



Munich Personal RePEc Archive

Spatial econometrics: an application to obesity indicators in Argentinian provinces

Viego, Valentina and Temporelli, Karina

CONICET - Departamento de Economía, Universidad Nacional del Sur

16 November 2010

Online at <https://mpra.ub.uni-muenchen.de/26878/>
MPRA Paper No. 26878, posted 21 Nov 2010 21:51 UTC

Econometría espacial: una aplicación a los problemas de sobrepeso y obesidad en las provincias de Argentina

Dra. Valentina Viego (UNSUR)
Dra. Karina Temporelli (CONICET, UNSUR)

Resumen

Los modelos espaciales, a pesar de mostrar una expansión vigorosa en los últimos 30 años en el campo de la econometría, tienen aún escasa difusión en los trabajos empíricos en Argentina focalizados en aspectos cuantitativos de alcance regional. Este trabajo tiene por objetivo presentar los lineamientos principales del análisis econométrico espacial aplicando sus herramientas al estudio de los factores que inciden en la problemática de sobrepeso y obesidad en la población utilizando datos provinciales de la Encuesta Nacional de Factores de Riesgo. Los resultados muestran que las herramientas econométricas convencionales son insuficientes para identificar con claridad el impacto individual del ingreso, la pobreza o la educación sobre el fenómeno del sobrepeso.

Abstract

Despite their vigorous expansion in the last 30 years in the field of econometrics, spatial models are still not widely disseminated in the empirical literature in Argentina focused on quantitative aspects of regional scope. This paper aims to present the central lines of spatial econometric analysis applied to the problem of overweight in the population and its determinants. We use provincial data from the National Survey on Risk Factors. The results show that conventional econometric tools are insufficient for identifying the individual impact of income, poverty or education on the phenomenon of overweight.

Palabras clave: Autocorrelación espacial, modelos espaciales, determinantes socioeconómicos de la obesidad

Keywords: Spatial dependence, spatial models, obesity´ economic determinants

Códigos JEL: *C21, I19*

Introducción

El principal objetivo de este trabajo es difundir el uso de las técnicas de econometría espacial. Estas técnicas fueron especialmente diseñadas para el análisis empírico de datos y procesos espaciales. Su importancia radica en que, en determinados contextos, su uso es indispensable para obtener resultados válidos. Sin embargo, el conocimiento y la utilización de estas técnicas es aún incipiente en la economía y especialmente en Argentina.

A fin de contribuir a su difusión, se describen algunas de las herramientas de econometría espacial, tomando como ejemplo datos de sobrepeso y obesidad en las provincias argentinas.

La primera parte se dedica a justificar el uso de técnicas específicas en el análisis de datos espaciales, tanto por las características propias de éstos como por los fenómenos que se trata de comprender, especialmente por la existencia de autocorrelación espacial. La segunda sección presenta un análisis exploratorio de la distribución espacial de los indicadores de problemas de peso en la población. El principal objetivo de ese apartado será determinar, al menos de modo tentativo, si los problemas de peso se distribuyen de forma aleatoria o si

presentan algún tipo de asociación espacial. Luego se exponen algunos modelos sencillos que explican los factores que inciden en la problemática de salud asociada al peso, que se contrastarán econométricamente tratando de recoger las pautas de asociación, en caso de que existan identificadas en el apartado anterior. Finalmente se repasan las conclusiones que se pueden derivar de los resultados econométricos, delimitando su alcance y limitaciones.

1. Técnicas de análisis espacial

Un aspecto común en el uso de métodos estadísticos espaciales es la necesidad de relajar el supuesto, a menudo restrictivo, de independencia entre las observaciones, bastante frecuente, por otro lado, cuando se tratan muestras recogidas con referencia a puntos o regiones localizadas en el espacio. Es por ello que el supuesto de independencia es reemplazado por otros que ofrecen un esquema que considera relaciones de independencia entre las observaciones.

En los últimos 30 años esta corriente econométrica ha tomado un vigor notable, especialmente a partir de la difusión de bases y paquetes de gestión de datos geo-referenciados (como *ArcGis*, *GRASS*, *MapWindow*, entre otros) y, más generalmente, con la propagación de los sistemas de información geográfica (GIS, por sus siglas en inglés)¹. Su aplicación tomó impulso además con el desarrollo de programas que contienen las herramientas específicamente diseñadas para tratar datos espaciales. *GeoDa*, desarrollado por el *Center for Geospatial Analysis and Computation* bajo la dirección de Luc Anselin (Universidad de Illinois) es el *software* más ampliamente difundido, debido a su libre disponibilidad. Hay variantes más completas, pero de acceso condicionado (*Space-Stat*, que emplea GAUSS como herramienta de cálculo, supervisada por el mismo centro de investigación; *Spatial-Econometrics* y *Spatial-Statistics*, basados en *MatLab* y supervisados por James LeSage). Esta proliferación de paquetes ha impulsado, además, el desarrollo de rutinas específicas en programas de análisis estadístico y econométrico más generales, como *R* y *Stata*², aunque su uso está menos difundido debido a una mayor complejidad de gestión.

