



EL COLEGIO DE MÉXICO, A.C.
CENTRO DE ESTUDIOS ECONÓMICOS

"ENSAYOS EMPÍRICOS DE LA INDUSTRIA MANUFACTURERA EN MÉXICO: SOBRE LA RELACIÓN ENTRE LA INNOVACIÓN DE PRODUCTOS, EL GASTO EN I&D Y LOS SPILLOVERS DE CONOCIMIENTOS Y SOBRE LA RELACIÓN ENTRE LA COAGLOMERACIÓN ECONÓMICA Y LAS FUERZAS DE AGLOMERACIÓN MARSHALLIANAS"

TESIS PRESENTADA POR:

ALBA VERÓNICA MÉNDEZ DELGADO

PARA OPTAR POR EL GRADO DE

DOCTOR EN ECONOMÍA

PROMOCIÓN 2010-2013

DIRECTOR DE TESIS

DR. JOSÉ ANGEL CALDERÓN MADRID



CENTRO DE ESTUDIOS ECONÓMICOS

CONSTANCIA DE APROBACIÓN

Doctorante: Alba Verónica Méndez Delgado

Tesis: *“Ensayos empíricos de la industria manufacturera en México: sobre la relación entre la innovación de productos, el gasto en I & D y los spillovers de conocimientos y sobre la relación entre la coaglomeración económica y las fuerzas de aglomeración Marshallianas”*

Director de Tesis: Dr. José Angel Calderón Madrid

Aprobada por el Jurado Examinador:

Dr. José Angel Calderón Madrid	Presidente	_____
Dr. Carlos Chiapa Labastida	Primer Vocal	_____
Dr. Ignacio N. Lobato García Miján	Vocal Secretario	_____
Dr. Eneas Caldiño García	Suplente	_____

México, D.F., 25 de septiembre de 2015

Dedicatoria

A mis papás Pedro Méndez González (†)
y Yolanda Delgado Padilla

A mis hermanas
Carolina y Patricia

A mis sobrinos
Erik y Dante

Agradecimientos

Gracias a mis papás, hermanas y sobrinos por su amor, comprensión y apoyo.

A David le agradezco su cariño y por acompañarme durante mi paso por el COLMEX, también a su familia por recibirme en su hogar.

Agradezco a mi asesor el Dr. Angel Calderón por su tiempo, disposición y comprensión.

Al Dr. Lobato y al Dr. Chipa agradezco el tiempo que dedicaron a leer mi tesis y sus comentarios.

Gracias a mi tía Lucía por todas las tortillas de harina que me hizo durante seis años y a mis tías Carmen, Silvia y Flora que también me hicieron tortillas. A mis primos Poncho, Alma y Oscar les agradezco todos los libros que me consiguieron.

A mis amigos que leyeron mi tesis, a Alan por su apoyo y a Briones por su amistad, gracias.

A los bibliógrafos Carolina Palacios y Tomás Bocanegra gracias por su apoyo y eficiencia.

Resumen

La tesis está compuesta de dos partes independientes. En la primera abordamos la relación entre I&D, *spillovers* de conocimientos y la producción de innovaciones. En la segunda parte estimamos la importancia relativa de las fuerzas de aglomeración Marshallianas sobre la coaglomeración económica en México.

El tema de cómo conceptualizar y medir los *spillovers* de conocimientos generados por las empresas que gastan en I&D, la relevancia que éstos tienen en la producción de conocimientos de las empresas que se benefician de ellos sin pago de por medio y cómo medir de una manera económicamente eficiente e insesgada la relación entre el gasto en I&D de una empresa y su desempeño en la innovación de productos, constituye la primera parte de la tesis.

Para analizar el efecto del gasto en I&D en las innovaciones de productos utilizamos el modelo de Crepon et al. (1998) (CDM), donde se emplea una función de producción de conocimientos o de innovaciones.¹ En este trabajo incluimos en el modelo CDM el efecto de los *spillovers* de conocimientos sobre las innovaciones y con ello demostramos que al no hacerlo se da lugar a sesgos de estimación de la relación del gasto en I&D de la empresa y su desempeño en la innovación de productos.

A diferencia de los trabajos anteriores, en éste incorporamos al análisis el efecto de los *spillovers* como otro insumo de la producción de innovaciones, tomando en cuenta que su omisión puede generar sesgos en las estimaciones. Además, en la variable de *spillovers* se incorporan aspectos objetivos de la proximidad tecnológica, la cercanía espacial y la capacidad de asimilación de tecnología externa de las empresas, a través de un sistema de ponderadores.

¹La ventaja de utilizar el modelo CDM es que corrige sesgos de selección y de endogeneidad.

Los resultados indican que los *spillovers* de conocimientos son relevantes y tiene un efecto positivo sobre la propensión a innovar de las empresas. También, encontramos que la omisión de los *spillovers* de conocimientos producen una sobreestimación del efecto que la intensidad del gasto en I&D tiene en la innovación de productos. La relevancia de los *spillovers* y la sobreestimación del coeficiente del gasto en I&D no cambian cuando aplicamos el modelo CDM en el Edo. Méx. y NL. Con los *spillovers* espaciales no encontramos efectos sobre la producción de innovaciones.

El objetivo de la segunda parte de la tesis es medir la aglomeración económica, a través de un índice de coaglomeración, y evaluar la importancia relativa que, sobre la coaglomeración, tienen las fuerzas o externalidades Marshallianas.

Para estimar la coaglomeración económica utilizamos el índice de Ellison and Glaeser (1997) y Ellison et al. (2010). Entre las ventajas que surgen de las economías de aglomeración tenemos las externalidades Marshallianas derivadas de las relaciones entre clientes y proveedores que reducen los costos de producción o de compartir la fuerza laboral de la región y de un ambiente en donde los conocimientos e ideas fluyan entre las industrias localizadas en una misma área.

Con el propósito de estudiar el efecto de las externalidades Marshallianas, se realiza el ejercicio a nivel entidad federativa, lo cual nos permite controlar por las características relacionadas a la localización geográfica. A diferencia de los trabajos anteriores, estimamos matrices insumo-producto regionales para aprovechar la desagregación espacial de la información a nivel entidad federativa.

Encontramos que la fuerza de aglomeración económica con mayor importancia sobre la coaglomeración es la relación a través del mercado de bienes (cliente-proveedor), entre los diferentes subsectores de la manufactura en México. En la región Norte, Golfo y Sur también sobresale la importancia del mercado de bienes como incentivo para la coaglomeración.

Índice general

Dedicatoria	iii
Agradecimientos	v
Resumen	1
PRIMERA PARTE. Determinantes de la Propensión a la Innovación de Productos en México: estimaciones microeconómicas de la relevancia de la inversión en I&D de las empresas manufactureras y de los <i>spillovers</i> de conocimientos que ésta genera	5
SEGUNDA PARTE. Fuerzas Marshallinas y coaglomeración económica: estimaciones econométricas para México, 2003-2008	95

PRIMERA PARTE

Determinantes de la Propensión a la Innovación de Productos en México: estimaciones microeconómicas de la relevancia de la inversión en I&D de las empresas manufactureras y de los *spillovers* de conocimientos que ésta genera

Índice

PRIMERA PARTE. Determinantes de la Propensión a la Innovación de Productos en México: estimaciones microeconómicas de la relevancia de la inversión en I&D de las empresas manufactureras y de los <i>spillovers</i> de conocimientos que ésta genera	5
Introducción	9
1. Antecedentes y descripción del modelo CDM	12
1.1. Literatura empírica	14
1.1.1. Función de producción de conocimientos: el modelo CDM	15
1.1.2. Los <i>spillovers</i> en el modelo CDM	17
1.2. Descripción del modelo CDM	18
2. <i>Spillovers</i> de conocimientos	23
2.1. Características de los <i>spillovers</i> de conocimientos	23
2.2. Antecedentes empíricos de la estimación de los <i>spillovers</i>	26
2.3. Sistema de ponderadores para la estimación de los <i>spillovers</i> de conocimientos	28
3. Estimación de los <i>spillovers</i> de conocimientos	32
3.1. Microdatos de la ESIDET 2012	32
3.2. Estadísticas descriptivas	33
3.3. <i>Spillovers</i> de conocimientos	35

4. El modelo CDM	37
4.1. Intensidad del gasto en I&D	37
4.1.1. Estadísticas descriptivas	38
4.1.2. Estimación de la intensidad del gasto en I&D	40
4.2. Función de producción de conocimientos	46
4.2.1. Estimación de la propensión a innovar en productos	47
5. El modelo CDM y su aplicación a nivel estatal: Nuevo León y el Estado de México	53
5.1. Intensidad de gasto en I&D	53
5.2. Función de producción de conocimientos	58
Conclusiones	62
Bibliografía	62
A. Especificaciones econométricas	71
A.1. Modelo de selección muestral	71
A.2. Modelo semiparamétrico de selección muestral	73
A.3. Modelo probit	75
A.4. Medidas de ajuste: <i>AIC</i> y <i>BIC</i>	77
B. Definición de variables	78
C. Otras regresiones	81
Índice de cuadros	92

Introducción

Las externalidades positivas producidas por el conocimiento generado por las empresas que gastan en investigación y desarrollo (I&D), denominadas *spillovers* de conocimientos, benefician a la economía en su conjunto, aunque para la empresa que las generan implica una menor redituabilidad al no poder cobrar por la totalidad de los resultados derivados de su I&D.

En este trabajo abordamos el tema de cómo conceptualizar y medir los *spillovers* de conocimientos generados por las empresas que gastan en I&D, la relevancia que éstos tienen en la producción de conocimientos de las empresas que se benefician de ellos sin pago de por medio y cómo medir de una manera econométricamente eficiente e insesgada la relación entre el gasto en I&D de una empresa y su desempeño en la innovación de productos.

Para analizar el efecto del gasto en I&D en las innovaciones de productos utilizamos el modelo de Crepon et al. (1998) (CDM), donde se emplea una función de producción de conocimientos o de innovaciones.¹ En este trabajo incluimos en el modelo CDM el efecto de los *spillovers* de conocimientos sobre las innovaciones y con ello demostramos que al no hacerlo se da lugar a sesgos de estimación de la relación del gasto en I&D de la empresa y su desempeño en la innovación de productos.

Los trabajos empíricos previos, basados en el modelo CDM, han considerado que los *spillovers* influyen sobre la producción de innovaciones. Las variables *proxies* de los *spillovers* se obtienen de la respuestas que las empresas dan a la pregunta sobre la importancia de las fuentes de información para la innovación. Existen dos problemas con esta estrategia, el primero es que ese tipo de variables se derivan de información subjetiva, obtenida de las

¹La ventaja de utilizar el modelo CDM es que corrige sesgos de selección y de endogeneidad.

encuestas de innovación, y depende de la percepción de quienes contestan las encuestas. El segundo problema es que no relacionan los *spillovers* con la producción de innovaciones de forma directa. La relevancia de la relación directa entre las innovaciones y los *spillovers* puede observarse si pensamos en una empresa que no realiza gasto en I&D pero tiene acceso al conocimiento derivado de los esfuerzos en I&D de otras empresas, y sin realizar ningún pago de por medio, la empresa produce alguna innovación.

Para poder diferenciar la capacidad que tienen las empresas de acceder y asimilar el conocimiento que se genera en la economía, así como su proximidad espacial, utilizamos un sistema de ponderadores (Kaiser, 2002; Aiello y Cardamone, 2008). De esta forma tomamos en cuenta que el aprovechamiento del conocimiento disponible para cualquier empresa se ve limitado cuando carece de personal calificado o cuando no pertenece a sectores tecnológicos afines a aquéllos en los que se generan los *spillovers* de conocimientos.

A diferencia de los trabajos anteriores, en éste incorporamos al análisis el efecto de los *spillovers* como otro insumo de la producción de innovaciones, tomando en cuenta que su omisión puede generar sesgos en las estimaciones. Además, en la variable de *spillovers* se incorporan aspectos objetivos de la proximidad tecnológica, la cercanía espacial y la capacidad de asimilación de tecnología externa de las empresas, a través de un sistema de ponderadores.

Para construir el índice que representa los *spillovers* de conocimientos y realizar las estimaciones econométricas de la relación entre los *spillovers*, la I&D y la innovación en productos, utilizamos información de la Encuesta sobre Investigación y Desarrollo Tecnológico (ESIDET) del año 2012.

Para el propósito de éste estudio la ESIDET 2012 tiene dos ventajas, la primera es que sigue la metodología de los manuales Frascati y Oslo de la OCDE, que ha sido ampliamente aceptada a nivel internacional y forma parte de una tendencia de homogeneización de las encuestas de innovación. La segunda es que por primera vez en México, la ESIDET correspondiente al año 2012 cuenta con representatividad a nivel entidad federativa.

La ESIDET contiene información detallada sobre el proceso de innovación de las empresas y es resultado de la colaboración entre el Instituto Nacional de Estadística y Geografía (INEGI) y el Consejo Nacional de Ciencia y Tecnología (CONACYT). La unidad de obser-

vación de la encuesta son las empresas de 20 o más empleados y nos concentramos en el sector manufacturero. La realización de este trabajo no hubiera sido posible sin el acceso al Laboratorio de Análisis de Datos (LAD) del INEGI, ya que requerimos de información desagregada a nivel empresa. Los resultados obtenidos a partir de la ESIDET que se presentan en este trabajo fueron revisados por el LAD para comprobar que los principios de confidencialidad de la encuesta fueron debidamente respetados.

Para alcanzar el objetivo de este trabajo primero estimamos la intensidad del gasto en I&D con y sin variables subjetivas (*proxies* de los *spillovers*). Después incluimos el efecto directo de los *spillovers* en las innovaciones, que diferencia a este trabajo de los anteriores, siguiendo del modelo CDM. Además, aprovechando que la ESIDET 2012 tiene información estatal, construimos *spillovers* que toman en cuenta la distancia geográfica entre las empresas y también es posible utilizar controles espaciales en el modelo CDM. Adicionalmente, estudiamos los efectos de los *spillovers* sobre las innovaciones a nivel entidad federativa, aplicando el modelo CDM para Nuevo León (NL) y el Estado de México (Edo. Méx.).

Los resultados indican que los *spillovers* de conocimientos son relevantes y tiene un efecto positivo sobre la propensión a innovar de las empresas. También, encontramos que la omisión de los *spillovers* de conocimientos producen una sobreestimación del efecto que la intensidad del gasto en I&D tiene en la innovación de productos. La relevancia de los *spillovers* y la sobreestimación del coeficiente del gasto en I&D no cambian cuando aplicamos el modelo CDM en el Edo. Méx. y NL. Con los *spillovers* espaciales no encontramos efectos sobre la producción de innovaciones.

La primera parte de la tesis se estructura de la siguiente manera, en el primer capítulo se presenta la revisión de literatura y descripción del modelo CDM. El segundo capítulo incluye las características de los *spillovers*, sus antecedentes empíricos y la metodología. El tercer capítulo contiene los datos y la estimación de los *spillovers* de conocimientos. La estimación de la intensidad del gasto en I&D y de la función de producción de conocimientos están en el cuarto capítulo. En el quinto capítulo se estima el modelo CDM para el Edo. Méx. y NL. Al final se presentan las conclusiones del trabajo.

Capítulo 1

Antecedentes y descripción del modelo CDM

La principal herramienta para estudiar el efecto del gasto en I&D sobre la generación de conocimientos es la función de producción de conocimientos, propuesta por Griliches (1979), donde supone que la producción de nuevo conocimiento depende del gasto en I&D, pasado y presente, y otros factores como los flujos de conocimientos externos a la empresa.¹ A partir de dicha función se puede examinar si los gastos en investigación, de otras industrias y de las universidades, incrementan la actividad innovadora. También se incorpora la dimensión espacial al análisis para estudiar el efecto del gasto en I&D sobre las innovaciones en una región a través de la función de producción de conocimientos de Griliches-Jaffe (Jaffe, 1989).

Por su parte, Crepon et al. (1998) proponen un modelo, conocido como CDM, que vincula la decisión que toman las empresas sobre gastar en I&D y el monto destinado a la I&D; éste gasto se relaciona con la producción de innovaciones, tomando como base una función de producción de conocimientos, y finalmente, se relaciona con la productividad de las empresas.² El modelo CDM incorpora aspectos como el hecho de que las empresas pueden destinar recursos a mejorar los procesos y productos, sin contar con un área especial para

¹Una revisión detallada de la evolución de la función de producción de Griliches está en el documento de Acs (2009).

²Para esta investigación nos vamos a centrar en la función de producción de conocimientos y es por esto que consideramos sólo las dos primeras etapas del modelo CDM.

la I&D y que no todas las innovaciones son registradas por las empresas pero sí afectan su productividad.

El progreso tecnológico, considerado como factor de producción, implica la creación de ideas cuya apropiación es diferente de los otros factores. Se puede difundir dado que es parcialmente no-exclusivo y, además, es no-rival (Keilbach, 2000).³

El monto no apropiable del conocimientos, producto de los esfuerzos de innovación de una empresa, se denominan *spillovers* tecnológicos (Kaiser, 2002). Las empresas pueden adquirir estos *spillovers* tecnológicos de otras empresas sin realizar ninguna transacción de mercado, es decir, sin pagar total o parcialmente por la información que utilizan (Grossman y Helpman, 1991).

Dadas las características del progreso tecnológico, la distancia espacial influye en el proceso de difusión de conocimientos entre las empresas, debido a la necesidad de proximidad geográfica para transmitir el conocimiento, como sugiere la teoría de localización (Feldman, 1994a). El conocimiento nuevo se compone en parte por conocimiento tácito que se trasmite mejor de persona a persona, y por sus características está limitado en el espacio (Acs, 2002).

Dentro del estudio de las economías de aglomeración se abordan aspectos relacionados a los *spillovers* de conocimientos. Cuando hay aglomeración económica se generan ventajas de la interacción personal para el intercambio de información, fenómeno decreciente con la distancia. Una de las fuerzas de aglomeración, según Marshall (1890), son los *spillovers* de conocimientos. La ventaja que obtienen las empresas al agruparse en el espacio es que pueden aprender de sus competidores estudiando sus productos o en algunos casos copiando sus diseños, y así no sólo se benefician de su propia I&D sino también de las externalidades de su entorno (Krugman y Obsfeld, 2005). Otra fuente importante del *know-how* es el intercambio informal de información e ideas entre los trabajadores, situación que se facilita con la cercanía de las empresas que permite la interacción entre ellos.

³El conocimiento es un bien no-rival porque puede ser utilizado por un agente sin limitar el uso de otros. También es parcialmente exclusivo porque se puede restringir su uso, al menos por un periodo de tiempo, a través de las patentes y otros instrumentos de protección de la propiedad intelectual (Keilbach, 2000). A partir de la no exclusividad del conocimiento, es decir, la imposibilidad de su completa apropiación, la I&D privada puede generar *spillovers* de conocimientos (Acs, 2002).

1.1. Literatura empírica

A partir de la función de producción de conocimientos se ha encontrado que las actividades de innovación crecen ante el aumento del gasto privado en I&D y de la inversión en investigación universitaria en una región determinada (Griliches, 1979). Si se compara la localización de los *spillovers* académicos e industriales a nivel regional, los *spillovers* universitarios están más localizados que los empresariales (Adams, 2002). Este fenómeno ocurre porque el conocimiento generado en las universidades es más abierto, mientras el industrial está protegido por derechos de propiedad intelectual. Hay evidencia de que los *spillovers* se facilitan por la coincidencia geográfica de investigación privada y universitaria, incluso cuando se incorpora al análisis la proximidad tecnológica se puede confirmar que los *spillovers* se localizan en una región y esto ocurre independientemente del tamaño de las empresas (Jaffe, 1989; Acs et al., 1991; Autant-Bernard, 2001; Anselin et al., 1997, 2000).⁴

En cuanto a las condiciones del mercado, la aglomeración económica tiene efectos sobre la intensidad de las transferencias locales de conocimientos, de las universidades hacia las empresas innovadoras (Varga, 2000). La concentración de las actividades de innovación puede estar determinada por una previa aglomeración de la producción (Jaffe et al., 1993). Sin embargo, aún controlando por la distribución geográfica de la producción, las actividades de innovación tienden a agruparse más y este fenómeno se acentúa en industrias donde los *spillovers* juegan un papel decisivo (Feldman y Kogler, 2010; Audretsch y Feldman, 1996b). La influencia de la aglomeración económica es positiva sobre la concentración de las actividades de innovación en las primeras etapas del ciclo de vida de la industria, después se convierte en efectos de congestión que generan dispersión de las actividades de innovación (Audretsch y Feldman, 1996a).

Al considerar el tamaño de las empresas en el estudio de los *spillovers*, el resultado es que las empresas pequeñas se benefician en mayor medida de recursos externos y de la cercanía

⁴Jaffe (1989) utiliza patentes para representar las innovaciones, sin embargo, Scherer (1983), Mansfield (1984) y Griliches (1990) advirtieron que utilizar el número de patentes no es una medida adecuada como variable *proxy* de la producción de innovaciones debido a que no todas las innovaciones se registran como patentes. Acs et al. (1991) utiliza el número de innovaciones registradas por las empresas.

de otras empresas (Feldman, 1994b). En cuanto al gasto en I&D privado, éste juega un papel relativamente más importante en la generación de innovación dentro de las grandes empresas (Acs et al., 1991). Las empresas pequeñas parecen tener más posibilidades de transformar el conocimiento obtenido de instituciones universitarias, respecto a las empresas grandes, y de ésta forma incrementan su actividad interna de I&D (Link y Rees, 1990). Las ventajas de las empresas grandes giran entorno a una mayor capacidad para invertir en I&D, acceso a financiamiento, economías de escala, diversificación de la investigación y mayores posibilidades de absorber los *spillovers* de conocimientos (Crespi y Zuñiga, 2012; Crespi et al., 2014).

1.1.1. Función de producción de conocimientos: el modelo CDM

Para estudiar la relación entre gasto en I&D, innovaciones y productividad, Crepon et al. (1998) propone un modelo estructural, conocido como modelo CDM. Las estimaciones económicas enfrentan algunos obstáculos, primero, hay un sesgo de selección en la muestra ya que las empresas se auto-seleccionan a participar en actividades relacionadas con la innovación. Además, existe endogeneidad entre el gasto en I&D, innovaciones y productividad. El modelo CDM considera estas dos dificultades en las estimaciones.

Los resultados que obtienen Crepon et al. (1998), para Francia, respaldan que el tamaño de la empresa afecta positivamente la probabilidad de hacer I&D, así como la participación de mercado y la diversificación. Las condiciones de demanda y las oportunidades tecnológicas afectan positivamente la decisión de hacer o no I&D y el monto destinado a ésta actividad. La producción de innovaciones también se ve afectada positivamente por estas dos variables y tiene una relación positiva con el gasto en I&D. Finalmente la productividad aumenta con las innovaciones controlando por la composición del trabajo y la intensidad del capital físico

Otro trabajo para Francia lo realizaron Mairesse y Mohnen (2005) donde estiman la intensidad del gasto en I&D y el impacto sobre diferentes medias de innovación así como una comparación interindustrial. Confirman una relación positiva entre el gasto en I&D y las medias de innovación como innovación en productos nuevos para la empresa, para el mercado, en procesos y patentes.

Crespi y Zuñiga (2012) analizan el vínculo entre gasto en innovación, producción de innovaciones y productividad laboral entre países de América Latina (Argentina, Chile, Colombia, Costa Rica, Panamá y Uruguay). En todos los países, las empresas que invierten en I&D tienen mayor probabilidad de introducir nuevas tecnologías y aquellas que innovan tienden a aumentar su productividad.

En Chile, no se observa un efecto contemporáneo de la innovación en productos sobre la productividad, pero sí un efecto positivo de la innovación en procesos, lo que podría ser una señal de un proceso de aprendizaje lento (Alvarez et al., 2010). Las empresas grandes y con mayor participación de mercado realizan más gasto en I&D y las empresas grandes tienden a innovar en mayor medida (Benavente, 2006). Además, las empresas que realizan exportaciones e inversión en I&D presentan un efecto positivo en la productividad (Bravo-Ortega et al., 2014).

En Colombia, la innovación fomenta las ventas por empleado y la productividad total de los factores. Además, la innovación es mayor en empresas medianas, grandes, exportadoras, con participación extranjera y con mayor gasto destinado a la I&D (Arbeláez y Parra, 2011).

Los gastos en I&D y la adquisición de tecnología tienen un impacto positivo en la introducción de nuevos productos al mercado argentino (Chudnovsky et al., 2006). Las empresas grandes son más innovadoras y las innovadoras reflejan mayores niveles de productividad.

En Brasil, la inversión en I&D aumenta la inversión en capital físico (De Negri et al., 2007). Las actividades que llevan a mayores niveles de productividad son el cambio organizacional, la cooperación con los clientes, el desarrollo de capital humano, el uso de tecnologías de comunicación y las exportaciones. En el largo plazo, también la I&D influye sobre nivel de productividad (Goedhuys, 2007).

Para México, Calderón-Madrid (2012) con el modelo CDM y utilizando la ESIDET 2010, encuentra que existe una relación positiva del tamaño de las empresas respecto a la decisión de gastar en I&D. Las exportaciones tienen efectos mixtos sobre el monto a gastar en I&D. Mientras que el tamaño de las empresas y el gasto en I&D tienen un efecto positivo sobre la innovación. Finalmente, el gasto en I&D y las innovaciones, por separado, así como el capital físico afectan positivamente la productividad en México.

Por su parte, Dutrénit et al. (2013) realizan un análisis de la innovación en sector de servicios utilizando también el modelo CDM, para México. Encuentran que las exportaciones y las fuentes de información afectan positivamente la intensidad del gasto relacionado con las actividades de innovación. En la función de producción de conocimientos observaron una relación positiva entre la intensidad del gasto en innovación y las exportaciones, con la probabilidad de innovar.

Raffo et al. (2008) compara el desempeño económico y la innovación en países de América Latina y Europa (Francia, España, Suiza, Argentina, Brasil y México). Las empresas tienden a innovar para mejorar su situación económica con bases similares entre países, pero su interacción con el sistema nacional es más débil en los países en desarrollo.

Por su parte, Griffith et al. (2006) hacen una comparación de los resultados del modelo CDM entre países europeos (Francia, Alemania, España y Reino Unido). Encontraron similitudes entre los países respecto a la decisión de realizar I&D y el monto a gastar en este rubro. Las empresas que tienen relación con el mercado internacional usan mecanismos de protección para sus innovaciones, reciben financiamiento gubernamental y son empresas grandes, con mayor probabilidad de invertir en I&D y su intensidad de gasto es más alta. También encuentran una relación positiva del tamaño de las empresas y la intensidad del gasto en I&D con la producción de innovaciones.

1.1.2. Los *spillovers* en el modelo CDM

En la mayoría de las investigaciones basadas en el modelo CDM se utilizan como variables *proxies* de los *spillovers* las fuentes de información que promueven la innovación. Los datos para aproximar dichos *spillovers* se derivan de información subjetiva de las encuestas de innovación, dependen de la percepción de quienes evalúan la importancia de las fuentes de información para hacer innovaciones. Además, se incorporan al modelo CDM en la etapa donde se determina el monto del gasto en I&D, así sólo capturan el efecto indirecto de los *spillovers* sobre la producción de innovaciones (Mairesse y Mohnen, 2005).

A diferencia de ellos, Goya et al. (2013) incluyen los *spillovers* en el CDM para estimar la decisión de gastar en I&D, utilizan como variable *proxy* la proporción de empresas que

hacen I&D y la suma del gasto en I&D de las empresas que hacen innovaciones. A partir de sus estimaciones, encuentran una relación positiva entre la decisión de gastar en I&D y la productividad con los *spillovers*.

Para esta parte de la tesis, en el contexto del modelo CDM, consideramos el impacto directo de los *spillovers* sobre la producción de innovaciones. A su vez, para estimar variables *proxies* de los *spillovers* utilizamos un sistema de ponderadores que hace posible incorporar la proximidad tecnológica de las empresas, la cercanía espacial y la capacidad de asimilación de tecnología externa, tomando en cuenta los esfuerzos de innovación de las empresas y sus características tecnológicas. Otra diferencia con los trabajos que se han revisado es que contamos con información espacial que permitirá aplicar el modelo CDM con controles espaciales y analizar los casos del Edo. Méx. y NL. La forma en que estimamos los *spillovers* se describe en el Capítulo 2.

1.2. Descripción del modelo CDM

Originalmente, Griliches (1979) propuso una función de producción de conocimientos para estudiar la relación entre la generación de innovaciones y la investigación, tanto privada (de empresas) como pública (de universidades). Después apareció otra propuesta para analizar la producción de conocimientos utilizando un modelo recursivo que relaciona los insumos con la producción de conocimientos y posteriormente, las innovaciones con la productividad.

El modelo CDM (Crepon et al., 1998) vincula los insumos para la producción de conocimientos (I&D y otras actividades) con la generación de innovaciones (producción de conocimientos), lo que se esperaría tuviera un impacto directo sobre la situación económica (productividad). El modelo CDM consiste en tres etapas: primero, las empresas toman la decisión de invertir o no en I&D y definen la intensidad de gasto en I&D, después, con la función de producción de conocimientos se vincula la intensidad del gasto en I&D y la producción de innovaciones, en la tercera etapa, se relaciona la productividad y las innovaciones a través de una función de producción. Para este trabajo nos concentramos en las dos primeras etapas.⁵

⁵El propósito de esta investigación es estudiar el efecto de los *spillovers* sobre las innovaciones, por lo que



Fuente: Ampliación con base al modelo de Crepon et al. (1998) y se ajusta con los modelos de Griffith et al. (2006), Crespi y Zuñiga (2012) y Calderón-Madrid (2012).

Figura 1.1: El modelo CDM ampliado

La Figura 1.1 es un diagrama de la relación de las variables dentro del modelo. El diagrama se basa en el de Crepon et al. (1998) y se ajusta con los modelos de Griffith et al. (2006), Crespi y Zuñiga (2012) y Calderón-Madrid (2012). La principal diferencia con los modelos anteriores es que se consideran explícitamente los *spillovers* de conocimientos como un factor que influye sobre la producción de innovaciones. Además, al disponer de información sobre la ubicación de las empresas es posible controlar por efectos espaciales (a nivel entidad federativa).

En la Figura 1.1 se presenta la relación de la I&D con el financiamiento, colaboración, participación de mercado, con el mercado externo y el tamaño de las empresas. Las últimas variables también influyen en la producción de innovaciones que está en función del valor estimado de la intensidad del gasto en I&D y de los *spillovers* de conocimientos.

estimamos la función de producción de conocimientos utilizando la propuesta de Crepon et al. (1998).

Los *spillovers* de conocimientos son un fenómeno que afecta la producción de innovaciones y se supone que son independientes de los esfuerzos de I&D de las empresas. En este trabajo se definen los *spillovers* de conocimientos como el monto de I&D indirecta disponible para la empresa, es decir, es la parte no apropiable del conocimiento producido a través del esfuerzo de innovación de las empresas (Aiello y Cardamone, 2005; Kaiser, 2002).

El modelo CDM considera algunas características de los datos cuando se relaciona la I&D y la producción de innovaciones. Primero tenemos que sólo una proporción de las empresas se auto-selecciona a realizar I&D. Además, la I&D es una variable endógena en la función de producción de conocimientos. El modelo CDM se ajusta a las características de los datos disponibles y propone una solución al problema de endogeneidad.

Especificación del modelo

El modelo CDM es un sistema de ecuaciones recursivas que relaciona los insumos para la generación de conocimientos y la producción del conocimiento de cada empresa ($i, j = 1, \dots, N$). La primera ecuación representa la intensidad del gasto en insumos para la producción de conocimientos (IID_i^*):

$$IID_i^* = \beta x_i + u_i \quad (1.1)$$

la intensidad del gasto en I&D (IID_i^*) es una variable latente, x es el vector de factores que influyen en la intensidad, β los coeficientes y u el término de error. Dado que no podemos observar IID_i^* utilizamos como *proxy* el gasto en I&D por empleado.

Pueden existir factores que influyan en la decisión de gastar en I&D por parte de las empresas. Para resolver este problema se utiliza un modelo Tobit generalizado o modelo de selección muestral en donde se incluye una ecuación de selección, para considerar la auto-selección a gastar en I&D, definida de la siguiente forma:⁶

$$SID_i = 1[\delta z_i + v_i > 0] \quad (1.2)$$

La Ecuación (1.2) indica si la empresa realizó o no gasto en I&D, donde SID_i toma valor de uno si el gasto es positivo y cero de otra forma. Si la empresa decide realizar I&D es posible observar la intensidad del gasto en I&D (IID_i).

⁶En el Apéndice A se detalla el modelo de selección muestral.

Los supuestos del modelo Tobit generalizado son que las variables independientes y la propensión a gastar en I&D (SID_i) siempre se observan, mientras que el gasto en I&D (IID_i) sólo se observa cuando $SID_i = 1$. Además, los errores son independientes de las variables explicativas. También se supone que (u_i, v_i) siguen una distribución normal bivariada con media cero, $var(u_i) = \sigma_1^2$, $cov(u_i, v_i) = \sigma_{12}$ y $var(v_i) = 1$.

Para obtener coeficientes estimados consistentes a través del modelo de selección muestral es necesario que se cumpla el supuesto sobre la distribución conjunta de los errores debido a que éste procedimiento es sensible a la especificación de la distribución de los errores (Newey et al., 1990). Ante esta situación han surgido modelos semiparamétricos que relajan el supuesto de la distribución conjunta de los errores y permiten obtener estimadores consistentes.⁷

En la primera etapa del CDM se estima la IID para todas las empresas. El supuesto detrás de este paso es que las empresas pueden estar realizando esfuerzos para innovar y no registran el gasto en esas actividades.

