



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

Carlos J. Vilalta y Perdomo*

Documento de Trabajo
Working Paper

EGAP-2004-07

Tecnológico de Monterrey, Campus Ciudad de México

*EGAP, Calle del Puente 222, Col. Ejidos de Huipulco, 14380 Tlalpan, México, DF, MÉXICO
E-mail: carlos.vilalta@itesm.mx

Sobre la espacialidad de los procesos electorales y una comparación entre las técnicas de regresión OLS y SAM[†]

Carlos J. Vilalta y Perdomo **

Tecnológico de Monterrey, Campus Ciudad de México

Resumen

El objetivo de este trabajo es demostrar dos cosas: (1) Que debido a que algunos fenómenos sociales presentan variación espacial, la modelación de los mismos tiene implicaciones metodológicas importantes, y por lo tanto (2) es recomendable utilizar una técnica de regresión que incorpore la espacialidad de los datos en su diseño. Se procede en dos pasos generales: Primero, se calcula el coeficiente de autocorrelación I de Morán para demostrar la espacialidad del comportamiento electoral en el México urbano entre 1994 y 2000. Segundo, 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 aún en la presencia de modelos con problemas de heterodasticidad.

Abstract

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.

*Carlos J. Vilalta y Perdomo obtuvo su doctorado en Estudios Urbanos en *Portland State University* y su maestría en Estudios Urbanos en el Colegio de México. Es profesor investigador en el Tecnológico de Monterrey, Campus Ciudad de México, y candidato a investigador nacional en el SNI desde 2004.

†Quiero agradecer expresamente a Carlos Urzúa por sus sugerencias y explicaciones metodológicas y a Juan Manuel Martínez por sus comentarios al presente documento.

Dirección: Calle del Puente 222, Col. Ejidos de Huipulco; 14380 Tlalpan, México D. F.
Correo-e: carlos.vilalta@itesm.mx

I. Introducción y objetivo del trabajo

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

El análisis de datos agregados geográficamente o espacializados,¹ comúnmente denominado Análisis Espacial, tiene implicaciones metodológicas notablemente importantes para los científicos sociales que tienen preferencia por la modelación estadística (Vilalta, 2003a). Históricamente hablando, la discusión al respecto se puede considerar reciente. Aunque el tratamiento de datos espacializados se empezó a tratar en revistas científicas (tales como *Biometrika* y la *Journal of the Royal Statistical Society*) en los alrededores de la década de los cincuentas (Moran, 1948; 1950), el tema no se empezó (1) ni a discutirse sistemáticamente (2) ni a popularizarse en la literatura académica sino hasta la década de los ochentas y principios de los noventas (Cliff and Ord, 1981; Anselin, 1988; Anselin & Griffith, 1988; O’loughlin y Anselin, 1991; Flint, 1995; King, 1996). Previo a esos años, los artículos que trataban este tema eran muy ocasionales, pese a ser de una excepcional calidad científica (Cliff y Ord, 1971; Ord, 1975).

Para el caso mexicano, la atención metodológica en el tratamiento de este tipo de datos es aún más reciente, pero aparentemente 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 por parte del gobierno en poner a fácil disposición del público

¹ Como variables que capturan una variación geográfica del fenómeno bajo estudio

información estadística actualizada y series de tiempo, y (3) a la popularización de las encuestas y mapas en materia sociodemográfica, económica y política.

Desde esta perspectiva metodológica, este trabajo se concentra estrictamente en demostrar las implicaciones estadísticas que tiene el uso de datos espacializados en el análisis de regresión. El propósito principal es ofrecer una demostración empírica de la pertinencia de utilizar técnicas adecuadas con este tipo de datos y las consecuencias de no hacerlo. Para este efecto, nos enfocamos en un área particular de investigación que es el comportamiento electoral urbano. Si bien, la discusión es principalmente metodológica y en concreto estadística, las implicaciones no están separadas de explicaciones teóricas. Indistintamente, las preguntas de investigación geográfica son evidentemente teóricas, y es por ello que existen técnicas estadísticas específicas y sofisticadas para probar estas teorías.

En este documento, procedemos de la siguiente manera: Primero, a fin de plantear claramente el problema, empíricamente se detecta y mide la variación espacial no aleatoria de los procesos electorales. Segundo, se advierte sobre la importancia que la geografía tiene y los problemas que conlleva la aplicación de técnicas estadísticas a-espaciales cuando analizamos fenómenos sociales que presentan, precisamente, variaciones geográficas. Adelantamos, concretamente, que los dos fenómenos de tipo estadístico que resultan son la autocorrelación y la heterogeneidad espaciales.² Tercero y de manera pragmática, hacemos una comparación entre dos técnicas de regresión diferentes, una ordinaria (OLS) y la otra espacial (SAM), que conllevan a resultados diferentes. Lo anterior se realiza para, de manera circular, demostrar la utilidad de aplicar técnicas de estadística espacial cuando los datos así lo requieren.

