ESSAYS ON CITY TRADE FLOWS

JORGE DÍAZ LANCHAS

UAM
UNIVERSIDAD AUTONOMA
DE MADRID

Ilustración por: Pak Morello (Pakmorelloarts)

Universidad Autónoma de Madrid

School of Economics and Management

Essays on City Trade Flows

Jorge Díaz Lanchas

A thesis submitted in partial fulfillment for
the degree of Doctor of Philosophy

Supervisors:
Jose Luís Zofío Prieto and Carlos Llano Verduras

2016

Department of Economics

THESIS COMMITTEE

Rafael Domenech
University of Valencia
BBVA Research

Raymond R. Flórax
Purdue University
Free University of Amsterdam

Andrés Rodríguez-Pose
London School of Economics
RSAI and ERC

Jesús Fernández-Huertas
Carlos III University
IAE(CSIC) and IZA

José M. Rueda-Cantuche
Institute for Prospective Technological Studies
European Commission

ALTERNATE COMMITTEE

Federico Steinberg
Autonomous University of Madrid
Elcano Royal Institute

Coro Chasco
Autonomous University of Madrid
Lawrence R. Klein Institute

EXTERNAL EVALUATORS

Horst Raff
Kiel University
Kiel Institute for the World Economy

Sofía Boza
Chile University
International Studies Institute

September 2, Madrid

A mis padres,
A mi hermano,
Y especialmente, a mis abuelos.

Agradecimientos

Más que el trabajo que representa esta Tesis, ésta muestra el fin de una etapa. Una etapa llena de acontecimientos, situaciones únicas e incluso aventurillas, que hacen que uno se tenga que parar unos minutos y mirar hacia atrás para hacer recuerdo de todo lo que ha ocurrido desde que en Diciembre de 2009, el día 16 concretamente, entrase a formar parte de la Universidad Auónoma de Madrid.

Esta fecha, no supuso el inicio del proyecto predoctoral, ni muchísimo menos. De hecho, puedo llegar a decir que en muy raras ocasiones he llegado a estar centrado completamente en la Tesis, pues muchas han sido las tareas y funciones que he desarrollado en estos años. Tareas que, por cierto, me han permitido no sólo aprender una serie de habilidades personales que no conocía de mi mismo cuand empecé el posgrado, sino que además me han servido para conocer a muchísimas personas y visitar lugares muy distintos que de no haber iniciado el posgrado, difícilmente hubiera conocido. Quizás ese sea el valor de una tesis doctoral. Quiero decir que quizás no sólo importa lo que realmente aporte con este trabajo o con los que están por llegar, eso el tiempo lo decidirá, sino que las vivencias y personas que he conocido en este tiempo posiblemente sean las que marquen el antes y el después de mi carrera profesional.

Como digo, empecé en la UAM a través del Instituto Klein/CEPREDE y en concreto gracias al proyecto C-intereg. Recuerdo que ya este inicio no me vino como caído del cielo, pues mi interés en realizar un posgrado venía de los primeros años de la carrera. Aquellos años en lo que pensaba que tras la carrera querría aprender economía "de verdad". Y lo cierto es que, viéndolo ahora en perspectiva, quizás no me equivocaba pues, aunque tuve profesores brillantes (de los que te marcan) y con los que pude aprender muchísimo, la curiosidad que sentía por profundir en aquellos fenómenos que estudiaba y que no terminaba de comprender, me empujó a realizar el posgrado una vez terminé la Licenciatura. No sé si hoy puedo decir que de verdad "sé algo de economía", posiblemente eso nunca lo sepa, pero creo que el posgrado, bien entendido por quien lo ejercita vocacionalmente, ayuda a estructurar los razonamientos y la forma de ver el mundo que dista mucho de la mentalidad juvenil con la que se termina la carrera. Es más, viendo la situación actual que están viviendo tanto España como Europa en el momento en el que escribo estas líneas, no puedo sino agradecer esta decisión de incursarme en el posgrado.

De esta época “undergraduate” me quedo con la sensación de haber conocido a profesores que me aconsejaron muy bien y a los que estoy muy agradecido. Desiderio Romero y María Arrazola fueron sin duda los más influyentes dentro de la Universidad Rey Juan Carlos, tanto por colaborar con ellos directamente, como por los consejos que me dieron. Sin merecer, claro está, las conversaciones con otros profesores a lo largo de toda la carrera. Pero incluso más relevante de esta época, es el haber conocido a uno de los mejores compañeros, y aún mejor persona (y la más vacilona), con los que no puedes parar de reir ni un minuto, Mario García.

Una vez ya en la UAM, no puedo sino recordar los inicios con compañeros del Klein. Empezando por Willy y siguiendo con Juan, Julián, Juanjo, Mila, José Antonio, Mar, los dos “Pepes” (con todos mis recuerdos hacia uno de ellos), e incluso al propio Antonio Pulido, muchos han sido los que han pasado por aquí y a los que tengo que agradecer su colaboración, de un modo u otro, como compañeros de trabajo. Todo ello sin olvidar especialmente a tres personas con los que he mantenido el contacto incluso después de marcharse del Klein, Martín, Beatriz Sánchez y Álvaro. Además, mención aparte desde la distancia merecen los inestimables socios del proyecto, que han antepuesto y peleado por mantener la financiacón a un proyecto como C-intereg haciendo que éste se mantenga y sobreviva durante los duros recortes en los primeros años de la crisis.

El proyecto C-intereg supuso la plataforma de empiece que se complementaría con la participación en el máster donde, por suerte, pude conocer a grandes compañeros y, mejor aún, amigos. Entre ellos destacan Javier, Bárbara y César. Respecto del primero, qué voy a decir. Compañero de máster, después de doctorado, y espero que a futuro, de investigación y trabajo. En estos años tengo que agradecerle su apoyo con multitud de programas (como éste con el que escribo esta tesis). A los segundos, simplemente les guardo un cariño muy especial por el trato y la relación que mantenemos, y por ese detalle que tuvieron al invitarme con el mejor recibimiento posible, tanto a su ciudad, Juárez (Méjico), como a su Universidad, donde me recibieron con todos los honores. A todos estos compañeros y profesores de Juárez, gracias.

Aparte de estos compañeros, otros han ido apareciendo a lo largo del doctorado y con los que mantengo una relación muy singular. A Pablo, Álvaro Remesal, Sofía y Ernesto, los he conocido en distintos momentos y bajo circunstancias distintas, y la verdad es que me alegro mucho de ello, no sólo como compañeros con los que espero poder llevar a cabo futuros proyectos juntos, sino también como colegas con los que poder debatir del mundo (Ernesto y Álvaro) o de cosas ya más personales (Pablo y Sofía). A los cuatro, un abrazo.

De entre todas estas personas, no he nombrado a dos personas muy especiales, sobretodo en los primeros años, Nuria y Tamara. Aunque con ambas he tenido que “lidiar” con el frente común que continuamente han hecho contra mi persona, a las dos las estoy

muy agradecido por ser grandes compañeras, y más aún amigas, aunque ello me implique ceder ahora (por un momento sólo) ante los "piques" constantes con Nury, cual hermanos postizos. Para ya terminar de formar este grupo de "doctorandos", es una alegría contar con nuevos integrantes como Santi, con quien seguro saldrán grandes proyectos a largo plazo.

Todo esto no hubiese sido posible de no ser por dos personas que desde distintos ángulos y enfoques han permitido que pudiese empezar mi carrera profesional, Carlos y José. Mientras que el primero ha supuesto ser como el mentor de todo el grupo de investigación, el segundo traía todo el rigor, la consistencia y la perseverancia en la manera de hacer las cosas. Los dos me han ayudado a desarrollarme como investigador, facilitando las estancias, la búsqueda de financiación y el apoyo, al fin y al cabo, técnico con el que empezar "from the scratch". Por ello, quiero darles mi más profundo aradecimiento, por estos años y los que están por llegar.

He de decir que durante estos años igualmente mi mundo no se ha cerrado únicamente a este grupo de investigación, si no que he procurado mantener relación constante con demás profesores del departamento, de entre los cuáles a algunos los considero algo más que simples colleagues (muy típica expresión del mundo académico). Entre éstos destacan, Raúl López, el hombre del eterno debate, Federico Steinberg, que tanto me ha ayudado y con el que podido iniciarme en las tareas como docente, y Julián Moral, la persona encargada de desmontar todas y cada una de las ideas que se me ocurren para investigar. A los tres, un fuerte abrazo. De hecho, y dentro de este grupo, bien podría nombrar a otros miembros como Coro Chasco, Inmaculada Álvarez y Alba Valdés por a los apoyos recibidos en ciertos momentos y de los que estoy muy agradecido.

En estos años he tenido la suerte de viajar en múltiples estancias y conocer a investigadores a cada cuál más bueno. Sin pretender hacer de menos a nadie, querría destacar a dos que me han sorprendido sobradamanente, Asier Minondo y Paco Requena. Creo que sin ningún tipo de dudas, esta tesis no hubiese sido la misma si no hubiese tenido la suerte de trabajar y recibir consejos tanto de Asier como de Paco. Es más, creo que en mi período de formación, hay un "salto" formativo y de experiencia, tras haber realizado la estancia doctoral en Deusto y haber podido aprender directamente de la forma de trabajar de estos dos grandes investigadores.

Querría remarcar que hasta ahora únicamente me he centrado en los aspectos profesionales, pero casi aún es más relevante el entorno personal que te rodea cuando uno está desarrollando una tesis doctoral. Tengo que decir que en eso he sido un privilegiado, pues el tipo de gente que me ha acompañado en estos años ha sido única. Empezando por los colegas de toda la vida de Coslada, tales como Paco (el artista de la portada de esta tesis), Fátima, Bego, Miki, Antonio, Brus, Itu, Dani, Diego, Miguel, a muchos otros que nombrarlos me llevaría mucho tiempo. Así mismo y siguiendo con los que he

llegado a conocer en tantas estancias destacaría a Svend, Raúl o Merce, a quienes pude conocer en la Universidad de Purdue, o Luis desde Amsterdam. Todos ellos conforman un grupo muy heterogéneo y con multitud de visiones que me han permitido mejorar mi visión y argumentos sobre muchos fenómenos que nos rodean. Mención aparte dentro de este grupo, resalta una personita única a la que querría nombrar. Dicen que no es bueno hacerlo en los agradecimienos de un tesis. No lo sé, quizás sea verdad, pero lo que sí sé es que, a día de hoy, y gracias a la paciencia y el apoyo aférrimo que ha mostrado tanto en estos dos últimos años como en los que están por llegar, querría agradecer a Cris, ya como amiga, todo su ánimo, buen hacer y ayuda. Un abrazo para ti también.

Finalmente y como las personas que realmente han estado ahí durante todos estos años, está mi familia. La paciencia y persevarancia que han mostrado en estos años mis padres y mi hermano, incluso cuando me he ido al extranjero sin que terminaran de entender del todo lo que hacía o el por qué lo hacía, merecen que mi más profundo agradecimiento sea para ellos. A los tres, muchísimas gracias por todo.

- **Acknowledgments in English**

  Apart, I would like to deeply acknowledge different professors outside Spain for their support when travelling abroad. First, I would like to thank Raymond Flórax for helping me and inviting me to Amsterdam to get his advise in one of the chapter of this Dissertation. On this regard, I fully appreciate the hospitality and support from Peter Mulder during my stay in Amsterdam. My experience with these great researchers is such that in the future I would like to have the opportunity to collaborate with them again.

  From the US, I would like to thank David Hummels for inviting me to Purdue University, but also to help me with my Dissertation. His advise and the possibility of attending his course were more than helpful.

  Finally, I would like to express my gratitude to Horst Raff for inviting me to Kiel University, and also Geoffrey Hewings for doing the same at Illinois University.

RESUMEN

Ensayos sobre flujos comerciales a nivel de ciudad

(Essays on city trade flows)

A lo largo de esta Tesis, se analizan tanto países desarrollados (España), como economías emergentes (Brasil), teniendo en cuenta, por un lado, cuestiones relacionadas con la jerarquía de ciudades y la propia estructura interna de las mismas, y por el otro, sus interacciones a través del comercio. Tras una primera introducción, la Tesis está compuesta por tres capítulos. Aunque los tres ensayos están estrechamente relacionados, cada uno de ellos se ha escrito de manera independiente para que puedan ser leídos de forma autónoma. Para introducir al lector en la temática de esta Tesis, a continuación se desarrolla un breve resumen de cada uno de estos capítulos.

- *Capítulo 2. Márgenes del comercio y áreas de mercado: Flujos municipales y jerarquía urbana.*

Mientras que la investigación reciente ha planteado múltiples modelos teóricos acerca de por qué los flujos de comercio tienden a aglomerarse en algunas ciudades específicas, el trabajo empírico en relación a esta literatura ha tendido a quedarse atrás. Haciendo uso de una base única de micro-datos sobre envíos de mercancías por carretera en España durante el período 2003-2007, conseguimos descomponer el valor total del comercio municipal en los dos márgenes del comercio, el extensivo y el intensivo. Gracias a ello, estudiamos el efecto obstaculizador que tienen los costes de transporte sobre dichos flujos comerciales atendiendo a varios niveles administrativos dentro de un país, NUTS-5 (municipal), NUTS-3 (provincial) y NUTS-2 (Comunidades Autónomas).

Con esto, confirmamos la acumulación de los flujos comerciales en distancias muy cortas. Al contrario de estudios anteriores (Hillberry and Hummels, 2008), encontramos que esta alta concentración de flujos comerciales no se explica por la existencia de límites administrativos que impiden el comercio (efecto de frontera), sino a cambios significativos en la relación entre el comercio y los costes de transporte. Para reforzar esta hipótesis, y recurriendo al Test de Chow endógeno, identificamos una serie de umbrales estadísticamente significativos para los cuáles se rompe de manera estructural la relación entre comercio y costes de transporte. En concreto, encontramos que los flujos comerciales tienden a aglomerarse en torno a las ciudades de mayor tamaño, siendo la mayor concentración a los 185€ de costes de transporte (alrededor de 130 kilómetros de distancia física). Más allá de esta distancia, la intensidad del comercio se reduce hasta un valor de costes de transporte de 330€ (unos 300 km), para posteriormente el comercio volverse prácticamente inexistente. De hecho, encontramos que para distancias muy largas, los flujos comerciales dentro de un país se llevan a cabo sólo entre las ciudades más importantes.

