



---

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS  
CAMPUS MONTERREY

”IMPACTO DE LA CERCANÍA A LA FRONTERA CON  
EE. UU. EN LA CREACIÓN Y ESPECIALIZACIÓN DE  
TRABAJO”

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

**Maestro en Cómputo Estadístico**

PRESENTA:

**Jorge Luis Ramos Zavaleta**

DIRECTOR:

Dr. Rafael Garduño Rivera

CODIRECTOR:

Dr. Rogelio Ramos Quiroga

Monterrey, Nuevo León, 2019



*A mi familia quienes me han apoyado durante todo el trayecto de esta aventura.  
Es gracias a ustedes que es posible el presente trabajo.  
En verdad, gracias.  
Yo.*



# Reconocimientos

---

También quisiera reconocer a ... por ...CONACYT, PAPIIT / etc. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.



# Resumen

---

Este trabajo esta basado en el de Blankespoor et al. (2017) haciendo uso de los datos utilizados en dicho trabajo, se busca establecer los impactos de la construcción y mejora de carreteras tanto en el empleo como en la especialización del trabajo. Para ello se considero la distancia de las cabeceras municipales a la frontera con EE. UU. más cercana haciendo uso de las carreteras disponibles durante el periodo 2004-2014.

Para realizar la modelación de dichos impactos los métodos elegidos son los llamados modelos de panel de datos espaciales, y debido a todas las distintas subespecificaciones que se desprenden de este tipo de modelos, en este trabajo se hara énfasis en las pruebas estadísticas para obtener el modelado correcto para este tipo de datos y con ello tener una interpretación mucho mas precisa.

Algo que cabe destacar es que la estimación de dichos modelos presenta variantes entre ellos, por lo que para mantener los resultados consistentes entre las distintas subespecificaciones solo se consideraron las estimaciones basada en Máxima Verosimilitud.

Para lograr establecer de mejor manera el impacto se hace uso de variables de control apropiadas como población con educación superior en el municipio, municipios petroleros y municipios que funcionan como capital estas ultimas como variables dummy. Con esto se encontraron posibles efectos causales positivos respecto del aumento de la distancia a la frontera con la especialización del trabajo pero negativos con respecto al empleo en general.





# Índice general

---

Índice de figuras	IX
Índice de tablas	XI
<b>1. Introducción</b>	<b>1</b>
<b>2. Trabajo en México</b>	<b>3</b>
2.1. Empleo en México . . . . .	3
2.2. Especialización del trabajo . . . . .	4
2.2.1. Definición . . . . .	4
2.2.2. Ventajas . . . . .	5
2.2.3. Desventajas . . . . .	7
2.3. Especialización del trabajo en México . . . . .	8
2.4. Distancia a la frontera con EE.UU. . . . .	9
2.5. Justificación, objetivos y alcance del proyecto . . . . .	9
2.5.1. Justificación . . . . .	10
2.5.2. Objetivos . . . . .	10
2.5.3. Alcance . . . . .	11
<b>3. Análisis descriptivo</b>	<b>13</b>
3.1. Infraestructura carretera en México . . . . .	13
3.2. Sectores económicos en México . . . . .	15
3.2.1. Comercio . . . . .	15
3.2.2. Manufactura . . . . .	16
3.2.3. Servicios . . . . .	16
3.2.4. Minería . . . . .	16
3.3. Índice de especialización de Krugman (KSI) . . . . .	16
3.4. Distancia a la frontera norte . . . . .	16
3.5. Variables de control . . . . .	16
<b>4. Regresión lineal espacial</b>	<b>17</b>
4.1. Correlación espacial . . . . .	17
4.2. Matriz de distancias . . . . .	18

## ÍNDICE GENERAL

---

4.3. Prueba de Moran . . . . .	18
4.4. Modelos de regresión lineal espacial . . . . .	20
4.5. Modelos espaciales de panel de datos . . . . .	21
4.5.1. Modelos de panel de datos espaciales con efectos aleatorios . . .	23
4.5.2. Modelos de panel de datos espaciales con efectos fijos . . . . .	24
4.5.3. Estimación de los parámetros . . . . .	25
<b>5. Análisis de resultados</b>	<b>29</b>
5.1. Modelo Pooling . . . . .	30
5.2. Pruebas LM . . . . .	30
5.3. Prueba de Hausman . . . . .	30
5.4. Efectos aleatorios . . . . .	30
5.4.1. Rezago espacial . . . . .	30
5.4.2. KKP . . . . .	30
5.5. Efectos fijos . . . . .	30
5.5.1. Rezago espacial . . . . .	30
5.5.2. Error espacial . . . . .	30
5.6. Interpretación de los modelos finales . . . . .	30
<b>6. Conclusiones</b>	<b>31</b>
<b>Referencias</b>	<b>33</b>

# Índice de figuras

---

3.1. Estructura de carreteras usada. Fuente: Blankespoor et al. (2017).	14
4.1. Ejemplos de matrices de distancias. El inciso (a) corresponde la región a la cual se le quiere obtener la matriz de distancias, los incisos (b), (c) y (d) corresponden a matrices de distancias correspondientes a la región en el inciso (a) bajo distintos criterios: adyacencia, vecino más cercano y distancia < 2 respectivamente. Fuente: Arbia (2014).	19



# Índice de tablas

---



---

## Capítulo 1

# Introducción

---

Debido a la relación comercial México-EE. UU. que se suscitó a partir de 1994 con la firma del TLCAN, se han realizado diversos estudios para conocer el impacto POS-TLCAN en la economía mexicana. En este sentido, y con base en el trabajo de Blankespoor et al. (2017), se explora la posible relación entre la distancia entre municipios a la frontera con EE. UU. con respecto al empleo y la especialización del trabajo.

Dado que se tiene información a nivel municipal y temporal el enfoque planteado es usar modelos de panel de datos espaciales, esto con el fin de explotar lo mejor posible la riqueza de dicha estructura. Uno de los retos que aparece inmediatamente en este marco de trabajo es la existencia de varios tipos de subespecificaciones para esta clase de modelos.

Con el fin de reducir el problema de la elección de la subespecificación correcta para el modelo se hizo uso de pruebas estadísticas que nos permiten conocer el tipo de efectos del modelo y si es significativa la autocorrelación espacial para continuar con esta clase de modelos. Otro problema a considerar es que estos modelos no tienen una medida de bondad de ajuste directa como en el caso de la regresión lineal por lo que se hizo uso de una alternativa (pseudo  $R^2$ ).

Para mantener la relación temporal entre los municipios se consideró la estructura de los municipios tal y como se presenta en Blankespoor et al. (2017), por lo que se trabajó con 2377 municipios unificando los datos de los nuevos municipios con los del municipio líder histórico.

Tal como se menciona anteriormente, se hizo uso de pruebas de hipótesis estadísticas con el propósito de tener un criterio que nos permita discernir en la elección de las subespecificaciones para cada modelo. Para este trabajo se hizo uso de dos tipos de pruebas: pruebas de Multiplicadores de Lagrange (LM) y una prueba de Hausman espacial.

## 1. INTRODUCCIÓN

---

Las pruebas LM utilizadas permiten establecer bajo un criterio de distribución estadística tanto si el modelo a utilizar presenta una estructura de efectos fijos o de efectos aleatorios dependiendo de la forma del estadístico utilizado en la implementación de la prueba. Una ventaja de este tipo de pruebas es que no requieren estimar el modelo, sino que solo requiere utilizar los residuales de un modelo de regresión clásico por lo que son muy útiles para cuando es requerido calcular varios modelos con el fin de establecer la elección de variables.

Por otro lado, la prueba de Hausman es aunque es una prueba estándar para la modelación de datos de tipo panel requiere calcular directamente el modelo, por lo que aunque en general es una prueba más robusta para conocer el tipo de especificación del modelo también es computacionalmente mas pesada.

Aplicando ambos tipos de pruebas estadísticas a los modelos finales se concluyó que el modelo tiene una especificación de efectos fijos, es decir que ninguna de las covariables estaba correlacionada con los residuales, y también se encontró que existía evidencia muy fuerte para considerar que el modelo debía presentar autocorrelación espacial de algún tipo.

