

II Congreso de la Asociación Latinoamericana de Población

Guadalajara, México, 3 – 5 de Septiembre de 2006

**La demografía latinoamericana del siglo XXI
Desafíos, oportunidades y prioridades**

Métodos para el análisis espacial. Una aplicación al estudio de la geografía de la pobreza

Landy L. Sánchez Peña

University of Wisconsin-Madison
llsanchez@wisc.edu

Mesa 10. Métodos, Técnicas y Datos en la
Investigación Demográfica en América Latina y el
Caribe

Métodos para el Análisis Espacial. Una Aplicación al Estudio de la Geografía de la Pobreza¹

Landy L. Sánchez Peña²

Mesa: *Métodos, Técnicas y Datos en la Investigación Demográfica en América Latina y El Caribe*

En los estudios demográficos han sido recurrentes las preguntas acerca de cómo se distribuyen espacialmente los grupos poblacionales (socioeconómicos, étnicos, de migrantes, etc.) y cuáles factores explican dicha distribución. Más aún, también ha existido el interés por entender cómo ciertos procesos demográficos ocurren de manera distinta en diversas regiones y hasta que punto localidades cercanas geográficamente se influyen entre sí en la adopción de prácticas sociales (por ejemplo, en el control de la fertilidad o los patrones de consumo). Este tipo de preguntas implican, sin embargo, retos metodológicos en la estimación de efectos espaciales como la *autocorrelación* y la *heterogeneidad*. Recientes desarrollos estadísticos y de software permiten modelar mejor dichos efectos espaciales y plantearnos nuevas preguntas sobre la dimensión espacial de los procesos demográficos.

Este trabajo presenta tres métodos para el análisis espacial: 1) Análisis Exploratorio de Datos Espaciales, 2) Modelos de Regresión Espacial y 3) Regresión Ponderada Geográficamente. Estas técnicas permiten visualizar la distribución geográfica de las variables, estimar la presencia de clústeres y la heterogeneidad en su comportamiento en el espacio y desarrollar modelos explicativos que consideren dicho comportamiento. Para mostrar su utilidad a los estudios demográficos, se aplican estas técnicas al estudio de la pobreza urbana en Guadalajara, México. Los datos provienen del Censo de Población y Vivienda 2000 agregados por Agebs (la unidad censal más pequeña) y son vinculados al mapa digital de la cartografía censal.³ A lo largo del texto se discutirán las aplicaciones, ventajas y limitaciones de estos métodos, a la par que se analizará la geografía de la pobreza en Guadalajara.

La geografía de la pobreza urbana

En la abundante literatura sobre la pobreza han recobrado centralidad preguntas concernientes a su localización y distribución espacial. Son más frecuentes los estudios que se enfocan a identificar variaciones regionales en los niveles de pobreza, la ubicación de zonas de alta marginalidad e incluso las características geofísicas de los lugares donde se asientan los pobres (zonas de riegos). Estos estudios se han beneficiado de los Sistemas de Información Geográfica que permiten el rápido mapeo de las variables, facilitando su análisis y la comunicación de los hallazgos a un público más amplio. Sin embargo, los métodos estadísticos de análisis espacial han sido menos utilizados para examinar la geografía de la pobreza. En parte, ello se debe

¹ Trabajo presentado en el II Congreso de la Asociación Latinoamericana de Población, realizado en Guadalajara, México, del 3 al 5 de septiembre de 2006

² Candidata a Doctora en Sociología por la Universidad de Wisconsin-Madison llsanchez@wisc.edu

³ Una Ageb incluye entre 20 y 50 manzanas y no debe estar dividida por carreteras, aeropuertos, ríos o grandes áreas deshabitadas.

a los problemas matemáticos y computacionales que tales métodos implicaban, así como a la poca información georeferenciada disponible.

Nuevos métodos y programas desarrollados en los últimos años facilitan esta tarea, permitiendo un análisis más completo de la dimensión espacial de los procesos demográficos. En este trabajo presentamos las posibilidades que ofrecen por un lado Geoda, un software gratuito disponible en línea⁴ que permite implementar *Análisis Exploratorio de Datos Espaciales y Modelos de Regresión Espacial*, y por otro lado, GWR⁵ que implementa *Regresiones Ponderadas Geográficamente*. Estos tres métodos consideran la ubicación del fenómeno de estudio tanto en términos absolutos (dónde ocurren) como relativos (distribución espacial, distancia). Permiten no sólo la fácil visualización de nuestras variables, sino explorar cómo éstas se “relacionan” en el espacio, qué tipo de distribución presentan (clústeres, dispersión) y permiten estimar modelos explicativos que consideren dicha distribución. En el estudio de la pobreza urbana estos métodos podrían ayudarnos a responder preguntas tales como ¿Cómo está distribuida la pobreza en el espacio urbano? ¿Es posible identificar clústeres de exclusión social? En tal caso, ¿a qué responden dichos clústeres? ¿Tienen las variables explicativas de la pobreza el mismo efecto en toda la ciudad? O ¿Es posible identificar regiones en que una variable explicativa importa más que en otras áreas?

Datos

Este trabajo explora las preguntas anteriores en el caso de Guadalajara, la segunda ciudad más grande de México. Los datos provienen del cuestionario básico del Censo de Población y Vivienda 2000 agregados por Ageb (Área Geoestadística Básica) y que es unidad geográfica censal más pequeña. En el 2000 el área metropolitana de Guadalajara tenía una población de 3.2 millones de personas, dividida en 7 municipios y 1146 Agebs.

Los datos por Ageb que el INEGI ha publicado incluyen variables poblacionales, educativas, de empleo, estado marital, hogares, migración, disponibilidad de servicios de salud y características de la vivienda. Las variables relacionadas al ingreso, sin embargo, son limitadas. De hecho, la única variable que mide los niveles de pobreza a nivel Ageb considera sólo pobreza *extrema*⁶, ya que se define como pobres a aquellos cuyo ingreso per capita del hogar es menor de dos dólares, la cantidad mínima necesaria para cubrir necesidades alimenticias de acuerdo al Banco Mundial. Bajo esta definición, 5.8 % de población en Guadalajara es clasificada como pobre. Sin embargo, cuando analizamos sus valores por Ageb esta variable muestra una distribución muy dispersa: su promedio es de 6.2 por ciento, pero su recorrido abarca un 65.9 por ciento. Esta definición es claramente insuficiente para dar cuenta de la pobreza urbana, pero es la única disponible al público. Para complementar el análisis, también se analiza la distribución espacial de trabajadores de bajo ingreso, aquí definida como aquellos que ganan menos de dos salarios mínimos por día (aproximadamente \$2270 pesos en el 2000). Ellos representan el 31 por ciento de aquellos que reciben ingresos por trabajo, teniendo una media por Ageb del 32.5 % y una desviación estándar del 13.2 por ciento. Un análisis a profundidad sin duda requiere mejores mediciones de pobreza, pero dado que el objetivo central de

⁴ <https://geoda.uiuc.edu/downloadin.php>

⁵ Geographically Weighted Regression por su nombre en inglés

⁶ Dicha variable está incluida en la base de datos Regiones Socioeconómicas de México, disponible en línea <http://www.inegi.gob.mx/est/contenidos/espanol/sistemas/regsoc/default.asp?>

este trabajo es ilustrar la utilización de métodos espaciales, por ahora estas dos variables son suficientes. El resto de las variables a utilizar en los modelos multivariados están definidas en el apéndice A.