Estos puntos de corte nos permiten dividir la muestra y controlar por el efecto que tienen tanto sobre el margen extensivo como el intensivo, sucesivas fronteras administrativas. Aplicando este procedimiento a un modelo gravitatorio estándar, podemos separar correctamente el impacto que los costes de transporte (y las fronteras administrativas) tienen sobre los flujos comerciales. Gracias a ello, argumentamos que esta manera de dividir los costes de transporte resulta ser un procedimiento más apropiado para capturar el efecto no-lineal de los costes de transporte sobre el comercio, mejorando así proposiciones anteriores en la literatura (Eaton and Kortum, 2002; Abbate, et al., 2013).

Por último y más interesante, basándonos en estos umbrales de costes de transporte definimos áreas de mercado que configuran un sistema de jerarquía urbana coherente y basado en flujos comerciales que claramente evidencia las predicciones planteadas por la Teoría del Lugar Central (Christaller, 1930).

- *Capítulo 3. Especialización exportadora de las ciudades*

¿Existen patrones de exportación distintos entre ciudades grandes y pequeñas? Utilizando datos de comercio muy desagregados a nivel de producto para las ciudades brasileñas en el año 2013, encontramos que las zonas urbanas más pobladas exportan bienes proporcionalmente más complejos e intensivos en cualificación que las zonas urbanas menos pobladas. A su vez, mostramos que las ciudades brasileñas que han aumentado más en población han sido las que han aumentado más que proporcionalmente las exportaciones de bienes complejos e intensivos en cualificación.

Usando técnicas paramétricas y no paramétricas, nuestros resultados empíricos apoyan modelos recientes (Davis y Dingel, 2014; Hausmann y Hidalgo, 2011) que argumentan que las grandes ciudades, por un lado, atraen a los trabajadores con mayor cualificación, y por el otro, exhiben una amplia gama de capacidades que les proporciona ventaja comparativa en la exportación de bienes complejos e intensivos en cualificación.

- *Capítulo 4. La aparición de ciudades especializadas y diversificadas*

Muchos economistas se han ocupado de la cuestión fundamental acerca de por qué algunas ciudades se especializan y otras se diversifican. Determinar por qué surgen estos fenómenos ha resultado ser un rompecabezas teórico, que hasta el momento ha sido abordado a través del desarrollo de modelos cuyos supuestos difieren en la manera en la que se han (micro) fundamentado las fuerzas de aglomeración y dispersión, la movilidad de bienes y hogares o los propios mecanismos de formación de las ciudades.

Este capítulo busca contribuir a esta literatura realizando un análisis empírico que tiene como objetivo evaluar los mecanismos y las hipótesis fundamentales que se proponen desde la literatura teórica (Abdel-Rahman, 1990a,b, 1996; Abdel-Rahman y Fujita, 1993; Duranton y Puga, 2001; Henderson, 2000; Anas y Xiong 2003, 2005). A diferencia de la mayor parte de la literatura empírica existente que estudia la composición, el tamaño y la ubicación de las ciudades de Estados Unidos, en este capítulo nos centramos en España. En nuestro análisis consideramos la posibilidad de que las ciudades puedan ser no sólo especializadas o diversificadas, sino también una combinación de ambas (economías de co-aglomeración) o ninguna de las dos. Para ello, no sólo consideramos las características internas de las ciudades, sino que atendemos al rol que las interacciones entre ciudades pueden llegar a tener en la estructura productiva de las mismas.

Con este objetivo proponemos un enfoque empírico novedoso basado en el desarrollo de modelos probit bivariantes usando los micro-datos de tres bases únicas: una base de datos a nivel de empresa que nos permite calcular los niveles de especialización y diversificación de las ciudades; una base de datos de comercio entre ciudades españolas; y una base de datos acerca del coste económico de desplazamiento entre ciudades. Con todo y en consonancia con anteriores análisis para el caso de EE.UU. (Henderson, 2003), encontramos que una ciudad tiene una mayor probabilidad de convertirse en especializada si su mano de obra está relativamente poco cualificada, su estructura de producción está orientada hacia la fabricación de productos estandarizados, y su tamaño en términos de población es relativamente pequeño. Por su parte, el patrón obtenido es el opuesto para el caso de la aparición de ciudades diversificadas, mientras que una combinación de estos resultados ayudaría a explicar el surgimiento de las denominadas economías de co-aglomeración (Ellison y Glaeser, 1997; Helsley y Strange, 2014).

Además, encontramos que la historia importa: el grado de especialización en el pasado resulta ser un importante predictor de los patrones de especialización actuales. Por último y más interesante, encontramos que la probabilidad de que una ciudad se convierta en especializada (diversificada) depende positivamente de bajos (altos) costes de transporte en el pasado, mientras que el efecto contrario se observa para niveles de coste de transporte actuales. Estos resultados dan apoyo a las propuestas teóricas previas en los que se fusionan modelos de crecimiento urbano (Anas y Xiong, 2003, 2005) y de nueva geografía económica (Fujita et al., 1999). Estos resultados sugieren que, aunque la estructura interna de las ciudades es un factor determinante de su especialización y diversificación, la localización de la ciudad juega un papel incluso más crucial en el desarrollo de las ciudades.

- Bibliografía (Resumen)

1. Abbate, A., De Benedictis, L., Fagiolo, G., Tajoli, L. (2012). The International Trade Network in Space and Time. Laboratory of Economics and Management (LEM), No. 2012/17, Sant'Anna School of Advanced Studies, Pisa, Italy.

2. Abdel-Rahman, H.M. (1990a). Agglomeration economies, types, and sizes of cities. Journal of Urban Economics, 27 (1), 25–45.

3. Abdel-Rahman, H. M. (1990b). Sharable inputs, product variety and city sizes. Journal of Regional Science, 30 (3), 359–374.

4. Abdel-Rahman, H. M. and Fujita, M., (1993). Specialization and diversification in a system of cities. Journal of Urban Economics, 33 (2), 189–222.

5. Abdel-Rahman, Hesham M. (1996). When do cities specialize in production? Regional Science and Urban Economics, 26 (1), 1–22.

6. Anas, A. and Xiong, K. (2003). Intercity trade and the industrial diversification of cities. Journal of Urban Economics, 54 (2), 258–276.

7. Anas, A. and Xiong, K. (2005). The formation and growth of specialized cities: efficiency without developers or malthusian traps. Regional Science and Urban Economics, 35 (4), 445–470.

8. Christaller, Walter (1933). Central places in Southern Germany. Translated Edition.Prentice-Hall.

9. Davis, D. R. and Dingel, J. I. (2014). The Comparative Advantage of Cities. Working Paper 20602, National Bureau of Economic Research.

10. Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. American Economic Review, 91 (5), 1454–1477.

11. Eaton, J., Kortum, S. (2002). Technology, geography, and trade. Econometrica, 70 (5), 1741-1779.

12. Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. Journal of Political Economy, 105 (5), 889–927.

13. Fujita, M., Krugman, P. R., and Venables, A. (1999). The spatial economy: Cities, regions, and international trade. MIT press.

14. Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. Journal of Economic Growth, 16 (4), 309–342.

15. Helsley, R. W. and Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. Journal of Political Economy, 122 (5), 1064–1093.

16. Henderson, V. (2000). Political economy of city sizes and formation. Journal of Urban Economics, 48, 453–484.

17. Henderson, V. (2003). Marshallian scale economies. Journal of Urban Economics, 53, 1--28.

18. Hillberry, R. and Hummels, D. (2008). Trade responses to geographic frictions: A decomposition using micro-data. European Economic Review, 52(3):527 – 550.

CONCLUSIONES

A lo largo de esta Tesis se han obtenido numerosos resultados empíricos acerca de la de tipología de las ciudades y los patrones de comercio entre las mismas. En esta sección se plasman las principales recomendaciones de política económica que se derivan de estos resultados, así como las futuras líneas de investigación que se desarrollarán a tenor de esta Tesis.

Tras un primer capítulo introductorio, el primer de los artículos de la Tesis se centra en el papel que tienen los costes de transporte sobre los flujos de comercio en distancias muy cortas (alrededor de 150 kilómetros–185€) y cómo, alrededor de las grandes ciudades españolas, surgen áreas de mercado que delimitan sus zonas de influencia geográfica. A la hora de plantear estrategias de crecimiento regional, los responsables políticos deberían ser conscientes de esta concentración del comercio. En este sentido, debido a que algunas ciudades son más predominantes (poseen mayor influencia comercial y jerarquía) que otras dentro del sistema de ciudades españolas, éstas consiguen beneficiarse en mayor medida que el resto cuando se llevan a cabo políticas de infraestructuras públicas. De hecho, mejoras en la dotación de carreteras pueden generar una mayor aglomeración de la actividad comercial en torno a estas grandes ciudades. Este problema se vuelve aún más acuciante si consideramos que algunas ciudades poseen mayores ventajas de localización que otras dentro de la red de carreteras. A modo de ejemplo, para los mismos umbrales de distancia (áreas de mercado), Madrid puede alcanzar distancias más largas, dentro de España, que Barcelona (Figura 2.5, Mapas a y b), lo que le permite beneficiarse en mayor medida de mejoras en las infraestructuras.

Yendo un paso más allá de esta concentración de la actividad en distancias cortas, la inversión en infraestructuras durante la burbuja inmobiliaria española ha sido fuertemente criticada tras la aparición de estudios que arrojan evidencia sobre las ineficiencias en la asignación de recursos que tal inversión ha conllevado. En este sentido, trabajos como el de De La Fuente, (2010) muestran que la inversión en carreteras ha llegado a alcanzar tasas marginales de retorno negativas. No obstante, bien es cierto que, aunque estos efectos sobre la eficiencia de recursos sean bastante robustos, los efectos positivos de la inversión en infraestructuras no llegan a considerarse correctamente dentro de la literatura. De este modo, la reducción de los costes de transporte a través de las infraestructuras de carreteras en el período 2003-2007, parece impulsar el comercio entre ciudades (Tabla 2.2). Es más, estas infraestructuras llegan incluso a reducir la acumulación de flujos comerciales a nivel municipal (medido a través de un menor coeficiente en el año 2007 para el límite administrativo municipal en la Tabla 2.4). 4 Así mismo, conectar más municipios a través de la red de carreteras ha permitido que más municipios puedan comerciar más productos (variedades) con más destinos (Tabla 2.9), es decir, los municipios han experimentado mejoras en dos dimensiones de la diversificación: la de productos y la de destinos.

Los resultados obtenidos en este primer capítulo abren la puerta a futuros análisis centrados en la dimensión de producto a nivel municipal. En concreto, resulta de interés entender por qué sólo algunos productos son comerciados en cortas distancias, mientras que otros llegan a alcanzar distancias muy largas, creando con ello una jerarquía de ciudades en las que algunas muestran un impacto enorme en el comercio internacional debido a los productos que consiguen comerciar, mientras que otras ciudades únicamente se limitan a abastecer mercados locales. El denominado como efecto *Alchian-Allen* proporciona un marco tanto teórico (Borcherding y Silberberg, 1978) como empírico (Hummels y Skiba, 2004) a través de cuál se puede analizar los motivos que llevan a que algunas ciudades puedan especializarse en productos de alta calidad, permitiéndoles mejorar dentro de la estructura jerárquica de ciudades.

El segundo capítulo estudia el grado de especialización en productos complejos e intensivos en conocimiento de las ciudades, y cómo esta especialización surge gracias a que las (grandes) ciudades consiguen a traer trabajadores de alta cualificación. Si los productos complejos son los que permiten generar crecimiento económico (Hausmann y Hidalgo, 2011), deberíamos de esperar que algunas ciudades crezcan mucho más que otras, lo que provoca la aparición de divergencias entre unas regiones y otras dentro de un mismo país. En términos de política económica, la principal conclusión que se deriva de este análisis es la incapacidad de los representantes políticos para promocionar la especialización en bienes complejos y de alto conocimiento de las ciudades pequeñas. Esta incapacidad surge tan sólo por el hecho de que los trabajadores más cualificados tienen el incentivo a emigrar a las ciudades más grandes en su búsqueda por mejores oportunidades laborales. De ahí que, al llegar una crisis, sean las ciudades más grandes las que mejor pueden responder a la misma tan sólo por el mero hecho de poseer una mayor dotación de capacidades derivadas de estos trabajadores con alto capital humano.

Para evitar estas divergencias de largo plazo, los responsables políticos (a nivel nacional y local) tendrían que centrar sus esfuerzos en generar más capacidades en las ciudades pequeñas, de modo que puedan atraer a trabajadores de alta formación y, con ello, favorecer la especialización de estas ciudades en productos complejos. Estas capacidades pueden generarse no sólo gracias de la creación de grandes complejos de investigación (que requieren de una fuerte inversión inicial y sus efectos positivos son de muy largo plazo), sino favoreciendo las conexiones entre individuos, empresas e instituciones.

Dadas las dotaciones actuales de factores en las ciudades, un uso más intenso de las redes de información entre estos agentes, de modo que la interacción entre los mismos sea más intensa, puede resultar una manera mucho más barata y eficiente de que los responsables políticos a nivel local creen nuevas actividades en las ciudades, permitiendo con ello la generación de más capacidades y, por último, el crecimiento económico.

Este capítulo pertenece a un área de investigación muy relevante, esta es, hasta qué punto las ciudades se especializan en función de su "estadio de desarrollo". Estudios previos (Imbs y Wacziarg, 2003) han encontrado evidencia de la existencia de relaciones no-lineales entre el nivel de desarrollo de un país (PIB) y su grado de diversificación. En primeras fases del desarrollo, los países de especializan en productos eminentemente agrícolas, después diversifican su estructura productiva mediante la creación de nuevos sectores para, finalmente, volver a especializarse en productos de alto conocimiento. Esta pauta que se observa a nivel nacional resulta de especial interés analizarla para el caso de las ciudades. Si a nivel urbano se observa un patrón similar que en el conjunto del país, podríamos inducir que la actual especialización de las ciudades serviría para predecir patrones futuros de especialización de los países. De ahí que se deberían impulsar proyectos de investigación que analicen con detalle la estructura productiva de las ciudades.

El último capítulo arroja luz sobre la aparición de ciudades y cuáles son los factores que llevan a este fenómeno. Quizás la implicación más importante de este capítulo sea la de hacer una llamada de atención a la agenda investigadora en su conjunto y, más concretamente, a la necesidad por realizar futuras investigaciones, tanto teóricas como empíricas, relacionadas con la aparición de economías de co-aglomeración. Dentro de la literatura no está claro por qué algunas ciudades presentan ventajas tanto en diversificación como en especialización (economías de co-aglomeración). Una mayor y mejor compresión de este tipo de fenómenos serviría para entender el rol que juegan tanto las economías de aglomeración como las interacciones entre ciudades en la aparición de este tipo de ciudades. Todo ello con el intento de ayudar a los responsables políticos a encontrar las ventajas comparativas a nivel local.

