalmente, proponemos el marco de la econometría espacial como el más idóneo en el tratamiento empírico de las economías externas en los estudios regionales.

## **3.2 LA CIENCIA ECONÓMICA REGIONAL**

### **3.2.1 Origen y evolución de la Ciencia Regional. Consideración económica del espacio**

Después de la II Guerra Mundial las raíces de la ciencia regional aún se encontraban dispersas entre distintas ramas del saber. Algunos de sus fundamentos ya estaban dados por las teorías clásicas espaciales, lideradas por la Geografía. Sin embargo, otras ciencias “no espaciales” también habían hecho suya desde antiguo la dimensión espacial. En concreto, la Economía se había dedicado a estudiar la localización de la actividad económica, los costes de transporte, el comercio internacional así como la especialización nacional y regional. Sirva como ejemplo las aportaciones de varios economistas alemanes a la teoría de la localización como von Thunen (1826) que propuso la primera teoría sobre la localización uniendo la componente espacial a los procesos económicos, y Weber (1909) que fundó el análisis teórico específico de la localización industrial.

No obstante, aún faltaba establecer una teoría del orden espacial en su conjunto dentro de la ciencia económica. Según algunos autores, August Losch fue el primero en ver claramente la necesidad de alcanzar, desde la perspectiva de la teoría de la localización, un desarrollo de una teoría económica espacial en los años cuarenta. Si bien él no lo llevó a cabo en su totalidad, sí que la teoría económica empezó a experimentar un nuevo cambio en sus intereses, en los que el aspecto espacial se tenía en cuenta seriamente. El verdadero despegue de todas estas aproximaciones en una disciplina autónoma no ocurrió hasta principios de los años cincuenta. Por una parte, en Alemania, Herbert Giersch contribuyó en 1950 a la explicación de las consecuencias distribucionales de las estructuras de localización, mientras que von Boverter (1962) realizaba una explicación integrada de la estructura espacial en base a las varias teorías disponibles entonces. En EEUU, Walter Isard (1956) daba una serie de propuestas para vincular los diferentes modelos económicos que estaban a disposición y que hasta principios de los cincuenta se habían considerado de forma aislada, destacando la importancia del factor transporte en la localización. Asimismo, creó una nueva plataforma científica, la Asociación de Ciencia Regional, a fin de poner en

común las nuevas ideas en pro de la creación de una nueva ciencia. De esta forma, a mediados de los cincuenta ya se podía hablar de un marco de investigación integral para el estudio del sistema económico con diferenciación geográfica, particularmente para los países desarrollados. Así la ciencia económica regional quedaba totalmente relacionada con la teoría del comercio internacional, las ciencias del transporte y las teorías de la integración económica.

Resulta necesario resaltar dos aspectos relativos al desarrollo desde finales de los años sesenta y principios de los setenta de esta nueva rama de la Economía: la Economía Regional. Por una parte, en EEUU, junto con el desarrollo inicial de la ciencia regional se avanzó en la investigación regional metodológica. Así por ejemplo, la aplicación de las tablas input-output de Leontief al ámbito regional fue popularizada por Isard en sus "Métodos de Análisis Regional" (1960). Con el paso de los años y la mejora tanto de la disponibilidad de información estadística como de los procedimientos estadísticos y econométricos, empezó a resultar común la aplicación de modelos regionales empíricos. Por otra parte, el desarrollo de la ciencia regional ha estado orientado a la vertiente política, es decir, su centro fue íntimamente ligado a la ayuda en los sistemas de toma de decisión regional gubernamental, de forma que se estudiaran las consecuencias objetivas de las distintas formas de actuación pública, ofreciendo a los agentes decisores una base de argumentación. Hay que tener en cuenta que en estos años se creaba el mercado común de la Europa de los Seis, la integración de la Europa del Este en un mercado común y la mayor relevancia de los servicios y la tecnología, creando multitud de nuevos aspectos de los que no se sabían las consecuencias concretas.

En la década de los ochenta y hasta el momento actual, la ciencia económica regional ha prestado una gran atención al análisis de los determinantes del desarrollo y crecimiento económico. El estudio de las economías externas y las ventajas de la aglomeración, el potencial endógeno y cómo reforzarlo a través de políticas públicas, han sido temas centrales en el tratamiento de la ciencia regional. Estos aspectos se reforzaron a finales de los ochenta, con la unificación de las dos Alemanias, la reapertura de los países del este de Europa y el proyecto de la Unión Economía y Monetaria (UEM), suponiendo nuevos retos en la investigación regional, principalmente en Europa.

En conclusión, la novedad principal de la Economía Regional y su diferencia con otras ramas de la Ciencia Económica es la consideración del espacio. Los economistas clásicos y neoclásicos habían negado desde antiguo la dimensión espacial, asumiendo que el espacio es isotrópico, es decir, que sus propiedades son idénticas en todas las direcciones. Sin embargo, la economía regional introduce el espacio a través de la distancia. Y la distancia se traduce en costes. A partir de este momento, la principal función de los medios de transporte será reducir esta distancia. De este modo se introduce el concepto de sistema-espacio, creando una visión dinámica de la economía espacial. El espacio económico según MacLoughlin (1972) “es un conjunto de elementos independientes, pero cada elemento puede ser considerado como un sistema y el sistema entero puede considerarse como un elemento de un sistema mayor”. Y cada elemento tiene sus características propias, de forma que “el crecimiento no aparece en todas partes al mismo tiempo; aparece en polos de crecimiento con intensidades variables y se extiende a través de diferentes canales con diferentes efectos en la economía global” (Perroux, 1954). Estas ideas fueron innovadoras en los años cincuenta, pero han sido retocadas y mejoradas en los trabajos de Lajugie *et al.* (1985), Paelinck (1985) y Nijkamp y Paelinck (1975).

### 3.2.2 La política regional

Según la teoría neoclásica convencional, si nos basamos exclusivamente en términos de eficiencia, no existe justificación alguna de la corrección de la localización espacial obtenida según las fuerzas del mercado. La única justificación que se puede encontrar para la política regional consiste en utilizarla como un instrumento redistribucional del tipo *second best*. Está claro que siempre hay espacio para la mejora de Pareto en la distribución de la localización de la actividad obtenida por el mercado. Sin embargo, según muchos autores, esto no significa que esté justificada la intervención pública. El coste de bienestar de los fallos de mercado tiene que ser comparado con los costes de bienestar de los fallos de la actuación pública.

Observando la política regional en la práctica, se encuentran varias normas de actuación que, sin embargo, no se sabe si suponen un incremento de la eficiencia. La política regional de la UE, por ejemplo, consiste en reunir dinero de las zonas más adelantadas, principalmente del centro, y pasarlo a la periferia. Siguiendo la nueva teoría de la economía espacial no se puede saber si es una buena norma en términos de eficiencia. En las nuevas teorías de la localización se observa la concentración de la actividad económica y se considera que es eficiente. Aunque puede suceder que se requiera más concentración para aumentar la eficiencia. Incluso puede suceder que las actividades estén concentradas en un lugar equivocado debido a la coincidencia histórica o que el número de ciudades en la economía sea ineficiente. Sin embargo, modelos espaciales ineficientes sobreviven. Quizás fuera preferible que todas las empresas se trasladaran de localización, pero no lo hacen porque individualmente les es más eficiente permanecer junto a las demás. El modelo específico de aglomeración puede estar equivocado pero no existe nada incorrecto en la aglomeración en sí misma. Según autores como Matsuyama y Takahashi (1993) existe una clara indicación de que el mercado produce más bien poca concentración. De hecho, según señalan dichos autores, a nivel de las regiones, la política regional acepta dicha proposición, acumulando las inversiones en lugares céntricos. Sin embargo, en un ámbito mayor, a nivel nacional o por ejemplo el europeo, la política que se persigue es la de la igualación. Y la cuestión de si la nueva teoría de la localización aboga en favor de la equidad es una cuestión diferente. La nueva teoría se limita a señalar la falta de confianza en la idea neoclásica de que los niveles de bienestar tienden a converger en una economía integrada. Los resultados han mostrado que en el curso de una mayor integración, las regiones periféricas pueden perder no únicamente en términos relativos sino absolutos, por lo que se requiere una compensación por razones de justicia social.

Respecto a la política regional, hay que partir de dos aspectos de la realidad. Por una parte, estamos asistiendo a un proceso de globalización de los mercados junto a la creación de empresas multinacionales que hacen que cada vez sea más necesario el abordar el análisis de la economía desde una perspectiva global, lo que hace disminuir las capacidades de los gobiernos para afectar la evolución de las economías regionales en el largo plazo. Por otra parte, no es menos cierto el incremento de la popularidad de movimientos de regionalización, y la petición de los gobiernos regionales de reformas federalistas en la

mayoría de países europeos. Este proceso ha hecho que las autoridades nacionales y de la UE vean cada vez más difícil el apoyar una política coherente que incremente la cohesión entre las regiones desarrolladas y las retrasadas. Todos estos procesos, que en un principio parecen antagónicos, han hecho cambiar de forma importante el entorno en el que la actividad económica se lleva a cabo. Así, el proceso de internacionalización de las empresas requiere una descentralización de las funciones operativas y la creación de alianzas flexibles con empresas extranjeras. Dicha internacionalización requiere que las empresas tengan capacidad suficiente para trabajar en diferentes entornos. Además, en tal entorno surgen modernas formas de organización de las empresas que implican el desarrollo de relaciones cooperativas entre los diversos actores locales en un sistema de producción local. De esta manera, la política regional sugiere que el papel de las instituciones públicas se base en la representación de un ente integrador del sistema y en la toma de iniciativas de proyectos estratégicos que requieren la cooperación de los diferentes actores regionales.

### **3.2.3 El nivel regional en Europa**

Se ha debatido ampliamente sobre el futuro del nivel regional en el proceso de decisión política de la UE. La política regional en la UE no fue introducida hasta mediados de los setenta con la creación del Fondo Europeo de Desarrollo Regional (FEDER). Dicho fondo empezó representando tan solo un 5% del presupuesto de la Comunidad en sus primeros años, por lo que no resultaba un instrumento efectivo. Según Armstrong (1995), la política regional tardó unos 15 años en crearse de verdad, después de una serie de reformas en 1988. Hasta entonces había estado centrada en el gasto agrícola y los estados miembros consideraban la política regional como un asunto nacional. Sin embargo, con el establecimiento de un Mercado Único, se hacía necesario crear una verdadera asistencia a aquellas economías regionales que necesitaban ser reestructuradas. De este modo, el presupuesto del FEDER se dobló entre 1988 y 1993, junto con el establecimiento de los principios en los cuales se asienta la política regional europea:

1. *Concentración* de la ayuda en las áreas con más necesidades: fomentar el desarrollo y el ajuste estructural en las regiones menos desarrolladas, reconvertir las regiones o partes de regiones gravemente afectadas por el declive industrial, combatir el paro de larga duración, facilitar la inserción profesional de los jóvenes, acelerar la adaptación de las estructuras agrarias y fomentar el desarrollo de las zonas rurales.
2. *Coordinación* y planificación de la actuación de la UE en los problemas regionales: estrecha *colaboración* entre la Comisión y todas las autoridades competentes a escala nacional, regional o local designada por cada Estado.
3. *Adicionalidad* de los fondos de la UE, complementando las políticas regionales de los estados miembros.
4. Establecimiento de la prioridad en los *programas de asistencia*.

Asimismo, en el ejercicio del poder a nivel regional y federal en muchos de los Estados miembros se ha incrementado la demanda de una mayor representación del nivel regional en la arquitectura de la UE. Como resultado se estableció el Comité de las Regiones, con representación de los cuerpos regionales y locales y con *status* consultivo. Dicho Comité debiera ser consultado por la Comisión en cada cuestión que tenga un impacto en el desarrollo regional y en cada iniciativa legislativa vinculada a la política regional de la UE. De esta manera, hay que diferenciar la Política Regional de la Política de las regiones. La primera tiene como finalidad reducir las diferencias regionales a través del apoyo a las inversiones públicas y privadas en el marco de programas regionales, es decir, conseguir un desarrollo armónico de todas las regiones. La segunda, sin embargo, es una política enfocada a permitir que todas las regiones, independientemente de su situación económica, geográfica o política, tengan una participación más integrada en el proceso de decisión política de la UE.

Entre los temas que han centrado la atención de los investigadores regionales estos últimos años caben citar los siguientes:<sup>1</sup>

---

<sup>1</sup> La consideración de estos aspectos ha sido especialmente relevante en el caso europeo dado el nuevo entorno regional creado con la UE. No obstante, dichos temas han sido tratados y aplicados en todo el mundo, pues el proceso de globalización de la economía no escapa a ningún país.

- la consecuencia de la integración de las regiones en el proceso de convergencia o divergencia, es decir, el aumento o disminución de las disparidades regionales
- el estudio de la distribución de la actividad económica en el marco de la nueva estructura regional creada
- los impactos de la integración regional en la división internacional e interregional del empleo
- el estudio de los requisitos para el desarrollo de las oportunidades económicas que ofrece el Mercado Único, prestando especial importancia a las infraestructuras

Si nos centramos en el caso español, se constata que el interés por el conocimiento de los patrones y los factores determinantes del crecimiento regional ha recibido un nuevo impulso en la última década por varios motivos, entre otros:

- el proceso de descentralización y consolidación del Estado de las Autonomías
- ampliación y profundización en el proceso de integración europeo del que España forma parte desde 1986
- constatación del comportamiento diferenciado entre la economía nacional y las regionales, es decir, la existencia de especificidades en el comportamiento regional
- debate acerca del efecto real de las políticas de desarrollo regional y de cohesión en la reducción de las disparidades
- aumento en la cantidad y calidad de la información estadística disponible a un mayor grado de desagregación territorial

En el ámbito regional español, múltiples aportaciones han destacado un proceso de crecimiento importante en todas las regiones españolas. Dicho proceso se ha caracterizado por una destacada capitalización y por un cambio en la composición de la estructura productiva (industrialización y posterior terciarización), junto a un importante trasvase de mano de obra entre regiones en las primeras fases del mismo. Sin embargo, en este proceso de crecimiento generalizado, también se ha evidenciado una cierta heterogeneidad en la evolución de la actividad económica entre regiones. Entre las posibles razones de la existencia de tal heterogeneidad se puede pensar en la diferente capacidad de atracción de

actividad, de competir en mercados internos y externos, composición de la estructura productiva, así como la existencia de una cultura empresarial o la estabilidad político-social de una región.

Por tanto, si a un nivel económico-político se ha observado en las últimas décadas que las regiones juegan un papel cada día más importante, a nivel de los estudios regionales dicho papel debe tenerse en cuenta. En nuestra opinión, se debe evitar el traslado automático de conceptos macroeconómicos, destacando el papel de la geografía y el espacio, con la consideración explícita de los vínculos interregionales. Éstos son tratados detalladamente en la siguiente sección.

### **3.3 LA PRESENCIA DE EFECTOS EXTERNOS EN LOS ESTUDIOS REGIONALES**

La economía mundial ha experimentado en las últimas décadas un importante proceso de globalización. Las relaciones comerciales son cada vez más internacionales, especialmente entre países de áreas comerciales integradas como la UE o la NAFTA. Los países comercian unos con otros, se comunican entre sí y aprenden de los demás como nunca había sucedido. Sin embargo, la teoría del crecimiento tradicionalmente ha tratado a cada país como si fuera una isla de forma que el crecimiento del producto o de la productividad de un país depende del crecimiento de sus propios factores de producción. Por el contrario, ¿no es probable que existan fuentes del crecimiento que no respeten fronteras y se encuentren fuera del ámbito propio de la región?, es decir, ¿no parece lógico pensar que el creciente intercambio de bienes, capitales e ideas puede haber llevado a una creciente interdependencia en el crecimiento de los países y/o regiones?

En este apartado se tratan detalladamente las externalidades, las nuevas teorías del crecimiento que las incorporan y los efectos que las mismas poseen sobre la localización de la actividad y la convergencia.

### 3.3.1 Las externalidades y las teorías del crecimiento endógeno

#### 3.3.1.1 El desbordamiento del crecimiento: Economías externas. Idea general

Modelos teóricos recientes han puesto especial énfasis en el papel de los efectos externos, economías externas o externalidades como factor de crecimiento económico. El incremento del stock de capital genera en sí mismo una mejora en el nivel tecnológico que no puede ser apropiada totalmente por el agente que realiza la inversión. En consecuencia, el rendimiento agregado o social de la inversión es superior al rendimiento que obtiene el agente individual, por lo que se asume que el conocimiento se difunde por toda la economía sin respetar fronteras y afectando, por tanto, al nivel tecnológico de cada empresa.

Asimismo, si se analizan las razones que influyen en las decisiones de localización de las empresas, o alternativamente, si se estudian los posibles determinantes de la distribución de la actividad económica en el espacio, se observa que existen patrones de localización de las actividades manufactureras a lo largo del tiempo. Pese a la libertad de localización de la actividad (especialmente en el momento presente en el que se disfruta de libre circulación de los flujos económicos), la misma no está distribuida aleatoriamente en el espacio económico. Así, con independencia del emplazamiento de los recursos naturales, existe una tendencia a la concentración en unas cuantas localizaciones que ofrecen ventajas a las empresas especializadas en actividades complementarias de un mismo proceso productivo. Dicha complementariedad indica, por una parte, la división del trabajo o especialización dentro de un territorio y de una industria. Por otra parte, Marshall (1920) explica la concentración de industrias especializadas en un territorio a través del concepto de economías externas como fuerza centrípeta. La concentración geográfica de las actividades económicas es comparado con el efecto bola de nieve.<sup>2</sup> Los agentes desean concentrarse dados los diversos factores que permiten obtener una mayor diversidad y una mayor especialización en el proceso productivo. El establecimiento de nuevas empresas en tales áreas da lugar a nuevos incentivos a los trabajadores para inmigrar porque pueden esperar mejores trabajos y, en consecuencia, mayores salarios. Asimismo, las empresas se beneficiarán de la disponibilidad de una mano de obra cualificada y de ciertos servicios, lo

---

<sup>2</sup> Este fenómeno es similar al que se trata en las externalidades de red (David y Greenstein, 1990).

cual hace el lugar atractivo para nuevas empresas. En conclusión, ambos agentes se benefician el uno del otro, tal como describe Marshall:

“Cuando una industria ha elegido una localización para situarse, es probable que permanezca en la misma durante largo tiempo. ... Una industria obtiene ventajas gracias al hecho de que un lugar le ofrece un mercado constante de trabajadores cualificados... Los empresarios están dispuestos a desplazarse al lugar donde pueden encontrar una buena disponibilidad de trabajadores con la preparación que dicha actividad requiere; asimismo, la gente busca los trabajos en aquellos lugares en los que las empresas necesitan empleados con dicha preparación y en donde puede encontrar por tanto un buen mercado.”

Se observa que las industrias tienden a concentrar sus actividades en ciertas localizaciones gracias a la existencia de una serie de efectos externos asociados a dichas ubicaciones pero a ninguna empresa en concreto. De forma más específica, las *externalidades marshallianas* explican la concentración geográfica de la actividad económica en un territorio gracias a la presencia de un mercado de mano de obra cualificada, una oferta de los factores concretos necesarios para el proceso productivo y la mayor rapidez en la difusión de la información y de la tecnología en dicho territorio. Las empresas se benefician por el simple hecho de encontrarse ubicadas en dicho entorno. Con todo ello, aparece la teoría marshalliana que trata de reconciliar los rendimientos crecientes a escala con la competencia perfecta. En los distritos industriales marshallianos las empresas son precio-aceptantes, obteniéndose rendimientos crecientes en el conjunto de la industria, a pesar de que los rendimientos de cada empresa puedan no serlo.

### 3.3.1.2 Las teorías del crecimiento endógeno y las externalidades<sup>3</sup>

La teoría neoclásica del crecimiento había dejado relegado a un segundo término el papel que el progreso tecnológico supone sobre las actividades económicas, considerándolo un mero determinante exógeno de las mismas. La corriente principal de la economía del siglo XIX consideraba que el cambio tecnológico estaba fuera del alcance de las leyes

---

<sup>3</sup> Los libros de Grossman y Helpman (1991a) y Barro y Sala-i-Martin (1995) son una buena referencia para la revisión de las nuevas teorías del crecimiento económico.

económicas, introduciéndolo de manera exógena en los modelos de crecimiento. Solow (1957) demostró que sólo una parte del crecimiento per cápita era consecuencia de un incremento del stock de capital y la fuerza del trabajo, argumentando que el progreso tecnológico debía explicar parte de ese residual no explicado. En palabras de Romer (1990):

"Si EEUU se hubiera administrado para alcanzar hacia 1860 el mismo nivel de capital por trabajador como el que lograron en 1960, parece totalmente implausible que el output por trabajador hubiera sido tan alto como lo fue en 1960. Algo más sobre la producción en 1960 fue manifiestamente diferente".

Sin embargo, Solow (1956) no explicaba el crecimiento en el largo plazo. La teoría de Solow únicamente permitía explicar la transición a un estado estacionario en el largo plazo, pero dicho estado estacionario es exógeno en su teoría. El progreso técnico es un maná caído del cielo que hace que el empleo sea más productivo, por lo que no ofrece ningún tipo de guía sobre como fomentar el crecimiento. Si la acumulación de conocimientos es en realidad la causa principal del progreso económico, y si a la vez es exógena, en palabras de de la Fuente (1992) en su revisión sobre el crecimiento y el progreso técnico "el crecimiento resultaba ser un almuerzo gratis, pero un almuerzo en el que los comensales no tenían nada que decir sobre el menú". Todos los intentos por convertir en endógenas las tasas del crecimiento en el largo plazo fueron un fracaso, dado que las bases neoclásicas se mantenían. El fracaso de dichos modelos hacía evidente la necesidad de enfrentarse con el cambio técnico. Y el crecimiento endógeno y la teoría neoclásica son incompatibles, pues esta segunda considera modelos de competencia perfecta, sin externalidades y con tecnologías convexas que impiden la existencia de economías de escala.

La nueva teoría del crecimiento introduce el stock del conocimiento como un factor adicional de producción. Las inversiones en tecnología se producen gracias al uso de recursos tomados de los bienes de producción. En este marco, el crecimiento en el largo plazo es posible si los factores de producción aumentables, como el conocimiento y el capital, presentan rendimientos no decrecientes a escala en la función de producción

agregada (por ejemplo, si sus respectivas elasticidades de producción parciales suman no menos que la unidad). Esto implicaría que la función de producción en su conjunto presenta rendimientos crecientes a escala ya que la función de producción contiene, aparte de los factores aumentables ya mencionados, factores de producción no aumentables como el empleo y la tierra, que presentan elasticidades parciales de producción estrictamente positivas. Por tanto, las elasticidades de todos los factores sumarán una cantidad estrictamente superior a la unidad.

El problema se presenta cuando se quiere reconciliar tales rendimientos crecientes con la teoría de la competencia perfecta. Bajo los supuestos clásicos de competencia perfecta y rendimientos constantes, los factores utilizados son pagados de forma que agotan el producto total. Y por tanto, resulta imposible invertir en actividades de desarrollo como la tecnología, que requieren de recursos que deben ser financiados. Marshall (1920) fue el primero en intentar combinar la existencia de rendimientos crecientes y una estructura competitiva de mercado a través de la presencia de un conjunto de externalidades. La principal idea marshalliana se basa en considerar que la concentración de la industria en un área determinada permite que las empresas puedan beneficiarse de la existencia de un conjunto de trabajadores cualificados, de un mercado de productos intermedios y de la más fácil difusión de los nuevos avances tecnológicos. De este modo, se acepta que existan rendimientos crecientes externos que se manifiestan en la industria o la economía pero que a nivel de empresa no son crecientes. En este caso, por ejemplo, el conocimiento y la tecnología serán el resultado de otras actividades, de forma que su producción no requiere compensación y los factores tradicionales pueden recibir sus productos marginales. De este modo, Marshall da un paso al alejarse de las hipótesis de convexidad en el crecimiento que limitaban el entendimiento del proceso de crecimiento, y los rendimientos crecientes son el resultado de una externalidad que entra sin problemas en un marco competitivo.

Las ideas de Marshall fueron las primeras en considerar la innovación tecnológica como un elemento clave no sólo para la industria individual sino como factor explicativo del crecimiento económico de la economía global. Siguiendo la idea marshalliana, Arrow (1962) describió un modelo de crecimiento que permitía rendimientos crecientes externos a la empresa. En dicho modelo, argumentaba que el conocimiento se adquiere a través de la

práctica (*learning by doing*), lo que permite conseguir importantes aumentos en la productividad. Asimismo, consideraba que la inversión acumulada, o stock de capital, es una medida utilizable del aumento de la experiencia. De este modo, según Arrow, el desarrollo tecnológico crece de forma paralela a la inversión.

Llegado este punto, dos son las ideas básicas en torno al progreso tecnológico: el efecto desbordamiento o difusión de las ideas y el aprendizaje por la práctica. Estos dos puntos han permitido transformar el modelo neoclásico en un modelo descriptivo de crecimiento endógeno que reconcilia la existencia de rendimientos crecientes y las leyes de la competencia perfecta: los modelos de crecimiento endógeno.