Por último, los modelos finales obtenidos resultaron tener la subespecificación de un modelo de rezago espacial en ambos casos. En ambos casos se obtuvo un componente espacial muy alto y bastante significativo lo cual implica que la presencia del empleo (especialización) en un municipio tiende a influir en los municipios cercanos.

En el caso del modelo para la especialización del trabajo se encontró una relación positiva con respecto a la distancia con la frontera, lo cual puede explicar la existencia de varios municipios en el sur del país que se dedican a una actividad en específico. Mientras que para el empleo se encontró una relación negativa como era de esperarse debido a la presencia de las empresas de manufactura en el norte del país.



---

## Capítulo 2

# Trabajo en México

---

En el presente capítulo se presentará el concepto de trabajo y de especialización del trabajo. Para el caso de la especialización del trabajo se indicarán sus ventajas y desventajas, y un breve bosquejo histórico de cómo ha cambiado la aglomeración de actividades en el contexto mexicano. Dentro del contexto de este capítulo también se presentarán los objetivos, antecedentes, justificación y el alcance de este trabajo.

### 2.1. Empleo en México

De acuerdo a la Organización Internacional del Trabajo una *persona empleada* es aquella que tiene la edad legal para llevar a cabo de manera formal en alguna de las actividades de la economía que se dedique a la producción de bienes y servicios por una remuneración **OIT (2019)**.

Existen dos tipos de empleo: formal e informal. El empleo formal incluye a los trabajadores que tienen una relación laboral reconocida y que hacen cumplir sus derechos laborales (tales como seguridad social, beneficios no salariales de liquidación o finiquito al término de la relación de trabajo). A la inversa, el empleo informal engloba a los trabajadores que, aunque reciben un pago por su trabajo, no tienen una relación laboral reconocida y no pueden hacer cumplir sus derechos laborales.

De acuerdo a la OCDE, el empleo formal es el que más impacto tiene en el crecimiento económico y en la reducción de la pobreza **OECD (2009)**, de aquí la importancia de promover su crecimiento y desarrollo plenos.

En México, el empleo formal se mide a través del número de trabajadores registrados en el Instituto Mexicano del Seguro Social (IMSS), que es la institución que brinda los servicios de salud y seguridad social a los trabajadores registrados (a los trabajadores formales).

## 2. TRABAJO EN MÉXICO

---

Una perspectiva sobre el empleo en México así como algunos artículos de interés respecto al tema pueden encontrarse en [Enríquez y Galindo \(2015\)](#).

### 2.2. Especialización del trabajo

El concepto de especialización es uno de los importante en la teoría económica, tanto así que hace hincapié en su rol en el comercio desde los primeros cursos. Por ejemplo en [Mankiw \(2014\)](#) que es uno de los libros mas usados para impartir los cursos iniciales de economía se presenta como uno de los principales motores para obtener ganancias del comercio desde sus primeras paginas.

#### 2.2.1. Definición

Existen diferentes definiciones del concepto de especialización del trabajo, dentro de la teoría económica podemos encontrar definiciones desde un punto de vista más enfocado en lo laboral como en [Murali \(2017\)](#) que define la especialización del trabajo de la siguiente manera

*Job specialization is defined as the impact of mismatch on match productivity, where mismatch is the distance between the skills/ability of a worker and the skill requirements of their job. If there is zero specialization, then any worker of any skill level is suitable for any job, so, mismatch has zero effect on the match productivity. On the other hand, if a job is highly specialized, even a small amount of mismatch can have a large negative impact on match productivity.*

Por otra parte, y para fines de este trabajo requerimos una definición de especialización del trabajo mucho más enfocada en el contexto regional. En [Kemeny y Storper \(2015\)](#) encontramos una definición de *región diversificada*

*Una región diversificada es definida como aquella que contiene un amplio arreglo de sectores no relacionados en su base económica, sin un sector específico dominante.*

Este concepto de región diversificada nos sirve de partida, debido a que el concepto de especialización del trabajo que buscamos puede considerarse el inverso de este concepto. Por lo que podríamos definir el concepto de especialización del trabajo a nivel regional de la manera siguiente

*Es un área que contiene en proporciones una mayor parte de sectores económicos, que se relacionan en su base económica y que funcionan para permitir el crecimiento y desarrollo de un sector específico dominante.*

Con este concepto establecido podemos comenzar enfatizar en las ventajas y desventajas que se han presentado alrededor de dicho concepto.

### 2.2.2. Ventajas

Existen varias causas por las que la especialización del trabajo en forma de aglomeración económica se da, en [Duranton y Kerr \(2015\)](#) y [Ellison et al. \(2010\)](#) se establecen algunas de ellas. Dichas causas pueden considerarse como posibles ventajas de establecer dicha especialización. Sin embargo, no necesariamente dichas causas terminan representando una ventaja.

En términos prácticos y con el fin de resumir dichas ventajas, en [Ellison et al. \(2010\)](#) se establecen seis puntos que se pueden distinguir como ventajas de establecer dichas aglomeraciones

- **Retornos crecientes bajo contexto de globalización.** Debido al proceso de aglomeración las empresas dentro de los clusters terminan generando ciertos factores de ventaja competitiva relacionados al concepto de proximidad: rivalidad local entre empresas, presencia de industrias especializadas, presencia de instituciones de apoyo a la producción, entre otras.

Como consecuencia de la presencia de estos factores se presentan retornos crecientes derivados de las propiedades inmersas dentro del sistema local sin la necesidad de generar incremento en el tamaño de las plantas o del volumen de la producción.

- **Reducción de costos de transacción.** La teoría económica que establece alrededor del concepto de los costos de transacción considera que las empresas nacen como una forma de reducir los costos de transacción internalizándolos [Coase \(1937\)](#).

De acuerdo a [Williamson \(2007\)](#) se pueden considerar tres distintos tipos de falla que dan nacimiento a estos costos de transacción:

1. **Racionalidad acotada en la toma de decisiones.** Se refiere a la limitación de la capacidad humana de adquirir y procesar información. Por ejemplo, al querer comprar una computadora existen muchas variedades entre modelos y las piezas específicas que las conforman así como la calidad de las piezas que integran dicha computadora, por lo que regularmente la decisión de compra se hace con base en un número limitado de opciones.
2. **Oportunismo.** Este concepto resulta del egocentrismo o auto interés de las personas. Por ejemplo, al vender un auto usado el vendedor conoce bien la maquina del automóvil incluyendo las reparaciones que se le han hecho por lo que es capaz de establecer un precio final mayor del real convenciendo al

comprador que las fallas que ha tenido el auto fueron muy leves o que no ha presentado fallas.

3. **Incertidumbre del mercado.** Es el resultado de dificultades no planeadas en cada transacción. Por ejemplo, al comprar un aparato electrónico usado puede ser que el aparato deje de funcionar mientras se le muestra al comprador, aunque el aparato no hubiese tenido ninguna falla hasta el momento.

La presencia de alguna de estas fallas propician que el precio de los productos no dé información real sobre la calidad del producto por lo que tanto comprador como vendedor terminan intercambiando a un precio distinto del real, estableciendo costos extras no considerados para el vendedor o una pérdida en la capacidad adquisitiva del comprador.

Al generarse aglomeración con respecto a un sector se generan redes de información que permiten tanto a vendedores como a compradores tener mejor idea del precio real de los productos, lo que implicaría una reducción de los costos de transacción.

- **Diversificación de producto y centramiento en el cliente.** En la actualidad existen muchos mercados que no se rigen directamente por el precio por lo que se presentan otros criterios como calidad del producto y el valor agregado que se le impone al producto lo que hace que las empresas tengan que centrar su producción en los requerimientos del usuario.

Al pertenecer a una aglomeración especializada en un producto o servicio las empresas generan redes de interacción con las que se puede comprobar el dinamismo de la industria y tendencias en los usuarios, así como acumulación de experiencia conjunta, en conjunto esto conlleva a que las empresas establezcan una competencia local propiciando la búsqueda de nuevos segmentos de mercado a través de una continua diversificación de sus productos enfocándose en las exigencias de los clientes.

