



Geolocalización y *Big Data*

Un modelo de comercio con **econometría espacial**

Marcos Herrera
CONICET-IELDE, Universidad Nacional de Salta

LA CRECIENTE DISPONIBILIDAD DE DATOS GEORREFERENCIADOS REQUIERE HERRAMIENTAS PARTICULARES QUE PERMITAN CAPTURAR LA INFORMACIÓN DISPONIBLE. LAS PREDICCIONES ECONÓMICAS SE POTENCIAN CON *MACHINE LEARNING* O RECONOCIMIENTO DE IMÁGENES SATELITALES. LA ECONOMETRÍA ESPACIAL OFRECE LA POSIBILIDAD DE DETECTAR LOS EFECTOS DE INTERACCIÓN PROVENIENTES DE ESPACIOS GEOGRÁFICOS, COMO SE MUESTRA CON UN EJEMPLO DE MODELO DE COMERCIO GRAVITACIONAL.

La incipiente disponibilidad de datos georreferenciados a nivel individual y la ofrecida por diferentes agencias estadísticas a nivel mundial han generado un creciente uso de técnicas de análisis y econometría espacial. La econometría espacial es una rama econométrica dedicada al análisis de efectos espaciales en modelos de regresión. Los efectos espaciales (dependencia y heterogeneidad) provienen de la proximidad o vecindad entre las unidades de observación y se entienden de manera amplia: geográfica (proximidad según posición geográfica), social (proximidad según vínculos familiares/sociales), económica (proximidad según interacción comercial/relaciones sectoriales) o combinaciones entre estas alternativas. Ya sea cualquiera de las relaciones de proximidad mencionadas, la econometría espacial considera de principal relevancia el impacto de tales interacciones y genera modelos que permiten estimar el efecto de las mismas.

Puede pensarse que estos desarrollos espaciales son recientes, sin embargo los comienzos datan del año 1979 (Anselin, 2010), cuando se iniciaron diferentes desarrollos metodológicos plasmados una década más tarde en el libro *Spatial Econometrics: Methods and Models* (Anselin, 1988). En los primeros años, la difusión en economía fue limitada debido a la escasa disponibilidad de este tipo de datos y también por la

falta de opciones de programas estadísticos para la estimación. En la actualidad, ambas barreras han sido derribadas, en particular por la disponibilidad de datos georreferenciados ofrecidos por usuarios individuales, por empresas que permiten su acceso mediante una interfaz de programación (API) o por técnicas de programación que logran extraer datos de sitios web (*web scraping*). Además, los métodos de estimación espacial se encuentran integrados a los programas estadísticos más populares, como Stata, R, Matlab y Python, entre otros.

Arribas-Bel (2014) destaca que la cantidad y diversidad de las nuevas fuentes de información tienen una característica accidental, disponible para investigación y en cualquier lugar del planeta. Es accidental en referencia a que los datos no fueron pensados desde una encuesta o censo específicamente creado para investigación o análisis de políticas sociales o económicas. Se une a esto la disponibilidad sin costo de acceso para el investigador. Las fuentes son tan diversas y a diferentes escalas que permiten reducir el error de localización de las observaciones, evitar la discretización de problemas continuos en el espacio y nutrir de información externa a datos recolectados de forma tradicional (encuestas y censos). Pero las técnicas estadísticas y econométricas convencionales no están prepara-

das para manejar el volumen de datos actual y por ello el surgimiento del *big data*. Asimismo, tampoco logran capturar de forma adecuada estructuras complejas, como son la interconectividad y la geolocalización de las observaciones, como sí lo puede realizar la econometría espacial.

La característica distintiva de un modelo econométrico espacial es la presencia de la matriz de contactos o red de conectividad. Esta red de conectividad puede verse como un punto en común con la literatura de redes sociales. La diferencia con esta literatura es que en redes sociales el foco de atención se centra en el análisis de la estructura de conectividad, lo que caracteriza a la red con diferentes medidas (centralidad, coeficiente de *clustering*, grado medio, densidad de la red, etcétera), en cambio, la econometría espacial focaliza su atención en el impacto de la red de conectividad en el modelo econométrico. Claramente, ambas literaturas tienen puntos en común y pueden complementarse.