En los modelos de crecimiento endógeno, el motor del crecimiento es algún tipo de proceso que se puede caracterizar como generador de progreso técnico. Estas actividades, como se explica posteriormente, incluyen la inversión en I+D dirigida al desarrollo de nuevos productos y procesos productivos y la acumulación de capital humano. El tratamiento del fenómeno tecnológico de forma endógena lleva a una visión del crecimiento más amplia que la que había dominado hasta entonces, desde el punto de vista de que los modelos construidos sobre esta base permiten la consideración de distintas políticas y sus efectos sobre el ritmo de desarrollo. Antes de pasar a explicar los modelos de crecimiento endógeno, hay que tener en cuenta dos ideas. En primer lugar, los modelos que se ofrecen se han de tratar como descripciones parciales y complementarias de una misma realidad compleja y no como teorías alternativas y mutuamente excluyentes. En segundo lugar, a pesar de que la teoría del crecimiento endógeno de finales de los años ochenta es la que, tal como se verá, da los modelos formales que incluyen la tecnología y las externalidades como verdaderos motores del crecimiento, el tema no es un fenómeno nuevo. Así, como ya se ha comentado anteriormente, autores como Marshall a principios de siglo y Arrow (1962), Kaldor (1957, 1961), Kaldor y Mirrlees (1962), Shell (1966, 1967) y Nordhaus (1967, 1969) en los años sesenta ya se habían preocupado por los mismos temas centrales e incluso dado modelos formales. La cuestión es que hubo un lapsus de más de diez años en dicha literatura que no empezó a reactivarse hasta 1985-1986. Entre las razones que argumenta de la Fuente (1992) para esta reactivación nos encontramos con la reducción de

los obstáculos de carácter técnico (como el avance en el desarrollo de modelos tratables de competencia monopolística y otras contribuciones en el área de la organización industrial) y también por el movimiento pendular de los intereses de los macroeconomistas. Tras una década de estudios sobre las implicaciones de la hipótesis de las expectativas racionales y los modelos de ciclo económico basados en problemas de información, el tema empezaba a agotarse por lo que en las conferencias de Lucas (1985) se vino a decir que la productividad marginal de la política de estabilización es mucho menor que la de la promoción del crecimiento, por lo que este segundo tema era el que tenía que desarrollarse de forma prioritaria.

Antes de analizar las diferencias entre los modelos de crecimiento endógeno, de forma genérica se puede decir que los mismos consideran una función de producción aumentada por un índice de eficiencia tecnológica,  $Y = f(K, L, A)$ , en la que la senda temporal de  $A$  se determina endógenamente, con diferentes formas de especificar la evolución de  $A$ , lo cual da lugar a los diferentes modelos de crecimiento endógeno.  $A$  puede ser un subproducto de otras actividades, es decir, el progreso técnico aparece como el resultado casi accidental de otras actividades económicas (como la adquisición del saber) o gracias a la asignación explícita de recursos (inversión en I+D o preparación del capital humano).

### **El capital físico**

Según Romer (1986) la inversión en capital físico puede ser considerada origen del crecimiento endógeno ya que genera externalidades que provienen de varias fuentes.<sup>4</sup> Por una parte, Romer habla del proceso de difusión del aprendizaje (*learning spillover*), por el que al acumular capital, la empresa acumula al mismo tiempo conocimientos de los que se benefician el resto de empresas. Esta idea reposa en la formulada por Arrow (1962) que argumenta que el conocimiento se adquiere a través de la práctica (*learning by doing*), lo que permite conseguir importantes aumentos en la productividad. La familiaridad creciente

---

<sup>4</sup> Romer considera que las externalidades tecnológicas positivas se producen como consecuencia de la acumulación de un factor capital, sin nombrarlo como capital físico; incluso emplea la expresión "conocimientos" y no la de "capital" en la formulación de su modelo, aunque implícitamente se refiere al capital físico.

con un proceso productivo lleva al aumento de la productividad a través del aprendizaje.<sup>5</sup> De este modo, el progreso técnico aparece como un subproducto de las actividades normales de producción e inversión. La inversión y el progreso técnico están completamente ligados. El desarrollo del conocimiento por la práctica implica, en consecuencia, que las experiencias previas de los trabajadores en un proceso productivo son importantes para incrementar la productividad ya que es la propia actividad la que va sometiendo a los trabajadores a nuevos problemas que requieren respuestas favorables en el tiempo. Arrow supone una tecnología de coeficientes fijos en la que la cantidad de trabajo necesaria para producir una unidad de output en una máquina decrece con la experiencia siguiendo una curva de aprendizaje. Una implicación de tal modelo es que el nivel de inversión en equilibrio será inferior al óptimo porque los agentes no tienen en cuenta la contribución de la inversión al aumento de la productividad. Otra consecuencia es que la participación de cada factor en el producto nacional ya no es igual a la elasticidad del output con respecto al mismo debido a la existencia de la externalidad, de forma que se estará subestimando la contribución del capital al aumento de la renta si se hace por el método tradicional de contabilidad del crecimiento. En el modelo de Arrow, la tasa de crecimiento de la renta per cápita en el estado estacionario aumenta con el coeficiente de aprendizaje y con la tasa de crecimiento de la población, ya que un aumento de cualquiera de estos dos parámetros supondrá un aumento de la velocidad del aprendizaje.

Asimismo, dicho autor considera que la inversión acumulada o stock de capital es una medida utilizable del aumento de la experiencia. La inversión bruta acumulada (producción acumulada de bienes de capital) representa la experiencia ya que cada nueva máquina puesta en funcionamiento cambia el ambiente en el que la producción tiene lugar y en segundo lugar, porque los bienes de capital nuevos son los que incorporan todo el conocimiento, entrando así el aprendizaje en la producción. El desarrollo tecnológico crecería, por tanto, de forma paralela a la inversión.

---

<sup>5</sup> Según de la Fuente (1992) el trabajo de Kaldor (1957, 1961) es anterior al de Arrow y trata las mismas cuestiones, pero sin embargo ha recibido menos atención en parte porque el autor trabaja con modelos que resultan poco familiares para los economistas.

No obstante, una vez que una empresa ha aumentado sus conocimientos, todas las empresas tienen acceso a éstos permitiendo un aumento de la productividad de las mismas, sin que éstas paguen ningún tipo de compensación. Este fenómeno se conoce como difusión tecnológica (*knowledge spillovers*), una de las externalidades marshallianas, y da explicación a la evidencia de que los sectores industriales se encuentran localizados en áreas geográficas determinadas (concentración de industrias). De este modo, cuando las empresas compran un producto a otra empresa, están pagando un precio que es menor que toda la información que dicho producto recoge, en el sentido de que no resulta internalizado en la empresa innovadora todo el beneficio que ha supuesto la innovación y mejora del producto o del proceso productivo. Es por esta razón que se conoce a este fenómeno como desbordamiento del conocimiento. Una posible razón explicativa de estos efectos es que existe poca correlación entre la unidad geográfica básica de análisis (el país o nación) y el área en que la tecnología tiene sus efectos. Así, las fronteras de los países son definidas de forma administrativa por razones políticas mientras que la unidad relevante a considerar sería aquella en la cual un determinado tipo de conocimiento se desborda.

La productividad de una empresa presentará, por tanto, una función creciente respecto a la inversión agregada acumulada en la industria, de forma que el conocimiento adquirido por el empleo es una función del stock de capital total. De esta manera, el simple acto de incrementar el stock de capital a través de la inversión de las empresas, incrementará el nivel del conocimiento en la industria en general. La economía en su conjunto estará operando bajo rendimientos crecientes a escala, lo cual es perfectamente coherente con la existencia de una productividad marginal decreciente del bien conocimiento. Esta noción de rendimientos crecientes externos a la empresa se encuentra en la tradición de Marshall (1920) y Young (1928) que reconcilia los rendimientos crecientes con la teoría de la productividad marginal de la distribución que implicaría, de otro modo, que a los factores se les estaría pagando más que el output total (en caso de que los factores fueran pagados según su producto marginal).<sup>6</sup>

---

<sup>6</sup> Existe un modelo reciente de King y Robson (1989) que vincula el crecimiento de la productividad directamente a la inversión. Dichos autores consideran una función del progreso técnico que enfatiza la idea del “*learning by watching*” (aprender mirando). Según dicha idea, los nuevos proyectos de inversión en un sector de la economía tienen un efecto demostración o efecto contagio sobre la eficiencia de otros sectores, permitiendo de este modo, un incremento de su output a través del uso de factores ya existentes.

El modelo de Arrow ha sido generalizado y extendido por Levhari (1966), Sheshinski (1967) y Black (1969) sin apartarse de la conclusión de que socialmente se invierte y se produce poco. Levhari sigue con el mismo modelo de Arrow pero abandonando el supuesto de coeficientes fijos y los otros dos trabajos estudian el problema de la acumulación óptima de capital. Stokey (1988) y Young (1991) también incorporan recientemente procesos de aprendizaje en estudios de crecimiento endógeno.

Romer (1986, 1989) ofrece una variante en estos modelos de “aprender haciendo” o “aprendizaje por la práctica”. El rasgo esencial del trabajo de Romer es que el conocimiento presenta una productividad marginal creciente. En concreto, el modelo de Romer considera que el *nuevo conocimiento*, el último determinante del crecimiento a largo plazo, se produce gracias a la inversión en tecnología de investigación que presenta rendimientos decrecientes. La creación de nuevo conocimiento por una firma aumenta las posibilidades de producción de otras firmas debido a la inadecuación de la protección de patentes. Pero, y aquí es donde viene el punto principal de su modelo, la producción de bienes que provienen de conocimiento incrementado sí presenta rendimientos crecientes, de forma que se puede concluir que el conocimiento presenta una productividad marginal creciente.

Según el modelo de Romer ya no resulta válida la conclusión neoclásica de la convergencia de los ingresos per cápita en dos sectores o modelos de comercio internacional. Las economías menos avanzadas experimentarán tasas de crecimiento menores que las economías más desarrolladas, ya que el capital y la inversión de los países pobres se trasladará a los ricos en donde su productividad marginal es mayor. De esta manera, se hace mayor la distancia entre países ricos y pobres.

En último término, existe otro mecanismo de externalidad ligado al capital físico referido al mismo capital y no a los conocimientos que éste engendraría. Esta idea, desarrollada por Durlauf (1991) se basa en la existencia de complementariedades entre actividades y entre

---

Según este modelo, existen múltiples sendas de crecimiento del estado estacionario, incluso para economías que tienen similares dotaciones iniciales.

empresas, ante mercados incompletos, produciéndose un desarrollo simultáneo de varias actividades que sería necesario para el crecimiento económico (por ejemplo, una empresa de construcción requiere de empresas proveedoras de materiales). Es una idea similar a la desarrollada en el trabajo de Murphy *et al.* (1989) referente al “big push”.

A través de las externalidades tecnológicas, nuevas parcelas del conocimiento se difunden a otras empresas de la economía, aumentando su propio nivel tecnológico. Azariadis y Drazen (1990) han puesto énfasis en el papel de la interacción entre el conocimiento individual o privado y el conocimiento agregado: los rendimientos de adicionales parcelas de conocimiento privado serían mayores para un nivel suficientemente grande de conocimiento agregado.

#### **Investigación y Desarrollo (I+D)**

En contraposición al modelo precedente en el que el crecimiento endógeno se atribuye a una externalidad debida al capital físico, Romer (1990) propuso una visión alternativa que se basa en el papel básico de la innovación tecnológica obtenida mediante recursos dedicados a investigación y desarrollo (I+D). De este modo, si bien es cierto que el progreso técnico se incorpora a través del capital físico, no proviene de la inversión en este tipo de capital, sino que es la aplicación de un nuevo input productivo, la tecnología, lo que permite aumentar la productividad del sector del bien final. La tecnología es, por tanto, un factor intermedio cuya producción requiere la utilización de recursos.

En el modelo de Romer, el conocimiento entra en la producción de dos maneras diferentes. Una nueva idea permite la producción de un nuevo factor intermedio. Pero también una nueva idea permite aumentar el stock total de conocimiento y por tanto, la productividad del capital humano empleado en el sector de la investigación.

Asimismo, la tecnología es un factor no rival en el sentido de que su uso por una empresa no impide que otras lo puedan utilizar. El tratamiento del conocimiento como un bien no rival explica los *spillovers* del conocimiento y niega los rendimientos constantes a escala en la función de producción dado que no es necesario replicar los factores no rivales. Es por esta razón que en el modelo de Romer, la innovación requiere de un poder de monopolio,

que es el que les permite obtener rentas asociadas a los nuevos productos durante un tiempo. Así, el monopolio sería el verdadero incentivo para que las empresas sigan un proceso de innovación, ya que permite la internalización de parte de las externalidades en el innovador, restringiendo el flujo de ideas sin compensación y permitiendo una aceleración de la innovación y del crecimiento. Según este modelo, el conocimiento se da en forma de patentes (Romer, 1990) o nuevos métodos (Aghion y Howitt, 1992). La falta de derechos de propiedad o de apropiación supondría una reducción de las inversiones de las empresas en nuevas tecnologías, ya que si el nuevo producto puede ser copiado rápidamente y sin costes, el innovador no puede recuperar los costes del desarrollo de la tecnología. Mediante los derechos de propiedad o patentes se reduce el problema.

Grossman y Helpman (1991a) presentan un modelo similar. La idea de ambos trabajos es, por tanto, que la investigación da lugar a dos productos: ideas y diseños para nuevos bienes e información que reduce el coste de innovaciones futuras. El valor del primer producto es apropiable pero el del segundo no. Sin embargo, se está generando una externalidad que es la que permitirá que se consiga crecimiento en el largo plazo. Esto es así ya que la reducción de costes que supone compensa la tendencia de la creciente saturación del mercado a reducir el valor de nuevas invenciones. De esta manera se generan estados estacionarios (con especificación concreta de la tecnología de investigación) en los que las dos fuerzas se anulan y la innovación continúa a un ritmo constante.<sup>7</sup>

No obstante, si los aspectos recién comentados diferencian al modelo de Romer (1990) de los anteriores, por el contrario, resulta similar la conclusión de que un crecimiento sostenido de la actividad acumulativa asentado en la existencia de rendimientos de escala crecientes resulta necesario para el crecimiento sostenido del producto final. Y estos rendimientos crecientes se obtienen como consecuencia de la naturaleza particular de los conocimientos. Asimismo, se concluye que el tipo de crecimiento óptimo es superior al tipo de crecimiento de equilibrio, dado que la fracción de capital humano dedicada a la investigación resulta más elevada en el óptimo que en el punto de equilibrio, es decir, que el rendimiento privado de la investigación es inferior a su rendimiento social. La

---

<sup>7</sup> De la Fuente (1992) ofrece una buena ilustración de estas ideas.

explicación de ello se basa en que los agentes considerados de manera aislada no tienen en cuenta la externalidad generada por los resultados de su investigación y, por el contrario, un planificador sí las tomaría en cuenta.

### **El capital humano**

Una tercera fuente de crecimiento endógeno es la acumulación de capital humano. Estos modelos tienen la misma base que los anteriores, es decir, el saber técnico se concibe como el resultado de la inversión de las empresa en formación, pero donde dicho saber técnico se mira no como un saber abstracto (ideas para nuevos productos o procedimientos) sino que es el resultado de la preparación de los trabajadores, para lo que se requiere tiempo y recursos destinados a la educación. Es lo que se conoce como el capital humano de una economía. Éste tiene la doble característica de estar constituido por el saber (como la tecnología) y de ser apropiable por los individuos (como el capital físico) ya que se encuentra incorporado a los mismos. Según Amable y Guellec (1992), al ser saber, se propaga de los profesores a los alumnos, los cuales utilizan sus conocimientos presentes para adquirir otros nuevos, emparentando el capital humano con el conocimiento técnico, pudiéndosele aplicar las reglas de la acumulación con rendimientos de escala dinámicos. Así, en palabras de Romer (1990):

"El capital humano puede crecer sin límite cuando se mide en una base per cápita. Como tal, difiere de forma crucial de la noción usual del capital humano medido por los economistas laborales en años de escolaridad o experiencia de la fuerza de trabajo. Este nuevo tipo de capital humano puede pasar de un individuo al siguiente, o de una generación a la siguiente. La posibilidad de transmisión sugiere que el capital humano definido de esta forma debe contener algo así como ideas científicas o el stock de conocimiento o el nivel de tecnología, que tiene una existencia fuera de todo individuo".

Por otra parte, al contrario del capital físico, que presenta rendimientos decrecientes, el capital humano puede, en sí mismo, suponer un proceso de crecimiento endógeno. La idea básica de este modelo es, por tanto, que el saber es productivo únicamente desde el

momento en que se incorpora a la fuerza laboral. De forma que la educación es un elemento importante en el proceso de crecimiento.

En el modelo de Lucas (1988)<sup>8</sup>, dado que el capital humano está incorporado en los individuos se puede decir que es un factor apropiable, de modo que no se puede producir una igualación de los niveles de capital humano entre las naciones. Es decir, la calidad de la mano de obra es, según dicho autor, un bien propio de una zona. De este modo, la eficiencia de cada individuo en la producción del bien final es tanto mayor a medida que aumenta el nivel medio de capital humano, lo que supone un efecto externo positivo en relación a la acumulación individual de dicho capital. Lucas (1988) enfatiza la distinción entre efectos internos del capital humano donde el rendimiento se refiere al individuo que está llevando a cabo la preparación y los efectos externos que se trasladan a cambios en el output.

La implicación del modelo de Lucas (1988) y el de Romer (1990) es que los beneficios de una nueva idea son parcialmente excluibles y retenidos por la parte que ha iniciado el estudio de la nueva idea. Sin embargo, existe una diferencia importante entre los modelos de Lucas y Romer. En el de Lucas, se asume que la producción del capital humano genera un bien no rival y no excluible, de forma que el conocimiento emerge como un bien público cuya producción no puede ser explicada en términos de inversión de la firma en I+D. A través de la introducción de la excluibilidad parcial, la inversión en I+D es tenida en cuenta en términos del comportamiento racional maximizador de los beneficios de la parte de las empresas capaces de disfrutar de cuasi-rentas.

La idea de los modelos del capital humano es que la educación genera una externalidad positiva que los individuos no tienen en cuenta. Los agentes invierten en capital humano para aumentar sus ingresos futuros, y en ese sentido existe una parte del valor de la

---

<sup>8</sup> Aunque el trabajo de Lucas (1988) es el que ha marcado el renacimiento del interés por el capital humano en la teoría del crecimiento endógeno, no sería justo olvidar trabajos anteriores como los de Schultz (1960, 1961) y Uzawa (1965). En el primero se justifica la importancia cuantitativa de la inversión en la fuerza laboral para el caso de los EEUU, mientras el segundo es el pionero en incorporar el capital humano en los modelos formales de crecimiento, añadiendo un sector educacional al modelo de crecimiento óptimo.

educación que puede ser apropiada por el agente, proporcionando una justificación para la inversión privada en formación. Sin embargo, como consecuencia de que la educación tiene un efecto positivo sobre el planificador que el individuo no tiene en cuenta, existirá una subinversión.

El trabajo de Azariadis y Drazen (1990) presenta unas ideas similares a las de Lucas. Al igual que en Lucas, parte de los beneficios de la educación no son directamente apropiables, pero en Azariadis y Drazen se destaca principalmente las externalidades entre generaciones, es decir, la producción de conocimientos es un proceso acumulativo en el que las nuevas generaciones aprovechan los esfuerzos de las anteriores.

El trabajo de Grossman y Helpman (1991b) considera el capital humano como la acumulación de esfuerzos dedicado a la escolarización y la preparación. Las habilidades que un individuo adquiere son aplicadas a unas tecnologías de producción que pueden ser mejoradas, por lo que el valor del capital humano continuará incrementando en el tiempo. Según Grossman y Helpman una manera de pensar en el capital humano de un lugar es considerar el tamaño de la fuerza laboral efectiva que posee. A mayor empleo, la economía tendrá más industria o más I+D o ambas. El inconveniente que señalan los propios autores es que en tal modelo se estaría diciendo que cuanto mayor es la economía, siempre se crece más rápido, lo cual no siempre es verdad. Para reconciliar su argumentación inicial con el hecho de que un mayor tamaño no siempre produce mayor crecimiento, ponen el ejemplo de que el resultado dependerá de la proporción de mano de obra con buena preparación y del potencial para mejorar la tecnología. Así, una economía grande que presente una población en su mayoría no preparada puede crecer más lentamente que otra con una población menor. El país con abundante mano de obra que se haya especializado en la producción intensiva en empleo puede llevar a cabo menos investigación industrial que un país pequeño con una ventaja comparativa en I+D.

Los modelos de I+D y de capital humano coinciden en la conclusión de que la investigación producida por el sector de investigación será subóptima ya que los beneficios que generan están disponibles para todos. Esta conclusión estaría sugiriendo un posible papel para el gobierno en cuanto a subsidiar la investigación. Sin embargo, el modelo de

Romer va más allá en sus implicaciones de política económica. Dado que la investigación se explica en términos del comportamiento maximizador de beneficios y supone avanzar parte de los beneficios futuros, entonces surge un claro papel para el tipo de interés.

Una vez vistos los tres tipos de modelos, decir que hay dos conclusiones idénticas en todos ellos: en primer lugar, que el nivel de inversión y de educación en equilibrio será inferior al óptimo porque los agentes no tienen en cuenta la contribución de su inversión y su educación al aumento de la productividad global de la economía; y en segundo lugar, la participación de cada factor tradicional en el producto nacional ya no es igual a la elasticidad del output con respecto al mismo, pues el progreso tecnológico hará aumentar las contribuciones de ciertos factores como el capital privado. Sin embargo, existe una diferencia importante entre los mismos. El progreso técnico y la acumulación de capital son típicamente actividades complementarias en el primer grupo de modelos (inversión en capital privado), mientras que son sustitutos en el grupo de I+D y capital humano. En este segundo grupo, la adquisición de conocimientos compite con la acumulación de capital por los recursos de que la empresa dispone, de forma que el diseño de una política pública de desarrollo no es evidente, ya que hay que tener en cuenta qué tipo de inversión resulta más productiva y los efectos que distintas políticas pueden tener sobre la asignación de recursos entre sectores. Por el contrario, en el primer grupo, toda política pública se debe basar en crear incentivos para el ahorro y la inversión que permitan acelerar la formación de capital.<sup>9</sup>

Existen otros autores que han destacado empíricamente estas nuevas fuentes de crecimiento en trabajos recientes.<sup>10</sup> Coe y Helpman (1995) muestran que los stocks de capital de conocimiento doméstico y extranjero, es decir, gastos acumulados en I+D por un país y por

---

<sup>9</sup> Muchas de las ideas básicas de la teoría del crecimiento endógeno están presentes, como se ha comentado anteriormente, en el trabajo de Marshall, así como en el de Fisher (1969) y Young (1928). Sin embargo, no se había dado un tratamiento modelizador y sistemático al estudio de las externalidades y del progreso técnico. Este aspecto es la novedad de los modelos que se desarrollan en los años ochenta y que han sido considerados la base de la teoría del crecimiento endógeno.

<sup>10</sup> Parte de la evidencia directa más convincente en favor de la consideración de la innovación industrial como el motor del crecimiento ha sido ofrecida por historiadores económicos. Landes (1969) describe el papel que las nuevas tecnologías jugaron en la revolución industrial; Rosenberg (1972) ofrece una revisión de la relación entre avances tecnológicos y crecimiento económico en EEUU desde 1800.

sus aliados comerciales, explican el crecimiento en la productividad total de los factores en los países de la OCDE. Eaton y Kortum (1993) obtiene que el número de científicos nacionales y de ingenieros entran de forma significativa a formar parte de la determinación del nivel de renta de un país, mientras que Lichtenberg (1992) obtienen el mismo resultado con el nivel de gasto en I+D. Caballero y Jaffe (1993) han empezado la tarea de estimar un modelo de equilibrio general completo de crecimiento basado en la innovación, para observar como el modelo puede tener en cuenta las tendencias en la productividad agregada y el crecimiento de consumo en los EEUU.

### 3.3.1.3 Aspectos concretos de las economías externas

Vistas las ideas principales sobre las externalidades, podemos concluir sobre las mismas que son un *bien intangible* ya que no pueden adquirirse en el mercado, del que se benefician únicamente las empresas localizadas en el territorio, las cuales participan en la creación de las propias externalidades. Un ejemplo claro se encuentra en el conocimiento y la información que hace que los productos o los procesos de producción mejoren en relación a lo que supondría la utilización de técnicas antiguas. Estos conocimientos son difícilmente separables de las organizaciones o de los individuos que los incorporan.

Asimismo, las externalidades son un *bien colectivo*, ya que la presencia de las mismas supone una ventaja comparativa únicamente para las empresas que se encuentran ubicadas en ese área (Kindelberger, 1983; Bellandi, 1985). De este modo, las empresas pueden obtener una mejora de la productividad por la proximidad a otras empresas y la facilidad de transmisión de conocimientos, bienes y mano de obra entre unas y otras. Se puede decir que no existen barreras de entrada a la información para aquellas empresas que ya se encuentran allí ubicadas pero de las que no podrían aprovecharse las no localizadas en el territorio. Es por eso que se conoce a este tipo de bienes como colectivos y no bienes públicos, ya que afectan a un grupo de empresas determinado y no a toda la sociedad por igual.