a. Especificidad de los procesos y datos espaciales

El trabajo con datos y fenómenos localizados tiene algunas características específicas que impiden la aplicación de las técnicas econométricas tradicionales, tales como las que se basan en los desarrollos basados en datos temporales. Teniendo en cuenta este detalle, Paelinck y Klaassen (1979) destacaron cinco peculiaridades que obligaban al uso de lo que desde entonces se conoce como econometría espacial:

- i. el papel de la interdependencia espacial
- ii. la asimetría en las relaciones espaciales
- iii. la importancia de factores explicativos radicados en otras localizaciones
- iv. la diferencia entre interacción *ex ante* y *ex post*
- v. la modelización explícita del espacio

La existencia de autocorrelación o dependencia espacial subyace en todos estos fenómenos. Así, mientras que la aleatoriedad espacial se refiere a valores observados en una localización que no dependen de valores observados en las localizaciones vecinas, cuando esta condición no se cumple aparece autocorrelación (o dependencia) espacial.

¹ La notable expansión de literatura que considera, aunque sea tangencialmente, el análisis espacial cuantitativo condujo al *Journal of Economic Literature* a crear el apartado C21 con el objeto de incluir en él este tipo de trabajos (modelos espaciales).

² La ventaja de *R* por sobre el resto de los programas específicos o adaptados de econometría espacial es que admite el tratamiento de datos de panel espaciales.

La autocorrelación espacial tiene dos causas fundamentales. Por un lado, la existencia de procesos espaciales, como se mencionó anteriormente. Y, por otro, por errores de naturaleza diversa como datos equívocos, desajuste entre el área y el fenómeno que se pretende explicar, etc. (Anselin, 1988).

Más allá de la existencia de dependencia espacial, se debe destacar la elevada heterogeneidad que suelen presentar los datos localizados (e.g. disponibilidad de recursos entre municipios o entre provincias).

Por otro lado, si bien es cierto que tanto la autocorrelación como la heterogeneidad no son exclusivos de los datos espaciales al presentarse también en series temporales, la autocorrelación en series de tiempo sólo toma una única dirección (hacia atrás en el tiempo) mientras que con datos espaciales puede ser multidireccional. Además, la heterogeneidad suele manifestarse en el análisis de regresión en forma de varianza no constante, que puede ser confundida con autocorrelación espacial, complicando el análisis (Anselin y Bera, 1998).

En suma, si se estima un modelo econométrico en el que existe dependencia espacial y ésta no es eliminada o modelizada correctamente, el ajuste, la inferencia y varios tests perderán confiabilidad. Dependiendo de la naturaleza de la autocorrelación, los estimadores podrán ser sesgados, inconsistentes o ineficientes.

2. Dependencia espacial: diagnóstico y técnicas de tratamiento

Tal como se mencionó en el punto anterior, el carácter multidireccional que toma la autocorrelación es uno de los rasgos distintivos del análisis espacial respecto del basado en series de tiempo. El modo más común de controlar este tipo de interacción es mediante el uso de matrices de ponderación o de ordenamiento espacial, W , en las que cada elemento w_{ij} registra la interacción entre observaciones.

Los elementos de la matriz pueden ser de diversa naturaleza. Por ejemplo, si W se expresa en términos de distancias, w_{ij} suele representar la inversa de la distancia entre las observaciones localizadas en i y j , bajo el supuesto de que en la mayoría de los procesos espaciales la interacción depende negativamente de la distancia. En otros casos, W representa una matriz de contactos, donde w_{ij} toma valores binarios, en función de si hay o no contigüidad (existencia de frontera común o si la distancia que las separa se encuentra por debajo de cierto umbral tolerado) entre las observaciones³.

Por razones operativas, la matriz W suele estar estandarizada por filas, es decir que cada elemento se suele dividir por la sumatoria de los elementos de la fila correspondiente.

El caso aquí presentado se basa únicamente en criterios de contigüidad y no de distancia, ya que este último calcula la distancia geodésica de los centros o centroides de los polígonos (provincias). Ocurre que en el caso provincial las capitales o centros urbanos de mayor tamaño no necesariamente se ubican próximos a los centroides de los polígonos, de modo que la distancia como indicador de cercanía no resulta enteramente satisfactoria. Por ello, el análisis se condujo a partir de indicadores de contigüidad.