Los desarrollos de *big data* también encuentran un punto en común bajo datos georreferenciados (*geo big data*) y ofrecen estrategias alternativas para capturar efectos espaciales. Por ejemplo, el proyecto de Livehoods busca definir áreas urbanas no solo por los tipos de lugares próximos geográficamente, sino también por las personas que hacen que el área forme parte de su rutina diaria. Usando más de 18 millones de registros del sitio Foursquare, agrupan los lugares cercanos en áreas según patrones en el grupo de personas que se registran en ellos. Al examinar los patrones en estos *check-ins* (clasificaciones y opiniones del lugar), se aprende sobre las diferentes áreas que componen la ciudad, lo que permite estudiar la dinámica social, la estructura y el carácter de las ciudades

a gran escala. Este tipo de propuesta construye vecindarios digitales combinando preferencias de usuarios e información georreferenciada.

Como puede verse, las alternativas de análisis de datos espaciales son amplias y se encuentran en pleno auge y desarrollo. En lo que resta del artículo, intentaré destacar las principales herramientas de la econometría espacial para capturar información compleja. Como punto inicial, debe reconocerse que los datos espaciales presentan características muy particulares, indistintamente de si se cuenta con una base de datos de dimensiones tradicionales o de una gran base de datos como las que manejan *big data*.

NATURALEZA DE LOS DATOS ESPACIALES

El primer aspecto para destacar es que existen diferentes tipos de datos espaciales y esto conlleva desarrollos estadísticos particulares para los mismos. Cressie (2015) realiza una taxonomía de datos según los supuestos del proceso estocástico espacial (*spatial random field*). Sin entrar en formalizaciones, los datos espaciales pueden clasificarse como geoestadísticos (*geostatistical data*), regionales (*lattice data*) y de patrón de puntos (*point pattern data*). A su vez, la representación y visualización de estos datos puede realizarse en formato vectorial o raster, que son dos tipos de capas de información espacial habituales en los Sistemas de Información Geográfica (SIG).

En el formato raster, el espacio se divide en celdas regulares y cada celda contiene un número que identifica al objeto. Este tipo de representación es la aconsejada para datos geoestadísticos como superficies de temperaturas, contaminación, precipitaciones, etcétera.

Las imágenes satelitales por teledetección (satélites o drones) son almacenadas bajo este formato.

En el formato vectorial, existen diferentes objetos representados por puntos, líneas y polígonos. Este tipo de representación es típica de datos regionales, como un mapa de un país dividido en provincias o municipios, o las líneas de ferrocarril y los puntos de viviendas en venta. Este es el tipo más habitual de datos georreferenciados en ciencias sociales en general y en economía en particular.

Pero ambos formatos pueden combinarse y hoy en día no es extraño encontrar aplicaciones mixtas. Por ejemplo, Patino y Duque (2013) revisan las aplicaciones en ciencia regional que utilizan imágenes satelitales en asentamientos urbanos. Las aplicaciones más comunes utilizan imágenes de Landsat, SPOT, ASTER, entre otros, para detectar zonas urbanas deprimidas, evaluación de calidad de vida, crecimiento urbano y vulnerabilidad social. En un trabajo de gran repercusión, Jean *et al.* (2016) analizan imágenes satelitales, datos vectoriales y datos de encuestas oficiales mediante técnicas de aprendizaje automático (*machine learning*) para estimar el gasto de consumo y riqueza en cinco países africanos. Mediante esta combinación de información y técnica, los autores logran focalizar zonas pobres que pueden ser objeto de políticas específicas.