La segunda etapa consiste en estimar la producción de conocimientos, o propensión a innovar, utilizando los resultados de la etapa anterior.⁸ La ecuación de innovación se calcula con un modelo probit cuando la variable dependiente es dicotómica:⁹

$$INN_i = I\hat{I}D_i\alpha + \gamma_c c_i + \epsilon_i \quad (1.3)$$

donde INN_i es la producción de innovaciones (o propensión a innovar cuando la variable dependiente es dicotómica) a partir de insumos como el gasto en I&D ($I\hat{I}D_i$), otras variables de control (c) y el término de error (ϵ).

En el modelo CDM, los *spillovers* de conocimientos se incorporan en la primera etapa, es decir, su efecto en las innovaciones es indirecto (Mairesse y Mohnen, 2005). Además, los *spillovers* son aproximados utilizando información subjetiva de las encuestas de innovación. En este trabajo vamos a estudiar los efectos directos de los *spillovers* en la producción

⁷Aunque las estimaciones semiparamétricas quedan fuera del objetivo de este trabajo en el apéndice A se presentan algunas referencias.

⁸Crepon et al. (1998) señala que la I&D es un factor endógeno en la producción de conocimientos así que al utilizar los valores estimados de la IID , en lugar de los observados, se busca corregir la endogeneidad.

⁹En el Apéndice A se detalla el modelo probit.

de conocimientos y los *spillovers* serán aproximados con un sistema de ponderadores que incorporan la cercanía tecnológica y espacial, así como la capacidad de asimilación de las empresas. Por lo que incluimos los *spillovers* en la ecuación de innovación:

$$INN_i = SPILL_i\alpha_1 + I\hat{I}D_i\alpha_2 + \gamma_c c_i + \epsilon_i \quad (1.4)$$

donde, $SPILL_i$ será una variable *proxy* de los *spillovers* de conocimientos que representan las derramas de conocimientos de las empresas. Para estimar esta variable utilizamos el gasto de las otra empresas así como sus diferencias tecnológicas y la distancia geográfica entre ellas.¹⁰

La ventaja de utilizar el modelo CDM es que considera las relaciones endógenas entre las variables al sustituir los valores observados por los estimados ($I\hat{I}D$) y esto se logra a través de la estimación del modelo por etapas. La intensidad de la I&D está en función del financiamiento y colaboraciones, donde se supone que sólo afecta la innovación de forma indirecta, a través del gasto en I&D. Cabe señalar que el modelo se estima para todas las empresas, las que gastan en I&D y las que no, porque se supone que pueden existir esfuerzos de innovación que las empresas no reportan como gasto en I&D.

¹⁰El modelo CDM original tiene una última etapa donde se estima la productividad. Esta se determina a partir del nivel de innovaciones de la empresa, la intensidad del capital físico, el tamaño de la empresa y algunos autores incluyen el nivel de calificación de los empleados. En este trabajo no consideramos esta etapa debido a que nos interesa la relación de los *spillovers* y la producción de innovaciones, además, la estimación de la productividad requiere información complementaria a la que existe en la ESIDET.

Capítulo 2

Spillovers de conocimientos

La tecnología externa utilizada por una empresa esta influenciada por los conocimientos generados a partir de la I&D que se desarrolla fuera de ella (Griliches, 1979). Los flujos de conocimientos o *spillovers* derivados de la I&D externa surgen debido a que las empresas no pueden apropiarse por completo de los conocimientos que generan en sus procesos de innovación. Si conocemos el monto de I&D indirecto disponible para la empresa, es decir, el gasto en I&D de otras empresas, es posible estimar un indicador que represente los *spillovers* de conocimientos (Aiello y Cardamone, 2005).

2.1. Características de los *spillovers* de conocimientos

El conocimiento se define como la información y habilidades que los individuos tienen para resolver problemas, tomar decisiones y procesar nueva información (Doring y Schnellenbach, 2006). Cuando una empresa genera nuevo conocimiento, a través de la I&D, se pueden presentar derramas de información hacia otras empresas, es decir, la empresa no puede apropiarse por completo de los conocimientos nuevos. Existen *spillovers* de conocimientos cuando un agente (empresa o individuo) accede al conocimiento creado por otro agente económico local, sin interacción en el mercado o compensación financiera (Karlsson et al., 2013).

Otra forma de entender los *spillovers* de conocimientos es pensando en ellos como ganancias intelectuales derivadas del intercambio de información, por las cuales no hay una compensación directa al productor, o es menor a su valor (Caniels, 2000; Feldman y Kogler, 2010). Los *spillovers* que surgen del proceso de I&D, sin pago de por medio, se denominan *spillovers* puros o de creación de ideas (Griliches, 1979).

Por otro lado, están los *spillovers* relacionados con el intercambio de bienes (*rent spillovers*) que se derivan del conocimiento que tiene por características la rivalidad y exclusividad (Griliches, 1979). Estos *spillovers* se presentan cuando hay transacciones económicas entre las empresas, a través del comercio, la inversión directa, el pago de tecnología, la contratación de trabajadores, colaboraciones y fusiones entre empresas (Hall et al., 2010).

En esta parte de la tesis nos enfocamos en los *spillovers* puros que surgen del conocimiento que se caracteriza por ser no-rival y parcialmente no-exclusivo. La no-exclusividad se refiere a que es difícil evitar que alguien lo consuma, es decir, los resultados de la investigación, realizada por algún agente, se pueden difundir a través de canales de comunicación, tales como publicaciones en revistas, libros, o en sitios web (Arrow, 1962). Por ello, el conocimiento forma parte del ámbito público y está disponible para cualquiera que lo busque. La no-rivalidad del conocimiento implica que no sólo puede ser consumido por muchos individuos al mismo tiempo, además, los usuarios de dicho conocimiento no disminuyen su cantidad o calidad disponible para otros (Feldman y Kogler, 2010).

La transmisión del conocimiento depende del nivel en que se puede comunicar. En este contexto tenemos dos tipos de conocimientos, el codificado y el tácito, el conocimiento codificado es información técnica que se puede encontrar en publicaciones y se transmite fácilmente a través de los medios de comunicación (Gertler, 2003). Este tipo de conocimiento se puede proteger a través de patentes y otros derechos de propiedad intelectual. El conocimiento tácito está relacionado a las capacidades específicas de los individuos y es el resultado de las características sociales e institucionales de un lugar. El conocimiento tácito se transmite mejor a través de la interacción personal y, en general, es difícil su difusión a distancia (Feldman y Kogler, 2010).

El flujo de la información dependerá del tipo conocimiento, además, el capital humano también influirá en la capacidad de absorción de la información nueva.¹ Por ejemplo, la transferencia de conocimiento tácito se facilita cuando hay confianza, cercanía cultural y cognitiva, incluyendo un lenguaje común y si se comparte el mismo campo científico (Karlsson et al., 2013).

La capacidad de absorber el conocimiento generado en una región también se ve afectada por las características de su actividad económica (Karlsson y Grasjo, 2014). La difusión de conocimientos puede facilitarse entre agentes que pertenecen a la misma industria, en un contexto de especialización. Por otro lado, tenemos el caso en que el flujo de conocimientos podría ocurrir entre agentes de diferentes industrias, en donde la diversidad es el motor de la innovación.

La cercanía geográfica de los agentes económicos favorece el flujo de conocimientos. Una implicación de la proximidad es que las transacciones y las interacciones entre los agentes son menos costosas y que la probabilidad de interactuar es mayor (Karlsson y Grasjo, 2014). Por otro lado, las empresas podrían optar por invertir en canales de comunicación y así aminsonar la relevancia de la proximidad geográfica. Sin embargo, la difusión del conocimiento ha mostrado que se dificulta a mayor complejidad del conocimiento, por lo que se podría considerar al conocimiento como espacialmente semifijo (*spatially sticky*) (Johansson, 2014).

Entre los factores que pueden influir en el flujo de los *spillovers* tenemos la distancia geográfica y las diferencias de los campos de investigación. También es necesario considerar que la diferencia de conocimientos entre los agentes no sea muy amplia para que se puedan aprovechar los *spillovers*. El conocimiento sólo se puede difundir si las partes involucradas exhiben una distancia cognitiva óptima porque sólo entonces es posible absorber e implementar conocimientos externos que resulten en cambio tecnológico e innovaciones (Cohen y Levinthal, 1990).

¹Feldman y Kogler (2010) definen capital humano como individuos que son capaces de comprender, integrar y crear nuevo conocimiento.

2.2. Antecedentes empíricos de la estimación de los *spillovers*

El estudio de los *spillovers* de conocimientos es un tema difícil de abordar; Krugman (1991) veía como obstáculo que no existiera un rastro en papel de los flujos de conocimientos que se pudiera seguir y medir. Si definimos los *spillovers* como el monto no apropiado de conocimientos de una empresa que puede utilizar otra, sin pago de por medio, la forma más sencilla de aproximar este tipo de *spillovers* es sumar el gasto en I&D de las otras empresas, es decir, el monto de I&D indirecto disponible para la empresa (Aiello y Cardamone, 2008). Esta medida no considera que las empresas se enfocan a diferentes actividades con diversos procesos tecnológicos, que su capacidad de absorber tecnología externa es heterogénea y que la distancia espacial también puede afectar los flujos de conocimientos.

Para refinar la forma en que se miden los *spillovers* se ha utilizado un sistema de ponderadores que incorporan la proximidad tecnológica entre las empresas, la capacidad de asimilación de tecnología externa y la distancia geográfica entre ellas. Kaiser (2002) clasifica en tres rubros los ponderadores según la (i) proximidad en el espacio tecnológico, (ii) proximidad geográfica y (iii) medidas directas calculadas a partir de las encuestas de innovación.

Para estimar los *spillovers* considerando la proximidad en el espacio tecnológico (i), Jaffe (1986, 1988) propuso aproximar el flujo de conocimientos con información sobre patentes. El flujo de conocimientos entre dos industrias son proporcionales a la participación de patentes de una industria en el área interés de la otra industria. Con el vector correspondiente a cada empresa se estima una correlación *uncentered* que mide la distancia tecnológica entre dos empresas. El ponderador toma valor de uno cuando los vectores de dos empresas son iguales, es decir, son cercanas en términos tecnológicos y de cero si no comparten ninguna característica.²

Kaiser (2002) utiliza un vector de características tecnológicas de la empresa para estimar la proximidad tecnológica que incluyen variables como la proporción de personal no calificado,

²Griliches (1990) señala que utilizar patentes podría estar generando omisiones debido a que no todas las invenciones se patentan y la calidad de las patentes varía en gran medida.

con calificación media y alta, respecto al total de empleados; el gasto en educación continua y capacitación por empleado; el costo laboral por empleado; la inversión y factores que obstaculizan la innovación de las empresas.

Dentro de los ponderadores que consideran la proximidad tecnológica (i) entre las empresas destaca la propuesta de Adams (1990). Este autor se basa en los trabajos de Jaffe (1986, 1988) y en lugar de utilizar patentes usa la proporción de científicos en cada campo de la ciencia. El objetivo es incorporar al análisis las diferencias de capital humano entre las empresas. Si dos empresas emplean una proporción alta de científicos, se esperaría que tuvieran más capacidad para aprovechar el conocimiento agregado de la economía.

La dimensión espacial se incorpora al análisis de los *spillovers* a través de una medida de coincidencia espacial que identifica la presencia de laboratorios universitarios e industriales en una misma región (Jaffe, 1989). Con estos indicadores se encuentra evidencia débil de la presencia de *spillovers* localizados. Acs et al. (1991) replican el trabajo de Jaffe (1989) utilizando el número de innovaciones como variable dependiente, un concepto más amplio para aproximar la producción de conocimientos, y encuentran evidencia de que los *spillovers* son más intensos cuando existe proximidad espacial.

Para la estimación de ponderadores que incluyan el aspecto espacial (ii) se ha utilizado la ubicación de las empresas y se estima la distancia entre ellas. Beise y Stahl (1999) utilizan la inversa de la distancia geográfica para calcular el ponderador.

Por último, las medidas directas calculadas a partir de las encuestas de innovación (iii), aprovechan la información sobre la percepción de las empresas respecto a los obstáculos a la innovación (Inkman, 2000). Para estimar este tipo de ponderador se hace un promedio de las respuestas de las empresas. En este rubro podemos incluir las fuentes de información para la innovación que se utilizan como *proxies* de los *spillovers* en la literatura empírica del modelo CDM (Crespi y Zuñiga (2012); Griffith et al. (2006); Calderón-Madrid (2012), entre otros).

El análisis de los *spillovers* se ha realizado en diferentes contextos lo que dificulta conocer la eficiencia de las diferentes formas de estimarlos. Kaiser (2002) realiza una comparación de los *spillovers* y señala que los más adecuados son aquéllos que consideran la proximidad

en el espacio tecnológico, seguidos por los que incorporan el aspecto espacial. En el trabajo de Aiello y Cardamone (2005, 2008) se utilizan dichas aproximaciones de los *spillovers* en el contexto de la función de producción y encuentran que los *spillovers* afectan positivamente la producción de las empresas y que la proximidad espacial si tiene relevancia.

2.3. Sistema de ponderadores para la estimación de los *spillovers* de conocimientos

Una empresa que gasta en I&D no puede conservar por completo la exclusividad de los conocimientos que produce, debido a las características inherentes del conocimiento. Una primera medida de los *spillovers* para la empresa i es el monto de I&D indirecta disponible:³

$$SPILL_{0,i} = \sum_{j=1, j \neq i}^N ID_j \quad i, j = 1, 2, \dots, N \quad (2.1)$$

donde, las empresas están indexadas por i y j , ID_j corresponde a la inversión en I&D de la empresa j . $SPILL_{0,i}$ es la suma del gasto en I&D de todas las empresas diferentes a i . Existen dos problemas cuando se estima $SPILL_0$: se suponen que todas las externalidades tecnológicas son relevantes para la empresa y que la capacidad de asimilar la tecnología no difiere entre empresas y sectores.

Para incorporar las diferencias tecnológicas y, además, considerar la dimensión espacial, se utiliza un sistema de ponderadores que definiremos a continuación (Kaiser, 2002; Aiello y Cardamone, 2008):⁴

a) Proximidad tecnológica de las empresas Se supone que las empresas que comparten el mismo tipo de tecnología pueden, con mayor facilidad, apropiarse de los flujos de conocimientos generados a partir de la I&D de otras empresas. Una forma de medir la cercanía tecnológica es con la siguiente correlación *uncentered*:

³La revisión de las medidas de los *spillovers* se basa en los trabajos de Aiello y Cardamone (2008) y Kaiser (2002).

⁴El supuesto base es que a mayor proximidad entre dos empresas, en términos tecnológicos y espaciales, la posibilidad de aprovechar los flujos de conocimientos es mayor.

$$\omega_i = \frac{X_i X_j'}{((X_i X_i')(X_j X_j'))^{1/2}} \quad (2.2)$$

donde, X es un conjunto de variables que definen el espacio tecnológico de las empresas (i, j) .⁵ El ponderador definido en la Ecuación (2.2) supone que la capacidad de absorción no varía en gran medida entre las empresas. Por lo que ω_i es una matriz simétrica que contiene la correlación de los vectores de características tecnológica de las empresas i y j , que es igual a la correlación de j con i .

Existe evidencia de que la dirección del flujo de conocimientos, es decir, de i a j o viceversa, afecta la forma en que se transmite la tecnología, debido a que cada empresa puede tener diferentes capacidades para absorber o asimilar los conocimientos externos (Kaiser, 2002).

b) Proximidad tecnológica y capacidad de asimilación de las empresas La heterogeneidad de las empresas hace que la absorción y el aprovechamiento del conocimiento externo difiera entre ellas. En la estimación de los *spillovers* se deben considerar estos factores y para ello se tiene el siguiente ponderador:

$$\hat{\omega}_{ij} = \left[\frac{X_i X_j'}{((X_i X_i')(X_j X_j'))^{1/2}} \right] \left[\frac{h_i}{\max(h_i, h_j)} \right] \quad (2.3)$$

$$\hat{\omega}_{ji} = \left[\frac{X_i X_j'}{((X_i X_i')(X_j X_j'))^{1/2}} \right] \left[\frac{h_j}{\max(h_i, h_j)} \right] \quad (2.4)$$

donde, h es una medida del capital humano. La idea es construir una correlación asimétrica *uncentered* ($\hat{\omega}_{ij} \neq \hat{\omega}_{ji}$) utilizando el capital humano, de tal forma que la capacidad de asimilación depende de la calidad del capital humano empleado por las empresas. El ponderador ($\hat{\omega}_{ij}$) considera que el flujo de conocimientos de la empresa i a j se ve afectado por la calificación del personal ocupado y viceversa.

c) Proximidad geográfica de las empresas Otra forma de ponderar la difusión del conocimiento entre las empresas es tomando en cuenta la distancia geográfica entre ellas.

⁵Si las empresas realizan investigación en áreas semejantes a otras empresas o producen patentes similares, se dice que son cercanas en el espacio tecnológico (Kaiser, 2002).

Para cada par de empresas i y j , se calcula la distancia entre ellas (d_{ij}) considerando como punto de referencia la capital administrativa de la entidad donde la empresa opera. El ponderador de proximidad geográfica es:

$$g_{ij} = 1 - \left[\frac{d_{ij}}{\max(d_{ij})} \right] \quad (2.5)$$

donde, d_{ij} es la distancia entre dos empresas y $\max(d_{ij})$ es la distancia máxima entre todos los pares de empresas. A través de la fórmula del círculo máximo se calcula la distancia más corta entre dos puntos en una esfera.⁶

d) Proximidad tecnológica y espacial Dado que es probable que empresas cercanas espacial y tecnológicamente aprovechen, en mayor medida, los *spillovers* generados por otras empresas, Aiello y Cardamone (2008) proponen promediar los ponderadores $\hat{\omega}_{ij}$ y g_{ij} para generar una medida agregada:

$$\nu_{ij} = \frac{\hat{\omega}_{ij} + g_{ij}}{2} \quad (2.6)$$

donde, ν_{ij} es una matriz asimétrica que toma valores de cero cuando no hay cercanía espacial o proximidad tecnológica y tiende a uno cuando son muy cercanos. Para este ponderador se supone que el peso de los componentes, espacial y sectorial, es el mismo.⁷

De esta forma, con los ponderadores ω_i , $\hat{\omega}_{ij}$, g_{ij} y ν_{ij} se genera una medida de los *spillovers* multiplicando el gasto en I&D (ID_j) con el ponderador (por ejemplo, por ν_{ij}) y se suma sobre las empresas diferentes a la empresa i :

$$SPILL_{\nu,i} = \sum_{j=1, j \neq i}^N \nu_{ij} ID_j \quad i, j = 1, 2, \dots, N \quad (2.7)$$

En la Ecuación (2.7), el gasto en I&D de otras empresas ($i \neq j$) está ponderado por las diferencias sectoriales, la distancia geográfica y la capacidad de asimilación de las empresas

⁶Dado que sólo conocemos la entidad donde se localizan las empresas, la distancia interna se define como: $d_{ii} = \frac{2}{3} \sqrt{\frac{A_i}{\pi}}$, donde A_i representa el área en kilómetros cuadrados de cada entidad federativa (Giacinto y Pagnini, 2011).

⁷Hasta el momento, no hay trabajos en donde se haya estimado la diferencia de los efectos de ambos ponderadores (Aiello y Cardamone, 2008).

($SPILL_\nu$). Con este ajuste se busca aproximar el flujo de conocimientos o los *spillovers* entre empresas, suponiendo que se facilitan entre similares, en términos tecnológicos y a mayor cercanía espacial.

La construcción de los diferentes ponderadores nos permitirá estimar el efecto de la similitud tecnológica de las empresas ($SPILL_\omega$) y de la capacidad de asimilación de tecnología ($SPILL_{\hat{\omega}}$) en la producción de conocimientos. Con la dimensión espacial se puede estimar el efecto de la cercanía geográfica ($SPILL_g$) y, con el último ponderador, el impacto combinado de la similitud tecnológica y la proximidad espacial ($SPILL_\nu$), en las innovaciones.

Capítulo 3

Estimación de los *spillovers* de conocimientos

3.1. Microdatos de la ESIDET 2012

Para estimar los *spillovers* de conocimientos, descritos en el Capítulo 2, ocupamos información sobre el gasto intramuros en I&D total, por tipo de investigación e indicadores que caractericen las diferencias tecnológicas entre las empresas. Al conocer la ubicación de las empresas se puede incorporar el aspecto espacial a los *spillovers* y la ESIDET, del año 2012, contiene la información que requerimos por entidad federativa, a nivel empresa.

La ESIDET 2012 captura 59 sectores de la actividad económica, entre servicios y manufactura, y es la primera encuesta de este tipo con representatividad a nivel entidad federativa. En este trabajo consideramos sólo el sector manufacturero y la unidad de observación son las empresas con 20 y más empleados. Las respuestas de las empresas hacen referencia a los años 2010 y 2011. Para procesar los datos a nivel empresa se trabajó en el Laboratorio de Análisis de Datos del INEGI, organismo que revisa y vigila el cumplimiento de la confidencialidad de la información.¹

¹Como parte de la limpieza de la base de datos eliminamos las empresas que reportan ingresos y no registraron empleados. La submuestra final está compuesta por 5,393 empresas que representan 11,679 de las empresas en la población total.

3.2. Estadísticas descriptivas

El gasto intramuros en I&D (ID_j) de las empresas es aquél que realizan las empresas en proyectos internos de I&D tecnológico y es la variable que utilizamos como gasto en I&D indirecto para estimar los *spillovers* (Ecuación (2.7)). Al no incluir el gasto extramuros, que se refiere a los pagos a terceros que realiza la empresa cuando ésta no conduce la investigación, sólo consideramos los *spillovers* puros, flujos de conocimientos generados por el gasto en I&D de las empresas, por los cuales no recibe compensación financiera.

Cuadro 3.1: Gastos en actividades relacionadas con la innovación y madurez tecnológica de las empresas (promedio)

Variable	Total	Innovadoras	No innovadoras
Gasto en I&D intramuros (miles de pesos)			
Investigación básica	56.37	357.83	16.08
Investigación aplicada	302.59	729.06	245.58
Desarrollo tecnológico	645.55	4,607.58	115.97
Total	1,004.52	5,694.47	377.64
Gasto en I&D extramuros (miles de pesos)			
Pagos al sector productivo	111.51	339.50	81.03
A instituciones de educación superior	22.82	114.01	10.63
Pago a otras instituciones	2.53	16.51	0.66
Gastos en servicios (miles de pesos)			
Científicos y tecnológicos	433.35	2,350.01	177.16
Madurez Tecnológica*			
Adquiere tecnología	3.21	2.74	3.27
Compra maquinaria y equipo	2.86	2.40	2.92
Asimila la tecnología	2.88	2.27	2.97
Adapta y modifica tecnología	3.03	2.31	3.13
Genera o desarrolla tecnología	3.43	2.63	3.54
Patenta	3.68	3.23	3.75
Vende tecnología	3.92	3.81	3.93
N	11,679	1,377	10,302

*Promedio de las respuestas de las empresas, donde uno indica que las empresas siempre realizan esa actividad y cuatro que casi nunca la hacen.

Para aproximar la similitud tecnológica (vector X , Ecuaciones (2.2) y (2.3)) entre las empresas se utiliza el gasto por tipo de investigación, es decir, si es básica, aplicada o se

dedica al desarrollo tecnológico.² También se incorpora el gasto extramuros según el sector que recibe el pago y el gasto por servicios científicos y tecnológicos, con la finalidad de capturar las diferencias tecnológicas entre las empresas. Para complementar el vector de características tecnológicas consideramos la madurez tecnológica de las empresas.

La madurez tecnológica está definida como la frecuencia con que la empresa compra tecnología y no realiza modificaciones, asimila la tecnología que adquiere, adapta o modifica la tecnología, genera o desarrolla tecnología, patenta o vende tecnología. Las empresas contestan uno si siempre realizan estas actividades, dos si lo hacen con frecuencia, tres si sólo a veces y cuatro si nunca hacen dichas actividades.

Otra característica deseable de los *spillovers* es que diferencien entre la capacidad de asimilación de tecnología de las empresas. Con la ESIDET se puede diferenciar dicha capacidad utilizando la escolaridad de los trabajadores (h) dedicados a la I&D.

En el Cuadro 3.1 presentamos la media de las variables utilizadas para construir los *spillovers*, distinguiendo por el estatus de innovación de las empresas. Donde, las empresas innovadoras son aquellas que han introducido al mercado productos o procesos nuevos o significativamente mejorados (INEGI, 2015). Casi el 12 por ciento de las empresas son innovadoras y son las que presentan un mayor nivel de gasto en I&D. El desarrollo tecnológico destaca entre los gastos intramuros y el pago al sector productivo es el que mayor peso tiene entre los gastos extramuros.

Las actividades definidas como madurez tecnológica de las empresas, si se realizan frecuentemente generan un promedio bajo.³ Para las empresas innovadoras la actividad que, en promedio, más practican es asimilar tecnología, y no queda lejos de ese nivel lo que declaran las no-innovadoras como actividad más frecuente. La actividad que menos se realiza, en general, es la venta de tecnología.

²Las variables que componen X se estandarizan respecto a la media del sector, por entidad federativa. La descripción de las variables se presenta en el Cuadro B.1 del Apéndice.

³Debido a que si la empresa indica uno es porque frecuentemente realiza la actividad relacionada con la madurez tecnológica.

3.3. *Spillovers* de conocimientos

En el Cuadro 3.2 se presenta el promedio de las estimaciones de los diferentes *spillovers*, que es la suma de los gastos ponderados de las empresas diferentes a i . Además, dado que el conocimiento puede fluir entre diferentes sectores o dentro de cada sector estimamos indicadores intrasectoriales e intersectoriales. Para calcular los *spillovers* intrasectoriales se consideran nueve sectores de la industria manufacturera.⁴ Los *spillovers* intersectoriales se estiman entre todas las empresas dentro de cada entidad federativa. Para Nuevo León (NL) y el Estado de México (Edo. Méx.) no se estiman los *spillovers* que incluyen la distancia geográfica porque no contamos con desagregación espacial de las variables al interior de las entidades federativas.

La estimación de los *spillovers* sin ponderador ($SPILL_0$) es el gasto en I&D agregado menos el de la empresa en cuestión. La magnitud de los *spillovers* depende del nivel de gasto de la empresa, produciendo una relación inversa entre el gasto en I&D y los *spillovers*, es decir, entre más gasta la empresa menor es el nivel de los *spillovers*. Al utilizar los ponderadores, en la estimación de los *spillovers*, la relación entre el nivel de gasto de la empresa y los *spillovers* es positivo porque va a depender de la similitud tecnológica entre las empresas y de su cercanía geográfica.

Al consideramos la proximidad tecnológicas, la diferencia entre las empresas ($SPILL_\omega$) según su estatus de innovación refleja que las innovadoras tienen mayores *spillovers* de conocimientos (Cuadro 3.2). Si además de las similitudes tecnológicas se incluye la capacidad de absorción ($SPILL_{\hat{\omega}}$) se acentúa la relación positiva.

Los *spillovers* espaciales y tecnológicos ($SPILL_g$ y $SPILL_\nu$) intrasectoriales sobresalen por tener cifras grandes y esto se debe a que la interacción espacial aumenta la magnitud de los flujos de conocimientos.

En el aspecto regional, observamos un mayor nivel de *spillovers* en NL, respecto al Edo. Méx. y al promedio nacional. El Edo. Méx. queda por debajo de los promedios nacionales, en la mayoría de los casos (Cuadro 3.2).

⁴Los sectores están en el Cuadro B.1 del Apéndice

Cuadro 3.2: Promedio de los diferentes tipos de *spillovers* (en miles de pesos)

Tipo de <i>spillovers</i> ^a	Intersectoriales			Intrasectoriales		
	Total	Innovadoras	No-innovadoras	Total	Innovadoras	No-innovadoras
Nacional						
$SPILL_0$	1,489	1,253	1,520	103	115	101
$SPILL_\omega$	432	400	436	527	707	503
$SPILL_{\hat{\omega}}$	38	196	17	71	372	31
$SPILL_g^b$	-	-	-	33,800	41,200	32,800
$SPILL_\nu^b$	-	-	-	16,900	20,600	16,400
N	11,679	1,377	10,302	11,679	1,377	10,302
Estado de México						
$SPILL_0$	1,241	1,236	1,242	71	79	70
$SPILL_\omega$	221	305	209	39	50	37
$SPILL_{\hat{\omega}}$	28	167	7	5	28	2
N	1,488	192	1,296	1,488	192	1,296
Nuevo León						
$SPILL_0$	2,086	2,078	2,087	251	143	268
$SPILL_\omega$	440	586	417	61	83	58
$SPILL_{\hat{\omega}}$	59	298	20	9	46	3
N	1,149	159	990	1,149	159	990

^a Los *spillovers* que no utilizan ponderador son $SPILL_0$, aquéllos que sólo incorporan la similitud tecnológica son $SPILL_\omega$ y los que además incluyen la capacidad de asimilación tecnológica de la empresa son $SPILL_{\hat{\omega}}$, aquéllos que incorporan el aspecto espacial son $SPILL_g$ y los que combinan la proximidad geográfica y tecnológica son $SPILL_\nu$.

^b En millones de pesos

Capítulo 4

El modelo CDM

4.1. Intensidad del gasto en I&D

En el modelo CDM primero se estima la intensidad del gasto en I&D (IID_i), que es uno de los insumos para la producción de conocimientos. Las empresas deciden gastar o no I&D y, si es afirmativa la respuesta, eligen el monto a gastar en I&D. Para estimar la intensidad del gasto en I&D se utiliza un modelo Tobit generalizado o de selección muestral:¹

Ecuación principal:

$$\begin{aligned} IID_i = & \beta_1 FO_i + \beta_2 EXD_i + \beta_3 CO_i + \beta_4 FIN_i + \beta_5 INFO_1 \\ & + \beta_6 INFO_2 + \beta_7 INFO_3 + u_i \end{aligned} \quad (4.1)$$

Ecuación de selección:

$$\begin{aligned} SID_i = & 1[\delta_1 FO_i + \delta_2 EXD_i + \delta_3 EM_i + \delta_4 CON + \delta_5 OBJ \\ & + \delta_6 OBST_1 + \delta_7 OBST_2 + \delta_8 OBST_3 + v_i > 0] \end{aligned} \quad (4.2)$$

donde, IID_i es la intensidad del gasto en I&D de las empresas, se aproxima con el logaritmo del gasto en I&D más el gasto en innovación y servicios científicos y tecnológicos por empleado.²

¹La descripción de las variables está en el Cuadro B.1 del Apéndice. En el Apéndice A se detalla el modelo Tobit generalizado.

²En países en desarrollo la inversión en I&D es baja y para capturar el efecto total del gasto en este rubro

El efecto del mercado externo se incorpora a través de la participación de capital extranjero en las empresas (FO) y las exportaciones (EXD), divididas en cuatro categorías que van de más de cero y aumentan 25 por ciento hasta llegar a 100. Dichas variables están presentes en la ecuación principal y en la de selección.

En la ecuación de intensidad del gasto en I&D se incluye la colaboración con otras empresas, institutos o universidades (CO) y si tiene acceso al financiamiento público (FIN).

En las primeras regresiones consideramos las fuentes de información de mercado ($INFO_1$), científicas ($INFO_2$) y públicas ($INFO_3$) para la innovación (Modelo (a)). Las variables relacionadas con la información han sido utilizadas para aproximar los *spillovers* de conocimientos (Mairesse y Mohnen (2005), Griffith et al. (2006), Crespi y Zuñiga (2012), entre otros). Dado que el objetivo de este trabajo es estudiar el efecto directo de los *spillovers* sobre la producción de innovaciones, en una segunda ronda de regresiones eliminamos este grupo de variables (Modelo (b)) para comparar los resultados.

En la ecuación de selección, la variable dependiente SID_i toma valor de uno cuando la empresa realizó algún proyecto de I&D tecnológico extramuros o intramuros, reportó gasto en innovación o en servicios científicos y tecnológicos, y cero de otra forma. Las variables utilizadas como explicativas incluyen condiciones de la demanda a través de la variable OBJ , que nos indica si para la empresa mantener la participación de mercado es un objetivo altamente significativo, y CON representa el grado de concentración económica por sector. También se incluyen algunos obstáculos para la innovación como costos elevados ($OBST_1$), los derivados de la legislación vigente ($OBST_2$) y la falta de financiamiento adecuado ($OBST_3$). Otra variable del modelo es el tamaño de la empresa (EM) que busca reflejar el acceso a financiamiento y economías de escala de las empresas (Mairesse y Mohnen, 2005).

4.1.1. Estadísticas descriptivas

Los datos necesarios para estimar la intensidad del gasto en I&D están en la ESIDET 2012 que contiene información detallada de las actividades de I&D y de innovación de las

se estima el gasto en I&D como la suma el gasto en I&D, en servicio científicos y tecnológicos, más el gasto de innovaciones (Crespi y Zuñiga, 2012). Se le llamará indistintamente gasto en I&D o gasto en innovación.

empresas, la base de microdatos está descrita en la sesión 3.1. En el Cuadro 4.1 presentamos las características de las empresas según su estatus de innovación, donde, las empresas innovadoras son aquellas que han introducido al mercado productos nuevos o significativamente mejorados para la empresa y para el mercado.

