

Sobre la espacialidad de los procesos electorales urbanos y una comparación entre las técnicas de regresión OLS y SAM*

Carlos J. Vilalta y Perdomo**

El propósito de este trabajo es demostrar que debido a que algunos fenómenos sociales presentan variación espacial, la modelación de los mismos tiene implicaciones metodológicas importantes, y por lo tanto es recomendable utilizar una técnica de regresión que incorpore la espacialidad de los datos en su diseño. Se siguen dos procedimientos generales: primero se calcula el coeficiente de autocorrelación I de Moran para demostrar la espacialidad del comportamiento electoral en el México urbano entre 1994 y 2000; después se comparan dos técnicas de regresión, OLS y SAM. De la comparación se concluye que la técnica SAM es más robusta que la OLS con base en que la primera obtiene coeficientes de regresión más precisos y detecta efectos regionales estadísticamente significativos aun en la presencia de modelos con problemas de heterodasticidad.

Palabras clave: geografía electoral, comportamiento electoral, electorado urbano, estadística espacial, autocorrelación espacial, análisis de regresión, heterodasticidad.

Fecha de recepción: 10 de agosto de 2004.

Fecha de aceptación: 4 de enero de 2005.

On the Spatial Distribution of Urban Electoral Processes and a Comparison between OLS and SAM Regression Techniques

The aim of this work is to demonstrate two things: 1) That due to the fact that some social phenomena presents spatial variation, its modeling also presents important methodological implications, and therefore 2) it is advisable to use a regression technique that incorporates data spatiality in its design. A two-step process is followed: First, the Moran autocorrelation coefficient is calculated to demonstrate the spatial concentration of electoral behavior in urban Mexico between 1994 and 2000. Secondly, two regression techniques are compared, OLS and SAM. From the comparison it is concluded that the SAM technique is a more robust technique than the OLS, based on that the former is able to compute more accurate regression coefficients and detects significant regional effects even in the presence of models dealing with the heteroskedasticity problem.

* Agradezco a Carlos Urzúa sus sugerencias y explicaciones metodológicas y a Juan Manuel Martínez sus comentarios.

** Profesor investigador en el Tecnológico de Monterrey, campus Ciudad de México. Correo electrónico: carlos.vilalta@itesm.mx.

Key words: electoral geography, electoral behavior, urban electorate, spatial statistics, spatial self-correlation, regression analysis, heterodasticity

Introducción y objetivos

¿Cuáles son las probabilidades de cometer un error del tipo II?; en concreto, ¿hasta qué punto podemos estar razonablemente seguros de que nuestras técnicas estadísticas más complejas, como sería el caso del análisis de regresión, son necesaria y suficientemente robustas contra este tipo de errores? En este documento nos abocamos a demostrar las serias limitaciones que existen al respecto, en particular cuando analizamos ciertos datos que han sido agregados geográficamente.

El análisis de los datos agregados geográficamente o espacializados,¹ comúnmente denominado “análisis espacial”, muestra implicaciones metodológicas notablemente importantes para los científicos sociales que tienen preferencia por la modelación estadística (Vilalta, 2003a). Históricamente hablando, la polémica al respecto se puede considerar reciente. Aunque el tratamiento de datos espacializados empezó a abordarse en revistas científicas (como *Biometrika* y *Journal of the Royal Statistical Society*) en los alrededores de la década de los cincuenta (Moran, 1948; 1950), el tema no comenzó a discutirse sistemáticamente ni a popularizarse en la literatura académica sino en la década de los ochenta y principios de los noventa (Cliff y Ord, 1981; Anselin, 1988; Anselin y Griffith, 1988; O’Loughlin y Anselin, 1991; Flint, 1995; King, 1996). Antes de esos años, los artículos referentes al asunto eran muy ocasionales, pese a que contaban con una excepcional calidad científica (Cliff y Ord, 1971; Ord, 1975).

Para el caso mexicano, la atención metodológica al tratamiento de ese tipo de datos es aún más reciente, pero aparentemente está en proceso de difusión (Vilalta, 2003a y 2004; Fernández-Durán, Poiré y Rojas-Nandayapa, 2004). Esta atención creciente y reciente se debe principalmente a: 1) la difusión en el uso de los Sistemas de Información Geográfica (SIG), 2) los avances del gobierno para facilitar que el público disponga de información estadística actualizada y series de tiempo, y 3) la popularización de las encuestas y mapas en materia sociodemográfica, económica y política.

¹ Como variables que capturan una modificación geográfica del fenómeno en estudio.

Desde tal perspectiva metodológica, este trabajo se centra estrictamente en demostrar las implicaciones estadísticas del uso de datos espacializados en el análisis de regresión. Su propósito principal es ofrecer una demostración empírica de la pertinencia de utilizar técnicas adecuadas con este tipo de datos y de las consecuencias de no hacerlo. Para tal efecto nos hemos enfocado sobre un área particular de investigación: el comportamiento electoral urbano. Si bien la discusión es principalmente metodológica y en concreto estadística, sus implicaciones no están separadas de las explicaciones teóricas. Indistintamente, las preguntas de investigación geográfica son evidentemente teóricas, de ahí que existan técnicas estadísticas específicas y complejas para probar dichas teorías.

En este documento seguimos cierto orden: Primero, a fin de plantear claramente el problema se detecta empíricamente y se mide la variación espacial no aleatoria de los procesos electorales. Segundo, se advierte sobre la importancia de la geografía y respecto a los problemas que conlleva la aplicación de técnicas estadísticas espaciales cuando analizamos fenómenos sociales que presentan, precisamente, variaciones geográficas. Adelantamos concretamente que los fenómenos de tipo estadístico que resultan son la autocorrelación y la heterogeneidad espaciales.² En tercer orden y de manera pragmática, hacemos una comparación entre dos técnicas de regresión diferentes: una ordinaria (OLS, Ordinary Least Squares) y la otra espacial (SAM, Spatial Autoregressive Model), que llevan a resultados diferentes. Lo anterior conduce a demostrar de manera circular la utilidad de aplicar técnicas de estadística espacial cuando los datos así lo requieren.

Planteamiento del problema estadístico

Esta parte se divide en dos secciones: la espacialidad de los procesos electorales y la aplicación equivocada de técnicas estadísticas espaciales para el análisis de datos autocorrelacionados espacialmente.

² Son problemas en cuanto a la aplicación de técnicas estadísticas, pero su existencia tiene razones claramente identificadas por las teorías geográficas.

La espacialidad de los procesos electorales en México

Un objetivo continuo e indispensable en los estudios electorales es conocer cómo se distribuye el voto geográficamente (Vilalta, 2003a). Los investigadores mexicanos han encontrado notables diferencias regionales en el comportamiento electoral (Ames, 1970; Domínguez y McCann, 1995; Klesner, 1993 y 1998; Molinar y Weldon, 1990). Esta variable regional ha demostrado durante varias décadas una consistencia explicativa que podría llevarnos a concederle un peso teórico. Es decir, en los estudios de geografía electoral para México resulta evidente que el contexto local o regional es determinante del voto.³ Igual sucede con la dicotomía urbano-rural: también se ha detectado sistemáticamente, desde los primeros estudios de geografía electoral en los setenta, una marcada diferencia entre las preferencias de los votantes urbanos y de los rurales (Reyna, 1971).

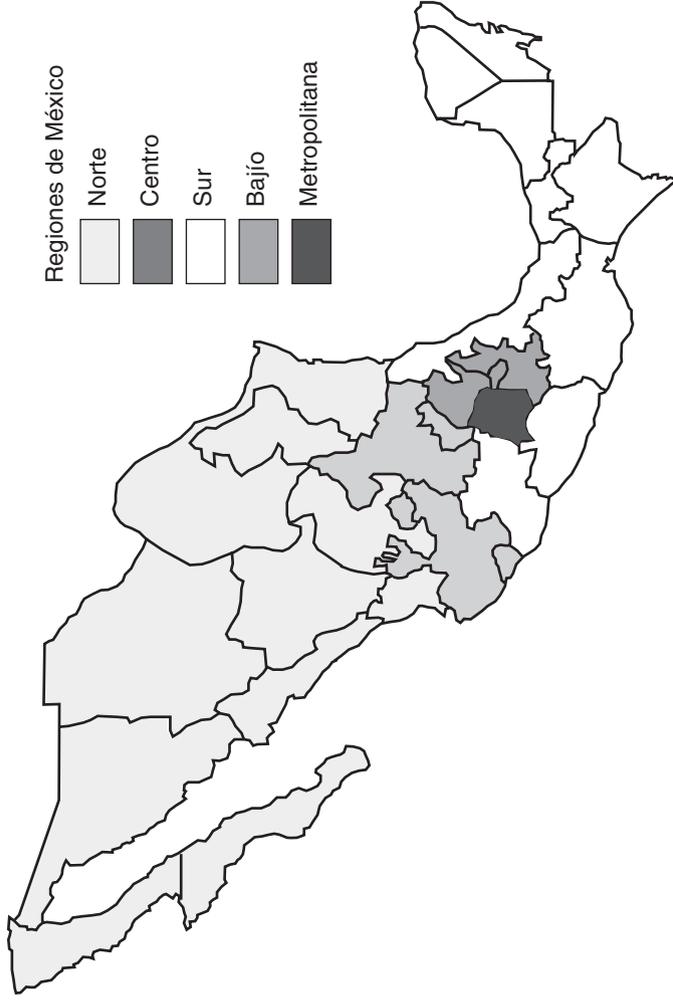
El cuadro 1 muestra la variación regional del comportamiento del grupo de ciudades con una población mayor de 100 000 habitantes, con base en el mapa correspondiente (véase el mapa 1). El porcentaje se refiere al promedio de la región. Al respecto es muy notable que en el periodo 1994-2000 el PAN haya contado con un gran apoyo principalmente en el Bajío y el Norte, en donde siempre ha recibido por lo menos el 30% del voto válido, pero es más interesante que durante esos seis años su avance se generalizara en todas las regiones. El PRD muestra un apoyo nacional menor que el PAN, pero en el sur es preferencial. El PRI, en cambio, muestra niveles de apoyo homogéneos entre las regiones, aunque son decrecientes en el tiempo; sin embargo cabe advertir que mantiene su mayor apoyo en el norte.