- **Pionero.** En el caso de que una empresa tenga un liderazgo tecnológico es posible para ella mantener dicho liderazgo debido a que las redes de información que se generan debido a la aglomeración le permiten tener acceso mas rápido a las nuevas tendencias.
- **Reducción de costos vía aprendizaje efectivo.** Debido a la concentración geográfica de empresas del mismo tipo se llegan a establecer sistemas organizacionales conjuntos, dichos sistemas permiten que los empleados de una empresa puedan emular e imitar las habilidades de los trabajadores de otra empresa y con ello reducir tanto el tiempo como los recursos en aprendizaje para los trabajadores.

Evidencia de esta afirmación se puede encontrar en [Asheim \(1999\)](#).

- **Innovación y desarrollo tecnológico debido a interacciones locales.** Debido al proceso de aglomeración tanto trabajadores como empresarios de las distintas empresas se encuentran en contacto regular lo que permite que se generen interacciones entre ellos y generándose un efecto de derrame de conocimiento sobre los procesos que se están manejando adentro de cada empresa permitiendo que se tenga acceso a nuevos procesos que en alguna de las empresas y que se pueden implementar dentro de las demás empresas del cluster.

Estas nuevas implementaciones permiten a las empresas generar innovación en sus productos y servicios, así como también permiten renovar mas frecuentemente los procesos incentivando un mayor desarrollo tecnológico.

### 2.2.3. Desventajas

En [Kemeny y Storper \(2015\)](#) se indica que los argumentos a favor de las regiones diversificadas se pueden resumir en tres justificaciones

- **Diversificación del riesgo.** En este punto se hace la referencia a las fluctuaciones en el ciclo de los negocios y como éstos pueden tener consecuencias extremas sobre los sectores en los que se especializa la región. Por ejemplo, en [Hill \(2006\)](#) se presenta el caso del precio del café en Uganda. En este estudio se indica que debido a las fluctuaciones del precio, los productores locales son los que se ven mayormente afectados.

Como consecuencia de estas afectaciones se pueden llegar a nulificar las posibilidades de sustento de los productores y debido a que es la única actividad en el área terminan por no tener otra opción que mudarse a otra área económica. Aunque puede darse el caso de que estas regiones deban su especialización a que tienen un costo de oportunidad mucho más bajo que la competencia por lo que el ciclo de los negocios tendrá un efecto negativo mucho menor.

- **Eficiencia en el suministro de insumos generales.** De acuerdo con esta justificación economías urbanizadas proveen de insumos generales en escalas eficientes que son útiles para muchas actividades en una región. Es decir, debido a que una región puede ser suficientemente grande entonces ésta puede proveer de distintos insumos a la economía en general por lo que puede soportar una diversificación de sus actividades y mantener su productividad. Sin embargo, el hecho de que una región sea grande y pueda proveer dicha diversificación puede deberse en realidad a un proceso inicial de especialización económica.
- **Dinámica de la economía regional.** En este sentido se piensa a la economía moderna como un ente muy amplio y complejo, y como tal para permitir la adopción de nuevas actividades económicas de manera más eficiente requiere que exista una diversificación en las habilidades de los trabajadores y en los insumos

que genera. De manera que una economía demasiada especializada puede ser muy ineficiente para la adopción de nuevas actividades económicas que requieran habilidades fuera de su sector de especialización.

### 2.3. Especialización del trabajo en México

En México desde antes de la firma del TLCAN ya se tenía conciencia de la presencia de algunos clusters pero la mayor parte de las actividades se realizaban alrededor de la Ciudad de México. De acuerdo con [Hanson \(1998\)](#) existía una concentración manufacturera alrededor de la Ciudad de México pero debido los intentos de Estados Unidos por impulsar el libre comercio entre ambos países se comenzó a impulsar una expansión de la industria manufacturera en la frontera norte atrayendo mano de obra bien calificada a través del pago de un salario Premium.

Dos de los clusters mas conocidos que han sido bien identificados y que cuentan con diferentes niveles de concentración de manera regional: Calzado y Automotriz, son estudiados en [Unger \(2003\)](#) donde se establece nivel de concentración que se representa regionalmente así como su variación en el tiempo.

Por otra parte, en [Dávila Flores \(2008\)](#) se identifican los distintos clusters industriales en el noreste de México y como se fue presentando su variaciones durante el período 1993-2003 con el fin de observar cambios debido a la firma del TLCAN. En dicho trabajo se encontró que los estados que conforman la frontera con Texas forman la base del comercio exterior debido a la acumulación de industria manufacturera que obtuvieron durante ese periodo.

De manera similar, en [Baylis et al. \(2012\)](#) se hace un estudio a nivel municipal para identificar cambios en la actividad económica como efecto del TLCAN y se encontró que el efecto de aglomeración disminuyó debido a la presencia de dicho tratado. Mientras que en [Alvarez et al. \(2017\)](#) se establece un esquema de frontera de producción para identificar la eficiencia técnica a nivel estado en un período anterior y posterior a la firma del TLCAN encontrando que los estados cercanos a la frontera con Texas son los que mayor eficiencia presentan.

La firma del TLCAN estableció nuevas reglas dentro de la forma en que estaba organizada la economía mexicana especialmente a lo que se refiere a la especialización del trabajo. Mientras que las actividades se comenzaron a aglomerar en algunos estados en otros no se presentaron cambios sustanciales en este sentido.

Debido a las ventajas expuestas anteriormente que representa la especialización del trabajo, es importante encontrar que causas son las que pueden propiciar dicha variable. En el tercer capitulo se mostrara visualmente el cambio en la especialización del

trabajo en México a lo largo del periodo establecido para este trabajo.

## **2.4. Distancia a la frontera con EE.UU.**

Debido a la heterogeneidad que presenta la infraestructura carretera mexicana tanto espacial como temporalmente se decidió usar una medida de accesibilidad que incorpore en algún sentido la información de la infraestructura carretera en México.

Por lo que se decidió por hacer uso de la distancia hacia los municipios que son frontera con los Estados Unidos, debido a la importancia comercial que tiene este país con México. Para ello se consideraron las distancias entre municipios a través de la infraestructura carretera disponible hasta el momento y usando la ruta mas corta a alguno de los municipios que se encuentra en la frontera solo considerando las distancias. A los municipios que se encuentran en la frontera se les asigno una distancia cero.

De acuerdo con los datos usados se tienen 5 tipos de caminos distintos entre municipios y aunque las velocidades en cada tipo de camino pueden variar bastante estas no se consideraron para tener un enfoque basado en la distancia con el fin de mantener la estructura espacial lo mayormente posible. Los tipos de caminos utilizados son:

- Multicarril dividido.
- Dos carriles o dividido.
- Pavimento.
- Grava o camino de tierra.
- Categoría desconocida.

## **2.5. Justificación, objetivos y alcance del proyecto**

Como cualquier proyecto de investigación es importante reconocer si este genera una aportación suficiente para resolver un problema, y a su vez debe reconocerse que no siempre es posible abarcar por completo el problema por lo que se vuelve importante acotarlo.

Siendo así en los siguientes puntos se establecen la relevancia del trabajo, los objetivos a perseguir y el alcance del trabajo con el fin de no generar expectativas mayores a las que se esperan resolver.

### 2.5.1. Justificación

La red carretera de un país tiene una importancia muy amplia, particularmente su impacto en la economía se puede ver directamente reflejado en una reducción de los costos de transporte lo cual permite a empresas alcanzar nuevos mercados con un costo menor. Una colección amplia de trabajos sobre este tema se puede encontrar en [Redding y Turner \(2015\)](#).

Si bien este hecho de la reducción de costos es bien conocido, la construcción de infraestructura carretera también permite obtener otros clase de beneficios como pueden ser una mayor eficiencia en el transito de personas, lo cual conlleva a mejorar la movilidad de mano de obra mejor calificada, así como también permite que aumente el empleo y los salarios de empleados en el sector formal. Un modelo teorico explicando estos hechos puede encontrarse en [Duranton et al. \(2014\)](#).