En línea con el segundo capítulo, las implicaciones derivadas de éste tercero suponen ser un reto para las intervenciones de política económica. Por un lado, la aparición de ciudades está dirigida por las economías de aglomeración internas, especialmente aquellas relacionadas con el tipo de trabajadores y el conocimiento (formación) que éstos poseen. Por otra parte, los patrones de dependencia del pasado (path dependence) parecen jugar un papel fundamental en el desarrollo urbano. En concreto, previos niveles de población y costes de transporte ("ventajas de primera naturaleza") suponen ser aquellos que en mayor medida determinan la tipología de las ciudades en el futuro. Al cumplirse estos efectos, el espacio que se deja a la intervención de políticas resulta ser muy reducido a la hora de impulsar el crecimiento en las ciudades. De hecho, puesto que bajos costes de transporte en el pasado parecen afectar positivamente a la especialización futura de las ciudades, los responsables de políticos continuamente se estarían enfrentando a disyuntivas (trade-offs). Si las intervenciones políticas van encaminadas a reducir los costes de transporte con el objetivo de impulsar el comercio de las ciudades (implicación del primer capítulo), estas tendrían que hacerlo a expensas de saber que con ello favorecerán la aparición de ciudades especializadas, generando así un patrón de centro-periferia à la NEG dentro del país, y enfatizando de este modo las desigualdades entre ciudades. Siguiendo con esto, si las ciudades diversificadas son aquellas que promueven el crecimiento económico en el largo plazo, mientras que las ciudades especializadas únicamente se centran en la producción de productos estandarizados (Duranton y Puga, 2001), el hecho de reducir los costes de transporte generaría que algunas áreas (ciudades especializadas) no pudiesen crecer de la misma manera que otras (ciudades diversificadas).

Las líneas de investigación que surgen a raíz de este artículo posiblemente se encuentran entre las más interesantes a explorar a futuro. En primer lugar, la agenda investigadora tendría que profundizar en el rol que la intervención gubernamental puede tener en el desarrollo de las ciudades a través de políticas de crecimiento regional o mediante el propio federalismo fiscal. Así, la búsqueda de cuasi-experimentos naturales, como la creación del Estado de las Autonomías en España al principio de los años 80, podría servir como estrategia empírica con la que desentrañar y diferenciar los efectos que tanto la aglomeración económica como los pertenecientes a la acción de los gobiernos regionales tienen sobre la estructura productiva de las ciudades. Por último, los resultados obtenidos entre la caída de los costes de transporte y la reducción de las desigualdades en la distribución de la población en España (Figura 4.5), abren la puerta a futuras investigaciones que analicen este inesperado y contradictorio patrón, pues va en contra de proposiciones teóricas planteadas en trabajos de referencia como el de Tabuchi y Thisse (2011). En este sentido, la investigación tendría que centrarse en estudiar si dicha pauta tiene lugar sólo en España, o por el contrario, se observa en más países. De e ser así, la literatura teórica estaría fallando, lo que requeriría de una nueva formulación de modelos acerca de la evolución jerárquica de las ciudades.

- Bibliografía (Conclusiones)

1. De La Fuente, A. (2010). Infrastructures and Productivity: an Updated Survey. Barcelona Graduate School of Economics, 475.

2. Barbero, J. y Zofío, J. L. (2012). The multiregional core-periphery model: The role of the spatial topology. Networks and Spatial Economics, 1-28.

3. Borcherding, T.E. y Silberberg, E. (1978) "Shipping The Good Apples Out: The Alchian And Allen Theorem Reconsidered." The Journal of Political Economy, 86 (1), 131-138.

4. Duranton, G. y Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. American Economic Review, 91 (5), 1454–1477.

5. Hausmann, R. y Hidalgo, C. A. (2011). The network structure of economic output. Journal of Economic Growth, 16 (4), 309–342.

6. Hummels, D. y Skiba, A. (2004). Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture. Journal of political economy, 112 (6), 1384-1402.

7. Imbs, J. y Wacziarg, R. (2003). Stages of diversification. American Economic Review, 63-86.

8. Tabuchi, T. y Thisse, J.-F. (2011). A new economic geography model of central places. Journal of Urban Economics, 6 (4), 240–252.

Contents

Agradecimientos **7**

1 Introduction **19**

1.1 Chapter 2. Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy 21

1.2 Chapter 3. Cities Export Specialization 22

1.3 Chapter 4. On the emergence of diversified and specialized cities 23

1.4 Bibliography 24

Introducción (In Spanish) **27**

1.5 Capítulo 2. Márgenes del comercio y áreas de mercado: Flujos municipales y jerarquía urbana 30

1.6 Capítulo 3. Especializacíon exportadora de las ciudades 31

1.7 Capítulo 4. La aparición de ciudades especializadas y diversificadas 31

1.8 Bibliografía 33

2 Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy **36**

2.1 Introduction 36

2.2 Trade flows and their extensive and intensive margins 41

2.2.1 Trade value data: The road freight transportation survey 41

2.2.2 The generalized transport cost (GTC) 42

2.2.3 Trade flows decompositions 43

2.2.4 Analysis in growth rates 44

2.2.5 Descriptive analysis and Kernel regressions 46

2.3 Econometric specification: The general gravity model 49

2.4 Structural trade patterns: breakpoints analysis 56

2.5 Trade areas and the hierarchy of Spanish cities 65

2.6 Conclusions 70

2.7 Bibliography 73

2.8 Appendix 2.A: Kernel Regressions . . . 76

3 Cities Export Specialization 73

3.1 Introduction . . . 73
3.2 Data . . . 75
3.2.1 Urban areas . . . 75
3.2.2 Trade flows . . . 75
3.2.3 The complexity and skill-intensity indicators . . . 77
3.3 Empirical analyses . . . 78
3.3.1 Descriptive analysis: Brazilian cities trade pattern . . . 78
3.3.2 A non-parametric analysis: The elasticity test . . . 79
3.3.3 A parametric analysis: The log-supermodularity regression . . . 82
3.4 Conclusions . . . 84
3.5 Bibliography . . . 85
3.6 Appendix 3.A: Spanish trading-cities . . . 87
3.6.1 Appendix 3.A: Data . . . 87
3.6.2 Appendix 3.A: Empirics . . . 89
3.7 Appendix 3.B: The pairwise comparisons test . . . 92

4 On the Emergence of Diversified and Specialized Cities 94

4.1 Introduction . . . 94
4.2 Methodology . . . 96
4.2.1 Indices for Diversification and Specialization in cities . . . 96
4.2.2 Indices for city size, agglomeration economies and cities' interactions 97
4.2.3 Econometric Specification . . . 100
4.3 Data . . . 102
4.3.1 Service and Manufacturing Data . . . 102
4.3.2 Urban Areas . . . 103
4.3.3 The Generalized Transport Cost (GTC) . . . 105
4.3.4 Trade value data: The road freight transportation survey . . . 108
4.4 Empirics . . . 109
4.4.1 Descriptive Analysis . . . 109
4.4.2 Results: Baseline model . . . 112
4.4.3 Path Dependency . . . 118
4.5 Discussion . . . 123
4.6 Conclusions . . . 125
4.7 Bibliography . . . 127
4.8 Appendix 4.A: Data Representativeness . . . 130
4.9 Appendix 4.B: Robustness Analysis . . . 134

5 Conclusions and Policy Implications **137**
5.1 Bibliography 140

Conclusiones (In Spanish) **141**
5.2 Bibliografía 145

List of Tables

2.1 Transport costs variation along the period. 45
2.2 Variation effect of transport costs on trade flows. 46
2.3 Fixed effects estimation with GTC. 53
2.4 Cross-section regressions for 2003 and 2007 (First Level Decomposition). 55
2.6 Total Value of Trade by GTC thresholds. 61
2.7 Extensive Margin (Number of Shipments) by GTC thresholds. 62
2.8 Intensive Margin (Average Value per Shipment) by GTC thresholds. . . . 63
2.9 Shipments distribution by municipalities and products. 67
2.10 Number of Cities by Cluster-Type. 68

3.1 Descriptive Statistics. 76
3.2 Results for the elasticity test (% of successful comparisons) 82
3.3 Log-supermodularity regressions. Regression results for 2013. 83
3.4 Log-supermodularity regressions. Panel regression results for 2000, 2005 and 2013. 84
3.5 Results for the elasticity test (% of weighted successful comparisons) . . . 91
3.6 Log-supermodularity regressions. OLS Regression results for 1988 and 2012. 91

4.1 Categorization of Cities. 97
4.2 Probability for each type of city. 101
4.3 Descriptive statistics. 110
4.4 Transition Matrix between type of cities (1995-2007). 112
4.5 Regression Results from Bivariate Probit Model (Baseline Model, 1995-2007). 114
4.6 Marginal Effects from Bivariate-Probit Model (Baseline Model, 1995-2007). 115
4.7 Regression Results from Probit Model (with Lagged Variables) 120
4.8 Marginal Effects from Bivariate-Probit Model (with Lagged Variables). . 122
4.9 Ratios used for representativeness of SABI Database. 130
4.10 Firm Level: SABI Vs.Official Database (by year) 131
4.11 Industry Level: SABI Vs.Official Database (all years) 132

4.12 Province Level, SABI Vs.Official Database (all years). 133
4.13 Regression Results from Probit Models 1995-2007 (using Av. Firm Size) 134
4.14 Marginal Effects from Bivariate-Probit Model, 1995-2007 (Av.Firm Size). 135

List of Figures

2.1 Standard deviation of the Total Trade for the 633 municipalities. Average Values (Exports+Imports). Period 2003-2007. 47
2.2 Kernel regressions: Total value of trade, number of shipments (extensive margin) and average value per shipment (intensive margin) on GTC. . . 48
2.3 Kernel regressions: number of commodities and number of shipments per commodity (frequency) on GTC. 48
2.4 Kernel regressions: average price and average tons on GTC. 49
2.5 Natural trade areas using GTC Breakpoints for the whole period 2003-2007. 59
2.6 Kernel regressions for different city-cluster. 69
2.7 Urban Hierarchy and Natural Trade Areas for Dominant Cities. First, Second and Third GTCs Breakpoints. 70
2.8 Trade areas competition for Madrid and Barcelona municipalities (NUTS-5). 76
2.9 Trade areas competition for Madrid and Barcelona provinces (NUTS-3). . 77

3.1 Urban areas in Brazil (2013). 76
3.2 Complexity Vs. Skill-Intensity Indicators (2013). 77
3.3 Cities trading-patterns (2013). 78
3.4 Ratio of exports shares big cities/small cities vs complexity and skill-intensity (2013). 79
3.5 The population elasticity of exports vs. complexity and skill intensity (2013). 81
3.6 Share of educated people over occupied urban population (1982-2012). . . 87
3.7 Spanish Urban Areas (2012). 88
3.8 The relationship between the number of exported products and population. 89
3.9 Industries' population elasticities and skill intensities. 90

4.1 Diversification and Specialization by Urban Area (1995 and 2007). 104
4.2 Population patterns between 1995 and 2007. 105
4.3 GTC reduction rates by Urban Area (1995-2007). 106
4.4 Average Annual Transport Cost Changes (1980-2007). 107

4.5 Population patterns between 1980 and 2007. 107
4.6 Inter-city trade growth rates by Urban Area (1995-2007). 108
4.7 Four type of cities in Spain for 1995 and 2007. 111
4.8 Marginal effects for each city by regressors, 1995-2007 (Baseline Model). . 117
4.9 Marginal effects for trading partners, 1995-2007 (Baseline Model). 118
4.10 Marginal effects for each city by lagged variables. 123
4.11 Marginal effects for each city by regressors, 1995-2007 (Bivariate Baseline Model). 136

Chapter 1

Introduction

The following statement gathers the main insights of this Dissertation:

> *"The scale at which specialization of individuals leads to diversification is the city. Larger cities are more diversified than smaller cities. Among cities with similar populations, more diversified cities are richer than less diversified cities. They tend to grow faster and become even more diversified, not only because they have larger internal markets, but also because they are more diversified in terms of what they can sell to other cities and countries."* (Ricardo Hausmann, *The specialization myth*, The Project Syndicate, December 30, 2013).

The globalization process has put cities in the spotlight. In today's world an increasingly share of world population is getting interconnected. Over half of the world population (54%) now lives in urban areas. Whereas this share only represented 30% of total population in 1950, the future becomes even more promising. The coming decades will bring further changes to the size and spatial distribution of population as 66% of total population is projected to live in cities by 2050, being the highest concentration in Asia and Africa with nearly 90% of total urban growth (United Nations, 2015).

Understanding not only the internal urban mechanisms but how cities relate has become a challenge for improving citizens' quality of life. In 2015, the World Bank recognizes the relevance of cities as key drivers for jobs improvements and economic growth. For an economy to be competitive, its cities must be competitive first. In 2012, 70% of world cities outperformed their national economies in productivity. However this trend entails considerable variations across countries, regions and cities. In the Middle East and South Asia, the top 10% best-performing cities increased productivity 3.0% faster than their peers. In Africa, the top 10 percent of city performers increased productivity by around 1.7%. In China the results are even more surprising as the top-performing cities, such as Tangshan (Hebei) and Dongguan (Guangdong), exceed by more than seven times the national average productivity. Even in the United States, with a mature urban system, in the most productive city, San Jose (California), the output per worker doubles that of the least productive one, Buffalo (New York), Kilroy, et al., (2015).

This disparity is driven by, among other things, differences in city economic structures and industry specializations. Depending on the cities' stage of development, their national economies will subsequently evolve and transform. For instance, at national GDP per capita below $2,500, cities are typically market towns that will need to industrialize and transform to increase their incomes; At GDP per capita from $2,500, cities are typically "production centers" that can grow, to around $20,000 GDP per capita, by increasing the value of their existing industry mix; To rise above national GDP per capita of $20,000, cities will typically need to move towards higher value creative and financial services.