Para corroborar la no aleatoriedad de estas diferencias regionales se aplicó a los resultados anteriores la prueba Kruskal-Wallis,⁴ la cual mostró diferencias regionales estadísticamente significativas para todos los partidos, excepto para el PRI en 1994, cuando contaba con un apoyo muy uniforme en todo el país (véase el cuadro 2). Paralelamente se observa en los coeficientes una tendencia a la disminución en las diferencias regionales para el PAN y el PRD, y un aumento para el PRI entre 1994 y 2000. Es decir, el PRI urbano parece estar perdiendo su presencia nacional para tener un perfil más regionalizado en el norte del país.

³ No conozco en la literatura especializada mexicana algún estudio que contradiga esta conclusión.

⁴ Prueba no paramétrica. El paquete utilizado para esta prueba fue el SPSS versión 10.0.

MAPA 1
Regiones de México



FUENTE: Vilalta (2002), con base en la propuesta regional electoral de Klesner (1998).

CUADRO 1
Porcentaje promedio de votos por región en las 89 ciudades más pobladas de México, según año de elección,
para diputados federales de mayoría relativa (1994-2000)

	1994			1997			2000		
	PAN	PRI	PRD	PAN	PRI	PRD	PAN	PRI	PRD
Centro	25	53	15	20	40	31	45	35	15
Norte	30	50	11	30	40	21	42	41	14
Bajío	34	50	8	42	35	14	54	33	9
Metropolitana	28	47	16	19	36	35	44	32	20
Sur	19	47	26	20	40	33	37	33	26
Promedio general	28	49	16	29	39	24	43	36	17

NOTA: Número de ciudades por región: Centro = 5, Norte = 33, Bajío = 20, Metropolitana = 2, y Sur = 29.

CUADRO 2

Resultados de la prueba Kruskal-Wallis sobre diferencias regionales en el nivel de apoyo para cada partido en las 89 ciudades más pobladas de México, según año de elección, para diputados federales de mayoría relativa (1994-2000)^a

	PAN			PRI			PRD		
	1994	1997	2000	1994	1997	2000	1994	1997	2000
Valor	25.6**	30.9**	17.5**	3.5	9.7*	18.1**	35.6**	26.3**	28.5**

^a En el año 2000 compitieron dos alianzas: 1) Alianza por el Cambio: coalición del Partido Acción Nacional (PAN) y el Partido Verde Ecologista de México (PVEM); 2) Alianza por México: coalición del Partido de la Revolución Democrática (PRD), Partido del Trabajo (PT), Convergencia, Partido de la Sociedad Nacionalista (PSN), y Partido Acción Social (PAS). Los tres últimos partidos fueron creados justamente antes de las elecciones del año 2000.

* Significativo a un nivel de 0.05.

** Significativo a un nivel de 0.01.

Se agrega a lo anterior la importancia del electorado urbano no sólo por su comportamiento particular, sino por su participación en el auge de los partidos no priístas, que ocupa sin duda un lugar central en el análisis del comportamiento electoral en México (véase el cuadro 3). Sin embargo, pese a su importancia, los análisis detallados sobre el comportamiento electoral urbano desde una perspectiva interurbana son muy escasos (Pacheco, 1997; Vilalta, 2004). Al respecto predominan los estudios de caso, sobre todo del electorado de la Ciudad de México.

CUADRO 3

Número de ciudades ganadas por cada partido mayoritario de las 89 más pobladas de México, según año de elección, para diputados federales de mayoría relativa (1994-2000)

	1994	1997	2000	Cambio absoluto 1994-2000
PAN	5	28	59	+54
PRI	81	41	19	-62
PRD	3	20	11	+8
Total	89	89	89	

FUENTE: Vilalta (2004).

La perspectiva espacial de los procesos electorales cuenta con un peso teórico probado y tiene una importancia política elemental. Dada la escasez de estudios al respecto, se requiere su mayor difusión en la literatura científica junto con el uso de técnicas estadísticas novedosas y apropiadas.

La aplicación equivocada de técnicas espaciales para la comprensión de fenómenos espaciales y sus implicaciones estadísticas

Entre las técnicas estadísticas utilizadas en los estudios de geografía electoral, los análisis de correlación y de regresión lineal múltiple son los más típicos y más avanzados en la literatura mexicana, con excepción de los escasos estudios que muestran la utilidad de las aplicaciones espaciales mediante regresiones lineales espaciales para variables dependientes continuas (Vilalta, 2004) o dicotómicas (Fernández-Durán, Poiré y Rojas-Nandayapa, 2004).

Fuera de esa escasa literatura que hace uso de técnicas espaciales, la mayor parte de los estudios sobre geografía electoral mexicana que cuentan con un enfoque cuantitativo han ignorado o no han advertido al lector las implicaciones metodológicas que conlleva ese tipo de técnicas estadísticas cuando se hace uso de datos espacializados. Aunque los lectores pueden efectivamente reconocer las limitaciones explicativas de los modelos estadísticos utilizados, los autores no indican que se está incurriendo en posibles violaciones a los supuestos estadísticos de los análisis de regresión, posiblemente por desconocimiento de causa. Es probable que tales omisiones se deban a que la competencia multipartidista es nueva en México, y sus respectivos análisis políticos y estudios académicos de geografía electoral no han incorporado aún herramientas de análisis geográfico y estadístico avanzado; parece prematuro solicitar discusiones y advertencias metodológicas cuando hablamos de fenómenos recientes de estudio.⁵

El punto central de este trabajo es que la aplicación de técnicas estadísticas —en concreto del análisis de regresión— a datos espacializados puede llevarnos a cometer errores de interpretación y por lo tanto de validez en las conclusiones, ya que se viola un supuesto elemental en el análisis de regresión: la independencia de las observaciones.

⁵ Como se mencionó en la introducción, los primeros estudios de geografía electoral en México datan de la década de los setenta y la producción en general ha sido modesta en cuanto a su volumen.

En el análisis de datos agregados geográficamente se suele encontrar que estas unidades de análisis están autocorrelacionadas espacialmente o son espacialmente dependientes. Autocorrelación y dependencia espacial significan lo mismo, pero la distinción en el uso de palabras estriba en que el primer término hace referencia simultáneamente a un fenómeno y a una técnica estadística, y la segunda a una explicación teórica. Concretamente, existe dependencia espacial cuando “el valor de la variable dependiente en una unidad espacial es parcialmente función del valor de la misma variable en unidades vecinas” (Flint, Harrower y Edsall, 2000: 4). Esto ocurre por una razón teóricamente importante que resume la primera ley geográfica de Tobler (1970): “Todo se relaciona con todo, pero las cosas más cercanas están más relacionadas que las cosas distantes”.

El coeficiente I de Moran (1950) es la técnica estadística más frecuentemente utilizada para probar esta ley geográfica; o para poner a prueba en una fase de investigación la hipótesis de la presencia de una autocorrelación espacial de un fenómeno y sus correspondientes grados de concentración o de dispersión. Su diseño es muy similar al del coeficiente de correlación r de Pearson. También sus valores varían entre +1 y -1, en donde el primero significa una perfecta concentración (o autocorrelación) espacial y el segundo una perfecta dispersión espacial. El cero significa un patrón espacialmente aleatorio o sin orden (véase la gráfica 1).

La diferencia básica entre el coeficiente I de Moran y el r de Pearson es que en el primero la asociación entre los valores de la variable dependiente es predeterminada por una matriz de unidades vecinas.

La fórmula del coeficiente I de Moran es:⁶

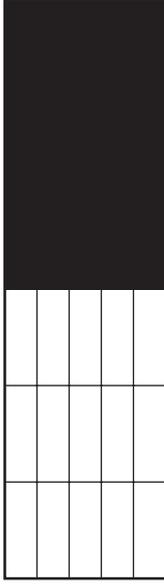
$$I = \frac{n}{\sum_{i=1}^{i=n} \sum_{j=1}^{j=n} W_{ij}} \cdot \frac{\sum_{i=1}^{i=n} \sum_{j=1}^{j=n} W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{i=n} (x_i - \bar{x})^2}$$

En esta fórmula n significa el número de las unidades (es decir, áreas o puntos) en el mapa, W_{ij} es la matriz de distancias que define si las áreas o puntos geográficos i y j son o no vecinos. El coeficiente I se sujeta a una prueba de significancia estadística de valores Z , es decir,

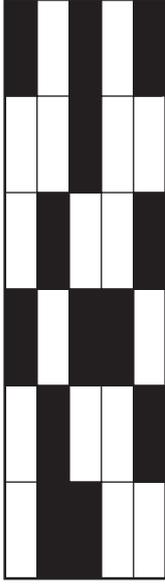
⁶ Aquí se muestra la fórmula para una variable continua; el coeficiente de Moran también puede ser calculado para variables medidas en una escala dicotómica.

GRÁFICA 1

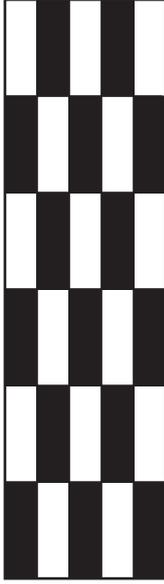
Representación visual de los tres diferentes patrones espaciales en que puede estar ordenada una variable en un mapa de celdas regulares y sus respectivos coeficientes I de Moran



$I = 1$ (perfecta concentración)



$I = 0$ (patrón aleatorio)



$I = -1$ (perfecta dispersión)

bajo el supuesto de una distribución normal de valores probables (Cliff y Ord, 1981; Goodchild, 1987).