Como ya se menciona anteriormente, la especialización del trabajo en su forma de aglomeración puede conllevar a muchos beneficios tanto para las empresas que integran el cluster como a los clientes que consumen sus productos o servicios. Debido a esto se vuelve importante determinar factores que puedan favorecer la creación de dichas aglomeraciones.

De esta manera, en este trabajo se busca verificar la existencia de una relación entre la construcción y mejora de infraestructura carretera con respecto a la especialización del trabajo y de la cantidad de trabajadores empleados a través de la distancia que se debe recorrer hasta la frontera con EE. UU., y en caso de existir dicha relación cuantificarla.

Con dicha cuantificación se pueden generar conclusiones que posteriormente pueden ser usadas una herramienta para la creación de políticas públicas tanto en el marco de infraestructura carretera como en el del crecimiento regional.

### 2.5.2. Objetivos

Es importante tener objetivos muy claros debido a que el problema de la medición del impacto puede hacerse de muchas formas por lo que este problema puede volverse intratable para la ventana de tiempo que se tiene pensada para terminarlo. Por lo que se establecieron objetivos general y especifico que fueran factibles de trabajar en dicha ventana de tiempo.

#### Objetivo general

El objetivo general del trabajo es con base en el trabajo de [Blankespoor et al. \(2017\)](#) es determinar si existe una relación entre la distancia hacia la frontera establecida a través de las carreteras con respecto tanto del número de trabajadores empleados como



de la aglomeración económica del trabajo.

Para ello se realizara un modelado estadístico explotando el potencial de la información haciendo uso de modelos de panel de datos espaciales y realizando las apropiadas pruebas estadísticas para la elección del modelado correcto.

### Objetivos específicos

Como parte de la modelación se utilizaran algunas variables de control como edad o tamaño de la población para permitir controlar de mejor manera la relación que existe entre especialización del trabajo y el potencial de mercado. Por ello uno de los objetivos específicos es plantear la pertinencia de posibles relaciones causales de dichas variables que permiten esclarecer de mejor manera el impacto que se quiere medir.

Debido a que las especificaciones del modelo contienen términos de error que pueden ser autorregresivos espacialmente el segundo objetivo específico es establecer una interpretación correcta de dicho resultado.

### 2.5.3. Alcance

De manera similar al trabajo de Blankespoor et al. (2017) los datos con los que se van a trabajar son de municipios de México en el periodo 2004-2014, y debido a los cambios que se presentan en dicho periodo sobre la creación o destrucción de municipios se tuvo que mantener alguna relación históricas para permitir que los resultados no se vean afectados por dichos cambios.

Para mantener la interpretabilidad de los modelos y permitir su comparación solo se va a considerar su estimación por el método de Máxima Verosimilitud, aun cuando alguna especificación pueda tener otro método de estimación puesto que las condiciones y supuestos que se deben verificar en esos casos serian distintas.



---

## Capítulo 3

# Análisis descriptivo

---

A lo largo de este capítulo se realizara un análisis descriptivo de los datos a utilizar y se establecerá una justificación del porque estos datos se consideran relevantes para el modelado que se plantea realizar.

De igual manera que en [Blankespoor et al. \(2017\)](#) con el fin de mantener la integridad de la división política de los municipios debido a la creación de nuevos municipios se considera la estructura histórica que se tenía en sus fronteras en 1988. Con lo que se tienen 2,377 municipios y las regresiones se hacen sobre dicha división, razón por la que los mapas que se muestran en este capítulo presentan algunos huecos, pues los mapas son generados a partir de la división que provee la INEGI en 2015.

### 3.1. Infraestructura carretera en México

Los caminos en México tiene una historia muy larga, pero su relevancia mas grande dió comienzo durante la época porfirista donde la contrucción de caminos y redes ferroviarias fueron los estandartes de su política de progreso económico. Sin embargo, al presentarse la Revolución Mexicana las obras se detuvieron e inclusive una gran parte de la infraestructura construída se daño, y con la presencia de los vehículos de automotor se inicio un proceso de demanda de caminos que permitieran el uso apropiado de dichos vehículos. Por lo que en 1925 se constituye la Comisión Nacional de Caminos.

La construcción de caminos durante las décadas de 1920 y 1930 fue un preámbulo del papel del estado sobre la construcción de las carreteras debido a que anteriormente dicha infraestructura era llevada a cabo principalmente por capital extranjero y particulares. A partir de las relaciones amistosas que se establecieron con EE. UU. debido a los procesos económicos que conllevo la Segunda Guerra Mundial se comenzaron a inyectar inversiones extranjeras que permitieron construir una mayor infraestructura, y también se impulso dicha medida como parte de la política de sustitución de importaciones por parte del presidente Miguel Alemán. Detalles mas específicos de los procesos

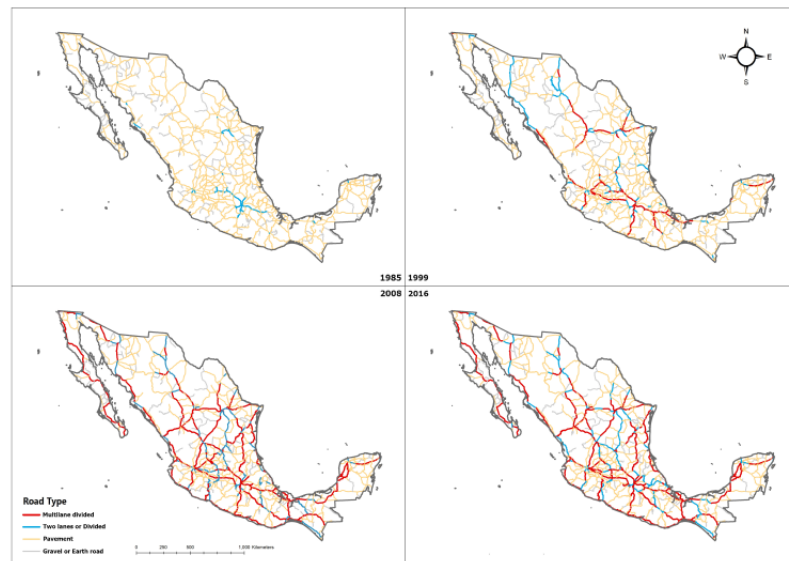
### 3. ANÁLISIS DESCRIPTIVO

---

que se llevaron a cabo desde 1920 pueden consultarse en [Gómez \(1990\)](#), [Bess \(2016\)](#) y en [Bess \(2014\)](#).

En 2014, como parte del Plan Nacional de Desarrollo 2013-2018 el gobierno mexicano publico el Programa Nacional de Infraestructura 2014-2018 ([DOF \(2014\)](#)), donde se detallan diversos proyectos en cuestión de infraestructura carretera principalmente en el sur y centro del país como una forma de mejorar las condiciones económicas de la región. También debido a este hecho INEGI en ese mismo año libera información digitalizada de la infraestructura carretera para apoyar la generación de políticas en este rubro.

Debido a que el periodo de estudio es anterior al 2014, entonces se requirió considerar una base de carreteras distinta a la que trabaja INEGI, por lo que se usaron los mismos datos utilizados en [Blankespoor et al. \(2017\)](#) que provienen de una base de datos utilizada en DeLorme(2014) y actualizada con datos de la American Automobile Association (AAA). En la siguiente figura se muestran los caminos usados con su respectiva clasificación.



**Figura 3.1:** Estructura de carreteras usada. Fuente: [Blankespoor et al. \(2017\)](#).

Puede observarse un incremento en las carreteras con límites de velocidad mayores, lo que implica una mejora en los tiempos de traslado y por supuesto también en las distancias a recorrer.

## 3.2. Sectores económicos en México

En México la actividad económica se divide en tres grandes sectores

- Sector primario. Este sector abarca de todas las actividades productivas que involucran la extracción y obtención de materias primas.
- Sector secundario. Este sector comprende las actividades productivas que se dedican al procesamiento y transformación de las materias en bienes y/o productos para el consumo.
- Sector terciario. Este sector engloba todas las actividades económicas relacionadas con los servicios, por lo que no produce bienes materiales.