Un inconveniente común a todos los datos espaciales es el problema de unidad de área modificable (*modifiable areal unit problem*, MAUP), versión geográfica de la falacia ecológica, en donde las conclusiones basadas en una agregación particular de las áreas o regiones pueden cambiar si se agrupan los mismos datos en un conjunto diferente de áreas o regiones. Esto tiene implicancia en los resultados in-

ferenciales y no siempre es destacado adecuadamente. Además, la utilización de diversas fuentes de datos conlleva riesgos; por ejemplo, es común que la información no esté disponible al nivel de escala deseada. La transformación necesaria de los datos espaciales a la escala de interés genera el problema de cambio de soporte (*COSP, change of support problem*). El término “soporte” hace referencia al tamaño y volumen de cada base, pero también incluye la forma, el tamaño y la orientación espacial de los objetos o campos representados. El paso de un soporte involucra generar nuevas variables relacionadas a las originales, pero con diferencias espaciales y estadísticas.

Estas particularidades de los datos espaciales son poco consideradas al trabajar con este tipo de datos. Piénsese en el modelo gravitacional de comercio, en donde los países (polígonos) son transformados en puntos (generalmente la ciudad capital) para medir la distancia y así utilizar dicha medición como variable. Esta simple transformación puede contener todos los problemas antes comentados. Puede ser cierto que, en este caso, la exactitud de la medición no modifique significativamente los resultados. Sin embargo, el avance en la geolocalización de los productos generará en poco tiempo la necesidad de una mayor precisión del origen y el destino de cada bien y replanteará cuestiones tan simples como la medición de esta variable.

MODELOS ECONOMÉTRICOS

Ya sea que el espacio relevante sea el geográfico, el digital o el socioeconómico, las herramientas desarrolladas por la econometría espacial permiten obtener una medida de la interacción y significancia de los efectos espaciales o

de red. La principal virtud descansa en reconocer desde el inicio la naturaleza dependiente de los datos, contrario a lo que se asume en los métodos estadísticos habituales.

La econometría estándar establece independencia de las observaciones y permite así un tratamiento estadístico manejable. Una implicación de este supuesto es que los valores observados y_i , para un individuo i , son estadísticamente independientes de otros valores y_j , para un individuo j , tal que $E(y_i y_j) = 0$, donde $E(\cdot)$ es el operador de esperanza, suponiendo valores centrados, por simplicidad. Sin embargo, cuando los individuos interactúan de manera tal que sus decisiones dependen de sus pares o vecinos, este supuesto debe relajarse para permitir alguna forma de dependencia.

La forma tradicional de dependencia espacial en un corte transversal es por medio de la autocorrelación espacial entendida como la similitud de valores en localizaciones próximas. La autocorrelación puede ser positiva, en el sentido de que un valor alto (bajo) de una variable aleatoria en una posición espacial es acompañado por un vecindario de valores altos (bajos). También es posible una autocorrelación espacial negativa con un valor elevado (bajo) en una posición geográfica y valores bajos (elevados) en el vecindario. La existencia de autocorrelación espacial implica que una muestra de datos autocorrelacionados contiene menos información

que una muestra no correlacionada. Esta pérdida de información debe ser explícitamente considerada en la estimación y este es el problema esencial en temas de econometría aplicada con datos espaciales.

LA MATRIZ DE PESOS ESPACIALES

La autocorrelación espacial puede verse como una versión reescalada de la covarianza: $Cov(y_i, y_j) = E(y_i y_j) - E(y_i)E(y_j) \neq 0, \forall i \neq j$, donde y_i, y_j son observaciones de una variable aleatoria en la localización i y j en el espacio, es decir, todo par (i, j) posee información geográfica específica medida por latitud y longitud. Sin embargo, para una muestra de n observaciones en un corte transversal, existen $(n^2 - n)/2$ covarianzas por ser estimadas bajo simetría y no hay suficientes datos para estimar todos los pares. La forma de resolver este problema es imponiendo restricciones sobre la forma en que las observaciones interactúan entre sí.

El principal enfoque en econometría espacial para imponer restricciones a la interacción es mediante la matriz de pesos espaciales, comúnmente denominada "W" (o "G" en redes sociales) y describe la conectividad entre las n unidades que se encuentran localizadas en un espacio bidimensional. La construcción de W está basada en al menos dos supuestos clave sobre la estructura espacial: (1) un criterio de conectividad que define qué unidades pueden ser consideradas vecinas entre sí y (2) un supuesto de pesos espaciales que operacionaliza cómo los vecinos se afectan entre sí.