Para seguir con el objetivo de este trabajo y probar la variación espacial no aleatoria o espacialidad de los procesos electorales en México, se calcularon diversos coeficientes *I* de autocorrelación espacial. El cuadro 4 presenta los coeficientes con base en datos agregados por entidad federativa ($n = 32$);⁷ el cuadro 5 los presenta para el caso de las ciudades más grandes del país ($n = 89$). En ambos casos los datos se refieren a las elecciones a diputados federales de mayoría relativa para cada elección entre 1994 y 2000.

Los coeficientes calculados para los resultados por entidad federativa muestran que a partir de 1997 el voto del PRI se empezó a concentrar geográficamente, mientras que los resultados del PAN y el PRD se concentran en 1997 y se vuelven a convertir en espacialmente aleatorios en las elecciones del año 2000. En cambio, para el caso del electorado urbano los coeficientes muestran claramente el proceso de reemplazo geográfico del PRI por el PAN y el PRD durante este periodo. Los coeficientes aquí evidencian una creciente concentración del voto para el PRI mientras que para el PAN y el PRD decrece su concentración para irse dispersando o nivelando en el sistema de ciudades.⁸

Estos resultados muestran que el comportamiento electoral en México tiene efectivamente una geografía no aleatoria de concentración en el ámbito urbano. Es decir, que las unidades de análisis (ciudades en este caso) se hallan autocorrelacionadas espacialmente, y que por lo tanto, potencialmente y en directa relación con este trabajo, los análisis de regresión realizados con este tipo de datos pueden carecer de validez porque se viola el supuesto de independencia entre las observaciones. Particularmente, el error estadístico al utilizar una regresión lineal de mínimos cuadrados ordinaria (OLS) con datos que presentan una autocorrelación espacial consiste en que los coeficientes estarán sesgados (Anselin, 1988). Es decir, no representarán adecuadamente la magnitud existente entre las variables independientes (VI) y la dependiente (VD).

Además de dependencia espacial, los procesos electorales pueden también presentar un patrón de heterogeneidad espacial. La heterogeneidad espacial se define como una variación en las relaciones de las variables de una región o lugar a otra (Lesage, 1998). En términos

⁷ En este caso se utilizaron los resultados totales del estado, esto es, para áreas urbanas de todos los tamaños junto con áreas rurales.

⁸ La concentración del voto urbano priísta se ha dado principalmente en el Norte.

CUADRO 4

Coefficientes de autocorrelación espacial *I* de Moran para las entidades federativas de México, según año de elección, para diputados federales de mayoría relativa (1994-2000)^a

	1994	1997	2000 ^b
PAN	0.050	0.152*	-0.060
PRI	-0.005	0.198**	0.252***
PRD	0.082	0.229**	-0.102

^a Para el cálculo del coeficiente, la definición de entidad vecina fue la de "primer orden", es decir, de estados estrictamente contiguos geográficamente.

^b En el año 2000 compitieron dos alianzas: 1) Alianza por el Cambio: coalición del Partido Acción Nacional (PAN) y el Partido Verde Ecologista de México (PVEM); 2) Alianza por México: coalición del Partido de la Revolución Democrática (PRD), Partido del Trabajo (PT), Convergencia, Partido de la Sociedad Nacionalista (PSN), y Partido Acción Social (PAS). Los tres últimos partidos fueron creados justamente antes de las elecciones del año 2000.

* Significativo a un nivel de 0.05 (prueba de dos colas).

** Significativo a un nivel de 0.01 (prueba de dos colas).

CUADRO 5

Coefficientes de autocorrelación espacial *I* de Moran para las ciudades grandes de México, según año de elección, para diputados federales de mayoría relativa (1994-2000)^a

	1994	1997	2000 ^b
PAN	0.167**	0.199**	0.072*
PRI	0.020	0.225**	0.251**
PRD	0.193**	0.155**	0.069*

^a Para el cálculo del coeficiente, la definición de ciudad vecina fue considerado un radio de 330 km entre centroides geográficos.

^b En el año 2000 compitieron dos alianzas: 1) Alianza por el Cambio: coalición del Partido Acción Nacional (PAN) y el Partido Verde Ecologista de México (PVEM); 2) Alianza por México: coalición del Partido de la Revolución Democrática (PRD), Partido del Trabajo (PT), Convergencia, Partido de la Sociedad Nacionalista (PSN), y Partido Acción Social (PAS). Los tres últimos partidos fueron creados justo antes de las elecciones del año 2000.

* Significativo a un nivel de 0.05 (prueba de dos colas).

** Significativo a un nivel de 0.01 (prueba de dos colas).

teóricos la heterogeneidad espacial se debe una variación real y sustantiva que evidencia la existencia y la validez del contexto local o regional en la definición del comportamiento social (O'Loughlin y Anselin, 1991) y de las preferencias electorales en este caso (Flint, 1995; Vilalta, 2004). Un ejemplo de lo anterior se presentaría cuando la población de cierta religión apoyara a un partido en una región, mientras que en otra región la población con la misma religión apoyara a un partido opuesto.

Lo teóricamente relevante en esta situación es que las preferencias de los electores son determinadas "por la socialización experimentada dentro del contexto de un lugar particular" (Flint, 1998: 1281). Es decir, los votantes agrupados conforme a sus características socioeconómicas o demográficas similares pueden mostrar una preferencia diferente porque están en lugares distintos y cuentan con otras experiencias políticas. Es razonable imaginar que la composición socioeconómica y demográfica de un partido no sea la misma dentro de la geografía nacional. Evidentemente en tal forma no serían necesariamente las características socioeconómicas lo que determinaría a un partido como ganador en una región, sino las circunstancias políticas de esa región.

En este caso, en términos estadísticos la heterogeneidad espacial causa la violación del supuesto de normalidad en los residuales. En específico, los dos errores estadísticos provenientes de no detectar la heterogeneidad espacial en los análisis de regresión OLS cuando existe son que: 1) al igual que en el caso de la dependencia espacial, los coeficientes serán ineficientes para mostrar la magnitud de la relación entre las variables, y 2) las pruebas de significancia estadística sobre esos mismos coeficientes pueden estar equivocadas debido a problemas de heterodasticidad (Anselin, 1988). La heterodasticidad se define como la variación no constante del error, lo que ocasiona una inflación de los errores estándar y resulta en pruebas de significancia estadística ineficaces.

Nótese que puede darse el caso de que la autocorrelación espacial venga acompañada de heterogeneidad espacial. Es decir, que los datos no sólo no sean independientes espacialmente, sino que además las relaciones entre las variables cambien de una región a otra.

En concordancia con lo anterior, una vez demostrada la existencia de un patrón de autocorrelación espacial sobre los resultados electorales en el grupo de las 89 ciudades más grandes del país (véase el cuadro 5), debemos probar si este fenómeno de dependencia está

afectando nuestros modelos de regresión. Para probar si esta dependencia espacial existe y en caso afirmativo cuantificarla, existe una variación a la regresión OLS, llamada regresión lineal espacial (SAM).⁹ Esta técnica se distingue de la OLS en que: 1) se le incorpora una estructura autorregresiva sobre la VD, limitada a una función de distancia entre observaciones, o unidades geográficas en este caso, y 2) se le agrega también la prueba *I* de Moran sobre los residuales. Nótese que la presencia de autocorrelación espacial entre los residuales se debe a alguna de las siguientes razones (Cliff y Ord, 1971):

- La presencia de relaciones no lineales entre las variables dependientes e independientes.
- La omisión de una o más variables explicativas.
- Que el modelo debe tener una estructura autorregresiva a fin de eliminar la autocorrelación en el error.

La técnica SAM tiene la siguiente forma matemática (Anselin, 1992):

$$y = \rho W y + x \beta + \varepsilon$$

En donde y es la variable dependiente, ρ o *rho* es el coeficiente autorregresivo de la variable dependiente (VD) espacialmente ligada o retrasada (*spatial lag*: $W y$), W es precisamente la matriz de unidades vecinas, x es la matriz de variables independientes, b son los respectivos coeficientes, y ε es el error.

Específicamente el coeficiente *rho* es el promedio de la VD en las áreas vecinas. Como se mencionaba, este coeficiente es un término autorregresivo de la VD dirigido a medir la autocorrelación espacial.¹⁰ Se le denomina “efecto espacial”, ya que al ser independiente de alguna VI causal, permite probar la hipótesis y medir el efecto que tienen los niveles de la VD en las áreas vecinas y sostener la inferencia de un efecto contextual (Puech, 2004). Su valor varía entre +1 y -1. En caso de que tenga un valor de cero o muy cercano y por ende no estadísticamente significativo, la ecuación evidentemente regresa a ser de tipo OLS. Ya que, si 1) $y = \rho W y + x \beta + \varepsilon$, y si 2) $\rho W y = 0$, entonces, 3) $y = x \beta + \varepsilon$.

En este trabajo realizamos un ejercicio comparativo entre las técnicas OLS y SAM con el propósito de probar si la dependencia espacial presente en las preferencias partidistas del electorado urbano en Méxi-

⁹ Véase Lesage (1998).

¹⁰ Precisamente de la misma VD.

co afecta a los coeficientes de regresión OLS, y si es o no pertinente utilizar la técnica SAM. Previo al ejercicio y a la explicación de los resultados, se describen las variables, la información estadística y los paquetes utilizados en este estudio.