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

II. Planteamiento del problema estadístico

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

1. La espacialidad de los procesos electorales en México

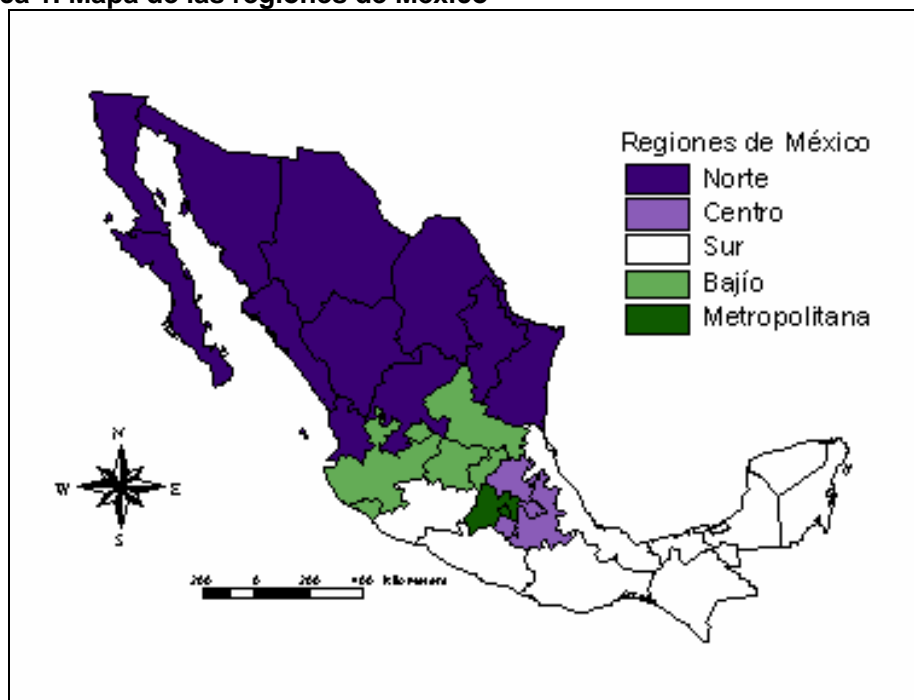
Un objetivo continuo e indispensable en los estudios electorales es el de conocer cómo se distribuye el voto geográficamente (Vilalta, 2003a). Los investigadores mexicanos han encontrado notables diferencias regionales en el comportamiento electoral (Ames, 1970; Domínguez y McCann 1995; Klesner, 1993,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

El cuadro 1 muestra la variación regional en el comportamiento del grupo de ciudades con una población mayor de 100,000 habitantes, con base en el mapa correspondiente (ver gráfica 1). El porcentaje se refiere al promedio de la región. En este respecto, es muy notable en el periodo 1994-2000 que el PAN ha 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 observar que durante este periodo de seis años su avance ha sido generalizado a través de todas las regiones. Igualmente el PRD muestra poseer un apoyo nacional menor comparado con el partido anterior, pero preferencial ubicado en el Sur. El PRI en cambio, muestra niveles de apoyo

³ No conozco en la literatura especializada Mexicana de algún estudio que contradiga esta conclusión

homogéneos entre las regiones, aunque decreciente en el tiempo; pero nótese que mantiene sus mayores niveles de apoyo en el Norte.

Gráfica 1. Mapa de las 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%
Metro	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 (n = 5), Norte (n =33), Bajío (n = 20), Metropolitana (n = 2) y Sur (n = 29).

Para corroborar la no aleatoriedad de estas diferencias regionales, a los resultados anteriores se les aplicó la prueba Kruskal-Wallis⁴, la cual mostró diferencias regionales estadísticamente significativas para todos los partidos, excepto para el PRI en 1994, el cual contaba con un apoyo muy uniforme a través de todo el país (ver cuadro 2). Paralelamente también se observa en los

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

coeficientes una tendencia a la disminución en las diferencias regionales para el PAN y el PRD, y de un aumento para el PRI entre 1994 y 2000. Es decir, el PRI urbano parece estar pasando de ser un partido de presencia nacional a tener un perfil más regionalizado en el Norte del país.

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**

** Significativo a un nivel de 0.01

* Significativo a un nivel de 0.05

^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 Mexico: 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 previamente a las elecciones del año 2000

Precisamente, se agrega a lo anterior la importancia del electorado urbano no sólo en cuanto a su comportamiento particular, sino por su participación en el auge de los partidos no priístas, el cual ocupa sin duda un lugar central en el análisis del comportamiento electoral en México (véase cuadro 3). Sin embargo, pese 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.

Es por lo anterior, es decir por contar con (1) un peso teórico probado, (2) una importancia política elemental y (3) la escasez de estudios al respecto, que la perspectiva espacial de los procesos electorales requieren precisamente una mayor difusión en la literatura científica junto con el uso de técnicas estadísticas novedosas y apropiadas.

Cuadro 3. Número de ciudades ganadas por cada partido mayoritario de las 89 ciudades 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

2. La aplicación equivocada de técnicas a-espaciales para la comprensión de fenómenos espaciales y sus implicaciones estadísticas

En cuanto a las técnicas estadísticas utilizados en los estudios de geografía electoral, los análisis correlación y de regresión lineal múltiple permanecen como los más típicos y los más avanzados dentro de la literatura Mexicana; con la excepción de algunos muy escasos estudios que muestran la utilidad de las aplicaciones espaciales a través de regresiones lineales espaciales para variables dependientes continuas (Vilalta, 2004) o bien dicotómicas (Fernández-Durán, Poiré y Rojas-Nandayapa, 2004).

A excepción de esa escasa literatura que hace uso de técnicas espaciales, la mayor parte de los estudios en geografía electoral mexicana que cuentan con un enfoque cuantitativo han ignorado y/o no han advertido al lector sobre las implicaciones metodológicas que conllevan 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 las posibles violaciones a los supuestos estadísticos de los análisis de regresión en que se está incurriendo, posiblemente por desconocimiento de causa. Es decir, es posible que estas omisiones se deban en parte a la novedad que tiene la competencia multipartidista en sí misma en México, sus respectivos análisis políticos y los estudios académicos de geografía electoral los cuales no han tenido tiempo de incorporar herramientas de análisis geográfico y estadístico avanzado; es tal

vez prematuro solicitar discusiones y advertencias metodológicas cuando hablamos de fenómenos recientes de estudio.⁵