Los paquetes estadísticos específicos que contienen técnicas de econometría espacial pueden derivar relaciones de vecindad entre polígonos. Se trata de evaluar si los polígonos (en este caso, provincias) comparten una frontera, en cuyo caso son considerados vecinos. Si bien parece un concepto trivial, la especificación de las relaciones de vecindad puede influir en el

³ Si bien los casos anteriores corresponden a los más utilizados, suele usarse en algunos casos matrices calculadas en función de la longitud de la frontera compartida.

resultado del análisis posterior⁴. Hay 2 variantes disponibles para identificar contigüidad: torre y reina (cuyos nombres provienen de los movimientos en las piezas de ajedrez). La torre puede moverse sólo hacia los polígonos que comparten una frontera de longitud positiva con el polígono actual. En cambio la reina puede desplazarse hacia cualquier polígono que comparta frontera de al menos un punto (esquina o vértice). Por ello, la torre es una definición de contigüidad más restrictiva que la reina.

2.1 Análisis exploratorio

En este punto se presentan algunas herramientas utilizadas para detectar la autocorrelación espacial en el análisis descriptivo y en el contexto de regresión.

Si bien potencialmente la autocorrelación espacial puede ser positiva o negativa, prácticamente no se registran casos empíricos de dependencia espacial negativa. Por lo que, en general, se trata de detectar la existencia de autocorrelación espacial positiva, que se corresponde con la existencia de valores similares en localizaciones próximas.

El estadístico más utilizado para la detección de autocorrelación espacial global es la I de Moran, que se define como:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} x_i x_j}{\sum_i x_i^2}$$

Donde N es el número de observaciones, w_{ij} es el elemento de la matriz de ponderaciones espaciales correspondiente a par de observaciones i, j y x_i y x_j son las observaciones (centradas en sus medias). Los valores de I se ubican en el rango $[-1, 1]$, desde la dispersión perfecta a la correlación perfecta. El valor esperado de I bajo la hipótesis nula de ausencia de autocorrelación espacial es:

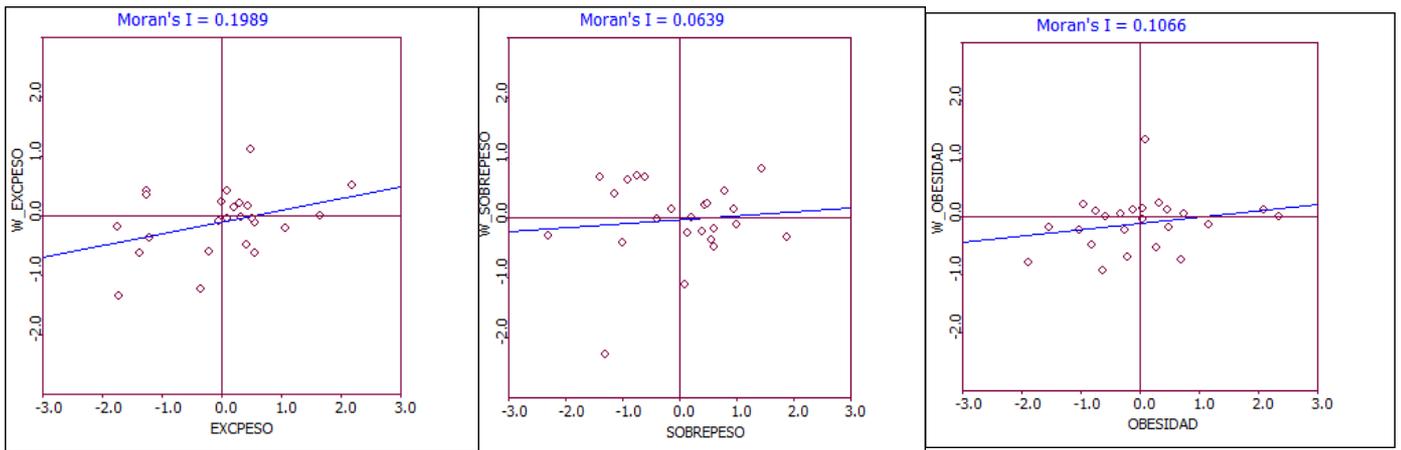
$$E(I) = \frac{-1}{N-1}$$

Así, un valor de I cercano a 0 indica un patrón espacial aleatorio.

El diagrama de dispersión de Moran permite su visualización (Anselin, 1995, 1996). Las variables están estandarizadas, de modo que las unidades en el gráfico representan desvíos estándar. Los 4 cuadrantes en el gráfico ofrecen una clasificación de los 4 tipos de autocorrelación espacial: alta-alta (superior derecha), baja-baja (inferior izquierda) en el caso de autocorrelación espacial positiva; alta-baja (inferior derecha) y baja-alta (superior izquierda) en el caso de autocorrelación espacial negativa. La pendiente de la línea de regresión es el I de Moran, informado en cada gráfico⁵. A continuación se presenta la aplicación de las herramientas de análisis espacial a indicadores de exceso de peso a nivel provincial y sus determinantes en Argentina.

⁴ Los errores en la geometría de los polígonos pueden interferir en la identificación de contigüidad entre polígonos en el archivo.