Variables, fuentes de información y *software*

El objetivo que perseguimos en este trabajo es comparar dos técnicas de regresión diferentes: la regresión OLS y la regresión espacial SAM. Con este propósito se definen tres variables dependientes: el voto para cada partido (PAN-PVEM, PRI, y PRD-6) en las elecciones para el año 2000. Se utilizan tres ecuaciones para cada técnica de regresión; o sea, una para cada partido. Los modelos incluyen variables causales que se han de probar para determinar la existencia de efectos sociodemográficos independientes de los efectos regionales (véase el cuadro 6). Todas las variables son continuas salvo la regional, que es de tipo nominal (*dummy*). Para cada partido se presenta el modelo más compacto y significativo con la intención de predecir su voto; por ende, los modelos varían entre los partidos.

En este estudio utilizamos una muestra de ciudades grandes: las 89 que en 1995 tenían 100 000 o más habitantes. La razón para elegir este umbral poblacional es contar con la posibilidad de tomar dichas ciudades como unidades de análisis que sean comparables en tamaño para así probar la importancia del contexto local urbano o regional-urbano con unidades similares, y paralelamente aumentar el poder de las pruebas inferenciales (Keppel y Zedeck, 1998).¹¹ Cabe mencionar que esta muestra urbana incluye 56.4% de la población total del país.

Si bien México no tiene una posición oficial respecto a la definición de ciudad, partimos de la definición institucional del Consejo Nacional de Población (Conapo) sobre las ciudades y las áreas metropolitanas incluidas en este estudio. Una vez adoptada la definición de ciudad, los datos sociodemográficos fueron agregados desde el ámbito municipal. Esta información la proporcionó el Instituto Nacional de Estadística, Geografía e Informática (INEGI). Finalmente, fueron obtenidos los resultados electorales municipales y posteriormente agregados con

¹¹ Otra forma de aumentar el poder de las pruebas inferenciales sería aumentar el tamaño de la muestra, pero esto conllevaría la inclusión de ciudades de tamaños más disímiles en las cuales el efecto contextual y la similitud serían más difíciles de argumentar.

CUADRO 6
VARIABLES EN LOS MODELOS DE REGRESIÓN

<i>Variables</i>	<i>Descripción</i>
Dependiente	
Preferencia electoral	% del voto para cada partido (PAN-PVEM, PRI, PRD-6) en las elecciones del año 2000
Independientes	
Población	Tamaño de la población en cientos de miles (2000)
Lengua indígena	% que habla una lengua indígena (2000)
Migración	% que vivía en otra entidad federativa en 1995 (2000)
Manufacturas	% empleado en el sector manufacturero (2000)
Educación	% > 15 años que sabe leer y escribir (2000)
Ingreso	% de la población ocupada que percibe > 10 salarios mínimos diarios (2000)
Catolicismo	% que se identifica como católico (2000)
Región	5 regiones: Central, Norte, Bajío, Área Metropolitana de la Ciudad de México, Sur

base en la definición de las ciudades; también se tomaron del sitio oficial del Instituto Federal Electoral (IFE).¹²

La localización de cada ciudad en el Sistema de Información Geográfica (SIG) es dada por los centroides geográficos. Para el cómputo del coeficiente de autocorrelación espacial I de Moran, se predefinió en la matriz de vecinos a las ciudades que se encontraran en un radio de 330 km, distancia mínima en que todas las ciudades de la muestra tienen por lo menos una ciudad vecina. El promedio de ciudades vecinas en la muestra es de 14. Para la técnica SAM se utilizó el paquete de SIG ArcView junto con el módulo Spatial-Statistics de S-Plus. Para la técnica OLS se utilizó el *software* Eviews.

Procedimiento y resultados de la comparación de técnicas OLS y SAM

Primeramente calcularemos los coeficientes de autocorrelación espacial de cada una de las VI utilizadas en este estudio. En el cuadro 7 podemos observar que con excepción de la *población* y la *migración*, todas las variables muestran un patrón geográfico de concentración.

¹² La información del INEGI y del IFE fue obtenida vía internet.

CUADRO 7

Coefficientes de autocorrelación espacial I de Moran para las ciudades grandes de México, en diferentes variables socioeconómicas (2000)^a

<i>Variables socioeconómicas</i>	<i>Coefficiente</i>
Población	-0.045
Lengua indígena	0.167**
Migración	0.052
Manufacturas	0.066*
Educación	0.291**
Ingreso	0.083**
Catolicismo	0.533**

^a Para el cálculo del coeficiente, la definición de ciudad vecina se consideró un radio de 330 km entre centroides geográficos.

* Significativo a un nivel de 0.05 (prueba de dos colas).

** Significativo a un nivel de 0.01 (prueba de dos colas).

Particularmente, se advierte una alta concentración espacial en el *catolicismo* y una moderada en la *educación*.

A continuación describiremos los análisis de regresión OLS y SAM que aplicamos a los partidos. Cabe reiterar que para cada uno se buscó el mejor modelo predictivo con base en las variables socioeconómicas seleccionadas. El ejercicio comparativo entre las dos técnicas siguió el siguiente procedimiento de tres pasos:

- a) Con base en la técnica OLS se buscó el mejor modelo socioeconómico para cada partido, es decir, aquel que sólo contara con variables significativas, ya fuera un modelo amplio (de seis o más variables y hasta siete posibles) o compacto (de dos o menos).
- b) Una vez definido el modelo socioeconómico, también empleando la técnica OLS se agregaron las variables (nominales o *dummies*) regionales para probar la hipótesis de efectos regionales diferenciados e independientes de las variables socioeconómicas. Aquí nos interesamos en observar los cambios en la R^2 y en vislumbrar problemas de heterodasticidad.
- c) Posteriormente se aplicó la técnica SAM al modelo socioeconómico y regional y se compararon los resultados con los anteriores, provenientes de la técnica OLS. SAM permite la inclusión de dos pruebas más en el modelo predictivo: una para detectar

y medir los efectos espaciales o de autocorrelación en la VD con base en el coeficiente ρ (ρ) y una prueba sobre autocorrelación espacial en los residuales.

Describimos los resultados en tres secciones, una para cada partido, y al final agregamos otra en donde resumimos los resultados de todos los modelos resultantes de la aplicación de las dos técnicas, OLS y SAM. Se adelanta que para cada partido se obtuvo un modelo predictivo diferente. Nótese que en cada sección por partido se encuentran tres cuadros que resumen los resultados de las tres regresiones llevadas a cabo. Dicho de nuevo y concisamente: dos regresiones con OLS (la primera sin efectos regionales y la segunda con efectos regionales), y una regresión con SAM (con efectos regionales y midiendo la autocorrelación espacial en la VD, además de una prueba de autocorrelación espacial en los residuales).

El caso del PAN-PVEM

El cuadro 8 muestra el mejor modelo con base en la técnica OLS para el PAN-PVEM utilizando como útiles para predecir el voto agregado geográficamente las variables socioeconómicas previamente elegidas y predefinidas en la literatura especializada (véase el cuadro 8). Podemos ver que este modelo: 1) presenta cuatro variables como estadísticamente significativas para predecir el comportamiento electoral para esta coalición partidista; 2) cuenta con una capacidad explicativa aceptable ($R^2 = .358$), y 3) no muestra problemas de inconstancia en el error (heterodasticidad).

Una vez definidas las variables socioeconómicas capaces de predecir el voto para esta coalición, incluimos las variables regionales (*dummies*) y realizamos otra regresión OLS. En este caso se excluyó la variable *catolicismo*, ya que dejó de ser estadísticamente significativa una vez incluidas las variables regionales, y se volvió a correr la regresión sin la misma variable sobre religión.

En los resultados (véase el cuadro 9) se aprecia que la significancia estadística de las variables socioeconómicas se mantuvo (salvo en el caso del *catolicismo*, que fue retirada del modelo). En cuanto a las variables regionales, el Bajío ofrece resultados positivos y significativamente superiores a los de la región Norte para la coalición PAN-PVEM; las demás regiones no presentan un patrón significativamente diferente al de la

CUADRO 8

Resultados del modelo de regresión OLS para la coalición del PAN-PVEM sin variables regionales

Variable dependiente: PAN-PVEM

Método: mínimos cuadrados (OLS)

n = 89

<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>	
Constante	-19.565	13.886	-1.408	0.162	
Socioeconómicas:					
Catolicismo	0.557	0.170	3.267	0.001	
Manufacturas	0.484	0.120	4.011	0.000	
Ingreso	6.417	1.834	3.497	0.000	
Lengua indígena	0.529	0.247	2.136	0.035	
R cuadrada	0.358	Media de la variable dependiente			43.438
R cuadrada ajustada	0.328	Desviación estándar de la variable dependiente			13.341
Error estándar de la regresión	10.934	Criterio de Akaike			7.676
Suma de los residuales cuadráticos	10043.45	Criterio de Schwarz			7.816
Logaritmo de verisimilitud	-336.594	Estadístico F			11.751
Prueba de Durbin-Watson	1.828	Sig. de estadístico F			0.000
Prueba de heteroscedasticidad de White:					
Estadístico F	0.873	Sig.			0.542

CUADRO 9

Resultados del modelo de regresión OLS para la coalición del PAN-PVEM con variables regionales

Variable dependiente: PAN-PVEM

Método: mínimos cuadrados (OLS)

n = 89

<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>	
Constante	18.138	4.492	4.037	0.000	
Socioeconómicas:					
Manufacturas	0.488	0.125	3.892	0.000	
Ingreso	6.712	1.799	3.730	0.000	
Lengua indígena	0.482	0.244	1.973	0.051	
Efecto regional:					
Bajío	14.568	2.972	4.900	0.000	
Centro	6.301	5.129	1.228	0.222	
Metropolitana	1.145	7.497	0.152	0.879	
Sur	2.429	3.337	0.727	0.468	
R cuadrada	0.452	Media de la variable dependiente			43.438
R cuadrada ajustada	0.405	Desviación estándar de la variable dependiente			13.341
Error estándar de la regresión	10.286	Criterio de Akaike			7.585
Suma de los residuales cuadráticos	8569.948	Criterio de Schwarz			7.808
Logaritmo de verisimilitud	-329.534	Estadístico F			9.578
Prueba de Durbin-Watson	2.094	Sig. de estadístico F			0.000
Prueba de heterodasticidad de White:					
Estadístico F	2.097	Sig.			0.034

NOTA: Se utiliza como referencia a la región Norte.

región Norte. Este modelo OLS tiene una mayor capacidad explicativa que el anterior ($R^2 = .452$), sin embargo muestra problemas de hetero-dasticidad (White test = 18.86, $\rho < .05$). Lo anterior se debe a la inclusión de las variables regionales, las cuales –recordemos– pueden ocasionar que: 1) los coeficientes sean ineficientes para mostrar la magnitud correcta de la relación entre las VI con la VD, o bien, 2) que las pruebas de significancia estadística sobre esos mismos coeficientes sean equivocadas. A este respecto, sin embargo, los coeficientes de las tres variables socioeconómicas –*manufacturas*, *ingreso* y *lengua indígena*– se mantuvieron moderadamente similares entre el primer modelo, que no mide los efectos regionales, y el segundo, que sí los incluye en la ecuación. Lo que cambió notablemente (en signo) fue la constante.