Para entender cómo la matriz W permite simplificar el problema de dependencia espacial, formalizaremos la discusión. Supóngase un proceso autorregresivo espacial SAR (*spatial auto-*

regressive) tal que la variable se distribuye espacialmente en tres regiones de la siguiente forma:

$$\begin{aligned} y_i &= \alpha_{ij} y_j + \alpha_{ik} y_k + u_i \\ y_j &= \alpha_{ji} y_i + \alpha_{jk} y_k + u_j \\ y_k &= \alpha_{ki} y_i + \alpha_{kj} y_j + u_k \end{aligned}$$

siendo $u_i, u_j, u_k \sim i.i.d.(0; \sigma^2)$. Es decir, en la primera ecuación, el valor en la región i depende del valor de y_j y y_k , más un término aleatorio que se distribuye de manera idéntica e independiente entre las localizaciones. Lo mismo puede decirse de las regiones j y k . En términos matriciales, el sistema puede reexpresarse como:

$$\begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{ij} & \alpha_{ik} \\ \alpha_{ji} & 0 & \alpha_{jk} \\ \alpha_{ki} & \alpha_{kj} & 0 \end{bmatrix} \begin{bmatrix} y_i \\ y_j \\ y_k \end{bmatrix} + \begin{bmatrix} u_i \\ u_j \\ u_k \end{bmatrix}$$

$$y = Ay + u$$

donde

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

El problema con este sistema es que, bajo un corte transversal, hay más parámetros que observaciones (3 observaciones y 7 parámetros, incluido el parámetro de dispersión). Puede imponerse simetría reduciendo el número de parámetros, pero, de todas formas, no podrá estimarse. Por lo tanto, es necesario buscar una solución alternativa para poder estimar la dependencia subyacente en los datos.

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

$$A = \rho W$$

donde ρ es un parámetro por estimar que captura el efecto promedio de inte-

racción entre todas las regiones y W es la matriz de pesos espaciales. Los elementos de W son:

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

tal que se ha intercambiado a los parámetros α (parámetros originales del modelo) por W (coeficientes exógenos al modelo) y el modelo queda expresado como $y = \rho Wy + u$, donde Wy es interpretado como el rezago espacial de y por analogía a series temporales.

Generalizando, la matriz W será de orden $n \times n$, donde n es el tamaño muestral. Cada elemento de W es denominado peso espacial, W_{ij} . Los pesos espaciales capturan la vecindad, y son diferentes de cero cuando las regiones i y j son consideradas vecinas. Por convención, ninguna región puede ser vecina de sí misma, lo que da como resultado que la diagonal principal de W contenga todos sus elementos iguales a cero, $W_{ii} = 0$.

La forma más tradicional para construir la matriz es bajo criterios geográficos, siguiendo la primera ley de la geografía (Tobler, 1970): "Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes". Entre los criterios geográficos, puede definirse vecindad por contigüidad, por alguna función de distancia, por k vecinos más cercanos o alguna combinación de estas opciones.

ESPECIFICACIONES HABITUALES

Una forma de aproximarse a la estimación de un modelo econométrico con datos espaciales es mediante un modelo no espacial, usualmente estima-

IMÁGENES
SATELITALES
SE USAN
COMO INSUMOS
EN PREDICCIONES

GEO BIG DATA

LA TÉCNICA PREDICTIVA
PONE ÉNFASIS EN LA
INTERCONECTIVIDAD Y
LA GEOLOCALIZACIÓN

do por mínimos cuadrados ordinarios (MCO). Usando los residuos del modelo no espacial, puede contrastarse la posible presencia de dependencia entre las observaciones con el test I de Moran, en el que la hipótesis nula del test es no autocorrelación espacial. El rechazo de la hipótesis nula abre la especificación a elementos espaciales.