These performance-growth patterns results in the emergence of a hierarchical structure among cities. In particular, whereas some cities get larger by taking advantage of first nature endowments (location, natural resources), hosting most productive firms and attracting the brightest minds in a country (or world), others lag behind with poor expectations for improvements. These disappointing paths arise in cities that fail in finding their *comparative advantages*. The main source of cities advantages comes from networks connecting people-to-people and people-to-ideas. This is the *"true secret"* of cities success (Glaeser, 2011). Those cities that are successful in "connecting people", will get diverse by allowing their citizens to realize their individual advantages. By contrast, those that fail in promoting human collaborations will worsen their future performance.

In a globalized world, connections do not only come from cities internal structure but from interactions with other cities in different parts of the world. These interactions mainly take the form of investment (FDI) and trade flows. While FDI flows are more easily accounted for at the city level because of their initial firm-nature (Head and Ries, 1996), trade flows are usually considered at the national level due to the lack of databases at a lower regional scale. It brings failures in out understanding of cities' interactions and how these interactions result in different cities' internal structure.

This Dissertation intends to make a contribution on this regard. In particular, it focuses on how cities structure evolves when trade flows are accounted for. From different perspectives and methodologies, each chapter tries to give insights and empirical evidence on the challenges that cities faces when they trade.

In the spirit of a new wave of research projects that analyze cities economic and trading structure,[1] the interest of this Dissertation is even greater as it considers trade

[1] In the recent years, new projects from prominent international reference institutions have arisen in an attempt for understanding cities performance. Among these projects, we highlight those from the OECD such as the "Metropolitan Explorer" (http://measuringurban.oecd.org/) that focuses

flows between cities not only internationally, but also internally when cities are located within the same country (Hillberry and Hummels, 2003, 2008; Wolf, 2000).

As pointed out before, trade flows usually are recorded at national level, but countries do not trade among themselves; it is their firms that undertake trading activities (Eaton and Kortum, 2002; Mayer and Ottaviano, 2008; Melitz, 2003). Even more, firms do not trade regardless of their surrounding contexts, as if they were isolated and independent points in space. They trade thanks to the economies of scale (agglomeration economies) embedded in the places where they are located (Behrens and Robert-Nicoud, 2015). That is the relevance of cities.[2]

Along this Dissertation, not only developed countries (Spain) but emerging economies (Brazil) are analyzed, considering issues related to cities hierarchical and internal structures, as well as their cities' interactions when trading. While the three essays are closely related, each one of them has been written so that they can be read independently. To introduce the reader to the scope of the Dissertation, a brief summary of each chapter is now included.

1.1 Chapter 2. Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy

Recent research has proposed theoretical foundations on why trade flows tend to agglomerate in some specific cities, but empirical work has lagged behind. Using a micro-database on road freight shipments within Spain for the period 2003-2007, we consistently decompose the total value of municipal freight flows into the extensive and intensive margins of trade to study the impeding effect of actual generalized transport costs on trade

on metropolitan indicators; the "City Programme" (`http://citiesprogramme.com/`) from the United Nations that extends the analysis not only to cities economic activity but to sustainability and governance issues; the "Urban Development" project from the World Bank (`http://www.worldbank.org/en/topic/urbandevelopment`) that moves to poverty and city design; or even promising projects at the country level such the "Metropolitan Policy Program" from Brookings Institution (`http://www.brookings.edu/about/programs/metro`) that analyzes US cities from multiple dimensions including trade flows, FDI or economic activity; or the Dataviva project (`http://dataviva.info/`) from MIT media Lab; and the "Atlas de Complejidad Económica" (`http://complejidad.datos.gob.mx/`) from the CID Harvard that do the same for the Brazilian and the Mexican cases, respectively.

[2] Agglomeration economies are crucial in explaining trade flows. As examples, for the US case, around 86% of total trade flows are developed by only 100 cities (data from the Export Monitor 2015, Brookings Institution). Among other cases different from the US, around 45.5% of the Brazil's total exports in 2013 were carried out by cities, whereas this proportion represented the 65% in 2000 (own elaboration from the Dataviva project). not only internationally, but intra-nationally, when attending to trade within a country, for the Spanish case, around 79% of total value of trade (with the origin and destination in the same country) is carried out by 26 municipalities (own elaboration from the EPTMC/RFTS Survey from the Ministerio de Fomento).

flows at the European Nuts-5 (municipal), 3 (provincial) and 2 (regional) administrative levels within a country.

We confirm the accumulation of trade flows at very short distances. Contrary to previous studies (Hillberry and Hummels, 2008), we find that this high density is not explained by the existence of administrative limits (border effects) but to significant changes in the trade flows-transport costs dependency. To support this hypothesis, we identify significant thresholds in the trade flows-transport costs relationship that are calculated by way of the endogenous Chow Test for structural changes. In particular, trade flows tend to agglomerate around larger cities, being the highest agglomeration at a transport cost value of 185€ (around 130km of physical distance).[3] Beyond this distance, trade intensity reduces up to a transport cost value of 330€ (around 300km), to latter being mainly non-existent. Indeed, we find that for very long distances, trade flows within a country are only carried out between the biggest cities.

These breakpoints allow us to split the sample and control for successive administrative borders in both the extensive and intensive margins. Applying this procedure to a naïve gravity model, allows us to correctly disentangle the effect that transport costs (and administrative borders) have on trade flows. We argue that this way of splitting transport costs is a more appropriate procedure to capture the non-linear effect of transport costs on trade flows, improving previous propositions in the literature (Eaton and Kortum, 2002; Abbate, et al.,2013).

Moreover and more interesting, relying on these thresholds we define relevant market areas corresponding to specific transport costs values that configure a consistent urban hierarchy system based on trade flows, thereby providing clear evidence of the predictions made by the Central Place Theory (Christaller, 1930).

1.2 Chapter 3. Cities Export Specialization

Do large and small cities exhibit different patterns of export specialization? Using highly disaggregated product-level trade data for Brazilian cities in year 2013, we find that more populated urban areas export proportionately more complex and skill-intensive goods than less populated urban areas. We also show that Brazilian urban areas that have increased more in population have also augmented more than proportionately the exports of complex and skill-intensive goods.

Resorting to non-parametric and parametric techniques, our empirical findings support recent models (Davis and Dingel, 2014; Hausmann and Hidalgo, 2011) which argue that large cities attract more skilled workers and exhibit a wide range of capabilities,

[3]For this first breakpoint (185€), Behrens, et al., (2015) obtain similar distance values for the geographical concentration of Canadian manufacturing industries.

providing them a comparative advantage in skill-intensive and complex goods.

1.3 Chapter 4. On the emergence of diversified and specialized cities

Many economists have dealt with the fundamental question as to why some cities become specialized and other diversified. Determining why this is so, has been approached mainly as a theoretical puzzle, through the development of models that differ in their assumptions about the (micro foundations of) agglomeration and dispersion forces, mobility of goods and households and city formation mechanisms.

We seek to contribute to this literature by means of an empirical paper that aims to assess the key mechanisms and assumptions proposed in the theoretical literature (Abdel-Rahman, 1990a,b, 1996; Abdel-Rahman and Fujita, 1993; Duranton and Puga, 2001; Henderson, 2000; Anas and Xiong 2003, 2005). In contrast to most of the existing empirical literature about the composition, size and location of cities, we present evidence from outside the US, namely Spain. In our analysis we consider the possibility that cities can be not only either specialized or diversified, but also either a combination (co-agglomeration economies) or none of this. We do not only consider internal cities characteristics but also the role of cities interactions as potential determinant of sector composition differences across the urban system.

To this aim, we propose an innovative empirical approach, where we develop bivariate probit models to data from three micro-datasets: a firm-level database that permits us to calculate levels of specialization and diversification at the city level; a trade database between Spanish cities; and a unique database on economic transport cost of moving between Spanish cities. We find, in line with previous evidence for the US (Henderson, 2003), a higher probability for a city to become specialized if its labor force is relatively low skilled, its production structure is biased towards manufacturing and standardized products, and its size is relatively small. The opposite results increase the probability that a city becomes diversified, while a mix of these results help explaining the emergence of so-called co-agglomeration cities (Ellison and Glaeser, 1997; Helsley and Strange, 2014).

In addition, we find that history matters: the degree of specialization in the recent past proves to be a strong predictor of current specialization patterns. Finally, and most interestingly, we find that the probability of becoming a specialized (diversified) city depends positively on low (high) transport costs in the past, while the opposite holds for more recent and present-day transport costs. These puzzling results give support for previous theoretical propositions that merge urban growth models (Anas and Xiong, 2003, 2005) and the new economic geography literature (Fujita et al., 1999). We hypothesize that these findings suggest that over time the internal structure of cities has become a more important driver of specialization and diversification patterns across the urban system, but the former location of the city plays even a more crucial role in city development.

1.4 Bibliography

1. Abbate, A., De Benedictis, L., Fagiolo, G., Tajoli, L. (2012). The International Trade Network in Space and Time. *Laboratory of Economics and Management (LEM)*, No. 2012/17, Sant'Anna School of Advanced Studies, Pisa, Italy.

2. Abdel-Rahman, H. M. (1990a). Agglomeration economies, types, and sizes of cities. *Journal of Urban Economics*, 27 (1), 25–45.

3. Abdel-Rahman, H. M. (1990b). Sharable inputs, product variety and city sizes. *Journal of Regional Science*, 30 (3), 359–374.

4. Abdel-Rahman, H. M. and Fujita, M., (1993). Specialization and diversification in a system of cities. *Journal of Urban Economics*, 33 (2), 189–222.

5. Abdel-Rahman, Hesham M. (1996). When do cities specialize in production? *Regional Science and Urban Economics*, 26 (1), 1–22.

6. Anas, A. and Xiong, K. (2003). Intercity trade and the industrial diversification of cities. *Journal of Urban Economics*, 54 (2), 258–276.

7. Anas, A. and Xiong, K. (2005). The formation and growth of specialized cities: efficiency without developers or malthusian traps. *Regional Science and Urban Economics*, 35 (4), 445–470.

8. Behrens, K., Bougna, T., Brown, W. M. (2015). The world is not yet flat: Transport costs matter!, *CEPR Discussion Paper*s, No. 10356.

9. Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In *G. Duranton, J.V. Henderson, and W. Strange (eds.) Handbook of Regional and Urban Economics*, Volume 5. Amsterdam: Elsevier.

10. Christaller, Walter (1933). *Central places in Southern Germany.* Translated Edition. Prentice-Hall.

11. Davis, D. R. and Dingel, J. I. (2014). The Comparative Advantage of Cities. *Working Paper 20602*, National Bureau of Economic Research.

12. Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91 (5), 1454–1477.

13. Eaton, J., Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70 (5), 1741-1779.

14. Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105 (5), 889–927.

15. Fujita, M., Krugman, P. R., and Venables, A. (1999). *The spatial economy: Cities, regions and international trade*. MIT press.

16. Glaeser, E. L. (2011). *Triumph of the city: How our greatest invention makes US richer, smarter, greener, healthier and happier*. Pan Macmillan.

17. Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16 (4), 309–342.

18. Head, K., Ries, J. (1996). Inter-city competition for foreign investment: Static and dynamic effects of China's incentive areas. *Journal of Urban Economics*, 40 (1), 38-60.

19. Helsley, R. W. and Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, 122 (5), 1064–1093.

20. Henderson, V. (2000). Political economy of city sizes and formation. *Journal of Urban Economics*, 48, 453–484.

21. Henderson, J. V. (2003). Marshallian scale economies. *Journal of Urban Economics*, 53, 1–28.

22. Hillberry, R., Hummels, D., (2003). Intranational home bias: Some explanations. *Review of Economics and Statistics*, 85 (4), 1089–1092.

23. Hillberry, R. and Hummels, D. (2008). Trade responses to geographic frictions: A decomposition using micro-data. *European Economic Review*, 52 (3), 527-550.

24. Kilroy, A., Francis L., Mukim, M. and Negri, S. (2015). Competitive cities for jobs and growth : what, who, and how. *Competitive cities for jobs and growth*. World Bank Group, Washington, D.C.

25. Mayer, T., Ottaviano, G. I., (2008). The happy few: The internationalisation of european firms. *Intereconomics*, 43 (3), 135–148.

26. Melitz, M. J., (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71 (6), 1695–1725.

27. United Nations (2015). World Urbanization Prospects: The 2014 Revision, *(ST/ESA/SER.A/366)*, Department of Economic and Social Affairs, Population Division .

28. Wolf, H. C., (2000). Intranational home bias in trade. *Review of economics and statistics,* 82 (4), 555–563.

Introducción (In Spanish)

La siguiente cita enmarca el contexto de esta Tesis Doctoral:

> "La escala en la cual la especialización de las personas conduce a la diversificación es la ciudad. Las ciudades más grandes son más diversificadas que las pequeñas. Entre las ciudades con poblaciones similares (...) las más diversificadas son más ricas que las menos diversificadas. Además, tienden a crecer más rápidamente y a diversificarse aún más, no solo porque cuentan con un mayor mercado interno, sino porque también son más diversificadas en términos de lo que pueden vender a otras ciudades y países."(Ricardo Hausmann, *El mito de la especialización*, The Project Syndicate, December 30, 2013[4]).

El proceso de globalización ha puesto las ciudades en el punto de mira. En la actualidad una proporción cada vez mayor de la población mundial vive interconectada. Más de la mitad de esta población (54%) vive en zonas urbanas. Mientras que este porcentaje tan sólo suponía el 30% de la población total en 1950, se espera que éste alcance el 66% de la población total en el año 2050. Es más, se preveé que en las próximas décadas estos cambios sean aún más pronunciados, tanto en el tamaño como en la distribución espacial de la población, llegando a concentrarse casi el 90% del crecimiento total de las ciudades en Asia y en África (Naciones Unidas, 2015).

Comprender tanto los mecanismos urbanos internos como el modo en el que se relacionan las ciudades, se ha convertido en un reto para mejorar la calidad de vida de los ciudadanos. En 2015, el Banco Mundial ya reconocía la importancia de las ciudades como motores del desarrollo económico y la generación de empleo. A este respecto, el Banco Mundial concluía que, para que una economía sea competitiva, primero sus ciudades han de ser competitivas. Así, en 2012 el 70% de las ciudades del mundo superaba en productividad a sus respectivas economías nacionales. Sin embargo, esta tendencia conlleva diferencias considerables entre países, regiones y ciudades. Mientras que en Oriente Medio y el Sur de Asia el 10% de las ciudades más competitivas superaban en más del 3% la productividad de ciudades similares en tamaño, en África, este mismo 10% alcanzaba

[4]Traducción realizada por Leopoldo Gurman dentro de la misma web The Project Syndicate

un productividad en torno al 1,7%. Por su parte, en el caso chino los resultados son aún más sorprendentes pues las ciudades con mejor rendimiento (Tangshan y Dongguan), superaban en más de siete veces la productividad media nacional. Incluso en los Estados Unidos, con un sistema urbano mucho más maduro, en la ciudad más productiva, San José (California), el producto por trabajador duplicaba el de la ciudad menos productiva, Buffalo (Nueva York), Kilroy, et al., (2015) .