Cuadro 4.1: Estadísticas descriptivas de las características de las empresas

Variables	Total	Inn.	No-inn.	Variables	Total	Inn.	No-inn.
Innovaciones	0.12	1.00	0.00	Capital extranjero, colaboración, financiamiento y concentración			
Gasto en I&D	1,879	8,890	941	<i>FO</i>	0.14	0.16	0.13
<i>SID</i>	0.20	0.89	0.11	<i>CO</i>	0.03	0.29	0.00
Tamaño de la empresa				<i>FIN</i>	0.04	0.23	0.01
<i>EM</i>	233	447	205	<i>CON</i>	0.32	0.35	0.32
<i>EM₁</i>	0.41	0.33	0.42	Fuentes de información			
<i>EM₂</i>	0.40	0.40	0.40	<i>INFO₁</i>	0.52	0.56	0.52
<i>EM₃</i>	0.07	0.11	0.07	<i>INFO₂</i>	0.44	0.47	0.43
<i>EM₄</i>	0.09	0.13	0.08	<i>INFO₃</i>	0.51	0.56	0.50
Exportaciones				Objetivos y obstáculos de la innovación			
<i>EXD</i>	0.32	0.45	0.30	<i>OBJ</i>	0.77	0.89	0.75
<i>EXD₁</i>	0.14	0.23	0.13	<i>OBST₁</i>	0.52	0.47	0.52
<i>EXD₂</i>	0.05	0.08	0.05	<i>OBST₂</i>	0.40	0.35	0.41
<i>EXD₃</i>	0.03	0.04	0.03	<i>OBST₃</i>	0.45	0.42	0.45
<i>EXD₄</i>	0.09	0.10	0.09				
N	11,679	1,377	10,302				

La descripción de las variables está en el Apéndice B.

El gasto en I&D de las empresas innovadoras supera el gasto promedio de las no-innovadoras. (Cuadro 4.1). El 20 por ciento de empresas realizan gastos en I&D e innovaciones (*SID*). Las empresas innovadoras tienen un tamaño (*EM*) promedio mayor que las no innovadoras, mientras, las empresas medianas y pequeñas (*EM₁* y *EM₂*) representan cerca del 70 por ciento de las empresas innovadoras (Cuadro 4.1).

En cuanto a la relación con el mercado externo, el 45 por ciento de las empresas innovadoras exportan (*EXD*) y sólo el 30 por ciento de las no-innovadoras exportan. La participación de capital extranjero (*FO*) en la empresa es muy similar entre los dos tipos de empresas y la

proporción de empresas con capital extranjero, mayor al 10 por ciento, es del 14 por ciento del total de empresas.

Sólo el cuatro por ciento de las empresas del país reciben financiamiento gubernamental (*FIN*) y alrededor del 20 por ciento de las empresas innovadoras se benefician de dichos recursos. En cuanto a la colaboración (*CO*) con otras instituciones para generar innovaciones, el 29 por ciento de las empresas innovadoras están involucradas en este tipo de convenios.

Las fuentes de información (*INFO*) se estiman como el promedio de respuestas de una escala de uno a cuatro, donde cuatro indica que es una fuente altamente significativa para la innovación.³ A mayor tamaño del indicador, más importancia tiene esa fuente de información para la empresa. Las empresas innovadoras siempre consideran a las fuentes de información como un recurso importante para la innovación, en menor medida lo hacen las no-innovadoras (Cuadro 4.1). La fuente de información de mayor peso para las empresas innovadoras son las de mercado (*INFO*₁), es decir, información proveniente de proveedores, clientes, competidores, consultorías y expertos. Después consideran importantes, las fuentes públicas (*INFO*₃), que incluye información de patentes, conferencias, seminarios, revistas especializadas, redes computarizadas de información, ferias y exposiciones industriales.

Para el 77 por ciento de las empresas mantener su participación de mercado es un objetivo de la innovación (*OBJ*) altamente significativo. Los obstáculos como los costos elevados de la innovación (*OBST*₁) o la falta de financiamiento (*OBST*₃) los consideran importantes poco más del 40 por ciento de las empresas innovadoras.

4.1.2. Estimación de la intensidad del gasto en I&D

El modelo de intensidad del gasto en I&D se compone de dos partes, en la primera se estima la decisión de invertir en I&D (ecuación de selección) y después el gasto en I&D (*IID*) (ecuación principal). En el Cuadro 4.2 presentamos los resultados, en la Columna (1) está la ecuación de selección que no cambia entre los diferentes modelos de intensidad del gasto en I&D. En las otras columnas se alternan variables dentro la ecuación principal, en la Columna (2) (Modelo (a)) incluimos las fuentes de información (*INFO*) y en las siguientes

³Para hacer el promedio se suman las respuestas y se divide entre el número máximo de respuesta.

se excluyen (Columnas (3-5)). En la Columna (4) incluimos los programas de financiamiento (Modelo (c)) y los mecanismos de financiamiento (Modelo (d)) están en la Columna (5).

Cuadro 4.2: Decisión e intensidad de gasto en I&D

Variables	Ecuación principal				
	Ec. de selección Var. dep.: <i>SID</i>	Var. dep.: <i>IID</i>			
		Modelo (a) con <i>INFO</i>	Modelo (b) sin <i>INFO</i>	Modelo (c) con <i>PFIN</i>	Modelo (d) con <i>MFIN</i>
(1)	(2)	(3)	(4)	(5)	
<i>FO</i> *	-0.031** (0.011)	0.471*** (0.141)	0.481*** (0.143)	0.329* (0.148)	0.576*** (0.143)
<i>EXD</i> ₁ *	0.071*** (0.012)	0.182 (0.110)	0.185 (0.110)	0.172 (0.113)	0.216* (0.108)
<i>EXD</i> ₂ *	0.055** (0.018)	0.091 (0.166)	0.087 (0.167)	0.199 (0.169)	0.177 (0.166)
<i>EXD</i> ₃ *	-0.002 (0.019)	0.601* (0.262)	0.597* (0.265)	0.718** (0.265)	0.612* (0.261)
<i>EXD</i> ₄ *	0.033* (0.017)	0.227 (0.162)	0.180 (0.162)	0.334 (0.174)	0.275 (0.162)
<i>CO</i> *	-	0.419*** (0.110)	0.470*** (0.110)	0.904*** (0.103)	0.339** (0.110)
<i>FIN</i> *	-	1.849*** (0.124)	1.909*** (0.122)	-	-
<i>INFO</i> ₁	-	0.053 (0.215)	-	-	-
<i>INFO</i> ₂	-	0.343* (0.148)	-	-	-
<i>INFO</i> ₃	-	0.366 (0.216)	-	-	-
<i>PFIN</i> ₁	-	-	-	1.200*** (0.130)	-
<i>PFIN</i> ₂	-	-	-	0.527** (0.174)	-
<i>PFIN</i> ₃	-	-	-	-1.795*** (0.238)	-
<i>MFIN</i> ₁	-	-	-	-	1.227*** (0.108)
<i>MFIN</i> ₂	-	-	-	-	1.944*** (0.121)
<i>EM</i> ₂ *	0.067*** (0.008)	-	-	-	-
<i>EM</i> ₃ *	0.123*** (0.018)	-	-	-	-

Continúa...

Modelo de intensidad del gasto en I&D... (Continuación)

Variables	Ecuación principal				
	Ec. de selección	Var. dep.: <i>IID</i>			
	Var. dep.: <i>SID</i>	Modelo (a) con <i>INFO</i>	Modelo (b) sin <i>INFO</i>	Modelo (c) con <i>PFIN</i>	Modelo (d) con <i>MFIN</i>
	(1)	(2)	(3)	(4)	(5)
EM_4^*	0.155*** (0.020)	-	-	-	-
<i>CON</i>	0.036 (0.024)	-	-	-	-
<i>OBJ</i> *	0.072*** (0.009)	-	-	-	-
$OBST_1^*$	-0.008 (0.007)	-	-	-	-
$OBST_2^*$	-0.019** (0.007)	-	-	-	-
$OBST_3^*$	0.007 (0.008)	-	-	-	-
<i>N</i>		11,679	11,679	11,679	11,679
<i>LL</i>		-10,424	-10,437	-10,496	-10,400
χ^2		933	913	731	1,009
ρ		0.619	0.587	0.677	0.627
χ_ρ^2		39	33	61	33
p_ρ		0.000	0.000	0.000	0.000
<i>AIC</i>		21,055	21,075	21,198	20,986
<i>BIC</i>		21,821	21,819	21,957	21,737

El asterisco en las variables indica que son dicotómicas. Los coeficientes de la ecuación de selección corresponden a los efectos marginales cuando *CON* está en la media, *OBJ* en uno y el resto de las variables están en cero. Los coeficientes de las ecuaciones principales corresponden a los efectos marginales cuando las variables *INFO* están en la media y el resto de las variables están en cero. En todos los casos incluimos variables *dummies* por sector y entidad federativa. La descripción de las variables está en Cuadro B.1 del Apéndice. Errores estándar robustos entre paréntesis. *** p<0.01, ** p<0.05, * p<0.1

En la literatura empírica, la mayoría de los modelos incluyen, en la ecuación de intensidad del gasto en I&D, variables que aproximan los *spillovers*, y de esta forma estiman su efecto indirecto sobre la producción de innovaciones. La propuesta que se hace en este trabajo es incorporar *spillovers*, que incluyan la proximidad tecnológica, capacidad de asimilación de tecnología y distancia espacial entre las empresas, en la ecuación de innovación y así capturar su efecto directo sobre la producción de conocimientos. La estimación de los Modelos (a) y (b) servirá para comparar los resultados y después con la estimación de la intensidad del gasto en I&D contrastamos las propensiones a innovar.

El modelo parte de un proceso de auto-selección a gastar en I&D, para comprobar si existe tal relación, se estima el coeficiente de correlación de los errores (ρ) de la ecuación principal y de selección. En el Cuadro 4.2, tenemos que la parte observada de la decisión de gastar en I&D está correlacionada con la parte no observada de la intensidad del gasto en I&D, es decir, se rechaza que la correlación entre los errores de las dos ecuaciones es estadísticamente igual a cero.

Con la ecuación de selección estimamos los efectos marginales que indican la probabilidad de gastar en I&D ante un cambio en las variables independientes (Cuadro 4.2, Columna (1)). Cada coeficiente mide el efecto parcial sobre la probabilidad de gastar en I&D manteniendo la concentración (*CON*) en la media, *OBJ* en uno y el resto de las variables, de la ecuación de selección, en cero.

La empresa base es una pequeña que no tiene relación con el mercado exterior, en un ambiente de concentración promedio y su objetivo al innovar es mantener su participación en el mercado. También se caracteriza por ser optimista dado que no considera significativos los obstáculos como el costo alto de innovar, la legislación vigente o la falta de financiamiento adecuado.

Una empresa mediana (EM_2) tiene una probabilidad siete por ciento mayor, que una empresa pequeña (EM_1), de gastar en I&D (Cuadro 4.2, Columna (1)).⁴ La probabilidad de gastar en I&D aumenta en 12 y 16 por ciento si consideramos a empresas grandes (EM_3 y EM_4) respecto a las pequeñas.

El impacto de ser una empresa exportadora (EXD_1 , EXD_2 y EXD_4) es positivo y significativo sobre la probabilidad de gastar en innovación respecto a las que no tienen relación con el mercado externo.

Una empresa pequeña, sin relación con el mercado externo, en un sector con concentración promedio y que no considera significativos los obstáculos a innovar, tiene una probabilidad de gastar en I&D mayor en siete por ciento, respecto a las empresas que no consideran importante el objetivo de mantener su participación en el mercado.

⁴En todas las interpretaciones de los coeficientes el resto de las variables se mantienen en los valores especificados en el Cuadro 4.2.

Por otro lado, una empresa que aprecia como obstáculo, para la innovación, la legislación vigente ($OBST_2$) reduce su probabilidad de invertir en I&D en cerca de dos por ciento, respecto a las que no consideran éste obstáculo.

Después de ajustar los coeficientes de la ecuación principal por la selección muestral, se obtiene el efecto de las variables sobre la esperanza del logaritmo del gasto en innovación cuando este se observa (Cuadro 4.2, Ecuación principal). Los coeficientes indican el efecto promedio sobre la intensidad del gasto en I&D (IID) cuando las fuentes de información ($INFO$) están en la media y el resto de las variables en cero.⁵

Si una empresa gasta en I&D y se convierte en una empresa con una participación de capital extranjero superior al 10 por ciento (FO), su gasto en I&D aumenta en alrededor de 60 por ciento. El efecto de las exportaciones sobre el monto de inversión en I&D no es significativo en la mayoría de los casos. Excepto el rango tres de exportaciones (EXD_3) que es significativo y positivo, lo que indica que el gasto en I&D es mayor cuando las empresas exportan entre el 50 y 70 por ciento de sus ventas totales.

El efecto marginal del financiamiento gubernamental (FIN) es superior a la unidad y nos indica una alta influencia de dicho factor para determinar el monto de gasto dirigido a la producción de innovaciones. Por otra parte, si una empresa que gasta en I&D, decide celebrar algún convenio de colaboración, aumentaría su gasto en I&D en poco más de 50 por ciento. La variable relacionada con las fuentes de información científicas ($INFO_2$) es significativa y presenta signo positivo.

Entre los Modelos (a) y (b) del Cuadro 4.2, no se puede elegir uno como el mejor debido a que las diferencias de su ajuste son pequeñas.⁶ Si consideramos el criterio de información de Akaike (AIC) el mejor es el Modelo (a), mientras que el criterio de información Bayesiano (BIC) indica evidencia débil de mejor ajuste del Modelo (b) sobre el Modelo (a). Con los estadísticos que se presentan no se puede confirmar que incluir las fuentes de información ($INFO$) que representan variables *proxies* de los *spillovers*, en la ecuación principal, mejoran el ajuste del modelo.

⁵La descripción de los efectos marginales está en el Apéndice A.

⁶Para realizar la comparación utilizamos el criterio de información de Akaike (AIC) y el Bayesiano (BIC), en donde, a menor es el criterio mejor el ajuste del modelo (Apéndice A).

En los resultados del Cuadro 4.2 (Columnas (2-3)), el coeficiente estimado de la variable relacionada con el financiamiento público (FIN) se caracteriza por ser significativo y tener un efecto grande sobre el monto destinado al gasto en I&D. Para explorar algunas variantes del Modelo (b) que excluye las fuentes de información, en las Columnas (4-5) se alternan variables relacionadas con los programas (Modelo (c)) y mecanismos (Modelo (d)) de financiamiento.⁷

La participación de capital extranjero (FO) en las empresas conserva su signo y significancia (Cuadro 4.2, Columnas (4-5)). Las exportaciones (EXD_3) que están en un rango entre 50 y 75 por ciento, son significativas. La colaboración (CO) con otras instituciones no cambia en signo y significancia respecto al Modelo (b).

El efecto de los programas de financiamiento público (Modelo (c)) es mixto.⁸ El financiamiento proveniente del CONACYT ($PFIN_1$) tiene un efecto positivo sobre el monto a gastar en I&D, también la influencia es positiva por parte de los programas de la Secretaría de Economía (SE) ($PFIN_2$). Por último, se presenta un efecto negativo de los programas de la Secretaría del Trabajo y Previsión Social (STPS), Banco Nacional de Comercio Exterior (BANCOMEXT) y STPS-Secretaría de Ecuación Pública (SEP) ($PFIN_3$).

Los mecanismos de financiamiento son positivos y significativos (Modelo (d), Cuadro 4.2, Columna (5)). Empresas que gastan en I&D y reciben financiamiento bancario ($MFIN_1$) aumentan el monto de inversión en I&D en más del doble, respecto a cuando no recibían el financiamiento. Empresas que ya habían decidido gastar en I&D y pasan de no recibir financiamiento público a recibirlo, aumentan su inversión en I&D.⁹

El Modelo (d) con mecanismos de financiamiento es el que mejor se ajusta a nivel nacional, comparando los criterios de información (AIC y BIC). En la siguiente etapa utilizamos estos resultados para estimar la función de producción de conocimientos o ecuación de innovación.

⁷La ecuación de selección no cambia, los resultados están en el Cuadro 4.2 Columna (1).

⁸La descripción de las variables está en el Cuadro B.1 del Apéndice.

⁹Los efectos promedio, ajustados por el proceso de selección, se estiman para empresas que no tienen participación en el mercado externo, pequeñas, en un sector con concentración promedio, sin colaboraciones, sin financiamiento, que no consideran relevantes los obstáculos a la innovación y que buscan mantener su participación de mercado.

4.2. Función de producción de conocimientos

La innovación se define como la introducción, al mercado, de un producto (bien o servicio) o proceso, nuevo o significativamente mejorado (INEGI, 2015). Algunas innovaciones son resultado de proyectos de innovación específicos, que incluyen I&D tecnológico como insumo, mientras que otras son resultado de mejoras rutinarias, ideas espontáneas u otros factores que llevan a las empresas a mejorar sustancialmente o desarrollar nuevos productos o procesos (INEGI, 2015). La segunda etapa del modelo CDM consiste en estimar la función de producción de conocimientos (ecuación de innovación) utilizando los resultados de la etapa anterior.¹⁰

En la literatura que usa el modelo CDM se ha abordado parcialmente la relación entre la producción de innovaciones y la presencia de *spillovers*. En este trabajo se utilizan variables *proxies* de los *spillovers* para determinar su influencia en la producción de innovaciones. La función de producción de conocimientos (ecuación de innovación) correspondiente a la segunda etapa del modelo CDM es la siguiente:

$$INN_i = \alpha_1 LSPILL_i + \alpha_2 I\hat{I}D_i + \alpha_3 FO_i + \alpha_4 EXD_i + \alpha_5 LEM_i + \epsilon_i \quad (4.3)$$

donde INN_i es la producción de innovaciones, variable que toma valor de uno si la empresa introdujo al mercado productos nuevos o significativamente mejorados y cero de otra forma. Las variables de control son la participación de capital extranjero (FO_i), las exportaciones (EXD_i) y el tamaño de la empresa en términos logarítmicos (LEM_i).¹¹

En los trabajos empíricos basados en el modelo CDM, tradicionalmente, se han utilizado las fuentes de información para la innovación como *proxies* de los *spillovers*, nosotros construimos variables *proxies* con el gasto en I&D ponderado por la cercanía tecnológica y espacial, además de la capacidad de asimilación de las empresas. A diferencia de la literatura

¹⁰Crepon et al. (1998) señala que la I&D es un factor endógeno en la producción de conocimientos así que con la primera etapa se corrige el problema de selección y de endogenidad, al sustituir el gasto observado por el estimado.

¹¹La descripción de variables está en el Cuadro B.1 del Apéndice.

existente, analizamos los efectos directos de los *spillovers* ($LSPILL_i$ es el logaritmo de los *spillovers*) sobre la producción de conocimientos.¹²

Estadísticas descriptivas

Los datos necesarios para estimar la ecuación de innovaciones están en la ESIDET 2012, descrita en la sesión 3.1. La media de las variables que se utilizan en la ecuación de producción de innovaciones están en los Cuadros 3.1 y 4.1. Para México, la propensión a innovar es de 12 por ciento. Sobresalen dos características de las empresas innovadoras. Primero, tienen un gasto promedio más alto en I&D que las empresas no-innovadoras y su nivel de empleo (EM) también es superior. Se considera un hecho estilizado que las empresas grandes sean las que más gastan en I&D, además, presentan un mayor nivel de innovación (Cohen y Klepper, 1996; Hall y Mairesse, 2006). Por lo que se espera que exista una relación positiva entre el gasto en I&D, el tamaño de las empresas y la propensión a innovar.

La relación del mercado extranjero con las empresas innovadoras y no-innovadoras, no es muy diferente, la proporción de empresas que exportan (EXD) es similar entre ellas (Cuadro 4.1). La participación de capital extranjero (FO) es ligeramente superior en las empresas innovadoras.

Los *spillovers* ($SPILL$) promedio de las empresas innovadoras son mayores respecto a las no-innovadoras, esto varía según el tipo de ponderador que se utilice para estimar los *spillovers* (Cuadro 3.1). Si consideramos que los *spillovers* son flujos de información que la empresa puede obtener sin que exista pago de por medio, esperaríamos que su efecto sobre la producción de innovaciones sea positivo.

4.2.1. Estimación de la propensión a innovar en productos

La estrategia a seguir es estimar una función de producción de conocimientos donde relacionamos la propensión a innovar en productos con el gasto en I&D de las empresas y

¹²Los *spillovers* en logaritmos que no utilizan ponderador son $LSPILL_0$, aquéllos que sólo incorporan la similitud tecnológica son $LSPILL_\omega$ y los que además incluyen la capacidad de asimilación tecnológica de la empresa son $LSPILL_{\hat{\omega}}$, aquéllos que incorporan el aspecto espacial son $LSPILL_g$ y los que combinan la proximidad geográfica y tecnológica son $LSPILL_\nu$.

los *spillovers* de conocimientos, que son las principales variables explicativas del modelo. La metodología utilizada corresponde a la segunda etapa del modelo CDM. Las regresiones realizadas fueron diferenciadas por tipo de *spillovers*, según el ponderador, y si son intra-sectoriales o intersectoriales. Los resultados con *spillovers* diferentes a $SPILL_{\hat{\omega}}$ están en el Apéndice C.

Entre las diferentes estimaciones de la ecuación de innovación sobresale el efecto de los *spillovers* ponderados por la proximidad tecnológica y la capacidad de asimilación de las empresas ($SPILL_{\hat{\omega}}$). Cuando se emplea esta variable, el modelo presenta un mejor ajuste y se observa una reducción en el coeficiente de la intensidad de gasto en I&D ($I\hat{I}D$).

Si en la estimación de los *spillovers* no se considera la capacidad de asimilación de las empresas, se sobrestima el efecto de los *spillovers* ($SPILL_{\omega}$ vs $SPILL_{\hat{\omega}}$) esto se puede ver al comparar los coeficientes estandarizados de las ecuaciones que los contienen y el mismo efecto se observa sobre el coeficiente de la intensidad del gasto en I&D.

En cuanto a las diferencias intrasectoriales e intersectoriales, encontramos que los modelos con *spillovers* intersectoriales tienen un mejor ajuste. Entonces, al considerar el flujo de conocimientos entre todos los sectores de la economía se obtiene una mejor estimación de la ecuación de innovación.

Para simplificar el análisis de los resultados nos concentramos en los *spillovers* ponderados con la proximidad tecnológica y la capacidad de asimilación de las empresas a nivel intersectorial ($SPILL_{\hat{\omega}}$). La elección de este indicador se hizo en función de los criterios de información (AIC y BIC) de las diferentes especificaciones de la ecuación de innovación (Cuadros C.2 y C.5 del Apéndice).

Relevancia de los *spillovers* en la ecuación de innovaciones

Cuando se incluyen *spillovers* en la ecuación de innovación, las regresiones tienen un mejor ajuste. En el Cuadro 4.3, los estadísticos AIC y BIC son menores cuando se utiliza la variable *spillovers*. El coeficiente estimado de los *spillovers* tiene signo positivo y es significativo. La probabilidad de innovar aumenta 12 por ciento cuando los *spillovers* aumentan en uno por ciento.¹³

¹³Efectos marginales estimados con $FO = 0$, $EXD = 0$ y el resto de las variables en la media.

Cuadro 4.3: Propensión a innovar en productos con $LSPILL_{\hat{\omega}}$

Variable Dependiente: Innovación en productos (INN_1)				
Variables	$\hat{I}\hat{D}$ del Modelo (a)		$\hat{I}\hat{D}$ del Modelo (b)	
	(1)	(2)	(3)	(4)
$\hat{I}\hat{D}$	0.680*** (0.029)	0.405*** (0.034)	0.811*** (0.032)	0.512*** (0.038)
	2.593	1.544	2.946	1.861
$LSPILL_{\hat{\omega}}^a$	-	0.124*** (0.005)	-	0.119*** (0.005)
		1.331		1.283
FO	-0.239*** (0.059)	-0.133* (0.063)	-0.291*** (0.059)	-0.169** (0.063)
	-0.273	-0.152	-0.333	-0.194
EXD	-0.086 (0.044)	0.027 (0.047)	-0.130** (0.045)	-0.012 (0.048)
	-0.133	0.042	-0.201	-0.018
LEM	0.084*** (0.016)	0.024* (0.017)	0.093*** (0.016)	0.031* (0.017)
	0.324	0.092	0.358	0.118
N	11,644	11,644	11,644	11,644
R_p^2	0.142	0.231	0.159	0.239
LL	-3,261	-2,921	-3,197	-2,892
χ^2	1,012	1,445	1,101	1,408
AIC	6,607	5,929	6,477	5,869
BIC	6,916	6,245	6,786	6,186
P_{corr}	90.41	91.13	90.47	91.33

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (*beta*). En todos los casos se controla con *dummies* por sector y por entidad federativa. El asterisco en las variables indica que son variables *dummy*. La descripción de las variables está en el Cuadro B.1 del Apéndice.

^a Los *spillovers* son intersectoriales.

*** p<0.01, ** p<0.05, * p<0.1

El coeficiente estimado de la intensidad del gasto en I&D ($I\hat{I}D$) disminuye si se incluye la medida de los *spillovers* que considera la proximidad tecnológica y la capacidad de asimilación de tecnología de las empresas. Cuando no se consideran los *spillovers*, el incremento de uno por ciento del gasto en I&D ($I\hat{I}D$) aumenta la probabilidad de innovar en 81 por ciento (Cuadro 4.3, Columna (3)). Si se incorpora el efecto directo de los *spillovers* en la ecuación de innovación, un incremento de uno por ciento del gasto en I&D aumenta la probabilidad de innovar en 51 por ciento (Columna (4)). El coeficiente estandarizado muestra que el efecto de $I\hat{I}D$ es más alto cuando se excluyen los *spillovers* de la ecuación de innovación. A través de la metodología de Karlson et al. (2012), que permite comparar entre modelos probit anidados, se confirma que la $I\hat{I}D$ está sobreestimada en el modelo que no incluye los *spillovers*.¹⁴

A partir de la estimación de la función de conocimientos obtenemos dos resultados importantes. Primero los *spillovers* son relevantes para la producción de innovaciones. El segundo es que al incluirlos en el modelo se observa que hay una sobreestimación del efecto del gasto en I&D.

Efecto directo de los *spillovers* sobre la producción de innovaciones

La presencia de *spillovers* afecta la producción de innovaciones y si se consideran en esta etapa del modelo CDM su ajuste mejora. En la sección anterior no se pudo determinar claramente qué modelo de la intensidad del gasto en I&D es el que tiene un mejor ajuste, por lo que estimamos la ecuación de innovaciones con los dos modelos.

A partir de las estimaciones de la IID con los Modelos (a) y (b) se comparan los resultados de la ecuación de innovación (Cuadro 4.3). El objetivo es analizar el impacto directo de los *spillovers* sobre la producción de innovaciones (Modelo (b)) y el efecto de incluir variables que aproximen a los *spillovers* en la ecuación de intensidad del gasto en I&D (Modelo (a)).

El coeficiente de los *spillovers*, cuando se estima IID con el Modelo (a) y (b), es positivo y significativo. En el Modelo (a), el crecimiento de uno por ciento de los *spillovers*, aumenta

¹⁴El modelo de Karlson et al. (2012) propone comparar un modelo probit restringido con el completo considerando la correlación de la variable nueva y la existente, además, del efecto de la variable nueva en la variable dependiente, de tal forma que los coeficientes de los modelos probit sean comparables.

la probabilidad de innovar en productos, 12 por ciento, similar al resultado obtenido con el Modelo (b), sus coeficientes estandarizados también son cercanos.¹⁵

Si sólo se consideran los *spillovers* en la segunda etapa del CDM, la probabilidad de innovar aumenta en 51 por ciento, cuando crece en uno por ciento el gasto en I&D (Columna (4), Cuadro 4.3). El efecto es mayor al encontrado cuando se incluyen los *spillovers* en las dos etapas (Columna (2), Cuadro 4.3).

En los resultados de la ecuación de innovación con las estimaciones del Modelo (b), también se observa que las variables de la relación con el mercado extranjero tienen efectos negativos. Mientras que el tamaño de las empresas tiene una relación positiva con la producción de innovaciones.

Los criterios de información *AIC* y *BIC* indican que utilizando el Modelo (b) para estimar el gasto en I&D y al incluir la variable de *spillovers*, se alcanza un mejor ajuste del modelo. Aunque en la etapa anterior del CDM no se pudo elegir entre los Modelos (a) y (b), la estimación derivada del Modelo (b) contribuye a un mejor ajuste de la ecuación de innovaciones. Con los resultados obtenidos se encuentra respaldo para utilizar los *spillovers* en la segunda etapa del CDM.

Dimensión espacial de los *spillovers*

En la revisión de literatura empírica relacionada con el efecto de los *spillovers* sobre la producción de innovaciones, se encuentra que los *spillovers* están localizados y tienen un efecto decreciente con la distancia geográfica. En las variables $SPILL_g$ y $SPILL_v$ incorporamos el aspecto espacial para estimar su influencia en la producción de innovaciones.

Los coeficientes estimados no son significativos y tampoco tienen el signo esperado.¹⁶ Por lo que, no se encuentra efecto de la dimensión espacial sobre la producción de innovaciones. Una posible explicación de estos resultados es que la ESIDET tiene como unidad de observación las empresas, lo que puede estar sesgando los efectos espaciales, debido a que no conocemos la ubicación de los establecimientos.

¹⁵El efecto sobre la probabilidad corresponde a empresas sin relación con el mercado extranjero, con nivel promedio del gasto en I&D, de los *spillovers* y del tamaño de la empresa.

¹⁶Las estimaciones están en el Cuadro C.5 del Apéndice.

Efecto directo de los *spillovers* sobre las innovaciones en productos, productos a nivel mercado y procesos

El efecto de los *spillovers*, ajustados por la proximidad tecnológica y la capacidad de asimilación de las empresas, sobre la producción de innovaciones está en el Cuadro C.8, del Apéndice. Para las funciones de producción de conocimiento que utilizan otras variables dependientes como la innovación en productos a nivel nacional y mundial (INN_2) y la innovación en procesos (INN_3), empleamos el mejor modelo de intensidad del gasto en I&D para la estimación de IID , en este caso es el Modelo (d) donde se incluyen los mecanismos de financiamiento.

La probabilidad de innovar en productos crece en 11 por ciento, cuando aumenta en uno por ciento los *spillovers*. El efecto sobre la probabilidad de innovar en productos para el mercado, nacional y mundial, no difiere del resultado anterior. La probabilidad de innovar en procesos aumenta en seis por ciento, cuando aumenta en uno por ciento los *spillovers*. Al utilizar otras variables que aproximen la producción de innovaciones, se conservan los resultados, es decir, los *spillovers* son relevantes y positivos.

En el Cuadro C.8 del Apéndice también se presentan las regresiones sin *spillovers* y se observa que existe una sobreestimación del coeficiente de la intensidad del gasto en I&D. Cuando utilizamos otras variables que aproximen la innovación encontramos que se confirma el resultado de la sobreestimación de \hat{IID} cuando se omiten los *spillovers* de conocimientos.

El modelo CDM permite analizar la decisión de gastar en I&D, el monto a invertir en este rubro y el impacto sobre la producción de innovaciones. Los resultados indican que incorporar los *spillovers* al modelo CDM mejora las estimaciones de la función de producción de innovaciones. El efecto de los *spillovers* es positivo, es decir, a mayor flujo de conocimientos aumenta el nivel de innovaciones de las empresas. Además, excluirlos de la ecuación de innovaciones genera una sobreestimación del efecto del gasto en I&D.

Capítulo 5

El modelo CDM y su aplicación a nivel estatal: Nuevo León y el Estado de México

La ESIDET 2012 tiene representatividad estatal y es por esto que se puede aplicar el modelo CDM para los casos de Nuevo León (NL) y el Estado de México (Edo. Méx.).¹ Este ejercicio permitirá observar si existen diferencias respecto a los resultados nacionales.

5.1. Intensidad de gasto en I&D

En la primera etapa del modelo CDM se estima la intensidad del gasto en I&D (IID_i). Las empresas deciden gastar o no I&D y, si es afirmativa la respuesta, eligen el monto a gastar en I&D. Para estimar la intensidad del gasto en I&D utilizamos un modelo Tobit generalizado representado en las Ecuaciones (4.1) y (4.2).²

Al igual que en la estimación a nivel nacional, para NL y el Edo. Méx. realizamos regresiones con (Modelo (a)) y sin fuentes de información (Modelo (b)), variables que en la literatura se han utilizado para aproximar los *spillovers* (Cuadro 5.1).

¹Se eligieron estas dos entidades federativas debido a que son las que cuentan con muestras grandes que permiten un buen funcionamiento del modelo CDM.

²La descripción de las variables está en el Cuadro B.1 del Apéndice y en el Apéndice A se detalla el modelo Tobit generalizado o de selección muestral.

El modelo parte de un proceso de auto-selección a gastar en I&D, para comprobar si existe tal relación, se estima el coeficiente de correlación de los errores de la ecuación principal y de selección (ρ). En el Edo. Méx. y NL, la parte observada de la decisión de gastar en I&D está correlacionada con la parte no observada de la intensidad del gasto en I&D (Cuadro 5.1).