⁵ Cada diagrama fue construido en base a la medida de contigüidad basada en el criterio de reina. El indicador y gráficos correspondientes no difieren sustancialmente cuando se emplea el criterio de vecindad de torre.



El análisis espacial global arroja un solo indicador que sintetiza al conjunto del área bajo estudio. En otros términos, supone homogeneidad. Pero aún en ausencia de correlación espacial global, es posible encontrar clusters a nivel local utilizando otros indicadores más apropiados que aprovechan que la I de Moran es la suma de productos cruzados individuales conocidos como “indicadores locales de asociación espacial” (LISA, por sus siglas en inglés). LISA examina si existe agrupamiento en unidades individuales calculando la I de Moran para cada unidad espacial y evaluando la significatividad estadística de cada I_i (Anselin, 1995). De la expresión anterior se obtiene que:

$$I_i = \frac{x_i N}{\sum_i x_i^2} \sum_j w_{ij} x_j$$

Luego,

$$I = \sum_i \frac{I_i}{N}$$

La Tabla 1 resume los resultados de la aplicación de la prueba de autocorrelación espacial local basada en los 2 criterios alternativos de contigüidad.

TABLA 1
Test de autocorrelación espacial local, LISA

Criterio de contigüidad	Exceso de peso	Sobrepeso	Obesidad
<i>Queen</i> (reina)	Buenos Aires**	Misiones**	Chubut**
	Tierra de Fuego***	Córdoba**	Tierra del Fuego***
		Río Negro**	
<i>Rook</i> (torre)	Tierra de Fuego***	Tierra del Fuego***	Corrientes**
		Misiones**	Chubut**
		Córdoba**	Tierra del Fuego***

** significatividad al 5%

*** significatividad al 1%

En suma, los estadísticos que diagnostican la presencia de autocorrelación espacial global advierten que este fenómeno no presenta particular contundencia en las problemáticas de peso en la población. Sin embargo, un examen más detallado revela la existencia de eventos de autocorrelación espacial localizada en algunas provincias como Buenos Aires, parte del litoral (Corrientes, Misiones) y de la región patagónica (Tierra del Fuego, Chubut, Río Negro).

2.2 Diagnóstico en análisis de regresión

En presencia de autocorrelación espacial, la utilización de técnicas convencionales, como mínimos cuadrados ordinarios, puede invalidar las pruebas de bondad de ajuste, significatividad individual y, en general, la inferencia afectando en algunos casos las propiedades de insesgadez y consistencia y, en otros, la de varianza mínima de los parámetros. En este punto se presentan las formas que puede adoptar la autocorrelación espacial en el análisis de regresión y los tests disponibles para detectarla.

La dependencia espacial se puede dar en la variable dependiente, en el término de error o en ambos a la vez. Al plantearse un modelo como

$$y = X\beta + \varepsilon \quad (1)$$

$$\text{con } E(\varepsilon) = 0; E(x_i, x_j) = 0, E(\varepsilon'\varepsilon) = s^2$$

si la variable dependiente está autocorrelacionada espacialmente, el modelo correcto será:

$$y = rWy + X\beta + \varepsilon \quad (2)$$

Donde el valor de y en cada punto del espacio se relaciona con los demás valores de y mediante el término autorregresivo espacial rWy . Cuando la autocorrelación espacial ocurre en la variable dependiente, las estimaciones que ignoran este fenómeno suelen arrojar estimadores sesgados e inconsistentes. Estos modelos se conocen como tipo "lag" por el carácter autorregresivo espacial de la dependiente.

Es posible, asimismo, que la dependencia espacial ocurra en los términos de error:

$$y = X\beta + \mu \quad (3)$$

$$\text{con } \mu = W\mu + \varepsilon \quad (3.a)$$

$$\mu = W\varepsilon + \varepsilon \quad (3.b)$$

En (3.a) el término de error sigue un proceso autorregresivo espacial mientras que en (3.b) se trata de una media móvil espacial. La consecuencia de no considerar estas modelizaciones del término de error aplicando técnicas econométricas tradicionales es la pérdida de eficiencia de los estimadores, por lo que las inferencias de allí extraídas no serán válidas⁶.

⁶ Es posible que la autocorrelación se presente tanto en la variable dependiente como en el término de perturbación, con consecuencias más graves sobre los estimadores. Sin embargo, no se cuenta a la fecha de un paquete econométrico que pueda modelizar ambos tipos de autocorrelación espacial. El software empleado en este trabajo permite evaluar la existencia

A continuación se presentan algunas de las pruebas empleadas para detectar dependencia espacial en el contexto de análisis de regresión y aplicadas a los indicadores provinciales de sobrepeso. Para ello, se propone un modelo uniecuacional sencillo donde el indicador de exceso de peso depende de una serie de factores individuales y socioeconómicos:

$$P_i = f(\text{conductas individuales}_i, \text{factores socioeconomicos}_i, \varepsilon_i)$$

donde i representa a la unidad espacial, P es el indicador testigo de problemas de peso y ε es el término de perturbación aleatorio. En este caso, P es aproximado con 3 variables: EXCPESO, SOBREPESO y OBESIDAD. Mientras que la primera considera el porcentaje de población con nivel de peso superior al saludable, las 2 restantes desdoblan este indicador entre tasa de población con sobrepeso moderado, por un lado y, severo (es decir, con síntomas de obesidad), por otro. Las 3 medidas permiten un panorama más detallado de la problemática de (sobre)peso en la población.