Efectos Espaciales: autocorrelación y heterogeneidad

La *autocorrelación* o *dependencia espacial* se presenta cuando una variable tiende a asumir valores similares en unidades geográficamente cercanas dando lugar al surgimiento de clústeres, por ejemplo áreas urbanas pobres tenderían a estar cerca de otras áreas pobres. Estimar la dependencia espacial en nuestros datos puede ser necesario por motivos metodológicos y teóricos. Por un lado, la presencia de clústeres espaciales violaría el supuesto de la independencia de las observaciones y generaría problemas en la correcta estimación de modelos de regresión lineal de mínimos cuadrados. Por otro lado, la dependencia espacial puede ser de interés en sí misma dado que puede expresar un proceso de “contagio” o influencia recíproca entre las unidades de observación o bien, puede ser producto de fuerzas económicas, sociales o políticas que tienden a agrupar a poblaciones con rasgos comunes en ciertas áreas (Voss et al 2005).

Por otro lado, la *heterogeneidad* espacial indica la presencia de diferencias *sistemáticas* en la ocurrencia de un fenómeno en distintas regiones geográficas, de tal forma que este podría tener diferentes distribuciones (media, varianza u otros parámetros) en un subgrupo espacial de los datos o bien, simplemente cambiar con la ubicación de las unidades (Anselin 1992). Entonces, podríamos hablar de heterogeneidad si la región Sur tiene niveles medios de pobreza distintos de la región Norte o si las variables explicativas tienen un efecto distinto en ciertas regiones (Voss et al 2005). Dichas diferencias regionales pueden ser el objeto de nuestro estudio o podemos sólo estar interesados en evaluar cuán estables estructuralmente son nuestras estimaciones.

Análisis Exploratorio de Datos Espaciales (AEDE) con Geoda

El AEDE proporcionado por Geoda es una serie de técnicas para visualizar y estimar la autocorrelación espacial. El estudio AEDE permite “mapear” cómo se distribuye la pobreza en el ámbito urbano e identificar la presencia de clústeres de exclusión social, es decir, es posible identificar zonas de la ciudad donde se agrupan Agebs con altos niveles de pobreza. En seguida, la cuantificación de dichos clústeres se realiza a través de dos medidas: I de Moran y Indicadores Locales de Asociación Espacial (LISA por sus siglas en inglés).

Las figuras 1 y 2 presentan la proporción de pobres y trabajadores de bajos ingresos en las Agebs de Guadalajara. Estos primeros mapas muestran una fuerte heterogeneidad en la distribución de los grupos por ingreso en la ciudad: el patrón general muestra un grado importante de “mezcla” entre las Agebs caracterizadas por niveles medios y bajos de pobreza. Dicho panorama concuerda con estudios que sugieren que las ciudades mexicanas tienen mayores niveles de heterogeneidad de lo que hubiésemos podido predecir basados sólo en sus niveles de ingreso (Sánchez-Peña 2006, Roberts 1995).

Aunque dicha heterogeneidad parece ser el rasgo fundamental de Guadalajara, también es clara la presencia de zonas de alta pobreza en las orillas de la ciudad (particularmente al este), mientras que una gran zona de baja pobreza se localiza en la zona centro y centro-occidente de la ciudad. Una imagen muy similar ofrece el mapa de trabajadores de bajo ingreso: las Agebs con altas proporciones de trabajadores de

bajos ingresos tienden a ubicarse en la periferia y las unidades con bajas proporciones están localizadas en el medio-oeste, mientras que el medio-este está mucho más mezclado. Dicho de otra manera, ambos mapas (figura 1 y 2) sugieran la presencia de clústeres ya sea de áreas donde la pobreza o el bienestar se concentran.

Una manera de cuantificar dicha tendencia es a través de la *I de Moran*, la cual mide la tendencia de valores similares a agruparse en el espacio, es decir, hasta que punto áreas con altos niveles de pobreza están cerca de otras áreas de alta pobreza mientras que zonas de poca pobreza están rodeadas de zonas de baja pobreza. Para calcularla, debemos definir primero una matriz de pesos geográficos la cual determina que observaciones son consideradas vecinas entre sí. En este caso usamos una estructura de “reina” de primer orden de contigüidad, la cual define como vecinas aquellas Agebs adyacentes entre sí que tienen puntos en común (fronteras o vértices)⁷.

Las figuras 3 y 4 muestran la gráfica y valores de *I de Moran* para nuestras dos variables de ingreso. En ambos casos hay evidencia de autocorrelación espacial positiva; es decir, existe una significativa aglomeración de Agebs con similares niveles de pobreza y trabajadores de bajos salarios⁸. La gráfica para la variable pobreza, muestra que las unidades con bajos índices contribuyen más a la autocorrelación espacial (cuadrante inferior izquierdo), mientras que las unidades con bajos niveles de los trabajadores pobres tienden a concentrarse tanto como aquellas con altos niveles (cuadrante izquierdo inferior y cuadrante superior derecho). La diferencia en la forma en que ambas variables se aglutinan puede reflejar los niveles de privación económica que cada una refleja: dado que una mide pobreza *extrema* es probable que ciertas colonias populares aparezcan como zonas de baja pobreza aún cuando sean zonas donde una alta proporción de los trabajadores gane menos de 2 salarios mínimos.

Una imagen más intuitiva y específica de la presencia de clústeres de pobreza es proporcionada por Indicadores Locales de Asociación Espacial (LISA). Estos indicadores miden la asociación espacial entre el valor que una variable asume en la unidad *i* y los valores que asume en las Agebs vecinas –vecindad definida también a través de la matriz de pesos geográficos-. Por ello, LISA ofrece una manera de identificar clústeres locales y de observar no-estacionariedad a través del espacio (Logley y Tobon 2004). Los clústeres rojos y azules mostrados por los mapas de LISA (figuras 5 y 6) señalan dónde es posible identificar estadísticamente la aglomeración de valores similares. Los mapas en tonos verdes muestran si dichos clústeres son estadísticamente significativos.

Las Agebs con altos niveles de pobreza (clústeres rojos) forman clústeres en diversas zonas periféricas de la ciudad, mientras que las áreas azules representan clústeres de baja pobreza. Por otro lado, las áreas coloreadas por rosa y azul pálido representan zonas donde Agebs con disímiles niveles de pobreza se agrupan (alto-bajo y bajo-alto respectivamente). LISA, entonces, puede ayudarnos a identificar zonas de alta y baja marginalidad (clústeres de valores similares) y la presencia de casos extremos o atípicos (áreas disímiles) (Logley y Tobon 2004). La figura 5 ilustra una mayor concentración de la pobreza y el bienestar de lo que hubiésemos supuesto basados exclusivamente en el mapa 1. Más aún, el mapa de LISA muestra la existencia de barrios altamente pobres junto a zonas de muy baja pobreza, ubicación

⁷ Geoda permite establecer distintas matrices de pesos geográficos, algunas basadas en adyacencia y otras en distancia.