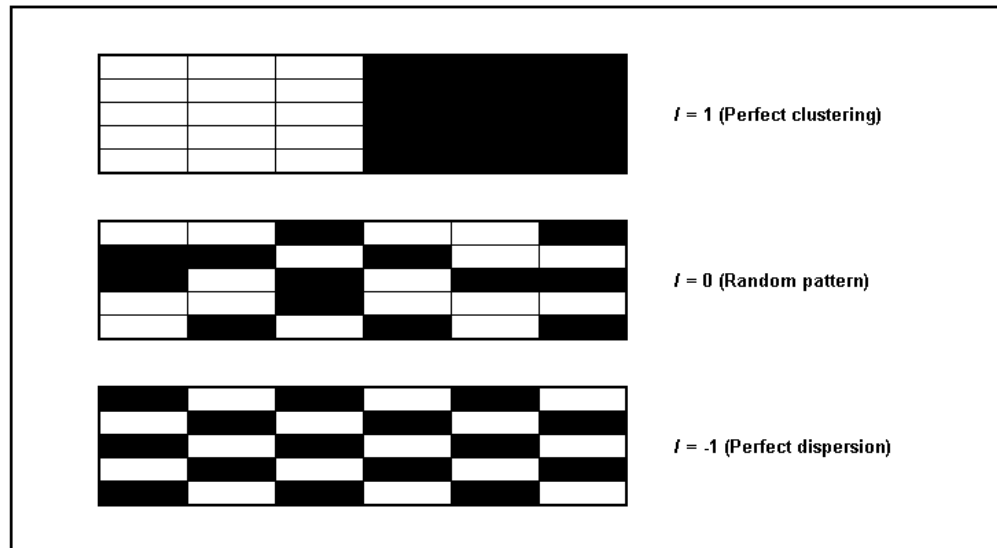
En este sentido, el punto central de este trabajo es que la aplicación de técnicas estadísticas, en concreto del análisis de regresión, a datos espacializados puede dirigirnos a cometer errores de interpretación y por lo tanto de validez en las conclusiones. La razón principal de lo anterior es la violación a un supuesto elemental en el análisis de regresión que es la independencia de las observaciones.

En el análisis de datos agregados geográficamente es natural encontrar que estas unidades de análisis estén autocorrelacionadas espacialmente o sean espacialmente dependientes. Autocorrelación y dependencia espacial significan lo mismo, pero la distinción en el uso de palabras estriba en que el primer término hace simultáneamente referencia a un fenómeno y 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, p. 4). Esto ocurre por una razón teóricamente importante que resume la primera ley geográfica de Tobler (1970): Todo se relaciona con todo, pero que las cosas más cercanas están más relacionadas que las cosas distantes.

El coeficiente I de Morán (1950) es la técnica estadística más frecuentemente utilizada para probar esta ley geográfica; o en una fase investigativa, poner a prueba la hipótesis de la presencia de una autocorrelación espacial de un fenómeno y sus correspondientes niveles 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 primer valor 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 2).

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

Gráfica 2. Representación visual de los tres diferentes patrones espaciales en que puede estar ordenada una variable en un mapa de celdas regular y sus respectivos coeficientes I de Morán



Fuente: Vilalta, 2002

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

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

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

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 las elecciones a diputados federales de mayoría relativa para cada elección entre 1994 y 2000.

Cuadro 4. Coeficientes de Autocorrelación Espacial I de Morán 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

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

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

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 justo previamente a las elecciones del año 2000

Cuadro 5. Coeficientes de Autocorrelación Espacial I de Morán 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*

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

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

a. Para el cálculo del coeficiente, la definición de ciudad vecina fue considerado un radio de 330 kilómetros 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 previamente a las elecciones del año 2000

Los coeficientes calculados para los resultados por entidad federativa muestran que es a partir de 1997 que el voto del PRI se empieza a concentrar geográficamente mientras que para el PAN y el PRD los resultados se concentran en 1997 y se vuelven a convertir en espacialmente aleatorios en las

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

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 muestran una creciente concentración del voto para el PRI mientras que el voto para el PAN y el PRD decrece en sus niveles de concentración para irse dispersando o nivelando a través del sistema de ciudades.⁸

Estos resultados muestran evidencia de que el comportamiento electoral en México tiene efectivamente una geografía no aleatoria de concentración al nivel 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 por la violación al supuesto de independencia entre las observaciones. Particularmente, el error estadístico en 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) con la dependiente (VD).

Además de dependencia espacial, los procesos electorales pueden también presentar un patrón de heterogeneidad espacial. La heterogeneidad espacial se define por una variación en las relaciones de las variables de una región o lugar a otro. (Lesage, 1998). En términos teóricos, la heterogeneidad espacial se debe a una variación real y substantiva que evidencia la existencia y la validez del contexto local o regional en la definición del comportamiento social (O'Loughlin y Anselin, 1992) y de las preferencias electorales en este caso (Flint, 1995; Vilalta, 2004). Un ejemplo de lo anterior sería 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.

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

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, p. 1281). Es decir los votantes agrupados con base en características socioeconómicas o demográficas similares pueden mostrar una preferencia diferente por razón de estar en lugares diferentes y contar con experiencias políticas diferentes. Es razonable imaginar que la composición socioeconómica y demográfica de un partido no es la misma a través de la geografía nacional. Evidentemente, en ese caso lo que determinaría a un partido como ganador en una región no serían necesariamente las características socioeconómicas 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 el en 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 ser equivocados 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 varíen de una región a otra.

En concordancia con lo anterior, una vez que ya demostramos la existencia de un patrón de autocorrelación espacial sobre los resultados electorales en el grupo de las 89 ciudades más grandes del país (ver 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 autoregresiva 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 Morán 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 unos o más variables explicativas; y/o
- Que el modelo debe tener una estructura autoregresiva a fin de eliminar la autocorrelación en el error.