Además, las innovaciones tecnológicas no pueden adquirirse en su totalidad en el mercado, ya que parte de sus componentes son de carácter tácito y no pueden extraerse de las

empresas en las que se ha adquirido un mayor aprovechamiento de las nuevas tecnologías y del aprendizaje por la experiencia.

Tal y como se ha visto, la presencia de las externalidades asociadas a la difusión tecnológica y del conocimiento (vía capital físico, inversión en I+D o inversión en capital humano) nos permiten considerar una economía con rendimientos crecientes que son externos a las firmas pero internos a la industria en su globalidad. De esta manera, la inclusión del conocimiento agregado como un input más en la producción provoca no-convexidades, llevándonos a un crecimiento en el largo plazo que de otra manera sería imposible. En este sentido, las externalidades tecnológicas son consideradas externalidades no pecuniarias ya que, a pesar de que permiten aumentar los rendimientos de las empresas obteniendo ventajas absolutas frente a la competencia global, no están basadas en los menores costes de producción sino en el aprovechamiento de un entorno que las convierte en impulsoras del cambio productivo y del crecimiento. Por el contrario, las economías externas asociadas a la disponibilidad de un mercado de trabajo y de inputs especializados son externalidades pecuniarias ya que el beneficio que las empresas obtienen de ellas son mejoras en los costes de producción.

Aparte de las externalidades tecnológicas, las externalidades marshallianas surgen como consecuencia de otras razones, que ya se han apuntado anteriormente: la existencia de un mercado de trabajo y de output especializado y los vínculos hacia delante y hacia atrás. Es lo que Scitovsky (1954) ha considerado como externalidades pecuniarias ya que se traducen en mejoras de costes para las empresas. Por una parte, la existencia de un mercado de trabajo especializado facilita la adaptación de la empresa a su ciclo productivo. Por otra parte, los vínculos hacia delante y hacia atrás con otras empresas que adquieren la producción u ofrecen inputs especializados, así como la presencia en la cantidad y variedad de empresas que suministran bienes no comercializables generan mejoras en los costes. Las externalidades de Marshall resultaban ser una mezcla de externalidades pecuniarias y tecnológicas (o no pecuniarias).

Son varios los autores (Tirole, 1990; Glaeser *et al.*, 1992) que han considerado las influencias que diferentes estructuras de mercado suponen sobre los procesos de innovación de las empresas. En su análisis, Tirole no obtiene un resultado concluyente dado que si bien un monopolio tendría más incentivos para innovar y así mantener su dominio absoluto en el mercado, existe en contrapartida, un efecto de reemplazo por el que la ausencia de competencia no permite la existencia de una lucha por obtener las tecnologías más novedosas. Los resultados que obtienen Glaeser *et al.* señalan que la innovación surge por el aumento de la presión a que se ven sometidas las empresas para adoptar las últimas técnicas de producción que les permiten continuar compitiendo con el resto de empresas. En consecuencia, se le da mayor importancia a la presión como estímulo a la innovación que a la posibilidad de internalización de los beneficios del proceso innovador. La cerámica y las industrias de joyas de oro de Italia son claros ejemplos de industrias obligadas a innovar si no quieren ver cerrado su negocio.<sup>11</sup>

Existe, no obstante, una idea central para Glaeser *et al.* que es común a todo tipo de externalidades: el hecho de que gran parte de las economías externas tienen lugar en las ciudades. "La aglomeración de individuos, oficios e industrias en territorios cercanos crea un ambiente en el que las ideas fluyen rápidamente de persona a persona". Dado que la gente puede apropiarse de este tipo de conocimientos sin pagar por ello, estos *spillovers* tecnológicos se consideran como externalidades. De este modo, dado que la proximidad geográfica facilita la fluidez de las ideas, cabría esperar que los desbordamientos del conocimiento fueran realmente importantes en las ciudades, considerándose a este tipo de externalidades como el "motor del crecimiento" (Romer, 1986), mientras que en las zonas rurales las externalidades serían menores porque la gente interactúa menos. De Lucio (1997) muestra también que la generación de nuevas ideas y la difusión de la tecnología es más probable en aglomeraciones urbanas, mostrando la significación del grado de urbanización en la explicación del crecimiento del producto per cápita para una amplia muestra de países.

---

<sup>11</sup> Otros trabajos en los que se estudia la estructura de mercado sobre la innovación son los de Scherer (1984) y Freeman (1986).

De hecho, es esta facilidad para la transmisión de ideas y las ventajas que pueden suponer para los individuos las que explican que las ciudades puedan sobrevivir a pesar de las más altas rentas que hay que pagar en ellas. Lucas (1988) resume la idea central en la formación de las ciudades con la siguiente pregunta:

“¿Por qué iba a estar pagando la gente los alquileres de Manhattan o del centro de Chicago si no fuera para estar cerca de otra gente?”

La literatura económica reciente ha resaltado la relación entre crecimiento y comercio, en la cual el comercio sería el principal canal en la difusión de la tecnología. Según el ejemplo de Grossman y Helpman (1991b), supongamos dos países y dos sectores productores de bienes de consumo. Uno de dichos sectores produce diversos productos que pueden ser mejorados gracias a la investigación, de forma que toda la inversión en I+D se dedica exclusivamente a los productos de dicho sector, por lo que nos referimos a él como el sector de alta tecnología. El otro sector produce un bien homogéneo bajo condiciones competitivas pero sin progreso tecnológico. Es un sector tradicional de bienes de consumo. Las actividades de producción y de investigación se llevan a cabo a través de dos factores principales: capital humano y trabajadores no cualificados.

Los productos de alta tecnología experimentarán mejoras gracias a la investigación llevada a cabo tanto en su propio país como en otros países. Las empresas son capaces, por tanto, de basar sus esfuerzos investigadores en el estado de la situación, de forma que los adelantos obtenidos en otros países son utilizados por los investigadores del propio país, bien a través del trasvase de tecnología *per se* como por la tecnología que se transmite de forma intrínseca con los mismos bienes comercializables. Las empresas se benefician así de la existencia de *spillovers* tecnológicos. De esta manera, el país que consiga la mayor abundancia de empleo cualificado se especializará en términos relativos en la actividad más capital-intensiva. En el largo plazo, adquirirá posición de líder en muchas de las industrias de alta tecnología y exportará dichos bienes a cambio del producto intensivo en trabajo procedente del sector manufacturero tradicional, de forma que las dotaciones relativas de factores determinarán el modelo de comercio en el largo plazo. Es el modelo de la ventaja

comparativa dinámica. Sin embargo, lo explicado hasta ahora no es sino parte de la teoría del comercio internacional de Heckscher-Ohlin. No obstante, el modelo de Grossman y Helpman (1991b) captura otros aspectos tales como la aparición de una red de comercio intraindustrial. Cada país exporta los productos de alta tecnología de la industria en la que es líder tecnológico mientras que importa productos de aquellas industrias en las que no está tan aventajado tecnológicamente. Incluso cabe pensar que si los costes de transporte son muy elevados o si las dotaciones relativas de los factores son muy diferentes entre los dos países, las empresas con alta tecnología de los países con salarios más elevados instalarán sus propias filiales en el país extranjero (inversión extranjera directa) o alquilarán su tecnología a los productores foráneos (licencias de patentes internacionales).

Asimismo, Coe y Helpman (1995) y Helpman (1997) muestran la relevancia de los *spillovers* internacionales de las inversiones en I+D para el crecimiento. La hipótesis de dichos autores es que la productividad total de los factores en un país no depende únicamente del capital en I+D doméstico sino también del capital en I+D de otros países.<sup>12</sup> Y el conocimiento que proviene de las inversiones en I+D se encuentra en los bienes comercializables por lo que los países que importan de un país se aprovechan de las innovaciones de dicho país. De esta manera, existen dos formas en las que el comercio y la inversión contribuyen a la productividad total de los factores: en primer lugar, indirectamente poniendo a disposición productos y servicios que incorporan en sí mismos conocimiento y que permiten un uso más efectivo de los recursos existentes aumentando así la productividad de los factores del país comprador; en segundo lugar, directamente, proveyendo tecnologías y otros tipos de conocimientos de otras economías (otras regiones y países) a las que de otra manera no se hubiera podido acceder o hubieran sido muy costosas de adquirir. Coe y Helpman encuentran que la I+D extranjera para 21 países de la OCDE más Israel tiene realmente efectos beneficiosos sobre la productividad doméstica y que el beneficio es mayor cuanto más abierta está la economía al comercio internacional y cuanto más pequeño es el país.

---

<sup>12</sup> En los trabajos de Griliches (1992) y Coe y Moghadan (1993) se encuentra evidencia empírica sobre la importancia de la acumulación de inversión en I+D como determinante de la productividad.

No obstante, algunos trabajos han relativizado la importancia del comercio en la difusión de la innovación entre las economías. Engelbrecht (1997) extiende el modelo de Coe y Helpman (1995) incluyendo una variable que representa el capital humano, que recogería otros aspectos que no son tenidos en cuenta en la I+D, como el “aprender haciendo”. La conclusión que obtiene es que, aunque la I+D tanto doméstica como extranjera sigue siendo significativa, los coeficientes disminuyen. Por tanto, el capital en I+D según dicho autor no es un concepto lo suficientemente amplio como para capturar la importancia del capital humano en el proceso de crecimiento económico, si bien reconoce la existencia de un cierto solapamiento entre ambas variables dado que el capital en I+D no es más que una forma específica de capital humano asociado con la innovación. Por otra parte, Keller (1997) ha puesto en duda la validez de la idea del comercio como fuente de difusión de la tecnología para el caso de las economías nacionales, de forma que según el autor existen otros canales por los cuales las innovaciones se difunden de unas economías a otras en el interior de un país. Así, en la ecuación de crecimiento que dicho autor estima, las inversiones extranjeras en I+D siguen siendo significativas cuando están ponderadas aleatoriamente, en lugar de usar el volumen de las importaciones en los países con los que comercian como ponderaciones. La conclusión que se obtiene es, por tanto, que otros canales pueden estar influyendo en la difusión del conocimiento además del comercio de bienes.

Vistas las principales ideas de las externalidades y los modelos que las incorporan, cabe preguntarse si sus conclusiones hacen cambiar la orientación de la política económica. Si no se introducen las externalidades y se considera un modelo neoclásico, según de la Fuente (1995) sólo existe un área en la que se puede intervenir estatalmente: la provisión de ciertos tipos de capital que presentan elementos de bien público, principalmente las infraestructuras. Sin embargo, cuando se tiene en cuenta la existencia de las externalidades en el capital humano y en el progreso tecnológico, se pueden llevar a cabo políticas públicas activas que corrijan la tendencia del mercado a producir subinversión en aquellas actividades cuyos beneficios no son totalmente apropiables.

Por una parte, deben crearse políticas que incentiven dos tipos de inversión: en capital en el sentido tradicional, es decir el aumento y modernización del stock de instalaciones y

equipos productivos, y también inversión en educación, investigación y desarrollo. A nivel general, existe amplia unanimidad entre economistas y políticos de que estos dos motores de crecimiento merecen la atención pública.

Por otra parte, según algunos autores como Tirole (1990), las políticas públicas deben estar encargadas del establecimiento de un sistema de patentes para proporcionar el grado óptimo de protección del producto y de las innovaciones de proceso. Esto es así ya que, dado que las innovaciones y la tecnología se transmiten entre las empresas o industrias permitiendo un aumento de la productividad de otras empresas sin que éstas paguen ningún tipo de compensación, se puede llegar a la situación de que las empresas vean como sus incentivos a innovar se han reducido. Es decir, existe una incompleta apropiación de ideas o innovaciones. De todo ello, dichos autores basan la importancia de los derechos de propiedad y patentes y la necesidad de dedicar recursos, públicos y privados, a la inversión en I+D y su difusión.

### **3.3.2 Efectos económicos de las externalidades**

#### **3.3.2.1 Externalidades: Crecimiento y localización**

En la sección anterior se ha comprobado como las teorías más recientes del crecimiento se han centrado de forma importante en el papel de las economías externas. Asimismo, las nuevas teorías de la localización de la actividad económica han recurrido también a la aportación marshalliana de las economías externas para explicar las pautas que sigue la industria para decidir su emplazamiento (Krugman, 1991, 1992). La distinción en los efectos de las externalidades sobre ambos, la localización de la actividad económica y el crecimiento, ha hecho que apareciera en la literatura una doble clasificación de las economías externas, estáticas y dinámicas, respectivamente, con la cual no toda la literatura está de acuerdo.

Respecto a la localización de la actividad económica, en Marshall una aglomeración económica está creada en base a externalidades tecnológicas y pecuniarias.<sup>13</sup> La fuerza de

---

<sup>13</sup> Para una revisión documentada sobre las fuerzas de aglomeración, ver Fujita y Thisse (1996).

aglomeración es la existencia de comunicaciones entre las firmas, lo que permite el intercambio de información (Saxenian, 1994). Dada la característica de bien público de la información, surge una externalidad creada por el intercambio de información a través de la comunicación entre un grupo de firmas. Dado que las empresas suelen poseer diferentes tipos de información, los beneficios de la comunicación suelen aumentar con el número de firmas. Además, como la comunicación suele presentar costes a medida que aumenta la distancia, los beneficios suelen ser mayores si las empresas se localizan cerca unas de otras. Por tanto, la actividad industrial tenderá a concentrarse para facilitar el flujo de información. Sin embargo, la concentración de la actividad en un área suele suponer aumentos en los precios del terreno y la vivienda y mayores salarios. Estos efectos suelen ser elementos desincentivadores de una mayor aglomeración de empresas en esa misma área. De esta manera, la distribución en equilibrio de las empresas estará determinada por un equilibrio entre estas fuerzas opuestas, fuerza centrípeta y centrífuga.

Marshall utiliza el concepto de distrito industrial para explicar la existencia de economías externas. Así, el distrito industrial marshalliano sería aquella aglomeración de industrias que integra todas las actividades verticalmente, lo cual está ligado directamente con la formación de economías externas (Becattini, 1979). Porter (1990) considera la misma concepción de industria que Marshall, utilizando la idea de sectores verticales para calcular las ventajas competitivas. La existencia de un distrito industrial marshalliano exige que exista un tamaño de mercado lo suficientemente grande como para permitir que se lleve a cabo dicha especialización del trabajo y desintegración vertical de las actividades de producción (Stigler, 1951). De este modo, según Costa y Callejón (1995), las empresas establecerán relaciones inter e intrasectoriales pero dentro del conjunto de actividades que todo producto industrial requiere a lo largo de su proceso productivo. Por el contrario, si se estuviera ante una estructura sectorial diversificada pero inconexa, no existirían economías externas y no se estaría ante una situación de estabilidad de las industrias. De esta manera, las externalidades son importantes en la explicación de la ubicación de la actividad económica. Milner y Westaway (1993) señalan la hipótesis de que el crecimiento consecuencia de la difusión tecnológica interna será menor en países pequeños. Un tamaño pequeño del sector industrial y un elevado grado de especialización puede limitar las

oportunidades para que la tecnología pueda transferirse vía inputs industriales tales como la maquinaria y el equipamiento. Asimismo, un mercado pequeño para los bienes industriales puede suponer un aumento de la concentración y de los precios. Sin embargo, esta hipótesis cada vez resulta menos sostenible dado que las oportunidades de exportar la tecnología aumentan en una economía más abierta e interdependiente, por lo que el crecimiento y la difusión del conocimiento resultan independientes del tamaño del país.

Sin embargo, la idea de las economías externas propugnada por Marshall ha sido la base de un conjunto de interpretaciones en la teoría de la localización de la actividad que no siempre son coherentes ni entre sí ni con la idea inicial. Así, mientras algunos autores propugnan que las externalidades pecuniarias son las verdaderas explicaciones de la aglomeración de las actividades, otros consideran que la razón principal se encuentra en el conocimiento y su difusión.

Por una parte, los trabajos de Krugman y Venables (1995), Puga y Venables (1996) y Venables (1996) basan la explicación de la aglomeración de la industria y su difusión en base a la existencia de economías externas pecuniarias principalmente. En dichos modelos se suponen diferentes países con idéntica tecnología y recursos con dos sectores, agricultura e industria. Las empresas que pertenecen al sector industrial son imperfectamente competitivas y están vinculadas por relaciones input-output, que son las que crean relaciones hacia delante y hacia detrás (*forward* y *backward linkages*). Si se supone la existencia de costes de transporte, entonces la proximidad a las empresas que proveen bienes intermedios permite reducir los costes de las empresas, creando lo que dichos autores denominan como vínculos de coste (o hacia delante). La presencia de empresas que utilizan bienes intermedios hace aumentar las ventas y los beneficios de las empresas proveedoras, creando vínculos de demanda (o hacia detrás). La interacción de estas fuerzas es la que crea las externalidades pecuniarias, creando ventajas para la aglomeración de la industria. Si los vínculos fueran muy fuertes, la industria tendería a concentrarse en una única ubicación. Los salarios, por tanto, serían mayores en dicho país pero las externalidades pecuniarias compensarían los mayores costes de los salarios. En dicho modelo, si una fuerza externa hiciera aumentar el tamaño del sector industrial en relación a la agricultura, aumentarían aún más los salarios en dicho país por lo que llega un punto en

que sería ventajoso para algunas industrias el trasladarse del país 1 a otro país 2. Si este proceso continúa, las empresas en el país 2 empezarán a beneficiarse de los vínculos hacia delante y hacia detrás, hasta el punto que se alcanzará una masa crítica en dicho país, con una expansión de la industria acompañada de aumentos en los salarios. El equilibrio ahora se encuentra en dos países industrializados y con mayores salarios que en el resto de países. Así es como la industria se difunde en dichos modelos, y el proceso podría seguir con la expansión de la industria en una serie de olas. Según Puga y Venables (1996) las industrias que se trasladarán más fácilmente son las intensivas en empleo, dado que se ven más perjudicadas por los aumentos de salarios en las aglomeraciones, así como las industrias que son el principio de una cadena de producción (*upstream industries*). Las industrias que pospondrán más su traslado a otras zonas son las que están más atadas a los vínculos que se han creado en una aglomeración. En la localización de la actividad económica, por tanto, según parte de la literatura, las externalidades que más influyen son las pecuniarias, ya que en sus modelos no introducen las externalidades debidas a la difusión de la tecnología.

Sin embargo, Krugman (1992) explícitamente comenta que los desbordamientos tecnológicos no son importantes en la localización ya que no existe diferencia entre la información sobre innovaciones tecnológicas y la transmisión de conocimientos entre industrias de alta tecnología. Dicho autor considera que las dos primeras razones explicativas marshallianas de la concentración geográfica de la actividad industrial, es decir, el aprovechamiento de un mercado de trabajo especializado y conjunto y la oferta de factores específicos, tienen una influencia muy notable incluso si las externalidades tecnológicas no parecen tener importancia. Por tanto, los ejemplos americanos de grandes núcleos de alta tecnología, el Silicon Valley, la Route 128 o el Triángulo de la Investigación de Carolina del Norte, poseen un mercado de trabajo conjunto que es el que hace que la transmisión de conocimientos fluya con rapidez, mientras que el hecho de que los conocimientos implicados sean de alta tecnología no tiene importancia. Se considera la tecnología como un bien más, al igual que la construcción de coches o la confección de tejidos. De hecho, algunos sectores poco avanzados no difieren en sus características básicas de los sectores de tecnología punta. En todos ellos, según Krugman, existe un importante elemento de casualidad: por un determinado accidente histórico, cierta actividad

empieza en una región y gracias a la presencia de externalidades se va extendiendo en la misma. Por tanto, para este autor los rendimientos crecientes externos son de carácter pecuniario. Esta idea está relacionada con la destacada en Glaeser *et al.* (1992) en la que también considera que resulta difícil separar la idea de los *spillovers* tecnológicos de la del mercado de trabajo conjunto, argumentando la existencia de relaciones de intercambio entre ambos.

Sin embargo, el trabajo de Audretsch y Feldman (1993) parece estar en desacuerdo con esta última idea. Así, basándose en datos sobre la actividad innovadora en los estados americanos e industrias específicas, los autores encuentran que la propensión de la actividad innovadora a agruparse espacialmente es mayor en las industrias donde la creación de nuevo conocimiento y los *spillovers* del conocimiento son relevantes. Y de esta manera, a pesar de que tales industrias tienden a presentar una mayor concentración geográfica de la producción, la evidencia empírica muestra que la propensión de la innovación a crear *clusters* es más atribuible al papel de los *spillovers* de conocimiento y no únicamente a la concentración geográfica de la producción.

A pesar de la discrepancia en cuanto a qué tipo de economías externas son las que más afectan a la localización de la actividad económica, existe consenso en considerar que son las externalidades las fuerzas centrípetas que mantienen la aglomeración.

Respecto al crecimiento económico, parte de la literatura se refiere a las economías externas dinámicas como aquellas que influyen en el crecimiento de la actividad económica y no únicamente en la localización de la industria.

La aportación de Glaeser *et al.* (1992) respecto a las economías externas dinámicas ha tenido una gran repercusión posterior. En dicho trabajo se realiza una clasificación de las mismas teniendo en cuenta tres cuestiones principales: en primer lugar, la forma de mercado, es decir, de competencia perfecta o imperfecta; en segundo lugar, la estructura sectorial de la industria local; y por último, el grado de especialización de la región considerada.

Las externalidades de tipo MAR (Marshall-Arrow-Romer), son entendidas por Glaeser *et al.* como los desbordamientos que tienen lugar entre empresas de una misma industria, que están localizadas de forma próxima en el territorio. De este modo, la concentración de una industria permite que haya una mayor difusión del conocimiento, a través de la imitación, espionaje y del trasvase de mano de obra altamente cualificada y especializada. Asimismo, consideran que el monopolio local es mejor para el crecimiento que la competencia, dado que permite la internalización de las externalidades en el innovador, restringiendo el flujo de ideas sin compensación y permitiendo una aceleración de la innovación y del crecimiento. Según esta teoría, la falta de derechos de propiedad o de apropiación supondría una reducción de las inversiones de las empresas en nuevas tecnologías.

Las externalidades de tipo Porter (1990) comparten con las anteriores las características de que consideran que la concentración de actividades económicas similares aceleran el proceso de crecimiento consecuencia de la facilidad en la transmisión de ideas y la rapidez en el proceso de imitación. De este modo, las externalidades tendrán lugar dentro de una misma industria. Pero a diferencia del caso anterior, Porter considera que es la competencia local la que acelera la adopción de la innovación y el consecuente mayor crecimiento. La innovación surge por el aumento de la presión a que se ven sometidas las empresas para adoptar las últimas técnicas de producción que les permitan continuar compitiendo con el resto de empresas.

Las externalidades de tipo Jacobs (1969) ofrecen importantes diferencias respecto a las anteriormente analizadas al considerar que la mayor parte de las innovaciones y la mano de obra especializada se transmiten entre industrias diferentes. Por esta razón, el crecimiento se encuentra promovido por la diversidad industrial (en contraposición a la idea más extendida de considerar la especialización regional como fuente de desarrollo). La idea base parte del hecho de que las industrias necesitan unas de otras, permitiendo que el mayor contacto suponga una transmisión de las mismas innovaciones con diferentes aplicaciones según el tipo de industria de que se trate. Un ejemplo de este tipo de externalidad se da en la industria textil que está relacionada con el sector de la maquinaria industrial, las instituciones financieras, etc., de forma que una innovación en uno de estos sectores puede

aumentar la productividad de toda la industria textil. Asimismo, la competencia perfecta resulta un mejor estimulante para el crecimiento económico ya que el monopolio suprime todas las capacidades de las economías para mejorar.

En el estudio de los efectos favorecedores de un mejor resultado en la innovación tecnológica y, por tanto, del crecimiento, muchos han sido los autores que han utilizado modelos en los que se estudia la relación entre una variable endógena representativa del incremento de la productividad o de la cantidad de innovaciones y un conjunto de variables exógenas que recogerían los diferentes aspectos de la economía que pueden resultar ventajas u obstáculos para la innovación.

Glaeser *et al.* (1992) realizan un estudio empírico en el que regresan la tasa de crecimiento del empleo de las industrias de cada ciudad en función de ciertos aspectos tratados en las diferentes teorías que analizan las externalidades (tipo MAR, Porter o Jacobs): la estructura de mercado, el grado de especialización de la industria y la diversidad de la ciudad. Los resultados permiten concluir, no sin ciertas limitaciones por las medidas utilizadas, que la competencia tiene un efecto elevado sobre la productividad, que la diversidad de la ciudad afecta positivamente y que el mayor o menor grado de concentración de la industria no es un aspecto relevante a tener en cuenta. Por tanto, la evidencia empírica parece dar apoyo a las externalidades de tipo Jacobs, con resultados mixtos respecto a las de Porter y desfavorables a las de tipo MAR.