⁸ Ambos fueron significativas con un valor de $p < .0001$. El test de significancia se realiza contra la hipótesis de distribución espacial aleatoria de las unidades, una hipótesis no muy restrictiva en el caso de los fenómenos sociales.

que podría responder a una búsqueda de oportunidades de empleo de los pobres (Portes 1989). De igual manera muestra la existencia de enclaves de bienestar en zonas altamente pobres, quizá mostrando el desarrollo de fraccionamientos cerrados, lo que ha sido documentado en el caso de Guadalajara (Cabrales 2001). El mapa de LISA de los trabajadores de bajo ingreso (figura 6) ofrece una imagen muy similar. La localización de los clústeres de valores parecidos (rojo y azul) es básicamente la misma que de los clústeres de pobreza, pero la extensión de los clústeres es mayor en el primer caso. De nuevo, esto podría reflejar las diferencias en el nivel de privación económica que estas variables miden: los bajos salarios están más extendidos que la pobreza extrema. Por ello, el clúster rojo (alto-alto) es más grande, mientras que el clúster azul mantiene proporciones similares.

En suma, el análisis exploratorio con las medidas de I de Moran y LISA nos permiten analizar el grado de concentración geográfica de la pobreza y la riqueza, identificar clústeres dentro de la ciudad y analizar la heterogeneidad por ingreso que caracteriza a las Agebs de Guadalajara.

Modelos de Regresión Espacial

Modelos de Regresión Espacial

En el análisis exploratorio anterior, las medidas de I de Moran para las variables de pobreza y de bajos ingresos fueron positivas y altamente significativas, lo cual fuertemente sugiere la presencia de autocorrelación espacial –es decir, la formación de clústeres de pobreza y bajos ingresos. La I de Moran, sin embargo, provee sólo una mirada limitada de este fenómeno, ya que el clúster de valores similares (bajo-bajo, alto-alto) pudiera estar sólo reflejando la distribución de variables explicativas de la pobreza. Por ejemplo, los clústeres de trabajadores de bajos ingresos en una zona podrían estar reflejando la aglomeración de trabajadores subempleados o con baja educación. Hipotéticamente, sin embargo, es posible que existan otros fenómenos que contribuyen a la formación de clústeres más allá de las características de la población que habita una determinada Ageb. Por ejemplo, estudios sobre el crecimiento de las áreas pobres muestran que un asentamiento original puede expandirse gracias a la migración de nuevos pobres a las zonas dónde hay tierra disponible. Este efecto muchas veces se genera por las redes sociales: los habitantes originales de una colonia popular informan de la disponibilidad de tierra e “invitan” a familiares y conocidos a mudarse en áreas adyacentes. En tales casos uno podría hablar de un efecto de “contagio” o de aprendizaje entre las unidades vecinas: la probabilidad de un asentamiento pobre en una zona aumenta las probabilidades de que la Ageb vecina también sea pobre. También es posible pensar la autocorrelación espacial como resultado de las regulaciones de crecimiento urbano que promueven el establecimiento de ciertos grupos en ciertas zonas y no otras. Dicha regulación actuaría como una fuerza estructural generando clústeres espaciales. El ejemplo más claro son las políticas de segregación residencial por raza y etnicidad que prevalecieron en Estados Unidos hasta tiempos recientes.

Substantivamente hay dos preguntas a hacerse: ¿El nivel de pobreza en una Ageb dada refleja sólo la composición social de dicha Ageb? o ¿Está influenciado por los niveles de pobreza en las unidades vecinas? Para examinar la dependencia espacial después de controlar otras características relevantes podemos implementar *modelos multivariados de regresión espacial*. Peor aún, si las medidas exploratorias revelan la presencia de dependencia espacial debemos implementar dichos modelos,

incluso si no tenemos ningún interés sustantivo en ella, esto porque ignorar la autocorrelación entre nuestras unidades geográficas podría comprometer la fiabilidad de nuestro análisis (Anselin 1992).

Anselin (1992) distingue dos tipos de dependencia espacial: una es ruido o un “estorbo” para nuestro análisis y la otra es sustantivamente relevante. La primera está en el término del error y viola el supuesto de la no correlación entre los errores, hecho en la regresión de mínimos cuadrados. Si ignoramos este tipo de autocorrelación los estimadores de nuestra regresión serán *ineficientes*, nuestras estadísticas de t y F estarán sesgadas y la bondad de nuestro ajuste (R^2) será engañoso (Anselin 1992). Es considerada como ruido porque no afectará los coeficientes del modelo y porque sólo nos interesa para mejorar nuestras estimaciones. La autocorrelación en el término del error puede ser manejada a través de un Modelo Espacial del Error (estimada con máxima verosimilitud). Este modelo asume que la dependencia espacial encontrada en nuestra variable dependiente es resultado de la distribución geográfica de nuestras variables explicativas y de la autocorrelación del término del error, esto último sugiere que nuestro modelo no está bien especificado (Anselin 1992a, Baller et al 2001).

Por otro lado, la autocorrelación espacial *substantiva* se presenta cuando el valor que toma la variable dependiente en cada unidad geográfica está realmente determinado por el valor que ésta asume en las unidades vecinas. Es decir, la autocorrelación estará todavía presente después de otras variables explicativas. Si ignoramos este tipo de dependencia espacial los coeficientes estimados con la regresión de mínimos cuadrados estarán sesgados. La alternativa para modelar este tipo de autocorrelación es el Modelo Espacial Lag, que considera la dependencia espacial introduciendo una variable espacial lag. De acuerdo con Baller et al (2001) un modelo espacial lag representa la relación interactiva entre las variables independientes y la dependiente en las unidades vecinas.

Geoda ofrece un camino para determinar que modelo necesitamos implementar⁹. Este software permite estimar una regresión estándar de mínimos cuadrados y una serie de diagnósticos espaciales con los cuáles podemos decidir qué tipo de autocorrelación espacial está presente en nuestros datos y, por tanto, el modelo a utilizar. La tabla 1 muestra los resultados para un modelo que estima el nivel de pobreza de las Agebs como una función de la inserción laboral, escolaridad y estructuras familiares y relación de dependencia prevalente en cada área¹⁰. Véase en el apéndice A la definición de cada una de las variables explicativas. El modelo implementado no pretende ser realmente exhaustivo, sino sólo ilustrativo del tipo de análisis a realizar con regresiones espaciales. Lo que buscamos es mostrar hasta dónde nuestros datos sugieren autocorrelación substantiva o no, y cuánto cambian nuestras estimaciones al implementar una regresión espacial.

En nuestro ejemplo, los resultados de la regresión de mínimos cuadrados sugieren que la proporción de trabajadores desempleados, los empleados menos de 32 horas a la semana y la proporción de hogares con jefatura femenina incrementan

⁹ Geoda también ofrece otra serie de diagnósticos sobre la presencia de heteroskedasticidad, sin embargo no tiene capacidad para modelar este efecto espacial de manera directa. Una posibilidad es hacerlo con SpaceStat, software comercial que también fue desarrollado por Luc Anselin, creador de Geoda. En nuestro modelo se detectó la presencia de heteroskedasticidad lo que implica que la autocorrelación espacial detectada podría estar “contaminada” por ella.