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

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

En donde “*y*” es la variable dependiente, “*p*” o *rho* es el coeficiente autoregresivo de la variable dependiente (VD) espacialmente ligada o retrasada (*spatial lag*; *Wy*), *W* es precisamente la matriz de unidades vecinas, “*x*” es la matriz de variables independientes, “*β*” 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 autoregresivo 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 una

⁹ En Inglés se le denomina como *Spatial Autoregressive Model*. Léase: Lesage (1998)

¹⁰ Precisamente de la misma VD

de tipo OLS. Ya que, si (1) $y = pWy + x\beta + \varepsilon$, y si (2) $pWy = 0$, entonces, (3) $y = x\beta + \varepsilon$

En este trabajo realizamos un ejercicio comparativo entre las dos técnicas OLS y SAM con el propósito de probar si la dependencia espacial presente en las preferencias partidistas del electorado urbano en México afecta a los coeficientes de regresión OLS, y la pertinencia de utilizar la técnica SAM. Previo al ejercicio y a la explicación de los resultados, se describen las variables, información estadística y los paquetes utilizados en este estudio.

III. Variables, fuentes de información y software

El objetivo que sigue en este trabajo es comparar dos técnicas de regresión diferentes: regresión “OLS” y regresión espacial “SAM”. Para este efecto 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 para probar para la existencia de efectos sociodemográficos independientes de los efectos regionales (ver cuadro 6). Todas las variables son continuas salvo la variable regional la cual es de tipo nominal (*dummy*). Para cada partido se presenta el modelo más compacto y significativo para predecir su voto; por ende, los modelos varían entre los partidos.

Cuadro 6. Variables en los modelos de regresión

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

Este estudio utiliza una muestra de ciudades grandes; las 89 ciudades que en 1995 tenían 100,000 o más habitantes. La razón de este umbral poblacional es analizar ciudades como unidades de análisis que sean comparables en tamaño para así (1) probar la importancia del contexto local urbano o regional-urbano con unidades similares y (2) paralelamente aumentar el poder de las pruebas inferenciales (Keppel y Zedeck, 1998).¹¹ Cabe mencionarse que esta muestra urbana contiene al 56.4% de la población total en el país.

Al respecto sobre la definición de ciudad, México no tiene una definición oficial al respecto, no obstante partimos de la definición institucional realizada por el Consejo Nacional de Población (CONAPO) el cual ha definido las ciudades y las áreas metropolitanas usadas en este estudio. Una vez realizada la definición de ciudad, los datos sociodemográficos fueron agregados desde un nivel municipal. Esta información fue obtenida del Instituto Nacional de Estadística, Geografía e Informática (INEGI). Finalmente, los resultados electorales fueron obtenidos a nivel municipal y posteriormente agregados con base en la definición de las ciudades, y obtenidos también desde el 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 Morán, se predefinieron en la matriz de vecinos a aquellas ciudades que se encontraran en un radio de 330 kilómetros, la cual es la distancia mínima en la cual todas las ciudades en 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*.

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

¹² La información del INEGI e IFE fue obtenida vía internet

IV. Procedimiento y resultados de la comparación de técnicas OLS y SAM

Primero que nada calculamos los coeficientes de autocorrelación espacial de cada una de las VI's utilizadas en este estudio. Podemos observar (ver cuadro 7) que con la excepción de Población y Migración, todas las variables muestran un patrón geográfico de concentración. Particularmente, se observa una alta concentración espacial en el caso de Catolicismo y moderada de Educación.

Cuadro 7. Coeficientes de Autocorrelación Espacial / de Morán para las ciudades grandes de México, en diferentes variables socioeconómicas (2000)^a

Variables Socioeconómicas	Coefficiente
Población	-0.045
Lengua Indígena	0.167**
Migración	0.052
Manufacturas	0.066*
Educación	0.291**
Ingreso	0.083**
Catolicismo	0.533**

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

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

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

A continuación realizamos los análisis de regresión OLS y SAM para cada partido. Se debe reiterar en que para cada partido se buscó el mejor modelo predictivo con base en las variables socioeconómicas seleccionadas. El ejercicio comparativo entre las dos técnicas entonces siguió el siguiente procedimiento de tres pasos:

a) Con base en la técnica OLS, se buscó el mejor modelo exclusivamente socioeconómico para cada partido; es decir aquel que sólo contara con variables significativas, ya fuera un modelo amplio (seis o más variables; de hasta siete posibles) o compacto (dos o menos).

b) Una vez definido el modelo socioeconómico, igualmente con el uso de la técnica OLS se agregaron a este modelo las variables (nominales o dummies) regionales para probar la hipótesis de efectos regionales diferenciados e independientes de las variables socioeconómicas. Aquí nos enfocamos en observar los cambios en la R^2 y en vislumbrar problemas de heterodasticidad.

c) Una vez teniendo los resultados anteriores, se aplicó la técnica SAM al modelo socioeconómico y regional, y se compararon los resultados con los anteriores provenientes de la técnica OLS. Esta técnica permite la inclusión de dos pruebas más en el modelo predictivo: Una prueba para detectar y medir efectos espaciales o de autocorrelación en la VD con base en el coeficiente rho (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 sección en donde resumimos los resultados de todos los modelos resultantes con 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).

1. El caso del PAN-PVEM

El cuadro 8 muestra el mejor modelo con base en la técnica OLS para el PAN-PVEM utilizando las variables socioeconómicas previamente elegidas y predefinidas en la literatura especializada como útiles para predecir el voto agregado geográficamente (ver cuadro 6). Podemos ver que el 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).

Cuadro 8. Resultados del modelo de regresión OLS para la coalición del PAN-PVEM sin variables regionales

Dependent Variable: PAN-PVEM				
Method: Least Squares				
Included observations: 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
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-squared	0.358	Mean dependent var		43.438
Adjusted R-squared	0.328	S.D. dependent var		13.341
S.E. of regression	10.934	Akaike info criterion		7.676
Sum squared resid	10043.45	Schwarz criterion		7.816
Log likelihood	-336.594	F-statistic		11.751
Durbin-Watson stat	1.828	Prob(F-statistic)		0.000
White Heteroskedasticity Test:				
F-statistic	0.873	Probability	0.542	
Obs*R-squared	7.151	Probability	0.520	

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 a 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.