Esta disparidad se debe, en parte, a las diferencias en la estructura económica de las ciudades y su especialización productiva. Dependiendo de la etapa de desarrollo en la que se encuentre cada ciudad, sus respectivas economías nacionales evolucionarán y transformarán posteriormente. Por ejemplo, para un PIB nacional per cápita por debajo de los $2.500, cabe esperar que las ciudades se vean caracterizadas como "pueblos mercantiles" (market towns) que necesitarán industrializarse para que sus economías nacionales crezcan. Para un PIB per cápita superior a los $2.500, las ciudades tienden a evolucionar hacia "centros de producción" que permiten crecer y con ello aumentar el PIB nacional hasta los $20.000 por habitante. Finalmente, para que el PIB nacional se eleve por encima de los $20.000 de renta per cápita, sus ciudades han que cambiar su estructura hacia sectores basados en los servicios financieros y, especialmente, en el desarrollo de actividades creativas (que permitan generar ideas).

Estos patrones de crecimiento se traducen en la aparición de estructuras jerárquicas entre ciudades. Así, mientras que algunas ciudades se quedan atrás bajo unas expectativas pésimas de crecimiento, otras se vuelven más grandes y productivas gracias, primero, al aprovechamiento de los factores que las rodean (ubicación geográfica y recursos naturales) y, posteriormente, a la atracción tanto de las empresas más eficientes como de los trabajadores con mayor talento dentro de su país (o incluso del resto mundo). Estas patrones tan divergentes surgen como consecuencia de existir ciudades que no logran encontrar sus *ventajas comparativas*.

La fuente de estas ventajas proviene de las redes de conexión entre agentes que tienen lugar dentro de las ciudades. El hecho de que las ciudades sean capaces de, en corta distancia, concectar indivíduos entre sí y que estos individuos a su vez puedan generar nuevas ideas, supone ser el *"verdadero secreto"* de las ciudades (Glaeser, 2011). De ahí que sean las ciudades que "conecten personas" entre sí las verdaderamente exitosas, pues se beneficiarán de la diversidad de ventajas comparativas que se dan a nivel individual. Por el contrario, las ciudades que fallan en la generaión de vínculos y colaboraciones entre individuos, serán las que empeoren sus crecimientos futuros.

En un mundo globalizado, las conexiones entre individuos, no sólo provienen de las propias ciudades, sino también de las interaciones que mantienen con otras ciudades situadas en diferentes partes del mundo. Estas interacciones toman la forma de inversiones (IED) y flujos comerciales. Mientras que los primeros pueden ser fácilmente contabilizados a nivel de ciudad debido a que su forma originaria parte de la propia empresa (Head y Ries, 1996), los flujos comerciales, por el contrario, suelen ser contemplados a nivel de país como consecuencia de la falta de datos a escalas regionales inferiores. Esta falla provoca que surjan problemas a la hora de comprender la naturaleza de las interacciones entre ciudades, y la forma en que éstas pueden llegar a modificar la propia estructura interna de las mismas.

Esta Tesis pretende hacer una contribución en este sentido. En concreto, se centra en cómo la estructura de las ciudades evoluciona cuando tenemos en cuenta los flujos comerciales. Desde diferentes perspectivas y metodologías, cada capítulo analiza diferentes ideas y resultados empíricos con la intención de plasmar los retos a los que se enfrentan la ciudades cuando éstas interaccionan a través del comercio.

En línea con una nueva ola de proyectos que estudian la estructura económica y comercial de las ciudades,[5] el interés de esta Tesis es aún mayor al considerar flujos de comercio tanto internacionalmente como entre ciudades dentro de un mismo país (Hillberry y Hummels, 2003, 2008; Wolf, 2000).

Como se ha señalado anteriormente, los flujos comerciales por lo general se registran a nivel nacional, pero los países no comercian entre sí, sino que son sus empresas las que llevan a cabo la actividad exportadora (Eaton y Kortum, 2002; Mayer y Ottaviano, 2008; Melitz, 2003). Es más, las empresas no llegan a comerciar con independencia del contexto que las rodea como si fuesen puntos aislados en el espacio, sino que consiguen comerciar gracias a las economías de aglomeración que se localizan en las ciudades donde estas empresas se ubican (Behrens y Robert-Nicoud, 2015). Es este hecho diferenciador es lo que realmente denota la importancia y relevancia de las ciudades.[6]

[5]En los últimos años, han ido surgiendo proyectos de referiencia por parte de las principales instituciones internacionales. Entre éstos, encontramos el "Metropolitan Explorer" de la OCDE (`http://measuringurban.oecd.org/`) centrado en indicadores a nivel de área urbana; el "City Programme" (`http://citiesprogramme.com/`) de las Naciones Unidas que lleva el análisis urbano a temáticas de sostenibilidad medioambietal y gobernanza; el "Urban Development" desarrollado por el Banco Mundial (`http://www.worldbank.org/en/topic/urbandevelopment`) especializado en pobreza y planificación urbana; o incluso interesantes proyectos centrados en países tales como el "Metropolitan Policy Program" de Brookings (`http://www.brookings.edu/about/programs/metro`) que analiza las ciudades de EE.UU. desde diferentes dimensiones (comercio, IED o PIB); o los proyectos Dataviva del MIT Lab (`http://dataviva.info/`) y el "Atlas de Complejidad Económica" del CID Harvard (`http://complejidad.datos.gob.mx/`) que atienden a los casos de Brasil y Méjico, respectivamente.

[6]Las economías de aglomeración son cruciales para explicar los flujos comerciales. A modo de ejemplo, para el caso de Estados Unidos, alrededor del 86% de los flujos totales de comercio son desarrollados por tan sólo 100 ciudades (datos de Export Monitor 2015, Brooking). Entre otros casos diferentes a los EE.UU., encontramos que alrededor del 45,5% de las exportaciones totales del Brasil en 2013 fueron llevadas a cabo por las ciudades, mientras que esta proporción representaba el 65% en 2000 (elaboración propia a partir de Dataviva project). De hecho, no sólo internacionalmente sino también intra-nacionalmente, vemos que en casos como el español, alrededor del 79% del valor total del comercio (con origen y destino en el mismo país) es llevado a cabo por 26 municipios (propia elaboración de la EPTMC Survey del Ministerio de Fomento).

A lo largo de esta Tesis, se analizan tanto países desarrollados (España), como economías emergentes (Brasil), teniendo en cuenta, por un lado, cuestiones relacionadas con la jerarquía de ciudades o la propia estructura interna de las mismas, y por el otro, sus interacciones a través del comercio. Aunque los tres ensayos están estrechamente relacionados, cada uno de ellos se ha escrito de manera independiente para que puedan ser leídos de forma autónoma. Para introducir al lector en la temática de esta Tesis, a continuación se desarrolla un breve resumen de cada uno de estos capítulos.

1.5 Capítulo 2. Márgenes del comercio y áreas de mercado: Flujos municipales y jerarquía urbana

Mientras que la investigación reciente ha planteado múltiples modelos teóricos acerca de por qué los flujos de comercio tienden a aglomerarse en algunas ciudades específicas, el trabajo empírico en relación a esta literatura ha tendido a quedarse atrás. Haciendo uso de una base única de micro-datos sobre envíos de mercancías por carretera en España durante el período 2003-2007, conseguimos descomponer el valor total del comercio municipal en los dos márgenes del comercio, el extensivo y el intensivo. Gracias a ello, estudiamos el efecto obstaculizador que tienen los costes de transporte sobre dichos flujos comerciales atendiendo a varios niveles administrativos dentro de un país, NUTS-5 (municipal), NUTS-3 (provincial) y NUTS-2 (Comunidades Autónomas).

Con esto, confirmamos la acumulación de los flujos comerciales en distancias muy cortas. Al contrario de estudios anteriores (Hillberry and Hummels, 2008), encontramos que esta alta concentración de flujos comerciales no se explica por la existencia de límites administrativos que impiden el comercio (efecto de frontera), sino a cambios significativos en la relación entre el comercio y los costes de transporte. Para reforzar esta hipótesis, y recurriendo al Test de Chow endógeno, identificamos una serie de umbrales estadísticamente significativos para los cuáles se rompe de manera estructural la relación entre comercio y costes de transporte. En concreto, encontramos que los flujos comerciales tienden a aglomerarse en torno a las ciudades de mayor tamaño, siendo la mayor concentración a los 185€ de costes de transporte (alrededor de 130 kilómetros de distancia física).[7] Más allá de esta distancia, la intensidad del comercio se reduce hasta un valor de costes de transporte de 330€ (unos 300 km), para posteriormente el comercio volverse prácticamente inexistente. De hecho, encontramos que para distancias muy largas, los flujos comerciales dentro de un país se llevan a cabo sólo entre las ciudades más importantes.

Estos puntos de corte nos permiten dividir la muestra y controlar por el efecto que

[7] Para este primer punto de ruptura (185€), Behrens, et al., (2015) obtienen valores de la distancia similares para el caso de la concentración geográfica de la industrias manufacturera canadiense.

tienen tanto sobre el márgen extensivo como el intensivo, sucesivas fronteras administrativas. Aplicando este procedimiento a un modelo gravitatorio estándar, podemos separar correctamente el impacto que los costes de transporte (y las fronteras administrativas) tienen sobre los flujos comerciales. Gracias a ello, argumentamos que esta manera de dividir los costes de transporte resulta ser un procedimiento más apropiado para capturar el efecto no-lineal de los costes de transporte sobre el comercio, mejorando así proposiciones anteriores en la literatura (Eaton and Kortum, 2002; Abbate, et al.,2013).

Por último y más interesante, basándonos en estos umbrales de costes de transporte definimos áreas de mercado que configuran un sistema de jerarquía urbana coherente y basado en flujos comerciales que claramente evidencia las predicciones planteadas por la Teoría del Lugar Central (Christaller, 1930).

1.6 Capítulo 3. Especializacíon exportadora de las ciudades

¿Existen patrones de exportación distintos entre ciudades grandes y pequeñas? Utilizando datos de comercio muy desagregados a nivel de producto para las ciudades brasileñas en el año 2013, encontramos que las zonas urbanas más pobladas exportan bienes proporcionalmente más complejos e intensivos en cualificación que las zonas urbanas menos pobladas. A su vez, mostramos que las ciudades brasileñas que han aumentado más en población han sido las que han aumentado más que proporcionalmente las exportaciones de bienes complejos e intensivos en cualificación.

Usando técnicas paramétricas y no paramétricas, nuestros resultados empíricos apoyan modelos recientes (Davis y Dingel, 2014; Hausmann y Hidalgo, 2011) que argumentan que las grandes ciudades, por un lado, atraen a los trabajadores con mayor cualificación, y por el otro, exhiben una amplia gama de capacidades que les proporciona ventaja comparativa en la exportación de bienes complejos e intensivos en cualificación.

1.7 Capítulo 4. La aparición de ciudades especializadas y diversificadas

Muchos economistas se han ocupado de la cuestión fundamental acerca de por qué algunas ciudades se especializan y otra se diversifican. Determinar por qué surgen estos fenómenos ha resultado ser un rompecabezas teórico, que hasta el momento ha sido abordado a través del desarrollo de modelos cuyos supuestos difieren en la manera en la que se han (micro) fundamentado las fuerzas de aglomeración y dispersión, la movilidad de bienes y hogares o los propios mecanismos de formación de las ciudades.

Este capítulo busca contribuir a esta literatura realizando un análisis empírico que tiene como objetivo evaluar los mecanismos y las hipótesis fundamentales que se proponen desde la literatura teórica (Abdel-Rahman, 1990a,b, 1996; Abdel-Rahman y Fujita, 1993; Duranton y Puga, 2001; Henderson, 2000; Anas y Xiong 2003, 2005). A diferencia de la mayor parte de la literatura empírica existente que estudia la composición, el tamaño y la ubicación de las ciudades de Estados Unidos, en este capítulo nos centramos en España. En nuestro análisis consideramos la posibilidad de que las ciudades puedan ser no sólo especializadas o diversificadas, sino también una combinación de ambas (economías de co-aglomeración) o ninguna de las dos. Para ello, no sólo consideramos las características internas de las ciudades, sino que atendemos al rol que las interacciones entre ciudades pueden llegar a tener en la estructura productiva de las mismas.

Con este objetivo proponemos un enfoque empírico novedoso basado en el desarrollo de modelos probit bivariantes usando los micro-datos de tres bases únicas: una base de datos a nivel de empresa que nos permite calcular los niveles de especialización y diversificación de las ciudades; una base de datos de comercio entre ciudades españolas; y una base de datos acerca del coste económico de desplazamiento entre ciudades. Con todo y en consonancia con anteriores análisis para el caso de EE.UU. (Henderson, 2003), encontramos que una ciudad tiene una mayor probabilidad de convertirse en especializada si su mano de obra está relativamente poco cualificada, su estructura de producción está orientada hacia la fabricación de productos estandarizados, y su tamaño en términos de población es relativamente pequeño. Por su parte, el patrón obtenido es el opuesto para el caso de la aparición de ciudades diversificadas, mientras que una combinación de estos resultados ayudaría a explicar el surgimiento de las denominadas economías de co-aglomeración (Ellison and Glaeser, 1997; Helsley and Strange, 2014).

Además, encontramos que la historia importa: el grado de especialización en el pasado resulta ser un importante predictor de los patrones de especialización actuales. Por último y más interesante, encontramos que la probabilidad de que una ciudad se convierta en especializada (diversificada) depende positivamente de bajos (altos) costes de transporte en el pasado, mientras que el efecto contrario se observa para niveles de coste de transporte actuales. Estos resultados dan apoyo a propuestas teóricas previas en las que se fusionan modelos de crecimiento urbano (Anas and Xiong, 2003, 2005) y de nueva geografía económica (Fujita et al., 1999). Estos resultados sugieren que, aunque la estructura interna de las ciudades es un factor determinante de su especialización y diversificación, la localización de la ciudad juega un papel incluso más crucial en el desarrollo de las ciudades.