Cuadro 5.1: NL y Edo. Méx.: Decisión e intensidad de gasto en I&D

Variables	Estado de México			Nuevo León		
	Ec. de selección	Ecuación principal		Ec. de selección	Ecuación principal	
		Var. dep.: <i>IID</i>			Var. dep.: <i>IID</i>	
Var. dep.: <i>SID</i>	Modelo (a) con <i>INFO</i>	Modelo (b) sin <i>INFO</i>	Var. dep.: <i>SID</i>	Modelo (a) con <i>INFO</i>	Modelo (b) sin <i>INFO</i>	
(1)	(2)	(3)	(4)	(5)	(6)	
<i>FO</i> *	-0.033 (0.024)	0.489 (0.387)	0.305 (0.391)	-0.028 (0.034)	0.392 (0.508)	0.658 (0.388)
<i>EXD</i> ₁ *	-0.001 (0.023)	0.413 (0.282)	0.491 (0.273)	0.118** (0.037)	-0.307 (0.456)	-0.576* (0.282)
<i>EXD</i> ₂ *	0.015 (0.041)	1.182* (0.470)	1.455** (0.503)	0.124* (0.053)	0.222 (0.531)	-0.245 (0.370)
<i>EXD</i> ₃ *	0.001 (0.056)	-0.239 (0.727)	0.422 (0.626)	-0.037 (0.048)	-1.150 (0.701)	-1.522** (0.586)
<i>EXD</i> ₄ *	0.104 (0.074)	0.099 (0.663)	0.430 (0.649)	-0.069* (0.034)	0.568 (0.687)	-0.051 (0.619)
<i>CO</i> *	-	0.577 (0.350)	0.827* (0.353)	-	0.220 (0.312)	0.329 (0.308)
<i>FIN</i> *	-	1.287** (0.443)	0.688 (0.415)	-	1.856*** (0.535)	1.852*** (0.377)
<i>INFO</i> ₁	-	0.499 (0.800)	-	-	0.686 (0.745)	-
<i>INFO</i> ₂	-	1.133** (0.438)	-	-	0.756 (0.542)	-
<i>INFO</i> ₃	-	0.185 (0.621)	-	-	0.231 (1.145)	-
<i>EM</i> ₂ *	0.080** (0.025)	-	-	0.082*** (0.021)	-	-
<i>EM</i> ₃ *	0.134** (0.046)	-	-	0.162** (0.051)	-	-
<i>EM</i> ₄ *	0.251*** (0.055)	-	-	0.252*** (0.053)	-	-
<i>CON</i>	-0.010 (0.059)	-	-	0.039 (0.059)	-	-

Continua...

Modelo de intensidad del gasto en I&D... (Continuación)

Variables	Estado de México			Nuevo León		
	Ec. de selección	Ecuación principal		Ec. de selección	Ecuación principal	
		Var. dep.: <i>IID</i>			Var. dep.: <i>IID</i>	
Var. dep.: <i>SID</i>	Modelo (a) con <i>INFO</i>	Modelo (b) sin <i>INFO</i>	Var. dep.: <i>SID</i>	Modelo (a) con <i>INFO</i>	Modelo (b) sin <i>INFO</i>	
(1)	(2)	(3)	(4)	(5)	(6)	
<i>OBJ</i> *	0.092*** (0.019)	-	-	0.055** (0.020)	-	-
<i>OBST</i> ₁ *	-0.042** (0.016)	-	-	0.086** (0.028)	-	-
<i>OBST</i> ₂ *	0.013 (0.021)	-	-	-0.085*** (0.014)	-	-
<i>OBST</i> ₃ *	0.041 (0.024)	-	-	-0.048** (0.016)	-	-
<i>N</i>		1,488	1,488		1,149	1,149
<i>LL</i>		-1,094	-1,104		-1,032	-1,036
χ^2		183	144		179	158
ρ		0.570	0.575		0.967	0.926
χ^2_ρ		6	5		9	27
p_ρ		0.017	0.022		0.003	0.000
<i>AIC</i>		2,274	2,288		2,149	2,152
<i>BIC</i>		2,502	2,501		2,366	2,354

El asterisco en las variables indica que son dicotómicas. Los coeficientes de la ecuación de selección corresponden a los efectos marginales cuando *CON* está en la media, *OBJ* en uno y el resto de las variables están en cero. Los coeficientes de las ecuaciones principales corresponden a los efectos marginales cuando las variables *INFO* están en la media y el resto de las variables están en cero. En todos los casos se incluyen variables *dummies* por sector. La descripción de las variables está en el Cuadro B.1 del Apéndice. Errores estándar robustos entre paréntesis.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A partir de la ecuación de selección estimamos los efectos marginales que indican la probabilidad de gastar en I&D ante un cambio en las variables independientes (Cuadro 5.1). Los efectos marginales representan el efecto parcial sobre la probabilidad de gastar en I&D manteniendo la concentración (*CON*) en la media, *OBJ* en uno y el resto de las variables, de la ecuación de selección, en cero.³

La probabilidad de gastar en I&D aumenta conforme aumenta el tamaño de las empresas, es decir, a mayor tamaño respecto a las empresas pequeñas, mayor propensión a gastar en I&D, tanto en el Edo. Méx. como en NL.

³Las ecuaciones de selección no se modifican entre las diferentes estimaciones de la intensidad del gasto en I&D y sus efectos marginales se presentan en el Cuadro 5.1, Columna (1).

Por otro lado, contar con capital extranjero (FO) no resulta significativo para ninguna de las dos entidades. En el caso de las exportaciones sólo en NL se observa que existe un efecto positivo de tener una proporción de ingresos por ventas al mercado extranjero de más de cero hasta el 50 por ciento, respecto a las empresas no exportadoras.

Una empresa pequeña, sin relación con el mercado externo, en un sector con concentración promedio y que no considera significativos los obstáculos a innovar, tiene una probabilidad a gastar en I&D mayor en nueve y seis por ciento, respecto a las empresas que no consideran importante el objetivo (OBJ) de mantener su participación en el mercado, en el Edo. Méx. y NL respectivamente (Cuadro 5.1).

Si la empresa considera que los costos de innovar son elevados ($OBST_1$) su probabilidad de gastar, en el Edo. Méx., se reduce en cuatro por ciento. En NL, una empresa que aprecia como un obstáculo la legislación vigente ($OBST_2$) reduce su probabilidad de invertir en I&D en cerca de nueve por ciento, respecto a las que no consideran éste un obstáculo. Si la empresa percibe que la falta de financiamiento adecuado ($OBST_3$) es un obstáculo para la innovación, su probabilidad de gastar en I&D disminuye cinco por ciento, en NL.

Después de ajustar los coeficientes de la ecuación principal por la selección muestral, se obtiene el efecto de las variables sobre la esperanza del logaritmo del gasto en innovación cuando éste se observa (Cuadro 5.1). Los coeficientes indican el efecto promedio sobre la intensidad del gasto en I&D (IID) cuando las fuentes de información ($INFO$) están en la media y el resto de las variables en cero.⁴

La relación con el mercado extranjero no tiene impacto sobre la intensidad del gasto en I&D a nivel entidad federativa. Se observan algunas excepciones, en el caso del Edo. Méx. donde las empresas que exportan entre 25 y 50 (EXD_2) por ciento sí aumentan el gasto en I&D respecto a las empresas que no exportan. En el caso de NL se presentan algunos coeficientes significativos pero no con el signo esperado. De acuerdo con Crespi y Zuñiga (2012) las empresas multinacionales son conservadoras en cuanto al gasto en I&D en los países en desarrollo, ya que se enfocan en sus ventajas comparativas y en la adaptación de productos para ese mercado, y es lo que podría estar reflejando los coeficientes negativos.

⁴La descripción de los efectos marginales está en el Apéndice A.

El efecto marginal del financiamiento gubernamental (FIN) es superior a la unidad y nos indica una alta influencia de dicho factor para determinar el monto de gasto dirigido a la producción de innovaciones. Sólo en el Modelo (b) del Edo. Méx. la variable no resultó significativa. La colaboración (CO) con otras empresas y universidades sólo resulta significativa en el Edo. Méx (Modelo (b)).

Entre los Modelos (a) y (b) del Cuadro 5.1, no se puede elegir uno como el mejor debido a que las diferencias en los criterios de información AIC y BIC son pequeñas. Si consideramos el criterio de información de Akaike (AIC) el mejor es el Modelo (a) para las dos regiones. El criterio de información Bayesiano (BIC) indica evidencia débil de mejor ajuste del Modelo (b) sobre el Modelo (a) para el Edo. Méx y hay evidencia fuerte en favor del Modelo (b) en NL. Con los estadísticos que se presentan no se puede confirmar que incluir las fuentes de información ($INFO$) que representan variables *proxies* de los *spillovers*, en la ecuación principal, mejoren el ajuste del modelo. Los resultados no difieren en gran medida de lo encontrado a nivel nacional.

Para NL y el Edo. Méx. también estimamos el modelo de intensidad del gasto en I&D alternando variables relacionadas con los programas (Modelo (c)) y mecanismos (Modelo (d)) de financiamiento (Cuadro C.1 del Apéndice).

En el Edo. Méx. la influencia es positiva por parte de los programas de la SE ($PFIN_2$). Mientras que en NL sólo es positivo y significativo el efecto de los programas del CONACYT. En el caso de los otros programas, su coeficiente no tiene el signo esperado.

En cuanto a los mecanismos de financiamiento se observa que los mecanismos de financiamiento son positivos y significativos en NL (Modelo (d)). Empresas que gastan en I&D y reciben financiamiento bancario ($MFIN_1$) aumentan el monto de inversión en I&D. Empresas que ya habían decidido gastar en I&D y pasan de no recibir financiamiento público a recibirlo, aumentan su inversión destinada a la innovación más de cinco veces.⁵

⁵Los efectos promedio, ajustados por el proceso de selección, se estiman para empresas que no tienen participación en el mercado externo, pequeñas, en un sector con concentración promedio, sin colaboraciones, sin financiamiento, que no consideran relevantes los obstáculos a la innovación y que buscan mantener su participación de mercado.

Al comparar los criterios de información encontramos que el Modelo (d) es el que mejor ajuste presenta para NL y el Modelo (c) es el que mejor se ajusta para el Edo. Méx. (Cuadro C.1 del Apéndice).

5.2. Función de producción de conocimientos

En el Cuadro 5.2 presentamos los resultados de la ecuación de innovación que utiliza la estimación de la intensidad del gasto en I&D del Modelo (b). Para analizar la relevancia de estudiar el impacto directo de los *spillovers* en la función de producción de conocimientos nos concentramos en las regresiones sin *spillovers* y la mejor regresión con *spillovers* (Cuadro 5.2). Según los criterios de información *AIC* y *BIC*, los *spillovers* que mejoran el ajuste del modelo son los que consideran la proximidad tecnológica y la capacidad de asimilación de las empresas ($SPILL_{\omega}$) a nivel intersectorial.⁶

Para el Edo. Méx. y NL, cuando se incluyen *spillovers* en la ecuación de innovación, las regresiones tienen un mejor ajuste. En el Cuadro 5.2, los estadísticos *AIC* y *BIC* son menores cuando se utiliza la variable *spillovers*. El coeficiente de los *spillovers* tiene signo positivo y es significativo en las dos regiones. La probabilidad de innovar aumenta entre 13 y 16 por ciento, según la región, cuando los *spillovers* crecen uno por ciento.⁷

Al igual que a nivel nacional, el coeficiente estimado de la intensidad del gasto en I&D disminuye cuando se incorpora a la ecuación de innovaciones la variable de los *spillovers* (Cuadro 5.2). Los coeficientes estandarizados muestran que el efecto de $I\hat{I}D$ es más alto cuando se excluyen los *spillovers* en la ecuación de innovación. Para confirmar que la $I\hat{I}D$ se sobreestima en el modelo que no incluye los *spillovers* se utilizó el modelo de Karlson et al. (2012), a partir del cual podemos decir que hay sobreestimación del coeficiente de la intensidad del gasto en I&D.

La relación con el mercado extranjero reporta un signo positivo en las exportaciones, en Nuevo León, mientras que la participación de capital extranjero en las empresas tiene un efecto negativo sobre las innovaciones. Para las dos entidades federativas, la probabilidad de innovar aumenta con el tamaño de la empresa.

⁶En el Apéndice C están las regresiones con los diferentes *spillovers*.

⁷Los efectos marginales se estiman con $FO = 0$, $EXD = 0$ y el resto de las variables en la media.

Cuadro 5.2: NL y Edo. Méx.: Propensión a innovar en productos

Variable Dependiente: Innovación en productos (INN_1)				
Variables	Edo. Méx.		NL	
	(1)	(2)	(3)	(4)
$I\hat{I}D^b$	0.875*** (0.132)	0.490*** (0.141)	0.733*** (0.098)	0.415*** (0.106)
	2.999	1.679	3.178	1.799
$LSPILL_{\omega}^a$	-	0.159*** (0.013)	-	0.129*** (0.014)
		1.788		1.465
FO	-0.273 (0.165)	0.048 (0.195)	-0.734*** (0.188)	-0.835*** (0.175)
	-0.322	0.057	-0.772	-0.879
EXD	-0.460** (0.170)	-0.605** (0.187)	0.507*** (0.119)	0.624*** (0.131)
	-0.761	-1.000	0.746	0.918
LEM	0.207*** (0.046)	0.192*** (0.051)	0.225*** (0.048)	0.120* (0.052)
	0.792	0.734	0.802	0.425
N	1,488	1,488	1,149	1,149
R_p^2	0.194	0.385	0.249	0.361
LL	-359	-274	-330	-281
χ^2	168	252	199	201
AIC	745	576	686	590
BIC	813	651	751	660
P_{corr}	91.47	94.29	89.99	91.38

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (*beta*). En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. La descripción de las variables está en el Cuadro B.1 del Apéndice. ^a Los *spillovers* son intersectoriales. ^b $I\hat{I}D$ estimado con el Modelo (b).

*** p<0.01, ** p<0.05, * p<0.1

Para ver la importancia de los *spillovers* y del gasto en I&D se utilizan los coeficientes estandarizados (Cuadro 5.2). Si se considera el efecto directo de los *spillovers*, en NL, el efecto del gasto en I&D sobre la probabilidad de innovar es superior al de los *spillovers* (Columnas (4)). En el Edo. Méx. los *spillovers* tienen un efecto mayor sobre las innovaciones, que el gasto en I&D (Columna (2)).

Cuando utilizamos otras variables dependientes en la ecuación de innovaciones, encontramos que los *spillovers* de conocimientos son significativos y tienen un efecto positivo sobre la propensión a innovar en producto a nivel mercado y en procesos (Cuadros C.9 y C.10 del Apéndice). También se confirma que independientemente de la variable que se utilice para aproximar la producción de innovaciones, hay una sobreestimación de la intensidad del gasto en I&D cuando se omiten los *spillovers* de conocimientos.

Al igual que a nivel nacional, los resultados indican que existen *spillovers*, son significativos y positivos. Además su ausencia en la ecuación de innovaciones genera una sobreestimación del efecto del gasto en I&D en la propensión a innovar. La única diferencia es que los *spillovers* tienen una importancia relativa mayor frente a la intensidad del gasto en I&D, en el Edo. Méx.

Conclusiones

Los *spillovers* de conocimientos son un concepto abstracto y su medición ha sido parte del debate en la literatura relacionada con el progreso tecnológico. En esta parte de la tesis estimamos indicadores de los *spillovers* de conocimientos utilizando microdatos de encuestas de innovación de las empresas. Con esta información fue posible incorporar a las *proxies* de los *spillovers* aspectos objetivos como la madurez tecnológica de las empresas y la distancia espacial, con el propósito de avanzar en la medición de los *spillovers* de conocimientos.

Las estimaciones llevadas a cabo permiten mostrar que, en el caso de México, los *spillovers* tienen un efecto relevante y positivo sobre la innovación de productos por parte de las empresas que se benefician de ellos.

Cuando incluimos los *spillovers* en la ecuación que determina la propensión a innovar de una empresa se observa que en ésta el efecto del gasto en I&D se reduce. Los resultados confirman que la exclusión de los *spillovers* genera una sobrestimación del efecto de la intensidad del gasto en I&D sobre las innovaciones.⁸ Es decir, cuando los *spillovers* se excluyen de la estimación de la relación entre la inversión en I&D y la propensión a innovar en productos, los parámetros obtenidos están sesgados. Dichos sesgos llevan a sobreestimar el efecto del gasto en I&D en la propensión a innovar en productos. Por ello, este sesgo podría llevar a sostener que el retorno privado del gasto en I&D para las empresas, en términos de resultados con valor comerciable, es mayor al que efectivamente obtienen.

A diferencia de lo encontrado para el caso del Estado de México, a nivel nacional y en Nuevo León la importancia del gasto en I&D supera al efecto de los *spillovers* sobre la producción de innovaciones.

⁸Los resultados se corroboran a través del modelo de Karlson et al. (2012).

Otros resultados obtenidos son que la probabilidad de producir innovaciones se incrementa cuando las empresas tienen características similares en términos tecnológicos como el tipo de gasto interno y externo en I&D, así como en actividades relacionadas a la compra, asimilación, modificación y desarrollo de tecnología, además de considerar las similitudes del capital humano.

A través de los *spillovers* que consideran la proximidad geográfica no se puede confirmar un impacto sobre la producción de innovaciones a nivel nacional. Entre las posibles explicaciones de la falta de evidencia del aspecto espacial de los *spillovers* está que la heterogeneidad del país impide capturar las relaciones entre las entidades en términos de generación de innovaciones. Por otro lado, la unidad de observación de la ESIDET es la empresa por lo que podrían generarse sesgos espaciales al no conocer la ubicación de los establecimientos.

De acuerdo con los fundamentos económicos hay un problema de política pública a resolver cuando hay actividades asociadas a la innovación que no se llevan a cabo, a pesar de que el costo económico que éstas implican es menor al beneficio que representan para la economía en su conjunto. Cuando esto sucede es reflejo de que el costo privado de estas actividades es menor al beneficio privado que obtendrían las empresas que realizan los gastos en I&D. En este caso se requiere implementar políticas pública para apuntalar el gasto en este rubro de tal manera que no haya una subinversión desde una perspectiva social.

Los resultados de esta parte de la tesis muestran que hay pertinencia para la elaboración de política pública dirigida a la innovación, ya que sugieren la necesidad de apoyos gubernamentales para compensar el rendimiento privado de la I&D, que se ve disminuido ante la dificultad de cobrar por los *spillovers* de conocimientos utilizados por otras empresas.

A partir de este trabajo podemos distinguir algunas líneas de investigación a futuro. Entre ellas está estudiar la relación entre las innovaciones en productos y procesos. Por otro lado, debido a que no se encontró evidencia de la influencia del aspecto espacial, se podrían crear algunas regiones innovadoras y estudiar en ellas las relaciones entre la I&D, los *spillovers* y la producción de innovaciones.

Bibliografía

- Acs, Z. J. (2002), *Innovation and the growth of cities*, Edward Elgar.
- Acs, Z. J. (2009), Jaffe-Feldman-Varga: the search for knowledge spillovers, *in* A. Varga, ed., ‘Universities, knowledge transfer and regional development: geography, entrepreneurship and policy’, Edward Elgar, chapter 3, pp. 36–56.
- Acs, Z. J., Audretsch, D. B. y Feldman, M. P. (1991), ‘Real effects of academic research: Comment’, *American Economic Review* **82**(1), 363–367.
- Adams, J. D. (1990), ‘Fundamental stocks of knowledge and productivity growth.’, *Journal of Political Economy* **98**(4), 673–702.
- Adams, J. D. (2002), ‘Comparative localization of academic and industrial spillovers’, *Journal of Economic Geography* **2**(3), 253–278.
- Ahn, H. y Powell, J. L. (1993), ‘Semiparametric estimation of censored selection models with a nonparametric selection mechanism.’, *Journal of Econometrics* **58**, 3 – 29.
- Aiello, F. y Cardamone, P. (2005), ‘R&D spillovers and productivity growth: evidence from italian manufacturing microdata’, *Applied Economics Letters* **12**(10), 625–631.
- Aiello, F. y Cardamone, P. (2008), ‘R&D spillovers and firms’ performance in Italy. Evidence from a flexible production function’, *Empirical Economics* **34**(1), 143–166.
- Alvarez, R., Bravo-Ortega, C. y Navarro, L. (2010), Innovation, R&D investment and productivity in Chile, Research Department Publications 4691, Inter-American Development Bank, Research Department.

- Anselin, L., Varga, A. y Acs, Z. (1997), 'Local geographic spillovers between university research and high technology innovations', *Journal of Urban Economics* **42**(3), 422–448.
- Anselin, L., Varga, A. y Acs, Z. J. (2000), 'Geographic and sectoral characteristics of academic knowledge externalities', *Papers in Regional Science* **79**(4), 435–443.
- Arbeláez, M. A. y Parra, M. (2011), Innovation, R&D investment and productivity in colombian firms, Working paper series IDB-WP-251, Inter-American Development Bank, Dep. of Research and Chief Economist.
- Arrow, K. J. (1962), 'The economic implications of learning by doing', *The Review of Economic Studies* **29**(3), 155–173.
- Audretsch, D. B. y Feldman, M. P. (1996a), 'Innovation clusters and the industry life cycle', *Review of Industrial Organization* **11**(2), 253–273.
- Audretsch, D. B. y Feldman, M. P. (1996b), 'R&D spillovers and the geography of innovation and production', *American Economic Review* **86**(3), 630–40.
- Autant-Bernard, C. (2001), 'The geography of knowledge spillovers and technological proximity', *Economics of Innovation and New Technology* **10**(4), 237–254.
- Beise, M. y Stahl, H. (1999), 'Public research and industrial innovations in Germany', *Research Policy* **28**(4), 397 – 422.
- Benavente, J. M. (2006), 'The role of research and innovation in promoting productivity in Chile', *Economics of Innovation and New Technology* **15**(4-5), 301–315.
- Bravo-Ortega, C., Benavente, J. M. y González, Á. (2014), 'Innovation, exports, and productivity: Learning and self-selection in Chile', *Emerging Markets Finance and Trade* **50**, 68–95.
- Calderón-Madrid, A. (2012), Análisis del impacto en productividad de las actividades de desarrollo tecnológico e innovaciones de las empresas: aplicación del modelo CDM en el caso de México. Working paper.

- Cameron, A. C. y Trivedi, P. K. (2005), *Microeconometrics. methods and applications.*, New York, N.Y. Cambridge University.
- Caniels, M. C. J. (2000), *Knowledge spillovers and economic growth: Regional growth differentials across Europe*, New Horizons in the Economics of Innovation.
- Chudnovsky, D., López, A. y Pupato, G. (2006), ‘Innovation and productivity in developing countries: A study of argentine manufacturing firms’ behavior (1992-2001)’, *Research Policy* **35**(2), 266–288.
- Cohen, W. M. y Klepper, S. (1996), ‘A reprise of size and R&D’, *Economic Journal* **106**(437), 925–51.
- Cohen, W. M. y Levinthal, D. A. (1990), ‘Absorptive capacity: A new perspective on learning and innovation.’, *Administrative Science Quarterly* **35**(1), 128 – 152.
- Cosslett, S. (1991), Semiparametric estimation of regression model with sample selectivity, *in* W. A. Barnett, J. L. Powell y G. Tauchen, eds, ‘Nonparametric and semiparametric methods in econometrics and statistics’, Cambridge University Press.
- Crepon, B., Duguet, E. y Mairessec, J. (1998), ‘Research, innovation and productivity: An econometric analysis at the firm level’, *Economics of Innovation and New Technology* **7**(2), 115–158.
- Crespi, G., Arias, E., Tacsir, E., Vargas, F. y Zuñiga, P. (2014), ‘Innovation for economic performance: the case of Latin American firms.’, *Eurasian Business Review* **4**(1), 31–50.
- Crespi, G. y Zuñiga, P. (2012), ‘Innovation and productivity: Evidence from six Latin American countries’, *World Development* **40**(2), 273 – 290.
- De Negri, J. A., Esteves, L. y Freitas, F. (2007), Knowledge production and firm growth in Brazil, Working Papers 0057, Universidade Federal do Paraná, Department of Economics.
- Doring, T. y Schnellenbach, J. (2006), ‘What do we know about geographical knowledge spillovers and regional growth?: A survey of the literature’, *Regional Studies* **40**(3), 375–395.

- Dutrénit, G., De Fuentes, C., Santiago, F., Torres, A. y Gras, N. (2013), Innovation and productivity in the service sector. The case of Mexico, Discussion paper IDB-DP-293, IDB.
- Feldman, M. P. (1994a), *The geography of innovation*, Vol. 2, Economics of Science, Technology and Innovation.
- Feldman, M. P. (1994b), 'Knowledge complementarity and innovation', *Small Business Economics* **6**(5), 363–372.
- Feldman, M. P. y Kogler, D. F. (2010), Stylized facts in the geography of innovation, *in* B. H. Hall y N. Rosenberg, eds, 'Handbook of The Economics of Innovation', Vol. 1, North-Holland, chapter 8, pp. 381–410.
- Gallant, A. R. y Nychka, D. W. (1987), 'Semi-nonparametric maximum likelihood estimation', *Econometrica* **55**(2), 363–390.
- Gertler, M. S. (2003), 'Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there)', *Journal of Economic Geography* **3**(1), 75–99.
- Giacinto, V. D. y Pagnini, M. (2011), 'Local and global agglomeration patterns: Two econometrics-based indicators', *Regional Science and Urban Economics* **41**(3), 266–280.
- Goedhuys, M. (2007), The impact of innovation activities on productivity and firm growth: evidence from Brazil, MERIT Working Papers 002, United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT).
- Goya, E., Vayá, E. y Suriñach, J. (2013), Do spillovers matter? CDM model estimates for Spain using panel data, IREA working papers, University of Barcelona, Research Institute of Applied Economics.
- Greene, W. H. (2003), *Econometric analysis*, Upper Saddle River, N.J. Prentice Hall.
- Griffith, R., Huergo, E., Mairesse, J. y Peters, B. (2006), 'Innovation and productivity across four European countries', *Oxford Review of Economic Policy* **22**(4), 483–498.

- Griliches, Z. (1979), 'Issues in assessing the contribution of research and development to productivity growth', *Bell Journal of Economics* **10**(1), 92–116.
- Griliches, Z. (1990), 'Patent statistics as economic indicators: A survey', *Journal of Economic Literature* **28**(4), 1661–1707.
- Grossman, G. M. y Helpman, E. (1991), *Innovation and growth in the global economy*, Cambridge, MA, MIT Press.
- Hall, B. H. y Mairesse, J. (2006), 'Empirical studies of innovation in the knowledge-driven economy', *Economics of Innovation and New Technology* **15**(4-5), 289–299.
- Hall, B. H., Mairesse, J. y Mohnen, P. (2010), Measuring the returns to R&D, in B. H. Hall y N. Rosenberg, eds, 'Handbook of The Economics of Innovation', Vol. 2, North-Holland, chapter 24, pp. 1033–1082.
- Heckman, J. y Robb, R. (1985), Alternative methods for evaluating the impact of interventions, in J. Heckman y B. Singer, eds, 'Longitudinal analysis of labor market data', Cambridge University Press, pp. 156–245.
- Hoffmann, R. y Kassouf, A. L. (2005), 'Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure', *Applied Economics* **37**(11), 1303–1311.
- Ichimura, H. y Lee, L. F. (1991), Semiparametric least squares single equation estimation of multiple index models, in W. A. Barnett, J. L. Powell y G. Tauchen, eds, 'Nonparametric and semiparametric methods in econometrics and statistics', Cambridge University Press, pp. 3–49.
- INEGI (2015), 'Instituto Nacional de Estadística y Geografía'.
URL: <http://www.inegi.org.mx>
- Inkmann, J. (2000), Horizontal and vertical R&D cooperation, CoFE Discussion Paper 00-02, Center of Finance and Econometrics, University of Konstanz.

- Jaffe, A. B. (1986), 'Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits, and market values.', *American Economic Review* **76**(5), 984.
- Jaffe, A. B. (1988), 'Demand and supply influences in R&D intensity and productivity growth.', *Review of Economics & Statistics* **70**(3), 431.
- Jaffe, A. B. (1989), 'Real effects of academic research', *American Economic Review* **79**(5), 957–70.
- Jaffe, A. B., Trajtenberg, M. y Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *The Quarterly Journal of Economics* **108**(3), 577–598.
- Johansson, B. (2014), Generation and diffusion of innovation, *in* M. M. Fischer y P. Nijkamp, eds, 'Handbook of Regional Science', Springer Berlin Heidelberg, chapter 21, pp. 391–412.
- Kaiser, U. (2002), 'Measuring knowledge spillovers in manufacturing and services: an empirical assessment of alternative approaches', *Research Policy* **31**, 125–144.
- Karlson, K. B., Holm, A. y Breen, R. (2012), 'Comparing regression coefficients between same-sample nested models using logit and probit: A new method', *Sociological Methodology* **42**(1), 286–313.
- Karlsson, C. y Grasjo, U. (2014), Knowledge flows, knowledge externalities, and regional economic development, *in* M. M. Fischer y P. Nijkamp, eds, 'Handbook of Regional Science', Springer Berlin Heidelberg, chapter 22, pp. 413–437.
- Karlsson, C., Johansson, B. y R Stough, R. (2013), Knowledge & innovation in space, Technical report, Royal Institute of Technology, CESIS-Centre of Excellence for Science and Innovation Studies.
- Keilbach, M. (2000), *Spatial knowledge spillovers and the dynamics of agglomeration and regional growth*, Contributions to Economics.
- Krugman, P. (1991), *Geography and Trade*, MIT Press, MA.

- Krugman, P. y Obsfeld, M. (2005), *International economics: Theory and policy.*, Pearson Addison-Wesley.
- Li, Q. y Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice.*, Princeton and Oxford.
- Link, A. N. y Rees, J. (1990), ‘Firm size, university based research, and the returns to R&D’, *Small Business Economics* **2**(1), 25–31.
- Long, J. S. (1997), *Regression models for categorical and limited dependent variables.*, Advanced quantitative techniques in the social sciences: 7, Thousand Oaks, California Sage Publications.
- Mairesse, J. y Mohnen, P. (2005), ‘The importance of R&D for innovation: A reassessment using french survey data.’, *Journal of Technology Transfer* **30**(1-2), 183–197.
- Mansfield, E. (1984), Using linked patent and R&D data to measure interindustry technology flows: Comment, *in* Z. Griliches, ed., ‘R&D, Patents, and Productivity’, University of Chicago Press, pp. 462–464.
- Marshall, A. (1890), *Principles of Economics*, MacMillan, London.
- Newey, W. K. (2009), ‘Two-step series estimation of sample selection models.’, *Econometrics Journal* **12**(1), S217–S229.
- Newey, W. K., Powell, J. L. y Walker, J. R. (1990), ‘Semiparametric estimation of selection models: Some empirical results’, *The American Economic Review* **80**(2), 324–328.
- Powell, J. L. (1987), Semiparametric estimation of bivariate latent variable models, Working paper 9704, University of Wisconsin.
- Powell, J. L. (1994), Estimation of semiparametric models, *in* ‘Handbook of Econometrics’, Vol. 4, Elsevier, chapter 41, pp. 2443–2521.

- Raffo, J., Lhuillery, S. y Miotti, L. (2008), 'Northern and southern innovativity: a comparison across European and Latin American countries', *The European Journal of Development Research* **20**(2), 219–239.
- Scherer, F. M. (1983), 'The propensity to patent', *International Journal of Industrial Organization* **1**(1), 107–128.
- Varga, A. (2000), 'Local academic knowledge transfers and the concentration of economic activity', *Journal of Regional Science* **40**(2), 289–309.
- Vella, F. (1998), 'Estimating models with sample selection bias: A survey', *The Journal of Human Resources* **33**(1), 127–169.
- Wooldridge, J. M. (2002), *Econometric analysis of cross section and panel data.*, Cambridge, Mass. MIT Press.

Apéndice A

Especificaciones econométricas

A.1. Modelo de selección muestral

Cuando una variable tiene información sólo para una parte de la muestra tenemos un problema de selección muestral.¹ Si una variable es resultado de otra, se dice que hay truncación incidental, es decir, si la población se auto-selecciona a participar o no en alguna actividad. Al utilizar MCO se obtendrían estimadores inconsistentes, sería equivalente a tener un problema de omisión de variables. Aquí se considera que la selección está determinada por un modelo probit. El modelo se compone de la ecuación principal (A.1) y la ecuación de selección (A.2):

$$y_1 = x_1\beta_1 + u_1 \quad (\text{A.1})$$

$$y_2 = 1[x\delta_2 + v_2 > 0] \quad (\text{A.2})$$

Los supuestos del modelo son: (a) existe un problema de selección dado que (x, y_2) siempre se observan y y_1 sólo se observa cuando $y_2 = 1$, (b) se supone exogeneidad, es decir, (u_1, v_2) son independientes de x con media cero, (c) $v_2 \sim N(0, 1)$, (u_1, v_2) tiene una distribución normal bivariada con media cero y un coeficiente de correlación ρ . Dado que y_1 sólo se observa cuando $y_2 = 1$, lo que se busca estimar es el valor esperado de y_1 dado que y_1 :

¹Las referencias de esta sección son: Greene (2003), Wooldridge (2002), Hoffmann y Kassouf (2005).

$$E[y_1|x, y_2 = 1] = E[y_1|x\delta_2 + v_2 > 0] \quad (\text{A.3})$$

$$= x_1\beta_1 + \sigma_u\rho\lambda\left(\frac{\gamma'x}{\sigma_v}\right) \quad (\text{A.4})$$

donde, $\lambda\left(\frac{\gamma'x}{\sigma_v}\right) = \frac{\phi\left(\frac{\gamma'x}{\sigma_v}\right)}{\Phi\left(\frac{\gamma'x}{\sigma_v}\right)}$ es la inversa de Mills. Cuando existe el problema de selección hay correlación entre los errores que se representa a través de ρ por lo que se utiliza un modelo de selección. Si se ignora este problema, se estaría omitiendo la variable $\sigma_u\rho\lambda\left(\frac{\gamma'x}{\sigma_v}\right)$ lo que genera estimadores sesgados.

Heckman propone estimar el modelo en dos etapas: primero se obtiene $\hat{\delta}_2$ utilizando un modelo probit sobre la muestra completa (N), se estima la inversa de Mills, y después, con una regresión de MCO se hace la estimación de los coeficientes de la ecuación principal, agregando la variable omitida, de tal forma que los coeficientes sean consistentes.