Con el fin de comparar las dos técnicas, OLS y SAM, de forma idéntica al modelo anterior, en el cuadro 10 se muestran los resultados de la regresión espacial SAM, la cual presenta diferencias pequeñas en comparación con los resultados del OLS en cuanto a la magnitud de los coeficientes, incluidas las variables regionales, salvo *metropolitana*, para la cual la diferencia entre una y otra técnica es notable, pero en ambos casos sin alcanzar una significancia estadística.

En este caso, en cuanto al efecto espacial el coeficiente autorregresivo *rho* resultó de pequeña magnitud y cercano a cero, por lo que muy probablemente no sea estadísticamente significativo.¹³ Así, tal modelo espacial se convertiría en uno típico de OLS. A esto se agrega que de la prueba *I* de Moran sobre los residuales también se desprende que los mismos son espacialmente aleatorios. Lo que esto demuestra, en conjunto, es que las VI utilizadas, socioeconómicas y regionales, neutralizan la autocorrelación espacial presente en la VD.¹⁴

Puede significar también que la autocorrelación espacial en el voto por esta coalición puede ser explicada por la autocorrelación espacial en las VI. Como lo habíamos mostrado en el cuadro 7, cada una de estas variables está concentrada en el espacio, y en este caso la no significancia estadística del coeficiente autorregresivo puede deberse a que éstas compartan en buena parte la misma geografía que el voto por el PAN-PVEM. Dicho de otra manera, la no existencia de un contex-

¹³ Desafortunadamente el módulo Space-Stat no calcula la significancia estadística de dicho coeficiente.

¹⁴ Cabe agregar que se utilizó una SAM con las mismas variables socioeconómicas sin incluir a las regionales, y resultó que el valor del coeficiente *rho* es igualmente de muy pequeña magnitud y no hay autocorrelación espacial en los residuales; es decir, las tres VI socioeconómicas utilizadas son suficientes para neutralizar los efectos espaciales en la VD.

CUADRO 10

Resultados del modelo de regresión SAM para la coalición del PAN-PVEM

<i>Variable dependiente: PAN-PVEM</i>				
<i>Método: regresión espacial (SAM)*</i>				
<i>n = 89</i>				
<i>Variable</i>	<i>Coficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	17.242	4.309	4.008	0.000
Socioeconómicas:				
Manufacturas	0.497	0.121	4.097	0.000
Ingreso	7.050	1.783	3.933	0.000
Lengua indígena	0.564	0.237	2.379	0.019
Efecto regional:				
Bajío	14.713	2.398	6.134	0.000
Centro	6.356	4.660	1.363	0.176
Metropolitana	0.121	7.293	0.016	0.986
Sur	2.262	3.017	0.749	0.455
Error estándar residual:				
ESR	10.160			
Efecto espacial:				
Rho	-0.031			
Prueba de autocorrelación espacial:**				
Coef. I de Moran	-0.005	0.033	0.171	0.864

* El paquete no calcula la R².

** Para este caso se computa y presenta la z-Statistic en vez de la t-Statistic.

NOTA: Se utiliza como referencia a la región Norte.

to local puede deberse a una concordancia espacial (*spatial match*) de tales variables socioeconómicas con las preferencias electorales.

Sin embargo no debemos perder de vista que tal modelo muestra problemas de heterodasticidad, por lo que los resultados habrán de tomarse con cautela. Y aunado a esto recordemos que la presencia de heterodasticidad en un modelo que incluye efectos regionales estadísticamente significativos puede ser aún más indicativa de heterogeneidad espacial, es decir, de variaciones en las relaciones entre las variables de una región a otra (Flint, 1995).

El caso del PRI

Para el caso del PRI seguimos exactamente el mismo procedimiento que para el partido o coalición anterior. Esto es, primero realizamos una regresión OLS con variables socioeconómicas para encontrar el

mejor modelo predictivo con base en información agregada geográficamente. Posteriormente, una vez detectado el mejor modelo socioeconómico, incluimos las variables regionales utilizando la misma técnica OLS. Una vez que tenemos el modelo ideal en OLS, el cual incluye tanto efectos socioeconómicos como regionales, procedemos a probarlo con la técnica SAM y observamos las diferencias entre ambas técnicas, poniendo énfasis en los valores de los coeficientes, los signos, evidencia de heterodasticidad, el coeficiente autorregresivo (efecto espacial) y la prueba *I* de Moran de autocorrelación espacial sobre los residuales.

En cuanto a los resultados de la regresión y a diferencia del PAN-PVEM, el PRI es más difícil de predecir con base en variables socioeconómicas agregadas geográficamente. El modelo muestra sólo una variable como estadísticamente significativa, *catolicismo*, y una varianza explicada muy baja ($R^2 = .073$). Es en síntesis un modelo muy insuficiente. Afortunadamente no muestra problemas de heterodasticidad.

Para seguir con nuestro ejercicio geográfico y comparativo entre técnicas, ahora a este modelo le agregamos la prueba de hipótesis sobre efectos regionales aplicando la misma técnica OLS. Seguimos utilizando como en el caso del partido anterior, a la región Norte como referencia.

En este caso (véase el cuadro 12) el modelo aumenta su capacidad explicativa ($R^2 = .270$) por razones obvias (simplemente por incluir más variables) y sigue sin mostrar problemas de heterodasticidad. La variable *catolicismo* mantiene su validez, y también su efecto mantiene una magnitud muy similar a la del modelo anterior sin variables regionales. A este respecto hay dos regiones que muestran significancia estadística: Bajío y Sur. En ambos casos hay efectos negativos sobre el PRI, lo que sugiere que éste obtiene resultados significativamente positivos en la región Norte sobre ambas y viceversa. Los resultados que obtiene en el Norte no se diferencian significativamente de los que obtiene en las ciudades de la región Centro y en la región Metropolitana (Zona Metropolitana de la Ciudad de México y Zona Metropolitana de Toluca).

A continuación utilizamos la técnica SAM (véase el cuadro 13) sobre el modelo anterior, y observamos que la variable *catolicismo* mantiene casi el mismo efecto sobre la VD, pero resulta interesante que se agregue la región Centro como significativamente diferente a la región de referencia Norte, igualmente con un efecto negativo como en los casos de las regiones Bajío y Sur.

CUADRO 11

Resultados del modelo de regresión OLS para el PRI sin variables regionales

Variable dependiente: PRI

Método: mínimos cuadrados (OLS)

n = 89

<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	58.879	8.647	6.808	0.000
Socioeconómica:				
Catolicismo	-0.297	0.112	-2.633	0.010
R cuadrada	0.073	Media de la variable dependiente		36.202
R cuadrada ajustada	0.063	Desviación estándar de la variable dependiente		7.868
Error estándar de la regresión	7.615	Criterio de Akaike		6.920
Suma de los residuales cuadráticos	5 046.034	Criterio de Schwarz		6.976
Logaritmo de verisimilitud	-305.964	Estadístico F		6.936
Prueba de Durbin-Watson	1.466	Sig. de estadístico F		0.009
Prueba de heterodasticidad de White:				
Estadístico F	0.986	Sig.		0.376

CUADRO 12

Resultados del modelo de regresión OLS para el PRI con variables regionales

Variable dependiente: PRI

Método: mínimos cuadrados (OLS)

n = 89

<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	64.809	9.118	7.107	0.000
Socioeconómica:				
Catolicismo	-0.319	0.120	-2.654	0.009
Efecto regional:				
Bajo	-5.045	2.162	-2.332	0.022
Centro	-5.496	3.322	-1.654	0.101
Metropolitana	-7.911	5.066	-1.561	0.122
Sur	-8.105	1.782	-4.546	0.000
R cuadrada	0.270	Media de la variable dependiente		36.202
R cuadrada ajustada	0.226	Desviación estándar de la variable dependiente		7.868
Error estándar de la regresión	6.919	Criterio de Akaike		6.771
Suma de los residuales cuadráticos	3 973.812	Criterio de Schwarz		6.939
Logaritmo de verisimilitud	-295.334	Estadístico F		6.159
Prueba de Durbin-Watson	1.703	Sig. de estadístico F		0.000
Prueba de heteroscedasticidad de White:				
Estadístico F	0.601	Sig.		0.728

NOTA: Se utiliza como referencia a la región Norte.