¹⁰ Los modelos espaciales fueron calculados con la matriz de pesos geográficos tipo reina, de primer orden de contigüidad, la misma utilizada para el cálculo de LISA e I de Moran. La variable dependiente fue transformada a logit ($p/1-p$) con el fin de asegurar que los valores predichos cayeron dentro del rango 0 y 1 y asegurar la confiabilidad del modelo

significativamente los niveles de pobreza por Ageb. Por el contrario, la escolaridad promedio, el trabajo en ocupaciones altamente informales y la relación de dependencia reducen dichos niveles. Los resultados no son sorprendentes, aunque si llama la atención el hecho de que ni la proporción de ocupados por cuenta propia ni la proporción de hogares ampliados sean significativos.

Los exámenes de diagnóstico confirman la presencia de autocorrelación espacial, pero sugieren que ésta se encuentra en el término del error¹¹. Es decir, los niveles de pobreza responden a la composición social de los Agebs y la autocorrelación en los errores sugeriría la necesidad de especificar otro modelo. Como se mencionó anteriormente, un Modelo Espacial del Error (MEE) considera dicha dependencia espacial y mejora la eficiencia de nuestra estimación. La tabla 1 muestra que los coeficientes no cambian mucho entre la regresión de mínimos cuadrados y el MEE, sin embargo la estimación del error estándar mejora. En nuestros datos, ello se traduce en un cambio en la significancia del coeficiente de hogares ampliados, el cual se vuelve significativo. Este parecería un cambio menor, y es probable que en nuestro caso lo sea, pero podría ser que dicha variable fuera la central en un estudio dado.¹²

En síntesis, el análisis multivariado sugiere que los clústeres de pobreza encontrados en el análisis exploratorio responden a la distribución geográfica de las variables explicativas incluidas en el modelo y otras no consideradas. La evidencia no sugiere procesos de aprendizaje o contagio entre las unidades o fuerzas estructurales guiando la concentración de la pobreza per se. Sin embargo, esta conclusión debe ser considerada con cuidado. Un análisis completo requeriría analizar la distribución de las variables a lo largo del tiempo, para confirmar o refutar la presencia de autocorrelación substantiva.

Además, el análisis anterior muestra una limitación importante del AEDE y la regresión lineal: incluso cuando es posible detectar dependencia espacial estos métodos no permiten explicarla. Ambos métodos ofrecen una mirada indirecta a los procesos de interacción entre las unidades geográficas, pero tienen poca capacidad para discriminar cuáles mecanismos están detrás de ella.

Regresión Geográficamente Ponderada

Mientras las Geoda se aboca a modelar la autocorrelación espacial, la regresión geográficamente ponderada (GWR por sus siglas en inglés) busca analizar la no-estacionariedad de los datos. Así esta técnica hace posible explorar si la asociación entre pobreza y sus variables explicativas es constante en toda la ciudad o si es posible identificar variaciones por zonas. Esto es posible porque una regresión geográficamente ponderada permite la estimación de parámetros locales y no sólo globales. Un estimado local es computado “tomando prestada” información de las unidades dentro de una distancia previamente establecida, donde las unidades más cercanas tienen mayor peso que las más lejanas (Fotheringham et al. 2001:52). Como tal, esta técnica cuestiona el supuesto implícito en las regresiones estándares de que un modelo explicativo puede aplicarse por igualmente a

¹¹ la regla es que se escoge el modelo con el valor del Multiplicador de Lagrange más grande y significativo.

¹² A pesar de que la serie de diagnósticos espaciales sugiere que el MEE es el modelo más adecuado, la tabla 1 muestra los resultados para el modelo lag, sólo con propósitos ilustrativos. El tamaño, signo y significancia de los coeficientes en éste último modelo no son muy distintos de la regresión de mínimos cuadrados, lo que es de esperarse ya que nuestros datos no muestran la presencia de autocorrelación substantiva.

toda el área geográfica analizada, cuando en realidad puede haber importantes variaciones tanto en el modelo completo como en la relación específica entre la variable dependiente y una de sus variables explicativas (Charlton et al. 2003, Fotheringham et al. 2001). Así, nuestro modelo explicativo podría funcionar mejor en ciertas zonas de la ciudad que en otras y algunas variables importar más en ciertas regiones que en otras o incluso tener efectos opuestos.

En esta apartado se implementa el mismo análisis multivariado que en la sección anterior y se analiza con especial atención cómo varía en la ciudad el efecto que las variables de inserción laboral tienen sobre los niveles de pobreza. GWR provee tanto los resultados de la regresión estándar (aquí llamada *global*) y los resultados de los cálculos locales. Los resultados de la regresión estándar generados por GWR son los mismos, por supuesto, que los de Geoda. La tabla 2 sólo presenta los resultados de las estimaciones locales, mostrando cuánto varían los coeficientes a través del espacio. Debido a la grande que sería un archivo con 1416 regresiones locales, GWR solo proporciona el resumen de 5 números de su distribución. De esta tabla podemos deducir que el valor que los coeficientes asumen en distintas áreas de la ciudad varía importantemente. Por ejemplo, el coeficiente para la variable de *trabajo por pocas horas* llega a tener un valor de -0.085 en algunas zonas, mientras en otras tiene un efecto positivo sobre los niveles de pobreza de hasta 3.97. En cambio, el coeficiente del desempleo varía en su tamaño pero siempre tiende a aumentar los niveles de pobreza.

Charlton et al. (2003:21) sugieren una manera tentativa de evaluar si la relación entre la variable dependiente y la explicativa es no-estacionaria: si el rango intercuartil es mayor que 2 errores estándares respecto de la media global, esto sugiere que la relación es no estacionaria. En nuestro ejemplo, la relación entre pobreza y trabajo por cuenta propia, trabajo por pocas horas, trabajo informal, fuerza femenina ampliada, hogares con jefatura femenina, hogares ampliados y la relación de dependencia muestran claros signos de no estacionariedad, es decir, que sus efectos no son los mismos a lo largo de la ciudad. Sin embargo, GWR provee un test formal para estimar si la variabilidad espacial de los parámetros es significativa. Los resultados muestran que sólo dos variables lo son: relación de dependencia y trabajo por cuenta propia. Dicha variabilidad puede ser más fácil de observar si mapeamos las parámetros locales calculados por GWR. La figura 7 muestra la variación que el parámetro *relación de dependencia* asume en Guadalajara. Aún cuando esta variable tiene un efecto negativo en prácticamente todas las zonas de la ciudad su efecto sobre la pobreza es mayor en el oeste y centro-oeste de la ciudad, las áreas donde clusters de bienestar fueron claramente identificados. Por su parte, la figura 8 presenta los coeficientes locales para la variable de trabajo por cuenta propia, misma que muestra todavía mayor variación. Mientras esta variable tiene un efecto negativo en las regiones oeste y centro-oeste, tiende a incrementar la pobreza en la región este de la ciudad. Esto podría simplemente representar la variedad de ocupaciones que se clasifican como por cuenta propia, reflejando puestos de mejor nivel en ciertas zonas y más precarios en las zonas más pobres. Si examinamos otras variables podemos apreciar que también muestran importantes variaciones, aún cuando el test de Monte Carlo sugiere que dicha variación podría ser aleatoria.

GWR también permite estimar cuán bien se ajusta nuestro modelo a los datos, calculando un R^2 local. La figura 9, muestra que nuestro modelo se desempeña mejor en la zona este que en el resto de la ciudad. Es decir, los niveles de pobreza extrema en dicha zona pueden ser mejor explicados por nuestras variables independientes, pero éstas tienen menos poder predictivo en otras zonas de Guadalajara. Diagnósticos tales como el coeficiente de determinación y el AIC sugieren que el modelo local es mejor que el modelo global.