Para la estimación es recomendable utilizar al menos una variable en la ecuación de selección que no esté en la principal. Si no se hace así la variación de la variable estimada en la primera etapa dependerá sólo de la no-linealidad de la inversa de Mills.

Efectos Marginales

El efecto marginal condicional de los coeficientes sobre y_1 tiene dos componentes: el directo sobre la media de y_1 que es β_1 y si la variable aparece en la ecuación de selección, también influye a y_1 a través de λ . Entonces, el efecto global sobre y_1 de un cambio en una variable que esté tanto en la ecuación principal como en la de selección, condicionado a que $y_1 > 0$ es:

$$\frac{\partial E[y_1|y_2 = 1]}{\partial x_k} = \beta_k - \gamma_k\left(\frac{\rho\sigma_u}{\sigma_v}\right)\gamma(\alpha_v) \quad (\text{A.5})$$

donde, $\alpha_v = -\frac{\gamma'x}{\sigma_v}$ y $\gamma = \lambda^2 + \alpha\lambda$. Los términos del efecto marginal se compensan dejando sólo el efecto donde $y_2 = 1$ de entrada.

Al aumentar en una unidad x_k , y_1 aumenta en $\frac{\partial E[y_1|y_2=1]}{\partial x_k} * 100$ por ciento, una vez ajustado por la selección muestral, en promedio, y manteniendo el resto de las variables constantes en su media. Si la variable dependiente está en logaritmos, el aumento de una unidad en x_k ,

aumenta y_1 en $\left(\exp\left(\frac{\partial E[y_1|y_2=1]}{\partial x_k}\right) - 1\right) * 100$ por ciento, con el resto de las variables en la media y para $y_2 = 1$.

El efecto marginal condicional de una variable binaria x_k , que toma valores de cero y uno es:

$$\frac{\Delta E[y_1|y_2 = 1]}{\Delta x_k} = \beta_k - \gamma_k \Delta\gamma \quad (\text{A.6})$$

donde, $\Delta\gamma = \frac{\phi(\gamma'\bar{\mathbf{x}}_{(1)}/\sigma_v)}{\Phi(\gamma'\bar{\mathbf{x}}_{(1)}/\sigma_v)} - \frac{\phi(\gamma'\bar{\mathbf{x}}_{(0)}/\sigma_v)}{\Phi(\gamma'\bar{\mathbf{x}}_{(0)}/\sigma_v)}$. En $\bar{\mathbf{x}}_{(1)}$ las variables están en la media mientras que la variable dicotómica toma valor de uno y en $\bar{\mathbf{x}}_{(0)}$ toma valor cero.

Después de corregir por la selección muestral y manteniendo el resto de las variables en la media, un cambio de cero a uno de la variable x_k , cambia a y_1 en $\frac{\Delta E[y_1|y_2=1]}{\Delta x_k} * 100$ por ciento, en promedio, para $y_1 > 0$. Si la variable dependiente está en logaritmos tenemos que el cambio de cero a uno en x_k , afecta a y_1 en $\left(\exp\left(\frac{\Delta E[y_1|y_2=1]}{\Delta x_k}\right) - 1\right) * 100$ por ciento, con el resto de las variables en la media y para $y_2 = 1$.

Cuando la variable x_k sólo se encuentra en la ecuación principal el efecto marginal es:

$$\frac{\partial E[y_1|y_2 = 1]}{\partial x_k} = \beta_k \quad (\text{A.7})$$

Después de corregir por la selección muestral y manteniendo el resto de las variables en la media, un cambio en la variable x_k , aumenta a y_1 en $\beta_k * 100$ por ciento, en promedio.

A.2. Modelo semiparamétrico de selección muestral

El modelo paramétrico de selección muestral presentado en la sección anterior supone la forma de la distribución conjunta de los términos de error de la ecuación principal (A.1) y la ecuación de selección (A.2). El problema se presenta si la distribución conjunta de los errores está mal especificada, es decir, si no se cumple la normalidad bivariada de los errores, los coeficientes ($\hat{\beta}$) serían inconsistentes (Ahn y Powell, 1993). La sensibilidad de los coeficientes estimados respecto a la especificación correcta de la distribución de los errores ha generado que se busquen alternativas menos restringidas para la estimación del modelo de selección muestral.

Los modelos semiparamétricos relajan el supuesto sobre la distribución de los errores. Entre las ventajas de dichos modelos está que permite una especificación más general de los parámetros, la estimación de los parámetros de interés son consistentes bajo supuestos menos restrictivos (como los derivados del modelo paramétrico), además, son más precisos que los derivados de modelos noparamétricos (Powell, 1994). Entre las desventajas de los estimadores obtenidos de modelos semiparamétricos están que son menos eficientes que las estimaciones de máxima verosimilitud de un modelo paramétrico bien especificado y son más sensibles a malas especificaciones de la ecuación estructural o de otros componentes paramétricos del modelo (Powell, 1994).

Heckman y Robb (1985) proponen una aproximación semiparamétrica donde en la primera etapa se estima la ecuación de selección de forma no-paramétrica y después, una función índice a través de una expansión de Fourier. Por su parte, Cosslett (1991) propone estimar primero los coeficientes de la ecuación de selección a través del método de máxima verosimilitud no-paramétrico y después estima la ecuación principal mientras aproxima la corrección de la selección con variables indicadoras (Vella, 1998).

El modelo de Powell (1987) consiste en una estimación de dos etapas, eliminando el sesgo por selección al diferenciar por pares las observaciones, mientras que en el modelo de Ichimura y Lee (1991) la selección muestral es un caso especial de un modelo de múltiples índices (Li y Racine, 2007). Estos modelos no hacen supuestos sobre la distribución conjunta de los errores y funcionan bajo el supuesto de que las variables explicativas son independientes de los errores.

En el caso del enfoque de Gallant y Nychka (1987) se sugiere aproximar la distribución conjunta de los errores a través de series de expansión (Li y Racine, 2007). Parten de una distribución conjunta normal como base y a través de la expansión de series se produce una forma más general de la función de distribución.

Para obtener coeficientes estimados consistentes, Newey et al. (1990) relajan el supuesto de normalidad bivariada de los errores, implementan una función noparamétrica en la ecuación de selección, lo anterior siguiendo una metodología de dos etapas similar a la propuesta por Heckman. Otro modelo propuesto por Newey (2009) propone la corrección de la selec-

ción con un modelo donde no se hace supuesto sobre la distribución de los errores y se utiliza un aproximación a través de series.

Los modelos que se presentan en esta sección permiten hacer estimaciones consistentes sin suponer que la distribución conjunta de los errores del modelo de selección es normal bivarianda.²

A.3. Modelo probit

Para variables dependientes binarias se puede abordar el modelo probit a través de un modelo de variable latente donde tenemos una variable continua y^* no observada que toma valores de uno y cero:³

$$y_i = \begin{cases} 1 & \text{si } y_i^* = \mathbf{x}_i\beta + \epsilon_i > 0 \\ 0 & \text{de otra forma} \end{cases} \quad (\text{A.8})$$

La probabilidad condicional de que y tome valor uno dado el vector de variables \mathbf{x} , si los errores siguen una distribución normal es:

$$Pr(y^* > 0|\mathbf{x}) = Pr(\epsilon > -\mathbf{x}\beta|\mathbf{x}) \quad (\text{A.9})$$

$$= Pr(\epsilon < \mathbf{x}\beta|\mathbf{x}) \quad (\text{A.10})$$

$$= F(\mathbf{x}\beta) \quad (\text{A.11})$$

donde, $F(\cdot)$ es una función específica, en el caso del modelo probit es una función de densidad acumulada (FDA) normal estándar. La probabilidad condicional del probit es:

$$Pr(y^* > 0|\mathbf{x}) = \Phi(\mathbf{x}'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z)dz \quad (\text{A.12})$$

donde, $\Phi(\cdot)$ es la FDA normal estándar. Cuando los errores (ϵ) son normales supone que $E[\epsilon|x] = 0$ y $Var[\epsilon|x] = 1$, valor arbitrario en el sentido de que no se puede probar y se utiliza para simplificar la expresión de la distribución.

²Una revisión general del tema están en los trabajos de Powell (1994), Vella (1998) y Li y Racine (2007).

³Las referencias de esta sección son: Long (1997) y Cameron y Trivedi (2005)

Efectos Marginales

El efecto del cambio de una variable independiente sobre la probabilidad condicional de $y = 1$ es el efecto marginal y el signo del efecto marginal está determinado por el coeficiente β_k . Los efectos marginales son la pendiente de la curva de probabilidad relacionando x_k a la $Pr(y = 1|\mathbf{x})$ dejando el resto de las variables constantes en su media. Para una variable continua, el efectos marginal en la media es:

$$\frac{\partial E[y = 1|\bar{\mathbf{x}}]}{\partial x_k} = \phi(\bar{\mathbf{x}}'\beta)\beta_k \quad (\text{A.13})$$

Si x_k aumenta en una unidad, la probabilidad de $y = 1$ aumenta en $\frac{\partial E[y=1|\bar{\mathbf{x}}]}{\partial x_k}$, manteniendo constante el resto de las variables en la media. Si x_k está en logaritmos, tenemos que un incremento de uno por ciento en x_k aumenta la probabilidad de $y = 1$ en $\frac{\partial E[y=1|\bar{\mathbf{x}}]}{\partial x_k}$ puntos porcentuales.

El efecto marginal de una variable *dummy* que cambia de cero a uno, manteniendo las otras variables en su media, es:

$$\frac{\Delta E[y = 1|\bar{\mathbf{x}}]}{\Delta x_k} = Pr(y = 1|\bar{\mathbf{x}}, x_k = 1) - Pr(y = 1|\bar{\mathbf{x}}, x_k = 0) \quad (\text{A.14})$$

Si $x_k = 1$, la probabilidad de $y = 1$ es $\frac{\Delta E[y=1|\bar{\mathbf{x}}]}{\Delta x_k}$ mayor que si x_k es cero, manteniendo el resto de las variables en su media.

Los efectos marginales difieren según el punto en donde se evalúen las variables independientes (\mathbf{x}) incluidas en la regresión. Se recomienda plantear un escenario eligiendo el valor de las variables, en especial cuando hay muchas variables binarias.

Coefficientes estandarizados

En las regresiones del modelo probit se presentan los coeficientes estandarizados (*fully standardized coefficient*). Si σ_k es la desviación estándar de x_k , el coeficiente estandarizado es:

$$\beta_k^S = \frac{\sigma_k \beta_k}{\sigma_{y^*}} = \sigma_k \beta_k^{S_{y^*}} \quad (\text{A.15})$$

Si x_k aumenta una desviación estándar, el incremento esperado en y^* es de β_k^S desviaciones estándar, manteniendo las otras variables constantes. Los coeficientes estandarizados homogeneizan las unidades de las variables independientes y pueden ayudarnos a jerarquizar el efectos sobre la variable dependiente.

A.4. Medidas de ajuste: *AIC* y *BIC*

Para comparar el ajuste de un modelo con otros, existen dos medidas de información que evalúan el ajuste de un modelo y sirven para elegir el modelo que mejor se ajusta a los datos:⁴

Criterio de información de Akaike (por sus siglas en inglés *AIC*) se calcula de la siguiente forma:

$$AIC = -2\ln\hat{L}(M_\beta) + 2k \quad (\text{A.16})$$

donde, $\hat{L}(M_\beta)$ es la estimación de verosimilitud del modelo y k el número de parámetros. El *AIC* se utiliza para comparar entre modelos de diferentes muestras o para comparar entre modelos no anidados. El modelo con menor *AIC* es el que tiene mejor ajuste.

Criterio de información Bayesiano (por sus siglas en inglés *BIC*) es una medida para evaluar el ajuste total del modelo, que permite hacer comparaciones entre modelos anidados y no anidados.

$$BIC = -2\ln\hat{L}(M_\beta) + k\ln N \quad (\text{A.17})$$

La diferencia en los *BIC* de dos modelos indica cuál es el que mejor ajusta a los datos, entonces, el modelo con menor *BIC* es el que se prefiere.

Cuadro A.1: Mejor modelo según la diferencia en valores absolutos de los *BIC*

Diferencia absoluta	Evidencia
0-2	Débil
2-6	Positiva
6-10	Fuerte
>10	Muy fuerte

Fuente: Long (1997)

⁴Esta sección está basada en Long (1997).

Apéndice B

Definición de variables

Cuadro B.1: Descripción de las variables

Variable	Categoría	Descripción	Periodo
INN_1	Innovación	Toma valor de uno si la empresa introdujo al mercado productos nuevos o significativamente mejorados	2011
INN_2	Innovación	Toma valor de uno si la empresa introdujo al mercado productos nuevos o significativamente mejorados para el mercado (novedad nacional y mundial)	2011
INN_3	Innovación	Toma valor de uno si la empresa introdujo al mercado procesos nuevos o significativamente mejorados	2011
Gasto intra-muros	Gasto en I&D	Son los gastos que hace la empresa para realizar un proyecto de I&D, en la misma empresa o en otras instalaciones.	2010
Gasto extra-muros	Gasto en I&D	Son los pagos que la empresa otorga a otros por realizar un proyecto de I&D, en la misma empresa o en otro lugar.	2010
IID	Gasto en I&D	Logaritmo de la suma del gasto de I&D, el gasto en innovaciones y el gasto en servicios científicos y tecnológicos entre el número de empleados de la empresa	2010-2011
SID	Gasto en I&D	Toma valor de uno si la empresa gasta en I&D	2010-2011
$SPILL_0$	<i>Spillovers</i>	<i>Spillovers</i> sin ponderador (suma del gasto intramuros de todas las empresas diferentes a i)	2010-2011
$SPILL_\omega$	<i>Spillovers</i>	<i>Spillovers</i> ponderados por la similitud tecnológica	2010-2011
$SPILL_{\hat{\omega}}$	<i>Spillovers</i>	<i>Spillovers</i> ponderados por la similitud tecnológica y capacidad de asimilación de tecnología	2010-2011
$SPILL_g$	<i>Spillovers</i>	<i>Spillovers</i> ponderados por la cercanía espacial	2010-2011
$SPILL_\nu$	<i>Spillovers</i>	<i>Spillovers</i> ponderados por la proximidad espacial y similitud tecnológica	2010-2011
$LSPILL$	<i>Spillovers</i>	Logaritmo de los diferentes tipos de <i>spillovers</i> ($0, \omega, \hat{\omega}, g$ y ν)	2010-2011

Continua...

Descripción de las variables (*Continuación*)

Variable	Categoría	Descripción	Periodo
<i>EM</i>	Empleo	Número de empleados	2010
<i>LEM</i>	Empleo	Logaritmo del número de empleados	2010
<i>EM₁</i>	Empleo	Toma valor de uno si el número de empleados es menor o igual a 50 (empresas pequeñas)	2010
<i>EM₂</i>	Empleo	Toma valor de uno si el número de empleados es mayor a 50 hasta 250 (empresas medianas)	2010
<i>EM₃</i>	Empleo	Toma valor de uno si el número de empleados es mayor a 250 hasta 500 (empresas grandes)	2010
<i>EM₄</i>	Empleo	Toma valor de uno si el número de empleados es mayor a 500 (empresas grandes)	2010
<i>EXD</i>	Exportaciones	Toma valor de uno si la empresa reportó ingresos por ventas en el mercado extranjero	2010
<i>EXD₁</i>	Exportaciones	Es uno cuando el porcentaje de exportaciones es mayor a cero hasta 25 %	2010
<i>EXD₂</i>	Exportaciones	Es uno cuando el porcentaje de exportaciones está en el rango de 25 % al 50 %	2010
<i>EXD₃</i>	Exportaciones	Es uno cuando el porcentaje de exportaciones está en el rango de 50 % al 75 %	2010
<i>EXD₄</i>	Exportaciones	Es uno cuando el porcentaje de exportaciones es superior al 75 %	2010
<i>FO</i>	Participación extranjera	Toma valor de uno si la participación de capital extranjero es mayor a 10 %	2011
<i>CON</i>	Índice de concentración	Ingreso de las 8 empresas más grandes entre el ingreso total del sector	2010
<i>INFO₁</i>	Información de mercado	Promedio de las respuestas sobre la importancia de los clientes, proveedores, competidores, consultorías y expertos como fuentes de información para hacer innovaciones	2010-2011
<i>INFO₂</i>	Información científica	Promedio de las respuestas sobre la importancia de las universidades, instituciones de investigación pública o privada como fuentes de información para hacer innovaciones	2010-2011
<i>INFO₃</i>	Información pública	Promedio de las respuestas sobre la importancia de las patentes, conferencias, redes, seminarios y ferias como fuentes de información para hacer innovaciones	2010-2011
<i>OBST₁</i>	Obstáculos a la innovación	Toma valor uno si los costos de innovación muy elevados son un obstáculo altamente o moderadamente significativos	2010-2011
<i>OBST₂</i>	Obstáculos a la innovación	Toma valor uno si los obstáculos derivados de la legislación vigente son altamente o moderadamente significativos	2010-2011
<i>OBST₃</i>	Obstáculos a la innovación	Toma valor uno si la falta de fuentes de financiamiento adecuadas son altamente o moderadamente significativos	2010-2011

Continúa...

Descripción de las variables (*Continuación*)

Variable	Categoría	Descripción	Periodo
<i>OBJ</i>	Objetivo de innovar	Toma valor uno si mantener la participación de mercado es un objetivo altamente o moderadamente significativos	2010-2011
<i>FIN</i>	Financiamiento	Toma valor de uno si la empresa recibe financiamiento público	2010-2011
<i>PFIN₁</i>	Financiamiento	Toma valor uno si recibe financiamiento de algún programa de CONACYT	2010-2011
<i>PFIN₂</i>	Financiamiento	Toma valor uno si recibe financiamiento de algún programa de la SE	2010-2011
<i>PFIN₃</i>	Financiamiento	Toma valor uno si recibe financiamiento de algún programa de STPS, BANCOMEXT o STPS-SEP	2010-2011
<i>MFIN₁</i>	Financiamiento	Toma valor uno si recibe financiamiento de créditos de instituciones bancarias	2010-2011
<i>MFIN₂</i>	Financiamiento	Toma valor de uno si recibe apoyos gubernamentales	2010-2011
<i>CO</i>	Colaboraciones	Toma valor de uno si la empresa colabora con otras empresas, instituciones o universidades	2010-2011
Sectores (OCDE)	Manufactura	4 Alimentos, bebidas y tabaco 7 Textiles, prendas de vestir, piel y cuero 11 Madera, papel, imprentas y publicaciones 15 Carbón, petróleo, energía nuclear, químicos y productos de caucho y plástico 21 Productos minerales no metálicos 22 Metales básicos 25 Productos fabricados de metal (excepto maquinaria y equipo) 26 Maquinaria, equipo, instrumentos y equipo de transporte 39 Muebles y otras manufacturas no especificadas en otra parte	

Todas las variables fueron construidas con información de la ESIDET 2012.

Apéndice C

Otras regresiones

Cuadro C.1: Intensidad del gasto en I&D: programas (*PFIN*) y mecanismos (*MFIN*) de financiamiento

Variable dependiente: *IID* (Ecuación principal)

Variables	Modelo (c) con <i>PFIN</i>			Modelo (d) con <i>MFIN</i>		
	Nacional (1)	Edo. Méx. (2)	NL (3)	Nacional (4)	Edo. Méx. (5)	NL (6)
<i>FO</i> *	0.329* (0.148)	0.411 (0.397)	0.409 (0.392)	0.576*** (0.143)	0.313 (0.412)	0.677 (0.384)
<i>EXD</i> ₁ *	0.172 (0.113)	0.562* (0.286)	-0.509 (0.300)	0.216* (0.108)	0.483 (0.287)	-0.468 (0.281)
<i>EXD</i> ₂ *	0.199 (0.169)	1.591** (0.485)	-0.060 (0.345)	0.177 (0.166)	1.453** (0.506)	-0.036 (0.385)
<i>EXD</i> ₃ *	0.718** (0.265)	0.503 (0.665)	-1.396* (0.659)	0.612* (0.261)	0.420 (0.624)	-1.335* (0.620)
<i>EXD</i> ₄ *	0.334 (0.174)	1.000 (0.545)	-0.197 (0.595)	0.275 (0.162)	0.435 (0.647)	0.129 (0.598)
<i>CO</i> *	0.904*** (0.103)	1.029*** (0.305)	0.623 (0.340)	0.339** (0.110)	0.819* (0.346)	0.229 (0.317)
<i>PFIN</i> ₁	1.200*** (0.130)	1.091 (0.622)	1.259*** (0.324)	-	-	-
<i>PFIN</i> ₂	0.527** (0.174)	3.567*** (0.315)	-1.757* (0.781)	-	-	-
<i>PFIN</i> ₃	-1.795*** (0.238)	-0.232 (0.434)	-1.338* (0.588)	-	-	-
<i>MFIN</i> ₁	-	-	-	1.227*** (0.108)	0.109 (0.689)	1.051** (0.351)
<i>MFIN</i> ₂	-	-	-	1.944*** (0.121)	0.684 (0.415)	1.728*** (0.360)
<i>N</i>	11,679	1,488	1,149	11,679	1,488	1,149
<i>LL</i>	-10,496	-1,084	-1,037	-10,400	-1,104	-1,031
χ^2	731		180	1,009	150	174
ρ	0.677	0.704	0.883	0.627	0.579	0.901
χ^2_ρ	61	11	25	33	5	21
p_ρ	0.000	0.001	0.000	0.000	0.023	0.000
<i>AIC</i>	21,198	2,250	2,158	20,986	2,290	2,144
<i>BIC</i>	21,957	2,468	2,370	21,737	2,508	2,351

El asterisco en las variables indica que son dicotómicas. Los resultados corresponden a la segunda etapa del modelo Heckman (Apéndice A). Los coeficientes corresponden a los efectos marginales cuando las variables *INFO* y *CON* están en la media, *OBJ* = 1 y el resto de las variables están en cero. En todos los casos se controla con variables *dummies* por sector y entidad federativa. Errores estándar robustos entre paréntesis.

*** p<0.01, ** p<0.05, * p<0.1

Cuadro C.2: Nacional. Propensión a innovar en productos (Modelo (a))

Variable Dependiente: Innovación en productos (INN_1)

Variables	<i>Spillovers</i> intersectoriales				<i>Spillovers</i> intrasectoriales				
	Sin $SPILL$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_g$	$SPILL_v$
$I\hat{I}D$	0.680*** (0.029) 2.593	0.676*** (0.029) 2.576	0.665*** (0.031) 2.535	0.405*** (0.034) 1.544	0.675*** (0.029) 2.571	0.648*** (0.030) 2.471	0.424*** (0.034) 1.617	0.675*** (0.029) 2.574	0.675*** (0.029) 2.574
$LSPILL$	-	-0.208 (0.132) -1.500	0.350** (0.107) 2.309	0.124*** (0.005) 1.331	0.010* (0.005) 0.169	0.361*** (0.077) 2.348	0.123*** (0.005) 1.695	-2.056 (1.422) -12.835	-2.053 (1.419) -12.813
FO^*	-0.239*** (0.059) -0.273	-0.237*** (0.059) -0.271	-0.238*** (0.059) -0.272	-0.133* (0.063) -0.152	-0.235*** (0.059) -0.268	-0.233*** (0.059) -0.267	-0.166** (0.063) -0.190	-0.236*** (0.059) -0.271	-0.236*** (0.059) -0.271
EXD^*	-0.086 (0.044) -0.133	-0.085 (0.044) -0.132	-0.075 (0.044) -0.116	0.027 (0.047) 0.042	-0.090* (0.044) -0.139	-0.072 (0.044) -0.112	-0.007 (0.047) -0.011	-0.087* (0.044) -0.134	-0.087* (0.044) -0.134
LEM	0.084*** (0.016) 0.324	0.083*** (0.016) 0.318	0.085*** (0.016) 0.326	0.024 (0.017) 0.092	0.082*** (0.016) 0.317	0.080*** (0.016) 0.308	0.033* (0.017) 0.127	0.078*** (0.016) 0.299	0.078*** (0.016) 0.299
N	11,644	11,644	11,644	11,644	11,644	11,644	11,644	11,644	11,644
R_p^2	0.142	0.142	0.147	0.231	0.142	0.147	0.225	0.144	0.144
LL	-3,261	-3,259	-3,241	-2,921	-3,259	-3,242	-2,943	-3,253	-3,253
χ^2	1,012	1,009	1,034	1,445	1,009	1,015	1,364	1,011	1,011
AIC	6,607	6,604	6,569	5,929	6,604	6,569	5,972	6,593	6,593
BIC	6,916	6,920	6,886	6,245	6,921	6,886	6,289	6,909	6,909
P_{corr}	90.41	90.41	90.43	91.13	90.36	90.28	91.24	90.46	90.46

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (β). La variable $I\hat{I}D$ está construida a partir del Modelo (a), Cuadro 4.2. En todos los casos se controla con *dummies* por sector y por entidad. El asterisco en las variables indica que son variables *dummy*. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Cuadro C.3: Estado de México. Propensión a innovar en productos (Modelo (a))

Variable Dependiente: Innovación en productos (INN_1)							
Variables	<i>Spillovers</i> intersectoriales				<i>Spillovers</i> intrasectoriales		
	Sin $SPILL$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$
$I\hat{I}D$	0.580*** (0.083)	0.493*** (0.084)	0.488*** (0.088)	0.367*** (0.084)	0.580*** (0.084)	0.620*** (0.084)	0.330*** (0.086)
	2.417	2.057	2.033	1.528	2.417	2.582	1.374
$LSPILL$	-	-115.689** (39.692)	1.153*** (0.313)	0.157*** (0.014)	0.059* (0.024)	-0.255 (0.137)	0.183*** (0.017)
		-4.622	0.956	1.774	0.997	-1.659	1.699
FO^*	-0.271 (0.162)	-0.332 (0.173)	-0.142 (0.165)	0.043 (0.187)	-0.276 (0.163)	-0.270 (0.166)	0.008 (0.184)
	-0.320	-0.392	-0.167	0.051	-0.326	-0.318	0.010
EXD^*	-0.125 (0.133)	-0.136 (0.136)	-0.281* (0.138)	-0.458** (0.164)	-0.128 (0.135)	-0.148 (0.132)	-0.347* (0.155)
	-0.207	-0.224	-0.464	-0.757	-0.212	-0.244	-0.573
LEM	0.166*** (0.048)	0.134** (0.050)	0.196*** (0.051)	0.171** (0.053)	0.159*** (0.048)	0.142** (0.049)	0.181*** (0.051)
	0.636	0.513	0.751	0.655	0.609	0.541	0.691
N	1,488	1,488	1,488	1,488	1,488	1,488	1,488
R_p^2	0.199	0.227	0.270	0.391	0.208	0.207	0.370
LL	-357	-345	-326	-271	-353	-354	-281
χ^2	158	145	205	249	146	167	243
AIC	740	717	679	571	734	735	589
BIC	809	792	753	645	808	810	664
P_{corr}	91.67	91.33	91.80	94.02	91.67	91.60	93.21

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (*beta*). La variable $I\hat{I}D$ está construida a partir del Modelo (a), Cuadro 4.2. En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. *** p<0.01, ** p<0.05, * p<0.1

Cuadro C.4: Nuevo León. Propensión a innovar en productos (Modelo (a))

Variable Dependiente: Innovación en productos (INN_1)

Variables	<i>Spillovers</i> intersectoriales				<i>Spillovers</i> intrasectoriales		
	Sin <i>SPILL</i>	<i>SPILL</i> ₀	<i>SPILL</i> _{ω}	<i>SPILL</i> _{$\hat{\omega}$}	<i>SPILL</i> ₀	<i>SPILL</i> _{ω}	<i>SPILL</i> _{$\hat{\omega}$}
$\hat{I}D$	0.458*** (0.068)	0.441*** (0.069)	0.401*** (0.073)	0.247*** (0.074)	0.458*** (0.068)	0.461*** (0.068)	0.248*** (0.073)
	2.167	2.086	1.897	1.168	2.167	2.182	1.173
<i>LSPILL</i>	-	-43.995 (22.865)	0.942*** (0.240)	0.138*** (0.014)	-0.001 (0.023)	-0.223 (0.178)	0.158*** (0.016)
		-2.196	0.726	1.562	-0.013	-1.711	1.503
<i>FO</i> *	-0.622** (0.191)	-0.562** (0.192)	-0.683*** (0.195)	-0.785*** (0.177)	-0.622** (0.191)	-0.621** (0.193)	-0.787*** (0.179)
	-0.654	-0.591	-0.719	-0.826	-0.655	-0.654	-0.829
<i>EXD</i> *	0.348** (0.123)	0.352** (0.125)	0.373** (0.124)	0.543*** (0.135)	0.348** (0.123)	0.324** (0.124)	0.551*** (0.135)
	0.512	0.518	0.549	0.800	0.513	0.477	0.812
<i>LEM</i>	0.215*** (0.047)	0.164** (0.051)	0.183*** (0.050)	0.107* (0.052)	0.215*** (0.047)	0.217*** (0.048)	0.122* (0.051)
	0.766	0.584	0.649	0.382	0.765	0.771	0.432
N	1,149	1,149	1,149	1,149	1,149	1,149	1,149
R_p^2	0.210	0.226	0.250	0.351	0.210	0.214	0.341
LL	-347	-340	-329	-285	-347	-345	-290
χ^2	180	162	160	209	180	183	203
<i>AIC</i>	720	708	687	599	722	719	607
<i>BIC</i>	786	779	758	669	793	790	678
P_{corr}	87.73	88.25	88.08	91.21	87.82	87.73	91.04

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (*beta*). La variable $\hat{I}D$ está construida a partir del Modelo (a), Cuadro 4.2. En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. *** p<0.01, ** p<0.05, * p<0.1

Cuadro C.5: Nacional. Propensión a innovar en productos (Modelo (b))

Variable Dependiente: Innovación en productos (INN_1)

Variables	<i>Spillovers</i> intersectoriales				<i>Spillovers</i> intrasectoriales				
	Sin $SPILL$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_g$	$SPILL_v$
$I\hat{I}D$	0.811*** (0.032) 2.946	0.807*** (0.032) 2.930	0.795*** (0.033) 2.888	0.512*** (0.038) 1.861	0.805*** (0.032) 2.926	0.778*** (0.033) 2.826	0.536*** (0.037) 1.947	0.806*** (0.032) 2.928	0.806*** (0.032) 2.928
$LSPILL$	-	-0.188 (0.122) -1.355	0.341** (0.106) 2.248	0.119*** (0.005) 1.283	0.009 (0.005) 0.158	0.337*** (0.079) 2.189	0.118*** (0.005) 1.634	-1.980 (1.321) -12.358	-1.977 (1.318) -12.339
FO^*	-0.291*** (0.059) -0.333	-0.289*** (0.059) -0.331	-0.289*** (0.059) -0.330	-0.169** (0.063) -0.194	-0.286*** (0.059) -0.328	-0.283*** (0.059) -0.324	-0.202** (0.063) -0.231	-0.288*** (0.059) -0.330	-0.288*** (0.059) -0.330
EXD^*	-0.130** (0.045) -0.201	-0.129** (0.045) -0.199	-0.119** (0.045) -0.184	-0.012 (0.048) -0.018	-0.134** (0.045) -0.207	-0.116* (0.045) -0.180	-0.047 (0.047) -0.073	-0.130** (0.045) -0.202	-0.130** (0.045) -0.202
LEM	0.093*** (0.016) 0.358	0.092*** (0.016) 0.352	0.093*** (0.016) 0.358	0.031 (0.017) 0.118	0.091*** (0.016) 0.351	0.089*** (0.016) 0.341	0.039* (0.017) 0.151	0.087*** (0.016) 0.333	0.087*** (0.016) 0.333
N	11,644	11,644	11,644	11,644	11,644	11,644	11,644	11,644	11,644
R_p^2	0.159	0.159	0.164	0.239	0.159	0.163	0.235	0.161	0.161
LL	-3,197	-3,195	-3,178	-2,892	-3,195	-3,180	-2,909	-3,189	-3,189
χ^2	1,101	1,097	1,113	1,408	1,097	1,093	1,326	1,099	1,099
AIC	6,477	6,475	6,443	5,869	6,475	6,446	5,903	6,464	6,464
BIC	6,786	6,792	6,759	6,186	6,792	6,763	6,220	6,781	6,781
P_{corr}	90.47	90.47	90.57	91.33	90.51	90.48	91.45	90.45	90.45

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (β). La variable $I\hat{I}D$ está construida a partir del Modelo (a), Cuadro 4.2. En todos los casos se controla con *dummies* por sector y por entidad. El asterisco en las variables indica que son variables *dummy*. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Cuadro C.6: Estado de México. Propensión a innovar en productos (Modelo (b))

Variable Dependiente: Innovación en productos (INN_1)