Entre los modelos espaciales más habituales puede mencionarse el modelo de rezago espacial (SLM):

$$y = \rho Wy + \alpha + X\beta + u \quad (1)$$

donde la variable dependiente es un vector ($n \times 1$), Wy representa el efecto de interacción espacial endógeno y ρ es el coeficiente de rezago espacial, u es un vector ($n \times 1$) de unos asociados a la constante α , X es una matriz de variables explicativas de orden ($n \times K$), β es un vector de parámetros desconocidos de orden ($K \times 1$) y $u = (u_1, u_2, \dots, u_n)'$ es un vector de términos aleatorios de dimensión ($n \times 1$), donde u_i es independiente e idénticamente distribuido para todo i con media cero y varianza σ^2 .

Otro modelo espacial habitual es el modelo de error espacial (SEM):

$$y = \alpha + X\beta + u, u = \lambda Wu + \varepsilon \quad (2)$$

donde Wu es el efecto de interacción espacial en el término de error o efecto espacial residual y el parámetro λ es el coeficiente de autocorrelación espacial de los errores.

La combinación entre ambos modelos (SLM+SEM) da lugar al modelo SARAR (*spatial autoregressive with autoregressive error*). Otros modelos más complejos pueden ser estimados incorporando rezagos espaciales en las variables explicativas (WX), conocidos como modelos de Durbin, o dinamizando el modelo de corte transversal en modelos

espacio-temporales. Un mayor detalle de estas opciones puede consultarse en Herrera (2017).

El modelo más ampliamente utilizado incorpora el rezago espacial endógeno y su preeminencia tiene al menos tres razones. Por una parte, bajo un esquema económico de teoría de juegos, con funciones de pago cuadráticas, el SLM puede justificarse como una aproximación a la función de reacción simultánea entre n individuos (Brueckner, 2003). Otra razón de peso es por cuestiones de econometría teórica, ya que la inclusión de un término espacial endógeno evita la inconsistencia del estimador, pero si se omite el rezago espacial en los errores, solo existe una pérdida de eficiencia. Por último, existe un motivo empírico para estimar coeficiente de interacción endógeno, el modelo estimado brinda una estimación del *spill-over* espacial que es muy atractiva en términos aplicados.

No es necesario tener una gran base de datos para estimar estos efectos espaciales. A continuación, se ejemplificará la estimación de estos modelos bajo una base de datos de tamaño estándar.

COMERCIO Y EFECTOS ESPACIALES

El modelo gravitacional en comercio permite predecir los flujos bilaterales entre países y es un modelo que ha sido ampliamente utilizado en los trabajos empíricos dentro del área (Feenstra, 2004; Helpman, Melitz y Rubinstein, 2008; Krisztin y Fischer, 2015). Utilizaré una especificación simple de este modelo para destacar que la naturaleza interdependiente del flujo comercial necesita ser considerada de una forma alternativa al tratamiento tradicional.

Los datos utilizados provienen del

trabajo realizado por Martin, Mayer y Thoenig (2008), en donde se utiliza el modelo gravitacional para medir el impacto sobre el flujo comercial de un conjunto de componentes separados entre variables no relacionadas a política (contigüidad, distancia bilateral, semejanza en lenguaje y vínculo colonial) y otras variables de tinte político (acuerdos comerciales y régimen comunista). Los autores utilizan una gran base de datos de comercio internacional entre 240 países desde el año 1950 al 2000. Para este ejercicio se seleccionaron 16 países del continente americano para los últimos 5 años observados (1996-2000).¹ Dentro de este subconjunto de datos, las variables políticas no muestran variabilidad, por lo que se trabajó con las variables no políticas. El modelo gravitacional no espacial por estimar es el siguiente:

$$\ln T_{ijt} = \beta_1 \ln GDP_{it} + \beta_2 \ln GDP_{jt} + \beta_3 \text{cont}_{ij} + \beta_4 \text{cont}_{ij} \times \ln \text{dist}_{ij} + \beta_5 \text{leng}_{ij} + u_{ijt}$$

donde GDP_{it} , GDP_{jt} son los productos brutos internos del país de origen i y del país de destino j ; cont_{ij} es una variable ficticia que toma valor 1 si los países son limítrofes y 0 en otro caso, dist_{ij} es la distancia bilateral entre capitales y leng_{ij} es un índice de semejanza en el lenguaje.