CUADRO 13

Resultados del modelo de regresión SAM para el PRI

<i>Variable dependiente: PRI</i>				
<i>Método: regresión espacial (SAM)*</i>				
<i>n = 89</i>				
<i>Variable</i>	<i>Coficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	63.684	9.600	6.335	0.000
Socioeconómicas:				
Catolicismo	-0.305	0.126	-2.408	0.018
Efecto regional:				
Bajío	-4.095	2.403	-1.704	0.092
Centro	-5.950	3.567	-1.667	0.099
Metropolitana	-8.432	5.174	-1.629	0.107
Sur	-7.758	1.966	-3.944	0.002
Error estándar residual:				
ESR	6.894			
Efecto espacial:				
Rho	0.015			
Prueba de autocorrelación espacial:**				
Coef. I de Moran	0.007	0.033	0.558	0.576

* El paquete no calcula la R^2 .

** Para este caso se computa y presenta la z-Statistic en vez de la t-Statistic.

NOTA: Se utiliza como referencia a la región Norte.

El coeficiente autorregresivo es de muy pequeña magnitud ($\rho = 0.015$) por lo que parece poco probable que tenga alguna significancia estadística, lo cual supone que la autocorrelación en la VD puede ser explicada por las variables independientes seleccionadas. Igualmente, el modelo no muestra problemas de autocorrelación espacial en los residuales ($I = -0.007$, n. s.).¹⁵

Por lo tanto, la carencia de efectos espaciales y la nula autocorrelación espacial en los residuales significan que el modelo elegido tiene una buena especificación para controlar la dependencia espacial de la VD, pese a que carece de una aceptable capacidad explicativa (véase el R^2 en el cuadro 11). Es decir, 1) la pobre capacidad explicativa aunada a 2) la existencia de efectos regionales en las regiones Centro, Bajío y Sur frente a la Norte y 3) la ausencia de heterogeneidad y dependencia espaciales, permiten sostener la hipótesis alternativa de que la variación espacial en la VD muy difícilmente se debe a la existencia

¹⁵ Al correr una SAM con la variable *catolicismo* como la única VI, sin variables regionales, mantuvo su capacidad para predecir el voto por el PRI, además de carecer de autocorrelación espacial en los residuales.

de un contexto regional independiente (McAllister, 1987), sino más bien a la falta de variables socioeconómicas adecuadas en el modelo para predecir el voto por el PRI.

El caso del PRD-6

Para el caso de la coalición PRD-6 podemos observar en el cuadro 14 los resultados del modelo socioeconómico con base en la técnica OLS. Salta a la vista que dos variables que son significativas para la coalición PAN-PVEM, *ingreso* y *manufacturas*, también lo son para el PRD-6, salvo que en este último caso tienen un efecto negativo (véase los cuadros 8 y 14 respectivamente).

Otro aspecto importante es que su capacidad explicativa ($R^2 = .337$) es superior a la del PRI y muy similar a la de la coalición PAN-PVEM, pese a que cuenta con sólo la mitad del número de VI: dos frente a cuatro. Pero lo que más llama la atención es que el modelo presenta problemas de heterodasticidad (White test = 8.243, $\rho < .10$). Esto es diferente a lo ocurrido con los partidos anteriores, en donde sólo la coalición PAN-PVEM presentaba problemas de inconstancia en el error, pero sólo cuando se incorporaban las variables regionales.

A continuación, de hecho observamos que una vez que incorporamos al modelo socioeconómico anterior la prueba de efectos regionales con la técnica OLS, se acentúa el problema de heterodasticidad (White test = 16.499, $\rho < .05$), por lo que la interpretación sobre la magnitud de los coeficientes en el modelo y su significancia estadística debe ser hecha con reservas (véase el cuadro 15). Puede verse que las variables socioeconómicas muestran efectos de una magnitud similar (aunque un poco más reducida) y de igual dirección que los del modelo anterior. Por otro lado, sólo el comportamiento de la región Bajío parece diferente al de la región Norte, en este caso con un efecto negativo.

La técnica SAM ofrece resultados diferentes. En primer lugar se encuentra el caso del efecto producido por la variable *ingreso*, notablemente menor que el registrado con la técnica OLS. La del Bajío se mantiene como la única región con una relación estadísticamente significativa, pero cuyo efecto es de una magnitud mayor que el doble ($b_{SAM} = -16.180$, $\rho < .01$) del registrado por la técnica OLS ($b_{OLS} = -7.823$, $\rho < .01$). Tanto el coeficiente autorregresivo como la prueba de autocorrelación en los residuales nos indican que la dependencia espacial

CUADRO 14

Resultados del modelo de regresión OLS para la coalición PRD-6 sin variables regionales

Variable dependiente: PRD-6

Método: mínimos cuadrados (OLS)

n = 89

<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	36.413	3.156	11.536	0.000
Socioeconómicas:				
Ingreso	-6.190	1.743	-3.551	0.000
Manufacturas	-0.514	0.117	-4.393	0.000
R cuadrada	0.337	Media de la variable dependiente		17.224
R cuadrada ajustada	0.321	Desviación estándar de la variable dependiente		13.044
Error estándar de la regresión	10.74	Criterio de Akaike		7.619
Suma de los residuales cuadráticos	9 921.569	Criterio de Schwarz		7.703
Logaritmo de verisimilitud	-336.051	Estadístico F		21.895
Prueba de Durbin-Watson	1.708	Sig. de estadístico F		0.000
Prueba de heteroelasticidad de White: Estadístico F	2.143	Sig.		0.082

CUADRO 15

Resultados del modelo de regresión OLS para la coalición PRD-6 con variables regionales

Variable dependiente: PRD-6

Método: mínimos cuadrados (OLS)

n = 89

<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	34.659	4.396	7.883	0.000
Socioeconómicas:				
Ingreso	-5.994	1.761	-3.402	0.001
Manufacturas	-0.401	0.123	-3.250	0.001
Efecto regional:				
Bajío	-7.823	2.925	-2.674	0.009
Centro	-3.322	5.029	-0.660	0.510
Metropolitana	4.918	7.385	0.666	0.507
Sur	3.440	3.078	1.117	0.267
R cuadrada	0.437	Media de la variable dependiente		17.224
R cuadrada ajustada	0.396	Desviación estándar de la variable dependiente		13.044
Error estándar de la regresión	10.134	Criterio de Akaike		7.544
Suma de los residuales cuadráticos	8 420.464	Criterio de Schwarz		7.740
Logaritmo de verisimilitud	-328.750	Estadístico F		10.635
Prueba de Durbin-Watson	1.866	Sig. de estadístico F		0.000
Prueba de heterodasticidad de White:				
Estadístico F	2.275	Sig.		0.030

NOTA: Se utiliza como referencia a la región Norte.

CUADRO 16
Resultados del modelo de regresión SAM para la coalición PRD-6

<i>Variable dependiente: PRD-6</i>				
<i>Método: regresión espacial (SAM)*</i>				
<i>n = 89</i>				
<i>Variable</i>	<i>Coefficiente</i>	<i>Error estándar</i>	<i>t</i>	<i>Sig.</i>
Constante	33.453	4.471	7.481	0.000
Socioeconómicas:				
Ingreso	-4.882	1.672	-2.919	0.004
Manufacturas	-0.435	0.121	-3.593	0.006
Efecto regional:				
Bajío	-16.180	4.368	-3.703	0.000
Centro	-8.152	6.048	-1.348	0.181
Metropolitana	-3.797	8.026	-0.473	0.637
Sur	-0.104	3.816	-0.027	0.978
Error estándar residual:				
ESR	10.160			
Efecto espacial:				
Rho	0.046			
Prueba de autocorrelación espacial:**				
Coef. I de Moran	-0.057	0.033	-1.362	0.173

* El paquete no calcula la R².

** Para este caso se computa y presenta la z-Statistic en vez de la t-Statistic.

NOTA: Se utiliza como referencia a la región Norte.

en la VD ha sido eliminada una vez incluidas las variables independientes seleccionadas; el modelo es adecuado para resolver posibles problemas provenientes de la autocorrelación espacial de la VD. A este respecto, el error estándar de la regresión, como una medida aproximada del buen ajuste del modelo, nos señala que el error de estimación es muy similar al obtenido con la técnica OLS (SER = 10.160). Cabe repetir que el modelo muestra problemas de heterodasticidad, por lo que los resultados deben ser interpretados con cautela.

Una recapitulación de los resultados obtenidos con las técnicas OLS y SAM

El objetivo del presente documento no es elaborar explicaciones teóricas sobre el comportamiento electoral, sino estrictamente comparar las dos diferentes técnicas y las implicaciones estadísticas, de ahí que nos concentremos precisamente en dos aspectos de igual importancia: 1) los valores de los coeficientes, y 2) las pruebas de significancia estadística

sobre los mismos.¹⁶ Una vez hecho lo anterior, y a la par en importancia, nos interesa enfocarnos en 3) la presencia de heterodasticidad como evidencia inicial de heterogeneidad espacial, y los resultados de las pruebas de 4) efectos regionales (*dummies* regionales), 5) efectos espaciales y 6) autocorrelación espacial en los residuales de los modelos.

El resumen comparativo de las técnicas OLS y SAM para cada partido se muestran en los cuadros 17 y 18. Iniciando la comparación en cuanto a los modelos predictivos (cuadro 17), lo que más llama la atención son:

- a) Las diferencias obtenidas para el caso del PRI con el modelo que incluye las variables regionales en cuanto a la magnitud de los coeficientes y la significancia estadística de las variables regionales.
- b) La diferencias obtenidas para el caso del PRD-6 con el modelo sin variables regionales en cuanto a la magnitud de los coeficientes.
- c) La diferencias obtenidas para el caso del PRD-6, pero ahora con el modelo que incluye las variables regionales en cuanto a la magnitud de los coeficientes.