Por su parte, las conductas individuales fueron aproximadas con el porcentaje de población que no realiza en forma regular actividad física (ACTIVFISBAJA) y los factores socioeconómicos fueron captados en un conjunto de variables: producto per cápita (PBIPC), porcentaje de población con ingresos bajos (YBAJO), tasa de población que completó el nivel secundario (SEC) y proporción de habitantes que poseen cobertura médica (COBERT). La definición de las variables y las fuentes se detallan en el Anexo. La muestra empleada contiene información de estas variables a nivel provincial y proviene de la Encuesta Nacional de Factores de Riesgo realizada en centros urbanos de más de 5000 habitantes durante 2005. El caso aquí presentado fue analizado utilizando el paquete *GeoDa* versión 0.9.5.i (*alpha*).

El test de Moran indica la presencia de autocorrelación espacial pero no hace referencia a la naturaleza de ésta. Este test se aplica sobre los residuos de la regresión mínimo-cuadrática. Requiere que los términos de error sigan una distribución normal (Anselin y Rey, 1991)⁷.

El resto de los tests, que detectan además el tipo de autocorrelación espacial presente, se basan en el multiplicador de Lagrange (LM) y requieren que el residuo siga una distribución normal. Se dispone de 2 pruebas que localizan dependencia espacial en la dependiente y se distribuyen como una χ^2 con 1 grado de libertad. La hipótesis nula en ambas es que en la ecuación (2) $r = 0$, es decir que no hay autocorrelación espacial en la endógena. Sin embargo, mientras que la primera versión supone que no hay autocorrelación espacial en los términos de error, $l = 0$, la segunda es robusta a ante la presencia de este otro tipo de dependencia espacial.

Algo análogo ocurre con los dos métodos diseñados para detectar autocorrelación espacial en los términos de error; se distribuyen asimismo como una χ^2 con 1 grado de libertad y la hipótesis nula en ambos es que en la ecuación (3) $l = 0$. De igual modo, la primera variante del test supone ausencia de dependencia espacial en la variable dependiente, $r = 0$, mientras que la segunda es robusta a la existencia de autorrelación espacial en y .

Finalmente, como ya se advirtió anteriormente, el test SARMA chequea la existencia de dependencia espacial en la dependiente y en las perturbaciones. En esta prueba la hipótesis nula es que $l = r = 0$, es decir, la ausencia de correlación espacial. Este estadístico se distribuye como una χ^2 con 2 grados de libertad.

de modelos del tipo SARMA, donde la dependencia espacial ocurre simultáneamente en la dependiente y el error sigue un proceso de medias móviles espaciales.

⁷ Si bien el test de Kelejian-Robinson tiene la ventaja, frente al de Moran, de no requerir normalidad en los errores ni linealidad del modelo no está recomendado para muestras pequeñas, como la presentada aquí. Los experimentos realizados por Anselin y Florax (1995) advierten también falta de potencia bajo algunas especificaciones.

En este caso, el objetivo final de este punto no es conseguir un modelo que explique totalmente los problemas de sobrepeso y obesidad en la población a nivel regional, sino averiguar si existe un componente espacial y calibrar su importancia.

En primer lugar se propone un modelo explicativo sencillo de los problemas de peso. Luego se someten a las variables que lo componen a un análisis exploratorio con el fin de indagar si existen pautas de asociación espacial.

TABLA 2
**Diagnóstico de dependencia y heterogeneidad espacial
en indicadores de problemas de peso**