Para este trabajo, en el caso del empleo se consideraron todos los sectores económicos en conjunto, pero para el caso de la especialización del trabajo aunque se consideraron los tres sectores económicos no se consideraron por completo o bien se dividió alguno en dos para eliminar algunos efectos que se pueden tener por cierto tipo de actividad económica.

La división de los sectores económicos usados para el cálculo del índice de especialización se da en las 4 siguientes subsecciones.

### 3.2.1. Comercio

El comercio es una actividad económica dentro del sector terciario. Particularmente para este trabajo se considero separarlo del resto de las demás actividades del sector terciario para identificar municipios que pueden estar especializadas en esta actividad económica, y con ello permitir establecer de mejor manera la heterogeneidad que se puede presentar en los datos a nivel municipal.

En la figura se presenta el número de empleados dedicados al comercio a nivel municipal para los años 2004, 2009 y 2014.

### 3. ANÁLISIS DESCRIPTIVO

---

**3.2.2. Manufactura**

**3.2.3. Servicios**

**3.2.4. Minería**

**3.3. Índice de especialización de Krugman (KSI)**

**3.4. Distancia a la frontera norte**

**3.5. Variables de control**

## Regresión lineal espacial

---

En este capítulo se presenta la técnica que dará fundamento al trabajo: Regresión espacial para datos de tipo panel. Pero antes de entrar en materia de dicha técnica se iniciará hablando de la importancia de la correlación espacial para los métodos de regresión y se explicará el funcionamiento de la regresión espacial para datos de sección cruzada lo cual nos permitirá usar esto como base para comprender mejor la versión para panel de datos.

### 4.1. Correlación espacial

Uno de los supuestos básicos de la regresión lineal es que los residuos obtenidos por el modelo usado deben ser estadísticamente independientes, es decir,  $E(\epsilon_i \epsilon_j) = E(\epsilon_i)E(\epsilon_j) = 0$  por lo que la modelación se simplifica al tener la forma

$$y_i = X_i \beta + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n$$

Pero en el caso de presentarse dependencia espacial en los datos el supuesto mencionado anteriormente se rompe y por tanto, de mantenerse el mismo sentido de la modelación que la usada para la regresión lineal simple se complicaría. Por ejemplo, considerando solo dos regiones que tienen una dependencia espacial tendríamos una modelación como la siguiente

$$y_i = \alpha_i y_j + X_i \beta + \epsilon_i$$

$$y_j = \alpha_j y_i + X_j \beta + \epsilon_j$$

$$\epsilon_i \sim N(0, \sigma^2), \quad i = 1$$

$$\epsilon_j \sim N(0, \sigma^2), \quad j = 2$$

Esto implica resolver simultáneamente ambos modelos, y en general si se tienen  $n$  posibles relaciones de dependencia espacial tendríamos  $n^2 - n$  posibles modelos que

deberíamos resolver simultáneamente. Esto debido a que restamos las relaciones de las observaciones consigo mismas debido a que no tienen ninguna relevancia espacial y pueden ser capturadas directamente en el modelo que las tienen específicamente como variables respuesta.

### 4.2. Matriz de distancias

Para enfrentar el problema anteriormente descrito se plantea una modelación autorregresiva. Para ello, se hace uso de una matriz de distancias que indica una relación de cercanía entre los posibles vecinos, donde bien pueden definirse distancias directas o hacer uso de algún tipo de métrica de similitud (disimilitud) que permite establecer la vecindad espacial entre las observaciones.

El caso más simple de este tipo de matrices es el mismo que el usado en la teoría de gráficas para indicar adyacencia, esto es, indicando un uno si el nodo es vecino de otro y un cero en caso contrario.

En la practica es usual que se estandaricen los pesos de dicha matriz de manera que la suma de los pesos en cada fila de la matriz sea 1. Esto último se hace generalmente para tener una mejor interpretación del proceso autoregresivo. Por ejemplo, si se define  $L(y) = W \times y$ , con  $W$  la matriz de pesos estandarizados entonces

$$L(y_i) = \sum_{j=1}^n w_{ij}^* y_j = \sum_{j=1}^n \frac{w_{ij} y_j}{\sum_{j=1}^n w_{ij}} = \frac{\sum_{j \in N(i)} y_j}{\#N(i)}$$

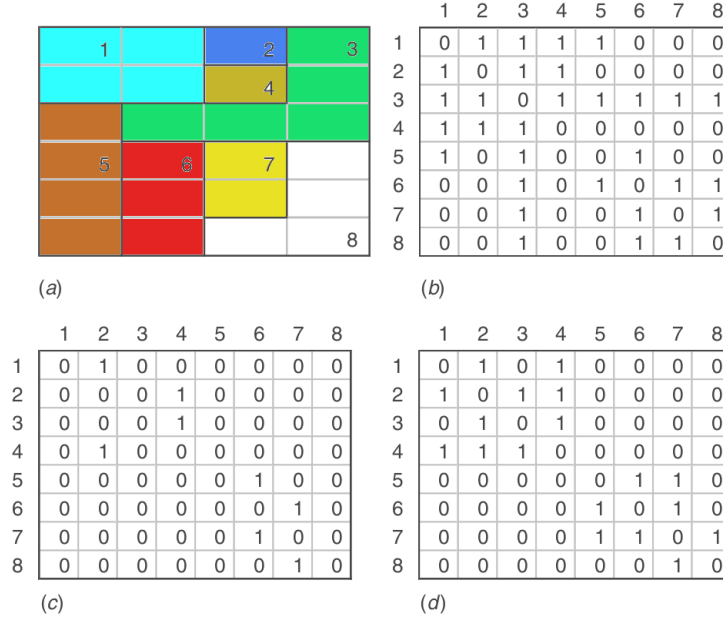
definido de esta forma  $L(y_i)$  representa el promedio de la variable  $y$  observado en todas las localizaciones que son vecinos a la localización  $i$ .

Un ejemplo simple de matrices de distancias se presenta en la figura 4.1, donde el inciso (a) representa un sistema de ocho regiones irregulares mientras que los demás incisos representan distintas matrices de distancias  $W$  calculadas a partir de dicho sistema bajo distintos criterios: (b) adyacencia, (c) vecino más cercano y (d) distancia  $< 2$ . Las distancias son medidas con base en los centroides de las regiones.

### 4.3. Prueba de Moran

Como ya se mencionó anteriormente, la dependencia espacial puede verse directamente en los residuos al realizar una regresión lineal. Para ello, usualmente se hace uso de la prueba de autocorrelación espacial de Moran, que tiene como hipótesis nula que no hay correlación entre los residuos de la regresión, pero que no tiene una hipótesis alternativa explícita.





**Figura 4.1:** Ejemplos de matrices de distancias. El inciso (a) corresponde la región a la cual se le quiere obtener la matriz de distancias, los incisos (b), (c) y (d) corresponden a matrices de distancias correspondientes a la región en el inciso (a) bajo distintos criterios: adyacencia, vecino más cercano y distancia < 2 respectivamente. Fuente: [Arbia \(2014\)](#).

Algo que cabe indicar es que la prueba de autocorrelación para series de tiempo conocida como Durbin-Watson se publicó en el mismo año que la de Moran y puede ser considerada un caso particular de ésta última, una prueba de esto puede encontrarse en [Arbia \(2006\)](#). La forma de la prueba de Moran es la de una correlación entre los residuos de la regresión y sus valores espacialmente rezagados

$$\text{corr}(\epsilon, L\epsilon) = \frac{\text{cov}(\epsilon, L\epsilon)}{\sqrt{\text{var}(\epsilon)\text{var}(L\epsilon)}}$$

asumiendo que

$$\text{var}(\epsilon) = \text{var}(L\epsilon)$$

tenemos que

$$\text{corr}(\epsilon, L\epsilon) = \frac{\text{cov}(\epsilon, L\epsilon)}{\text{var}(\epsilon)} = \frac{\epsilon^T W \epsilon}{\epsilon^T \epsilon}$$

Existen versiones mas especializadas de esta prueba, ya que como se indico la prueba

de Moran no tiene una hipótesis alternativa explícita. Dichas versiones contienen como hipótesis alternativas el tipo de dependencia espacial que se espera probar por lo que se puede elegir el modelado correcto a partir del uso de dichas pruebas.