En la especificación original de los autores, estos utilizan el criterio de contigüidad limítrofe como una variable que permite capturar el efecto de proximidad geográfica y de esta forma controlar los efectos espaciales posiblemente presentes en la base de datos.

El foco de interés en este ejercicio es sobre las variables de control que buscan capturar el efecto de la vecindad geográfica, cont_{ij} y Indist_{ij} . En nuestra especificación se incluye el logaritmo de la distancia entre capitales de países limítrofes usando un término de interac-

ción entre el criterio de contigüidad y la distancia medida en kilómetros. Es esperable que el coeficiente de contigüidad sea positivo y refleje el mayor flujo comercial entre países limítrofes, aunque la distancia entre capitales debería tener un impacto y afectar este flujo comercial de forma negativa.

Los modelos estimados se presentan en el cuadro 1. La cantidad de valores cero es de 103, con 80 de estos valores producto de la reestructuración necesaria para la estimación espacial, ya que debe considerarse el flujo tomando como origen y destino al mismo país. Dependiendo del objetivo de la investigación, estos 80 valores podrían ser diferentes de cero tal que se capture el flujo interno a cada país. Aquí se sigue la estrategia de considerar este flujo igual a 0, supuesto habitual para este tipo de modelos también conocidos como de origen-destino (LeSage y Pace, 2009, cap. 8). En definitiva, existen solo 23 valores sin flujos entre países (1,8% de la base) y esto permite una estimación simple para estos datos.

La matriz de contacto W fue construida siguiendo el criterio de contigüidad de los países de destino, estandarizada por fila. Esta construcción captura formalmente la noción de que los flujos asociados con un destino mejoran o empeoran según la atracción de los países de destino vecinos (para mayor detalle, ver LeSage y Pace, 2009).

CUADRO 1
ESTIMACIONES ALTERNATIVAS DEL MODELO DE GRAVEDAD

MODELOS	MCO	SLM	SEM	SARAR
ln (GDP origen)	1,208 ***	0,685 ***	1,064 ***	0,686 ***
ln (GDP destino)	0,633 ***	0,609 ***	0,638 ***	0,619 ***
Contig	10,733 ***	8,753 ***	7,925 ***	8,732 ***
Contigxln(dist)	-1,052 ***	-0,804 ***	-0,726 ***	-0,797 ***
Índice leng	0,146	-0,043	-0,286	-0,102
W x ln T	0,365 ***	0,350 ***		
W x u	0,476 ***	0,089		
Constante	-17,354 ***	-12,605 ***	-15,602 ***	-12,663 ***
Controles origen-destino	sí	sí	sí	sí
N	1.280	1.280	1.280	1.280
corr ² (ln T, ln T)	82,374	85,452	81,631	85,396
IdeMoran	13,452 ***			
L M error robusto	25,625 ***			
L M lag robusto	17,442 ***			

Nota: *** denota significancia al 1%. Estimación espacial bajo G2SLS con errores robustos.

Fuente: Elaboración propia.

Los resultados muestran la diferencia entre el método de estimación tradicional con respecto al espacial. En ambos casos se reconoce el impacto de los países limítrofes en la especificación del modelo al incorporar variables geográficas. En el modelo básico (primera columna), la especificación descansa en el supuesto de observaciones *i. i. d.* (independientes e idénticamente distribuidas) aproximando la característica espacial de los datos mediante una variable *proxy*. Bajo esta especificación, los diferentes test espaciales detectan que los errores no cumplen el supuesto de independencia y el modelo necesita ser reespecificado.

En los modelos espaciales (columnas 2, 3 y 4) hay un reconocimiento de la dependencia espacial entre los datos mediante la incorporación de la matriz de contigüidad, además de incluir variables geográficas. En todos los modelos, puede observarse que los elementos es-

paciales son significativos y su inclusión es relevante. El modelo SARAR brinda mayor información y muestra que el efecto por considerar es el rezago espacial endógeno. Es decir, el modelo MCO omite una variable que genera inconsistencia de las estimaciones.