Como podemos ver, las técnicas dieron resultados similares para la coalición PAN-PVEM; en cambio para el caso del PRI la diferencia estriba en la magnitud de los coeficientes y en la capacidad de la técnica SAM para detectar tres regiones (frente a dos con OLS) como significativamente diferentes de la región de referencia, que es la Norte. Para el PRD-6 las técnicas dieron resultados notablemente diferentes con ambos modelos, particularmente en cuanto a la magnitud de los coeficientes. Pero en todos los casos en que se observaron diferencias entre OLS y SAM, sin excepción la técnica SAM fue la más robusta de las dos debido a la constante detección de coeficientes con una mayor significancia estadística (y en consecuencia magnitud) y a la mayor capacidad en la detección de efectos regionales significativos. Esto se debió a la inclusión del coeficiente *rho* en la ecuación.¹⁷

La heterodasticidad estuvo presente en dos de los tres modelos (cuadro 18): PAN-PVEM y PRD-6; precisamente en aquellos partidos que

¹⁶ Para una detallada explicación de la importancia teórica de cada VI para predecir el comportamiento electoral en este mismo ejemplo, véase otros estudios al respecto, como Vilalta (2004).

¹⁷ Aun careciendo de significancia estadística en el modelo.

CUADRO 17

Comparación entre las técnicas OLS y SAM para cada partido según los modelos con variables socioeconómicas y con variables socioeconómicas y regionales

	<i>Modelo con variables socioeconómicas</i>			<i>Modelo con variables socioeconómicas y variables regionales</i>		
	<i>Magnitud y signo de los coeficientes</i>	<i>Significancia estadística</i>	<i>Magnitud y signo de los coeficientes</i>	<i>Significancia estadística</i>	<i>Regiones significativas y su efecto*</i>	
PAN-PVEM	OLS ≈ SAM	OLS ≈ SAM	OLS ≈ SAM	OLS ≈ SAM	OLS: Bajío (+) SAM: Bajío (+)	
PRI	OLS ≈ SAM	OLS ≈ SAM	OLS ≠ SAM	OLS ≠ SAM	OLS: Bajío (-) Sur (-) SAM: Bajío (-)	
PRD-6	OLS ≠ SAM	OLS ≈ SAM	OLS ≠ SAM	OLS ≈ SAM	OLS: Centro (-) Bajío (-) SAM: Bajío (-)	

* Se utiliza como referencia a la región Norte.

CUADRO 18

Detección de un efecto espacial, autocorrelación espacial en los residuales y heterodasticidad para cada partido

	<i>Efecto espacial (ρ)</i>	<i>Autocorrelación en los residuales (I de Moran)</i>	<i>Heterodasticidad en el modelo*</i>
PAN-PVEM	n. s.	n. s.	sí
PRI	n. s.	n. s.	no
PRD-6	n. s.	n. s.	sí

* Con base en la prueba White de heterodasticidad al utilizar la técnica OLS.

estuvieron menos concentrados espacialmente en el año 2000 (véase el cuadro 5). Esto significa que los coeficientes de la regresión pueden no estar capturando la magnitud real de la relación y además estar cometiendo potencialmente errores del tipo II en las pruebas de hipótesis sobre los mismos coeficientes (Anselin, 1988 y 1992). Esto lo puede ocasionar la existencia de heterogeneidad espacial sustantiva, es decir, patrones regionales de comportamiento electoral distintivos (O'Loughlin y Anselin, 1991; Flint, Harrower y Edsall, 2000).

En esta cuestión el modelo del PRI es diferente del de los otros partidos. El del PRI no presenta problemas de heterodasticidad, por lo que podemos confiar en que los coeficientes no son inestables ni las pruebas de hipótesis sobre los mismos ineficientes. Cabe considerar que con la técnica SAM se pudo detectar una región más como significativamente diferente de la Norte (cuadro 17) frente a la OLS, debido a la inclusión del coeficiente autorregresivo. Se agrega, ahora sí, de manera coincidente con los modelos de los otros dos partidos-coaliciones, que el efecto espacial medido por el coeficiente ρ es cercano a cero y no presenta una autocorrelación en los residuales. Esto nos lleva a concluir que pese a su pobre capacidad explicativa (sólo pudimos detectar que tiende a ganar en ciudades con baja proporción de población católica), el modelo para el PRI captura de forma suficiente la variación espacial en la VD y las variables regionales posiblemente están capturando una variación relacionada con las variables socioeconómicas faltantes en el modelo (McAllister, 1987); la variación geográfica difícilmente se debería a una dependencia espacial proveniente de un efecto contextual, al menos para el caso de la elección del año 2000.

En cuanto a la medición de un efecto espacial con la técnica SAM, el coeficiente ρ , que se refiere al valor promedio de la variable dependiente en las ciudades vecinas, no produjo efecto alguno para ningún partido (cuadro 18). Recordemos que el voto estaba espacialmente concentrado para todos los partidos, y muy notablemente para el PRI (cuadro 5). Para todos ellos, la no significancia del efecto espacial después de incorporar las variables socioeconómicas y regionales en sus modelos de regresión respectivos, sugiere que las explicaciones socioeconómicas de los modelos son efectivas y suficientes para aclarar la concentración espacial del comportamiento electoral. Así, el voto en cada ciudad no es dependiente de las ciudades vecinas sino de las variables socioeconómicas incluidas en el modelo; el voto se concentra debido a la agrupación de población con ciertas características socioeconómicas.¹⁸ Para demostrar esta aseveración y no dejar pendiente una prueba sobre un efecto regional-espacial se corrieron regresiones SAM a cada partido sin la inclusión de variables regionales, es decir, utilizando un modelo exclusivamente socioeconómico, y en todos los casos estas variables socioeconómicas fueron igualmente efectivas y suficientes para anular el efecto espacial.¹⁹ Sin embargo, recordemos que para los casos del PAN-PVEM y del PRD-6 los modelos presentaban problemas de heterodasticidad, por lo que la tesis anterior sigue abierta a posteriores pruebas.

Por otro lado, las pruebas I de Moran sugieren la casi nula probabilidad de contar con residuales autocorrelacionados espacialmente para todos los partidos (cuadro 18). Sin embargo aquí es necesario considerar que el hecho de que los residuales no estén autocorrelacionados espacialmente no implica que los efectos espaciales estén siendo detectados correctamente. De nuevo no perdamos de vista el problema de heterodasticidad, y de que efectivamente hay fallas en el cálculo de la magnitud de los coeficientes regionales, en especial para el caso del PRD-6.

En síntesis, las diferencias en los coeficientes y las fallas en las pruebas de significancia estadística se deben a problemas de heterodasticidad en dos modelos: PAN-PVEM y PRD-6. Claramente la técnica SAM es más robusta que el OLS en este sentido, lo cual se debe a que su

¹⁸ La evidencia permite sostener esta tesis al menos para efectos inferenciales; no se puede argumentar que esto ocurra en todas las ciudades de la muestra.

¹⁹ El mayor efecto que se obtuvo de un coeficiente ρ fue para el PRI ($\rho = 0.110$), difícilmente significativo. Por razones de brevedad en este documento sólo se incluyeron los resultados de la técnica SAM aplicada al modelo socioeconómico y regional.

diseño tiene un enfoque autorregresivo con base en una matriz de distancias para la definición de las unidades vecinas. En este sentido la heterodasticidad latente en los modelos de regresión puede estar reflejando indirectamente la existencia de heterogeneidad espacial, y por lo tanto la posible existencia de regímenes espaciales (Anselin, 1992). Los regímenes espaciales son subáreas geográficas en donde las relaciones entre las variables son diferentes a las del resto del territorio o a otras subáreas o regiones. Pero en este trabajo esto queda para la especulación; al proseguir con estas pruebas es necesario realizar regresiones para cada región y observar si las variables incluidas predicen los resultados electorales para cada partido en la misma dirección y en una magnitud similar. El punto probado en este trabajo es que el conocimiento del problema puede llevarnos a la interpretación correcta de la realidad.

Discusión y conclusiones

Muchas discusiones metodológicas son dogmáticas, están atiborradas de evidencia anecdótica y la argumentación parece poseer una secuenciación irónicamente aleatoria. En estadística suele confiarse excesivamente en los resultados que se obtienen, aunque en ocasiones se analicen datos con base en criterios poco meditados y prejuicios ambivalentes; frecuentemente se aplican técnicas sin conocimiento de causa y efecto, y el interés principal parecen ser los resultados impresos en el *output* más que el proceso de investigación.

Debido a lo anterior este estudio persiguió tres fines pragmáticos:

- a) Presentar evidencia sobre la espacialidad de un proceso social, en este caso del comportamiento electoral.
- b) Realizar una explicación detallada de las implicaciones estadísticas del uso de modelos de regresión lineal al utilizar datos espacializados.
- c) Finalmente también llevar a efecto una comparación efectiva entre dos técnicas de regresión parcialmente diferentes y realizar una demostración de cómo esas implicaciones estadísticas afectan nuestros resultados de investigación y las interpretaciones.

Para cumplir estos fines y atendiendo a que la estadística es una disciplina principalmente empírica, se utilizaron como fuente de información para el ejercicio de demostración los resultados de las elecciones del año 2000, agregados geográficamente en ciudades. Es decir, se recurrió al examen del electorado urbano para tales efectos. Esos datos ya se han utilizado en otra investigación en donde se elaboraron explicaciones detalladas sobre el comportamiento electoral urbano mexicano (Vilalta, 2004).

Con base en estos datos de comportamiento electoral, la discusión gira alrededor de varios puntos esencialmente metodológicos pero no desconectados de la teoría. Primero, se demostró que el comportamiento electoral urbano en México posee una dimensión espacial, la cual pudo ser detectada inicialmente por medio del coeficiente de autocorrelación espacial I de Moran (1950); asimismo se explicaron las causas y los efectos de utilizar este tipo de datos en el análisis de regresión. Segundo, y en un orden lógico de ideas, se procedió a realizar una comparación entre dos técnicas de regresión parcialmente diferentes: OLS y SAM.