Con el mismo objetivo, Zoltan y Audretsch (1988) estiman una regresión en la que buscan una explicación de la cantidad de outputs de innovación. Sus conclusiones refuerzan parte de los resultados que se obtienen en Glaeser *et al.* al obtener un efecto negativo del grado de concentración de la industria y positivo para la diversidad y la competencia. Sin embargo, estos autores consideran otras variables potencialmente explicativas de la actividad innovadora. Así, obtienen que a mayor proporción de industrias formadas por empresas de más de 500 trabajadores, a mayores gastos in I+D, y a mayor porcentaje de empleo cualificado, mayor es la actividad innovadora. Por el contrario, la innovación industrial tiende a disminuir a medida que aumenta el nivel de sindicalismo. Respecto al efecto positivo que el mayor tamaño de las empresas en una industria puede tener sobre la

actividad innovadora, los autores aclaran que ésta tenderá a surgir de las empresas pequeñas, dados los mayores esfuerzos innovadores que se verán obligadas a hacer si quieren seguir siendo rentables.

En el trabajo de De Lucio *et al.* (1996) se aplica un análisis similar al de Glaeser *et al.* para el caso industrial español. Sus resultados corroboran la presencia de un papel negativo de la especialización (externalidades tipo MAR), un efecto positivo de la competencia (externalidades tipo Porter) y una evidencia nada clara sobre el papel de los efectos *spillovers* tecnológicos basados en la diversidad (externalidades tipo Jacobs). Asimismo, sólo encuentran evidencia de la existencia de economías de urbanización en la explicación de la diversidad de las actividades en un territorio.

Si bien esta clasificación de las externalidades, en estáticas y dinámicas, es la más ampliamente considerada, han aparecido críticas a la misma. Así, Costa y Callejón (1995) argumentan que la distinción entre economías estáticas y dinámicas no es válida ya que la localización (y el tamaño) y el crecimiento son variables que no pueden considerarse de forma separada. De este modo, ambos tipos de economías externas estarían íntimamente relacionadas, llevando una a la otra.<sup>14</sup> Asimismo, las autoras consideran que los efectos de las economías externas sobre la localización y crecimiento no son únicos sino que permiten variar la organización empresarial, dado que "en la medida que impulsan la expansión de los mercados finales e intermedios, las empresas tenderán a realizar funciones especializadas y a aumentar sus relaciones de compra".

### 3.3.2.2 Externalidades: Reducción de las disparidades

Las implicaciones de los resultados de las teorías del crecimiento endógeno implican que aquellos países con una mayor dotación de capital (manteniendo el resto de características

---

<sup>14</sup> Pueden existir externalidades que expliquen la especialización regional y la urbanización pero que no expliquen el crecimiento, como por ejemplo, la existencia de abundantes recursos naturales que resultan costosos de transportar o una localidad con una demanda local elevada (economías de urbanización). Son factores que pueden explicar la localización de la industria en un área pero que no generan por sí mismas un proceso de crecimiento, sino que es su posterior especialización o diversidad la que inducirá el crecimiento. Sin embargo, Costa y Callejón (1995) no están considerando este tipo de economías ya que en su crítica sólo se refieren a las economías externas.

idénticas) y con rendimientos crecientes tendrán una ventaja que tenderá a aumentar con el tiempo ya que para una tasa de inversión constante, la tasa de crecimiento de la renta es una función creciente del stock de capital por trabajador acumulado. Si se permiten flujos de factores entre países, puede suceder que se refuerce más la divergencia entre los mismos, ya que quizás el capital se dirija de los países pobres a los ricos, aumentando más aún las disparidades. Asimismo, el progreso tecnológico y sus implicaciones pueden conducir hacia un proceso de aumento de la divergencia. La acumulación de conocimientos no está sujeta a la ley de rendimientos marginales decrecientes que tan unánimemente está aceptada para los factores tradicionales de producción. Así, el coste de innovaciones adicionales se reduce cuanto mayor es la experiencia, por lo que la rentabilidad de la inversión podría no ser decreciente en función del conocimiento acumulado, manteniéndose las diferencias entre las tasas de crecimiento de los distintos países.

Sin embargo, las teorías del crecimiento endógeno también nos ofrecen razones para pensar que la difusión del conocimiento puede reducir las disparidades entre países. La difusión de la tecnología ha sido considerada por algunos autores (Abramovitz, 1979, 1986; Baumol, 1986; Dowrick y Nguyen, 1989; Wolff, 1991; Nelson y Wright, 1992) como la posibilidad para los países menos desarrollados de alcanzar a los más adelantados, como consecuencia de que los primeros se aprovechan de las innovaciones que han llevado a cabo los segundos, copiándolas o comprándolas, pero en ningún caso llegando a internalizar todos los costes que esa innovación ha supuesto. Dicho fenómeno se conoce como convergencia en el sentido de *catching-up*. Según el mismo, les resulta mucho más fácil a los países o industrias menos desarrollados económicamente el imitar las tecnologías ya existentes, que a los países líderes avanzar en su innovación. Por esta razón, se esperaría que existiera una tendencia a la convergencia de la renta per cápita o de los niveles de productividad en el largo plazo, de modo que los países más pobres estarían creciendo a una tasa mayor que los más ricos (Hansson y Henrekson, 1994).

Centrándonos en esta última orientación, la idea principal de la convergencia basada en el *catch-up* es la siguiente. La edad tecnológica del stock de capital en el país líder está muy próxima a su edad cronológica; por el contrario, en los países seguidores, con una productividad menor, la edad tecnológica del stock de capital es mucho mayor que su edad

cronológica pero con una potencialidad elevada para incrementar su productividad con gran rapidez al adoptar las tecnologías desarrolladas del país líder. En este sentido, dos son los factores señalados por Hansson y Henrekson (1994) como necesarios para que dicho efecto *catch-up* pueda llevarse a cabo: la existencia de un *gap* tecnológico entre los países ricos y pobres y la presencia de una cierta capacidad social en los países pobres que les permita adoptar la tecnología superior.

La explicación de la necesaria existencia de un *gap* tecnológico para que se produzca un proceso de convergencia resulta evidente. Retrocediendo hasta Gerschenkron (1952) ya se apuntaba la idea de que existe una cierta ventaja para los países pobres por su "cierto retroceso" respecto a los ricos. El autor argumenta que cuando el nivel de productividad es mucho mayor en uno o más países que en el resto, estos últimos tienen la oportunidad de embarcarse en un proceso de *catching-up* mediante la imitación de las técnicas de producción superior que utilizan las economías más avanzadas. De este modo y, según dicho autor, las economías más retrasadas tecnológicamente deberían crecer más rápido que los países líderes en tecnología.

El grado en que el proceso de *catch-up* se lleva a cabo depende de que el país que vaya a imitar la tecnología presente una cierta capacidad social, que sea lo suficientemente sofisticado o avanzado como para adoptar la tecnología superior. Se considera que la capacidad de los países pobres para beneficiarse de las ventajas de estar en la retaguardia, depende de los modelos de consumo del país y de la existencia de un cierto nivel de desarrollo infraestructural. Las economías bajo dicho nivel, con insuficientes stocks de capital físico y humano y con una escasa demanda doméstica en las áreas productivas en las que se concentran los avances tecnológicos, no son capaces de seguir un proceso de crecimiento y de convergencia con las economías desarrolladas y fracasan en la búsqueda de unos elevados niveles de productividad (Dowrick y Gemmell, 1991).

La capacidad social necesaria para converger es un concepto muy impreciso. Así, Abramovitz (1986) enumera una serie de factores que son determinantes de la capacidad social de un país: el nivel de educación o capital humano, la organización de las empresas,

la apertura a la competencia internacional, la facilidad que tienen las empresas nuevas para establecerse, el funcionamiento del mercado de trabajo, el grado de competitividad de los mercados de productos domésticos. Otros autores como Stern (1991) y Baumol (1986) destacan factores tales como una cierta competencia en la dirección de empresas y en la presencia de un cierto nivel de desarrollo infraestructural, que para algunos llega a incluir factores sociales como la honestidad, la benevolencia de la burocracia y la claridad en la definición de los derechos de propiedad (Hansson y Henrekson, 1994).

Dowrick y Gemmell (1991) destacan que los países más pobres del África Sub-Sahariana no han aprovechado las ventajas de las nuevas tecnologías en su industria a pesar de las mejoras que se han obtenido en la productividad agrícola. La razón de ello, según los autores, se debe a la complementariedad de las nuevas tecnologías con el capital humano y con las infraestructuras, que provocan unas tasas de rendimiento del capital muy bajas en los países pobres. Los países pobres no sólo deben poder aprovechar los adelantos que produce la investigación en otros países, sino que también tienen que ser capaces de satisfacer las exigencias concretas de su economía (la tasa de innovación doméstica por la que los países pueden desarrollar nuevas tecnologías adecuadas a su producción, Romer, 1990).

Recientemente se han llevado a cabo estudios que ofrecen un resultado favorable a la hipótesis de *catch-up*. Dowrick y Nguyen (1989) modelizan el *catch-up* tecnológico como una función del ratio de la productividad del empleo de un país respecto a la del país líder. El nivel de productividad de un país captura el lado de la oferta del efecto *spillover* tecnológico: a mayor *gap* tecnológico, mayor es el potencial para copiar, comprar o transferir los avances tecnológicos de las economías más desarrolladas, lo que sugiere una relación negativa entre niveles de productividad y crecimiento económico. Abramovitz (1986) y Baumol (1986) encuentran una correlación negativa entre el nivel inicial de productividad del empleo y su tasa de crecimiento en el período de 1870-1979 para 16 países desarrollados. Entre otros estudios que han encontrado evidencia en favor del efecto *catch-up*, se encuentran los de Dollar y Wolff (1988), Dowrick (1989), Dowrick y Gemmell (1991) y Hansson y Henrekson (1994).

De todo lo dicho parece concluirse que el desarrollo de la tecnología de que dispone un país resulta de vital importancia para mejorar la productividad y la competitividad de las empresas, pudiéndose llegar a traducir en aumentos del bienestar. Sin embargo, se ha de matizar el hecho de que no es el nivel de tecnología lo que supone ventajas competitivas para un sector sino la posibilidad de adquirir rápidamente los avances que surgen en la tecnología, y la capacidad para adaptarlos a las condiciones específicas de cada sector obteniendo, de este modo, el máximo provecho. Se ha de tener en cuenta, por tanto, que las mejoras en la competitividad de una economía se deben a la rapidez con que la tecnología punta utilizada en los sectores de alto nivel tecnológico se transfiere al resto de los sectores.

El modelo neoclásico también predice la convergencia entre los diferentes países. La convergencia tratada en el modelo neoclásico de crecimiento se basa en el hecho de que la tasa de crecimiento en la inversión y la tasa de crecimiento del producto per cápita son funciones decrecientes del stock de capital per cápita. De este modo, los países ricos y pobres tenderían a converger en términos de renta per cápita. Por tanto, si bien los resultados de ambas teorías, la neoclásica y la basada en el *catch-up*, son idénticos, los puntos de partida teóricos son completamente distintos. Ambas predicen la existencia de convergencia entre países pero mientras en la teoría del crecimiento neoclásico se supone que todos los países tienen acceso a la misma tecnología, en la hipótesis del *catching-up* se presupone que los países han alcanzado diferentes niveles tecnológicos y que la difusión de la tecnología de los ricos a los pobres es la principal fuerza hacia la convergencia. Por tanto, la introducción del progreso técnico permite el crecimiento sostenido en el largo plazo (conclusión imposible de alcanzar en los modelos neoclásicos) pero no hace cambiar las conclusiones respecto a la convergencia, que siguen siendo optimistas, siempre que en los modelos endógenos se considere la tecnología como un bien público en el sentido de que todos los países tienen acceso a los mismos conocimientos técnicos en el largo plazo.

En conclusión, la hipótesis de las teorías del crecimiento endógeno de que el crecimiento es posible en el largo plazo implica también que es posible el crecimiento divergente. Si el crecimiento en una economía integrada por varias regiones con interacciones entre ellas va a permitir una reducción de las disparidades depende, según Grossman y Helpman (1991b),

de las hipótesis que se hagan sobre la movilidad de los factores y sobre la difusión del conocimiento. Si el efecto de la acumulación del conocimiento en el crecimiento es puramente local y si existe cierta movilidad en los factores no aumentables, entonces el crecimiento debe diverger entre las regiones. El otro extremo vendría dado por los efectos globales de la acumulación del conocimiento. Es decir, cuando cada nueva idea tiene el mismo efecto en el output agregado en cada localización. En tal caso se obtiene la convergencia si algunos factores constantes son inmóviles.

### **3.3.3 Interdependencias sectoriales y regionales**

#### **3.3.3.1 Externalidades sectoriales o vínculos intersectoriales**

La presencia de los vínculos intersectoriales se asienta en la idea de que el dinamismo del resto de sectores influye positivamente en el output de un sector dado. Dichas relaciones intersectoriales estarían causadas por las economías externas que se han comentado previamente en este apartado. Los vínculos hacia delante y detrás, la transmisión de tecnología bien por medio del comercio de bienes o por el hecho de compartir un mismo mercado laboral, hacen que en la evolución de un sector esté influyendo, de forma importante, la evolución del resto de sectores de la economía, principalmente aquellos con los que se comparten mayores similitudes. En el trabajo de Bartelsman *et al.* (1994) se clasifican dichos vínculos intersectoriales en externalidades de demanda y de oferta. Las primeras incluirían los efectos derivados de la mayor demanda que realizan otros sectores sobre el output de un sector dado, de forma que cuanto mayor sea el crecimiento del resto de sectores, más bienes intermedios requerirán, parte de los cuales los obtendrán de dicho sector que verá así aumentar la demanda de sus bienes. Las segundas incluyen los vínculos creados por la compra de bienes intermedios a otros sectores, y que recogen el conocimiento y la tecnología que se adquiere de forma intrínseca a los bienes comprados. De esta manera, el crecimiento se transmite entre los sectores, tanto más a mayor relación comercial entre los mismos, es decir, mayores relaciones input-output.

Varios trabajos han contrastado empíricamente la existencia y el tamaño de las externalidades entre las industrias de una economía desde finales de los años ochenta. Para el caso

americano, Caballero y Lyons (1990, 1992) y Burnside (1996) han tratado de explicar la evolución de la productividad y de los rendimientos a escala de la industria no sólo en función de la variación en los propios inputs sino como respuesta a la existencia de factores externos a las unidades de producción, que sólo se internalizan a nivel agregado de la industria.

La idea principal de estos trabajos es que al utilizar datos agregados nacionales y no sectoriales se está impidiendo la identificación de las economías de escala que son externas al sector y que se internalizan a nivel nacional confundándose, en este caso, las economías de escala y las economías externas en un mismo parámetro. De esta manera, en caso de que las externalidades intersectoriales sean importantes, las economías de escala ligadas a los inputs agregados de la nación estarán sobrevaloradas. Por esta razón, conviene separar los rendimientos a escala propios de la producción de aquéllos derivados de las interrelaciones entre sectores. Por tanto, si se parte de una función de producción:

$$Y_{it} = A_{it} \cdot f(L_{it}, Kp_{it}) \quad (3.1)$$

en la que  $Y$  representa el valor añadido bruto,  $L$  el trabajo,  $Kp$  el capital privado,  $A$  una medida del progreso técnico exógeno y los subíndices  $i$  y  $t$  representan los sectores y el momento temporal, respectivamente.

Pero si estamos interesados en estimar la relevancia de las economías externas, esta misma especificación puede ser estimada introduciendo efectos externos en la producción, representados por la actividad económica en el resto de sectores ( $Y_{\rho_{it}}$ ):

$$Y_{it} = A_{it} \cdot f(L_{it}, Kp_{it}) \cdot g(Y_{\rho_{it}}) \quad (3.2)$$

En la consideración de las externalidades que afectan a la función de producción, podemos suponer que las mismas están directamente relacionadas con el valor añadido bruto del nivel de agregación superior (VAB nacional,  $Y_T$ ):

$$g(Y_{pit}) = \delta \Delta Y_T + u_{it} \quad (3.3)$$

De esta manera en la función de producción se pueden separar los rendimientos a escala que se obtienen como consecuencia de los factores que son propios de cada sector industrial de aquellos rendimientos que se deben a las interrelaciones mantenidas entre los diferentes sectores, es decir, los rendimientos que provienen de las externalidades intersectoriales.

Los distintos autores que analizan los efectos externos difieren en la manera de modelizarlos. Así, Caballero y Lyons (1992) utilizan el output agregado industrial como índice para el efecto externo, tal como se ha mostrado aquí. Sin embargo, estos mismos autores en el trabajo de 1989 y Burnside (1996) utilizan como índice de las economías externas una medida que recoge los inputs agregados de la industria.

Los trabajos que han utilizado tales especificaciones (con ciertas variaciones) en la economía americana, obtienen evidencia en favor de rendimientos crecientes a nivel agregado que se convierten en constantes cuando se desciende a nivel sectorial. A la luz de los resultados, por tanto, el hecho de ignorar las economías externas ligadas al output supone una sobreestimación de las economías de escala internas, dado que las externalidades se internalizan a nivel agregado, mezclándose con las internas. En el caso español, la contribución de Suárez (1992) pone de relieve la importancia de las economías externas, no pudiéndose rechazar la hipótesis de rendimientos constantes, pero evidenciando la presencia de fuertes externalidades, principalmente en el sector industrial, que permiten obtener rendimientos a escala agregados superiores a la unidad.

No obstante, la mayoría de los modelos empíricos que estudian el crecimiento con datos para diferentes economías nacionales o regionales no consideran la existencia de *spillovers* entre tales economías. Implícitamente se estaría asumiendo que existen externalidades entre industrias de una economía pero no entre economías, lo cual no resulta del todo plausible. Así, en caso de que existan especificidades importantes en sectores concretos (tecnologías propias, materiales concretos, etc.) se puede pensar que las externalidades en una misma

industria localizada en diferentes regiones pueden ser más importantes y más fuertes que las externalidades entre industrias de una misma región. El siguiente apartado trata tales aspectos.

### 3.3.3.2 Externalidades regionales o vínculos interregionales

El crecimiento de una región puede traspasar a menudo las fronteras que las separan del resto de regiones, pudiendo ejercer un efecto en el crecimiento de las regiones más próximas geográficamente. De hecho, una región no es más que una delimitación administrativa de manera que cuando se considera a cada región de forma aislada, los vínculos interregionales se mezclan con los exclusivos de la región atribuyendo a las regiones un efecto inadecuado. De esta manera, se puede pensar que una parte del crecimiento experimentado por una región puede ser debido a un efecto de contagio por el que las tasas de crecimiento de la economía regional son mayores cuando las regiones vecinas están creciendo a una tasa también elevada, y menores cuando los vecinos están estancados o creciendo a tasas muy bajas. La explicación de tales efectos se puede encontrar en la presencia de externalidades que provienen tanto por los vínculos hacia delante y hacia atrás con empresas en otras regiones como por las externalidades tecnológicas o pecuniarias.

Si consideramos que las externalidades cruzan las fronteras de las economías regionales que pertenecen a un mismo área económica, nos encontramos con la idea de las interacciones entre economías que propone Lucas (1993). Según el autor, se puede pensar que existen canales más directos de interacción como, por ejemplo, la existencia de mercados comunes de empleo cualificado, un similar capital privado o unas mismas normas institucionales.

En tanto que estas relaciones interregionales pueden influir en el crecimiento, los modelos que tratan de explicar el crecimiento económico deberían incluir alguna medida que tomara en cuenta explícitamente dichos vínculos. La omisión de los *spillovers* regionales en los modelos de crecimiento pueden llevar a conclusiones erróneas dado que se imputarán los

efectos externos a los factores internos regionales. La existencia de importantes vínculos interregionales tendría importantes implicaciones para la política regional. Por una parte, las acciones que se lleven a cabo para empujar el crecimiento regional en las regiones menos desarrolladas deben tener en cuenta que parte de este efecto se difundirá a las regiones contiguas. Por otra parte, la existencia de externalidades interregionales hace que las regiones individualmente escojan un nivel de inversión en I+D, en capital humano y en capital privado que es inferior al óptimo, dado que no tienen en cuenta el efecto positivo que supone en el resto de regiones. Este resultado justificaría la existencia de agencias supra-regionales que fomenten el crecimiento a través del fomento de la inversión en los factores antes comentados.

Sin embargo, la literatura que considera dichos vínculos interregionales es muy reciente y escasa, destacando las aportaciones de Barro y Sala-i-Martin (1995, cap.12), Quah (1996), Ales y Chua (1997), Ciccone (1997) y López-Bazo *et al.* (1998a). Quah, por ejemplo, muestra como una vez que se ha condicionado a los niveles en las regiones vecinas, la distribución del producto per cápita en las regiones de la UE estaría más intensamente concentrados que la distribución real. En López-Bazo *et al.* se obtiene una fuerte correlación espacial entre regiones contiguas en la UE, detectando la existencia de *clusters* espaciales de regiones que muestran valores por encima y por debajo de la media, concluyendo a favor de la concentración de la actividad económica en el territorio.

Si bien teórica y empíricamente parecen existir externalidades relevantes entre economías, dichas interacciones no se dan en el espacio de forma ilimitada. Por el contrario son varios los autores que consideran que las externalidades, y en concreto la difusión de las innovaciones, serán más importantes entre grupos o clubs de economías cerradas. Durlauf y Quah (1998) comentan que si se permitiera que grupos de economías se formaran de forma natural, de manera que las economías en un grupo interaccionaran entre ellas más que con las de fuera del grupo, entonces el promedio del capital humano (en el caso de Durlauf y Quah) al que convergen variaría en general entre los diferentes grupos.

De este modo, en el caso de las economías regionales en un área integrada, cabe esperar que el papel del comercio y de otros canales de transmisión de externalidades y de

crecimiento sean más intensos. Los vínculos hacia delante y hacia atrás entre economías regionales parecen ser más importantes que entre países, de la misma manera que la existencia de mercados comunes de output e inputs es más probable a nivel regional en un país que entre países. Y dado que dichos factores pueden contribuir a la difusión del conocimiento, parece deducirse que las externalidades serían mayores entre regiones que entre países. Tomando el ejemplo de López-Bazo *et al.* (1998b) cabe pensar que en una economía nacional, unas pocas regiones concentrarán el mayor número de laboratorios y centros de I+D, aunque el resultado de la investigación puede ser aplicado por empresas localizadas en diferentes regiones. Incluso en el caso de centros de I+D públicos, la administración buscará la difusión de los resultados de la inversión en investigación por todo el territorio y no únicamente para las empresas localizadas en las regiones en las que se lleva a cabo tal actividad.

Una razón por la que se puede pensar que las externalidades fluyen con mayor facilidad entre grupos comerciales o países es la existencia de condiciones sociales locales que juegan un papel importante en la manera en que cada economía incorpora y adapta las innovaciones (Rodríguez-Pose, 1998). Por tanto, si las regiones contiguas comparten condiciones locales similares, las transferencias de conocimiento entre ellas pueden ser más intensas.

Parece lógico pensar que a mayor proximidad entre las regiones, mayor probabilidad de que las mismas compartan condiciones locales similares, bien porque pertenezcan a un mismo país o a un área comercial integrada. Sin embargo, ¿existen barreras geográficas para las externalidades?, es decir, aparte de la similitud entre regiones, ¿se reducen las externalidades cuando aumenta la distancia entre las regiones? Sin contestar directamente a dichas preguntas, varios autores han señalado la importancia de la proximidad geográfica para la transmisión de innovaciones (Henderson, 1992; Glaeser *et al.*, 1992). Sin embargo, parece ser que la distancia sería más importante para las externalidades pecuniarias que para la difusión del conocimiento y la tecnología. La proximidad permite, asimismo, el que las regiones contiguas puedan compartir los mercados de trabajo y de bienes finales.

Asimismo, las regiones vecinas a regiones con economías saturadas pueden aprovecharse de tal contigüidad. Así, podría ser beneficioso para algunos proveedores el localizarse en una región vecina con un bajo nivel de aglomeración (con salarios y precios del terreno relativamente bajos) y beneficiándose de la ventaja de la proximidad, siempre que los costes de transporte no sean elevados (Puga y Venables, 1996). Mori (1995) obtiene un resultado comparable en un modelo en el que las empresas están dispuestas a localizarse lejos de las ciudades dados los menores salarios a pagar en áreas agrícolas y porque los costes de aprovisionamiento de los bienes industriales son bajos y los trabajadores están dispuestos a abandonar la ciudad dados los menores costes en las zonas urbanas. De esta manera, nuevas ciudades pueden llegar a crearse si es que la población es lo suficientemente elevada, obteniéndose un modelo de actividad económica más dispersa. Krugman y Venables (1995a) también predicen el colapso de la estructura de centro-periferia y la convergencia de las regiones cuando los costes de transporte son bajos.

Antes de concluir, resaltar que si bien a nivel teórico y empírico resulta clara la existencia de vínculos interregionales e intersectoriales, ¿cuáles son más fuertes? Costello (1993) muestra como el crecimiento de la productividad total de los factores está más correlacionada entre industrias en un país que entre países considerando tan sólo una industria. Sin embargo, Kollmann (1995) obtiene que el crecimiento de la productividad está más intensamente correlacionado entre las regiones de EEUU que entre los países del Grupo de los Siete. En el trabajo de López-Bazo *et al.* (1998b), analizando las correlaciones entre los crecimientos del residuo de Solow en las distintas regiones españolas y en 12 sectores industriales, muestran como las correlaciones entre regiones son tan importantes como entre sectores. Parece haber, por tanto, una evidencia mixta en referencia a si las externalidades entre industrias de una misma región serían más fuertes o más débiles que a través de regiones en una misma industria.