	Exceso de peso	Sobrepeso	Obesidad
CONSTANTE	42.6378 (0.0107)	39.9791 (0.003)	2.6587 (0.7792)
ACTIVFISBAJA	0.0268 (0.7810)	-0.0029 (0.9593)	0.0299 (0.6219)
PBIPC	-0.1541 (0.8009)	-0.0086 (0.8128)	-0.0068 (0.8584)
YBAJO	0.0144 (0.9251)	-0.0536 (0.5595)	0.0681 (0.4807)
SEC	-0.1748 (0.3407)	-0.0661 (0.5421)	-0.1087 (0.3421)
COBERT	0.2212 (0.2284)	0.0029 (0.9786)	0.2183 (0.0641)
logML	-61.3726	-48.9236	-50.0498
AIC	134.745	109.847	112.1
SIC	141.814	116.915	119.68
Índice de multicolinealidad	61.0138	61.0138	61.0138
Jarque-Bera sobre residuos (normalidad)	1.6815 (0.4314)	0.2068 (0.9017)	0.0467 (0.9769)
Breusch-Pagan (varianza constante)	1.5181 (0.9109)	2.6617 (0.7520)	8.9604 (0.1106)
I de Moran	1.4026 (0.1607)	0.8248 (0.4095)	0.9417 (0.3463)
LM (lag)	2.1591 (0.1417)	0.0273 (0.8687)	5.5315 (0.0187)
LM Robust (lag)	3.0019 (0.0832)	0.0134 (0.9079)	8.2062 (0.0042)
LM (error)	0.5112 (0.4746)	0.0479 (0.8267)	0.1020 (0.7495)
LM Robust (error)	1.3454 (0.2446)	0.0339 (0.8537)	2.7767 (0.0956)
SARMA	3.5130 (0.1726)	0.0613 (0.9697)	8.3082 (0.0157)

Entre paréntesis se expresa el *p-value* asociado al estadístico que evalúa la hipótesis nula correspondiente.

En negrita se destacan los coeficientes y estadísticos significativos al 5% y en gris los significativos al 10%.

El modelo tiene como variable dependiente a la tasa de población con exceso de peso (en 3 variantes: la proporción de población considerada obesa, la tasa de población con sobrepeso que no alcanza el status de obesidad y la suma de los anteriores). Las variables explicativas son el porcentaje de población con bajos niveles de actividad física (ACTIVFISBAJA), el producto bruto *per cápita* (PBIPC), la proporción de población con ingreso bajo (YBAJO), el

porcentaje de población adulta que ha completado el nivel medio de estudios (SEC) y la tasa de población con cobertura médica (COBERT). En el Anexo se detalla la definición y fuente de las variables.

Se supone que conductas sedentarias inciden positivamente sobre los problemas de peso. Inversamente, el nivel educativo colabora en reducir esta problemática en la salud de la población. Respecto del ingreso, la literatura considera 2 efectos separados: uno que indica que los problemas de peso aumentan a medida que aumenta el ingreso *per cápita*, debido a que en los tramos superiores de ingreso la publicidad adquiere operatividad como determinante del patrón de consumo de alimentos. Ello podría impulsar el consumo de alimentos ricos en grasas. El segundo efecto ocurre en estratos inferiores de ingreso, donde los problemas de peso aparecen no tanto como respuesta a pautas publicitarias, sino como conducta racional a restricciones presupuestarias severas que favorecen el consumo de alimentos de menores precios relativos, pero ricos en calorías (carbohidratos, fiambres, arroz, polenta). Finalmente, respecto del nivel de cobertura médica, el signo esperado es ambiguo; por un lado, la cobertura médica puede ampliar el acceso de la población a chequeos médicos y la divulgación de conductas alimenticias más saludables, pero, por otro, podría favorecer dietas poco saludables en los sujetos asegurados, además, ante posibles eventos de salud.

En el caso de exceso de peso, si se emplea la matriz de vecindad basada en el criterio “torre”, el test de Moran indica dependencia espacial al 5%.

El ajuste en los distintos modelos estimados es bastante bajo. Si bien con la evidencia disponible se pueden aceptar las hipótesis de homocedasticidad y normalidad de los residuos, las variables explicativas no resultan significativas probablemente debido a la presencia de multicolinealidad entre los regresores⁸. Además, algunos de los tests de dependencia espacial resultan significativos (especialmente los aplicados a obesidad), por lo que la significatividad y el ajuste del modelo podrían estar cuestionados. Esto amerita la aplicación de técnicas de análisis espacial.

2.3 Tratamiento de la dependencia espacial

En este punto se presentan las principales formas de modelizar la autocorrelación espacial, conocidas como modelos tipo “lag” y el modelo de error autorregresivo (AR) espacial.

Dada la simultaneidad subyacente en el modelo lag y la presencia de errores no esféricos en el modelo de error AR, no es posible aplicar mínimos cuadrados ordinarios, sino que se debe acudir a otras técnicas, como la estimación por máxima verosimilitud (en adelante, MV) si se cumple la hipótesis de normalidad o la utilización de variables instrumentales. En el caso de aplicarse MV los parámetros no reciben la misma interpretación que los estimadores mínimo-cuadráticos. Asimismo, aunque el software empleado arroja el estadístico tradicional de bondad de ajuste, R^2 , éste no tiene el mismo significado que bajo mínimos cuadrados ordinarios ni es comparable entre modelos.

Con todo, las medidas de ajuste disponibles, como logML, los criterios de información de Akaike (AIC) y Schwartz (SIC), resultan comparables de un modelo a otro. El ajuste será mejor cuanto mayor sea el logMV y menores sean AIC y SIC. Se debe tener en cuenta la advertencia de Anselin (1992) sobre la tendencia de AIC y SIC a favorecer los modelos de error AR frente a los lag.