Variables	<i>Spillovers</i> intersectoriales				<i>Spillovers</i> intrasectoriales		
	Sin $SPILL$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$
$I\hat{I}D$	0.875*** (0.132)	0.740*** (0.143)	0.819*** (0.147)	0.490*** (0.141)	0.857*** (0.131)	0.914*** (0.132)	0.444** (0.141)
	2.999	2.537	2.808	1.679	2.939	3.134	1.522
$LSPILL$	-	-116.799** (39.583)	1.231*** (0.300)	0.159*** (0.013)	0.054* (0.024)	-0.223 (0.138)	0.185*** (0.016)
		-4.667	1.021	1.788	0.909	-1.453	1.716
FO^*	-0.273 (0.165)	-0.333 (0.178)	-0.175 (0.174)	0.048 (0.195)	-0.276 (0.168)	-0.267 (0.168)	0.015 (0.190)
	-0.322	-0.393	-0.206	0.057	-0.325	-0.315	0.018
EXD^*	-0.460** (0.170)	-0.407* (0.176)	-0.623*** (0.176)	-0.605** (0.187)	-0.448** (0.170)	-0.493** (0.167)	-0.480** (0.180)
	-0.761	-0.674	-1.030	-1.000	-0.741	-0.815	-0.795
LEM	0.207*** (0.046)	0.170*** (0.048)	0.230*** (0.050)	0.192*** (0.051)	0.203*** (0.046)	0.190*** (0.047)	0.198*** (0.049)
	0.792	0.649	0.877	0.734	0.775	0.725	0.758
N	1,488	1,488	1,488	1,488	1,488	1,488	1,488
R_p^2	0.194	0.220	0.274	0.385	0.201	0.200	0.365
LL	-359	-348	-324	-274	-356	-357	-283
χ^2	168	149	201	252	153	182	248
AIC	745	723	675	576	740	742	594
BIC	813	797	750	651	814	816	668
P_{corr}	91.47	91.94	92.14	94.29	91.67	91.53	93.21

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (β). La variable $I\hat{I}D$ está construida a partir del Modelo (a), Cuadro 4.2. En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Cuadro C.7: Nuevo León. Propensión a innovar en productos (Modelo (b))

Variable Dependiente: Innovación en productos (INN_1)							
Variables	<i>Spillovers</i> intersectoriales				<i>Spillovers</i> intrasectoriales		
	Sin $SPILL$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$	$SPILL_0$	$SPILL_\omega$	$SPILL_{\hat{\omega}}$
$\hat{I}D$	0.733*** (0.098)	0.714*** (0.098)	0.657*** (0.105)	0.415*** (0.106)	0.738*** (0.100)	0.748*** (0.100)	0.417*** (0.105)
	3.178	3.094	2.848	1.799	3.199	3.242	1.809
$LSPILL$	-	-43.883 (23.500)	0.901*** (0.250)	0.129*** (0.014)	-0.015 (0.023)	-0.292 (0.183)	0.148*** (0.017)
		-2.190	0.694	1.465	-0.215	-2.239	1.403
FO^*	-0.734*** (0.188)	-0.670*** (0.187)	-0.779*** (0.192)	-0.835*** (0.175)	-0.740*** (0.189)	-0.733*** (0.191)	-0.839*** (0.176)
	-0.772	-0.705	-0.820	-0.879	-0.779	-0.771	-0.883
EXD^*	0.507*** (0.119)	0.507*** (0.121)	0.515*** (0.118)	0.624*** (0.131)	0.516*** (0.119)	0.474*** (0.120)	0.632*** (0.131)
	0.746	0.747	0.759	0.918	0.760	0.698	0.931
LEM	0.225*** (0.048)	0.176*** (0.052)	0.194*** (0.050)	0.120* (0.052)	0.223*** (0.048)	0.229*** (0.049)	0.133** (0.051)
	0.802	0.627	0.688	0.425	0.792	0.813	0.473
N	1,149	1,149	1,149	1,149	1,149	1,149	1,149
R_p^2	0.249	0.264	0.284	0.361	0.250	0.255	0.351
LL	-330	-324	-315	-281	-330	-327	-285
χ^2	199	180	165	201	199	203	198
AIC	686	675	657	590	687	682	598
BIC	751	746	728	660	758	753	669
P_{corr}	89.99	90.08	90.25	91.38	89.47	89.38	91.04

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (β). La variable $\hat{I}D$ está construida a partir del Modelo (a), Cuadro 4.2. En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Cuadro C.8: Nacional. Propensión a innovar en productos (INN_1), productos a nivel mercado (INN_2) y procesos (INN_3)

Variables Dependientes: INN_1 , INN_2 e INN_3

Variables	INN_1		INN_2		INN_3	
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{I\hat{D}}^a$	0.872*** (0.033)	0.618*** (0.037)	0.830*** (0.032)	0.567*** (0.036)	0.869*** (0.032)	0.719*** (0.037)
	3.305	2.339	3.349	2.287	3.848	3.183
$LSPILL_{\omega}^b$	-	0.108*** (0.005)	-	0.106*** (0.005)	-	0.062*** (0.006)
		1.156		1.211		0.782
FO^*	-0.333*** (0.061)	-0.212*** (0.064)	-0.336*** (0.063)	-0.211** (0.065)	-0.262*** (0.067)	-0.168* (0.067)
	-0.381	-0.242	-0.409	-0.257	-0.354	-0.226
EXD^*	-0.220*** (0.046)	-0.103* (0.049)	-0.106* (0.048)	0.023 (0.050)	-0.238*** (0.052)	-0.182*** (0.053)
	-0.340	-0.160	-0.174	0.039	-0.435	-0.332
LEM	0.101*** (0.016)	0.042* (0.017)	0.125*** (0.017)	0.066*** (0.018)	0.082*** (0.018)	0.044* (0.018)
	0.389	0.161	0.513	0.271	0.371	0.200
N	11,644	11,644	11,615	11,615	11,610	11,610
R_p^2	0.200	0.262	0.213	0.277	0.219	0.241
LL	-3,038	-2,805	-2,710	-2,490	-2,302	-2,238
χ^2	1,098	1,365	1,171	1,421	1,064	1,124
AIC	6,160	5,695	5,502	5,064	4,685	4,561
BIC	6,469	6,012	5,804	5,373	4,987	4,870

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados (*beta*). En todos los casos se controla con *dummies* por sector y por entidad. El asterisco en las variables indica que son variables *dummy*. ^a $\hat{I\hat{D}}$ estimado con el Modelo (d) que utiliza mecanismos de financiamiento. ^b *Spillovers* intersectoriales. *** p<0.01, ** p<0.05, * p<0.1

Cuadro C.9: Estado de México. Propensión a innovar en productos (INN_1), productos a nivel mercado (INN_2) y procesos (INN_3)

Variables Dependientes: INN_1 , INN_2 e INN_3						
Variables	INN_1		INN_2		INN_3	
	(1)	(2)	(3)	(4)	(5)	(6)
$I\hat{I}D^a$	0.214*** (0.049)	0.061 (0.047)	0.241*** (0.051)	0.104* (0.048)	0.199*** (0.044)	0.113** (0.040)
$LSPILL_{\omega}^b$	-	0.166*** (0.013)	-	0.156*** (0.013)	-	0.120*** (0.012)
FO^*	-0.093 (0.155)	0.165 (0.184)	-0.065 (0.159)	0.177 (0.185)	-0.052 (0.152)	0.044 (0.172)
EXD^*	-0.110 (0.126)	0.195 (0.158)	-0.083 (0.131)	0.226 (0.156)	-0.057 (0.121)	0.048 (0.143)
LEM	0.036 (0.126)	-0.315* (0.158)	0.113 (0.131)	-0.177 (0.156)	-0.361** (0.121)	-0.589*** (0.143)
	0.059 (0.046)	-0.521 (0.052)	0.202 (0.049)	-0.317 (0.054)	-0.558 (0.043)	-0.911 (0.047)
	0.213*** (0.046)	0.195*** (0.052)	0.241*** (0.049)	0.233*** (0.054)	0.100* (0.043)	0.062 (0.047)
	0.815	0.744	0.995	0.963	0.360	0.223
N	1,488	1,488	1,488	1,488	1,456	1,456
R_p^2	0.156	0.375	0.174	0.389	0.072	0.174
LL	-376	-279	-326	-241	-446	-397
χ^2	137	244	154	243	81	150
AIC	778	585	678	510	916	820
BIC	847	659	747	585	980	889

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados ($beta$). En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. ^a $I\hat{I}D$ estimado con el Modelo (c) que utiliza programas de financiamiento. ^b *Spillovers* intersectoriales. *** p<0.01, ** p<0.05, * p<0.1

Cuadro C.10: Nuevo León. Propensión a innovar en productos (INN_1), productos a nivel mercado (INN_2) y procesos (INN_3)

Variables Dependientes: INN_1 , INN_2 e INN_3

Variables	INN_1		INN_2		INN_3	
	(1)	(2)	(3)	(4)	(5)	(6)
$I\hat{I}D^a$	0.854*** (0.101)	0.544*** (0.113)	0.849*** (0.100)	0.537*** (0.106)	0.597*** (0.089)	0.363*** (0.093)
	3.624	2.309	3.786	2.393	2.977	1.808
$LSPILL_{\omega}^b$	-	0.118*** (0.014)	-	0.110*** (0.015)	-	0.078*** (0.015)
		1.344		1.309		1.046
FO^*	-0.774*** (0.191)	-0.855*** (0.176)	-0.748*** (0.195)	-0.809*** (0.177)	-1.016*** (0.227)	-0.945*** (0.219)
	-0.815	-0.900	-0.827	-0.895	-1.256	-1.168
EXD^*	0.475*** (0.123)	0.594*** (0.133)	0.618*** (0.126)	0.760*** (0.141)	0.490*** (0.132)	0.563*** (0.138)
	0.700	0.875	0.957	1.176	0.847	0.974
LEM	0.218*** (0.050)	0.124* (0.052)	0.236*** (0.052)	0.140* (0.055)	0.202*** (0.053)	0.119* (0.054)
	0.777	0.440	0.882	0.524	0.843	0.496
N	1,149	1,149	1,149	1,149	1,149	1,149
R_p^2	0.292	0.377	0.306	0.383	0.251	0.289
LL	-311	-274	-283	-251	-258	-245
χ^2	209	190	218	209	154	157
AIC	648	575	592	531	542	518
BIC	714	646	657	601	608	588

Modelo Probit. Efectos marginales con $FO = 0$, $EXD = 0$ y el resto de las variables en la media. Errores estándar robustos entre paréntesis. En el tercer renglón coeficientes estandarizados ($beta$). En todos los casos se controla con *dummies* por sector. El asterisco en las variables indica que son variables *dummy*. ^a $I\hat{I}D$ estimado con el Modelo (d) que utiliza mecanismos de financiamiento. ^b *Spillovers* intersectoriales. *** p<0.01, ** p<0.05, * p<0.1

Índice de cuadros

3.1. Gastos en actividades relacionadas con la innovación y madurez tecnológica de las empresas (promedio)	33
3.2. Promedio de los diferentes tipos de <i>spillovers</i> (en miles de pesos)	36
4.1. Estadísticas descriptivas de las características de las empresas	39
4.2. Decisión e intensidad de gasto en I&D	41
4.3. Propensión a innovar en productos con $LSPILL_{\hat{\omega}}$	49
5.1. NL y Edo. Méx.: Decisión e intensidad de gasto en I&D	54
5.2. NL y Edo. Méx.: Propensión a innovar en productos	59
A.1. Mejor modelo según la diferencia en valores absolutos de los <i>BIC</i>	77
B.1. Descripción de las variables	78
C.1. Intensidad del gasto en I&D: programas (<i>PFIN</i>) y mecanismos (<i>MFIN</i>) de financiamiento	82
C.2. Nacional. Propensión a innovar en productos (Modelo (a))	83
C.3. Estado de México. Propensión a innovar en productos (Modelo (a))	84
C.4. Nuevo León. Propensión a innovar en productos (Modelo (a))	85
C.5. Nacional. Propensión a innovar en productos (Modelo (b))	86
C.6. Estado de México. Propensión a innovar en productos (Modelo (b))	87
C.7. Nuevo León. Propensión a innovar en productos (Modelo (b))	88
C.8. Nacional. Propensión a innovar en productos (INN_1), productos a nivel mercado (INN_2) y procesos (INN_3)	89

C.9. Estado de México. Propensión a innovar en productos (INN_1), productos a nivel mercado (INN_2) y procesos (INN_3)	90
C.10. Nuevo León. Propensión a innovar en productos (INN_1), productos a nivel mercado (INN_2) y procesos (INN_3)	91

SEGUNDA PARTE

**Fuerzas Marshallinas y coaglomeración económica:
estimaciones econométricas para México, 2003-2008**

Índice

SEGUNDA PARTE. Fuerzas Marshallianas y coaglomeración económica: estimaciones econométricas para México, 2003-2008	95
Introducción	99
1. Las fuerzas Marshallianas y la coaglomeración económica	103
1.1. Antecedentes empíricos de la coaglomeración económica	108
2. Índice de coaglomeración económica	111
2.1. Datos	111
2.2. Descripción del índice de coaglomeración	112
2.3. Patrones de coaglomeración	112
3. Fuerzas Marshallianas: <i>Bienes, Ocupación e Ideas</i>	116
3.1. Variable <i>proxy</i> : <i>Bienes</i>	116
3.1.1. Estimación de la variable <i>Bienes</i>	117
3.2. Variable <i>proxy</i> : <i>Ocupación</i>	118
3.2.1. Estimación de las variables <i>Ocupación y Escolaridad</i>	119
3.3. Variable <i>proxy</i> : <i>Ideas</i>	120
3.3.1. Estimación de la variable <i>Ideas</i>	121
4. Estimación de la importancia relativa de las fuerzas Marshallianas en la coaglomeración económica	123

4.1. Especificación econométrica	123
4.2. Resultados	124
Conclusiones	131
Bibliografía	131
A. Definición de sectores, ocupaciones, entidades y regiones	136
B. Método de regionalización indirecto de matrices insumo-producto	140
C. Regresiones adicionales	144
Índice de cuadros	146
Índice de figuras	147

Introducción

La actividad económica se caracteriza por estar espacialmente concentrada. Algunas ciudades tienden a especializarse en una industria, mientras que en otras su actividad económica está diversificada. Generalmente, lo que se observa es una gama intermedia, es decir, existen ciudades que concentran dos o más industrias sin llegar a la diversificación o especialización total. Este fenómeno se puede explicar por las ventajas que surgen de la aglomeración económica, como la disminución de los costos de transporte de bienes finales o intermedios, disponer de un mercado laboral amplio y compartir información e ideas que ayudan a la innovación tecnológica de las empresas, a lo que se conoce como fuerzas de aglomeración Marshallianas.

De lo anterior surge la interrogante de en qué medida las industrias de diferentes giros se agrupan en el espacio y cuáles son los factores relacionados con la concentración económica más relevantes. El objetivo de este trabajo es medir la aglomeración económica, a través de un índice de coaglomeración, y evaluar la importancia relativa que, sobre la coaglomeración, tienen las fuerzas o externalidades Marshallianas.

La información para construir el índice de coaglomeración y las variables que representan las fuerzas de concentración económica Marshallianas se obtiene de los Censos Económicos 2004 y 2009, la Encuesta Nacional de Empleo Urbano (ENEU) 2003 y la Encuesta Nacional de Ocupación y Empleo (ENOE) 2008, del Instituto Nacional de Estadística y Geografía (INEGI).

Con el propósito de estudiar el efecto de las externalidades Marshallianas, se realiza el ejercicio a nivel entidad federativa, lo cual nos permite controlar por las características relacionadas a la localización geográfica. A diferencia de los trabajos anteriores, estimamos

matrices insumo-producto regionales para aprovechar la desagregación espacial de la información a nivel entidad federativa. Además, utilizamos una desagregación sectorial de tres dígitos (10 subsectores) del Sistema de Clasificación Industrial de América del Norte (SCIAN) para el año 2003 y 2008.

En la literatura de la geografía económica se han desarrollado principalmente dos formas de medir el grado de concentración. Por un lado, se encuentran los índices de aglomeración, que estiman el grado en que una industria se agrupa espacialmente.¹ Por otro lado, los índices de coaglomeración miden la concentración de industrias diferentes en una misma región.² La ventaja de utilizar el concepto de coaglomeración es que permite estudiar la concentración económica considerando la existencia de más de una industria en un área geográfica, hecho más cercano a lo observado en las ciudades.

Entre los índices de coaglomeración más utilizados está el propuesto por Ellison y Glaeser (1997) y Ellison et al. (2010), el cual mide el grado en que se concentran dos industrias diferentes. A partir de este índice podemos conocer el patrón de coaglomeración de las empresas y así estudiar cuáles son los incentivos que tiene en común las diferentes industrias para ubicarse en la misma región.

La localización, de más de una industria en una zona determinada, se podría explicar por la existencia de diversos factores que las lleva a ubicarse una cerca de otra. Ellos dependen de las ventajas que pueden obtener las empresas al instalarse cerca de otras que pertenecen a una industria diferente. Dichas ventajas surgen de las externalidades derivadas de las relaciones entre clientes y proveedores que reducen los costos de producción o de compartir la fuerza laboral de la región y de un ambiente en donde los conocimientos e ideas fluyan entre las industrias localizadas en una misma área.

Las fuerzas Marshallianas que vamos abordar en este trabajo son las externalidades que se generan a través de la compra y venta de bienes, del mercado laboral y de los flujos de conocimientos. Para estudiar su efecto sobre la concentración económica, correlacionamos

¹Algunos índices de aglomeración han sido propuestos por Hoover (1948), Krugman (1991), Ellison y Glaeser (1997), Maurel y Sédillot (1999) y Duranton y Overman (2005).

²Los índices de coaglomeración fueron propuestos por Ellison y Glaeser (1997) y los autores Ellison et al. (2010) hacen una adaptación del índice de Duranton y Overman (2005) para medir la coaglomeración.

las características de las externalidades Marshallianas que comparten las industrias ubicadas en una misma región

La relación de bienes se estima con información de la matriz insumo-producto, de donde obtenemos la proporción de insumos que compra una industria al resto de la economía y también la proporción de ventas. Con las proporciones se estima la relación en el mercado de bienes entre las industrias y de ésta forma, se detecta si comparten proveedores o clientes.

Para aproximar las relaciones en el mercado laboral se utiliza el tipo de ocupación que tienen los empleados en las diferentes industrias. Con la proporción de trabajadores en cada tipo de ocupación sobre el total de empleados de una industria se estima la relación con otras industrias. El objetivo es aproximar el grado de similitud entre los tipos de trabajadores que se ocupan en las diferentes industrias de la región. Alternativamente, con los años de escolaridad, también se correlacionan las proporciones de empleados con determinado nivel de educación entre las diferentes industrias.

A través del grado de correlación de las características de las actividades de innovación entre las empresas se busca estimar los flujos de ideas o *spillovers* de conocimientos. El supuesto es que a mayor correlación en sus características tecnológicas se facilita el flujo de conocimientos en esa región.

La estimación de las fuerzas Marshallianas a través de correlaciones supone que a mayor grado de correlación, más similares entre ellas, es más fácil que las empresas de los diferentes sectores aprovechen las externalidades derivadas de las economías de aglomeración. A través del análisis de la importancia de las fuerzas de aglomeración Marshallianas nos aproximaremos a las posibles explicaciones de la concentración de la actividad económica en ciertas regiones del país.

Las principales conclusiones del trabajo son que la fuerza de aglomeración económica con mayor importancia sobre la coaglomeración es la relación a través del mercado de bienes (cliente-proveedor), entre los diferentes subsectores de la manufactura en México. En la región Norte, Golfo y Sur también sobresale la importancia del mercado de bienes como incentivo para la coaglomeración.

La estructura de la segunda parte de la tesis es la siguiente: en el primer capítulo se presentan antecedentes de la coaglomeración económica y las fuerzas Marshallianas. En el segundo capítulo se describe el índice de coaglomeración y se presentan los patrones de coaglomeración económica. El tercer capítulo incluye la descripción de las variables de las fuerzas Marshallianas y sus estimaciones. El cuarto capítulo contiene las estimaciones de la importancia relativa de las fuerzas Marshallianas en la coaglomeración económica. Las conclusiones del trabajo están al final.

Capítulo 1

Las fuerzas Marshallianas y la coaglomeración económica

En este capítulo se plantean las definiciones de aglomeración y coaglomeración económica, debido a que ambas miden la concentración de la actividad económica en una región, pero la segunda considera la presencia de más de un tipo de industria. Además, se relaciona la aglomeración económica con las posibles explicaciones de este fenómeno, es decir, presentamos las ventajas que las regiones ofrecen para que ahí se localicen las empresas. Las fuerzas de aglomeración Marshallianas explican la concentración económica y también se definen en este capítulo. Por último, se presenta una revisión de la literatura empírica relacionada con la coaglomeración económica.

La aglomeración económica se define como la concentración geográfica de las actividades productivas en las ciudades, la cual también se observa en la formación de agrupamientos industriales o centros de trabajo (Strange, 2008). La actividad económica se concentra en ciertas regiones, donde el nivel de aglomeración es mayor al que se esperaría si su localización hubiera sido asignada aleatoriamente o si reflejara sólo las diferencias derivadas de las ventajas comparativas entre las áreas geográficas (Ellison y Glaeser, 1997, 1999; Duranton y Overman, 2005).

Existen diversas herramientas para conocer la magnitud de la aglomeración económica, entre ellas encontramos los índices de concentración de Gini y Theil.¹ Entre los índices más utilizados está el propuesto por Ellison y Glaeser (1997), que miden el grado de concentración de una industria más allá de lo que se observaría si las empresas se ubicaran aleatoriamente. La ventaja de este índice es que controla por las diferencias en el tamaño de la distribución de las empresas y por el tamaño de las áreas geográficas.² Para analizar los patrones de concentración se ha utilizado el índice de Duranton y Overman (2005), basado en la distancia de localización de las empresas. Con esta medida se puede comprobar la significancia estadística del índice de aglomeración, en donde, el espacio es continuo, en lugar de utilizar un área geográfica delimitada.

Los índices de aglomeración miden la concentración de una industria en una región determinada, sin embargo, nos interesa conocer la concentración de más de una industria. La alternativa es utilizar una medida de coaglomeración, que estime el grado de aglomeración de más de una industria en un área geográfica. Ellison y Glaeser (1997) proponen un índice de coaglomeración cuya principal ventaja es que refleja en mayor medida la distribución real de la actividad económica.

El modelo en el que subyace el índice de coaglomeración de Ellison y Glaeser (1997) parte de la maximización de beneficios que realizan las empresas, a partir de las ventajas naturales de la región, las externalidades tecnológicas y la presencia de empresas de diferentes industrias en la región.³ A partir del modelo, la coaglomeración económica se define como la concentración de diferentes industrias en una región.

Abordar el tema de la concentración de la actividad económica desde ésta perspectiva permite estudiar la agrupación de diferentes industrias en una región. Debido a que la distribución de la actividad económica en una región puede ir desde la especialización hasta la

¹Algunos trabajos que utilizan este tipo de índices son los realizados por Brülhart y Torstensson (1996); Amiti (1999); Brülhart y Traeger (2005); Cutrini (2009) entre otros.

²Por su parte, Maurel y Sédillot (1999) modifican el índice de Ellison y Glaeser (1997) para ver la correlación entre las decisiones de localización de dos empresas que pertenecen a la misma industria.

³Por ello, las ventajas que obtienen diferentes industrias al localizarse en la misma región pueden ser diversas y más adelante revisaremos algunas teorías al respecto.

diversificación industrial, este enfoque considera una estructura más compleja de la aglomeración económica. Además, cuando se toma en cuenta la agrupación de diferentes industrias se pueden estudiar las relaciones que surgen de su interacción en una región, y los incentivos que conllevan a este tipo de concentración económica.

La existencia de concentración de la actividad económica se explica por la presencia de rendimientos crecientes externos, también conocidos como economías de aglomeración (Strange, 2008). Las economías de aglomeración representan las ventajas productivas que las empresas obtienen de localizarse unas cerca de otras, y la concentración económica derivada de ellas en un mercado grande permite compartir infraestructura e instalaciones, disponer de una gran variedad de proveedores y de trabajadores calificados. Además, un mercado amplio facilita el aprendizaje, promoviendo el desarrollo y adopción de nuevas tecnologías y prácticas de negocios (Puga, 2010).

Dentro del estudio de las economías de aglomeración existen dos vertientes: las economías de urbanización y de localización. Cada uno de los enfoques determinan las ventajas productivas que obtienen las empresas al concentrarse en una región. Por una parte, las economías de urbanización asociadas al tamaño y diversidad de las ciudades, representan economías externas a la empresa y a la industria (Strange, 2008). Entre las ventajas derivadas de este tipo de economías están la infraestructura, las relacionadas con la naturaleza de un mercado grande en la ciudad, la capacidad de la ciudad de generar factores productivos y un mercado de insumos. Esta corriente sostiene que la creación de nuevos conocimientos, avances tecnológicos e innovaciones surgen a través de la interacción de agentes de diferentes industrias (Jacobs, 1969). También que la generación de derramas de conocimientos surge de la diversidad industrial (Gleaser et al., 1992).

Las economías de localización se relacionan a la concentración de una industria particular. Marshall (1890) se enfoca en aquéllos aspectos de la concentración geográfica de la industria que no se explican por la existencia de recursos naturales. Dentro de éstos, identifica tres razones por las que un grupo de empresas podrían ser más eficientes que una empresa individual: la presencia de proveedores especializados, un mercado de trabajo amplio y la difusión (*spillovers*) de conocimientos (Krugman y Obsfeld, 2005).

Un grupo de empresas que utilizan insumos especializados pueden sostener un mercado lo suficientemente grande para mantener una amplia diversidad de proveedores. Además, una segunda fuente de las economías de aglomeración es el modo en que un conjunto de empresas puede crear un mercado de trabajadores especializados. La ventaja de contar con un mercado laboral amplio es que permite suavizar choques negativos en la productividad, al facilitar que los trabajadores puedan encontrar empleo en otras empresas que no tienen problemas de baja productividad (Krugman y Obsfeld, 2005). Sucede algo parecido de parte de los empresarios, ellos podrían tener la oportunidad de contratar más trabajadores en temporadas de alta productividad. Un mercado laboral amplio permite compartir riesgos derivados de las fluctuaciones económicas (Marshall, 1890).

Cada empresa puede generar nueva tecnología a través de su propia investigación y desarrollo, sin embargo, al localizarse en una región con más empresas también aprende de sus competidores estudiando sus productos y en algunos casos copiando sus diseños. Se ha encontrado que una fuente importante del *know-how* es el intercambio informal de información e ideas entre los trabajadores, situación que se presenta cuando las empresas donde laboran se encuentran cerca (Krugman y Obsfeld, 2005). La aglomeración económica ofrece un ambiente apropiado para la difusión de conocimientos e innovaciones tecnológicas, conocido como *spillovers* de conocimientos o tecnológicos. También se pueden generar intercambios de ideas entre los directivos de negocios en las concentraciones industriales y aprovechar que los trabajadores aprenden rápidamente en este tipo de ambiente (Marshall, 1890).

Otra forma de plantear las externalidades Marshallianas es considerar los beneficios de la aglomeración a través de la reducción de costos de transporte de bienes, personas e ideas (Ellison et al., 2010). Las empresas se localizan cerca de otras para reducir los costos de obtener insumos o enviar sus productos. Si los insumos están lejos, las empresas tendrían que elegir entre la distancia a los clientes y a los proveedores, dependiendo del costo de transporte de la materia prima y los bienes finales (Marshall, 1890). Si es alto el costo de transportar los bienes finales, la empresa va a preferir localizarse en el mercado de bienes finales, mientras que si el costo de transportar los insumos de producción es mayor, entonces, preferirá ubicarse cerca de sus proveedores.

Las anteriores son algunas de las posibles explicaciones de la aglomeración económica que en adelante llamaremos fuerzas, externalidades o teorías Marshallianas de la aglomeración económica.

En general, Marshall aborda marginalmente el tema de coaglomeración económica cuando habla de industrias suplementarias localizadas en una misma región (Helsley y Strange, 2012). Un efecto de la suplementariedad de las industrias es la mitigación de crisis regionales, las cuales se podrían presentar cuando la región está especializada en una sola industria y ante algún choque económico negativo no existen otras industrias que puedan sostener la economía local. Por ello se considera a la coaglomeración como una forma de compartir el riesgo entre varias industrias.

Las fuerzas Marshallianas no son las únicas que generan y determinan la aglomeración económica. Existen otras como la disponibilidad de *amenities*, las ventajas naturales y las economías internas de escala con costos de transacción (Strange, 2008).⁴ Esto último ha sentado las bases para el desarrollo de la Nueva Geografía Económica, donde Krugman (1998) define dos fuerzas que promueven o desalientan la concentración geográfica. Las fuerzas centrípetas o economías externas Marshallianas, son los vínculos hacia atrás y hacia adelante que surgen en un mercado grande, así como un mercado laboral amplio para la mano de obra y los *spillovers* de información. Y las fuerzas centrífugas que representan la inmovilidad de factores, donde la concentración de la actividad económica incrementa la renta de la tierra desalentando el proceso de aglomeración (deseconomías externas).

La mayoría de la literatura no considera las fuerzas que incentivan la coaglomeración, ya que generalmente el análisis se centra en los casos extremos, especialización o diversidad (Helsley y Strange, 2012). Se han detectado algunas fuerzas que generan la coaglomeración y en este trabajo exploramos su relación con las externalidades Marshallianas.

⁴*Amenities* son los bienes y servicios que generan cierto bienestar o comodidades.

1.1. Antecedentes empíricos de la coaglomeración económica

Existen tres patrones de coaglomeración según Helsley y Strange (2012): la concentración por compartir ventajas naturales y *spillovers* tecnológicos (Ellison y Glaeser, 1997), la especialización según funciones (administración y servicios *vs* producción) que sustituye la especialización sectorial (Duranton y Puga, 2005) y la concentración por complementariedades de trabajadores con diferentes niveles de educación (Ciccone y Peri, 2006) o la diferenciación horizontal de trabajadores según su tipo de habilidad (Bacolod, Blum y Strange, 2009).

La evidencia confirma que existen incentivos para que industrias diferentes se ubiquen en la misma región o exista coaglomeración económica. Las ventajas productivas de concentrarse giran entorno a las externalidades tecnológicas entre las industrias, la oportunidad de contar con capital humano diverso así como la posibilidad de reducir costos al transitar de una especialización sectorial a una funcional.

Si bien existen estímulos para la coaglomeración económica, resulta complicado determinar sus causas, al respecto hay una amplia literatura que trata de identificar la importancia de las ventajas productivas que generan concentración económica. Algunos trabajos se enfocan en las ventajas de compartir proveedores en ciudades grandes y muy especializadas (Amiti y Cameron, 2007; Overman y Puga, 2010). Otros en la importancia de un mercado laboral amplio sobre las economías de aglomeración (Overman y Puga, 2010) y sobre la generación de mejores emparejamientos empleado-empleador (Gan y Li, 2004; Costa y Kahn, 2000). El flujo de conocimientos también se ha estudiado debido a su influencia sobre la decisión de ubicarse en regiones con concentración económica (Jaffe et al., 1993; Audretsch y Feldman, 1996).⁵

A partir de la relación entre la aglomeración económica y las características de los sectores se ha encontrado que las industrias más innovadoras tienden a estar más concentradas espacialmente y ante esto se supone la presencia de *spillovers* de conocimientos (Audretsch y Feldman, 1996). Explorando la misma relación e incorporando las externalidades Marsha-

⁵En el trabajo de Puga (2010) hay más ejemplos de este tipo de artículos.

llianas se observa que la fuerza de aglomeración más importante es contar con un mercado laboral amplio (Rosenthal y Strange, 2001).

En cuanto a la literatura de la coaglomeración, los trabajos de Ellison y Glaeser (1997) y Ellison et al. (2010) son los referentes. Ellison et al. (2010) relacionan los índices de coaglomeración con el grado en el cual, pares de industrias, comparten bienes, trabajadores e ideas. Para reducir la causalidad inversa utilizan datos de Reino Unido como instrumentos de los vínculos insumo-producto y ocupaciones de los trabajadores. También utilizan variables instrumentales para controlar aspectos espaciales, los cuales incluyen las características de la industria en diferentes lugares de Estados Unidos. Encuentran que todas las fuerzas Marshallianas explican en cierta medida los patrones de coaglomeración, destacando los vínculos insumo-producto. En el mismo contexto, Arhansya (2010) evalúa la importancia de las externalidades derivadas de las fuerzas Marshallianas sobre la coaglomeración espacial. A partir de sus estimaciones observa, para Indonesia, que el mercado laboral amplio y diverso es el de mayor efecto, seguido por las relaciones insumo-producto.

Kolko (2010) analiza si las industrias que comercian tienden a aglomerarse y si el efecto del comercio en la coaglomeración es diferente para la manufactura y los servicios. También estudia las fuerzas de Marshallianas (mercado laboral amplio y diverso, *spillovers* de conocimiento y compartir insumos) que afectan la coaglomeración. Entre sus principales resultados encuentra que la tecnología de información aumenta la coaglomeración de los servicios a nivel código postal, mientras que en la industria manufacturera lo hace a nivel código postal y municipal. Además, el mercado laboral amplio y diverso afecta en mayor medida a la manufactura, debido a que industrias con trabajadores similares tienden a coaglomerarse.

Los dos artículos siguientes tienen en común con el trabajo de Ellison et al. (2010) el uso de variables que relacionan las características de dos industrias, pero en estos buscan explicar la atracción de nuevas empresas a una región. Glaeser y Kerr (2009) estudian porqué en algunos lugares hay más empresarios que en otros. Para ello, analizan los factores que influyen en la decisión de localización de las empresas nuevas. Las variables que utilizan representan las fuerzas Marshallianas y las ventajas naturales frente a la entrada de empresas. Para controlar los problemas de sesgo utilizan efectos fijos por ciudad y por industria, y concluyen

que los factores más importantes en la concentración de nuevas empresas es la existencia de un mercado laboral amplio y diverso así como un mercado para los productos.

Por su parte, Jofre-Monseny et al. (2011) analizan la importancia relativa de los mecanismos Marshallianos de aglomeración, examinando la localización de nuevas empresas manufactureras en España. Estos autores, en lugar de usar una medida de aglomeración, utilizan como variable dependiente el número de empresas nuevas, mientras que las independientes son variables *proxy* de las fuerzas Marshallianas ponderadas por el nivel de empleo existente. Para evitar el problema de omisión de variables incluyen efectos fijos por localización. A partir de sus resultados concluyen que la fuerza Marshalliana relacionada con el mercado laboral amplio y diverso es la más importante al explicar la localización de las empresas nuevas.