### 3.4 LA ECONOMETRÍA ESPACIAL COMO HERRAMIENTA EN EL ESTUDIO DE LAS EXTERNALIDADES

#### 3.4.1 Introducción

Recientemente, como consecuencia de la aparición de los modelos de crecimiento endógeno y la nueva geografía económica de Krugman (1991), se ha puesto de manifiesto la necesidad de incorporar las relaciones entre las unidades económicas, entendiéndose éstas en un sentido amplio que incluiría las regiones o las ciudades. Los conceptos de externalidades y efectos *spillover* hacen que el tratamiento actual que se ha dado en economía regional de considerar los recursos y factores propios de un área como los más explicativos del crecimiento del mismo no sea del todo adecuado. Asimismo, en parte importante de los trabajos de economía aplicada, el problema de la escasa correspondencia entre los objetivos de estudio y la arbitraria delimitación administrativa de las unidades de observación hace que el ámbito de los efectos de la variable objeto de estudio no coincida con las unidades espaciales que se están considerando, desbordando dicho efecto las fronteras de las mismas. Además, se ha mostrado que la actividad económica no está distribuida en el espacio de forma aleatoria, sino que aparecen *clusters* en los que la actividad es elevada y por otra parte, existen verdaderos desiertos económicos en los que el desarrollo económico es mínimo. Parece ser, por tanto, que el espacio y las relaciones entre regiones y ciudades es importante, y así lo demuestra la aparición de una serie de trabajos que intentan modelizar dichas relaciones tanto interregionales como intersectoriales.

Sin embargo, ha sido obviado en muchos de estos trabajos que un *mismatch* entre las unidades espaciales de observación y el ámbito espacial del fenómeno bajo consideración supone la aparición de problemas de autocorrelación espacial (Anselin, 1988a). De este modo, en economía aplicada regional se suelen utilizar datos transversales agregados que se caracterizan por la presencia de los llamados efectos espaciales, la dependencia o autocorrelación espacial y la heterogeneidad o estructura espacial. En el primer caso se está haciendo referencia a que el valor de una variable en un lugar del espacio está interrelacionado con el valor de la misma en otro u otros lugares, mientras que el segundo

efecto surge cuando utilizamos regiones extremas para explicar un mismo fenómeno económico. Dichos efectos son importantes en el análisis econométrico aplicado ya que pueden llegar a invalidar ciertos resultados metodológicos estándar de la econometría tradicional, pedir adaptaciones en otros casos y en según qué contextos, necesitar del desarrollo de un grupo especializado de técnicas. Es en este punto en el que los economistas aplicados vieron la necesidad de tratar explícitamente los problemas causados por la autocorrelación espacial en los datos transversales usados en la implementación de los modelos econométricos regionales. Así, a principios de los años setenta, Jean Paelinck utilizó el término “econometría espacial” para denominar al campo de la econometría aplicada que trata las soluciones al problema de la presencia de efectos espaciales.

La econometría espacial se puede definir como la subdisciplina de la econometría general, surgida a finales de los años cincuenta, que proporciona las técnicas de contrastación y estimación necesarias para trabajar en presencia de efectos espaciales. La característica fundamental de un modelo econométrico espacial es que trata de explicar cierta variable, observada en una serie de puntos del espacio, a través de un conjunto de variables observadas en los mismos. Es decir, un modelo que, a pesar de que pueda utilizar datos temporales, está definido principalmente por su dimensión espacial, por lo que deberá tenerse en cuenta que estos datos son fuertemente interdependientes. A ello se refieren Cliff y Ord (1981) cuando dicen que "la propiedad básica de los datos localizados espacialmente es que el conjunto de valores  $\{x_{ij}\}$  están seguramente relacionados sobre el espacio".

Así, por ejemplo, teniendo en cuenta los temas analizados hasta este punto del trabajo, se puede pensar que la elevada interdependencia entre las variables de distintas regiones es la causa de que, a nivel regional y metropolitano, la elasticidad del crecimiento económico ante las infraestructuras sea mucho menor que a nivel nacional. En efecto, a medida que descendemos en el ámbito territorial, empieza a tomar forma la falta de independencia que suele caracterizar a los grupos de datos *cross-section*. Es en este sentido que pueden aparecer problemas derivados de la autocorrelación espacial consecuencia de la presencia de externalidades espaciales o efectos *spillover* o difusión, resultado de las arbitrarias delimitaciones de las unidades de observación (las fronteras de las regiones).

En un principio, la dependencia o autocorrelación espacial puede recordarnos a la dependencia temporal de tipo unidireccional por la que el presente puede venir explicado por el pasado pero nunca por el futuro. Sin embargo, en las muestras de datos *cross-section*, como es el caso que nos ocupa, la dependencia surge como consecuencia de una red de relaciones multidireccionales, ya que el valor de una variable en un lugar puede estar relacionada a la vez con valores de esa o de otras variables en diferentes puntos del espacio. Esta interdependencia multidireccional imposibilita, por tanto, la aplicación de las simplificaciones normales de la econometría estándar y requiere métodos alternativos de especificación, contrastación y estimación, que es el objetivo de la econometría espacial.

En esta sección se hace referencia a los conceptos básicos de la econometría espacial. Se analizan los principios que rigen dicho campo, así como la especificación formal de los efectos espaciales, prestando una mayor importancia a la problemática de la dependencia espacial. Asimismo, se analiza con detalle la especificación de la estructura de dependencia espacial, los contrastes de hipótesis que se pueden realizar para detectar dicha problemática y los métodos de estimación a utilizar en caso de que se verifique dicho fenómeno espacial.

### **3.4.2 La razón de ser de la econometría espacial. Breve reseña histórica.**

En los últimos cuarenta años, la economía regional y urbana ha experimentado un fuerte desarrollo metodológico basado en la necesidad de trabajar con la especial naturaleza de los datos de corte transversal o *cross-section*. Cuando se tratan este tipo de datos suelen aparecer los denominados *efectos espaciales*: la heterogeneidad y la dependencia espacial.

El primer efecto aparece cuando se utilizan datos de unidades espaciales muy distintas para explicar un mismo fenómeno económico. En ese caso aparecen problemas como la heteroscedasticidad o la inestabilidad estructural que pueden ser resueltos, no obstante, mediante las técnicas econométricas existentes para series temporales. La dependencia o autocorrelación espacial, sin embargo, no puede ser tratada por la econometría estándar. Es el caso en el que el valor de una variable en un lugar del espacio está relacionado con el valor de

ella misma o de otras variables en otro u otros lugares del espacio. A diferencia del contexto temporal donde las relaciones de dependencia son unidireccionales, en el contexto espacial las relaciones de interdependencia pueden ser multidireccionales (el valor de una variable en un punto viene explicado por el valor de esa variable u otras en muchos puntos). Esta interdependencia imposibilita la utilización de las técnicas econométricas estándar.

En el supuesto de tratar con la metodología econométrica tradicional un modelo que presentase algún tipo de efecto espacial, obtendríamos resultados erróneos que nos conducirían a falsas conclusiones. Por esta razón, con el objetivo de resolver los problemas que supone la presencia de efectos espaciales surgió en la década de los cincuenta la econometría espacial. La econometría espacial es una subdisciplina de la econometría general, que proporciona las técnicas de contrastación y de estimación necesarias para trabajar con datos que presentan problemas de heterogeneidad o dependencia espacial. Paelinck (1985) señala asimismo que la econometría espacial integraría los preceptos de la econometría general y las enseñanzas de la teoría económica espacial.

En realidad, el conocimiento de los problemas causados por la estructura y la dependencia espacial y sus efectos en la validez de los métodos estadísticos tradicionales puede remontarse a Student (1914). Sin embargo, los comienzos en el desarrollo de un campo separado de la estadística espacial se atribuyen a la obtención de los primeros índices formales para detectar la presencia de autocorrelación espacial en los trabajos de Moran (1948) y Geary (1954).

Tras esta fase inicial de reconocimiento del problema, es en la década de los setenta cuando surge el término econometría espacial, acuñado por Paelinck y originariamente referido a los esfuerzos realizados para abordar la autocorrelación espacial en los términos de perturbación de las regresiones. En Anselin (1988a) la econometría espacial se define más concretamente como la colección de técnicas que tratan las peculiaridades causadas por el espacio en el análisis estadístico de los modelos de ciencia regional. Según Anselin y Bera (1998) en esta última definición se estarían incorporando específicamente las regiones, la localización y la interacción espacial, formando la base necesaria de los trabajos empíricos en economía regional y urbana.

El gran desarrollo de la econometría espacial se ha dado en las décadas de los ochenta y noventa. Unos primeros trabajos analizando de manera general los aspectos metodológicos más importantes son los de Cliff y Ord (1981), Blommestein (1983) y Anselin (1980, 1988a, b). Posteriormente han ido apareciendo en las revistas de economía regional contribuciones concretas al campo de la econometría espacial, así como la aparición de varias colecciones de artículos contenidos en Anselin (1992a), Anselin y Florax (1995a) y Anselin y Rey (1997). Sin embargo, a pesar de los avances recientes en econometría espacial, persiste una casi total ignorancia a nivel genérico de los efectos espaciales en la principal corriente de la literatura econométrica y estadística. En la mayoría de los manuales de estas dos ciencias no se hace referencia al problema de los efectos espaciales y cuando se menciona, se limita a un breve apunte, como es el caso del manual de econometría de Johnston (1984) y el de Novales (1993). Este último, al hablar de la autocorrelación apunta que "aunque esta propiedad está naturalmente asociada a los datos de series temporales, puede presentarse en muestras de sección cruzada una situación similar: si la ordenación de las observaciones en la muestra sigue unos criterios geográficos, como pueden ser las distintas áreas de un país, o las familias de un vecindario, entonces los términos de error correspondientes a observaciones sucesivas pueden estar correlacionados entre sí; esta situación se denomina autocorrelación espacial, para la que se han desarrollado algunos procedimientos específicos".

Esta falta de consideración de los efectos espaciales se ve reflejada en que las técnicas espaciales (contrastes y métodos de estimación) no se encuentran presentes en los paquetes estadístico-econométricos más utilizados, por lo que quien ha querido utilizarlos se ha visto obligado a implementarlos previamente. Sin embargo, a principios de los años noventa, Anselin (1992b) puso a disposición de los economistas aplicados el programa SpaceStat en el que están implementados, bajo el entorno Gauss, los diferentes contrastes de detección de autocorrelación espacial en los modelos de regresión así como los métodos de estimación que incorporan dicha problemática. La aparición de dicho paquete econométrico ha facilitado en gran medida la aplicación de dichos métodos por lo que han empezado a surgir trabajos

aplicados que analizan la problemática de los datos transversales no sólo en el campo de la economía regional y urbana.<sup>15</sup>

### 3.4.3 Principios básicos de econometría espacial

Paelinck y Klaessen (1979) destacan cinco principios básicos en el campo de la econometría espacial y el tratamiento de datos *cross-section* en general:

- 1) *Interdependencia*: cualquier modelo espacial es interdependiente, es decir, las variables endógenas (económicas, sociales, demográficas, etc.) observadas en varios lugares del espacio geográfico están relacionadas entre sí. Tanto si se trata de fenómenos de consumo y de producción como de inversión, las influencias espaciales son incalculables. Por ejemplo, se puede pensar en la interdependencia espacial de gastos y de ingresos por la que unos mayores ingresos en un área pueden conllevar mayores gastos no sólo en ese área sino también en áreas vecinas por lo que dichas regiones experimentarán un crecimiento económico que reportará mayores ingresos en aquellas áreas que en un principio no los tenían.
- 2) *Asimetría*: las relaciones espaciales son, en principio asimétricas. Basta con pensar en los comportamientos de compra en donde existe una jerarquía de centros, lo que implica que la probabilidad de que un residente de las afueras de la ciudad vaya a comprar al centro es superior a la probabilidad de que un residente en la ciudad vaya a comprar a las afueras.
- 3) *Alotopía*: se ha de buscar a priori "la causa" de un fenómeno espacial en otro lugar. Un ejemplo clásico es el triángulo de Weber en el que la solución de localización óptima está en el interior del mismo y no en los vértices, en los que se encuentran los mercados de productos y de factores. De la misma manera, los fenómenos migratorios se explican por la comparación de ventajas y de inconvenientes no sólo

---

<sup>15</sup> Entre los trabajos aplicados en economía regional y urbana que tienen en cuenta la autocorrelación espacial se encuentran los de Case *et al.* (1993), Holtz-Eakin (1994), Molho (1995), Chesire y Carbonaro (1995), Anselin *et al.* (1997), Vayá *et al.* (1998) y López-Bazo *et al.* (1998b).

en los espacios de origen y de destino (disponibilidad de comercios, calidad de residencia) sino en causas que se encuentren en otros lugares del espacio (nivel de salarios, por ejemplo).

- 4) *No linealidad*: la no linealidad de soluciones espaciales óptimas *ex-ante* conduce a modelos econométricos *ex-post* que requieren una atención particular en lo que respecta a su especificación; además, esta última deberá ser no lineal, en general.
- 5) *Inclusión de variables topológicas*: dado que la vida económica se desarrolla necesariamente en el espacio geográfico, un modelo espacial debe incorporar variables topológicas: coordenadas, distancias, superficies, densidades, etc. Este aspecto ha sido olvidado frecuentemente, ya que la utilización de un índice  $r$  de región da la impresión de haber incorporado con éxito la dimensión espacial. Por tanto, el espacio bidimensional, donde posiblemente se lleven a cabo las actividades económicas, debe aparecer de manera explícita en los modelos; en otras palabras, deben ser topológicos.

Según Paelinck, no siempre será posible observar estos cinco principios de construcción de modelos espaciales y probablemente pueden haber otros además de los aquí especificados. No obstante, es importante recordar que la característica fundamental de un modelo econométrico espacial es que trata de explicar cierta variable, observada en una serie de puntos del espacio, a través de un conjunto de variables observadas en los mismos puntos. Es decir, cualquier modelo econométrico viene definido por su dimensión espacial por lo que deberá tenerse en cuenta que estos datos son fuertemente interdependientes.

### 3.4.4 Los efectos espaciales

#### 3.4.4.1 Heterogeneidad espacial

Existen dos aspectos distintos de heterogeneidad espacial: la inestabilidad estructural y la heteroscedasticidad. En el primer caso, la heterogeneidad espacial es la falta de estabilidad en

el espacio del comportamiento de la variable bajo estudio. En otras palabras, la forma funcional y los parámetros de una regresión pueden variar con la localización siendo no homogéneos en toda la muestra. Es probable que este efecto surja, por ejemplo, cuando se utilizan datos de regiones extremas para explicar un mismo fenómeno económico como pueden ser regiones ricas del norte y pobres del sur o la jerarquía de los lugares centrales. El segundo aspecto, la heteroscedasticidad, proviene de la omisión de variables u otras formas de errores de especificación que llevan a la obtención de errores de medida.

En econometría, la heterogeneidad espacial se puede tratar por medio de técnicas econométricas estándar, en concreto, a través de la consideración explícita de parámetros variantes, coeficientes aleatorios (Hildreth-Houck, 1968) y *switching regressions* (Quandt, 1958) o las técnicas de filtraje adaptativo espacial (Foster y Gorr, 1983, 1984, 1986). Sin embargo, en algunas ocasiones, el conocimiento teórico inherente a la estructura espacial de los datos puede conducirnos a procedimientos más complejos pero más eficientes como la expansión espacial de parámetros (Casetti, 1986). Por tanto, dado que este efecto puede tratarse mediante las técnicas econométricas tradicionales, centraremos nuestra atención en la autocorrelación espacial.

#### 3.4.4.2 Autocorrelación espacial

La autocorrelación o dependencia espacial se podría definir como la existencia de una relación funcional entre lo que sucede en un punto del espacio y lo que sucede en el resto de localizaciones, es decir, la falta de independencia casi siempre presente en la utilización de datos *cross-section*. De este modo, en presencia de autocorrelación espacial, el valor de la variable endógena en un punto del espacio está funcionalmente relacionada con el valor de la misma variable en otras localizaciones del sistema.

La autocorrelación espacial puede ser positiva o negativa. Es positiva cuando los objetos similares tienden a agruparse en el espacio, es decir, cuando la presencia de una determinada característica en una región favorece la aparición de esa misma característica en regiones circundantes a aquélla. En cambio, la autocorrelación negativa se presenta cuando un determinado atributo en una región perjudica la presencia del mismo en el resto de regiones

colindantes, o lo que es lo mismo, objetos distintos tienden a concentrarse en el espacio. De estos dos tipos de autocorrelación, la positiva es la más intuitiva, mientras que la negativa supondría un modelo de valores como un tablero de damas y no siempre con una fácil interpretación.<sup>16</sup> En concreto, cuando se está en presencia de autocorrelación positiva, se puede decir que la muestra contiene menos información que una similar pero sin correlación, por lo que esta pérdida de información debe de ser tenida en cuenta explícitamente en la estimación y en los contrastes.

Después de observar estas definiciones da la impresión de que la dependencia espacial sólo puede venir determinada por una noción del espacio o localización relativo, es decir, sólo atribuible al orden geográfico de los datos espaciales. Si bien esto es cierto en sentido literal, está claro que existen otras posibles ligazones tanto económicas como sociales que pueden hacer que dos regiones lejanas en el espacio sean interdependientes. Tobler (1979) deja este aspecto bien claro en la que denominó primera ley de la Geografía según la cual "*cualquier cosa está relacionada con cualquier cosa, pero las cosas cercanas están más relacionadas que las distantes*". De este modo, la distancia influye en las relaciones entre las regiones si bien no son el único criterio a considerar, por lo que todo se puede relacionar con todo.

Dos son las posibles razones que pueden crear una estructura de dependencia espacial al trabajar con datos *cross-section*: los errores de medida y la amplia variedad de fenómenos de interacción espacial.

1. Entre los errores de medida se puede hablar de:

- *La arbitraria delimitación de las unidades espaciales de observación*: en muchas ocasiones encontradas en la práctica, los datos se dan solamente a un nivel agregado por lo que suele haber poca correspondencia entre el ámbito de los efectos de la variable objeto de estudio y la delimitación arbitraria de las unidades espaciales de observación (que generalmente son dadas a nivel administrativo). Como consecuencia es probable que surjan errores de medida.

---

<sup>16</sup> En Whittle (1954) se ofrece una discusión sobre el tema.

- *Las externalidades espaciales (positivas o negativas) y los efectos spillover*: como consecuencia de la poca correspondencia entre los propósitos especiales del fenómeno bajo estudio y la delimitación de las unidades de observación, los efectos de las variables analizadas pueden desbordar las fronteras de las unidades espaciales. Consecuencia de ello, los términos de perturbación de una observación, por ejemplo, en la unidad espacial  $i$ , pueden estar relacionados con los términos de perturbación en una unidad vecina  $j$  con el consecuente problema de no esfericidad del término de perturbación.

2. El segundo factor que puede causar la dependencia espacial surge de la importancia del espacio como elemento explicativo del comportamiento humano y económico. Las teorías de interacción espacial, los procesos de difusión y las jerarquías espaciales comparten la idea de que lo observado en un punto del espacio se encuentra determinado, en parte, por lo que sucede en el resto del sistema. Expresado de esta manera:

$$Y_i = f(Y_1, Y_2, \dots, Y_R) \quad (3.4)$$

vemos como la observación de la variable  $Y$  en  $i$  está relacionada funcionalmente con la magnitud de esa misma variable en otras unidades espaciales del sistema.

Sin embargo, esta expresión no es muy útil a nivel empírico ya que es un sistema no identificable, con más parámetros ( $N^2 - N$ ) que observaciones ( $N$ ), es decir, es imposible estimar los potenciales  $N \times N$  términos de correlación directamente de los datos. Aumentar el tamaño muestral no sirve de solución ya que el número de interacciones aumenta con  $N^2$ , o dicho de otra manera, aparece un problema incidental de los parámetros. Este problema es vital en el tratamiento de la autocorrelación espacial y necesita la imposición de una estructura. En concreto, para que el problema pueda resultar tratable, es necesario imponer suficientes restricciones en la matriz de interacción espacial  $N \times N$  de forma que un número finito de parámetros que caractericen la correlación puedan ser estimados eficientemente. Existen dos aproximaciones dadas en la literatura sobre la imposición de restricciones en la interacción: la representación directa y la perspectiva de la *lattice*.

En la primera aproximación, los elementos de la matriz de correlación se modelizan directamente como funciones de distancias. En geoestadística, por ejemplo, todos los pares de localizaciones se ordenan según la distancia que los separa, y la fuerza de la correlación entre ellos se expresa como una función continua de su distancia por medio de un variograma o semi-variograma (Cressie, 1993, Cap. 2). Sin embargo, dicha perspectiva no se suele utilizar en economía aplicada dado que requiere un proceso subyacente continuo en el espacio. De este modo, las observaciones formarían una muestra de un proceso espacial continuo subyacente que no es fácil de mantener cuando los datos se refieren a regiones, por ejemplo.

Según la segunda aproximación, para cada punto de la base de datos se define un grupo de contigüidad (grupo de vecinos) relevante que consiste en las otras localizaciones que interaccionan potencialmente con él. Para cada observación  $i$ , se obtiene un grupo de localizaciones  $S_i$  que es utilizado para especificar un proceso estocástico espacial. De esta forma, la estructura de la correlación entre las observaciones no se modeliza directamente sino que proviene de la forma particular del proceso estocástico. Para ello es necesario analizar la especificación operacional del grupo de contigüidad para cada observación por medio de la denominada matriz de contactos.

#### **3.4.5 Modelización de la autocorrelación espacial. La matriz de contactos**

Uno de los aspectos más importantes en econometría espacial es el problema de cómo expresar formalmente la manera en que la estructura de dependencia espacial puede ser incorporada en el modelo de regresión. En este sentido, el rasgo característico de esta subdisciplina econométrica consiste en que la interdependencia espacial de las observaciones se introduce explícitamente.

Con carácter general, un modelo econométrico regional autocorrelacionado espacialmente puede escribirse como:

$$Y = \rho W_1 Y + X\beta + W_1 X^* \xi + \varepsilon \quad (3.5)$$

$$\varepsilon = \lambda W_2 \varepsilon + \mu$$

$$\mu \rightarrow N(0, \Omega) \quad \Omega_{ii} = h_i(z\alpha) \quad h_i > 0$$

siendo  $Y$  el vector de  $N$  observaciones de la variable endógena sobre las  $N$  observaciones en las que se ha dividido el espacio relevante,  $X$  la matriz de orden  $(N \times K)$  observaciones en esas mismas regiones para las  $K$  variables exógenas,  $X^*$  la matriz de variables explicativas excepto el término independiente  $(N \times (K-1))$ ,  $\beta$  el vector de coeficientes de las variables exógenas  $(K \times 1)$ ,  $\rho$  el coeficiente de la variable endógena retardada espacialmente  $(1 \times 1)$ ,  $\lambda$  el coeficiente de la estructura espacial autorregresiva para el término de perturbación  $\varepsilon$ ,  $\xi$  el coeficiente de autocorrelación,  $W_1$  y  $W_2$  las matrices de contacto (en ocasiones coincidentes),  $\varepsilon$  el vector de  $N$  perturbaciones aleatorias, autocorrelacionado espacialmente y  $\mu$  el vector de perturbaciones aleatorias con distribución  $N(0, \Omega)$ .

El modelo reflejado en la ecuación (3.5) no difiere demasiado del modelo econométrico tradicional con datos de series temporales y perturbación autorregresiva. La característica diferenciadora del mismo procede de su dimensión espacial. Por una parte, los parámetros  $\rho$  y  $\lambda$  son parámetros de autocorrelación en la variable endógena y en el término de perturbación, respectivamente, que traducen la intensidad de las dependencias. Por otra parte, la matriz  $W$  (tanto  $W_1$  como  $W_2$ ) es la denominada matriz de contactos, cuya función es indicar, para cada región, el conjunto de regiones con las que ésta es interdependiente.

### 3.4.5.1 La matriz de contactos

La matriz de contactos es la piedra angular de la econometría espacial, dado que a través de ella se representa la potencial interacción entre las diferentes unidades espaciales. De ahí que la determinación de la correcta especificación de los elementos de la misma,  $w_{ij}$ , sea uno de los aspectos metodológicos más difíciles y controvertidos. Una matriz de contactos es una matriz  $W$  de tamaño  $N \times N$  en la que para cada observación (fila) se representan aquellas

localizaciones (columnas) que están correlacionadas espacialmente, como elementos diferentes de cero. Generalmente,  $w_{ij}=1$  cuando  $i$  y  $j$  están correlacionados y  $w_{ij}=0$  cuando no lo están, aunque en caso de conocerse la magnitud concreta de las correlaciones entre las observaciones, éstas pueden aplicarse directamente. Por convención, los elementos de la diagonal son ceros. A pesar de que existe un elevado grado de arbitrariedad en la especificación de la matriz de contactos, dos son los criterios más comúnmente utilizados:

1. *Criterio de contigüidad.* Da lugar a la denominada matriz de contigüidad espacial binaria en la que los elementos de la matriz,  $w_{ij}$ , tomarán el valor 0 en caso de que ambas regiones no estén en contacto físico y 1 en caso de que lo estén (contigüidad de primer orden). Esta solución, la más empleada, proviene de la propuesta de Moran (1950) de asignar un 1 al elemento  $w_{ij}$  cuando la región  $i$  recibe influencias de la  $j$  y 0 en otro caso, para lo que es necesario conocer el entramado de dependencias subyacentes entre las regiones. Cuando esto no es posible, es preferible usar el criterio de contigüidad o de contacto físico como indicador de interdependencia. En consecuencia, la matriz  $W$  será simétrica y de tipo binario, con la diagonal de ceros.