### 4.4. Modelos de regresión lineal espacial

Como se estableció en la introducción del capítulo, primero se discutirán los modelos de uso para datos de tipo sección cruzada. Esto es, modelos que incorporan la información de varios sujetos u objetos pero considerando que todas las observaciones se encuentran tomadas en el mismo punto en el tiempo.

En nuestro caso, pueden ser observaciones correspondientes a distintas localizaciones como municipios o estados, pero pensando que todas ellas fueron tomadas de manera simultánea. Esto nos sirve para permitirnos comparar el efecto de algunas variables sobre dichas localizaciones en un punto específico del tiempo.

Consideraremos un modelo en su forma más general de manera que se viole el supuesto de independencia en los residuos debido a la presencia de autocorrelación espacial. Dicho modelo tiene la forma

$$\begin{aligned} y &= \lambda W y + X\beta_{(1)} + W X\beta_{(2)} + u & |\lambda| < 1 \\ u &= \rho W u + \epsilon & |\rho| < 1 \end{aligned}$$

donde  $X$  es una matriz de regresores,  $W$  es la matriz de distancias o pesos dada de manera exógena, y  $\beta_{(1)}, \beta_{(2)}, \lambda, \rho$  los parámetros a estimar. Las restricciones sobre los parámetros  $\lambda$  y  $\rho$  se deben tener si la matriz de distancias se encuentra estandarizada por filas. Y además de esto, se debe cumplir que  $\epsilon|X = i.i.d. N(0, \sigma_\epsilon^2 I_n)$ .

Cabe mencionar que la matriz  $W$  que aparece 3 veces en ambas ecuaciones puede no ser igual en los tres casos. Sin embargo, regularmente es complicado justificar el caso en que se deba tener distintas matrices de pesos.

Una observación con respecto a las ecuaciones definidas anteriormente es que pueden ser reescritas como

$$\begin{aligned} y &= (I - \lambda W)^{-1} [X\beta_{(1)} + W X\beta_{(2)} + u] \\ u &= (I - \rho W)^{-1} \epsilon \end{aligned}$$

siempre que ambas inversas existan. Para asegurar la existencia de dichas inversas se hace referencia a el trabajo de [Kelejian et al. \(2004\)](#) que indica que bajo el supuesto de

que la matriz de distancias esta normalizada por filas entonces dichas matrices inversas existen al cumplir las restricciones  $|\rho| < 1$  y  $|\lambda| < 1$ .

Asimismo, el modelo descrito anteriormente es conocido como SARAR(1,1), o bien como *modelo autorregresivo con estructura adicional de error autorregresiva*.

Restringiendo por casos los parámetros a estimar para que algunos de ellos sean cero se pueden obtener otros modelos que pueden considerarse más simples:

- Si  $\beta = 0$  y  $\lambda = 0$  o  $\rho = 0$ , se conoce como Modelo Autorregresivo Espacial Puro.
- Si  $\lambda = \rho = 0$ , se le llama Modelo de Variable Independiente Rezagada.
- $\lambda = 0$ ,  $\rho \neq 0$  es un Modelo de Rezago Espacial.
- $\lambda \neq 0$ ,  $\rho = 0$  es conocido como Modelo de Error Espacial (SEM).
- $\lambda, \rho \neq 0$  se conoce como SARAR.

Es importante establecer la existencia de estos cinco casos puesto que cada uno tiene sus peculiaridades para establecer una solución y permitir el cálculo de los parámetros. Aunque todos ellos pueden ser resueltos basándose en un enfoque de maximización de la función de verosimilitud, en algunos casos se puede hacer uso de otros enfoques como mínimos cuadrados en dos etapas o mínimos cuadrados generalizados factibles que pueden representar una ligera ventaja con respecto a la versión de máxima verosimilitud.

## 4.5. Modelos espaciales de panel de datos

Los datos de panel espacial son un caso especial de datos de tipo panel en el que las observaciones se encuentran indexadas en dos dimensiones: a lo largo de las diversas unidades espaciales que se quieren estudiar y con respecto al tiempo, un tratamiento mas amplio de este tema se puede encontrar en [Elhorst \(2014\)](#) y en [Hsiao \(2014\)](#).

Este tipo de especificación del modelo permite tener un mayor control con respecto a posibles tendencias que se pueden presentar en los datos ya sea por cuestiones espaciales como temporales.

En [Hsiao \(2013\)](#) se establecen 3 ventajas principales de los modelos de panel de datos sobre los de sección cruzada así como varios ejemplos que respaldan estos hechos:

- **Mayor precisión en la inferencia sobre los parámetros del modelo.** Los modelos de panel de datos contiene mas grados de libertad y mayor variabilidad que los de sección cruzada lo que mejora la eficiencia de los estimaciones econométricas.

- **Mayor capacidad para capturar la complejidad del comportamiento humano.** Al permitir la incorporación de mas dimensiones que los datos de sección cruzada permite establecer patrones muchos mas generales.
- **Simplificación computacional y de la inferencia estadística** En algunos casos particulares como en el caso del análisis de series de tiempo no estacionarias donde los estimadores de máxima verosimilitud ya no siguen una distribución normal, si se agrega información de sección cruzada y entre dichas unidades de sección cruzada existe independencia se puede hacer uso del Teorema del Límite Central para mostrar que las distribuciones limite permanecen asintoticamente normales.

Para comprender mejor el modelo espacial de panel de datos analizaremos la versión no espacial primero, para mas detalles puede consultarse [Wooldridge \(2010\)](#) o [Hsiao \(2013\)](#). El modelo general para el caso no espacial tiene la forma

$$y_{it} = \alpha + \beta^T X_{it} + \epsilon_{it} \quad i = 1, \dots, n, \quad t = 1, \dots, S$$

donde  $i$  se refiere a los individuos a analizar,  $t$  al índice de tiempo,  $\alpha$  y  $\beta$  son los parámetros a estimar y  $\epsilon_{it}$  el termino de innovaciones que debe cumplir que  $\epsilon|X = i.i.d. N(0, \sigma_\epsilon^2 I_n)$ .

Para modelar la heterogeneidad individual regularmente se debe asumir que el término de error esta separado en dos componentes, con uno de ellos específico de los individuos y que no cambia con el tiempo. Para que esto suceda el modelo toma la siguiente forma

$$y_{it} = \alpha_i + \beta^T x_{it} + u_{it} = \alpha_i + \beta^T x_{it} + (\mu_i + \epsilon_{it})$$

donde el término  $\mu_i$  representa el componente de error individual y el término  $\epsilon_{it}$  representa el componente de error idiosincrático. Éste último se asume que es bien comportado e independiente tanto de las variables independientes como del componente de error individual [Wooldridge \(2010\)](#).

El método de estimación apropiado para el modelo anterior va a depender de las propiedades de los dos componentes de error en el modelo. Si el componente de error individual esta correlacionado con las variables independientes, el estimador de mínimos cuadrados ordinarios  $\beta$  sería inconsistente, por lo que regularmente se trata a  $\mu_i$  como un conjunto de  $n$  parámetros a ser estimados como si se considerara en el modelo general tener  $n$  interceptos. En este caso a este tipo de modelos se les conoce como de *efectos fijos*. De manera inversa, es decir, en el caso en que no haya correlación entre las variables independientes y el componente de error individual entonces se conoce como modelo de *efectos aleatorios*, un tratamiento mas amplio de esto puede encontrarse en [Elhorst \(2014\)](#) o en [Arbia \(2014\)](#).

De igual manera que la descrita anteriormente, en el caso de los modelos de panel

de datos espaciales también se tiene esta distinción entre modelos de efectos fijos y modelos de efectos aleatorios. Y dependiendo de la especificación del modelo se va a tener un método distinto de calcular los parámetros que nos interesan.

Sin embargo, la literatura de econometría espacial se encuentra caracterizada por usar como punto de partida la especificación de efectos aleatorios en lugar de la especificación de efectos fijos, puede verse en [Elhorst \(2014\)](#) para tener referencia de varios de estos trabajos.