⁸ El cociente entre los autovalores máximo y mínimo de la matriz $X'X$ es mayor a 30 en todos los modelos estimados, umbral considerado como signo de problemas severos de multicolinealidad (Dudek, 2005).

TABLA 3
**Determinantes de los problemas de peso en la población:
 modelo espacial autorregresivo**

	Exceso de peso	Sobrepeso	Obesidad
CONSTANTE	-10.5696 (0.5074)	11.8344 (0.2117)	-19.9597 (0.0302)
ACTIVFISBAJA	0.2751 (0.0018)	0.1783 (0.0007)	0.0867 (0.0901)
PBIPC	0.1851 (0.0032)	0.8448 (0.0236)	0.9130 (0.0117)
YBAJO	0.4982 (0.0008)	0.1577 (0.0751)	0.3195 (0.0002)
SEC	-0.3009 (0.0139)	-0.010 (0.4865)	-0.2469 (0.0006)
COBERT	0.4904 (0.0004)	0.0968 (0.2403)	0.3809 (0.0000)
LAMBDA	-1.4499 (0.0000)	-1.4141 (0.0000)	-1.3638 (0.0000)
logML	0.0000	0.0000	0.0000
AIC	12.0000	12.0000	12.0000
SIC	19.0683	19.0683	19.0683
Breusch-Pagan (varianza constante)	1.5364 (0.9088)	3.2629 (0.6595)	2.5366 (0.7710)
Test LR sobre el error espacial	122.7452 (0.0000)	97.8471 (0.0000)	100.0995 (0.0000)

Entre paréntesis se expresa el *p-value* asociado al estadístico que evalúa la hipótesis nula correspondiente.

En negrita se destacan los coeficientes y estadísticos significativos al 5% y en gris los significativos al 10%

La validez del parámetro autorregresivo se testea con el ratio de verosimilitud, LR. Además, cuando se trabaja con muestras pequeñas, se verifica que los resultados de los tests de Wald, W, LR y el multiplicador de Lagrange, sigan el orden $W \geq LR \geq LM$ como prueba indirecta de especificación correcta.

En el caso aquí analizado, se cree que el modelo espacial más apropiado es el de tipo autorregresivo, ya que los estadísticos de Moran sugerían baja dependencia espacial en las variables explicativas, mientras que las variables explicativas (especialmente la tasa de población con ingresos bajos y cobertura) reflejaban mayores guarismos de autocorrelación espacial.

La estimación de modelo de error espacial refleja una mejor bondad de ajuste que la arrojada por MCO (el logaritmo de la verosimilitud y los criterios de información disminuyen). La mejora ocurre sustancialmente al discriminar el exceso de peso en la población entre sobrepeso y obesidad. En particular, el fenómeno de obesidad, que alertaba en MCO la presencia de dependencia espacial en las observaciones, puede ser mejor ajustado. Las variables explicativas adquieren ahora significatividad especialmente como determinantes de exceso severo de peso.

No se descarta, sin embargo, su posible incidencia como factores incidentes en el sobrepeso moderado. La multicolinealidad entre los regresores puede obstaculizar la identificación del impacto separado de la educación, seguro médico e ingreso bajo.

3. Conclusiones

Se aplicó el modelo espacial autorregresivo a las 3 ecuaciones utilizando matrices de ponderaciones basada en contigüidad tipo reina y tipo torre⁹.

La modelización de la heterogeneidad espacial permitió avanzar en el análisis de los factores que inciden en los problemas de (exceso) peso entre la población. Así mientras que las técnicas econométricas convencionales no permiten identificar el impacto individual de variables como el nivel de ingreso, condiciones de vida, nivel educativo y conductas de los sujetos, el modelo espacial autorregresivo sugiere una serie de aspectos que enriquecen la comprensión de los problemas de sobrepeso y obesidad, como:

i) la actividad física rutinaria desfavorece el exceso de peso, especialmente cuando éste no excede el umbral de obesidad. Cuando la obesidad se hace presente, la realización de actividades físicas es menos relevante frente a otros factores, esencialmente ligados a nivel de ingreso, educativo y condiciones de vida.

ii) el nivel de ingreso *per cápita* estimula la aparición de problemas de exceso de peso, especialmente de obesidad. Simultáneamente, niveles de ingreso bajos inciden con más notoriedad en la aparición de eventos de obesidad. La elevada correlación entre la proporción de población con ingreso bajo, secundario completo y cobertura médica no permite identificar el impacto que esta variable (ingreso bajo) tiene sobre el exceso moderado de peso en los sujetos.

iii) la educación permite menguar la aparición de obesidad entre la población.

iv) la posesión de seguros de salud alienta la aparición de problemáticas de exceso de peso, especialmente de obesidad probablemente por el predominio de conductas alimenticias poco saludables producto del riesgo moral *ex ante*¹⁰.