La definición de contigüidad requiere de la existencia de un mapa en el cual se puedan observar las fronteras comunes. De todas maneras, el concepto de frontera común puede considerarse de diferentes maneras: lado común, vértice común o combinación de ambos conceptos. Asimismo, en el caso de que las unidades espaciales sean puntos (por ejemplo, ciudades que pueden estar regular o irregularmente esparcidas en el sistema) el significado de contigüidad puede consistir en la noción del camino más corto en una red por el que se considera que dos puntos de dicha red son contiguos si están en el interior de una distancia máxima dada, por ejemplo,  $w_{ij}=1$  si  $d_{ij} \leq \delta$ , donde  $d_{ij}$  es la distancia entre las unidades  $i$  y  $j$ , y  $\delta$  es un valor de corte de la distancia.

2. *Criterio de potencial interacción.* Da lugar a la matriz general de contactos que no surge de la mera observación de un mapa físico sino que incluye una medida general de la potencial interacción entre dos unidades espaciales. Han aparecido diferentes sugerencias sobre la manera de especificar los elementos de la matriz general de contactos:

- i) *Matriz de Cliff-Ord* (Cliff y Ord, 1973, 1981). Proponen la utilización de medidas de distancia como la inversa de la distancia en dos unidades espaciales  $i$  y  $j$ , o exponenciales negativos de la distancia, y la longitud relativa de frontera común entre dos unidades espaciales en el sentido de la proporción de la frontera interior de la unidad  $i$  que está en contacto con  $j$  respecto al total de su extensión-longitud fronteriza, es decir:

$$w_{ij} = [d_{ij}]^{-a} [\beta_{ij}]^b$$

siendo  $d_{ij}$  la distancia entre  $i$  y  $j$ ,  $\beta_{ij}$  la proporción de la frontera interior de  $i$  que está en contacto con  $j$ , y  $a$  y  $b$  parámetros a estimar. Sin embargo, los parámetros de la distancia suelen ser dados a priori y no estimados conjuntamente con los otros coeficientes en el modelo, ya que en este segundo caso la especificación se complica con la presencia de no linealidades (Anselin, 1980, Cap.8).

- ii) Siguiendo la misma idea, se ha sugerido introducir el área relativa de las unidades espaciales, tal que

$$w_{ij} = d_{ij} \alpha_i \beta_{ij}$$

siendo  $\alpha_i$  la proporción de la unidad  $i$  en el área total de todas las unidades espaciales del sistema.

En estos dos casos ( $i$  y  $ii$ ), las matrices de contactos obtenidas serán asimétricas. Estas matrices, no obstante, aun siendo más representativas de las relaciones de interdependencia, no dejan de estar relacionadas con rasgos físicos de las unidades espaciales. Como sucedía con las matrices binarias de contigüidad, son poco útiles cuando las variables espaciales consisten en puntos y no tienen sentido cuando el fenómeno de interacción espacial considerado está determinado por factores tales como variables puramente económicas. Para solventar este problema surgen los siguientes tipos de matrices de contacto:

- iii) Utilización de ponderaciones con una relación más directa con el fenómeno bajo estudio: así, por ejemplo, se ha utilizado un criterio de accesibilidad general (calibrado entre 0 y 1) que combina en una función logística la influencia de varios canales de comunicación entre regiones como podrían ser las carreteras, el ferrocarril y otros medios de transporte, que formalmente presentarían:

$$w_{ij} = \sum_n K_n \{a / [1 + b \exp(-c_j d_{ij})]\}$$

siendo  $K_n$  la importancia relativa del medio de comunicación  $n$ ,  $d_{ij}$  la distancia entre dos regiones  $i$  y  $j$ , y  $n$  el número de medios de comunicación;  $a$ ,  $b$  y  $c_j$  son parámetros a estimar. El hecho de que se incorporen parámetros en el cálculo de los pesos  $w_{ij}$  es un problema importante ya que son determinados a priori separadamente del resto del análisis espacial. Esto supone que la validez de las estimaciones está condicionada a la adecuada estimación de dichos parámetros.

- iv) *Matrices basadas en distancias económicas.* En Case *et al.* (1993) se sugiere, por ejemplo, la utilización de contactos entre dos unidades basados en la forma  $w_{ij} = 1/|x_i - x_j|$ , donde  $x_i$  y  $x_j$  son observaciones de características socioeconómicas, tales como la renta per cápita o el porcentaje de población en un grupo étnico. De esta forma, y siguiendo con el primer ejemplo, se estaría suponiendo que a mayor diferencia en los niveles de renta de dos unidades de observación, las relaciones entre las mismas son menores y viceversa.

En muchas ocasiones, la matriz que se utiliza se estandariza de tal manera que los elementos de cada fila sumen uno, es decir,  $w_{ij}^s = w_{ij} / \sum_j w_{ij}$ . A pesar de que no existe una exigencia matemática o estadística para ello, en muchos casos facilita la interpretación de los coeficientes del modelo, haciendo que todos los pesos estén entre 0 y 1. Asimismo, también asegura que los parámetros espaciales en muchos procesos estocásticos espaciales sean comparables entre los distintos modelos, lo cual está relacionado con las restricciones impuestas en el marco de estimación máximo-verosímil. Así, para que se pueda utilizar la estimación verosímil, los parámetros autorregresivos espaciales tienen que encontrarse en el

intervalo  $1/v_{p_{\min}}$  y  $1/v_{p_{\max}}$ , siendo  $v_{p_{\min}}$  y  $v_{p_{\max}}$  respectivamente el menor y el mayor valor propio de la matriz de contactos  $W$  (Anselin, 1982a). Dado que para una matriz de contactos estandarizada por filas, el mayor valor propio es siempre +1, esto facilita la interpretación del coeficiente autorregresivo como una correlación. Un efecto consecuencia de la estandarización es que la matriz resultante se convierte en asimétrica (no necesariamente, pero así sucede casi siempre dado que  $\sum_j w_{ij} \neq \sum_i w_{ji}$ ) aunque la matriz original lo fuera, complicando los cálculos de algunos estadísticos y estimadores. De todas maneras, el tema de las diferentes implicaciones que una u otra matriz de contactos podrían suponer en el poder de los contrastes de autocorrelación ha sido obviado en la literatura econométrica por lo que queda abierto a nuevas interpretaciones y conclusiones.

Independientemente del tipo de matriz de contactos que se utilice, el proceso espacial resultante debe satisfacer ciertas condiciones de regularidad de forma que se pueda utilizar la teoría asintótica para obtener las propiedades de los estimadores y los estadísticos. Esto requiere restricciones en el rango de interacción y/o el grado de heterogeneidad implicado por las matrices de contacto (lo que se conoce como condiciones mixtas, Anselin, 1988a, Cap. 5). Así, los pesos deben ser no-negativos, finitos y deben corresponder a una métrica adecuada. Con algunas matrices esto puede suponer un problema. Así, por ejemplo, resulta problemático si se considera una matriz con sentido económico en la que suceda que  $x_i = x_j$  para algunos pares de observaciones, por ejemplo, cuando dos regiones tienen exactamente la misma renta o un mismo porcentaje de personas de un grupo determinado. O cuando múltiples observaciones pertenecen a una misma unidad, como varios bancos localizados en una misma región, por lo que la distancia entre ellos debe establecerse en algo diferente a cero.

En la especificación y estimación de la estructura autorregresiva espacial nos vemos enfrentados, por tanto, a la elección de diversas matrices de contacto para describir la estructura de dependencia. Dicha estructura se suele dar a priori por lo que su elección no se incorpora explícitamente en el análisis estadístico. En consecuencia, la validez de cualquier estimación o conclusión obtenida utilizando las técnicas de econometría espacial, se encuentra limitada al hecho de que la estructura espacial incorporada sea correcta. En este sentido, a pesar de su importancia, no existe acuerdo en cuanto al tipo de matriz que debe utilizarse. No obstante, el criterio más utilizado parte de la idea de con qué objetivo se va a

usar dicha matriz de contactos. En este sentido, si únicamente la necesitamos para detectar la existencia de autocorrelación espacial a través de tests de hipótesis, la matriz de contigüidad binaria será suficiente. Esto es así ya que dado que la hipótesis nula es de independencia espacial (no autocorrelación espacial), la matriz de contactos estará relacionada con las diferentes hipótesis alternativas de dependencia espacial. En este caso, aunque estuviéramos ante una matriz incorrectamente especificada, una interpretación conservadora del rechazo de la hipótesis de dependencia espacial sólo implicaría la falta de independencia y no el tipo concreto de dependencia. A pesar de que se puede argumentar que el poder del contraste sí que se verá afectado según el tipo de matriz, la probabilidad de obtener falsas conclusiones es mucho menor que en el caso de buscar la especificación funcional del modelo. Por el contrario, cuando lo que se busca es la especificación del modelo espacial, la matriz de contactos a utilizar es la general de contactos. Con este objetivo, debe intentarse que la estructura de dependencia espacial sea incorporada a través de matrices de contacto que tengan una relación directa con los conceptos generales de la teoría de interacción espacial, tales como la accesibilidad y el potencial de interacción y no basarnos únicamente en la existencia o no de contacto físico entre dos regiones.

### 3.4.5.2 Modelización de la autocorrelación espacial

La matriz de contactos nos permite introducir la dependencia espacial a través del concepto de *retardo espacial*, que consiste en un promedio ponderado de los valores en las localizaciones vecinas. Formalmente, un retardo espacial se obtiene premultiplicando un vector de observaciones de una variable aleatoria  $Y$  por una matriz de contactos  $W$ , es decir,  $WY$ . Cada elemento de la variable retardada espacialmente es  $\sum_j w_{ij}y_j$ , es decir, la suma de los valores de  $Y$  en el conjunto de vecinos  $S_i$ , dado que  $w_{ij} = 0$  cuando  $j \notin S_i$ .<sup>17</sup>

La autocorrelación espacial se puede modelizar de dos formas diferentes:

<sup>17</sup> En el caso de que se utilice una matriz estandarizada, el retardo espacial de una variable recogería un promedio (y no la suma) de los valores de la misma en el conjunto de vecinos.

1. Como una relación funcional entre una variable  $Y$  y su asociado retardo espacial  $WY$ . De esta manera se representa la relación de dependencia multidireccional por la que el valor de una variable en un lugar del espacio depende del valor de esa misma variable en otros puntos. El caso más común es cuando la variable que se retarda espacialmente es la variable endógena de un modelo, lo que se conoce como modelo autorregresivo mixto regresivo espacial o dependencia del retardo espacial (*spatial lag dependence*), en el que, siguiendo la nomenclatura dada anteriormente,  $\lambda = 0$ ,  $\alpha = 0$ ,  $\xi = (0, \dots, 0)$ :

$$Y = \rho W_1 Y + X\beta + \varepsilon \quad (3.6)$$

siendo  $\varepsilon$  el vector de términos de perturbación ( $N \times 1$ ) con  $E[\varepsilon] = 0$  y  $E[\varepsilon\varepsilon'] = \sigma_\varepsilon^2 I$ .

Existen dos interpretaciones diferentes para el coeficiente autorregresivo espacial  $\rho$ . Por una parte, el término del retardo espacial puede estar indicando una dependencia espacial sustantiva, es decir, una medida de efectos *spillover* o difusión que aparecerían como resultado de un modelo teórico y en el que las unidades bajo consideración coinciden con las unidades espaciales en las que el fenómeno bajo estudio tiene lugar. Éste sería el ejemplo de las externalidades o efectos externos marshallianos. Sin embargo, otra situación se daría en el caso en el que la autocorrelación espacial surgiera a resultas de una incorrecta adecuación del fenómeno bajo estudio y de la escala espacial que se está considerando. Éste es el caso, por ejemplo, del trabajo de Can (1992) en el que estudia los mercados de hipotecas y de viviendas, dado que dichos mercados operan a un nivel diferente que las estadísticas de los censos, pueden aparecer problemas de autocorrelación espacial positiva. En tal caso, la inclusión de una variable endógena retardada espacialmente permitirá una interpretación correcta de la significación de las variables exógenas en el modelo, una vez que se han tenido en cuenta explícitamente los retardos espaciales.

2. A través de la especificación de un proceso de dependencia espacial en el término de perturbación,  $W\varepsilon$ , dando lugar a covarianzas de los errores no esféricas. Existen

diferentes procesos espaciales a utilizar para el término de perturbación: la especificación de un esquema de Markov parametrizado o esquema autorregresivo, un proceso de medias móviles espaciales y el modelo de componentes de los errores espaciales, siendo el primero el más utilizado a nivel empírico. La dependencia espacial en los errores (*spatial error dependence*) recogerá la posible autocorrelación espacial que no haya sido capturada por la endógena retardada, de forma que puede ser interpretada como un ruido dado que refleja la autocorrelación espacial en errores de medida o en variables que no son cruciales para el modelo<sup>18</sup>. En este modelo,  $\rho = 0$ ,  $\alpha = 0$ ,  $\xi = (0, \dots, 0)$ :<sup>19</sup>

$$Y = X\beta + (I - \lambda W_2)^{-1} \mu \quad (3.7)$$

bajo el supuesto de que  $(I - \lambda W_2)$  es invertible y  $|\lambda| < 1$  por razones de estacionariedad.

Si bien estos dos modelos son los más ampliamente considerados en econometría espacial, existen otras modelizaciones. Todas ellas se derivan de la formulación general (3.5) mediante la imposición de diferentes hipótesis sobre los parámetros:

1. Modelo cross-regresivo mixto regresivo espacial:  $\rho = 0$ ,  $\lambda = 0$ ,  $\alpha = 0$ .

$$Y = X\beta + W_1 X^* \xi + \varepsilon \quad (3.8)$$

siendo  $\varepsilon$  el término de perturbación definido anteriormente.

2. Modelo regresivo mixto regresivo espacial:  $\lambda = 0$ ,  $\alpha = 0$ , como resultado de la combinación de los modelos (3.6) y (3.8).

<sup>18</sup> Tal como se ha comentado anteriormente, el esquema autorregresivo espacial en el término de perturbación no es el único que puede especificarse, pero sí el más utilizado. Posteriormente se comentan las especificaciones básicas de los otros procesos espaciales.

<sup>19</sup> Dado que  $\varepsilon = \lambda W_2 \varepsilon + \mu$ , se obtiene que  $\varepsilon = (I - \lambda W_2)^{-1} \mu$ . Sustituyendo en  $Y = X\beta + \varepsilon$  se obtiene  $Y = X\beta + (I - \lambda W_2)^{-1} \mu$

$$Y = \rho WY + X\beta + W_1 X^* \xi + \varepsilon \quad (3.9)$$

3. Modelo autorregresivo mixto regresivo espacial con término de perturbación autorregresivo espacial:  $\lambda = 0, \xi = (0, \dots, 0)$ .

$$Y = \rho W_1 Y + X\beta + (I - \lambda W_2)^{-1} \mu \quad (3.10)$$

Desarrollando esta expresión se obtiene:

$$Y = \rho W_1 Y + \lambda W_2 Y - \lambda \rho W_2 W_1 Y + X\beta - \lambda W_2 X\beta + \mu \quad (3.11)$$

es decir, una forma ampliada de la especificación Durbin espacial, pero con un conjunto de restricciones no lineales adicionales en los parámetros. En el caso en el que  $W_1$  y  $W_2$  sean la misma matriz de contactos, los parámetros  $\rho$  y  $\lambda$  estarán identificados únicamente si al menos una variable exógena está incluida en  $X$  (aparte del término constante) y cuando las restricciones no lineales estén impuestas (Anselin, 1980), convirtiéndose el modelo en:

$$Y = (\rho + \lambda) WY - \lambda \rho W^2 Y + X\beta - \lambda W X\beta + \mu \quad (3.12)$$

En esta especificación, los coeficientes de  $WY$  y de  $W^2 Y$  no permiten una identificación separada de  $\rho$  y de  $\lambda$ , de forma que no surge un único estimador para cada parámetro a no ser que las restricciones no lineales estén impuestas. Esto complica en gran manera la estrategia de estimación para este modelo.

4. Modelo cross-regresivo mixto regresivo espacial con término de perturbación autorregresivo espacial:

$$Y = X\beta + W_1 X^* \xi + (I - \lambda W_2)^{-1} \mu \quad (3.13)$$

5. Modelo regresivo mixto regresivo espacial con término de perturbación autorregresivo espacial:

$$Y = \rho W_1 Y + X\beta + W_2 X^* \xi + (I - \lambda W_2)^{-1} \mu \quad (3.14)$$

Con respecto a los diferentes modelos de dependencia espacial se ha de decir que podríamos obtener una mayor clasificación de modelos espaciales mediante la introducción de diferentes matrices de contactos así como diferentes órdenes de contigüidad. Asimismo, ha de tenerse en cuenta que la clasificación que se ha dado se basa en la especificación de la dependencia espacial en los modelos de regresión. Sin embargo, la heterogeneidad espacial puede también aparecer por lo que debería ser incorporada en el modelo. Para ello se pueden especificar los términos apropiados en  $h(z\alpha)$  para la heteroscedasticidad.

Por otra parte, tal como se ha comentado anteriormente, si bien la estructura autorregresiva del término de perturbación es la más utilizada a nivel tanto teórico como empírico, existen otras estructuras. El proceso de medias móviles espacial se especifica como (Cliff y Ord, 1981):

$$\varepsilon = \gamma W\mu + \mu \quad (3.15)$$

donde  $\gamma$  es el coeficiente de medias móviles espacial y  $\mu$  es el término de perturbación incorrelacionado. En el mismo se está indicando la presencia de un término de error específico en cada localización,  $\mu_i$ , y una media ponderada de los errores en las localizaciones vecinas,  $W\mu$ .

Otra estructura del término de perturbación diferente a la autorregresiva es la dada por Kelejian y Robinson (1993, 1995), el modelo de componentes de los errores espaciales. En dicha estructura, el término de perturbación consiste en la suma de dos términos de error independientes, uno asociado con la unidad de observación de que se trata y otro específico de la localización:

$$\varepsilon = W\mu + \psi \quad (3.16)$$

siendo  $\mu$  y  $\psi$  términos de error independientes.

A pesar de la existencia de un gran número de posibilidades de especificación de la autocorrelación espacial, el modelo del retardo espacial y el modelo del error espacial son los que más han sido estudiados a nivel teórico. Para ellos se han diseñado la mayor parte de los contrastes de autocorrelación espacial y métodos de estimación que explícitamente los tienen en cuenta, por lo que también son los que más se han utilizado en los estudios empíricos. Es por ello que a partir de este punto nos centramos en ellos.

De hecho, existe una gran similitud entre un modelo de retardo espacial y un modelo de error espacial. Así, si partimos del modelo autorregresivo (3.7) y premultiplicamos ambos lados por  $(I-\lambda W)$ , se obtiene un modelo Durbin espacial (Anselin, 1980):

$$Y = \lambda WY + X\beta - \lambda W X\beta + \mu \quad (3.17)$$

de forma que el modelo resultante tiene una estructura de retardo espacial tanto de la variable endógena como de las explicativas, con una serie de restricciones.<sup>20</sup>

Otro aspecto interesante a tener en cuenta es que el efecto que se obtiene como consecuencia de la introducción de un retardo espacial o un error espacial se puede conseguir mediante el filtrado espacial de variables. Cuando se filtra una variable espacialmente se está quitando el efecto de la autocorrelación espacial. Por ejemplo, el modelo del retardo espacial se puede expresar como:

$$(I-\rho W)Y = X\beta + \varepsilon \quad (3.18)$$

donde  $(I-\rho W)Y$  es una variable dependiente filtrada espacialmente, es decir, se quita explícitamente el efecto de la autocorrelación espacial. De forma análoga, el modelo del error

<sup>20</sup> Para un mayor detalle de tales restricciones, véase Anselin, 1988a, pág. 226-229.

espacial se puede expresar en términos de variables filtradas espacialmente. Para ello, partimos de la expresión (3.17) y trasladando la variable del retardo espacial al lado izquierdo:

$$(I-\lambda W)Y = (I-\lambda W)X\beta + \mu \quad (3.19)$$

De esta manera se obtiene un modelo de regresión en el que las variables, tanto las explicativas como las endógenas, se han filtrado espacialmente y se obtiene un término de perturbación incorrelacionado,  $\mu$ .

La importancia de la no consideración de los efectos espaciales en los modelos de regresión que pueden presentar dicho efecto radica en que las técnicas econométricas estándar de estimación dejan de ser válidas en dicho contexto. Así, si estamos ante un modelo correlacionado espacialmente y estimamos un modelo de regresión lineal por los métodos tradicionales de mínimos cuadrados, obtendríamos los siguientes resultados:

1. En caso de estar ante un modelo de retardo espacial, los estimadores mínimo cuadráticos serían sesgados e inconsistentes. Ello es consecuencia de que la presencia del término  $WY$  en la ecuación (3.6) implica la presencia de correlación de cada uno de sus elementos  $(WY)_i$  con el término de perturbación,  $\varepsilon$ , no sólo en  $i$  sino con todos los términos de error en todas las otras localizaciones, dejando de ser consistente. Para analizarlo, partimos de la reformulación del modelo en:

$$Y = (I-\rho W)^{-1} X\beta + (I-\rho W)^{-1} \varepsilon \quad (3.20)$$

en la que la matriz inversa  $(I-\rho W)^{-1}$  es una matriz completa y no triangular como sucedía en el caso de series temporales cuando se trataba un retardo de la endógena (donde la dependencia es únicamente unidireccional). En el caso presente:

$$E[(WY)_i \varepsilon_j] = E\{[W (I-\rho W)^{-1} \varepsilon]_i \varepsilon_j\} \neq 0 \quad (3.21)$$

Por tanto, cuando una variable dependiente retardada espacialmente es ignorada en un modelo de regresión, pero está presente en el proceso generador de datos, el error de especificación que se comete es del tipo de omisión de variable relevante y, en consecuencia, los estimadores MCO en un modelo espacial serían sesgados e inconsistentes.

2. Ante un término de perturbación correlacionado espacialmente, los estimadores mínimo cuadráticos serían insesgados y consistentes pero ineficientes. Este resultado es consecuencia de la obtención de una covarianza de los errores que no sería esférica:

$$E[\varepsilon\varepsilon'] = \sigma^2(I - \lambda W)^{-1} (I - \lambda W')^{-1} = \sigma^2[(I - \lambda W)' (I - \lambda W)]^{-1} \quad (3.22)$$

De este modo, una estructura autorregresiva espacial lleva a una covarianza de los errores diferente de cero pero decreciente en magnitud con el orden de contigüidad. La varianza del término de perturbación, por tanto, podría estar seriamente infraestimada en el caso de estimar el modelo con las técnicas estándar, siendo el sesgo mayor en el caso de estar ante una autocorrelación positiva. Las varianzas de los parámetros estimados también estarían infraestimadas por lo que la fiabilidad de las estimaciones se encontraría sobrevalorada. Asimismo, la estructura de las matrices inversas en la expresión (3.22) supone una diagonal con elementos no constantes en la matriz de covarianzas de los errores, induciendo problemas de heteroscedasticidad en  $\varepsilon$ .

Teniendo en cuenta la problemática que supone la no consideración de los efectos espaciales y partiendo de la idea de que existen muchos trabajos empíricos que utilizan datos en los que a nivel teórico no parece descabellado suponer la existencia de problemas de dependencia espacial, parece resultar de vital importancia el contrastar la presencia de los mismos. En el siguiente apartado se ofrecen los principales estadísticos que han sido desarrollados a tal efecto.

### 3.4.6 Contrastación de la dependencia espacial

Para encontrar el origen de la contrastación en econometría espacial nos tenemos que remontar al contraste ad-hoc de Moran (1950). Sin embargo, la mayor parte de la literatura econométrica espacial pionera se centró en la utilización de los estadísticos basados en el principio máximo-verosímil (Brandsma y Ketellapper, 1979a, 1979b; Anselin, 1980), dentro del cual se han desarrollado la mayoría de los contrastes que han aparecido en los últimos quince años. La mayor parte de los mismos son tratados en el presente apartado. De forma genérica en todos los contrastes de autocorrelación espacial la hipótesis nula es la ausencia de dependencia espacial, en otras palabras, la especificación clásica de un modelo de regresión:

$$Y = X\beta + \varepsilon$$

en el que, siguiendo la nomenclatura dada en el apartado anterior,  $\rho = 0$ ,  $\lambda = 0$ ,  $\alpha = 0$ ,  $\xi = (0, \dots, 0)$ .