Tal y como podemos ver en los resultados (ver cuadro 9), la significancia estadística de las variables socioeconómicas se mantuvo (salvo el caso de Catolicismo que como se mencionó en el párrafo anterior fue retirada del modelo). En cuanto a las variables regionales, el Bajío ofrece resultados positivos y significativamente superiores a la región Norte para la coalición PAN-PVEM; las demás regiones no presentan un patrón significativamente diferente a la región Norte. Este modelo OLS tiene una mayor capacidad explicativa que el anterior ($R^2 = .452$), sin embargo muestra problemas de heterodasticidad (White test = 18.86, $p < .05$). Lo anterior se debe a la inclusión de las variables regionales las cuales, recordemos, pueden causar 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 equivocados. 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.

Cuadro 9. Resultados del modelo de regresión OLS para la coalición del PAN-PVEM con variables regionales

Dependent Variable: PAN-PVEM				
Method: Least Squares				
Included observations: 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
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
Metro	1.145	7.497	0.152	0.879
Sur	2.429	3.337	0.727	0.468
R-squared	0.452	Mean dependent var	43.438	
Adjusted R-squared	0.405	S.D. dependent var	13.341	
S.E. of regression	10.286	Akaike info criterion	7.585	
Sum squared resid	8569.948	Schwarz criterion	7.808	
Log likelihood	-329.534	F-statistic	9.578	
Durbin-Watson stat	2.094	Prob(F-statistic)	0.000	
White Heteroskedasticity Test:				
F-statistic	2.097	Probability	0.034	
Obs*R-squared	18.862	Probability	0.042	

Nota: se utiliza como región de referencia a la Norte

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

Cuadro 10. Resultados del modelo de regresión SAM para la coalición del PAN-PVEM

Dependent Variable: PAN-PVEM				
Method: Spatial Regression*				
n = 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
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
Metro	0.121	7.293	0.016	0.986
Sur	2.262	3.017	0.749	0.455
Error Estándar Residual:				
SRE	10.160			
Efecto Espacial:				
Rho	-0.031			
Prueba de Autocorrelación Espacial:**				
Coef. I de Morán	-0.005	0.033	0.171	0.864

Nota: se utiliza como región de referencia a la Norte

* El paquete no calcula la R^2

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

En este caso, en cuanto al efecto espacial, el coeficiente autoregresivo ρ muestra ser de pequeña magnitud y cercano a cero por lo que muy probablemente no sea estadísticamente significativo.¹³ En este caso, este modelo espacial se convertiría en uno típico de OLS. A esto se agrega que la prueba I de Morán sobre los residuales también muestra 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's. Es decir, tal y como lo habíamos mostrado en el cuadro 7, cada una de estas variables están concentradas en el espacio, y en este caso, la no significancia estadística del coeficiente autoregresivo puede deberse a que éstas compartan en buena parte la misma geografía que el voto por el PAN-PVEM. Dicho de

¹³ Desafortunadamente, el módulo Space-Stat no calcula la significancia estadística de este 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 ρ es igualmente de muy pequeña magnitud y no hay autocorrelación espacial en los residuales; es decir, las tres VI's socioeconómicas utilizadas son suficientes para neutralizar los efectos espaciales en la VD.

otra manera, la no existencia de un contexto local puede deberse a una concordancia espacial (*spatial match*) entre estas variables socioeconómicas con las preferencias electorales.

Sin embargo, no perdamos de vista de que tal modelo muestra problemas de heterodasticidad por lo que los resultados deben ser tomados 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 indicativo de la presencia de heterogeneidad espacial, es decir, de variaciones en las relaciones entre las variables de una región a otra (Flint, 1995).

2. El caso del PRI

En 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 autoregresivo (efecto espacial) y la prueba / de Morán 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.

Cuadro 11. Resultados del modelo de regresión OLS para el PRI sin variables regionales

Dependent Variable: PRI				
Method: Least Squares				
Included observations: 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constante	58.879	8.647	6.808	0.000
Socioeconómica:				
Catolicismo	-0.297	0.112	-2.633	0.010
R-squared	0.073	Mean dependent var		36.202
Adjusted R-squared	0.063	S.D. dependent var		7.868
S.E. of regression	7.615	Akaike info criterion		6.920
Sum squared resid	5046.034	Schwarz criterion		6.976
Log likelihood	-305.964	F-statistic		6.936
Durbin-Watson stat	1.466	Prob(F-statistic)		0.009
White Heteroskedasticity Test:				
F-statistic	0.986	Probability		0.376
Obs*R-squared	1.996	Probability		0.368

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, como región de referencia, a la región Norte.

En este caso (ver cuadro 12) el modelo aumenta en 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 al modelo anterior sin variables regionales. A este respecto, hay dos regiones que muestran significancia estadística: Bajío y Sur. En ambos casos, tienen efectos negativos sobre el PRI, lo que sugiere que el PRI obtiene resultados significativamente positivos en el Norte sobre ambas y viceversa. Los resultados que obtiene en el Norte escasamente no se diferencian significativamente de los resultados que obtiene en las ciudades del Centro y en la región Metropolitana (zona metropolitana de la Ciudad de México y zona metropolitana de Toluca).