Así, la aplicación de técnicas de econometría espacial ha permitido ilustrar el rol que los factores socioeconómicos tienen en el fenómeno del sobrepeso, superando incluso a las conductas individuales.

No obstante, se deben señalar algunas limitaciones aún existentes en los datos que impiden un mejor aprovechamiento de los modelos espaciales. Las herramientas de diagnóstico y tratamiento tienen propiedades asintóticas, por lo que su potencia está garantizada en muestras grandes. Su aplicación a datos provinciales condiciona los resultados. Ello se refleja en el caso aplicado, la importancia de la dependencia espacial cambia aplicando los tests del multiplicador de Lagrange respecto del test LR sobre el parámetro que capta la dependencia espacial. Esta limitación informativa podría ser soslayada mediante algunas alternativas: a) extendiendo el análisis a provincias de otros países, b) modificando la escala espacial desde provincias a departamentos o municipios. El costo de ello, sin embargo, es que en Argentina se cuenta con un número reducido de indicadores desagregados a nivel municipal o c) adoptando un esquema de datos de panel, que contenga observaciones espacio-temporales. Esta última línea cuenta todavía con un desarrollo incipiente y probablemente se convierta en la ruta que siga el progreso en la econometría espacial Parent y LeSage, 2009, 2010).

Referencias

Anselin L (1988); *Spatial Econometrics: Methods and Models*. Boston, Kluwer Academic.

⁹ Se ensayó asimismo el modelo espacial tipo lag en las ecuaciones. Sin embargo, y tal como se preveía, esta variante no resulta satisfactoria ya que produce heterocedasticidad en los residuos, además de no variar sustancialmente los resultados del modelo convencional estimado por MCO.

¹⁰ Una vez que los pacientes conocen su cobertura disminuyen los incentivos para protegerse de la enfermedad. Esto puede resultar en un cambio de comportamiento por parte del individuo que una vez que adquiere un seguro vuelve más probable el evento contra el cual se aseguró.

- _____ (1992); Space and applied econometrics: Introduction; *Regional Science and Urban Economics*, 22(3): 307-316.
- _____ (1995); Local indicators of spatial association – LISA; *Geographical Analysis*, 27: 93-115.
- _____ (1996); The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association; en M Fischer, H Scholten y D Unwin (eds); *Spatial Analytical Perspectives on GIS*; pp: 111-125; London: Taylor and Francis
- Anselin L y Bera A (1998); Spatial dependence in linear regression models with an Introduction to spatial econometrics; en A Ullah y D. Giles (Eds.); *Handbook of Applied Economic Statistics*, pp. 237-289; New York: Marcel Dekker.
- Anselin L y Florax R (1995) (eds); Small sample properties of test for spatial dependence in regression models: some further results; en L Anselin y R Florax (1995) (eds); *New Directions in Spatial Econometrics*. Berlin: Springer-Verlag.
- Anselin y Rey (1991); Properties of tests for Spatial dependence in linear regression models; *Geographical Analysis*; 23: 112-131.
- Dudek H (2005); The Relationship between a Condition Number and Coefficients of Variation, *Statistical Review*, 52 (1): 75-85.
- Paelinck J y Klaassen L (1979); *Spatial econometrics*; Farnborough: Saxon House.
- Parent O y LeSage J (2009); A space-time filter for panel data models containing random effects; *University of Cincinnati, Economics Working Papers Series*.
- _____ (2010); A Spatial Dynamic Panel Model with Random Effects Applied to Commuting Times; *University of Cincinnati, Economics Working Papers Series*.

APÉNDICE

Definición y fuente de variables

Variable	Definición	Fuente
EXCPESO	Población con índice de masa corporal, $IMC \geq 25 \text{ kg/m}^2$, %	Encuesta Nacional de Factores de Riesgo, 2005
SOBREPESO	Población con $25 \text{ kg/m}^2 < IMC < 30 \text{ kg/m}^2$, %	Encuesta Nacional de Factores de Riesgo, 2005
OBESIDAD	Población con $IMC \geq 30 \text{ kg/m}^2$, %	Encuesta Nacional de Factores de Riesgo, 2005
ACTIVFISBAJA	Población que realiza actividad física con baja frecuencia, %	Encuesta Nacional de Factores de Riesgo, 2005
PBIPC	Producto Bruto Geográfico per cápita, en miles de pesos, 2005	Centro de Estudios Nueva Mayoría
YBAJO	Ingreso mensual menor a \$600	Encuesta Nacional de Factores de Riesgo, 2005
SEC	Población adulta con secundario completo, %	Encuesta Nacional de Factores de Riesgo, 2005
COBERT	Población con cobertura médica, %	Encuesta Nacional de Factores de Riesgo, 2005