Además de la idea de que las variables tienen efectos diferenciados a través de la ciudad, los mapas de GWR permiten identificar la presencia de los llamados *regímenes*

espaciales que pueden definirse como regiones geográficas donde se observan cambios discretos en la distribución de las variables (media, varianza) y donde un tipo distinto de relación parece establecerse entre las variables consideradas. En nuestro caso, la región Este de Guadalajara parece constituir un régimen espacial, hipótesis que puede ser explorada con mayor detenimiento utilizando Geoda y modelos de regresión lineal con regímenes espaciales, los cuales calculan coeficientes específicos para cada subgrupo de los datos (estos modelos pueden implementarse con SpaceStat o R).

A manera de conclusión: límites y posibles contribuciones del Análisis Espacial

Los métodos estadísticos que consideran la ubicación y distribución espacial de los procesos sociales están desarrollándose rápidamente, pero aún tienen limitaciones importantes tanto por el desarrollo de los métodos en sí mismos como por la disponibilidad de datos georeferenciados. La literatura señala tres problemas comunes en el análisis espacial:

- a) los datos georeferenciados disponibles tienden a ser de áreas geográficas administrativamente delimitadas (estados, municipios, localidades, agebs) que pueden o no ser las más adecuadas para el fenómeno en cuestión. Es probable, por ejemplo, que los mercados de trabajo no sean bien capturados con estas fronteras administrativas, pero sí lo sean las políticas públicas de empleo que utilizan dichas unidades como referencia de planeación (Voss et al 2005, Messner y Anselin 2002). De cualquier manera, existe el riesgo de que el fenómeno poblacional estudiado no se ajuste a las unidades geográficas en las que la información esté disponible o no ocurra en la escala para la que tenemos datos. Ello limita no sólo el tipo de análisis que podemos emprender sino también las conclusiones que podemos obtener con los métodos anteriormente expuestos. Por ejemplo, es común que el Modelo Espacial del Error sea favorecido por los datos cuando existe dicho desajuste entre los datos que tenemos y la “verdadera” dimensión espacial en la que un fenómeno ocurre.
- b) Dado que la mayor parte de los métodos se han abocado a modelar datos agregados (por área o polígono) se corre el riesgo de la *falacia ecológica*, es decir de tratar de inferir los comportamientos de los individuos a partir de las características del área en la que viven. Claramente, esto no puede hacerse. Los modelos sólo tienen sentido para entender que pasa en el nivel agregado y tratar de inferir comportamientos “colectivos” (Messner y Anselin (2002). Es posible, sin embargo, mezclar métodos de análisis espacial con otros métodos que pueden descomponer ambos niveles de análisis, por ejemplo, los modelos multinivel o jerárquicos (Sampson et al 1997).
- c) Más allá del riesgo de falacia ecológica está el problema de cuál puede ser la interpretación substantiva de la autocorrelación o heterogeneidad espacial, particularmente cuando sólo tenemos información para un punto en el tiempo. En el caso de la dependencia espacial, sólo tendremos mayores certezas con datos de más de un punto en el tiempo, ya que con datos de un momento en el tiempo no es posible distinguir cuando un efecto espacial es aparente y cuando existe un efecto real de contagio, aprendizaje o “spillover” (Messner y Anselin 2002). El análisis de la heterogeneidad es un poco más permisivo al respecto, en este caso el mayor reto es identificar el tipo de variación en el espacio que se observa (continua o discreta) y desarrollar explicaciones de porque dicha heterogeneidad está presente.

A pesar de estas limitaciones, los métodos de análisis espacial pueden ser de gran utilidad para los estudios de población. En el diseño de políticas de combate a la pobreza se vuelven cada vez más necesarios, sobre todo ahora que muchas son focalizadas a veces con bases geográficas. Este tipo de métodos puede identificar claramente áreas donde los pobres

tienden a estar concentrados, lo que puede facilitar la provisión de servicios a poblaciones desfavorecidas. Hipotéticamente, es posible desarrollar políticas urbanas que disminuyan la concentración geográfica de la pobreza, un objetivo deseable dado los estudios que documentan que dicha concentración puede favorecer la estigmatización de sus habitantes y un deterioro mayor de sus condiciones de vida. Además, ya que podríamos saber cuáles son las variables que más influyen en los niveles de pobreza de cada zona (por ejemplo educación o empleo o estructuras familiares) es posible desarrollar políticas de pobreza más sensitivas y, quizás, más eficientes.

Referencias

Anselin, L. (1992). *Spatial Data Analysis with GIS: an Introduction to Application in the Social Science*, Santa Barbara: University of California Santa Barbara. Technical Report 92-10.

Anselin, L. (1995). *Exploring Spatial Data with Geoda: a Workbook*. Electronic document. Available on-line at <https://geoda.uiuc.edu/pdf/geodaworkbook.pdf>. Downloaded, June 20, 2005.

Anselin, L. (1992b). *SpaceStat Tutorial. A workbook to using SpaceStat in the Analysis of Spatial Statistics*. Urbana, Ill: University of Illinois-Champaign.

Ariza M and Solis P (2005) “Dinámica de la desigualdad social y segregación espacial en tres áreas metropolitanas de México” Paper presented at the XIII Internacional Meeting of Population Studies, France, July 2005.

Baller, R, L Anselin, S Messner, G Deane, D Hawkins (2001). “Structural covariates of U.S. county homicide rates: incorporating spatial effects” *Criminology* 39: 561-590.

Cabrales L (2001) *Latinoamerica, Paises abiertos, ciudades cerradas*. Guadalajara, Mexico:Universidad de Guadalajara-UNESCO

Charlton, Martin, Stewart Fotheringham, Cris Brunsdon. 2003. *Geographically Weighted Regression and Associated Statistics Workbook*. Santa Barbara California: CSISS Summer Workshop.

Fotheringham, Stewart, Martin E. Charlton and Chriss Brunsdon. 2001. “Spatial Variations in School Performace: a Local Analysis Using Geographically Weighted Regression. *Geographical & Environmental Modelling*, Vol. 5, No.1, 2001. 43-66

Garcia B y Oliveira O (2003) “Trabajo e Ingreso de los Miembros de las Familias en el México Metropolitano” in E. De la Garza y C. Salas (Ed) *La situación del Trabajo en Mexico 2003*, Mexico City: IET, UAM, Plaza y Valdez.