#### 4.5.1. Modelos de panel de datos espaciales con efectos aleatorios

Como ya se estableció anteriormente en una especificación de efectos aleatorios, los efectos individuales no observados son asumidos como no correlacionados con las otras variables explicativas del modelo, y por lo tanto pueden ser consideradas como parte del término de error.

En el contexto espacial de modelos de efectos aleatorios se consideran tres subespecificaciones: el Modelo de Error Espacial con Efectos Aleatorios (SEM-RE), el Modelo KKP y el Modelo de Efectos Aleatorios con Rezago Espacial para mas detalles de estos modelos puede consultarse [Arbia \(2014\)](#). En términos prácticos el modelo SEM-RE es muy costoso computacionalmente por lo que no se considerara como parte de las subespecificaciones a modelar, pero se detalla un poco en las siguientes líneas para cubrir la teoría de las tres subespecificaciones.

##### Modelo de Error Espacial con Efectos Aleatorios (SEM-RE)

Este tipo de modelo asume que  $\mu_i \sim i.i.d.N(0, \sigma_\mu^2)$  y que el término de error idiosincrático  $\epsilon$  obedece a una formulación de Error Espacial, de manera que para cada momento del tiempo ( $t=1, \dots, S$ ), tenemos que

$$\epsilon_{it} = \rho W \epsilon_{it} + \eta_i$$

de manera que como consecuencia directa tendremos que  $\epsilon_{it} = (I_\rho W)^{-1} \eta_i$ . Así al definir  $B = (I_n - \rho W)$  con  $I_n$  la matriz identidad de tamaño  $n \times n$ ,  $W$  la matriz de distancias y  $\rho$  el parámetro de dependencia del error espacial. Por lo que para todo el panel el componente de error idiosincrático se escribe como

$$\epsilon = (I_S \otimes B^{-1}) \eta$$

con  $\eta$  un vector de tamaño  $nS$  tal que  $\eta \approx n.i.d.N(0, \sigma_\eta^2)$ . Debido a lo anterior, tenemos que el término de error compuesto  $u_i = \mu_i + \epsilon_{it}$  se puede escribir de forma compacta de la siguiente manera

$$u = (I_S \otimes I_n) \mu + (I_S \otimes B^{-1}) \eta$$

## 4. REGRESIÓN LINEAL ESPACIAL

---

con  $i_S$  un vector de unos de dimensión  $S$ . Si definimos  $J_S = i_S i_S^T$  como una matriz de  $S \times S$  de unos. De aquí tendremos que la matriz de covarianzas del error compuesto puede entonces ser escrita como

$$_{nS}\Omega_{nS} = \sigma_\mu^2(J_S \otimes I_n) + \sigma_\eta^2(I_S \otimes B_n^T B_n)^{-1}$$

que nos permitirá derivar la función de verosimilitud que requerimos, el calculo explícito de la función se puede encontrar en [Elhorst \(2014\)](#).

### La especificación KKP

Una forma alternativa para modelar los errores fue considerada por [Kapoor et al. \(2007\)](#). En esta alternativa se asume que tanto a los efectos individuales como a los otros componentes de error se les puede aplicar una estructura de correlación espacial. En este sentido, el término de error compuesto  $u = (i_S \otimes I_n)$  se asume que sigue un proceso autorregresivo espacial de primer orden de la forma

$$u = \rho(I_S \otimes W)u + \eta$$

a partir de aquí se puede obtener la matriz de covarianzas del error compuesto  $u$  que se expresa en forma abreviada como

$$_{nS}\Omega_{nS} = (I_S \otimes B^{-1})\Omega_\epsilon(I_S \otimes B^{-S})$$

con  $\Omega_\epsilon = \sigma_\mu^2 J_S + \sigma_\eta^2 I_S \otimes I_n$ .

### Modelo de Efectos Aleatorios con Rezago Espacial

Esta subespecificación es una extensión natural del caso espacial presentado anteriormente pero incorporando la estructura de panel de datos, por lo que el término del rezago espacial tiene en esencia la misma interpretación que en el caso espacial de sección cruzada.

#### 4.5.2. Modelos de panel de datos espaciales con efectos fijos

Tal como se indico anteriormente, si los efectos individuales no se encuentran correlacionados con las variables independientes, entonces dichos efectos pueden ser considerados como componentes del término de error y caracterizados como un modelo de efectos aleatorios. Pero al considerar que existe una correlación de los efectos individuales con las variables independientes entonces los coeficientes obtenidos usando mínimos cuadrados generalizados se vuelven inconsistentes.

En este caso, para obtener una buena estimación se requiere estimar cada uno de los

efectos individuales por separado o bien eliminar dichos efectos generando algunas diferencias en los datos, siendo este último procedimiento conocido como *time-demeaning*.

Este procedimiento de *time-demeaning* consiste en transformar los datos restando la media temporal a cada observación con el fin de eliminar los efectos espaciales individuales, por lo que tendremos las nuevas variables

$$y_{it}^* = y_{it} - \frac{1}{S} \sum_{t=1}^S y_{it} \quad y \quad x_{it}^* = x_{it} - \frac{1}{S} \sum_{t=1}^S x_{it}$$

con dicha transformación impuesta es posible hacer uso de los métodos de estimación estándar para el Modelo Rezago Espacial o bien para el Modelo de Error Espacial.

### 4.5.3. Estimación de los parámetros

Para todas las especificaciones mencionadas anteriormente se pueden estimar sus parámetros a través del Método de Maximización de la Función de Verosimilitud. Aunque también puede usarse el Método Generalizado de Momentos para estimar los parámetros, en el caso de los modelos de efectos aleatorios no es posible usarlo más que en el caso de KKP, por lo que solo se considerara para efectos de este trabajo el uso del método de Máxima Verosimilitud (MV).

Los métodos de estimación que a continuación se describirán pueden ser consultados en [Elhorst \(2003\)](#) para mas detalles técnicos sobre su funcionamiento.

### Modelos de efectos fijos

Primero consideraremos el Modelo de Rezago Espacial, para ello debemos definir  $B = (I_n - \lambda W)$  con  $\lambda$ ,  $W$  y  $I_n$  definidas como hasta ahora. El método de estimación consiste en una estimación iterativa, para ello primero generamos los vectores  $X^*$  y  $y^*$  usando la técnica de *time-demeaning*.

A partir de aquí consideramos los residuales derivados del modelo con un filtro espacial sobre  $y$  definido como

$$\eta^* = (I_S \otimes B)y^* - X^*\beta$$

Ahora tenemos que la verosimilitud concentrada con respecto a  $\beta$  y  $\sigma_\eta^2$  es

$$l = c - \frac{nS}{2} \ln(\sigma_\epsilon^2) + S \ln|B| - \frac{nS}{2} \ln(\eta^{*T} \eta^*)$$

donde  $c$  es una constante, entonces podemos maximizar dicha función con respecto a  $\lambda$ . Con este valor máximo de  $\lambda$  procedemos a utilizar Mínimos Cuadrados Generalizados, imponiendo las siguientes condiciones de primer orden

$$\hat{\beta} = (X^{*T} X^*)^{-1} X^* (I_S \otimes B) y^*$$

#### 4. REGRESIÓN LINEAL ESPACIAL

---

y

$$\sigma_\eta^2 = \frac{\eta^{*T} \eta^*}{nS}$$

y estos nuevos valores son reemplazados en la expresión de los residuales que se mencionó anteriormente. Este proceso se mantiene iterando hasta que se obtiene convergencia o se alcanza un punto de paro.

Ahora para el caso del Modelo de Error Espacial se puede seguir un procedimiento casi idéntico pero considerando ahora la nueva ecuación de residuales

$$\eta^* = (I_S \otimes A)y^* - X^*\beta$$

con  $A = (I - \rho W)$ , y la función de verosimilitud concentrada con respecto a  $\beta$  y  $\sigma_\eta^2$

$$l = c - \frac{nS}{2} \ln(\sigma_\eta^2) + S \ln|A| - \frac{nS}{2} \ln(\eta^{*T} \eta^*)$$

y ahora maximizando esta función con respecto a  $\rho$  procedemos a usar Mínimos Cuadrados Generalizados imponiendo las siguientes condiciones de primer orden

$$\hat{\beta} = (X^{*T} X^*)^{-1} X^* y^*$$

y

$$\sigma_\eta^2 = \frac{\eta^{*T} \eta^*}{nS}$$

estos nuevos valores se sustituyen en la ecuación de residuales y se sigue el proceso de manera iterativa hasta alcanzar algún tipo de convergencia.