Como hipótesis alternativas, los modelos para los cuales se han diseñado la mayoría de estadísticos de autocorrelación espacial son los modelos del error espacial (3.7) y del retardo espacial (3.6).

En términos generales vamos a distinguir dos grupos diferentes de contrastes de dependencia espacial: el primero de ellos, los contrastes *ad-hoc*, incluye una serie de contrastes basados en estadísticos previamente desarrollados en un contexto temporal y que han sido adaptados al ámbito espacial; los del segundo grupo son el resultado de aplicar el principio máximo-verosímil al modelo econométrico formulado en (3.5).

#### 3.4.6.1 Contrastes ad-hoc

Desde antiguo existen tres contrastes de autocorrelación aplicables a datos transversales, el estadístico de la  $I$  de Moran (1950), el de Geary (1954) y el de Dacey (1968), denominados estadísticos  $I$ ,  $c$  y  $d$  respectivamente. La obtención de los momentos de estos estadísticos, así

como sus distribuciones probabilísticas es compleja dada la multidireccionalidad de las conexiones. Sin embargo, en el caso de la  $d$  de Dacey no es posible calcularlas por lo que los únicos contrastes factibles de utilizar en el campo espacial son la  $I$  de Moran y la  $c$  de Geary, cuyas expresiones son, respectivamente:

$$I = \frac{N}{S} \frac{\sum_{ij}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.23)$$

$$C = \frac{N-1}{2S} \frac{\sum_{ij}^N w_{ij} (x_i - x_j)^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (3.24)$$

siendo  $x_j$  la observación de la variable  $x$  en la región  $j$ -ésima,  $\bar{x}$  la media muestral de la variable  $x$ ,  $s = \sum_{ij} w_{ij}$  la suma de todos los elementos de la matriz de contactos ( $\sum_{ij}$  representa el doble sumatorio) y se le conoce como factor de estandarización,  $N$  el número de observaciones.

Con respecto a estos estadísticos se ha comprobado que la  $I$  de Moran es preferible. Desde principios de los ochenta, Cliff y Ord (1981) demostraron que la  $I$  es asintóticamente más potente que la  $c$  de Geary, conclusión que se ve corroborada por los resultados de King (1981) que demuestran que el estadístico de Moran es el contraste localmente más potente para contrastar la hipótesis nula  $H_0: \rho = 0$ , frente a alternativas del tipo  $H_A: \rho > 0$ . Posteriormente, tal como se verá en el apartado 3.4.6.3, en varios trabajos que analizan el comportamiento de los distintos tests en muestras finitas a través de técnicas de Montecarlo, la  $I$  de Moran resulta ser el más potente ante multitud de casos. Es por ello que es el más utilizado de todo el grupo de contrastes *ad-hoc*.

La manera más común de presentar el contraste de la  $I$  de Moran para el caso de autocorrelación espacial residual es la siguiente (Cliff y Ord, 1972, 1973):

$$I = \frac{N}{S} \frac{e' W e}{e' e} \quad (3.25)$$

donde  $e$  es el vector de residuos MCO del modelo especificado sin tener en cuenta ningún tipo de correlación espacial,  $W$  es la matriz de contactos de tal manera que, en el caso de encontrarnos ante una matriz de contactos previamente estandarizada tal que la suma de los elementos de cada fila sea uno, la expresión se simplifica a:

$$I = \frac{e' W e}{e' e} \quad (3.26)$$

En dicho contraste, la matriz de contactos espacial se utiliza para representar el tipo de potencial interacción que causa la dependencia. Sin embargo, la naturaleza del proceso estocástico subyacente no está especificada. En consecuencia, si bien la hipótesis nula representa la ausencia de dependencia espacial, no existe una expresión concreta para la hipótesis alternativa.

La distribución del estadístico de Moran en un contexto asintótico y bajo el supuesto de residuos MCO incorrelacionados, corresponde, a través de la adecuada transformación, a la normal estándar. Dicha transformación es la estandarización del estadístico:

$$z_I = \frac{I - E[I]}{V[I]^{1/2}} \rightarrow N(0,1) \quad (3.27)$$

siendo  $E[I]$  la media del estadístico de Moran, momento de primer orden y  $V[I]$  la varianza del mismo, momento de segundo orden. La inclusión en (3.25) de los residuos complica la expresión de los momentos que, en el caso de una matriz de contactos general no estandarizada tienen las siguientes expresiones, respectivamente:

$$E[I] = \frac{N}{S} \frac{\text{tr}(MW)}{N - K} \quad (3.28)$$

$$V[I] = E[I^2] - (E[I])^2 = \frac{\left(\frac{N}{S}\right)^2 \text{tr}(MWMW') + \text{tr}(MW)^2 + [\text{tr}(MW)]^2}{(N - K)(N - k + 2)} - (E[I])^2$$

siendo  $M$  la matriz  $I - X(X'X)^{-1} X'$ .<sup>21</sup>

A pesar de que Moran no derivó su estadístico de ningún principio básico, se ha sugerido, por una parte, la similaridad entre la  $I$  de Moran y el estadístico de autocorrelación en series temporales de Durbin y Watson (Anselin y Bera, 1998), así como el vínculo con los test del ratio de verosimilitud (Cliff y Ord, 1972).

Un test que ha sido desarrollado recientemente es el de Kelejian y Robinson (1992) que, como el test de Moran, no se basa en una especificación concreta del proceso generador del término de perturbación. Asimismo, dicho test no requiere que el modelo sea lineal o que el término de perturbación esté normalmente distribuido. Se deriva de una regresión auxiliar en la que se usan los productos cruzados de los residuos de las observaciones que están potencialmente correlacionadas espacialmente (según el esquema de la matriz de contactos) y los productos cruzados de las variables explicativas correspondientes a dichas observaciones. De forma concreta, la variable dependiente en la regresión auxiliar es:

$$C_h = e_i e_j \tag{3.29}$$

donde  $h$  es un índice para cada producto cruzado,  $e$  es el término residual y  $i, j$ , son las observaciones contiguas. Las variables explicativas en la regresión auxiliar,  $Z_h$ , están formadas por los productos cruzados de  $X_i$  y  $X_j$ . Considerando  $\gamma$  el vector de coeficientes obtenido de la estimación MCO en una regresión de  $C$  en  $Z$ , y  $\alpha$  el vector asociado de residuos, el estadístico de KR resultante es:

<sup>21</sup> En el caso de considerar la matriz de contactos estandarizada, las expresiones de los momentos del test serían las mismas pero quitando los componentes  $N/S$  y  $(N/S)^2$  respectivamente.

$$K - R = \frac{\gamma' Z' Z \gamma}{\alpha' \alpha / h_R} \quad (3.30)$$

donde  $h_R$  es el número de observaciones en el vector auxiliar (3.29). El estadístico se distribuye como una  $\chi^2$  con  $K$  grados de libertad, donde  $K$  es el número de variables explicativas en  $Z$ .<sup>22</sup>

### 3.4.6.2 Contrastes basados en el principio máximo-verosímil

Este segundo grupo de contrastes de dependencia espacial se obtienen de la aplicación del principio máximo-verosímil al modelo econométrico (3.5). Los tres tipos de tests, el contraste de Wald (W), el contraste del ratio de verosimilitud (LR) y el de los multiplicadores de Lagrange (LM)<sup>23</sup> están basados en las propiedades óptimas del estimador máximo-verosímil (MV). Formalmente, siendo  $h$  el estimador máximo-verosímil del vector de parámetros  $\theta$  la diferencia entre el estimador y el parámetro poblacional converge en distribución a una normal con media cero y varianza la inversa de la matriz de información.

$$N^{1/2} (h - \theta) \rightarrow N [0, \lim_N (I(\theta)/N)^{-1}] \quad (3.31)$$

Los contrastes basados en la función de verosimilitud (FV), al contrario de lo que sucede con el test de Moran, están rigurosamente estructurados en términos de una hipótesis nula y alternativa específicas. De hecho, lo que se contrasta es la significación de unos coeficientes determinados. En términos generales, los tres tests asintóticos están basados en diferentes medidas de la distancia que existe entre el estimador no restringido (bajo la  $H_A$ ) y el estimador que satisface la hipótesis nula, o estimador restringido. Intuitivamente, si la distancia entre estos dos resultados es demasiado grande, la restricción no podrá aceptarse por lo que la hipótesis nula será rechazada.

<sup>22</sup> Existe una conexión algebraica entre el estadístico de KR y la  $I$  de Moran (Anselin y Bera, 1998).

<sup>23</sup> Este tipo de contraste también se denomina *Rao's Score test*.

La diferencia entre uno y otro contraste estriba en la manera en que miden la distancia entre el estimador restringido y el no restringido. En el caso del test de  $W$ , el modelo que se estima es el modelo completo por lo que se usan las estimaciones no restringidas. El test de los LM utiliza el modelo estimado bajo la imposición de la hipótesis nula, es decir, la especificación más sencilla. Por último, en el test del LR se utilizan ambas estimaciones, la restringida y la no restringida. La literatura inicial sobre la contrastación en econometría espacial estuvo dominada por los contrastes de  $W$  y LR.<sup>24</sup> Sin embargo, dado que este último requiere la estimación del modelo alternativo por medio de una optimización no lineal, en seguida se observaron las ventajas de basar un test en la regresión mínimo cuadrática del modelo nulo, tal como sucede en el caso de los LM, siendo el que más se ha desarrollado en los últimos años para diversas alternativas de especificación (Anselin, 1988a, 1988b).

#### Contrastes para el modelo de autocorrelación espacial en el término de perturbación

Tal como se ha comentado anteriormente, los contrastes basados en la FV están rigurosamente estructurados en términos de una hipótesis nula y alternativa específicas. De hecho, lo que se contrasta es la significación de unos coeficientes determinados. Formalmente, en el caso de contrastar la existencia de un término de perturbación correlacionado espacialmente, las hipótesis nula y alternativa son:

$$H_0: \lambda = 0$$

$$H_A: \lambda \neq 0$$

El *contraste de Wald* se expresa en términos generales como:

$$W = [R\tilde{\theta} - r_0]' [R\theta R']^{-1} [R\tilde{\theta} - r_0] \rightarrow \chi_s^2 \quad (3.32)$$

donde  $\theta$  es el vector de parámetros estimados,  $R$  es la matriz de constantes de orden  $s(K+2)$ ,  $r_0$  es el vector de constantes de orden  $s*1$ ,  $R\tilde{\theta} - r_0$  será un vector  $s*1$  de valores que resultan de evaluar las restricciones para las estimaciones máximo verosímiles,  $(\theta)^{-1}$  es la inversa de

<sup>24</sup> Ver, por ejemplo, Brandsma y Ketellapper (1979) y Anselin (1980)

la matriz de información y  $s$  es el número de restricciones. Este estadístico se distribuye asintóticamente como una chi-cuadrado con  $s$  grados de libertad. Y dado que la raíz cuadrada del contraste corresponde a una normal estándar, es equivalente a una  $t$  de Student.

En el caso concreto del modelo con término de perturbación correlacionado espacialmente, el contraste de autocorrelación espacial residual de Wald resulta (Anselin, 1988a):

$$W = \lambda^2 [t_2 + t_3 - 1 / N(t_1)^2] \rightarrow \chi_1^2 \quad (3.33)$$

con la siguiente notación  $t_1 = \text{tr}(WB^{-1})$ ,  $t_2 = \text{tr}(WB^{-1})^2$ ,  $t_3 = \text{tr}(WB^{-1})'(WB^{-1})$ , con  $B = I - \lambda W$  y en donde se sustituye el coeficiente  $\lambda$  por su estimación MV. Alternativamente, tal como se expuso anteriormente, la raíz cuadrada de esta expresión se distribuye como una normal estándar.

*El test del LR* se basa en la diferencia entre el logaritmo de la FV del modelo no restringido y del restringido:

$$RV = 2[L(\tilde{\theta}) - L(\tilde{\theta}_R)] \rightarrow \chi_s^2 \quad (3.34)$$

siendo  $L$  los correspondientes logaritmos de la FV. Dicho test también está distribuido asintóticamente como una chi-cuadrado con  $s$  grados de libertad.

La expresión final del contraste para el modelo de autocorrelación espacial residual (Anselin, 1988a) resulta ser:

$$LR = N[L(\sigma_0^2) - L(\sigma_1^2)] + 2 \ln|I - \lambda W| \rightarrow \chi_s^2 \quad (3.35)$$

siendo  $\sigma_0^2$  la varianza residual estimada para el modelo bajo la  $H_0$  (sin autocorrelación espacial en el término de perturbación) y  $\sigma_1^2$  la varianza residual estimada para el modelo bajo la  $H_A$  (modelo espacial). Lo que diferencia dicho resultado de aquél en presencia de

correlación serial es la presencia del determinante Jacobiano en (3.35) donde queda reflejada la estructura espacial.

*El contraste de los LM* se basa solamente en la estimación bajo la hipótesis nula, por lo que resulta muy fácil de aplicar mediante los residuos MCO. Partimos de la expresión general de los contrastes de los LM:

$$LM = d'(\tilde{\theta})\mathfrak{I}(\tilde{\theta})^{-1}d(\tilde{\theta}) \quad (3.36)$$

siendo  $d(\theta) = \partial L(\theta) / \partial \theta$ ,  $\mathfrak{I}(\theta) = -E[\partial^2 L(\theta) / \partial(\theta)\partial(\theta)]$  la matriz de información,  $L(\theta)$  es la función del logaritmo de verosimilitud y  $\tilde{\theta}$  es el estimador máximo verosímil restringido (bajo la hipótesis a contrastar) del vector de parámetros  $\theta$ .

Para el caso concreto del modelo de autocorrelación en los residuos, la expresión final del contraste fue sugerida por Burridge (1980) y obtenida a partir de las primeras y segundas derivadas de la FV del modelo bajo la hipótesis nula:

$$LM = \frac{1}{T} \left( \frac{e' We}{\sigma^2} \right)^2 \rightarrow \chi^2(1) \quad (3.37)$$

donde  $T = \text{tr}\{(W+W')W\}$ . A este contraste se le nombrará *LM-ERR* para diferenciarlo de otros casos que se darán a continuación y que también se basan en los principios de los LM. Presenta una clara hipótesis alternativa que bien puede referirse a un esquema AR(1) para el término de perturbación autocorrelacionado espacialmente, o también un esquema MA(1); es decir, el test LM-ERR es igualmente óptimo para las alternativas de término autorregresivo o media móvil. Este aspecto hizo surgir la pregunta de si dicho test sería inferior a los otros tests asintóticamente equivalentes, el de  $W$  y el del  $LR$ , en lo que se refiere a poder, dado que no usa la información concreta contenida en la hipótesis alternativa. Sin embargo, los resultados de los experimentos de Montecarlo de Godfrey (1981) y Bera y McKenzie (1986) para el contexto del modelo de regresión estándar, sugieren que no existe ninguna penalización en el poder de los tests de LM en comparación con el test de  $LR$ .

De la expresión de contraste de los LM se puede deducir fácilmente una relación formal entre el estadístico de la  $I$  de Moran y el test de LM-ERR. En concreto, cuando la matriz  $W$  se encuentra estandarizada:

$$LM - ERR = \frac{(NI)^2}{T} \quad (3.38)$$

siendo  $T$  la expresión dada anteriormente y  $N$  el tamaño muestral.

Asimismo, el estadístico de los LM puede ser fácilmente generalizable para procesos espaciales de mayor orden. Si consideramos un modelo autorregresivo espacial de orden  $q$ :

$$\varepsilon = \lambda_1 W_1 \varepsilon + \lambda_2 W_2 \varepsilon + \dots + \lambda_q W_q \varepsilon + \xi \quad (3.39)$$

y se pretende contrastar  $H_0: \lambda_1 = \lambda_2 = \dots = \lambda_q = 0$ , el test de LM correspondiente vendrá dado por:

$$LM - ERR(l) = \sum_{l=1}^q \frac{[e' W_l e / \hat{\sigma}^2]^2}{T_l} \rightarrow \chi_l^2 \quad (3.40)$$

siendo  $T_l = tr[W_l' W_l + W_l^2]$ ,  $l=1,2,\dots,q$ , siendo  $W_l$  la matriz de contactos de orden  $l$ . Como se puede observar, el test para procesos espaciales de mayor orden es el resultado de la suma de los tests unidireccionales correspondientes a los esquemas autorregresivos o medias móviles de orden uno hasta  $l$ .

Si se observan las expresiones finales de los tres contrastes  $MV$  para la contrastación de autocorrelación espacial residual [(3.33), (3.35), (3.37)] se concluye que el estadístico LM es mucho más sencillo que el  $LR$  o  $W$ , ya que para su aplicación basta con obtener los estimadores MCO de los parámetros del modelo, mientras que para la utilización de los otros dos es necesario resolver un proceso de optimización iterativa no lineal bastante complejo. Es

por ello que se aconseja la elección del estadístico LM frente a los otros dos. Los tres cumplen, no obstante, las propiedades asintóticas deseables siendo equivalentes ante muestras de elevado tamaño, aunque pueden diferir en muestras de tamaño finito.

### Contrastes para el modelo del retardo espacial de la variable endógena

Los contrastes que nos permiten contrastar la existencia de un retardo espacial de la variable endógena, presentan la hipótesis nula de  $H_0: \rho = 0$  frente a  $H_A: \rho \neq 0$ . Dado que el contraste de los LM es el más fácil de implementar, nos centramos en éste, si bien se podrían derivar también los de LR y de  $W$  correspondientes a la FV que incorpora el retardo espacial de la variable endógena. En concreto, Anselin (1988c) derivó dicho test explícitamente:

$$LM - LAG = \frac{[e'Wy / \hat{\sigma}^2]^2}{RJ_{\rho-\beta}} \rightarrow \chi_1^2 \quad (3.41)$$

siendo  $RJ_{\rho-\beta} = [T_1 + (WX\beta)' M(WX\beta) / \sigma^2]^{-1}$  y  $T_1$  tal como se ha definido antes.

### Contrastes robustos en presencia de un error de especificación local

A través de los contrastes anteriores se puede contrastar la presencia de autocorrelación espacial tanto residual como del modelo de retardo espacial. Sin embargo, podría suceder que en el modelo del error espacial analizado estuviese presente un retardo espacial de la variable endógena, y que en el caso del modelo del retardo estuviéramos ante una posible especificación errónea en el término de perturbación, alterando así los resultados de los contrastes LM que se han dado. Así, siguiendo la aproximación general de Bera y Yoon (1993), Anselin *et al.* (1996) han sugerido contrastes que son robustos a errores de especificación locales.<sup>25</sup> La idea de estos contrastes es la de ajustar el test individual de los LM correspondiente teniendo en cuenta que dichos tests convergen a una  $\chi^2$  no central como consecuencia de la presencia de un error de especificación.

El test de LM para los residuos espaciales, es decir,  $H_0: \lambda = 0$  ante la presencia local de  $\rho$  tiene la siguiente expresión:

<sup>25</sup> No es posible construir contrastes robustos en presencia de errores de especificación globales.

$$LM - EL = \frac{[(e'We / \hat{\sigma}^2) - T(R\tilde{J}_{\rho-\beta})^{-1}(e'Wy / \hat{\sigma}^2)]^2}{T - T^2(R\tilde{J}_{\rho-\beta})^{-1}} \rightarrow \chi_1^2 \quad (3.42)$$

donde  $WX\beta$  es un retardo espacial de los valores predichos de la regresión MCO del modelo sin ningún tipo de problema espacial y en donde los demás términos ya han sido presentados.

Al igual que en el caso del test LM-ERR, el test LM-EL necesita únicamente de la estimación del modelo bajo la hipótesis nula de no dependencia espacial residual, coincidiendo ambos contrastes en caso de que  $RJ$  se anule.

El contraste LM-LE permite contrastar la presencia de un retardo espacial de la variable endógena,  $H_0 : \rho = 0$  ante la presencia local de  $\lambda$ , es decir, ante la presencia local de un error con problemas de dependencia espacial. La expresión del contraste es la siguiente:

$$LM - LE = \frac{[(e'Wy / \hat{\sigma}^2 - e'We / \hat{\sigma}^2)]^2}{R\tilde{J}_{\rho-\beta} - T} \rightarrow \chi_1^2 \quad (3.43)$$

donde todas las expresiones ya han sido presentadas.<sup>26</sup>

### 3.4.6.3 Propiedades finitas de los contrastes

Todos los contrastes basados en los LM se basan en las propiedades asintóticas, es decir, su justificación se deriva de la presencia de muestras muy grandes. Sin embargo, éste no suele ser el caso más normal en la mayoría de las aplicaciones, por lo que se han realizado varios trabajos para analizar el comportamiento de los tests de dependencia espacial en muestras de tamaño finito, tanto en cuanto a su tamaño como a su poder, así como las consecuencias que

<sup>26</sup> Anselin (1998) ofrece un contraste de multiplicadores de Lagrange para detectar la presencia de autocorrelación espacial residual en la forma de un modelo de componentes del error espacial (modelo dado por Kelejian y Robinson, 1993). Asimismo, en Anselin y Moreno (1998) se ofrecen las propiedades de dicho contraste y de los comentados en este apartado en muestras finitas a la hora de detectar dicho esquema de dependencia espacial. No obstante, dado que en el presente trabajo no se va a hacer uso de los mismos, no se describen y se remite al lector a las referencias originales.

sobre ellos puede tener una alteración de la distribución normal del término de perturbación o una especificación errónea de la matriz de contactos.

Brandsma y Ketellapper (1979a) encontraron, mediante un experimento de Montecarlo aplicado a las regiones de Holanda, que el test del LR presenta un menor poder en comparación con el test de Moran. De forma más detallada, Mur (1990, 1992), en un ejercicio de simulación investiga el comportamiento del contraste de la  $I$  de Moran y el de LM-ERR mientras que Anselin y Rey (1991) añaden a su estudio el de LM-LAG para diferentes matrices de contacto y distintas distribuciones de los errores. Dado que ambos trabajos ofrecen resultados comunes, se resumen de forma conjunta:

1. En lo referido al tamaño de los contrastes, se observa que el ajuste de la distribución empírica de dichos estadísticos a la distribución teórica respectiva (normal y chi-cuadrado) mejora al aumentar el tamaño de la muestra. Salvo en el caso de un tamaño muestral reducido (16-25 observaciones) en el que el ajuste de los tres tests es pobre, en el de la  $I$  de Moran se observa una mejora significativa al utilizar un tamaño muestral mediano (unas 50 observaciones), mientras que en el caso de los test LM sólo se produce la mejora con un tamaño grande (100 o más observaciones).
2. En términos de poder de los estadísticos, la conclusión es clara. La potencia estimada del estadístico de Moran es superior, para ambos tipos de dependencia, a la estimada para el de los LM, haciéndose muy significativa la diferencia para tamaños pequeños de la muestra y/o valores del parámetro de autocorrelación espacial intermedios o bajos. Sin embargo, resulta necesario matizar esta última conclusión porque, ante muestras pequeñas y valores poco elevados del coeficiente  $\rho$ , la potencia estimada de ambos estadísticos es muy pobre. Los estadísticos LM-ERR y LM-LAG tienen el poder mayor para las alternativas específicas y respectivas para las que han sido diseñados.