Para el caso de México, hasta el momento, no se han encontrado trabajos que estudien la relación de la coaglomeración económica y las fuerzas de aglomeración Marshallianas. Con el objetivo de coadyuvar en el análisis de estas relaciones en la región, en este trabajo vamos a construir variables que aproximen las fuerzas Marshallianas y el índice de coaglomeración de Ellison et al. (2010).

Capítulo 2

Índice de coaglomeración económica

A través del índice de coaglomeración económica buscamos medir la concentración de más de una industria en una región y en este capítulo describimos el índice. Además, hacemos un primer acercamiento al comportamiento de la coaglomeración económica en el país a través del análisis del índice de coaglomeración por subsector económico, a nivel estatal y por región.

2.1. Datos

El estudio de las fuerzas marsallianas y la coaglomeración se hace para la industria manufacturera a nivel subsector de la actividad económica. Los subsectores tienen una desagregación de 3-dígitos del SCIAN (10 subsectores de la industria manufacturera).¹ La desagregación espacial es a nivel entidad federativa, aunque el índice de coaglomeración también utiliza información municipal.² Los datos para estimar el índice de coaglomeración se derivan de los Censos Económicos de 2004 y 2009 del INEGI, de donde también se obtiene la información sobre las actividades relacionadas con la innovación, para construir la variable *proxy* de ideas (*spillovers* de conocimientos). Para la externalidad relacionada con los bienes se estiman Matrices Insumo-Producto (MIP) estatales, a partir de las MIPs domésticas de los años 2003 y 2008, y del Producto Interno Bruto (PIB) 2003 y 2008 por entidad federativa, a

¹La clasificación de subsectores está en el Cuadro A.1 del Apéndice.

²El mapa con las entidades federativas del país están en la Figura A.1 del Apéndice.

precios constantes del 2003 (INEGI, 2015). La externalidad relacionada con los trabajadores se aproxima con información de la ENEU 2003 y la ENOE 2008, ambas corresponden al tercer trimestre del año.

2.2. Descripción del índice de coaglomeración

El índice de coaglomeración de Ellison et al. (2010) mide la tendencia de diferentes industrias a localizarse una cerca de otra. Para el cálculo utilizamos el personal ocupado total y por subsector de la industria manufacturera, a nivel municipal y por entidad federativa, de los Censos Económicos 2004 y 2009 (INEGI, 2015). El índice se calcula para todos los pares de los 10 subsectores de la industria manufacturera y por entidad federativa. El índice de coaglomeración de Ellison et al. (2010) es el siguiente:

$$Coagl_{ijs} = \frac{\sum_{m=1}^M (e_{mi} - E_m)(e_{mj} - E_m)}{1 - \sum_{m=1}^M E_m^2} \quad (2.1)$$

donde, i, j indexan los subsectores de la industria manufacturera, s representa la entidad federativa y m el municipio. $Coagl_{ijs}$ es el índice de coaglomeración de los subsectores i y j en la entidad federativa s . Se construye con la proporción del personal ocupado en la industria i contenida en cada municipio m (e_{mi}) y con la proporción del empleo municipal respecto al estatal (E_m).

Cuando el índice toma valores positivos, ubicados en la parte superior de su distribución, indica una alta coaglomeración entre los subsectores i y j de la industria manufacturera. Los valores negativos aparecen cuando los subsectores están aglomerados en diferentes áreas geográficas, es decir, no hay coaglomeración o es muy baja.

2.3. Patrones de coaglomeración

El índice de coaglomeración (Ecuación (2.1)) tiene una media de 0.016 con una desviación estándar cercana a 0.076 entre los años 2003 y 2008 a nivel nacional (Figura 2.1(a)). En éste

periodo se observa un aumento del valor máximo del índice. En cuanto a la distribución de los índices tenemos que la mayoría se concentran en las cercanías del punto de origen (Figura 2.1(c) y 2.1(d)). Esto se debe a la estructura del índice, donde la desviación de cada industria del contexto nacional, en general, no está correlacionado con la desviación de otras industrias (Ellison et al., 2010).

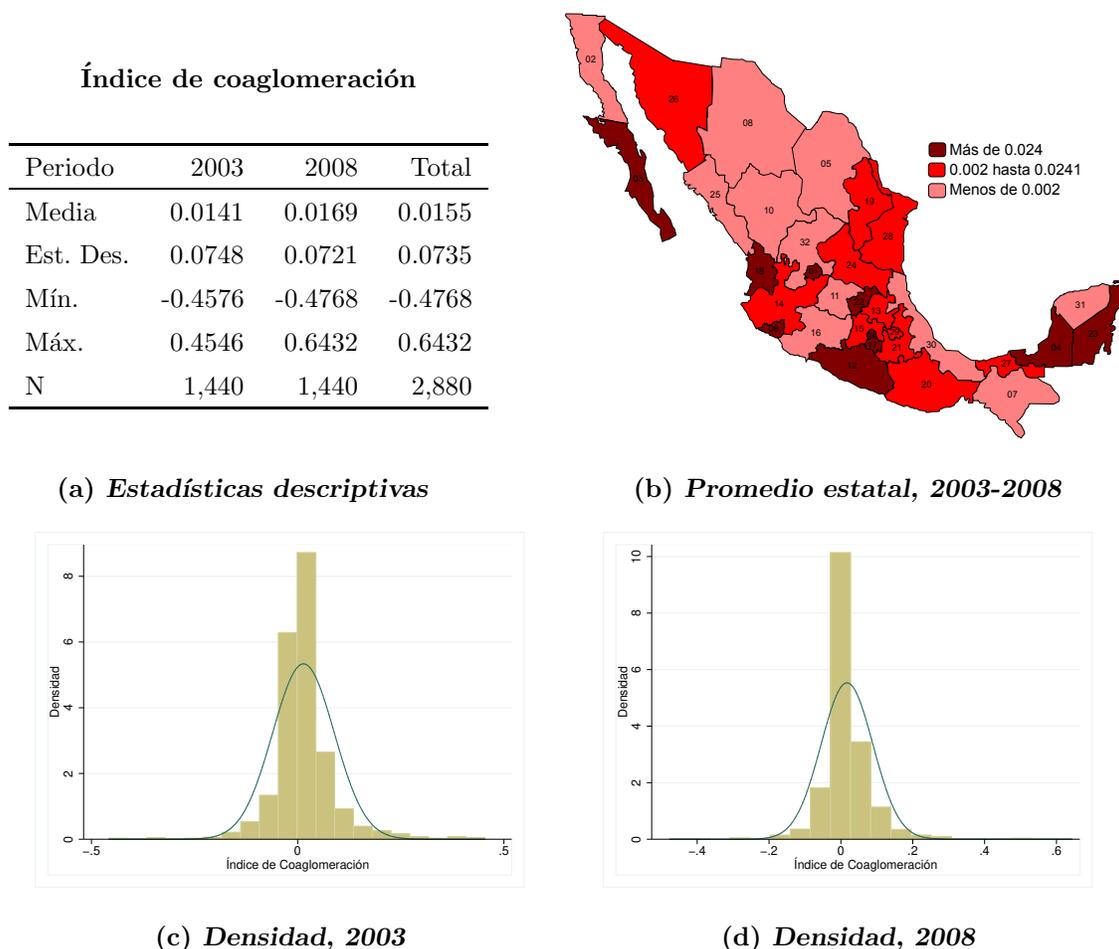


Figura 2.1: Características del índice de coaglomeración nacional, 2003-2008

En el mapa de México (Figura 2.1(b)) están representados los diferentes niveles de coaglomeración de las entidades federativas para los años 2003 y 2008. Los estados con mayor nivel de coaglomeración son Aguascalientes, Querétaro y Distrito Federal (DF). En el otro extremo vemos a varios estados del norte como Chihuahua, Durango, Sinaloa, Baja California

y Coahuila, con niveles bajos coaglomeración. Una baja coaglomeración indica que existe aglomeración de las industrias en diferentes regiones.

Una alternativa para estudiar los patrones de coaglomeración es a través de regiones, así que consideramos siete, divididas según su contigüidad geográfica y características económicas.³ El DF y el estado de México conforman la región Capital, presentan el mayor promedio del índice de coaglomeración y en el año 2008 es la región Centro Norte la que sobresale (Cuadro 2.1). En el otro extremo está la región Norte con la media más baja del índice de coaglomeración, donde el grueso de la producción se concentra en dos subsectores de la actividad económica, en la Industria metálica (VII) y de Maquinaria y equipo (VIII).

Cuadro 2.1: Índice de coaglomeración por regiones, 2003-2008

Período	2003		2008	
	Media	Des. Est.	Media	Des. Est.
Capital	0.0234	0.0298	0.0204	0.0257
Centro	0.0155	0.0579	0.0224	0.0612
Centro Nte.	0.0175	0.1000	0.0353	0.1015
Golfo	0.0114	0.1041	0.0128	0.0901
Norte	0.0050	0.0492	0.0020	0.0566
Pacífico	0.0179	0.0677	0.0202	0.0635
Sur	0.0150	0.0545	0.0056	0.0351
Total	0.0141	0.0748	0.0169	0.0721

En el Cuadro 2.2 se presenta el promedio de los índices de coaglomeración por subsector.⁴ El subsector de Maquinaria y equipo (VIII) presenta altos niveles de coaglomeración con los subsectores de Derivados del petróleo, carbón, industria química y del hule (V) y el de Otras industrias manufactureras (X). Mientras, el sector VIII también destaca por una baja coaglomeración con la Industria alimentaria, de bebidas y tabaco (I), la Industria de la madera (III) y la Fabricación de muebles y productos relacionados (IX). El grado de especialización del subsector de Maquinaria y equipo (VIII) refleja una localización fuerte respecto a industrias relacionadas a su actividad.

³En la Figura A.2 del Apéndice está la agrupación por regiones.

⁴La descripción de los subsectores está en el Cuadro A.1 del Apéndice.

Cuadro 2.2: Promedio nacional del índice de coaglomeración por subsectores de la industria manufacturera, 2003-2008

Sector i	Sector j									
	I	II	III	IV	V	VI	VII	VIII	IX	X
I	-	0.0258	0.0176	-0.0037	0.0143	0.0278	0.0112	-0.0083	0.0163	-0.0131
II	-	-	0.0331	0.0024	0.0214	0.0255	0.0093	-0.0001	0.0215	-0.0023
III	-	-	-	0.0003	0.0020	0.0172	0.0028	-0.0058	0.0365	0.0094
IV	-	-	-	-	0.0391	0.0154	0.0111	0.0094	0.0039	0.0189
V	-	-	-	-	-	0.0404	0.0280	0.0464	0.0170	0.0218
VI	-	-	-	-	-	-	0.0240	0.0119	0.0257	-0.0002
VII	-	-	-	-	-	-	-	0.0292	0.0149	0.0184
VIII	-	-	-	-	-	-	-	-	-0.0025	0.0518
IX	-	-	-	-	-	-	-	-	-	0.0119
X	-	-	-	-	-	-	-	-	-	-

Las parejas de subsectores más coaglomeradas sombreado rojo y menos coaglomeradas sombreado amarillo.

El subsector de Derivados del petróleo y del carbón (V) también presenta una alta coaglomeración con la Fabricación de productos a base de minerales no metálicos (VI) y con la Industria del papel, impresión e industrias conexas (IV). Este último sector tiene una baja coaglomeración con la Industria alimentaria (I).

En general, los patrones de coaglomeración económica en México son muy diversos. A partir del análisis regional se observa que la elección de una determinada área geográfica sí puede afectar los resultados del estudio. A mayor proximidad, las relaciones entre las industrias pueden cambiar. Por lo tanto, cuando se estima la relación entre las fuerzas Marshallianas y la coaglomeración consideramos las diferentes regiones que conforman el país.

Capítulo 3

Fuerzas Marshallianas: *Bienes*, *Ocupación e Ideas*

3.1. Variable *proxy*: *Bienes*

Entre los motivos por los cuales las empresas se localizan en la misma región está reducir los costos de obtener insumos de proveedores o de enviar sus productos al consumidor final. El objetivo de la variable *proxy Bienes* es estimar si existe una relación cliente-proveedor entre cada par de subsectores de la industria manufacturera, es decir, si entre ellos existen lazos comerciales a través del intercambio de insumos o productos. Para estimar esta fuerza Marshalliana se utiliza la MIP doméstica y el PIB de los años 2003 y 2008, información obtenida del INEGI (2015).

Un paso intermedio para calcular este indicador es la construcción de MIPs estatales y para la estimación de las matrices utilizamos la metodología de Flegg et al. (1995) y Flegg y Webber (1997).¹ De las matrices obtenemos información de los flujos insumo-producto a nivel de entidad federativa. La principal ventaja de realizar la regionalización de las MIPs es que incorporamos al análisis la dimensión espacial y así estudiar la importancia de las fuerzas Marshallianas a nivel regional.

¹En el Apéndice se presenta el método de regionalización indirecto de MIPs.

Para medir la relación comercial, a través de la compra o venta de bienes, entre las industrias utilizamos la variable *proxy Bienes* propuesta por Ellison et al. (2010):

$$Bienes_{ijs} = \max\{Insumo_{ijs}, Producto_{ijs}\} \quad (3.1)$$

donde, i, j indexan los subsectores de la industria manufacturera y s es la entidad federativa. $Bienes_{ijs}$ es la variable *proxy* para la relación insumo-producto, entre los subsectores i y j en la entidad federativa s y toma el valor máximo de $Insumo_{ijs}$ o $Producto_{ijs}$, estas variables se estiman de la siguiente forma:

$$Insumo_{ijs} = \max\{Insumo_{i\leftarrow j}, Insumo_{j\leftarrow i}\} \quad (3.2)$$

donde, $Insumo_{i\leftarrow j}$ es la proporción de los insumos del subsector i que vienen del subsector j , mientras que $Insumo_{j\leftarrow i}$ es la proporción de insumos del subsector j que se producen en i . Obtenemos éste valor al calcular las proporciones de insumos de todos los sectores respecto al PIB total del sector y después elegimos el valor máximo. La variable $Producto_{ijs}$ se calcula con la siguiente ecuación:

$$Producto_{ijs} = \max\{Producto_{i\rightarrow j}, Producto_{j\rightarrow i}\} \quad (3.3)$$

donde, $Producto_{i\rightarrow j}$ es la proporción de los productos del subsector i que se venden al subsector j y $Producto_{j\rightarrow i}$ es la proporción de productos de j que se venden al subsector i . Para estimar esta variable se calculan las proporciones de producto de cada sector en relación al PIB total del sector y después se elige el valor máximo.

El objetivo de esta variable es capturar la relación de venta o compra de bienes más fuerte entre los subsectores o industrias, debido a esto es que elegimos su valor máximo. La variable toma valores entre cero y uno, a mayor es *Bienes* más intensa la relación a través del intercambio de bienes. Con este indicador se busca aproximar las relaciones cliente-proveedor que puedan facilitar el aprovechamiento de las externalidades derivadas de las economías de aglomeración a través de un mercado de bienes amplio.

3.1.1. Estimación de la variable *Bienes*

En la Cuadro 3.1(a) se presentan las estadísticas descriptivas de la variable que aproxima la relación insumo-producto, su media es cercana a 0.01. La variable *Bienes* tiene un rango

pequeño, en el caso del año 2008 va de cero a 0.1. Por lo que la proporción máxima de intercambio de bienes es menor al 10 por ciento de los insumos o productos que comercian entre los subsectores. En la Figura 3.1(b), se observa que entre las entidades federativas donde las relaciones cliente-proveedor son más fuertes están el DF, el Estado de México, Veracruz, Tamaulipas y Nuevo León (NL).

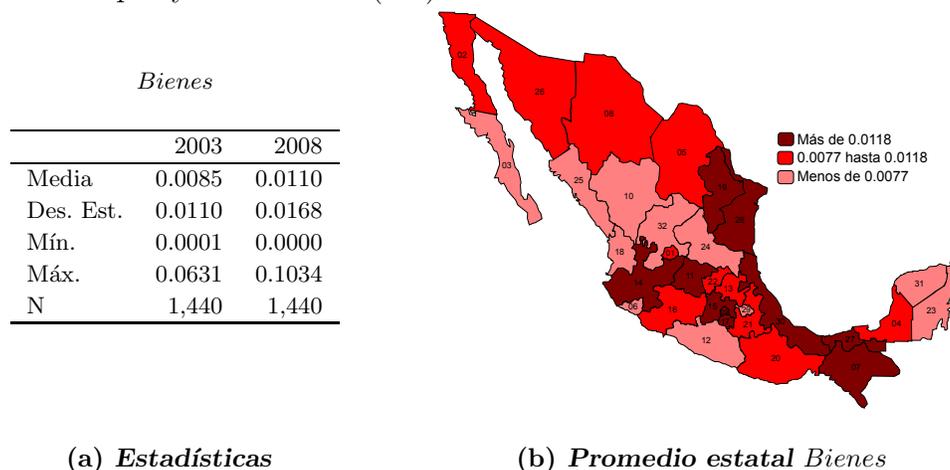


Figura 3.1: Estadísticas descriptivas: *Bienes*, 2003-2008

3.2. Variable *proxy*: *Ocupación*

Otra explicación de porqué los empresarios se aglomeran en una región es que existe una fuerza laboral conveniente o adecuada para sus procesos productivos. Un mercado laboral amplio y diverso permite reducir los efectos de choques económicos a través de la movilidad laboral (Ellison et al., 2010).

Para aproximar la fuerza Marshalliana relacionada con las personas se construyen dos variables. En la estimación de la variable que representa las relaciones según el tipo de trabajadores, utilizamos los empleados según la ocupación en el trabajo.² En la otra variable consideramos la escolaridad de los trabajadores en los diferentes sectores de la industria manufacturera. La información se obtiene de la ENEU 2003 y de la ENOE 2008 del INEGI (2015).

²Se utilizan 11 ocupaciones de la Clasificación Mexicana de Ocupaciones (CMO) (Cuadro A.2 de Apéndice).

El objetivo es determinar el grado en que los diferentes subsectores de la industria manufacturera utilizan trabajadores en ocupaciones similares o con niveles de escolaridad semejantes. Las variables se construyen de las siguiente forma (Ellison et al., 2010):

$$Ocupación_{ijs} = Corr(Proporción\ ocupación_{iso}, Proporción\ ocupación_{jso}) \quad (3.4)$$

$$Escolaridad_{ijs} = Corr(Proporción\ escolaridad_{ise}, Proporción\ escolaridad_{jse}) \quad (3.5)$$

donde, i, j indexan los subsectores de la industria manufacturera, s es la entidad federativa, o indexa la ocupación de los trabajadores. $Ocupación_{ijs}$ es la variable *proxy* de un mercado laboral amplio y diverso, donde se correlacionan las proporciones de empleados en la ocupación o entre las industria i y j de la entidad federativa s ($Proporción\ ocupación_{iso}$). Cuando dos subsectores comparten el mismo tipo de trabajadores, el indicador se aproxima a uno y en el caso contrario tiende a menos uno.

La otra variable *proxy* es $Escolaridad_{ijs}$, que estima la correlación de los trabajadores según escolaridad entre los subsectores i y j . Para su cálculo se obtiene la correlación de las proporciones de trabajadores con escolaridad e en las industrias i y j de cada entidad federativa s ($Proporción\ escolaridad_{ise}$).

A través de estos indicadores suponemos que a mayor correlación respecto al tipo de ocupación o escolaridad entre los subsectores, es más fácil aprovechar las ventajas derivadas de las externalidades producidas por contar con un mercado laboral amplio y diverso, lo que implica un incentivo para la coaglomeración económica.

3.2.1. Estimación de las variables *Ocupación* y *Escolaridad*

En la Cuadro 3.2(a) están las estadísticas descriptivas de la variable que aproxima la relación del mercado laboral amplio y diverso a través de la ocupación de los trabajadores (Ecuación (3.4)), cuya media es de aproximadamente 0.8 en 2003 y 2008. Entre las entidades federativas en donde es mayor la correlación, entre el tipo de trabajadores que se utilizan en la industria manufacturera, están NL, Querétaro, Jalisco, Puebla y Guanajuato (Figura 3.2(b)).

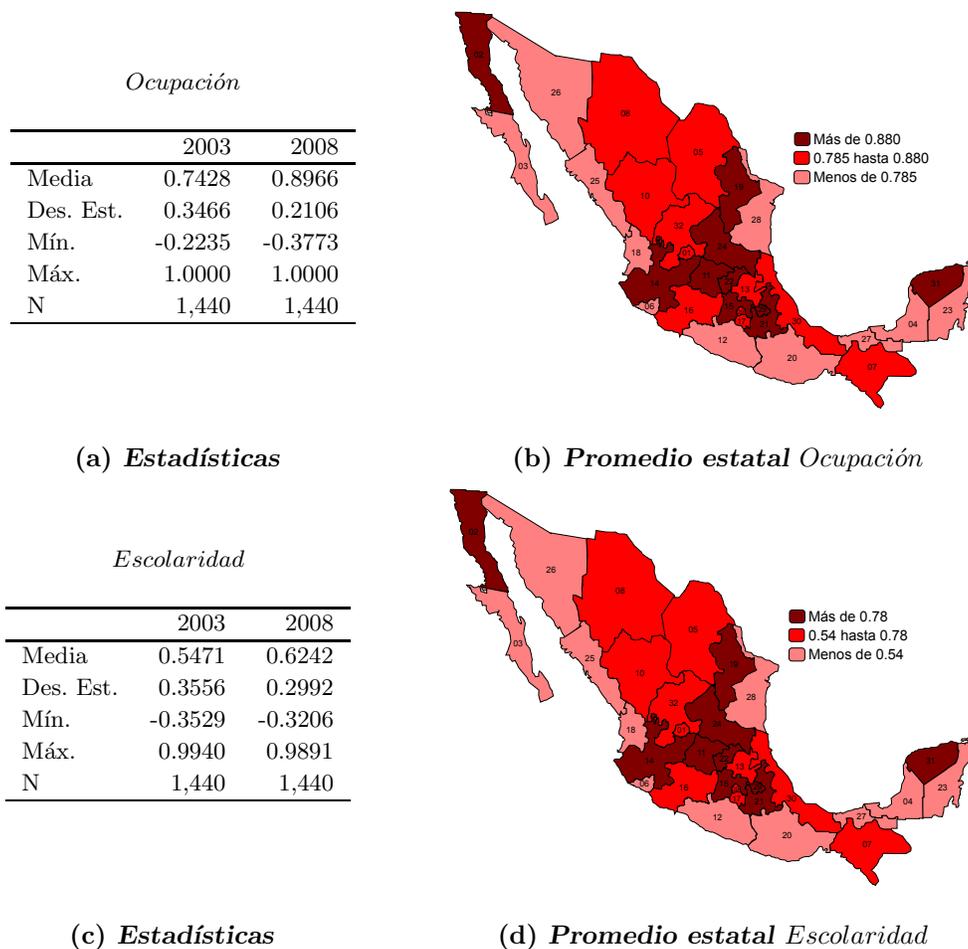


Figura 3.2: Estadísticas descriptivas: *Ocupación* y *Escolaridad*, 2003-2008

En la Cuadro 3.2(d) se presentan estadísticos de la variable *proxy* de personas que utiliza información sobre la escolaridad del personal ocupado (Ecuación (3.5)). La media de esta variable es de 0.54, en 2003 y aumenta 0.62 en 2008, una correlación menor a la obtenida con la variable *Ocupación*. La distribución de entidades según su nivel de correlación no cambia. Mientras que las entidades que sobresalen por tener el menor nivel de correlación en escolaridad son Nayarit, Oaxaca, Guerrero y Baja California Sur (Figura 3.2(d)).

3.3. Variable *proxy*: Ideas

Finalmente, consideramos que el flujo de ideas es otra razón por la que los empresarios ven redituable agruparse en cierto lugar. Marshall señala que los trabajadores aprenden

habilidades de otros de forma más rápida, en un agrupamiento industrial (Ellison et al., 2010). El objetivo de construir esta variable es ver si las industrias que se localizan cerca unas de otras tienen en común aspectos relacionados con la innovación e investigación, rubros clave para el flujo de ideas. Suponemos que a mayor proximidad tecnológica habrá más facilidad para que los conocimientos fluyan entre las industrias (Kaiser, 2002).

La variable *Ideas* se construye con información del módulo de innovación e investigación del Censo Económico 2004 y con variables relacionadas con la innovación del Censo Económico 2009 (INEGI, 2015). Entre las características que se incluyen están si las empresas cuentan con personal calificado para buscar soluciones técnicas, mejorar la eficiencia o innovar, si la empresa patentó, celebra convenios de colaboración con universidades y hace investigación propia. La variable que aproxima la fuerza Marshalliana de *Ideas* se construye de la forma siguiente:

$$Ideas_{ijs} = Corr(Características\ de\ innovación_{is}, Características\ de\ innovación_{js}) \quad (3.6)$$

donde, i, j indexan los subsectores de la industria manufacturera y s representa la entidad federativa. La variable *proxy* de *Ideas* ($Ideas_{ijs}$) correlaciona las características de las actividades de innovación de las industrias i y j en el estado s ($Características\ de\ innovación_{is}$).

La variable oscila entre menos uno y uno, a mayor nivel, más similitud entre las dos industrias, en términos de características de innovación e investigación. Suponemos que a mayor similitud entre los vectores de características tecnológicas de los subsectores mayor facilidad de aprovechar las externalidades derivadas de las economías de aglomeración.

3.3.1. Estimación de la variable *Ideas*

En el Cuadro 3.3(a) se presentan los estadísticos de la variable *proxy* de *Ideas* donde la media es de 0.85 en 2003 y de 0.52 en 2008.³ En cuanto al aspecto espacial, las entidades federativas en donde se encuentra mayor correlación son NL, Jalisco, DF, Baja California y Puebla.

³Para esta variable el vector de características que se consideraron difieren entre los Censos Económicos, sin embargo, se trató de utilizar información similar para evitar que influya en los resultados.

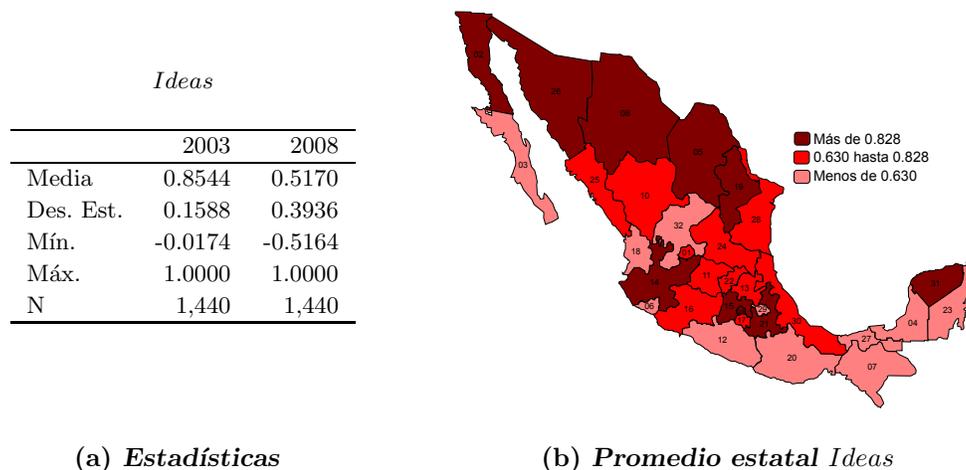


Figura 3.3: Estadísticas descriptivas: *Ideas*, 2003-2008

Una vez estimadas las variables *proxy* de las fuerzas Marshallianas y con el cálculo previo del índice de coaglomeración de Ellison et al. (2010), podemos evaluar la importancia relativa de las diferentes fuerzas Marshallianas en la coaglomeración de los subsectores de la industria manufacturera de México.

A partir las estadísticas descriptivas de las variables observamos que el cambio entre 2003 y 2008 es bajo, en general. En cuanto al aspecto espacial del análisis encontramos que la relación entre subsectores cuando se consideran las características de las actividades de innovación son más fuertes en los estados del norte del país, Jalisco y en entidades cercanas a la capital. La relación en el mercado de bienes es intensa en la zona del Golfo de México y en menor media en el Norte. El Centro y algunas entidades del Norte presentan una alta correlación entre los tipos de ocupaciones y la escolaridad de los trabajadores. Las fuerzas Marshallianas presentan ciertos patrones de localización que se pueden observar en los mapas (Figuras 3.1(b), 3.2(b), 3.2(d) y 3.3(b)), sin embargo, el índice de coaglomeración no tiene un patrón tan claro con excepción del centro del país que se caracteriza por estar coaglomerado (Figura 2.1(b)).

Capítulo 4

Estimación de la importancia relativa de las fuerzas Marshallianas en la coaglomeración económica

4.1. Especificación econométrica

La estrategia para estimar la importancia de las fuerzas Marshallianas sobre la coaglomeración económica es mediante regresiones con Mínimos Cuadrados Ordinarios (MCO), utilizando como variables independientes las variables *proxies* de las teorías Marshallianas. La especificación econométrica base es:

$$Coagl_{ijs} = \beta_0 + \beta_1 Bienes_{ijs} + \beta_2 Ocupación_{ijs} + \beta_3 Ideas_{ijs} + \epsilon_{ijs} \quad (4.1)$$

donde, i, j indexan los subsectores de la industria manufacturera y s es la entidad federativa. $Coagl_{ijs}$ es el índice de coaglomeración de los subsectores i y j de la entidad federativa s . La variable *proxy* para la relación de insumo-producto es $Bienes_{ijs}$, la relación de los tipos de trabajadores según su ocupación esta representada por $Ocupación_{ijs}$ y alternativamente utilizamos la escolaridad ($Escolaridad_{jis}$). La variable *proxy* de ideas o *spillovers* tecnológicos es $Ideas_{ijs}$ y por último, ϵ_{ijs} es el término de error.¹

¹Siguiendo a Ellison et al. (2010) aplicamos errores *bootstrap*.

Existen dos problemas de identificación al tratar de encontrar el efecto de las fuerzas Marshallianas sobre la coaglomeración económica (Glaeser y Kerr, 2009; Ellison et al., 2010; Jofre-Monseny et al., 2011): el primero es la simultaneidad o causalidad inversa que se presentaría si las industrias se localizan en un lugar y a partir de este hecho comienzan a compartir el mismo tipo de trabajadores, conocimientos o inician relaciones cliente-proveedor o sucede a lo contrario. El segundo es la omisión de variables, un problema común en econometría, y está presente en casos donde existen variables observadas o no observadas que generan dependencia en la localización de dos industrias, por ejemplo, ciertas ventajas naturales.

Para este trabajo, la estrategia de identificación es incluir variables de control por entidad federativa para considerar el problema de variables observadas y no observadas relacionadas con las ventajas naturales. El otro control es por subsector de actividad económica cuyo objetivo es tratar de minimizar el problema de causalidad inversa.

4.2. Resultados

En esta sección se presentan los resultados de las regresiones que se hicieron con las variables *proxies* de las fuerzas Marshallianas y el índice de coaglomeración. En las regresiones se utilizan como variables de control *dummies* por entidad federativa y por industria, con la finalidad de minimizar los problemas de identificación que se mencionaron anteriormente. También se presentan los coeficientes estandarizados (*beta*), los cuales permiten realizar comparaciones entre las variables debido a que están expresados en desviaciones estándar, de tal forma que podamos identificar la fuerza Marshalliana con mayor importancia relativa sobre la coaglomeración económica, que es el objetivo de este trabajo.

En el Cuadro 4.1 se presentan regresiones lineales de las externalidades Marshallianas y el índice de coaglomeración. En 2003, se encuentra que existe una relación significativa y positiva entre las variables *proxies* de la relación comercial (*Bienes*) entre sectores y del flujo de conocimientos (*Ideas*), con la coaglomeración económica (Columnas (1)-(2) del Cuadro 4.1). A través de los coeficientes estandarizados encontramos que la variable de mayor jerarquía es la de *Bienes*, es decir, es la que mayor impacto tiene sobre la coaglomeración.

Para el año 2008 (Columnas (3-4) del Cuadro 4.1) sólo son significativas las variables *Ideas* y *Escolaridad*.

Cuadro 4.1: Regresiones MCO a nivel nacional, 2003-2008

Variable dependiente: Índice de Coaglomeración						
Variables	2003		2008		2008 con rezago ^a	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Bienes</i>	0.416** (0.061)	0.403** (0.060)	0.138 (0.032)	0.137 (0.032)	0.442*** (0.068)	0.434** (0.066)
	0.170	0.192	0.093	0.145	0.159	0.181
<i>Ocupación</i>	-0.012 (-0.055)		-0.009 (-0.025)		-0.007 (-0.034)	
	0.008		0.010		0.005	
<i>Escolaridad</i>		0.012 (0.058)		0.015* (0.062)		0.004 (0.018)
		0.010		0.008		0.007
<i>Ideas</i>	0.079*** (0.168)	0.079*** (0.169)	0.017** (0.092)	0.017** (0.094)	0.066*** (0.146)	0.067** (0.146)
	0.025	0.028	0.007	0.007	0.025	0.027
N	1,440	1,440	1,440	1,440	1,440	1,440
<i>R</i> ²	0.098	0.098	0.167	0.168	0.179	0.179

Errores *bootstrap* robustos entre paréntesis. En el tercer renglón están los coeficientes estandarizados. En todos los casos se usan variables *dummies* por entidad y por sector. ^a En estas regresiones se utilizan las variables dependientes del año 2003 y la dependiente de 2008. *** p<0.01, ** p<0.05, * p<0.1

El efecto de contar con un mercado amplio de trabajadores, de proveedores y clientes, así como un ambiente propicio para la difusión de conocimiento, sobre la coaglomeración, podría tardar tiempo en verse reflejado en el contexto económico. Para considerar este efecto rezagado presentamos una regresión manteniendo las variables independientes en el año 2003 y la dependiente en el año 2008 (Columnas (5-6) del Cuadro 4.1). Los resultados muestran que los coeficientes relacionados con las variables del mercado de bienes y del flujo de ideas son significativos y positivos. La fuerza Marshalliana *Bienes* es la de mayor jerarquía, según la comparación de los coeficientes estandarizados.