Logley, Paul and Carolina Tobón. 2004. “Spatial Dependence and Heterogeneity in Patterns of Hardship: An intra-Urban Analysis” *Annals of the Association of American Geographers*, 94 (3): 503-519

Messner y Anselin (2002) “Spatial Análisis of Homicida with Areal Data” mimeo, disponible en línea <http://sal.uiuc.edu/users/anselin/papers/smla.pdf>, consultado Marzo del 2006

Portes A and Roberts B (2004) “The free market city: Latin American urbanization in the years of the neoliberal experiment”. Draft

Portes A et al (1997) *The Urban Caribbean. Transition to the new global economy*. Baltimore: The John Hopkins Press

Portes, A (1989) “Latin American Urbanization during years of the crisis” *Latin American Research Review*, 24, (3), 7-44

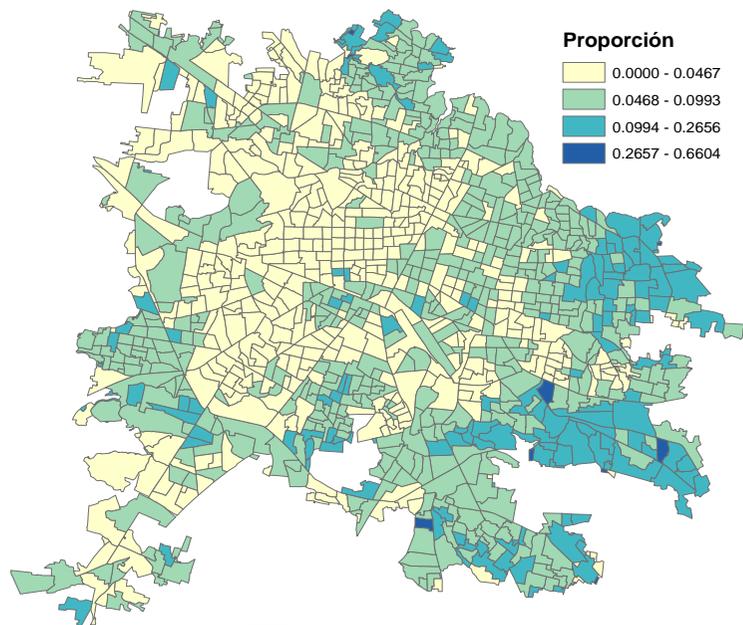
Roberts B R (1995) *The Making of Citizens*, New York: Halsted Press

Sampson R J, Raudenbush S and F Earls (1997) “Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy” *Science* 277, 918-924

Sánchez Peña, L “Precarious Employment and the Geography of Poverty in Guadalajara” Paper presented at the 2006 Population Association of America Meeting, Los Angeles, April

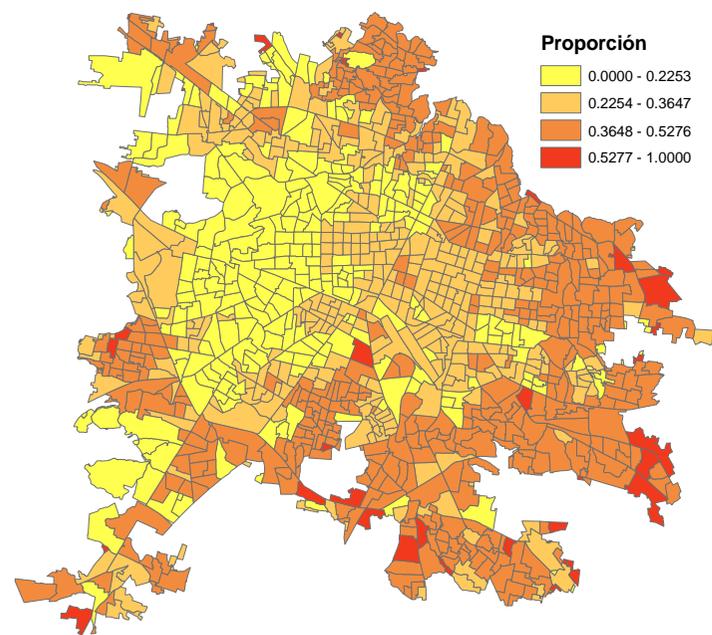
Voss P, D Logg, R Hammer y S Friedman (2005) “County Child Poverty Rates in the US: A Spatial Regression Approach” mimeo