Más recientemente, los trabajos de Anselin y Florax (1995b) y Florax y Rey (1995), entre otros, han ofrecido simulaciones más exhaustivas, incluyendo más estadísticos y para distintas matrices de contacto y distintas distribuciones de los errores, con las siguientes conclusiones:

1. Respecto al tamaño de los contrastes, para cualquier tamaño muestral, los tests unidireccionales LM-ERR, LM-EL, LM-LAG, LM-LE presentan frecuencias empíricas correctas, a diferencia de lo que ocurre con el de LM-ERR(2) que sub-rechaza la hipótesis nula cuando se utiliza el criterio *queen* para la matriz de contactos (contigüidad física tal como describe el movimiento de la reina en el juego del ajedrez) y con el test de K-R que sobre-rechaza la hipótesis nula cuando se utiliza el criterio *rook* (basado en el movimiento de la torre en el ajedrez). Asimismo, cuando se modifica la distribución normal del término de perturbación por una distribución log-normal, se alteran los resultados anteriores, principalmente los referidos a la estructura del término de perturbación. Especialmente, el contraste de K-R sobre-rechaza la hipótesis nula para cualquier tamaño muestral, es decir, se rechaza la hipótesis nula de no dependencia espacial en algunos casos en los que es cierta, aumentando innecesariamente la complejidad de la especificación y estimación del modelo espacial.
2. En referencia al poder de los estadísticos, las conclusiones se pueden resumir en los siguientes puntos:
  - i. Se confirman los resultados de Mur (1992) y Anselin y Rey (1991) sobre el mayor poder de la  $I$  de Moran para cualquier caso de dependencia espacial y el buen comportamiento de los tests de LM contra las alternativas concretas para las que han sido diseñados.
  - ii. En general, los tests unidireccionales y multidireccionales presentan mayor poder contra alternativas autorregresivas que contra estructuras medias móviles.
  - iii. Los contrastes para el modelo del retardo de la endógena presentan un mayor poder que los del modelo del término de perturbación espacial. Este resultado es importante dado que las consecuencias sobre las estimaciones MCO del modelo cuando se ignora un retardo espacial de la variable endógena

- (inconsistencia) son más graves que las de obviar una estructura de dependencia espacial en el error (ineficiencia).
- iv. El contraste de K-R presenta un escaso poder. Los autores resaltan una doble posible explicación para este resultado. En primer lugar, debido a sus mayores grados de libertad. En segundo lugar, al hecho de que su poder dependa del grado de autocorrelación en las variables explicativas. Sin embargo, en el trabajo de Anselin y Moreno (1998) se ha estudiado más detenidamente el escaso poder de dicho test ante una posible autocorrelación y/o multicolinealidad de las variables explicativas, concluyéndose que éstas no son las razones del bajo poder del contraste.<sup>27</sup>
- v. Buen comportamiento de los contrastes robustos ante posibles errores de especificación, LM-EL y LM-LE, mostrando un poder casi indistinguible entre ellos y sus tests análogos no robustos para tamaños muestrales elevados y con tamaños empíricos razonables. Es decir, parece ser que dichos contrastes robustos resultan escasamente penalizados cuando su utilización es innecesaria. A su vez, los tests ajustados parecen ser más apropiados para contrastar la dependencia del retardo de la endógena en presencia de correlación en los errores que en el caso inverso, lo cual resulta importante dado que ignorar la dependencia de un retardo de la endógena tiene consecuencias más importantes. Teniendo en cuenta estos resultados, Anselin y Florax (1995b) sugieren la siguiente estrategia de decisión del mejor modelo espacial. Cuando el test LM-LAG resulte más significativo que el test LM-ERR y el LM-LE sea significativo a diferencia del test LM-EL, la presencia de un retardo espacial de la variable endógena será la hipótesis correcta. De

---

<sup>27</sup> En el trabajo de Anselin y Moreno (1998) se propone una variación en el contraste de K-R que permite aumentar el poder del mismo. En concreto, el cambio consiste en derivar la regresión auxiliar no sólo con los productos cruzados de los residuos de las observaciones que están potencialmente correlacionadas según la matriz de contactos ( $W$ ), sino también considerar aquéllas que resultarían correlacionadas según el esquema dado por el producto,  $WW'$ . De esta manera, se estarían considerando las contigüidades de primer y segundo orden, detectando con mayor precisión los problemas de autocorrelación espacial.

forma análoga se puede identificar la presencia de dependencia espacial en el término de error mediante el test LM-EL.

- vi. El test LM-ERR(2) tiene un elevado poder para detectar la presencia de órdenes superiores de dependencia espacial en el término de perturbación, aunque proporcionan una escasa guía para la búsqueda de la especificación correcta.
- vii. La diferente especificación de la matriz de contactos supone un cambio en el resultado de todos los tests, con un mayor poder para el caso de *rook*.<sup>28</sup>
- viii. La alteración de la distribución normal del término de error tiene escasas consecuencias sobre el poder de los estadísticos, si bien son los estadísticos para el error los que se ven más afectados, a diferencia de los tests de los retardos, más robustos ante este tipo de alteraciones.

### 3.4.7 Métodos de estimación ante problemas de autocorrelación espacial

Como se ha comentado anteriormente, la estimación mínimo cuadrática en modelos que presentan dependencia espacial no es apropiada. En el caso de las series temporales, se sabe que el estimador MCO sigue siendo consistente incluso ante la presencia de una variable dependiente retardada en el tiempo, siempre que el término de perturbación no muestre correlación serial. En consecuencia, aunque las propiedades de los estimadores en muestras de tamaño pequeño se ven afectadas (dejan de ser insesgadas) pueden seguirse utilizando como base para la inferencia asintótica. ¿Se obtienen los mismos resultados con los modelos espaciales?

---

<sup>28</sup> Florax y Rey (1995) obtienen que una sobreespecificación de la matriz de contactos (por ejemplo, utilizar el criterio *queen* en lugar del *rook*) disminuye de forma importante el poder de los tests LM, sobre todo para valores negativos y elevados de los coeficientes espaciales. En caso de la subespecificación, la disminución del poder de los tests es superior para valores positivos de los coeficientes espaciales para la *I* de Moran, siendo igual para los contrastes de LM.

Consideremos inicialmente un modelo autorregresivo espacial de primer orden:

$$Y = \rho WY + \varepsilon \quad (3.44)$$

donde  $\varepsilon$  es un término de perturbación distribuido independiente e idénticamente. A pesar de que este modelo es excesivamente sencillo, captura todos los efectos que puede suponer la presencia de una variable dependiente retardada espacialmente en los estimadores MCO, y por tanto, sin pérdida de generalidad.

Si llamamos  $r$  a la estimación mínimo-cuadrática de  $\rho$ :

$$r = (Y_L' Y_L)^{-1} Y_L' Y \quad (3.45)$$

con  $Y_L = WY$ , la variable dependiente retardada espacialmente.

Substituyendo la expresión de  $Y$  de (3.44) en (3.45) se obtiene:

$$r = (Y_L' Y_L)^{-1} Y_L' (\rho Y_L + \varepsilon) = \rho + (Y_L' Y_L)^{-1} Y_L' \varepsilon \quad (3.46)$$

resultado muy similar a la situación en series temporales, ya que el valor esperado del segundo término no es igual a cero y por tanto, es sesgada la estimación mínimo-cuadrática, ya que la esperanza del estimador no es igual al parámetro poblacional, sino que presenta un sesgo.

Asintóticamente, la consistencia del estimador mínimo-cuadrático depende de las dos condiciones siguientes:

$$\text{plim } N^{-1} (Y_L' Y_L) = Q, \text{ matriz finita y no singular}$$

$$\text{plim } N^{-1} (Y_L' \varepsilon) = 0$$

Mientras que la primera condición puede verse satisfecha con las restricciones adecuadas en el valor del parámetro  $\rho$  y la estructura de la matriz de contactos, la segunda condición no se mantiene en el caso espacial. En realidad (Anselin, 1988a):

$$plim N^{-1} (Y_L' \varepsilon) = plim N^{-1} \varepsilon W (I - \rho W)^{-1} \varepsilon$$

Así, la presencia de la matriz de contactos en esta expresión nos hace obtener una forma cuadrática en los términos de perturbación. Por consiguiente, excepto en el caso de que  $\rho = 0$ , la probabilidad en el límite de esta expresión no sería cero. En consecuencia, el estimador MCO de los parámetros de un modelo autorregresivo espacial será sesgado además de inconsistente, independientemente de las propiedades del término de perturbación.

En cambio, si analizamos los efectos de la autocorrelación residual espacial en las propiedades de los estimadores MCO se observa que son bastante similares a los resultados obtenidos con series temporales. Los estimadores serán insesgados pero ineficientes debido a la estructura no diagonal de la matriz de varianzas de los términos de perturbación. En realidad, en lo que se refiere a una estructura autorregresiva espacial en los términos de perturbación, no es posible obtener una estimación consistente del parámetro autorregresivo mediante la estimación mínimo-cuadrática. Al contrario de lo que sucede con el método de Mínimos Cuadrados Generalizados Factibles (MCGF) desarrollados para la dependencia serial en el dominio temporal (que permite obtener estimadores consistentes), la naturaleza multidireccional de la dependencia espacial hace que no haya un estimador iterativo o en dos etapas que sea consistente para este tipo de modelos. Para comprobar tal afirmación, partimos de que la dependencia espacial en el término de perturbación de un modelo lineal causa la no esfericidad de la matriz de covarianzas de los errores, siendo:

$$E[\varepsilon \varepsilon'] = \sigma^2 \Omega(\lambda) \tag{3.47}$$

$$\text{con } \Omega(\lambda) = (I - \lambda W)^{-1} [(I - \lambda W)^{-1}]' = [(I - \lambda W)' (I - \lambda W)]^{-1}$$

Consecuencia de ello, la varianza asociada a los estimadores mínimo-cuadráticos de los coeficientes tampoco tendrá la forma tradicional de  $\sigma^2 (X'X)^{-1}$ , sino que será una función

compleja del parámetro  $\lambda$ . Por tanto, aunque el estimador MCO mantiene la insesgadez, ya no es eficiente y la inferencia basada en estimadores MCO para la varianza y el ajuste del modelo puede llevarnos a resultados erróneos.

Si se conoce la covarianza del término de perturbación, el mejor estimador lineal insesgado es el de *Mínimos Cuadrados Generalizados de Aitken (MCG)*, lo que implica que debería conocerse la estructura de dependencia espacial (la matriz de contactos así como el coeficiente asociado,  $\lambda$ ). Pero en la práctica no sucede así. Para resolver este problema se utiliza el método de estimación de *Mínimos Cuadrados Generalizados Factibles (MCGF)* que consiste en la aplicación del principio de MCG con estimaciones consistentes de los parámetros en  $\Omega$  sustituyendo a los valores poblacionales desconocidos. En el modelo espacial, éste es el estimador resultante:

$$b_{MCGF} = [X' (I - \lambda W)' (I - \lambda W) X J^I X' (I - \lambda W)' (I - \lambda W) Y]^{-1} X' (I - \lambda W)' (I - \lambda W) Y \quad (3.48)$$

Dicha estimación es equivalente numéricamente a la estimación MCO de una transformación adecuada de las variables:

$$b_{MCGF} = [X^* X^* J^I X^* Y^*]^{-1} X^* Y^*$$

siendo  $X^* = (I - \lambda W) X$ ,  $Y^* = (I - \lambda W) Y$ , en donde  $\lambda$  está reemplazado por una estimación consistente.

La matriz de varianzas de los coeficientes tendrá la forma:

$$Var[b_{MCGF}] = \sigma^2 [X' (I - \lambda W)' (I - \lambda W) X J^I]^{-1} \quad (3.49)$$

y

$$\sigma^2 = (Y - X b_{MCGF})' (Y - X b_{MCGF}) / (N - K)$$

Sin embargo, resulta necesario decir que los resultados de la estimación MCGE son de naturaleza asintótica, por lo que en muestras finitas la distribución de los estimadores no está

bien definida, y ni siquiera es entonces superior a la estimación MCO en el sentido de un menor error cuadrático medio. Asimismo, las propiedades del estimador MCGE son muy sensibles a una incorrecta especificación de  $\Omega(\lambda)$ , es decir, la elección de la correcta matriz de contactos es un aspecto muy importante.

Al contrario de lo que sucede en el caso de los varios procedimientos MCGE desarrollados para la dependencia serial en el dominio temporal, no existe un estimador iterativo o en dos etapas que sea consistente para el modelo residual espacial. La mejor alternativa es, por tanto, la estimación MV.

En conclusión, la inadecuación del estimador MCO en modelos que incorporan dependencia espacial ha hecho centrar la atención de la literatura econométrica espacial en la aproximación máximo verosímil como alternativa principal aunque existen otros métodos aplicables pero menos utilizados como el de Variables Instrumentales (VI) o Método Generalizado de los Momentos (GMM).<sup>29</sup>

#### 3.4.7.1 Estimación máximo verosímil (MV)

La estimación MV de los modelos de regresión que incorporan problemas de autocorrelación espacial, bien en la forma de un término de perturbación espacial bien como un retardo espacial, fue tratada inicialmente por Ord (1975). Partiendo de una función de densidad normal conjunta para los errores en el modelo, se obtiene una función de verosimilitud (FV), uno de cuyos aspectos más interesantes es el Jacobiano de la transformación, que toma la forma de  $|I - \lambda W|$  y  $|I - \rho W|$  para el modelo del error autorregresivo espacial y del retardo espacial, respectivamente. Al contrario de lo que sucede en el caso del dominio temporal, el Jacobiano espacial no es el determinante de una matriz triangular, sino de una matriz completa, lo que complica los cálculos de forma considerable. Sin embargo, Ord (1975) mostró que dicho término Jacobiano puede ser expresado en función de los valores propios  $v_p$ , de la matriz de contactos, tal que:

---

<sup>29</sup> Para una explicación detallada del método de estimación generalizado de los momentos aplicado al campo espacial, véase Anselin y Bera (1998).

$$|I - \rho W| = \prod_{i=1}^N (1 - \rho v p_i) \quad (3.50)$$

La estimación MV de los modelos que presentan autocorrelación espacial en los errores pueden considerarse como un caso especial de los términos de error no esféricos parametrizados tal como se ha dado en (3.47). La FV del modelo con término de perturbación autorregresivo espacial parte de la hipótesis de normalidad y contiene un término Jacobiano  $|I - \lambda W|$ . Esta expresión puede simplificarse en términos de los valores propios de la matriz de contactos espacial (Anselin, 1992a). Utilizando dicha simplificación, el logaritmo de la FV es:

$$L = \ln |I - \lambda W| - (N/2) \ln 2\pi - (N/2) \ln \sigma^2 - \frac{(Y - \lambda WY - X\beta + \lambda WX\beta)'(Y - \lambda WY - X\beta + \lambda WX\beta)}{2\sigma^2} \quad (3.51)$$

Esta expresión muestra claramente el porqué, al contrario que en el caso de series temporales, la estimación MCO no es máximo verosímil, dado que se ignoraría el término Jacobiano. Puede mostrarse que las estimaciones de  $\beta$  y de  $\sigma^2$  pueden expresarse en función del coeficiente autorregresivo  $\lambda$ :

$$b_{MV} = [(X - \lambda WX)'(X - \lambda WX)]^{-1} (X - \lambda WX)'(Y - \lambda WY) \quad (3.52)$$

$$\hat{\sigma}^2 = (Y - \lambda WY - Xb + \lambda WXb)'(Y - \lambda WY - Xb + \lambda WXb) / N$$

La estimación MV es muy similar a la estimación MCGE en muchos aspectos, con la diferencia de que la estimación del parámetro que acompaña al término de perturbación  $\lambda$  en MV está especificado claramente e incluido en la inferencia de manera consistente. El estimador  $\lambda$  se encuentra por una minimización de la forma cuadrática de los errores. De hecho, Anselin (1988a) demuestra que los resultados MCGF para  $\beta$  y  $\sigma^2$  son también MV (siguiendo una de las condiciones de orden de la maximización del logaritmo de la FV con dependencia espacial). De este modo, después de substituir  $b$  y  $\sigma^2$  en la FV se obtiene una FV concentrada (ignorando la constante en  $2\pi$ ) que sólo contiene el parámetro  $\lambda$  y los residuos  $e = (Y - Xb_{MCGF})$ :

$$L_C = -N/2 \ln e'(I - \lambda W)'(I - \lambda W)e + \ln |I - \lambda W| \quad (3.53)$$

Esta expresión, condicionada a los valores  $b_{MCGF}$  que dan lugar a los residuos, puede maximizarse para obtener un nuevo grupo de  $b_{MCGF}$ , los que darían lugar a nuevos residuos, proceso que puede seguir iterándose hasta alcanzar la convergencia.

La varianza asintótica de las estimaciones MV se ajusta a la forma general de Breusch (1980) y es diagonal en bloques entre los parámetros de la regresión ( $b_{MV}$ ) y los parámetros de la varianza de los errores  $\sigma^2$  y  $\theta$ . Así, en el caso del error autorregresivo espacial, la varianza asintótica para los coeficientes tiene la expresión:

$$AsyVar[b] = \sigma^2 [(X - \lambda WX)'(X - \lambda WX)]^{-1} \quad (3.54)$$

mientras que el bloque de la varianza para los parámetros de los errores es:

$$AsyVar[\sigma^2, \lambda] = \begin{bmatrix} N / 2\sigma^4 & tr(W_B) / \sigma^2 \\ tr(W_B) / \sigma^2 & tr(W_B)^2 + tr(W_B'W_B) \end{bmatrix}^{-1} \quad (3.55)$$

donde  $W_B = W(I - \lambda W)^{-1}$ . Dada la forma diagonal en bloque de la matriz de varianzas asintótica, la precisión de  $\lambda$  no afecta la precisión de la estimación de los regresores. Por tanto, si el principal objetivo es la estimación de los regresores, no resulta necesario calcular (3.55).

Un estadístico apropiado para contrastar la significación del parámetro del error espacial puede basarse en un test del LR aplicado directamente.

Existen, sin embargo, dos procedimientos iterativos en los modelos espaciales, Cochrane-Orcutt espacial y Durbin espacial, que evitan el tener que realizar las múltiples iteraciones en el proceso MV no lineal.

El procedimiento *Cochrane-Orcutt espacial* tiene tres pasos:

1. Estimación MCO del modelo lineal
2. Utilización de los residuos obtenidos por MCO para conseguir una estimación del coeficiente autorregresivo espacial  $\lambda$
3. Utilización del parámetro  $\lambda$  en el procedimiento MCGE.

Este proceso iterativo se repite hasta la convergencia.

El procedimiento *Durbin espacial* se basa en la equivalencia formal del modelo con dependencia residual a la especificación autorregresiva mixta regresiva espacial. Formalmente a partir de:

$$Y = X\beta + (I - \lambda W)^{-1} \mu \quad (3.56)$$

podemos obtener, directamente:

$$Y = \lambda WY + X\beta - \lambda WX\beta + \mu \quad (3.57)$$

donde  $\mu$  es un término de perturbación esférico. De este modo, una estimación MV o por VI nos dará estimaciones consistentes de  $\lambda$  y  $\beta$ , siendo mucho más fácil de calcular que la regresión con dependencia espacial en los términos de perturbación.

La FV para el modelo con retardo espacial es similar a la vista para el modelo con término de perturbación autocorrelacionado espacialmente, conteniendo el término Jacobiano  $|I - \rho W|$ . Dicha expresión también puede simplificarse obteniéndose la FV logarítmica:

$$L = \ln (I - \rho W) - N/2 \ln 2\pi - N/2 \ln \sigma^2 - (Y - \rho WY - X\beta)' (Y - \rho WY - X\beta) / 2\sigma^2 \quad (3.58)$$

De la misma manera, las estimaciones de  $\beta$  y  $\sigma^2$  pueden expresarse en función del coeficiente de autocorrelación  $\rho$  (para mayor detalle, véase Ord, 1975; Anselin, 1988a):

$$b_{MV} = (X'X)^{-1} X' (Y - \rho WY) \quad (3.59)$$

$$\hat{\sigma}^2 = (Y - \rho WY - Xb)(Y - \rho WY - Xb) / N$$

La sustitución de las expresiones (3.59) en la FV nos da un logaritmo de verosimilitud concentrado como una función no lineal de un parámetro singular  $\rho$ :

$$L_C = -(N/2) \ln[(e_0 - \rho e_L)'(e_0 - \rho e_L) / N] + \ln|I - \rho W| \quad (3.60)$$

donde  $e_0$  y  $e_L$  son los residuos en la regresión de  $Y$  sobre  $X$  y de  $WY$  sobre  $X$ , respectivamente. Se puede obtener un estimador LM para  $\rho$  mediante una optimización numérica de la FV concentrada. Los estimadores resultantes mantienen las propiedades típicas asintóticas de consistencia, normalidad y eficiencia asintótica. La matriz de varianzas asintótica se obtiene como la inversa de la matriz información:

$$\text{AsyVar}[\rho, \beta, \sigma^2] = \quad (3.61)$$

$$\begin{bmatrix} \text{tr}(W_A)^2 + \text{tr}(W_A' W_A) + (W_A X \beta)' + (W_A X \beta) / \sigma^2 & (X' W_A X \beta) / \sigma^2 & (\text{tr}(W_A)) / \sigma^2 \\ (X' W_A X \beta) / \sigma^2 & (X' X) / \sigma^2 & 0 \\ (\text{tr}(W_A)) / \sigma^2 & 0 & N / (2\sigma^4) \end{bmatrix}$$

siendo  $W_A = W(I - \rho W)^{-1}$ . Así, mientras la covarianza entre  $\beta$  y la varianza del error es cero, como en el modelo de regresión estándar, éste no es el caso de  $\rho$  y la varianza del error. Esta falta de diagonalidad en bloque en la matriz de información del modelo del retardo espacial es otro rasgo diferenciador entre el caso espacial y el análogo en series temporales.

El ajuste de los modelos espaciales estimados por métodos MV no debe basarse en las medidas tradicionales como el  $R^2$  ya que nos llevaría a conclusiones erróneas en presencia de autocorrelación espacial. En su lugar, se pueden utilizar medidas como la comparación del logaritmo de verosimilitud, o una forma ajustada de tener en cuenta el número de parámetros en el modelo, tales como la medida de AIC (Anselin, 1988b).

Cuando se está en presencia de variables dependientes retardadas especialmente, las propiedades asintóticas del estimador MV se suelen considerar como aptas. Sin embargo, recientemente se han obtenido unas condiciones formales para la consistencia, eficiencia y normalidad asintótica del estimador máximo verosímil con variables dependientes. En términos generales estas condiciones son:

1. la existencia de la FV para los valores de los parámetros bajo consideración (una FV no degenerada)
2. la continua diferenciabilidad del logaritmo de la FV (hasta 2º y 3º orden)
3. la existencia de una matriz de covarianzas definida positiva y/o no singular
4. que el número de parámetros se fije de manera independiente del número de observaciones, con el objetivo de evitar el problema incidental de los parámetros, común en modelos espacio-tiempo. Por ejemplo, una situación donde cada unidad espacial tuviera un parámetro separado para indicar sus rasgos característicos (variable *dummy*) no sería admisible en un contexto donde lo asintótico se refiere a la dimensión transversal (*cross-section*)

Estas condiciones suelen cumplirse cuando la estructura de interacción espacial, expresada tanto por el coeficiente autorregresivo como por la matriz de contactos, no es explosiva. Esto se puede comprobar estudiando las propiedades del término Jacobiano asociado a cada modelo.

El caso del modelo con dependencia espacial en el término de perturbación podemos considerarlo como un caso especial del modelo con una matriz de covarianzas de los errores no escalar. Dicha matriz es generalmente desconocida y suele estar expresada en términos de un número finito de parámetros, tales como el parámetro autorregresivo en el caso que nos ocupa. Si se conoce la matriz de varianzas de los errores, el estimador MCG es el mejor estimador lineal insesgado, así como el MV. Sin embargo, en la situación más realista, en donde los parámetros de la varianza del término de perturbación se basan en una estimación, el estimador MV bajo condiciones generales alcanza las propiedades deseables de consistencia, eficiencia asintótica y normalidad asintótica. Además, en muchas situaciones, las estimaciones de los parámetros del modelo también son insesgadas.

Las condiciones de regularidad que deben cumplirse para que así sea son muy similares a las expuestas anteriormente y son diseñadas para garantizar el buen comportamiento de la FV así como una matriz de covarianzas de los errores definida positiva. En este modelo espacial estas condiciones se cumplen cuando nos encontramos ante una estructura de interacción en la matriz de contactos no explosiva, e imponiendo restricciones de no negatividad en los elementos de la diagonal de la matriz de covarianzas de los errores estimados.

### 3.4.7.2 Estimación por Variables Instrumentales

La consideración de la variable dependiente retardada espacialmente  $WY$  en el modelo del retardo espacial como una forma de simultaneidad o de endogeneidad ha motivado la utilización de una aproximación por variables instrumentales (VI) en su estimación. Este método de estimación presenta la ventaja de no requerir la hipótesis de normalidad de los residuos y la mayor facilidad de los cálculos cuando las matrices de contactos que se utilizan son de un tamaño considerable (en torno a las 2000 observaciones o más). Dado que el principal problema es la correlación entre  $WY$  y el término de error, la elección de unos instrumentos adecuados que estén incorrelacionados con el término de perturbación ofrecerán estimadores consistentes. No obstante, como resulta usual en la utilización de VI, la eficiencia de estos estimadores dependerá en forma importante de la elección de los instrumentos y puede ser pobre en muestras pequeñas.

Utilizando los resultados estándar del método de VI y considerando  $Q$  como la matriz ( $P \times N$ ) de instrumentos que incluyen las  $K$  variables exógenas de  $X$ ), el método de VI o Mínimos Cuadrados en dos etapas se estima tal que:

$$\beta_{VI} = [Z' Q(Q' Q)^{-1} Q' Z]^{-1} Z' Q(Q' Q)^{-1} Q' Y \quad (3.62)$$

siendo

$$\begin{aligned} Z &= [WY \quad Z] \\ \text{AsyVar}(\beta_{VI}) &= \sigma^2 [Z' Q(Q' Q)^{-1} Q' Z]^{-1} \\ \sigma^2 &= (Y - Z\beta_{VI})'(Y - Z\beta_{VI}) / N \end{aligned}$$

Como se ha comentado, un aspecto crucial en la utilización del método de VI es la elección de los instrumentos. Si bien se han dado varias sugerencias, en econometría espacial, para tratar de guiar en la elección de los instrumentos para  $WY$ , recientemente, Kelejian y Robinson (1993) demuestran la consistencia de un estimador basado en el método de los momentos generalizados para el modelo del retardo espacial con instrumentos que consisten en los retardos espaciales de primer orden y de órdenes superiores de las variables explicativas ( $WX$ ,  $W^2X$ , etc.). No obstante, se ha demostrado que la utilización de los retardos de primer orden ya es suficiente para la obtención de estimaciones consistentes.