En forma comparativa y resumida se pudo ver en este trabajo que la técnica SAM ofreció resultados ligeramente más robustos que la técnica OLS. En la presencia de datos espacializados pudo medir con mayor precisión los coeficientes de regresión y además permitió la detección de un efecto regional para el PRI. Lo anterior se debió a que la técnica SAM incorpora la dependencia espacial de la VD en el modelo de regresión, mientras que la OLS no lo hace. La técnica SAM tiene una estructura autorregresiva que disminuye el efecto de la heterodasticidad en el modelo; en este caso fue particularmente evidente al momento de registrar con mayor precisión la magnitud en las relaciones entre las variables de la ecuación y en la existencia de un mayor número de efectos regionales significativos.

La discusión se torna fundamental cuando vemos que estas demostraciones nos indican de manera directa y evidente la necesidad de analizar el comportamiento electoral en diferentes niveles de agregación. Se mencionó anteriormente que una forma de continuar este ejercicio de heterogeneidad espacial sobre un proceso electoral, sería realizar un análisis de regresión para cada región (o grupo de unidades geográficas) y comparar los resultados de cada modelo regional. Si se encontraran relaciones inversas (con signos diferentes) significativas entre variables entre una región y otra u otras, esto constituiría una evidencia de heterogeneidad espacial.

A este respecto es indispensable reiterar que el fenómeno de la autocorrelación o dependencia espacial no debe ser entendido en una forma exclusivamente empírica o considerado un “problema” estadístico. La dependencia espacial posee una razón teórica, y es que los fenómenos sociales son función del tiempo y el espacio. Como se dijo anteriormente, una teoría geográfica elemental es que las cosas más cercanas son más similares que las cosas distantes (Tobler, 1970). En este orden de ideas, cuanto más nos alejamos más diferentes somos, y esto incluye naturalmente las diferencias en nuestro comportamiento político y en nuestras preferencias electorales.

Con base en tales puntos, la conclusión central de este trabajo es que la autocorrelación y la heterogeneidad espaciales deben ser atendidas cuidadosamente en cualquier análisis de regresión que haga uso de datos espacializados. Las implicaciones de la prueba de teorías sociales con el uso del análisis de regresión no nos permiten ser elusivos o irreflexivos al respecto. En el ejercicio se demostraron fallas en las pruebas de significancia estadística sobre los coeficientes de regresión OLS, lo que nos hubiera llevado a conclusiones erróneas y tal vez a debates académicos insustanciales producto del empleo de información equivocada como fundamento.

Sin embargo cabe aclarar que las conclusiones obtenidas a partir de este ejercicio comparativo no sugieren un rechazo a la técnica OLS; cada base de datos requiere un tratamiento especial, al igual que cada pregunta de investigación requiere un enfoque metodológico particular. La técnica OLS es correctamente aplicable al análisis de datos sobre procesos sociales en donde las observaciones efectivamente son independientes y aleatorias. Por otro lado, la técnica SAM sólo es una alternativa metodológica y una modalidad entre las técnicas de regresión espacial (Lesage, 1998). Y manteniendo una perspectiva más crítica y abierta en este aspecto metodológico, considérese la posibilidad de que el hallazgo de un contexto regional o urbano estadísticamente significativo se pueda deber simplemente a que el comportamiento atípico del electorado en alguna o algunas ciudades representa un límite dentro de una curva de probabilidad normal.

Como conclusión definitiva, de los resultados obtenidos en este trabajo de tipo metodológico se infieren tres cuestiones prácticas e importantes relacionadas con el ejercicio serio de la investigación social: 1) precisamente que la realidad social tiene una dimensión geográfica (además de temporal) ineludible que debe considerarse preliminarmente en los diseños estadísticos de investigación, 2) que nuestras

técnicas estadísticas actuales tienen limitaciones que en ocasiones omite la literatura científica pero que son reales y precisas, y 3) que los resultados de este trabajo y otros similares nos comprometen a ser fieles al método científico y al análisis inteligente de datos mediante el enlace veraz de la teoría con la evidencia empírica.

Bibliografía

- Agnew, J. (1987), *Place and Politics: The Geographical Mediation of State and Society*, Londres, Allen and Unwin.
- Ames, B. (1970), "Bases of Support for Mexico's Dominant Party", *American Political Science Review*, vol. 64, núm. 1, pp. 153-167.
- Anselin, L. (1992), *Spacestat: A Program for Statistical Analysis of Spatial Data*, Santa Bárbara, CA, National Center for Geographic Information and Analysis (NCGIA).
- (1988), *Spatial Econometrics: Methods and Models*, Dordrecht, Kluwer Academic.
- y D. Griffith (1988), "Do Spatial Effects Really Matter in Regression Analysis?", *Papers of the Regional Science Association*, núm. 65, pp. 11-34.
- Cliff, A. y J. Ord (1981), *Spatial Processes: Models and Applications*, Londres, Pion Limited.
- y J. Ord (1971), "Testing for Spatial Autocorrelation among Regression Residuals", *Geographical Analysis*, vol. 4, núm. 3, pp. 267-284.
- Domínguez, J. y J. McCann (1995), "Shaping Mexico's Electoral Arena: The Construction of Partisan Cleavages in the 1988 and 1991 Elections", *American Political Science Review*, vol. 89, núm. 1, pp. 34-48.
- Fernández-Durán J., A. Poiré y L. Rojas-Nandayapaa (2004), "Spatial and Temporal Effects in Mexican Direct Elections for the Chamber of Deputies", *Political Geography*, vol. 23, núm. 5, pp. 529-548.
- Flint, C. (1998), "Forming Electorates, Forging Spaces: The Nazi Party Vote and the Social Construction of Space", *American Behavioral Scientist*, vol. 41, núm. 9, pp. 1282-1303.
- (1995), *The Political Geography of Nazism: The Spatial Diffusion of the Nazi Party Vote in Weimar Germany*, tesis de doctorado, University of Colorado at Boulder.
- , M. Harrower y R. Edsall (2000), "But How Does Place Matter? Using Bayesian Networks to Explore a Structural Definition of Place", ponencia presentada en la New Methodologies for the Social Sciences Conference, University of Colorado at Boulder.
- Goodchild, M. (1987), "A Spatial Analytical Perspective on Geographical Information Systems", *International Journal of Geographical Information Systems*, núm. 1, pp. 327-334.

- Keppel, G. y S. Zedeck (1998), *Data Analysis for Research Designs*, Nueva York, Freeman and Company.
- King, G. (1996), "Why Context Should Not Count", *Political Geography*, vol. 15, núm. 2, pp. 159-164.
- Klesner, J. (1998), "Electoral Alignment and the New Party System in Mexico", ponencia presentada en el 1998 Congress of the Latin American Studies Association, Chicago.
- (1993), "Modernization, Economic Crisis, and Electoral Alignment in Mexico", *Mexican Studies/Estudios Mexicanos*, vol. 9, núm. 2, pp. 187-224.
- Lesage, J. (1998), *Spatial Econometrics Using MATLAB*, consultado en agosto, de 2001 en: www.econ.utoledo.edu.
- McAllister, I. (1987), "Social Context, Turnout, and the Vote: Australian and British Comparisons", *Political Geography Quarterly*, núm. 6, pp. 17-30.
- Molinar, J. y J. Weldon (1990), "Elecciones de 1988 en México: Crisis del autoritarismo", *Revista Mexicana de Sociología*, vol. 52, núm. 4, pp. 229-362.
- Moran, P. (1950), "Notes on Continuous Stochastic Phenomena", *Biometrika*, vol. 37, núm. 1 y 2, pp. 17-23.
- (1948), "The Interpretation of Statistical Maps", *Journal of the Royal Statistical Society*, vol. 10, núm. 2, pp. 243-251 (Series B, Methodological).
- O'Loughlin, J. y L. Anselin (1991), "Bringing Geography Back to the Study of International Relations: Dependence and Regional Context in Africa, 1966-1978", *International Interactions*, núm. 17, pp. 29-61.
- Ord, J. (1975), "Estimation Methods for Models of Spatial Interaction", *Journal of the American Statistical Association*, núm. 70, pp. 120-126.
- Pacheco, G. (1997), "Un caleidoscopio electoral: ciudades y elecciones en México, 1988-1994", *Estudios Sociológicos*, vol. 15, núm. 44, pp. 319-350.
- Puech, Frédéric (2004), "How Do Criminals Locate? Crime and Spatial Dependence in Minas Gerais", documento presentado en la conferencia internacional Economic Policies in the New Millennium, Coimbra, Portugal. Disponible en: <http://www4.fe.uc.pt/30years/papers/68.pdf>.
- Reyna, J. (1971), *An Empirical Analysis of Political Mobilization: The Case of Mexico*, tesis de doctorado, Cornell University.
- Tobler, W. (1970), "A Computer Movie Simulation Urban Growth in the Detroit Region", *Economic Geography*, vol. 46, núm. 2, pp. 234-240.
- Vilalta, C. (2004), "The Local Context and the Spatial Diffusion of Multiparty Competition in Urban Mexico, 1994-2000", *Political Geography*, vol. 23, núm. 4, pp. 403-423.
- (2003a), "Una aplicación del análisis espacial al estudio de las diferencias regionales del ingreso en México", *Economía, Sociedad y Territorio*, vol. 4, núm. 14, pp. 317-340.
- (2003b), "Perspectivas geográficas en la sociología urbana: la difusión espacial de las preferencias electorales y la importancia del contexto local", *Estudios Demográficos y Urbanos*, vol. 18, núm. 3, pp. 147-177.

——— (2002), *Local Contextual Effects and the Spatial Diffusion of Multiparty Competition in Urban Mexico (1994-1997)*, tesis de doctorado, Portland State University.

Sitios Web

Consejo Nacional de Población (Conapo).

Instituto Nacional de Estadística, Geografía e Informática (INEGI).

Instituto Federal electoral (IFE).