Cuadro 12. Resultados del modelo de regresión OLS para el PRI con variables regionales

Dependent Variable: PRI				
Method: Least Squares				
Included observations: 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constante	64.809	9.118	7.107	0.000
Socioeconómica:				
Catolicismo	-0.319	0.120	-2.654	0.009
Efecto Regional:				
Bajío	-5.045	2.162	-2.332	0.022
Centro	-5.496	3.322	-1.654	0.101
Metro	-7.911	5.066	-1.561	0.122
Sur	-8.105	1.782	-4.546	0.000
R-squared	0.270	Mean dependent var		36.202
Adjusted R-squared	0.226	S.D. dependent var		7.868
S.E. of regression	6.919	Akaike info criterion		6.771
Sum squared resid	3973.812	Schwarz criterion		6.939
Log likelihood	-295.334	F-statistic		6.159
Durbin-Watson stat	1.703	Prob(F-statistic)		0.000
White Heteroskedasticity Test:				
F-statistic	0.601	Probability		0.728
Obs*R-squared	3.749	Probability		0.710

Nota: se utiliza como región de referencia a la Norte

A continuación utilizamos la técnica SAM (ver cuadro 13) sobre el modelo anterior, y obtenemos que la variable Catolicismo mantiene casi el mismo efecto sobre la VD, pero un cambio interesante es que se agrega la región Centro como significativamente diferente a la región de referencia Norte; igualmente teniendo un efecto negativo como son los casos de la región Bajío y Sur.

El coeficiente autoregresivo es de muy pequeña magnitud ($\rho = 0.015$) por lo que es 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

¹⁵ Igualmente corrí una SAM con la variable Catolicismo como la única VI, sin variables regionales, y mantenía su capacidad para predecir el voto por el PRI además de carecer de autocorrelación espacial en los residuales.

Cuadro 13. Resultados del modelo de regresión SAM para el PRI

Dependent Variable: PRI				
Method: Spatial Regression*				
n = 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constante	63.684	9.600	6.335	0.000
Socioeconómicas:				
Catolicismo	- 0.305	0.126	- 2.408	0.018
Efecto Regional:				
Bajo	- 4.095	2.403	-1.704	0.092
Centro	- 5.950	3.567	- 1.667	0.099
Metro	-8.432	5.174	-1.629	0.107
Sur	- 7.758	1.966	- 3.944	0.002
Error Estándar Residual:				
SRE	6.894			
Efecto Espacial:				
Rho	0.015			
Prueba de Autocorrelación Espacial:**				
Coef. I de Morán	0.007	0.033	0.558	0.576

Nota: se utiliza como región de referencia a la Norte

* El paquete no calcula la R^2

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

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

3. El caso del PRD-6

Para el caso de la coalición PRD-6, en el cuadro 14 podemos observar 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 (ver cuadros 8 y 14 respectivamente).

Cuadro 14. Resultados del modelo de regresión OLS para la coalición PRD-6 sin variables regionales

Dependent Variable: PRD-6				
Method: Least Squares				
Included observations: 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
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-squared	0.337	Mean dependent var		17.224
Adjusted R-squared	0.321	S.D. dependent var		13.044
S.E. of regression	10.74	Akaike info criterion		7.619
Sum squared resid	9921.569	Schwarz criterion		7.703
Log likelihood	-336.051	F-statistic		21.895
Durbin-Watson stat	1.708	Prob(F-statistic)		0.000
White Heteroskedasticity Test:				
F-statistic	2.143	Probability		0.082
Obs*R-squared	8.243	Probability		0.083

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 contar con sólo la mitad del número de VI's; dos versus cuatro. Pero lo que más llama la atención es que el modelo presenta problemas de heterodasticidad (White test = 8.243, $p < .10$). Esto es diferente a los casos de 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, el problema de heterodasticidad se acentúa (White test = 16.499, $p < .05$), por lo que la interpretación que hagamos sobre la magnitud de los coeficientes en el modelo y su significancia estadística debe ser hecha con reservas (ver cuadro 15). A este respecto se puede ver que las variables socioeconómicas muestran tener efectos de una magnitud similar (aunque un poco más reducida) y dirección igual al modelo anterior. Por otro lado, sólo la región Bajío parece tener un comportamiento diferente a la región Norte, en este caso, con un efecto negativo.

Cuadro 15. Resultados del modelo de regresión OLS para la coalición PRD-6 con variables regionales

Dependent Variable: PRD				
Method: Least Squares				
Included observations: 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
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
Metro	4.918	7.385	0.666	0.507
Sur	3.440	3.078	1.117	0.267
R-squared	0.437	Mean dependent var		17.224
Adjusted R-squared	0.396	S.D. dependent var		13.044
S.E. of regression	10.134	Akaike info criterion		7.544
Sum squared resid	8420.464	Schwarz criterion		7.740
Log likelihood	-328.750	F-statistic		10.635
Durbin-Watson stat	1.866	Prob(F-statistic)		0.000
White Heteroskedasticity Test:				
F-statistic	2.275	Probability		0.030
Obs*R-squared	16.499	Probability		0.035

Nota: se utiliza como región de referencia a la Norte

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

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

Dependent Variable: PRD				
Method: Spatial Regression*				
n = 89				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
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
Metro	-3.797	8.026	-0.473	0.637
Sur	-0.104	3.816	-0.027	0.978
Error Estándar Residual:				
SRE	10.160			
Efecto Espacial:				
Rho	0.046			
Prueba de Autocorrelación Espacial:**				
Coef. I de Morán	-0.057	0.033	-1.362	0.173

Nota: se utiliza como región de referencia a la Norte

* El paquete no calcula la R^2

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

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

Ya que 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, nos concentramos 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 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:

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)

- 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.