En numerosas publicaciones se intenta controlar la interacción espacial de la forma presentada en la estimación MCO sin avanzar más allá. Las estimaciones del cuadro 1 evidencian que este tipo de estrategia no es siempre útil y que debe realizarse un mayor esfuerzo. Además, en dichas publicaciones, los test espaciales no suelen presentarse como evidencia de control efectivo.

SINTONÍA FINA

La mayor disponibilidad de información georreferenciada se presenta como una oportunidad nunca antes ex-

perimentada para realizar análisis empíricos. Sin embargo, las herramientas convencionales en econometría no logran aprovechar la riqueza de la información. En este artículo se presentan algunas herramientas de econometría espacial que pueden ser aprovechadas para capturar la complejidad de la información.

Los datos espaciales presentan características particulares al provenir de diferentes fuentes de observación. La agregación altera la variabilidad e impacta en la inferencia por realizar. También el soporte de la información geográfica es importante cuando se combinan diferentes tipos de datos espaciales. El avance en la información permitirá gradualmente identificar y trabajar a escalas cada vez más desa-

gregadas con fuentes de información alternativas, de manera tal que trabajos empíricos como el presentado en este escrito serán cada vez más detallados y podrán identificarse actores locales que realizan exportaciones o importaciones de bienes.

Este mayor detalle desafiará las técnicas estándares de análisis estadístico, lo que requiere el reconocimiento de la particularidad de los datos; en este punto es donde la econometría espacial adquiere relevancia y posibilita el tratamiento de la información. El uso de este instrumental es independiente de la dimensión de la base de datos que seguramente tenderá a ser del tipo *geo big data*. El ejemplo empírico evidencia que la estrategia clásica no captura la complejidad de la interacción espacial. ✓

NOTAS

¹ Los países son Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras,

México, Panamá, Perú, Paraguay, El Salvador, Uruguay y Venezuela.

BIBLIOGRAFÍA

- Anselin, L. 1988.** *Spatial Econometrics: Methods and Models*, Volume 4. Dordrecht: Kluwer Academic Publishers.
- . **2010.** "Thirty Years of Spatial Econometrics". *Papers in Regional Science*. 89 (1): 3-25.
- Arribas-Bel, D. 2014.** "Accidental, Open and Everywhere: Emerging Data Sources for the Understanding of Cities". *Applied Geography*. 49: 45-53.
- Brueckner, J. 2003.** "Strategic Interaction Among Governments: An Overview of Empirical Studies". *International Regional Science Review*. 26 (2): 175-188.
- Cressie, N. 2015.** *Statistics for Spatial Data*. Edición revisada. Nueva York: Wiley.
- Elhorst, J. 2014.** *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Heidelberg: Springer.
- Feenstra, R. 2004.** *Advanced International Trade: Theory and Evidence*. Princeton: Princeton University Press.
- Helpman, E., Melitz, M., y Rubinstein, Y. 2008.** "Estimating Trade Flows: Trading Partners and Trading Volumes". *The Quarterly Journal of Economics*. 123

(2): 441-487.

Herrera, M. 2017. "Fundamentos de econometría espacial aplicada". *MPRA Paper No. 80871*. Munich: University Library of Munich, Universidad de Múnich.

Jean, N., Burke, M., Xie, M. et al. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty". *Science*. 353 (6301): 790-794.

Krisztin, T. y Fischer, M. 2015. "The Gravity Model for International Trade: Specification and Estimation Issues". *Spatial Economic Analysis*. 10 (4): 451-470.

LeSage, J. y Pace, R. K. 2009. *Introduction to spatial econometrics*. Boca Ratón: Chapman and Hall/CRC Press.

Martin, P., Mayer, T. y Thoenig, M. 2008. "Make Trade Not War?". *The Review of Economic Studies*. 75 (3): 865-900.

Patino, J. y Duque, J. 2013. "A Review of Regional Science Applications of Satellite Remote Sensing in Urban Settings". *Computers, Environment and Urban Systems*. 37:1-17.

Tobler, W. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region". *Economic Geography*. 46: 234-240.