#### Modelos de efectos aleatorios

En el caso de estos modelos se considera una transformación parcial de datos (*partial time-demeaning*) con el fin de eliminar la estructura de los efectos aleatorios y con ello poder aplicar los métodos estándar para modelos de rezago espacial o error espacial.

La función de general de verosimilitud en el caso de los modelos de rezago espacial con efectos aleatorios combinada con una estructura de covarianza del error ( $\Sigma$ ) es

$$l = c - \frac{nS}{2} \ln(\sigma_\eta^2) + \frac{1}{2} \ln|\Sigma| + S \ln|B| - \frac{1}{2\sigma_\eta^2} \frac{|(I_S \otimes B)y - X\beta|^T \Sigma^{-1} |(I_S \otimes B)y - X\beta|}{nS}$$

con  $\Sigma$  la matriz de covarianzas del error compuesto.

Con esto se inicia un proceso iterativo para calcular los estimados de máxima verosimilitud para los parámetros desconocidos. Para ello, se dan valores iniciales para  $\lambda$  y



los parámetros de covarianza del error compuesto. Con esto se obtienen valores para  $\beta$  y  $\sigma_\eta^2$  a partir de las condiciones de primer orden

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} (I_S \otimes B) y$$

y

$$\sigma_\eta^2 = \frac{|(I_S \otimes B)y - X\beta|^T \Sigma^{-1} |(I_S \otimes B)y - X\beta|}{nS}$$

Tanto para el caso de la especificación *KKP* como para *SEM – RE* se sigue esta misma idea pero se hace una distinción recordando que el término de error compuesto en ambas es distinto, para más detalles de esto se puede consultar [Arbia \(2014\)](#).



---

## Capítulo 5

# Análisis de resultados

---

Debido a la relación comercial México-EE. UU. que se suscitó a partir de 1994 con la firma del TLCAN, se han realizado diversos estudios para conocer el impacto POSTLCAN en la economía mexicana. En este sentido, y con base en el trabajo de [Blankespoor et al. \(2017\)](#), se explora la posible relación entre la distancia entre municipios a la frontera con EE. UU. con respecto al empleo y la especialización del trabajo.

**5.1. Modelo Pooling**

**5.2. Pruebas LM**

**5.3. Prueba de Hausman**

**5.4. Efectos aleatorios**

**5.4.1. Rezago espacial**

**5.4.2. KKP**

**5.5. Efectos fijos**

**5.5.1. Rezago espacial**

Empleo

Especialización del trabajo

**5.5.2. Error espacial**

Empleo

Especialización del trabajo

**5.6. Interpretación de los modelos finales**

---

Capítulo 6

# Conclusiones

---

fdsfs



# Referencias

---

Alvarez, A., Garduño-Rivera, R., y Nuñez, H. M. (2017). Mexico's north-south divide: The regional distribution of state inefficiency 1988–2008. *Papers in Regional Science*, 96(4), 843–858.

[8](#)

Arbia, G. (2006). *Spatial econometrics: statistical foundations and applications to regional convergence*. Springer Science & Business Media.

[19](#)

Arbia, G. (2014). *A primer for spatial econometrics: with applications in r*. Springer.

[ix](#), [19](#), [22](#), [23](#), [27](#)

Asheim, B. (1999). The territorial challenge to innovation policy: Agglomeration effects and regional innovation systems. i asheim, b. and k. smith (red.). *Regional Innovation Systems, Regional Networks and Regional Policy*.

[6](#)

Baylis, K., Garduño-Rivera, R., y Piras, G. (2012). The distributional effects of nafta in mexico: Evidence from a panel of municipalities. *Regional Science and Urban Economics*, 42(1-2), 286–302.

[8](#)

Bess, M. K. (2014). Routes of conflict: building roads and shaping the nation in mexico, 1941–1952. *The Journal of Transport History*, 35(1), 78–96.

## REFERENCIAS

---

[14](#)

Bess, M. K. (2016). Revolutionary paths: Motor roads, economic development, and national sovereignty in 1920s and 1930s mexico. *Mexican Studies/Estudios Mexicanos*, 32(1), 56–82.

[14](#)

Blankespoor, B., Bougna, T., Garduno-Rivera, R., y Selod, H. (2017). *Roads and the geography of economic activities in mexico*. The World Bank.

[v](#), [ix](#), [1](#), [10](#), [11](#), [13](#), [14](#), [29](#)

Coase, R. H. (1937). The nature of the firm. *economica*, 4(16), 386–405.

[5](#)

Dávila Flores, A. (2008). Los clusters industriales del noreste de México (1993-2003): Perspectivas de desarrollo en el marco de una mayor integración económica con Texas. *Región y sociedad*, 20(41), 57–88.

[8](#)

DOF. (2014). *Programa nacional de infraestructura 2014-2018*. Diario Oficial de la Federación.

[14](#)

Duranton, G., y Kerr, W. R. (2015). *The logic of agglomeration* (Inf. Téc.). National Bureau of Economic Research.

[5](#)

Duranton, G., Morrow, P. M., y Turner, M. A. (2014). Roads and trade: Evidence from the US. *Review of Economic Studies*, 81(2), 681–724.

[10](#)

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International regional science review*, 26(3), 244–268.

[25](#)



Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels* (Vol. 479). Springer.

[21](#), [22](#), [23](#), [24](#)

Ellison, G., Glaeser, E. L., y Kerr, W. R. (2010). What causes industry agglomeration? evidence from coagglomeration patterns. *American Economic Review*, 100(3), 1195–1213.

[5](#)

Enríquez, A., y Galindo, M. (2015). Empleo. *Serie de Estudios Económicos*, 1.

[4](#)

Gómez, O. G. (1990). Construcción de carreteras y ordenamiento del territorio. *Revista Mexicana de Sociología*, 49–67.

[14](#)

Hanson, G. H. (1998). North american economic integration and industry location. *oxford review of Economic Policy*, 14(2), 30–44.

[8](#)

Hill, V. R. (2006). Coffee price risk in the market: Exporter, trader and producer. data from uganda. *Working paper*.

[7](#)

Hsiao, C. (2013). Panel data analysis-advantages and challenges. *Working paper*.

[21](#), [22](#)

Hsiao, C. (2014). *Analysis of panel data* (Vol. 54). Cambridge University Press.

[21](#)

Kapoor, M., Kelejian, H. H., y Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of econometrics*, 140(1), 97–130.

[24](#)

## REFERENCIAS

---

Kelejian, H. H., Prucha, I. R., y Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. En *Spatial and spatiotemporal econometrics* (pp. 163–198). Emerald Group Publishing Limited.

20

Kemeny, T., y Storper, M. (2015). Is specialization good for regional economic development? *Regional Studies*, 49(6), 1003–1018.

4, 7

Mankiw, N. G. (2014). *Principles of economics*. Cengage Learning.

4

Murali, S. (2017). Job specialization and labor market turnover. *Working paper*.

4

OECD. (2009). *Promoting pro-poor growth employment*. Autor.

3

OIT. (2019). *Employment*. Descargado 2019-05-20, de [https://www.ilo.org/global/statistics-and-databases/statistics-overview-and-topics/WCMS\\_470295/lang--en/index.htm](https://www.ilo.org/global/statistics-and-databases/statistics-overview-and-topics/WCMS_470295/lang--en/index.htm)

3

Redding, S. J., y Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. En *Handbook of regional and urban economics* (Vol. 5, pp. 1339–1398). Elsevier.

10

Unger, K. (2003). Los clusters industriales en México: especializaciones regionales y la política industrial. 8

Williamson, O. E. (2007). The economic institutions of capitalism. firms, markets, relational contracting. En *Das summa summarum des management* (pp. 61–75). Springer.

[5](#)

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

[22](#)