1.8 Bibliografía

1. Abbate, A., De Benedictis, L., Fagiolo, G., Tajoli, L. (2012). The International Trade Network in Space and Time. *Laboratory of Economics and Management (LEM)*, No. 2012/17, Sant'Anna School of Advanced Studies, Pisa, Italy.

2. Abdel-Rahman, H. M. (1990a). Agglomeration economies, types, and sizes of cities. *Journal of Urban Economics*, 27 (1), 25–45.

3. Abdel-Rahman, H. M. (1990b). Sharable inputs, product variety and city sizes. *Journal of Regional Science*, 30 (3), 359–374.

4. Abdel-Rahman, H. M. and Fujita, M., (1993). Specialization and diversification in a system of cities. *Journal of Urban Economics*, 33 (2), 189–222.

5. Abdel-Rahman, Hesham M. (1996). When do cities specialize in production? *Regional Science and Urban Economics*, 26 (1), 1–22.

6. Anas, A. and Xiong, K. (2003). Intercity trade and the industrial diversification of cities. *Journal of Urban Economics*, 54 (2), 258–276.

7. Anas, A. and Xiong, K. (2005). The formation and growth of specialized cities: efficiency without developers or malthusian traps. *Regional Science and Urban Economics*, 35 (4), 445–470.

8. Behrens, K., Bougna, T., Brown, W. M. (2015). The world is not yet flat: Transport costs matter!, *CEPR Discussion Papers*, No. 10356.

9. Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In *G. Duranton, J.V. Henderson, and W. Strange (eds.) Handbook of Regional and Urban Economics*, Volume 5. Amsterdam: Elsevier.

10. Christaller, Walter (1933). *Central places in Southern Germany.* Translated Edition.Prentice-Hall.

11. Davis, D. R. and Dingel, J. I. (2014). The Comparative Advantage of Cities. *Working Paper 20602*, National Bureau of Economic Research.

12. Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91 (5), 1454–1477.

13. Eaton, J., Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70 (5), 1741-1779.

14. Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105 (5), 889–927.

15. Fujita, M., Krugman, P. R., and Venables, A. (1999). *The spatial economy: Cities, regions, and international trade.* MIT press.

16. Glaeser, E. L. (2011). *Triumph of the city: How our greatest invention makes US richer, smarter, greener, healthier and happier.* Pan Macmillan.

17. Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16 (4), 309–342.

18. Head, K., Ries, J. (1996). Inter-city competition for foreign investment: Static and dynamic effects of China's incentive areas. *Journal of Urban Economics*, 40 (1), 38-60.

19. Helsley, R. W. and Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, 122 (5), 1064–1093.

20. Henderson, V. (2000). Political economy of city sizes and formation. *Journal of Urban Economics*, 48, 453–484.

21. Henderson, J. V. (2003). Marshallian scale economies. *Journal of Urban Economics*, 53, 1–28.

22. Hillberry, R., Hummels, D., (2003). Intranational home bias: Some explanations. *Review of Economics and Statistics*, 85 (4), 1089–1092.

23. Hillberry, R. and Hummels, D. (2008). Trade responses to geographic frictions: A decomposition using micro-data. *European Economic Review*, 52(3):527 – 550.

24. Mayer, T., Ottaviano, G. I., (2008). The happy few: The internationalisation of european firms. *Intereconomics*, 43 (3), 135–148.

25. Melitz, M. J., (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71 (6), 1695–1725.

26. United Nations (2015). World Urbanization Prospects: The 2014 Revision, *(ST/ESA/SER.A/366)*, Department of Economic and Social Affairs, Population Division .

27. Wolf, H. C., (2000). Intranational home bias in trade. *Review of economics and statistics,* 82 (4), 555–563.

28. Kilroy, A., Francis L., Mukim, M. and Negri, S. (2015). Competitive cities for jobs and growth : what, who, and how. *Competitive cities for jobs and growth.* Washington, D.C. : World Bank Group.

Chapter 2

Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy

"The distribution of population and activity across the landscape is radically uneven."
(Masahisa Fujita, Paul Krugman and Anthony J. Venables,
The Spatial Economy, 1999)

2.1 Introduction

Why and to what extent trade flows tend to agglomerate in some specific places within a country are usually topics left out of the analysis due to the absence of very detailed micro-data on interregional trade flows.[1] There are some studies in which trade flows agglomerate as a result either of intermediate inputs flows (Hillberry and Hummels, 2003; Hillberry, 2002a) or because of the location of firms close to a real border as to reduce transport costs when trading with other countries—Chen (2004); Llano-Verduras et al.(2011). Conversely, there is not a systematic analysis on why trade flows concentrate in cities within a country, specially using very detailed trade flows. In this sense, the study by Hillberry and Hummels (2008) sheds light on the appearance of an internal "home bias" at the municipal level once accurate interregional trade flows and very precise measures of internal distance are used, although they concluded that this (local) "internal home bias" is a *reductio ad absurdum* of the home bias effect observed at the

[1] Recently, there have appeared studies such as Borraz et al. (2012); Llano et al. (2010); Yi (2010); Yilmazkuday (2012) from an empirical perspective; meanwhile from a theoretical approach Behrens et al. (2013) provide a promising attempt to model how trade frictions affect goods shipped between cities.

international level.

In this paper, we extend the analysis by these last authors incorporating new econometric techniques related to trade and urban economics, explaining why trade flows concentrate in some specific places resulting in a hierarchical pattern of cities. In the process, we bear in mind that most of the trade literature focuses on international trade flows, even if the largest share of the trade activity is performed within a country, and specifically, between cities, showing how pertinent the present study is.

Going beyond the study by Hillberry and Hummels (2008), we argue that this "internal home bias effect" is an *illusion* created by the existence of transport costs thresholds which shape a series of trade or market areas driven by the biggest trading-cities within a country. Indeed, in a further step, we present empirical evidence by which this concentration of trade flows around bigger cities responds to the existence of a hierarchical urban system as it is predicted by the Central Place Theory—McCann (2001); Mulligan et al. (2012); Parr (2002).

To perform this analysis, we make use of micro-data at the highest possible level of spatial disaggregation corresponding to individual shipments at the Nuts-5 municipal level, for the period 2003-2007. This particular database allows us to determine the existence of a concentration of trade flows in small areas (in this case cities) defining *natural* trade areas in the spirit of the central place and location theory—Tabuchi and Thisse (2011), and whose geographical reach is directly related to these transport costs thresholds, which may even exceed alternative levels of administrative aggregation of spatial units; particularly, Nuts-5 and Nuts-3 territorial units. This allows us to maintain that it is precisely the existence of these transport-related breakpoints (thresholds)—not always coinciding with relevant administrative geographical limits—what really explains the appearance of an *"internal home bias effect"*.

This paper presents several empirical novelties that link and extend different literatures: 1) the study of the internal home bias effect (internal border effect) within a country: Hillberry and Hummels (2003, 2008); Hillberry (2002a) against which we argue that administrative borders do not play an effective role in halting trade as administrative impediments to trade have been removed since long within countries as a result of "single-market" agreements; 2) the trading-cities theoretical literature whose empirical contrast is still pending: Anas and Xiong (2003); Behrens (2005); Cavailhes et al. (2007), and to which we provide relevant illustrations and complementary insights; 3) the study of market areas as regional boundaries depending on the geographical reach of trade: Löffler (1998), and, finally, 4), the re-emergence of the Central Place Theory literature—McCann (2001); Mulligan et al. (2012); Parr (2002); Tabuchi and Thisse (2011).

To accomplish these goals we rely on: 1) the compilation of a very detailed panel dataset (2003-2007) on individual freight trade flows; 2) the definition and calculation of a precise and realistic economic measure of transport costs associated to these flows: the Generalized Transport Cost (GTC) performs better that its usual distance and travel time counterparts in panel data contexts; 3) the adoption of a new methodological approach within the trade literature based on structural econometric tests (endogenous Chow Test) to determine transport cost thresholds, which in turn 4) allows us to define natural market trade areas at different spatial levels of aggregation for the overall value of trade, and its extensive and intensive margins; 5) the interpretation of results in terms of the central place and location theory and their discussion by way of suitable Geographic Information Systems (Arc/GIS) illustrations; and, finally, 6) the study of all these issues using panel data econometrics that enable us to capture the dynamics of the relationship between trade flows, transport costs and internal border (home bias) effects.

To perform this study we adopt a sequential strategy. The first step requires the compilation of a novel database on road freight shipments consisting of micro-data for individual deliveries between Spanish municipalities.[2] The statistical information reported in the RFTS allows us to decompose the value of trade flows into the extensive and the intensive margins following the trade flows definition proposed by Hillberry and Hummels (2008).

The next step is based on RFTS data reporting the origin and destination of a shipment. For each bilateral freight service we calculate the actual monetary measure of transport costs, i.e., generalized transport cost, GTC. This GTC corresponds to the minimum cost of joining any origin and destination, defined as the sum of the cost related to distance (e.g., fuel, toll, tires) and time (e.g., salaries, insurance, taxes). To make these calculations we have resorted to programming techniques using Geographic Information Systems (Arc/GIS) that allow the optimization of the least cost routes through the existing road network in the years 2003-2007. In contrast to all previous studies using the standard and non-monetary transport cost proxies of distance and time, we introduce a *real euro* measure of the spatial related frictions affecting trade.

Subsequently, to study the dependence of trade flows on transport costs and the magnitude (either real or illusory) and significance of the "internal home bias effect" we rely on the pseudo-poisson maximum likelihood (PPML) estimator proposed by Santos Silva and Tenreyro (2006, 2010, 2011), as the most efficient way to control for pervasive zero value trade flows and heteroskedasticity problems within the gravity model. The results obtained using the PPML estimation including time-varying origin and destination fixed

[2]The *Road Freight Transportation Survey*, RFTS, differs in several ways from the American *Commodity Flow Survey* undertaken through a partnership between the Census Bureau and the Bureau of Transportation Statistics (DOT). Relevant for this study is that in the latter case, the surveyed statistical units are production establishments (wholesalers and retailers), while the RTFS surveys road freight companies producing the transportation service.

effects, show that "municipal boundaries" have a stronger impact on trade flows and the extensive and intensive trade margins than the results reported by Hillberry and Hummels (2008); i.e., trade inside municipalities or between contiguous municipalities is much more important than trade between non-contiguous or long distant municipalities, especially in the extensive margin. This is explained by the changing pattern (elasticity) of the effect of transport costs on trade, which is particularly intense for short distances (euros) rather than any other border effect impediment. In this sense, considering administrative boundaries not as borders haltering trade, but as likely distances where the trade flow-transport cost relationship changes, they do reflect different effects on trade. Indeed the borders between provinces (Nuts-3) have a seizable effect on all the trade decomposition variables, but lower than the municipal level (Nuts-5), while regional borders (Nuts-2) have no significant or even an inverse effect on trade flows.

These findings corroborate the idea that the impact of bigger administrative boundaries (Nuts-2) on trade flows is not as important as the trade literature has emphasized: Chen (2004); Hillberry and Hummels (2008); Requena and Llano (2010); Wolf (2000). Indeed, this result suggests that the trade-concentration effect at the municipal level is highly related to agglomeration economies (Chen, 2004; Hillberry, 2002a; Puga, 2010).

Going in a further step as to understand this localized trade pattern, we argue that these agglomeration economies only take place around large cities because of the existence of an urban hierarchy system that emanates from the Central Place Theory (Mulligan et al., 2012; Parr, 2002; Tabuchi and Thisse, 2011). Hence, the so called "home bias effect" arises at the municipal level because high-order cities act as supply centers either for theirs metropolitan areas or for the lower rank surrounding cities. A relationship not explored thoroughly in the trade and urban empirical literature.

To study this hypothesis we determine to what extend the extensive margin of trade (number of shipments) is geographically located at very short distances, while the intensive margin (average value per shipment) remain basically constant, reflecting that the effect of transport costs on each component is quite different. To achieve this goal, we introduce a new methodology in the trade literature proposed by Berthelemy and Varoudakis (1996) for endogenous economic growth models. These authors conjecture the existence of different structural models between specific *breakpoints*, and propose an endogenous Chow Test that we adopt to divide the sample in sub-samples in order to determine the existence of differences in the trade flows-transport costs relationship.

In the current framework, as trade flows are geographically localized in terms of low transport costs (short distances), we perform this test to determine if GTCs (transport costs) are conditioning this particular trade pattern. Our hypothesis is based on the idea that transport costs and GDPs in the gravity equations cannot have the same effect on trade as the geographical distance of trade associated to transport cost increases. We conjecture that this localized trade pattern in short distances is due to the extensive margin, but once trade flows achieve a relevant breakpoint (a GTC threshold), the extensive margin and the intensive margin present two diverging and even opposite patterns; i.e., the intensive margin begins to gain relevance while the extensive margin becomes flat at rather low values. That is why we cannot estimate the effect of administrative boundaries over trade flows as a *mean effect* for the whole spectrum of distances mainly because not all the administrative boundaries appear in every shipment and are not equally important,[3] i.e., neither the extensive and intensive margins always dominate at the same GTCs ranges.

Furthermore, we have performed the structural Chow Test several times in order to determine what are the transport cost thresholds that condition interregional trade flows. Once we have determined the existence of different breakpoints, we split our PPML regressions in terms of these new thresholds—in contrast, for example, to Eaton and Kortum (2002) who divide trades flows considering arbitrary distances. With these new breakpoints, we conclude that the border (home bias) effect is not unique, while it does not have an average nor a linear impact on trade as the trade literature emphasizes. Indeed, as these breakpoints are not affected by arbitrary administrative borders, we argue that it is precisely the use of these breakpoints what really matters when studying the border (home bias) effect on trade flows as they spill out over consecutive administrative boundaries.

Based on the statistical determination of these transport cost thresholds we argue that they define *natural market areas* within a country. As this methodology on structural breaks has never been considered in the trade literature, we provide the first illustration of its potential in relation to the Central Place Theory (CPT) models and its associated market areas. Different thresholds create concentric *iso*-GTCs rings representing the geographical reach of trade for a location (city), which is increasing on its size and eventually overlaps with the market area of other locations. These market areas naturally define an urban hierarchy providing an empirical justification of the central place theory.