En el Cuadro 4.2 presentamos regresiones donde se consideran efectos del tiempo, espacio y de la actividad económica. La estimación se hizo a través de MCO agrupados (*pooled*

Cuadro 4.2: Regresiones MCO agrupados y con *dummies* a nivel nacional, 2003-2008

Variable dependiente: Índice de Coaglomeración				
Variables	MCO agrupados ^a		MCO con <i>dummies</i> ^{bc}	
	(1)	(2)	(3)	(4)
<i>Bienes</i>	0.177*	0.178*	0.195*	0.188**
	(0.034)	(0.035)	(0.038)	(0.037)
	0.101	0.101	0.107	0.087
<i>Ocupación</i>	-0.010**		-0.008	
	(-0.039)		(-0.031)	
	0.005		0.005	
<i>Escolaridad</i>		0.004		0.011*
		(0.020)		(0.048)
		0.004		0.006
<i>Ideas</i>	0.005	0.003	0.022***	0.024***
	(0.024)	(0.012)	(0.105)	(0.110)
	0.007	0.007	0.007	0.008
N	2,880	2,880	2,880	2,880
<i>R</i> ²	0.003	0.002	0.113	0.114

^a Errores robustos agrupados por sector (*ij*) y entidad, se incluye una variable *dummy* de tiempo. ^b Errores *bootstrap* robustos entre paréntesis. ^c Variables *dummies* por entidad, sector y año. En el tercer renglón están los coeficientes estandarizados. *** p<0.01, ** p<0.05, * p<0.1

OLS) donde incluimos el efecto fijo del tiempo y al estimar los errores controlamos la heterogeneidad de las observaciones, en este caso entre los subsectores dentro de cada entidad (Columnas (1-2)). Los resultados indican que la variable *Bienes* es significativa y positiva, además, la *Ocupación* también resulta significativa con signo negativo, lo que indica que favorece más la aglomeración de un subsector.

En las Columnas (3-4) del Cuadro 4.2 se presentan los coeficientes estimados de las fuerzas Marshallianas, donde incluimos en la ecuación las siguientes variables de control: la entidad, el sector y el año. Los coeficientes estimados de la relación insumo-producto e ideas son significativos y positivos, también la variable *Escolaridad* resulta significativa.

A nivel nacional, la fuerza Marshalliana que mayor peso tiene, sobre la coaglomeración económica, es la relacionada con el intercambio de productos (*Bienes*). En la mayoría de los casos, la variable relacionada con los *spillovers* de conocimientos tiene un efecto positivo y significativo sobre el índice de coaglomeración. Dichas relaciones positivas nos indican que cuando existe un mercado amplio de clientes y proveedores, además, de un ambiente adecuado para el flujo de ideas se puede dar la concentración de industrias de diferentes giros en una región.

Los modelos con mejor ajuste, según el coeficiente de determinación (R^2), son: el de variables explicativas rezagadas y el que incluye variables *dummies* por entidad, sector y año, utilizaremos estas dos especificaciones para estudiar los efectos a nivel regional.

Cuadro 4.3: Regresiones MCO con variables rezagadas, por región 2003-2008

Variable dependiente: Índice de Coaglomeración (2008)							
Variables (2003)	Norte (1)	Pacífico (2)	Golfo (3)	Centro Nte (4)	Centro (5)	Capital (6)	Sur (7)
<i>Bienes</i>	0.729** (0.150)	0.567 (0.081)	0.768* (0.095)	-0.323 (-0.030)	0.150 (0.025)	-0.015 (-0.008)	0.391* (0.139)
	0.320	0.474	0.433	0.669	0.487	0.207	0.204
<i>Ocupación</i>	0.003 (0.020)	-0.016 (-0.102)	0.010 (0.040)	-0.020 (-0.054)	-0.011 (-0.056)	0.154** (0.257)	-0.002 (-0.018)
	0.011	0.013	0.018	0.022	0.016	0.073	0.011
<i>Ideas</i>	-0.029 (-0.042)	0.012 (0.038)	0.231*** (0.571)	-0.096 (-0.104)	0.029 (0.056)	0.042 (0.062)	-0.017 (-0.082)
	0.038	0.048	0.050	0.086	0.037	0.067	0.022
N	270	225	225	270	180	90	180
R^2	0.196	0.245	0.351	0.292	0.294	0.740	0.398

Errores *bootstrap* robustos entre paréntesis. Variables *dummies* por entidad, sector y año. La variable dependiente corresponde al año 2008 y las explicativas al año 2003. En el tercer renglón están los coeficientes estandarizados. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

En el Cuadro 4.3 se presentan las regresiones utilizando el índice de coaglomeración del año 2008 y las externalidades correspondientes al año 2003. Las relaciones cliente-proveedor dentro de las regiones Norte, Golfo y Sur son significativas y tienen un impacto positivo sobre la coaglomeración. En la Capital se observa que la externalidad de amplio y diverso mercado de trabajo es la única con significancia y tiene un efecto positivo sobre la coaglomeración. En

el Golfo, el coeficiente estimado de los *spillovers* de conocimientos también es significativo, pero su importancia relativa es menor a la externalidad asociada a la variable *Bienes*, según sus coeficientes estandarizados.²

Cuadro 4.4: Regresiones MCO con *dummies*, por región 2003-2008

Variable dependiente: Índice de Coaglomeración							
Variables	Norte (1)	Pacífico (2)	Golfo (3)	Centro Nte (4)	Centro (5)	Capital (6)	Sur (7)
<i>Bienes</i>	0.500** (0.141)	0.003 (0.000)	0.236 (0.035)	-0.108 (-0.014)	-0.101 (-0.021)	0.034 (0.023)	0.403*** (0.142)
	0.233	0.250	0.277	0.300	0.285	0.121	0.138
<i>Ocupación</i>	0.007 (0.041)	-0.028** (-0.154)	-0.008 (-0.025)	-0.022 (-0.046)	-0.009 (-0.036)	-0.009 (-0.067)	0.016* (0.114)
	0.008	0.013	0.012	0.023	0.014	0.009	0.009
<i>Ideas</i>	0.020* (0.057)	-0.016 (-0.102)	0.064*** (0.260)	-0.007 (-0.018)	0.008 (0.036)	0.018 (0.098)	0.025* (0.211)
	0.012	0.014	0.024	0.018	0.014	0.011	0.014
N	540	450	450	540	360	180	360
R^2	0.165	0.161	0.175	0.205	0.263	0.714	0.366

Errores *bootstrap* robustos entre paréntesis. Variables *dummies* por entidad, sector y año.

En el tercer renglón están los coeficientes estandarizados. *** p<0.01, ** p<0.05, * p<0.1

En el Cuadro 4.4 controlamos por los efectos de tiempo, espacio y actividad económica. En la región Norte contar con un mercado amplio de clientes y proveedores (*Bienes*) incentiva la coaglomeración, de igual forma hay un efecto positivo de la presencia de industrias con tendencia a innovar que permite el flujo de conocimientos (*Ideas*). En el Golfo la variable *Ideas* resulta positiva y significativa. Mientras que en la región Sur todas las externalidades son significativas y positivas. En el Pacífico se encontró una relación negativa de la variable relacionada al mercado laboral amplio y diverso y la coaglomeración económica.³

Las relaciones que encontramos a nivel regional son débiles y esto podría explicarse por la forma de la regionalización. Se utiliza una regionalización según la vecindad espacial y

²Si en lugar de ocupación utilizamos escolaridad los resultados se modifican en el caso de la significancia de la variable *Bienes* en el Golfo y el efecto de *Escolaridad* en la Capital es negativo (Cuadro C.1 del Apéndice).

³Al utilizar la variable escolaridad sólo difiere la significancia de ésta variable en la Capital y su efecto es negativo (Cuadro C.2 del Apéndice).

dado que se están analizando externalidades, podría no ser la forma óptima de formar las regiones. Sin embargo, cuando la variable *Bienes* resultó significativa fue posible confirmar que es la que mayor influencia tiene sobre la coaglomeración.

Conclusiones

En las ciudades se concentra gran parte de la actividad económica de un país. En México, el 50 por ciento de la producción total está en seis de las 32 entidades federativas, y en siete se encuentra la mitad de la población nacional (INEGI, 2015). En cuanto a las actividades productivas, en el DF, Jalisco y NL hay presencia de los 10 subsectores de la industria manufacturera, lo que refleja un alto nivel de coaglomeración. En este trabajo se utiliza el enfoque de coaglomeración económica para estudiar la importancia de las fuerzas Marshallianas sobre la concentración de más de una industria en una región.

Para estudiar los patrones de la actividad económica en el espacio calculamos el índice de coaglomeración que considera la concentración de más de una industria en una región. De los patrones de concentración se distinguen algunos de industrias que esperaríamos sería más benéfico que se ubicaran juntas ya que pueden reducir el costo de transporte de la materia prima, por ejemplo, el sector de Maquinaria y equipo (VIII) y el de Derivados del petróleo, carbón, industria química y del hule (IX). Por otro lado, las entidades federativas que destacan por su índice de coaglomeración son el DF, Querétaro y Aguascalientes, todas entidades del centro del país en donde se concentra la mayoría de la población. También se realizó un análisis por región destacando la Capital y la región Centro Norte como las más coaglomeradas.

La existencia de la concentración económica se explica por la presencia de economías de aglomeración. Marshall (1890) identifica tres fuerzas que generan beneficios para las empresas que se aglomeran, más allá de las ventajas comparativas y naturales de la región. Para explicar la coaglomeración económica en México construimos variables *proxy* de las fuerzas Marshallianas que se resumen en contar con un mercado amplio para poder tener proveedores especializados, fuerza laboral especializada y densa así como *spillovers* tecnológicos.

Para reducir los problemas de identificación, a los que nos enfrentamos cuando relacionamos el índice de coaglomeración con las teorías Marshallianas, utilizamos variables de control por entidad federativa, por subsector de la industria manufacturera y por el periodo. Los resultados encontrados señalan que la fuerza Marshalliana relativa a la relación cliente-proveedor entre los subsectores es la que mayor importancia presenta. Cuando se hace al análisis a nivel regional, los resultados son mixtos, con las estadísticas descriptivas tampoco se observa que exista un patrón regional de coaglomeración. En la región Norte, Golfo y Sur, la fuerza Marshalliana relacionada con el mercado de bienes es la más importante para explicar la coaglomeración económica de esas regiones.

El objetivo del trabajo fue buscar cuál fuerza Marshalliana es la que tiene más importancia respecto al nivel de coaglomeración de las diferentes áreas geográficas. A través de herramientas econométricas tratamos de subsanar los problemas de especificación. En el caso de las regresiones con *dummies*, las variables corresponden al mismo año lo que omite las relaciones rezagadas entre las externalidades Marshallianas y la coaglomeración, limitando la posibilidad de identificar relaciones causales. Para mejorar los resultados necesitamos de más observaciones en el tiempo.

Dentro del tema de la coaglomeración económica existe todavía mucho trabajo por hacer debido a que hasta el momento no hay estudios que utilicen este enfoque para analizar la concentración económica en México. También se pueden mejorar los resultados que hemos obtenido si utilizamos un nivel mayor de desagregación sectorial de tal forma que podamos capturar con mayor detalle la diversidad en los patrones de coaglomeración.

Bibliografía

- Amiti, M. (1999), ‘Specialization patterns in Europe’, *Review of World Economics (Weltwirtschaftliches Archiv)* **135**(4), 573–593.
- Amiti, M. y Cameron, L. (2007), ‘Economic geography and wages’, *Review of Economics and Statistics* **89**(1), 15–29.
- Arhansya, S. (2010), Industrial agglomeration and coagglomeration in Indonesia. The University of Melbourne, Working paper.
- Audretsch, D. B. y Feldman, M. P. (1996), ‘R&D spillovers and the geography of innovation and production’, *American Economic Review* **86**(3), 630–40.
- Bacolod, M., Blum, B. S. y Strange, W. C. (2009), ‘Skills in the city’, *Journal of Urban Economics* **65**(2), 136 – 153.
- Brühlhart, M. y Torstensson, J. (1996), ‘Regional integration, scale economies and industry location in the European Union’, *CEPR Discussion Paper No. 1435* .
- Brühlhart, M. y Traeger, R. (2005), ‘An account of geographic concentration patterns in Europe’, *Regional Science and Urban Economics* **35**(6), 597–624.
- Ciccone, A. y Peri, G. (2006), ‘Identifying human-capital externalities: Theory with applications’, *Review of Economic Studies* **73**(2), 381–412.
- Costa, D. L. y Kahn, M. E. (2000), ‘Power couples: Changes in the locational choice of the college educated, 1940–1990’, *The Quarterly Journal of Economics* **115**(4), 1287–1315.

- Cutrini, E. (2009), 'Using entropy measures to disentangle regional from national localization patterns', *Regional Science and Urban Economics* **39**(2), 243–250.
- Dávila Flores, A. (2002), 'Matriz de insumo-producto de la economía de Coahuila e identificación de sus flujos intersectoriales más importantes', *economía mexicana NUEVA ÉPOCA* **XI**(1), 79–162.
- Duranton, G. y Overman, H. G. (2005), 'Testing for localization using micro-geographic data', *Review of Economic Studies* **72**(4), 1077–1106.
- Duranton, G. y Puga, D. (2005), 'From sectoral to functional urban specialisation', *Journal of Urban Economics* **57**(2), 343–370.
- Ellison, G. y Glaeser, E. L. (1997), 'Geographic concentration in U.S. manufacturing industries: A dartboard approach', *Journal of Political Economy* **105**(5), 889–927.
- Ellison, G. y Glaeser, E. L. (1999), 'The geographic concentration of industry: Does natural advantage explain agglomeration?', *American Economic Review* **89**(2), 311–316.
- 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.
- Esquivel, G. (1999), 'Convergencia regional en México, 1940-1995', *El Trimestre Económico* **46**(4), 725–761.
- Flegg, A. T. y Tohmo, T. (2013a), 'A comment on Tobias Kronenberg's "Construction of regional input-output tables using nonsurvey methods: The role of cross-hauling"', *International Regional Science Review* **36**(2), 235–257.
- Flegg, A. T. y Tohmo, T. (2013b), 'Regional input-output tables and the FLQ formula: A case study of Finland', *Regional Studies* **47**(5), 703–721.
- Flegg, A. T. y Webber, C. D. (1997), 'On the appropriate use of location quotients in generating regional input-output tables: Reply', *Regional Studies* **31**(8), 795–805.

- Flegg, A. T. y Webber, C. D. (2000), 'Regional size, regional specialization and the FLQ formula', *Regional Studies* **34**(6), 563–569.
- Flegg, A. T., Webber, C. D. y Elliott, M. V. (1995), 'On the appropriate use of location quotients in generating regional input–output tables', *Regional Studies* **29**(6), 547–561.
- Gan, L. y Li, Q. (2004), Efficiency of thin and thick markets, NBER Working Papers 10815, National Bureau of Economic Research, Inc.
- Glaeser, E. L. y Kerr, W. R. (2009), 'Local industrial conditions and entrepreneurship: How much of the spatial distribution can we explain?', *Journal of Economics and Management Strategy* **18**(3), 623 – 663.
- Gleaser, E., Kallal, H. D., Sheinkman, J. A. y Shkeifer, A. (1992), 'Growth in cities', *Journal of Political Economy* **100**(6), 1126–1152.
- Helsley, R. W. y Strange, W. C. (2012), Coagglomeration and the scale and composition of clusters. Working paper.
- Hoover, E. M. (1948), *The Location of Economic Activity*, NY: McGraw Hill.
- INEGI (2015), 'Instituto Nacional de Estadística y Geografía'.
URL: <http://www.inegi.org.mx>
- Jacobs, J. (1969), *The Economy of Cities*, Vintage, New York.
- Jaffe, A. B., Trajtenberg, M. y Henderson, R. (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *The Quarterly Journal of Economics* **108**(3), 577–598.
- Jofre-Monseny, J., Marín-López, R. y Viladecans-Marsal, E. (2011), 'The mechanisms of agglomeration: Evidence from the effect of inter-industry relations on the location of new firms', *Journal of Urban Economics* **70**(2-3), 61–74.
- Kaiser, U. (2002), 'Measuring knowledge spillovers in manufacturing and services: an empirical assessment of alternative approaches', *Research Policy* **31**, 125–144.

- Kolko, J. (2010), Urbanization, agglomeration, and coagglomeration of service industries, *in* E. L. Glaeser, ed., 'Agglomeration Economics', NBER Chapters, The University of Chicago Press, pp. 151–180.
- Kronenberg, T. (2009), 'Construction of regional input-output tables using nonsurvey methods: The role of cross-hauling', *International Regional Science Review* **32**(1), 40–64.
- Krugman, P. (1991), *Geography and Trade*, MIT Press, MA.
- Krugman, P. (1998), 'What's new about the new economic geography?', *Oxford review of economic policy* **14**(2).
- Krugman, P. y Obsfeld, M. (2005), *International economics: Theory and policy.*, Pearson Addison-Wesley.
- Marshall, A. (1890), *Principles of Economics*, MacMillan, London.
- Maurel, F. y Sédillot, B. (1999), 'A measure of the geographic concentration in french manufacturing industries', *Regional Science and Urban Economics* **29**, 575–604.
- Overman, H. G. y Puga, D. (2010), Labor pooling as a source of agglomeration: An empirical investigation, *in* E. L. Glaeser, ed., 'Agglomeration Economics', NBER Chapters, National Bureau of Economic Research, Inc, chapter 133-150.
- Puga, D. (2010), 'The magnitude and causes of agglomeration economies', *Journal of Regional Science* **50**(1), 203–219.
- Rosenthal, S. S. y Strange, W. C. (2001), 'The determinants of agglomeration', *Journal of Urban Economics* **50**(2), 191–229.
- Strange, W. C. (2008), Urban agglomeration, *in* S. N. Durlauf y L. E. Blume, eds, 'The New Palgrave Dictionary of Economics', Palgrave Macmillan.

Apéndice A

Definición de sectores, ocupaciones, entidades y regiones

Cuadro A.1: Industria manufacturera por subsector agregado

Id.	No.	Subsector
	31-33	Industrias manufactureras
I	311-312	Industria alimentaria, de las bebidas y del tabaco
II	313-316	Textiles, prendas de vestir y productos de cuero
III	321	Industria de la madera
IV	322-323	Industria del papel, impresión e industrias conexas
V	324-326	Derivados del petróleo y del carbón; industrias química, del plástico y del hule
VI	327	Fabricación de productos a base de minerales no metálicos
VII	331-332	Industrias metálicas
VIII	333-336	Maquinaria y equipo
IX	337	Fabricación de muebles y productos relacionados
X	339	Otras industrias manufactureras

Fuente: INEGI (2015).

Cuadro A.2: Clasificación de ocupaciones

Ocupación	
1	Profesionales, técnicos y trabajadores del arte
2	Trabajadores de la educación
3	Funcionarios y directivos
4	Oficinistas
5	Trabajadores industriales artesanos y ayudantes
6	Comerciantes
7	Operadores de transporte
8	Trabajadores en servicios personales
9	Trabajadores en protección y vigilancia
10	Trabajadores agropecuarios
11	No especificado

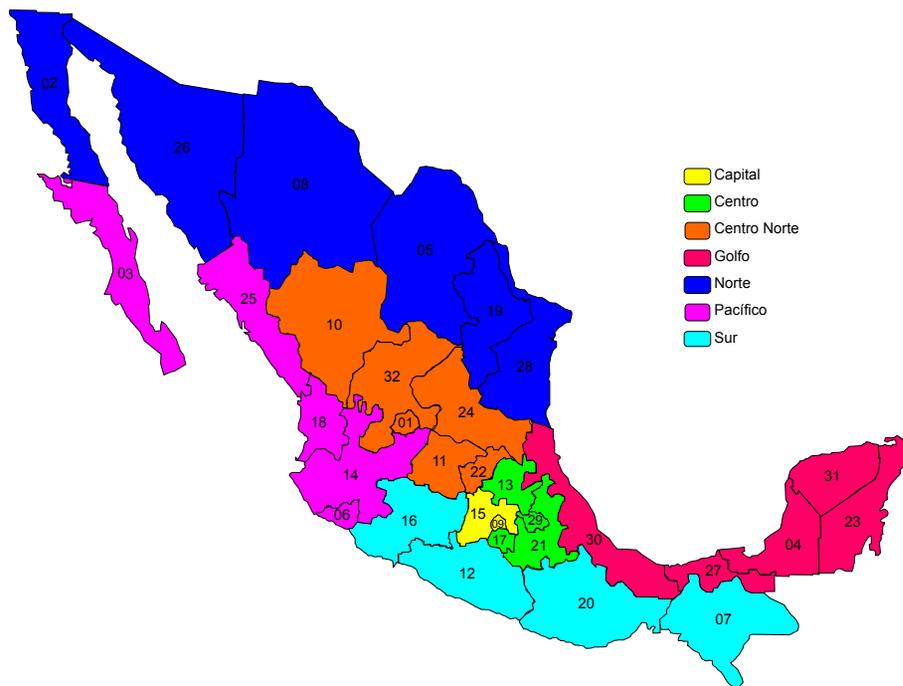
Fuente: CMO, INEGI (2015).



No.	Entidad	No.	Entidad
01	Aguascalientes	17	Morelos
02	Baja California	18	Nayarit
03	Baja California Sur	19	Nuevo León
04	Campeche	20	Oaxaca
05	Coahuila	21	Puebla
06	Colima	22	Querétaro
07	Chiapas	23	Quintana Roo
08	Chihuahua	24	San Luis Potosí
09	Distrito Federal	25	Sinaloa
10	Durango	26	Sonora
11	Guanajuato	27	Tabasco
12	Guerrero	28	Tamaulipas
13	Hidalgo	29	Tlaxcala
14	Jalisco	30	Veracruz
15	México	31	Yucatán
16	Michoacán	32	Zacatecas

Fuente: INEGI (2015).

Figura A.1: División política de la República Mexicana



Entidad	Región	Entidad	Región
09 Distrito Federal	Capital	27 Tabasco	Golfo
15 México	Capital	05 Coahuila	Norte
17 Morelos	Centro	19 Nuevo León	Norte
13 Hidalgo	Centro	28 Tamaulipas	Norte
21 Puebla	Centro	02 Baja California	Norte
29 Tlaxcala	Centro	08 Chihuahua	Norte
11 Guanajuato	Centro norte	26 Sonora	Norte
22 Querétaro	Centro norte	06 Colima	Pacífico
01 Aguascalientes	Centro norte	14 Jalisco	Pacífico
10 Durango	Centro norte	03 Baja California Sur	Pacífico
24 San Luis Potosí	Centro norte	25 Sinaloa	Pacífico
32 Zacatecas	Centro norte	18 Nayarit	Pacífico
04 Campeche	Golfo	07 Chiapas	Sur
23 Quintana Roo	Golfo	12 Guerrero	Sur
30 Veracruz	Golfo	16 Michoacán	Sur
31 Yucatán	Golfo	20 Oaxaca	Sur

Fuente: Esquivel (1999)

Figura A.2: Regionalización

Apéndice B

Método de regionalización indirecto de matrices insumo-producto

A través de las matrices insumo-producto (MIP) podemos conocer los flujos comerciales entre los diferentes sectores de la actividad económica. Este tipo de información generalmente la encontramos de forma agregada, es decir, a nivel nacional. El costo de producir MIPs regionales es alto así que para regionalizar la matriz nacional lo más viable es utilizar métodos de estimación indirectos.

Los métodos indirectos utilizan coeficientes de localización (LQ) para estimar los coeficientes regionales insumo-producto. Entre los LQ que se utilizan están (Flegg y Webber, 2000):

a) Coeficiente de localización simple (SLQ_i)

$$SLQ_i = \frac{\frac{E_{ri}}{E_r}}{\frac{E_{ti}}{E_t}} = \frac{E_{ri}}{E_{ti}} * \frac{E_t}{E_r}$$

donde, E_r es el empleo en la región r , E_{ri} empleo del sector i en la región r , E_t empleo total (nacional) y E_{ti} empleo del sector i a nivel agregado. Cuando el $SLQ_i < 1$ indica que el sector i está subrepresentado a nivel regional, es decir, la región no tienen la capacidad para cubrir los requerimientos de insumos.

b) Coeficiente de localización de industria cruzada ($CILQ_i$)

$$CILQ_{ij} = \frac{SLQ_i}{SLQ_j} = \frac{\frac{E_{ri}}{E_{ti}}}{\frac{E_{rj}}{E_{tj}}}$$

donde SLQ_i es el coeficiente de localización simple del sector i y SLQ_j es el coeficiente del sector j .

c) Fórmula FLQ_{ij} (Flegg et al., 1995; Flegg y Webber, 1997)

$$FLQ_{ij} = CILQ_{ij} * \lambda^*$$

donde, $\lambda^* = [\log_2(1 + \frac{E_r}{E_t})]^\delta$, $\delta \in [0, 1]$. Para $i = j$, $CILQ_{ij} = SLQ_i$ y con $\delta = 0$ tenemos $\lambda^* = 1$ y $CILQ_{ij} = FLQ_{ij}$.

d) Fórmula FLQ_{ij} aumentada (Flegg y Webber, 2000)

$$AFLQ_{ij} = CILQ_{ij} * \lambda^* [\log_2(1 + SLQ_i)]$$

donde el último término se incluye para incorporar el efecto de la especialización regional. Si sólo permitimos $SLQ_j > 1$, entonces $AFLQ_{ij} > FLQ_{ij}$ para $SLQ_j > 1$, y $AFLQ_{ij} = FLQ_{ij}$ para $SLQ_j \leq 1$.

Entre las diferentes alternativas de LQ para regionalizar las MIPs, la fórmula FLQ_{ij} es la que arroja resultados menos sesgados y más precisos, incluso que los generados con el $AFLQ_{ij}$ (Flegg y Webber, 2000; Flegg y Tohmo, 2013b).

Existe otro método de regionalización ajustada, el *cross-hauling* de Kronenberg (2009) (CHARM), ésta metodología toma en cuenta el hecho de que las regiones pueden importar y exportar los mismos *commodities*. Los resultados de las estimaciones regionales con el CHARM son mejores, ya que es más efectivo al momento de incorporar los flujos interregionales (Flegg y Tohmo, 2013a).

En este trabajo utilizaremos las matrices regionales para conocer los vínculos insumo-producto al interior de las regiones, la alternativa para la regionalización que usamos es la Fórmula FLQ_{ij} de Flegg y Webber (1997) debido a que es sencillo y sólo nos interesa la

matriz de transacciones sectoriales. Entonces, para tener un buen ajuste utilizamos la MIP simétrica doméstica 2003 y 2008 del INEGI (2015) de tal forma que podamos concentrarnos en los flujos comerciales intraregionales.

Regionalización con la fórmula FLQ_{ij}

El objetivo es estimar coeficientes técnicos de comercio intersectoriales a partir de la MIP nacional (A) (Flegg y Webber, 1997, 2000; Dávila Flores, 2002).

- El primer paso es obtener los coeficientes técnicos (a_{ij}) que son la proporción entre los insumos intermedios y los insumos totales de cada sector.
- Después calculamos el coeficiente de localización simple (SLQ_i).¹
- El siguiente paso es estimar λ^* de la fórmula de Flegg y Webber (1997).²
- Enseguida construimos una matriz con los coeficientes de localización de industria cruzada ($CILQ_{ij}$) y en la diagonal utilizamos los coeficientes de localización simple (SLQ_i).
- Ahora podemos calcular la fórmula $FLQ_{ij} = CILQ_{ij} * \lambda^*$. En la matriz que generamos tenemos los coeficientes de comercio (t_{ij}), que miden la proporción en que cada insumo es abastecido localmente. Donde $t_{ij} \in [0, 1]$, por tanto, si $FLQ_{ij} \geq 1$, entonces, $t_{ij} = 1$ y si $FLQ_{ij} < 1$, $t_{ij} = FLQ_{ij}$.
- Para construir los coeficientes técnicos de comercio intersectorial (r_{ij}) multiplicamos los coeficientes técnicos nacionales y los de comercio:

$$r_{ij} = a_{ij} * t_{ij}$$

- Finalmente para obtener la MIP regional multiplicamos r_{ij} por el valor bruto de la producción de cada sector.

¹Aquí en lugar de empleo utilizamos el PIB nacional y el estatal.

²El valor de δ que generalmente se utiliza es 0.3 (Flegg y Webber, 1997; Flegg y Tohmo, 2013b). En esta estimación también utilizamos el PIB en lugar del empleo.

Con la regionalización que se ha descrito en los puntos anteriores podemos obtener la matriz de transacciones sectoriales que nos servirá de base para la construcción de la variable relacionada con el mercado de bienes de las externalidades Marshallianas.

Apéndice C

Regresiones adicionales

Cuadro C.1: Regresiones MCO con variables rezagadas, por región 2003-2008
(con *Escolaridad*)

Variable dependiente: Índice de Coaglomeración (2008)

Variables (2003)	Norte (1)	Pacífico (2)	Golfo (3)	Centro Nte (4)	Centro (5)	Capital (6)	Sur (7)
<i>Bienes</i>	0.730* (0.150)	0.605 (0.087)	0.794 (0.098)	-0.377 (-0.036)	0.141 (0.024)	0.005 (0.003)	0.424* (0.151)
	0.400	0.466	0.518	0.617	0.585	0.172	0.217
<i>Escolaridad</i>	-0.003 (-0.017)	0.015 (0.086)	0.006 (0.025)	0.013 (0.041)	0.006 (0.030)	-0.033*** (-0.339)	0.018 (0.158)
	0.013	0.020	0.028	0.014	0.014	0.011	0.012
<i>Ideas</i>	-0.029 (-0.042)	0.016 (0.048)	0.231*** (0.571)	-0.094 (-0.102)	0.029 (0.056)	0.026 (0.038)	-0.017 (-0.082)
	0.045	0.055	0.066	0.076	0.038	0.053	0.026
N	270	225	225	270	180	90	180
R^2	0.196	0.243	0.350	0.291	0.293	0.765	0.412

Errores *bootstrap* robustos entre paréntesis. Variables *dummies* por entidad y sector. La variable dependiente corresponde al año 2008 y las explicativas al año 2003. En el tercer renglón están los coeficientes estandarizados. *** p<0.01, ** p<0.05, * p<0.1

Cuadro C.2: Regresiones MCO con *dummies*, por región 2003-2008 (con *Escolaridad*)

Variable dependiente: Índice de Coaglomeración							
Variables	Norte (1)	Pacífico (2)	Golfo (3)	Centro Nte (4)	Centro (5)	Capital (6)	Sur (7)
<i>Bienes</i>	0.507** (0.143)	0.002 (0.000)	0.206 (0.031)	-0.119 (-0.015)	-0.102 (-0.022)	-0.009 (-0.006)	0.421*** (0.149)
	0.241	0.231	0.285	0.318	0.307	0.106	0.125
<i>Escolaridad</i>	0.011 (0.064)	0.005 (0.027)	0.023 (0.079)	-0.001 (-0.003)	0.007 (0.033)	-0.030*** (-0.290)	0.024*** (0.158)
	0.012	0.015	0.018	0.020	0.010	0.009	0.007
<i>Ideas</i>	0.020* (0.057)	-0.006 (-0.037)	0.067** (0.272)	-0.006 (-0.017)	0.010 (0.046)	0.006 (0.032)	0.023* (0.192)
	0.012	0.013	0.028	0.022	0.015	0.010	0.012
N	540	450	450	540	360	180	360
R^2	0.167	0.146	0.179	0.203	0.263	0.743	0.376

Errores *bootstrap* robustos entre paréntesis. Variables *dummies* por entidad, sector y año.

En el tercer renglón están los coeficientes estandarizados. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Índice de cuadros

2.1. Índice de coaglomeración por regiones, 2003-2008	114
2.2. Promedio nacional del índice de coaglomeración por subsectores de la industria manufacturera, 2003-2008	115
4.1. Regresiones MCO a nivel nacional, 2003-2008	125
4.2. Regresiones MCO agrupados y con <i>dummies</i> a nivel nacional, 2003-2008 . .	126
4.3. Regresiones MCO con variables rezagadas, por región 2003-2008	127
4.4. Regresiones MCO con <i>dummies</i> , por región 2003-2008	128
A.1. Industria manufacturera por subsector agregado	136
A.2. Clasificación de ocupaciones	137
C.1. Regresiones MCO con variables rezagadas, por región 2003-2008 (con <i>Escolaridad</i>)	144
C.2. Regresiones MCO con <i>dummies</i> , por región 2003-2008 (con <i>Escolaridad</i>) . .	145

Índice de figuras

2.1. Características del índice de coaglomeración nacional, 2003-2008	113
3.1. Estadísticas descriptivas: <i>Bienes</i> , 2003-2008	118
3.2. Estadísticas descriptivas: <i>Ocupación y Escolaridad</i> , 2003-2008	120
3.3. Estadísticas descriptivas: <i>Ideas</i> , 2003-2008	122
A.1. División política de la República Mexicana	138
A.2. Regionalización	139