Figura 1. Pobreza Extrema. Guadalajara 2000



Censo de Población y Vivienda 2000

Figura 2. Trabajadores de Bajos Ingresos Guadalajara 2000



Censo de Población y Vivienda 2000

Figura 3

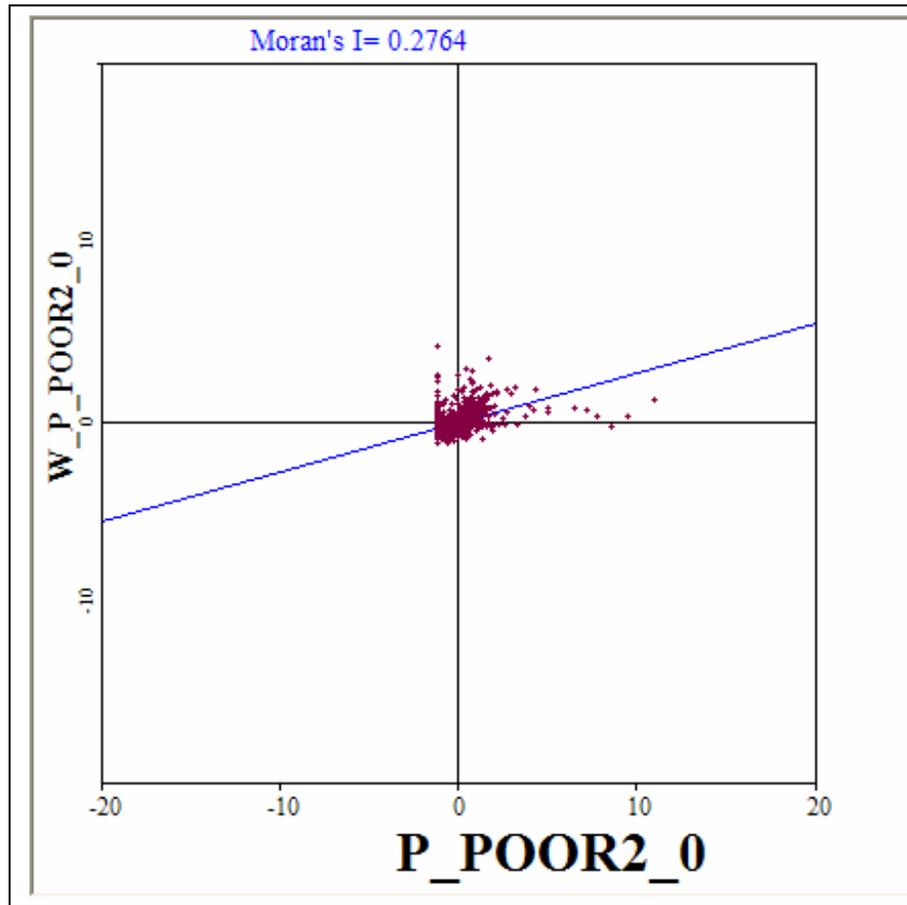


Figura 4

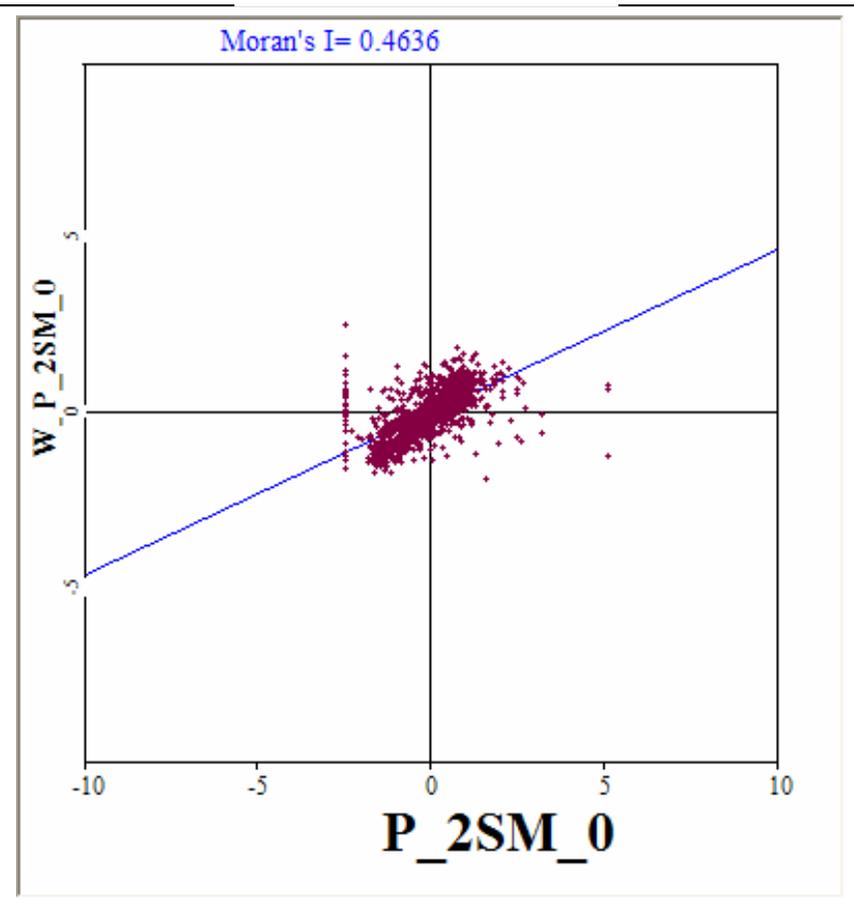


Figura 5

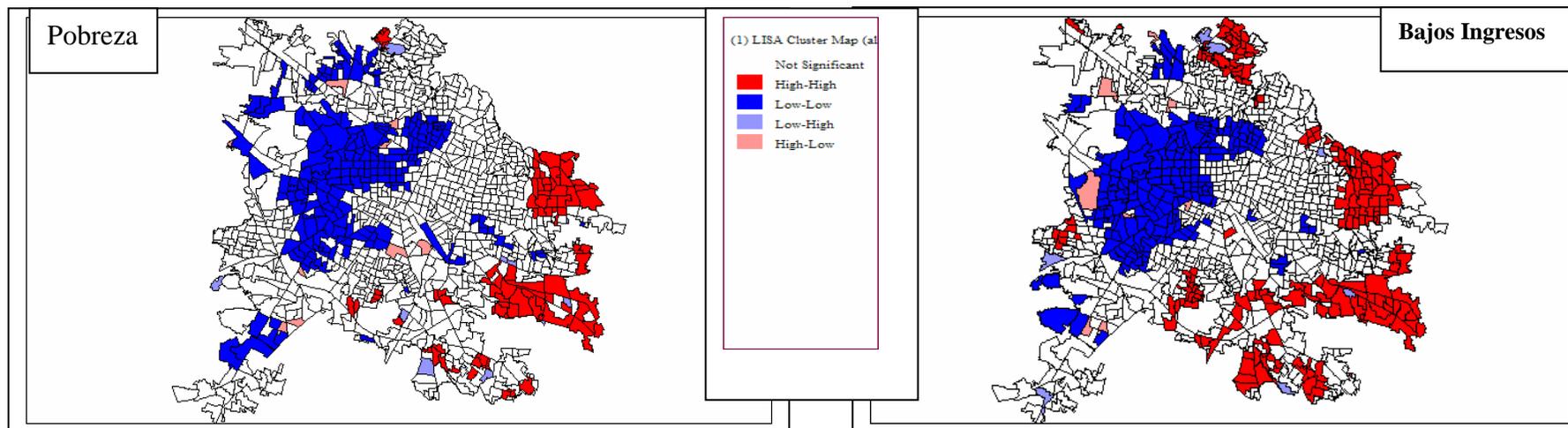


Figura 6

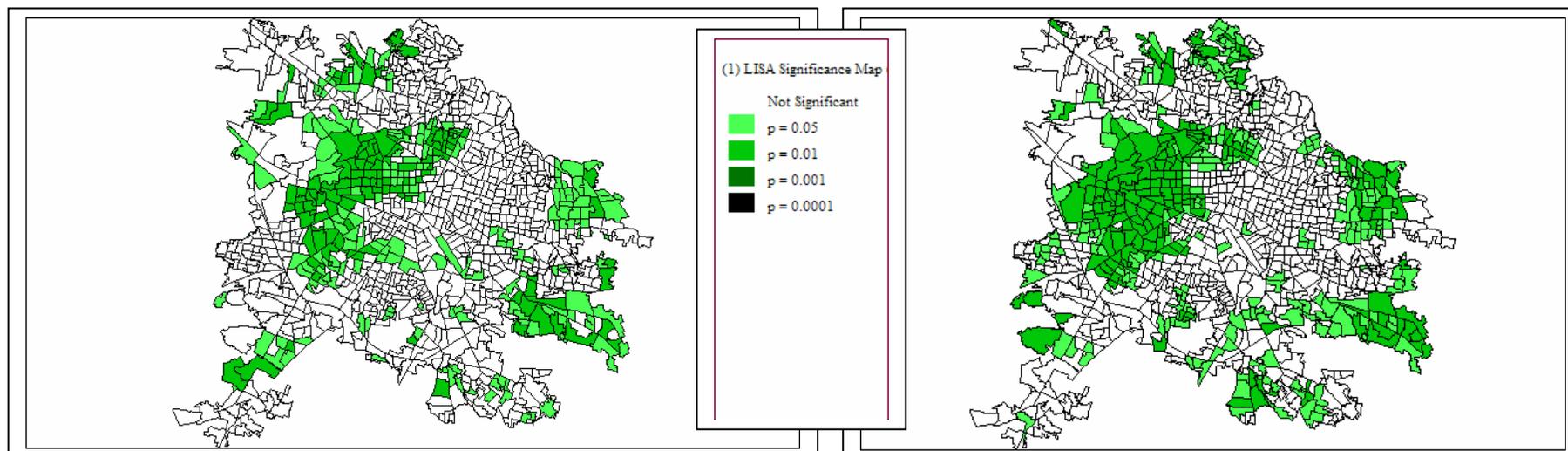


Figura 7. GWR Coeficientes Locales.
Relación de Dependencia

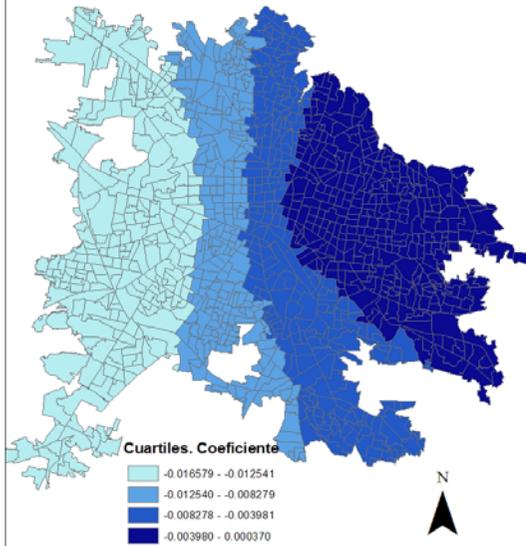


Figura 7. GWR Coeficientes Locales.
Trabajo por cuenta propia

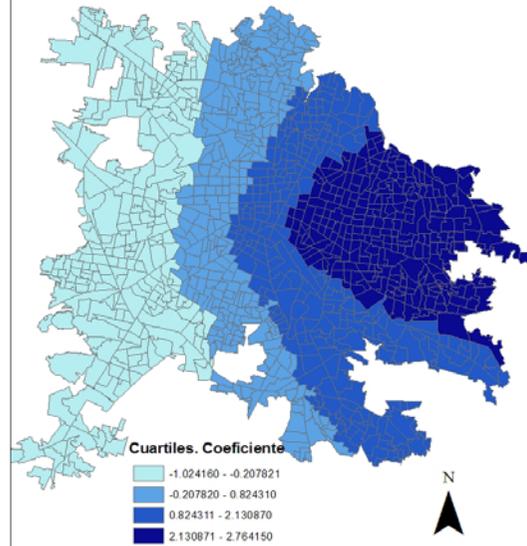


Figura 9. GWR R Cuadrado Local

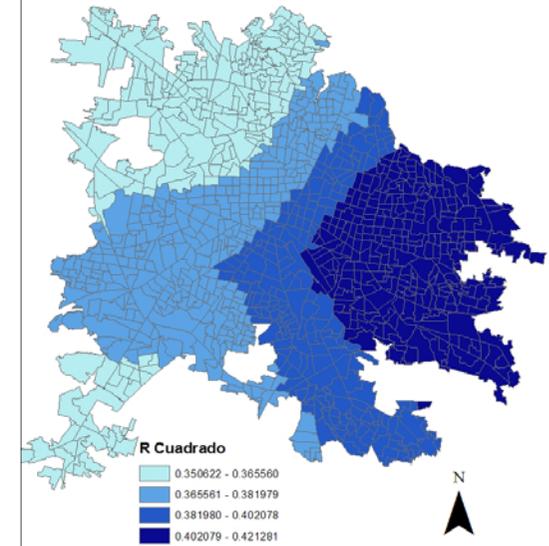


Tabla 1. Resultados de Mínimos Cuadrados y Espaciales

| Nivel de Pobreza Log p/(1-p) | Regresión de mínimos cuadrados | Modelo Espacial del Error | Modelo Espacia Lag |
|---|--------------------------------------|------------------------------|-----------------------|
| Constante | -0.622 *** | -0.609 *** | -0.0605 |
| Trabajo por cuenta propia | 0.574 | 0.562 | 0.289 |
| Trabajo por pocas horas | 2.312 *** | 2.408 *** | 2.393 *** |
| Trabajo en el sector informal | -1.052 *** | -1.162 *** | -1.135 *** |
| Desempleo | 20.984 *** | 21.559 *** | 20.211 *** |
| Grado de Escolaridad Promedio | - 0.270 *** | -0.267 *** | -0.2501 *** |
| Hogares de jefatura femenina | 1.432 * | 1.419 ** | 1.432 ** |
| Hogares Extendidos | 0.811 | 1.012 * | 0.698 |
| Relación de dependencia | - 0.009 *** | -0.009 *** | -0.010 *** |
| Población Femenina Ocupada | -0.516 *** | -0.458 ** | -0.492 ** |
| Termino Espacial Lamda | - | 0.2308 *** | - |
| Termino Espacial Lag | - | - | 0.1881 *** |
| Bondad del Ajuste | | | |
| R ² - Ajustado | 0.30067 | 0.3291 | 0.3238 |
| Log-Likelihood | - 1579.13 | -1565.73 | -1568.26 |
| AIC | 3178.25 | 3151.46 | 3158.52 |
| SC | 3228.69 | 3201.90 | 3214.01 |
| <i>Diagnósticos de Dependencia Espacial</i> | | | |
| - Moran's I (error) | 5.4204 *** | - | - |
| - LM Error | 27.4066 *** | - | - |
| - Robust LM Error | 4.9242 * | - | - |
| - LM Lag | 22.4828 *** | - | - |
| - Robust LM Lag | 0.00014 | - | - |
| -Likelihood Ratio Test | - | significativo | significativo |

Tabla 2. Resultados de GWR

| Label | Mínimo | Primer Cuartil | Mediana | Tercer Cuartil | Máximo |
|--------------------------------|-----------|----------------|----------|----------------|----------|
| Constante | -1.209719 | -1.00797 | -0.77728 | -0.503 | -0.24701 |