Moreover, according to Parr (2002), O'sullivan (2003) and Mulligan et al. (2012), although central place theory predicts the existence of market areas and explains how cities supply their surrounding areas by way of trade flows, the existing empirical studies fall short from this reference, since they provide illustrations based on population agglomeration rather than actual trade flows. Related to this goal we propose a new mapping representation based on municipal product diversification, and how these calculated market areas shape the urban hierarchy proposed by the CPT by which larger municipalities (*Rank* 1 and *Rank 2* cities) determine an area of influence which is more

[3] For instance, if a shipment is delivered at a cost of 500€ (about 455 km or 5 hours at the legal speed limit of 90 km/h), we must analyse what margins predominate and what are the spatial limits in which each one of them takes place (Nuts 5, 3 or 2).

extensive than the one obtained for smaller cities (*Rank 3* or *Rank 4* cities). The empirical regularities that we find are a promising result for future research linking trade and central place theories.

Once the motivation of the study and the review of the literature have been introduced, we lay out the structure of the article. The next section discusses the database on municipal trade flows and the generalized transport costs, justifying why the GTC is the most suitable measure of transport cost in a panel data framework. Section 2.2.3 presents the decomposition of the value of trade into the extensive and intensive margins, analyzes their values based on nonparametric kernel density distributions, and shows the results obtained with different specifications of the gravity model using the pseudo-poisson distribution. Here we discuss the determinants of the value of trade and its margins in terms of our monetary measure of transport costs. Section 2.4 discusses the structural breakpoint methodology based on the endogenous Chow Test that allows to determine the transport cost thresholds, and successively replicate the PPML regressions for each range of trade values according to these thresholds. In section 2.5 we interpret and illustrate our results in terms of the central place theory, mapping a hierarchical system of cities and their overlapping regional boundaries. The last section draws the conclusions.

2.2 Trade flows and their extensive and intensive margins

2.2.1 Trade value data: The road freight transportation survey

For this study we rely on a micro-database on shipments by road within Spain during the period 2003-2007 elaborated within the research project C-intereg.[4]. This database is based on the annual Road Freight Transportation Survey, RFTS, compiled by the transport division of the Ministerio de Fomento, which randomly surveys a sample of freight companies and independent truckers, with vehicles over 3.5 tons and operating within the national territory. This database surveys almost 85% of all Spanish internal trade flows; the remaining 15% corresponding to rail, maritime and air modes. It includes information about the characteristics of the vehicle and shipments, such as the number of tons carried out by the truck,[5] the number of shipments between the origin i and

[4]This project estimates, as far as we know, the largest database on interregional trade flows estimated in Spain. It includes bilateral trade flows specifying the region (Nuts-2) and the province (Nuts-3) of origin and destination, both in tones and euros. For further information see Llano et al. (2010)

[5]This corresponds to the real load of the truck in tons. Note that the truck load may range from zero to 100%, so the database may record empty truck movements as a result of the vehicle moving to a destination where it will be eventually loaded.

the destination j, the type of product,[6] the operations performed by the truck in each shipments, as well as the actual travel distance in kilometers between the geographical origin and destination of each shipment (recorded at the Nuts-5 municipal level). As a result, for each shipment, it is not necessary to approximate the distance as done in other studies working with databases that record the origin and destination of the shipment by municipal or ZIP codes (e.g., distance between the centroids of these areas), as the true *door-to-door distance* travelled by the vehicle is reported.[7] Therefore, and thanks to this distinctive feature of the database, we can also research intra-municipal trade flows (trade within municipalities); a relevant micro level flow that is normally left out of the analysis when the database does not record the real distance of shipments.[8]

For the 2003-2007 period, the database contains more than 1.890.000 records involving, on average for all these years, 7.178 municipalities of origin from where a freight service is made and 7.913 destinations. However, most of these origins and destinations are municipalities of little relevance in terms of population and trade volumes. Therefore, so as to ease computations we have considered only municipalities that, on average for the period, had over 10.000 inhabitants. As a result, we get a sample of 633 municipalities whose trade volume by road represents a 75.5% of the total.

Since the survey does not provide information about the value of the traded goods, product prices (in euros per ton) are needed so as to obtain a magnitude of the total value moved once they are multiplied by the tons carried. These prices are not available at the municipal level (Nuts-5) in any of the official databases because of statistical confidentiality. To overcome this limitation, we rely on the alternative interregional trade flows database compiled by the C-intereg project. With this database we calculate a price vector, measured in euros per ton, for the whole period. However, these prices are calculated at the provincial level, therefore, we are forced to assume that prices at the municipal and provincial levels are equal. This assumption implies that the pricing rules determining their level at municipal level, i.e., costs and mark-ups over costs, are similar to those observed at provincial level, e.g., similar labor and intermediate costs.

2.2.2 The generalized transport cost (GTC)

Another novel aspect of this study is the use of a real (monetary valued) measure of transport cost that clearly improves those normally used as approximations, mainly geo-

[6] In the micro-database, commodities are classified attending to the Eurostat classification NST-R which differentiates between 180 products.

[7] Unfortunately, the database does not compile any information on the firms involved in the shipments. However, since we know the type of product being shipped we can approximate the production sector of the firm.

[8] We highlight that the RFTS database reports the precise *door-to-door distance* reported by the freight company for each shipment, offering a new level for intra-city or inter-city transport cost, which improves in itself the distance measures used in other studies that are based on area centroids.

graphical distance. This variable corresponds to a Generalized Transport Cost (GTC) definition corresponding to the least cost itineraries between an origin and a destination. The GTC is calculated using GIS software (Arc/GIS) with the digitalized road network, as discussed in Zofío et al. (2011). GTCs differentiate economic costs related to both distance and time. The *distance economic cost* (euros per kilometer) includes the following variables: Fuel costs (fuel price); Toll costs (unit cost per km, multiplied by the length of the road); Accommodation and allowance costs; Tire costs; and Vehicle maintenance and repairing operating costs. On the other side, the *time economic cost* (euros per hour) includes the following variables: Labor costs (gross salaries); Financial costs associated to the amortization; Insurance costs; Taxes; Financing of the truck (assuming that it remains operative only for a certain number of hours/year); and indirect costs associated to other operating expenses including administration and commercial costs.[9] In contrast with other studies that use national level operating costs for GTC measurement, the GTCs employed in this study are calculated considering prices at the provincial level; specifically those observed in the province where the shipments originates.

2.2.3 Trade flows decompositions

Thanks to our micro-dataset on trade, we decompose the value of trade flows into the extensive and the intensive margins. This procedure allows solving potential specification errors in the gravity model when trade flows are not decomposed into these two margins and when analyzing how trade barriers (frictions) affect them (Chaney, 2008; Felbermayr and Kohler, 2006; Melitz, 2003). We rely on Hillberry and Hummels (2008) and define the total value of shipments between each origin-destination ij pair by T_{ij}, which can be decomposed in the following way:

$$T_{ij} = N_{ij}\overline{PQ}_{ij}, \tag{2.1}$$

where N_{ij} represents the total number of shipments (extensive margin) and $\overline{PQ}_{ij}$ is the average value per shipment (intensive margin). At a second level, the previous expression can be further broken down so that the total number of shipments N_{ij} equals the number of commodities (k) sent within the same pair ij (N^k_{ij}), multiplied by its frequency or *trading pair* (F); that is, the average number of shipments per commodity per ij (N^F_{ij}):

$$N_{ij} = N^k_{ij}N^F_{ij}. \tag{2.2}$$

With this expression, the extensive margin is decomposed according to the *product extensive margin* (N^k_{ij}) and the *product intensive margin* (N^F_{ij}), Mayer and Ottaviano

[9]The minimum cost itinerary among the set of possible itineraries defines as follows: $GTC_{ij} = min(DistCost_{ij} + TimeCost_{ij})$.

(2008). Meanwhile, the intensive margin can be decomposed into the average price ($\overline{P}_{ij}$) and the average quantity ($\overline{Q}_{ij}$) for each pair ij:

$$\overline{PQ}_{ij} = \frac{(\sum_{s=1}^{N_{ij}} P_{ij}^s Q_{ij}^s)}{N_{ij}} = \frac{(\sum_{s=1}^{N_{ij}} P_{ij}^s Q_{ij}^s)}{(\sum_{s=1}^{N_{ij}} Q_{ij}^s)} \frac{(\sum_{s=1}^{N_{ij}} Q_{ij}^s)}{N_{ij}} = \overline{P}_{ij}\overline{Q}_{ij}, \tag{2.3}$$

where s indexes unique shipments. Additionally, to obtain regression results based only in physical units (tons) leaving aside prices, we also consider as dependent variable the shipped quantity:

$$Q_{ij} = (\sum_{s=1}^{N_{ij}} Q_{ij}^s) = N_{ij}\overline{Q}_{ij}. \tag{2.4}$$

In the database the observations are recorded for each origin-destination pair (ij), type of commodity transported and year.[10] To obtain yearly values of trade for each ij we aggregate observations such that, following Eq.(2.3), we calculate the average quantity ($\overline{Q}_{ij}$) and the average price ($\overline{P}_{ij}$) for all the shipments between each ij, thereby obtaining the average value per shipment by multiplication, ($\overline{PQ}_{ij}$). Afterwards, we multiply the average value per shipment by the total number of shipment (N_{ij}) obtaining the total value of shipments Eq.(2.1). Additionally, we calculate the maximum number of different commodities transported between each ij and multiply it by its frequency (average number of shipments per commodity), so as to obtain Eq.(2.2). For the total trade in quantities Eq.(2.4), we multiply the average quantity moved between ij by the extensive margin Eq.(2.2).

2.2.4 Analysis in growth rates

The advantage of the GTC over its distance and time proxies is that it accounts for their associated transportation operating costs—see Combes et al. (2005); Zofío et al. (2014), thereby capturing the change in the transport service market as a result of changes in input prices (e.g., particularly fuel) and regulatory conditions (e.g., related to wages in the labor market).[11] We present evidence arguing that the GTC is the most suitable measure of transport cost over the years because it is the only one capturing simultaneously the improvements in road infrastructure as well as changing regulations. As shown in Table 2.1, distance remains mainly unaffected by road improvements over years (and it could even increase as a result of business by-pass routes). Because of this

[10] Commodities are classified in ten groups, from agricultural products to manufactured goods, including products such as metallurgical, minerals, chemicals and fertilizers, and heavy machinery.

[11] Due to the lack of space, we only report results on GTCs as it is one main contribution. In Díaz-Lanchas et al., (2013) we also present the results for the distance and time measures (real distance in km and time in hours) testing the robustness of all results when the usual proxies of transport costs are confronted to a GTC measure in gravity equations.

lack of variability in the distance variable over time, even if it may constitute a good proxy for transport costs in cross section studies, it is certainly inadequate in panel data studies where transport costs are expected to vary significantly as a result of improvements in road infrastructure and changes in operating costs. Correspondingly, while the time proxy for transports costs captures the improvements in road infrastructure, it cannot account for changes in the operating costs.[12]

Table 2.1: Transport costs variation along the period.

Transport Costs	Average 2003-2007	Mean by years 2003	2005	2007	Growth Rates
GTC	333.61	347.69	343.87	305.97	-12.00%
Distance	313.51	313.02	313.39	313.49	0.15%
Travel Time	287.19	290.21	286.58	284.16	-2.08%

Table 2.1 shows the average of each transport cost for the period 2003-2007, and for the individual years 2003, 2005 and 2007, plus their growth rates. GTC is the measure with the highest variation, while distance and travel time have changed to a lesser extent. By individual years, again the GTC has reduced more than its proxies. As a result, we should expect a higher impact of the GTC on trade flows along the period; i.e., distance and travel time have lower variation during 2003-2007.

To reinforce the suitability of the a GTC measure over its proxies we perform a set of regressions based on *growth rates* of trade flows to study to what extent the variability of the total value of trade is explained by the variability of the alternative transport cost measures. Indeed, and according to the Table 2.1, we should expect a more significant and negative impact of the GTC on trade flows in comparison to its proxies.

$$\Delta T_{ijt} = \beta_0 + \beta_1 \Delta cost_{ijt} + year + \eta_{it} + \eta_{jt} + \epsilon_{ijt} \tag{2.5}$$

Table 2 shows the results of regressing the growth rate of trade flows and their margins against the growth rate of each measure, and including year fixed effects and origin and destination invariant fixed effect to control for unobserved heterogeneity and the effect of business cycles: Eq.(5).[13]

As expected, distance and travel time do not have significant effects either on the growth rates of trade flows or its margins, because their changes from one year to another

[12]See Zofío et al. (2014) for a detailed discussion on how the variation of GTCs can be consistently decomposed into infrastructure and cost components using the economic theory approach to index numbers.

[13]We have chosen to include fixed effects in this equation instead of GDP by origin and destination because of the reduced variability shown by GPDs between years, resulting in non-significant estimations of these two variables.

Table 2.2: Variation effect of transport costs on trade flows.

		Variables in Growth Rates		
	Variables	Total Value	Extensive Margin	Intensive Margin
Costs in	*GTC*	-0.0296**	-0.0355**	-0.0275*
Growth	*Distance*	-0.00181	-0.00336	0.00433
Rates	*Travel Time*	0.00281	0.000761	0.00461

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

are quite reduced. By contrast, the GTC has a negative and significant impact on the three endogenous variables, as contended. Indeed, a reduction in the GTC leads to an increase in trade flows; that is, only reductions in the shipment's real economic cost results in larger trade flows, particularly those related to truck efficiency such as fuel saving, as well as the reduction of salaries –both the main components of trucking operating costs. These results confirm the GTC as the most suitable measure of transport costs in panel data studies, while distance and time are suitable proxies of trade costs valid only for cross-section estimations of gravity equations.

2.2.5 Descriptive analysis and Kernel regressions

Figure 2.1 shows the 633 municipalities finally included in the regressions. The map shows the standard deviation of the municipal total trade value, i.e., exports plus imports, both in average values during the whole period. This sub-sample captures all trade between the largest cities (and metropolitan areas) in Spain. It also includes cities where main ports are located, as areas with high levels of trading activity.

The largest shipments are delivered from the most populated areas with the highest levels of economic activity (Madrid, the Mediterranean coast and the Basque Country), while less populated areas (south-west and north-west areas) only record trade around the largest cities. Also, trade volumes follow the corridors corresponding to major high capacity roads, indicating the strong inertia between trade and road infrastructures; i.e., firms choose locations with large accessibility defined in terms of market potential, Duranton et al. (2014).