Cuadro 17. Comparación entre las técnicas OLS y SAM para cada partido según los modelos con variables socioeconómicas y con variables socioeconómicas y regionales

	Modelo con Variables Socioeconómicas		Modelo con Variables Socioeconómicas y Variables Regionales		
	Magnitud y signo de los coeficientes	Significancia Estadística	Magnitud y signo de los coeficientes	Significancia Estadística	Regiones significativas y su efecto*
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 (-) Sur (-) Centro (-)
PRD-6	OLS ≠ SAM	OLS ≈ SAM	OLS ≠ SAM	OLS ≈ SAM	OLS: Bajío (-)
					SAM: Bajío (-)

*Nota: se utiliza como región de referencia a la Norte

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 de detectar a tres regiones (versus dos con OLS) como significativamente diferentes de la región de referencia, que es la región Norte. Para el PRD-6, la 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 fue la técnica SAM 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 también 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 aquellos partidos que estuvieron menos concentrados espacialmente en el año 2000 (ver cuadro 5). Esto significa que los coeficientes de la regresión pueden no estar capturando la magnitud real de la relación además de potencialmente estar cometiendo errores del tipo II en las pruebas de hipótesis sobre los mismos coeficientes (Anselin, 1988 y 1992). Esto se puede deber a la existencia de heterogeneidad espacial substantiva, es decir, a patrones de comportamiento electoral regionales distintivos (O'Loughlin and Anselin, 1992; Flint, Harrower y Edsall, 2000).

Cuadro 18. Detección de un efecto espacial, autocorrelación espacial en los residuales y heterodasticidad para cada partido

	Efecto Espacial (ρ)	Autocorrelación en los residuales (I de Morán)	Heterodasticidad en el modelo*
PAN-PVEM	n .s.	n .s.	Si
PRI	n. s.	n. s.	No
PRD-6	n. s.	n. s.	Si

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

En esta cuestión, el modelo del PRI es diferente al de los otros partidos. El modelo 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 sean ineficientes. No dejemos de considerar que con la técnica SAM se pudo detectar una región más como significativamente diferente de la región Norte (cuadro 17), frente a la OLS, debido a la inclusión del coeficiente autoregresivo. Se agrega, ahora sí, de manera coincidente con los modelos de los otros dos partidos-coaliciones, que el efecto espacial medido por el coeficiente *rho* (ρ) es cercano a cero y no presenta una autocorrelación en los residuales. Esto nos permite concluir que pese a su pobre capacidad explicativa (sólo pudimos detectar que tiende a ganar en ciudades con bajos porcentajes de población católica) el modelo para el PRI

¹⁷ Aún careciendo de significancia estadística en el modelo

captura de forma suficiente la variación espacial en la VD y en el cual las variables regionales posiblemente están capturando una variación relacionada con 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 rho, que recordemos se refiere al valor promedio de la variable dependiente en las ciudades vecinas, no produjo ningún efecto 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 los partidos, 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 explicar la concentración espacial del comportamiento electoral. Es decir, que 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 concentración de población con ciertas características socioeconómicas.¹⁸ Para probar a detalle 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 el caso del PAN-PVEM y PRD-6, los modelos presentaban problemas de heterodasticidad, por lo que la tesis anterior sigue abierta a posteriores pruebas.

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

¹⁹ El mayor efecto que se obtuvo por parte de un coeficiente rho 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.

Por otro lado, las pruebas *t* de Morán 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 el hecho de 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 diseño tiene un enfoque autoregresivo 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 indirectamente reflejando la existencia de heterogeneidad espacial y por lo tanto a la posible existencia de regímenes espaciales (Anselin, 1992). Los regímenes espaciales son subáreas geográficas en donde las relaciones entre las variables son diferentes al del resto del territorio o a otras subáreas o regiones. Pero esto queda a la especulación en este trabajo; para proseguir con estas pruebas, es necesario realizar regresiones para cada región y observar si las variables incluidas predicen en la misma dirección y en una magnitud similar los resultados electorales para cada partido. El punto probado en este trabajo es que el conocimiento de este problema puede llevarnos a la interpretación correcta de la realidad.

V. Discusión y conclusiones

Muchas discusiones metodológicas son dogmáticas, están atiborradas de contra-ejemplos, evidencia anecdótica y la argumentación parece poseer una secuenciación irónicamente aleatoria.²⁰ En estadística, además de que se

²⁰ Al menos en muchas en las que ha participado el autor tanto en clases, como talleres y seminarios

tiene una excesiva confianza en los resultados que se obtienen, en ocasiones se analizan datos con base en criterios poco meditados y prejuicios ambivalentes; 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.

Es por lo anterior que este estudio persiguió tres fines empíricos y 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 realizar una comparación efectiva entre dos técnicas de regresión parcialmente diferentes y una demostración de cómo esas implicaciones estadísticas afectan nuestros resultados de investigación y las interpretaciones.

Para cumplir estos fines y debido a que la estadística es una disciplina principalmente empírica, se utilizó 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 estos efectos. Esos datos ya se han utilizado en otra investigación y en ella se elaboran explicaciones detalladas sobre el comportamiento electoral urbano mexicano (Vilalta, 2004). Y en este respecto, la pregunta novedosa y substancial en este trabajo es la siguiente: ¿por qué deben preocuparnos el uso de técnicas apropiadas de regresión?

Con base en estos datos de comportamiento electoral, la discusión a continuación 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 a través del coeficiente de autocorrelación espacial I de Morán (1950). Unido al punto anterior, 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.

De 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, la técnica SAM a diferencia de la técnica OLS pudo medir con mayor precisión coeficientes de regresión además de permitir 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, al tener una estructura autoregresiva, disminuyó el efecto de la heterodasticidad en el modelo, en este caso, lo que fue particularmente evidente al momento de efectivamente registrar con mayor precisión la magnitud en la relaciones entre las variables de la ecuación y en la existencia de un mayor número de efectos regionales significativos.

La discusión se torna imprescindible al ver 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 encuentran relaciones inversas (con signos diferentes) significativas entre variables entre una región y otra/s, esto constituiría evidencia de heterogeneidad espacial.

A este respecto, es indispensable redundar en que el fenómeno de la autocorrelación o dependencia espacial no debe ser entendido de una forma exclusivamente empírica o como 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 diferencias en nuestro comportamiento político y en nuestras preferencias electorales.