| Relación de dependencia | -0.01658 | -0.01239 | -0.00769 | -0.00315 | 0.00037 |
| Escolaridad Promedio | -0.303658 | -0.26845 | -0.25732 | -0.24887 | -0.23423 |
| Desempleo | 11.747485 | 15.04923 | 18.03723 | 21.4718 | 26.44027 |
| Trabajo por cuenta propia | -1.024161 | -0.21282 | 0.809797 | 2.125121 | 2.764146 |
| Trabajo por pocas horas | -0.085434 | 0.209669 | 0.891281 | 2.470724 | 3.975698 |
| Hogares de jefatura femenina | 0.173921 | 0.715762 | 1.169987 | 2.071613 | 2.973819 |
| Hogares extendidos | -0.227132 | 0.856677 | 1.477303 | 1.874226 | 2.25706 |
| Ocupaciones altamente informal | -1.829128 | -1.45605 | -1.10439 | -0.87724 | -0.56977 |
| Población Femenina Ocupada | -0.735187 | -0.68406 | -0.55815 | -0.34441 | -0.07782 |

| Apéndice A. Definición de variables | |
|--|---|
| Población Femenina Ocupada | Proporción de la población femenina ocupada. Población femenina de 12 años o más que trabajó en la semana de referencia, incluyendo la aquellas ocupadas pero que no trabajaron la semana de referencia. |
| Escolaridad Promedio | Años de escolaridad promedio. Resultado de dividir los años de educación aprobados por la población de 15 años o más entre el total de la población de ese rango de edad. |
| Hogares de Jefatura femenina | Proporción de hogares con jefatura femenina |
| Hogares extendidos | Proporción de hogares extendidos, definidos como aquellos que son incluyen un hogar nuclear y sus parientes o un jefe y sus parientes. |
| Relación de Dependencia | Número de individuos menores de 15 y mayores de 64 por cada 100 habitantes entre las edades de 15 y 64 años. |
| Desempleo | Proporción de trabajadores desempleados. Población que no trabajo pero buscó trabajo en la semana de referencia. |
| Trabajo por cuenta propia | Proporción de trabajadores por cuenta propia, definidos como aquellos que tienen su propio negocio sin contratar trabajadores. |
| Trabajo por pocas horas | Proporción de trabajadores que trabajaron 32 horas o menos en la semana de referencia |
| Ocupaciones altamente informal | Proporción de trabajadores ocupados en el sector servicios en ocupaciones con altos niveles de informalidad. Población ocupada de 12 años y más que labora en actividades del sector terciario (códigos SCIAN del 430 al 999) con excepción de las siguientes actividades: 461 al 469, 564, 710 al 713, 721, 810, 811 y 816 Variable construida por el INEGI para el proyecto regiones socioeconómicas de México. |