Con base en los puntos anteriores, la conclusión central de este trabajo es que la autocorrelación y la heterogeneidad espaciales deben ser atendidos cuidadosamente en cualquier análisis de regresión que haga uso de datos espacializados. Las implicaciones que tiene la prueba de teorías sociales con el uso del análisis de regresión no nos permite ser elusivos o irreflexivos al respecto. En el ejercicio se demostraron fallos 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 por utilizar como fundamento información equivocada.

Sin embargo, sí se debe 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 respecto 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 (s) ciudades representa un límite dentro de una curva de probabilidad normal.

Como conclusión definitiva, los resultados obtenidos en este trabajo de tipo metodológico presentan 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 en ocasiones omitidas en la literatura científica pero que son reales

y precisas, y (3) que los resultados en este trabajo nos comprometen a ser fieles con el método científico y con el análisis inteligente de datos a través del enlace veraz de la teoría con la evidencia empírica.

Bibliografía y sitios Web

Libros y artículos:

Agnew, J. (1987). *Place and Politics: The Geographical Mediation of State and Society*. London: Allen and Unwin.

Ames, B. (1970). Bases of Support for Mexico's Dominant Party. *American Political Science Review*, 64(1), 153-167.

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht: Kluwer Academic.

Anselin, L. y Griffith, D. (1988). Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association*, 65, 11-34.

Anselin, L. 1992. *Spacestat: A program for statistical analysis of spatial data*. Santa Barbara, CA: National Center for Geographic Information & Analysis (NCGIA)

Cliff, A. y Ord, J. (1971). Testing for spatial autocorrelation among regression residuals, *Geographical Analysis*, 4(3), 267-284.

Cliff, A. y Ord, J. (1981). *Spatial Processes: Models and Applications*. London: Pion Limited.

Domínguez, J. y McCann, J. (1995). Shaping Mexico' electoral arena: the construction of partisan cleavages in the 1988 and 1991 elections, *American Political Science Review*, 89(1), 34-48.

Fernández-Durán J., Poiré A. y Rojas-Nandayapaa L. (2004) Spatial and temporal effects in Mexican direct elections for the chamber of deputies. *Political Geography*, 23(5), pp. 529-548.

Flint, C. (1995). *The political geography of Nazism: the spatial diffusion of the Nazi party vote in Weimar Germany*. Tesis doctoral. University of Colorado at Boulder.

Flint, C. (1998). Forming Electorates, Forging Spaces: The Nazi Party Vote and the Social Construction of Space. *American Behavioral Scientist*, 41(9), 1282-1303.

Flint, C., Harrower M., y Edsall, R. (2000). But How Does Place Matter? Using Bayesian Networks to Explore a Structural Definition of Place. Paper presented at the 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*, 1, 327-334.

Keppel, G. y Zedeck S. (1998). *Data Analysis for Research Designs*. New York: Freeman and Company.

King, G. (1996) Why Context Should Not Count. *Political Geography*, 15(2), 159-164.

Klesner, J. (1993). Modernization, Economic Crisis, and Electoral Alignment in Mexico. *Mexican Studies/Estudios Mexicanos*, 9(2), 187-224.

Klesner, J. (1998). *Electoral Alignment and the New Party System in Mexico*. Paper presented at the 1998 Congress of the Latin American Studies Association. Chicago, IL.

LeSage, J. (1998). *Spatial Econometrics using MATLAB*. Consultado en Agosto, 2001, en: www.econ.utoledo.edu

McAllister, I. (1987). Social context, turnout, and the vote: Australian and British comparisons. *Political Geography Quarterly*, 6, 17-30.

Molinar, J. y Weldon, J. (1990). Elecciones de 1988 en México: Crisis del autoritarismo. *Revista Mexicana de Sociología*, 52(4), 229-362.

Moran, P. (1948) The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*. Series B (Methodological), 10(2), 243-251.

Moran, P (1950) Notes on continuous stochastic phenomena. *Biometrika*, 37(1 y 2), pp. 17-23.

O'Loughlin, J. y Anselin, L. (1991). Bringing geography back to the study of international relations: dependence and regional context in Africa, 1966-1978. *International Interactions*, 17, 29-61.

Ord, J. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, 120-126.

Pacheco, G. (1997) Un caleidoscopio electoral: ciudades y elecciones en Mexico, 1988-1994 . *Estudios Sociológicos*, 15(44), 319-350. El Colegio de México.

Reyna, J. (1971). *An Empirical Analysis of Political Mobilization: The Case of Mexico*. Tesis doctoral. Cornell University.

Tobler, W. (1970). A Computer Movie Simulation Urban Growth in the Detroit Region. *Economic Geography*, 46(2), 234-240.

Vilalta, C. (2002). *Local Contextual Effects and the Spatial Diffusion of Multiparty Competition in Urban Mexico (1994-1997)*, Tesis doctoral, Portland State University.

Vilalta, C. (2003a). Una Aplicación del Análisis Espacial al Estudio de las Diferencias Regionales del Ingreso en México. *Economía, Sociedad y Territorio*, 4(14), 317-340. El Colegio Mexiquense.

Vilalta, C. (2003b). Perspectivas Geográficas en la Sociología Urbana: La Difusión Espacial de las Preferencias Electorales y la Importancia del Contexto Local. *Estudios Demográficos y Urbanos*, 18(3), 147-177. El Colegio de México.

Vilalta, C. (2004). The Local Context and the Spatial Diffusion of Multiparty Competition in Urban Mexico, 1994–2000. *Political Geography*, 23(4), 403-423.

Sitios Web:

Consejo Nacional de Población (CONAPO)
Instituto Nacional de Estadística, Geografía e Informática (INEGI)
Instituto Federal electoral (IFE).