We use a non-parametric estimation (kernel analysis) to study how each trade variable in the decomposition behaves when considering the GTCs (in euros) between the 633 municipalities. The first level of the trade decomposition is illustrated in Figures 2.2a, 2.2b, and 2.2c, presenting total value, number of shipments, and average shipment value, respectively. The total value of shipments is lower in 2003 than in 2007, falling sharply in density as transport costs increase (2.2a). This same pattern is observed in the extensive margin (2.2b), where the number of shipments drops rapidly for all years as the GTC reaches the threshold of 150 euros, while the intensive margin (2.2c) shows a trend that

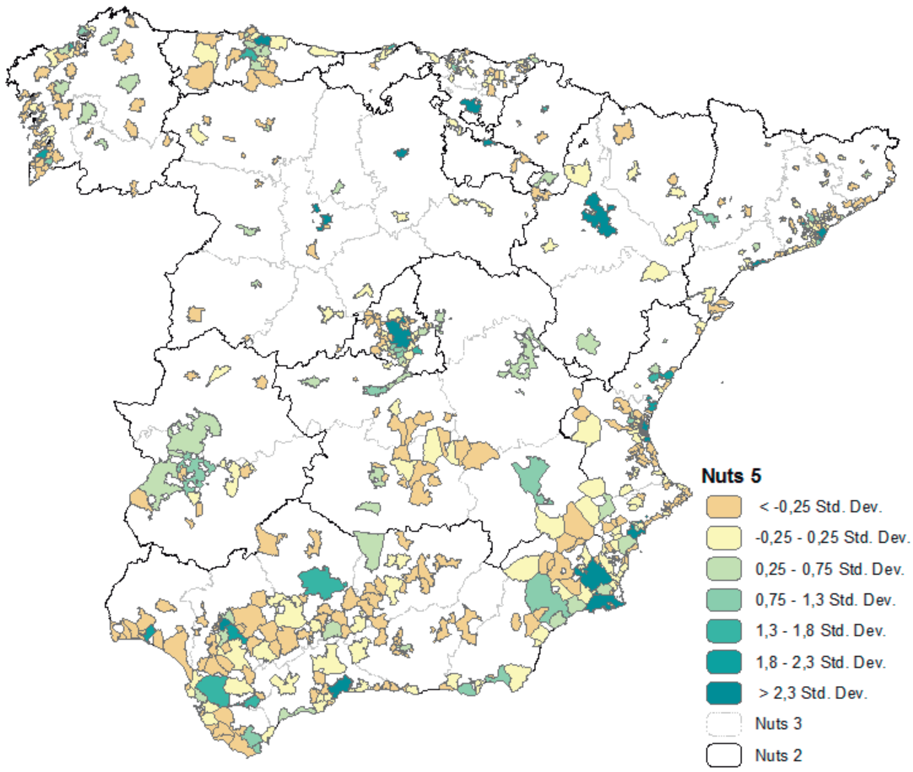
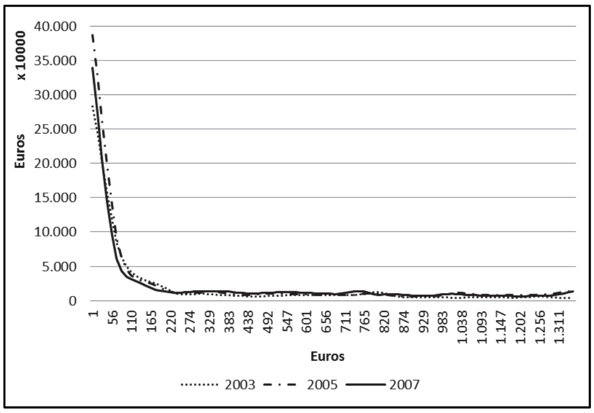


Figure 2.1: Standard deviation of the Total Trade for the 633 municipalities. Average Values (Exports+Imports). Period 2003-2007.

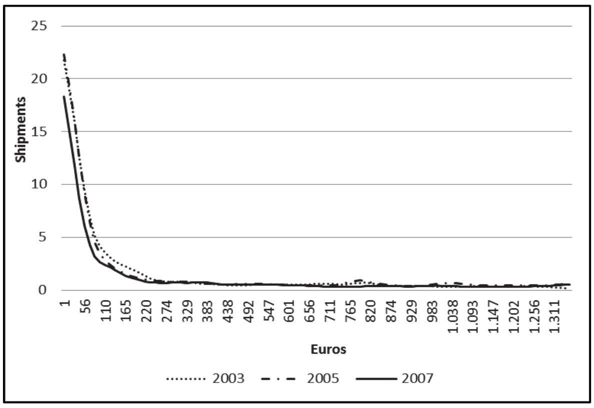
even increases in transport costs. This increasing behavior in the intensive margin is due to its composition. As we will see below, delivered prices naturally increase with GTCs, while there exists greater density of tons at very short distances, mirroring the extensive margin behavior. Subsequently, after a certain threshold, they drop sharply remaining stable at medium distances. This trend reflects the accumulation of shipments within the main Spanish metropolitan areas (Madrid, Barcelona and Valencia), while beyond these ranges the number of tons reduces. As it can be observed, this price and tons composition is creating an increasing pattern on the intensive margin.

In sections 2.4 and 2.5 we argue that this trade pattern over short and long distances is precisely the result of the existence of market areas at different transport cost thresholds that mainly respond to the existence of an urban hierarchy between trading cities.

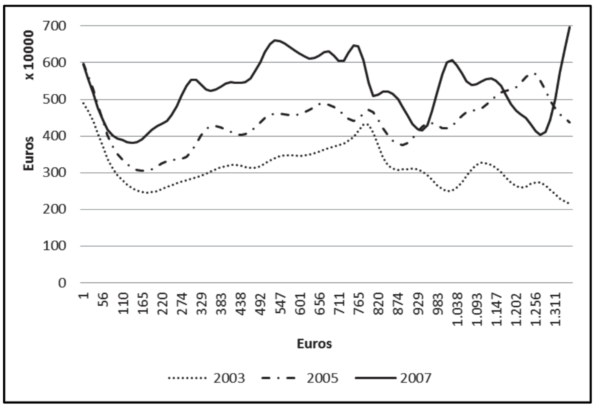
Figures 2.2 and 2.3 show the kernel estimations for the second level decomposition. Considering the extensive margin decomposition in Eq.(2.2), the number of commodities (Fig.2.3a) and its frequency (Fig.2.3b) exhibit a remarkable similar pattern and evolution; i.e., they rapidly reach their minimum values for GTC with an increase in density either at middle or high values. Decomposing the intensive margin into its average price and its average quantity as in Eq.(2.3), allows us to observe greater price variability, either between years or in one year. This variability may be due the *specific product* or the *product mix* of the shipment, which results in higher prices as a result of transport



(a) Total Value of Trade

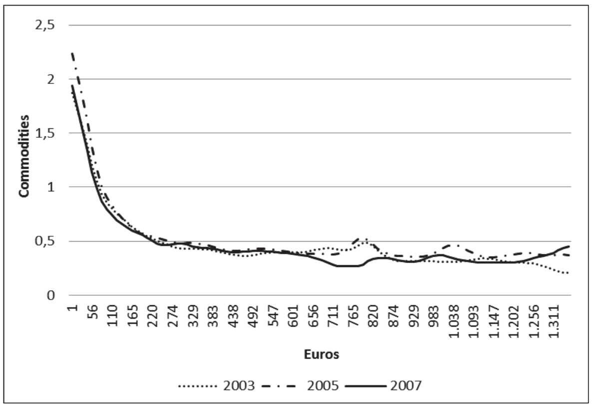


(b) Extensive Margin

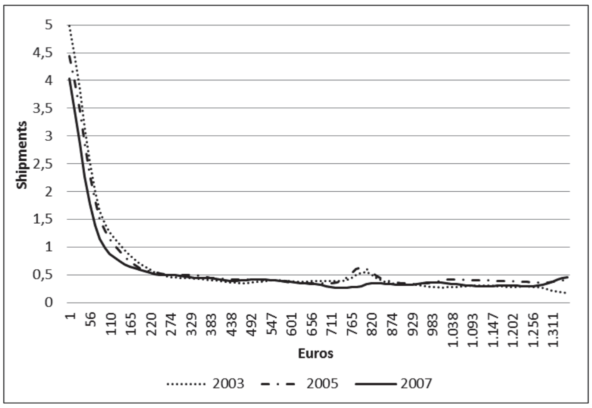


(c) Intensive Margin

Figure 2.2: Kernel regressions: Total value of trade, number of shipments (extensive margin) and average value per shipment (intensive margin) on GTC.



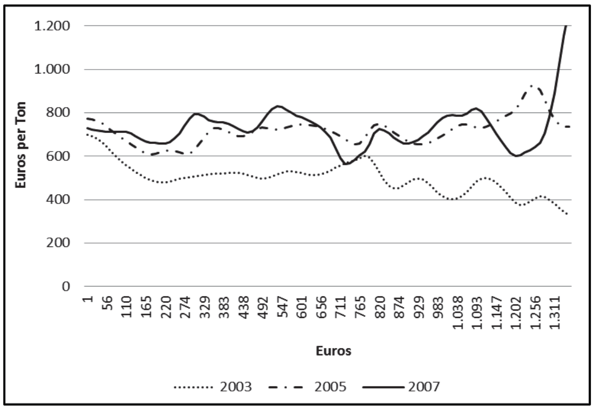
(a) Number of commodities



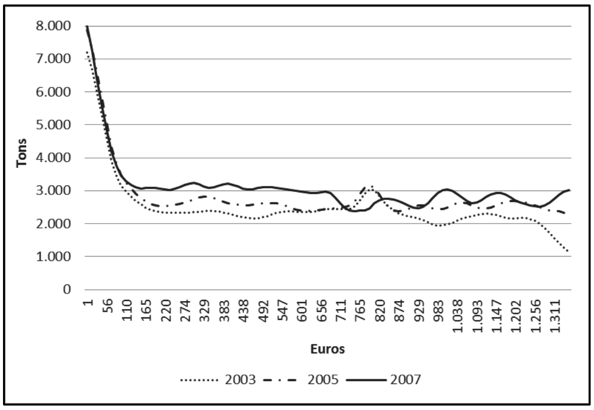
(b) Number of Shipments per Commodity

Figure 2.3: Kernel regressions: number of commodities and number of shipments per commodity (frequency) on GTC.

costs increases (Fig.2.4c); a sensible result for goods for which transport costs make up a large proportion of overall costs, that later on are passed on to delivered prices. Fo-



(a) Average Price



(b) Average Tons

Figure 2.4: Kernel regressions: average price and average tons on GTC.

cusing on average tons (Fig.2.4d), they follow the pattern previously described; i.e., they are highly accumulated on short distances until a certain CGT threshold is reached, for which they rapidly reduce and remain stable over medium and long distances. Indeed, in section 2.3 we study the existence of increasing returns in transport in the intensive (tons) margin; suggesting that for long distances it is more profitable from a logistics perspective to group shipments and send trucks with higher capacity and fully loaded, than making many individual shipments with low volume of tons, McCann (2005). Finally, the kernel regressions for total trade in physical quantities, Eq.(2.5), not presented here to save space, reflect the same trends and behavior as total trade in monetary units, i.e., the total amount falls steeply with increasing transport costs.

2.3 Econometric specification: The general gravity model

To structure the analysis in this section we present results for the standard *naïve gravity* model yielding average estimates for the whole sample without taken into account transport cost thresholds when determining their negative impact on trade flows. Specifically, we explore in detail the hypothesis about changing patterns in trade flows due to administrative boundaries. For this purpose we propose a set of regressions using equations (2.1) through (2.4) that allows us study how geographical frictions shape trade flows while taking into account different administrative boundaries, as well as to test whether these frictions may end up inducing a border effect in each of the different trade margins.

The set of PPML estimations regressing the value of trade and its components on geographical variables considering GTCs, a municipal contiguity variable and the three types of administrative boundaries: Regions (Nuts-2), provinces (Nuts-3), and municipalities (Nuts-5), are reported in Table 3.[14]

While trade data and transport cost measures have been extensively presented in section 2.2, we now discuss the remaining variables. To calculate the *Contiguity* variable, we use the GEODA software to code if the municipalities share a common border (*first-order queen contiguity*). It takes the value one if the origin and the destination of the shipment share a border, and also if the shipment takes place within the same municipality to correctly isolate the effect of the municipal boundary (Nuts-5). It permits us to capture how important are intra-municipal shipments over inter-municipal ones Hillberry (2002b), that is, coding shipments within municipalities as adjacent allows the next intradummy (Nuts-5) to isolate the additional local intensity of this type of shipments, relative to the local intensity tied to neighboring municipalities.[15]

For the administrative boundaries, we define three dummy variables as in Requena and Llano (2010). In this sense, the *Nuts_5* variable (municipal boundary) is set to one if the shipment is performed within the same municipality, and zero otherwise. Subsequently, the *Nuts_3* (provincial boundary) is one if the shipment is carried out between two municipalities that are in the same province but whose origin and destination are different.[16] Finally, the *Nuts_2* variable (the regional boundary) captures if the shipment takes place between two municipalities that are located in different provinces, but in the same *Nuts_2* region; in this case, it will take the value of one and zero otherwise. Additionally, the time dimension is reflected by a dummy variable for each year in the sample, Baldwin and Taglioni (2006). Finally, we include fixed effects by origin and by destination (Anderson and van Wincoop, 2003; Benedictis and Taglioni, 2011). As for the estimation method we rely on the pseudo-poisson maximum likelihood estimation, PPML—see Santos Silva and Tenreyro (2006, 2010, 2011), considering the endogenous variables in levels. By resorting to the PPML we can accommodate the zero trade flows problem and correct for heterokedasticity. Thus, the final specification to be estimated has the form:[17]

$$\begin{gathered} X_{ijt} = \beta_0 + \beta_1 GTC_{ijt} + \beta_2 GTC\_sq_{ijt} + \beta_3 contiguity + \\ \beta_4 Nuts\_5 + \beta_5 Nuts\_3 + \beta_6 Nuts\_2 + year + \eta_i + \eta_j + \epsilon_{ijt} \end{gathered} \tag{2.6}$$

In this specification, the GTC enters in levels and in a quadratic form to capture the

[14]In Díaz-Lanchas et al. (2013), we present robustness of our results controlling for GDP also. The municipal GDP has been obtained from the *Servicio de Estudios de La Caixa*.

[15]Coding the contiguity dummy as zero for intra-municipal shipments results in a larger value for the Nuts-5 dummy (and the remaining coefficients do not change), but the former effect cannot be decomposed.

[16]If the shipment is within the same municipality, the *Nuts_5* variable will take the value one while the *Nuts_3* and *Nuts_2* variables are assigned a value of zero. Clearly, shipments between municipalities located in different regions are assigned zero values for all administrative boundaries.

[17]In order to get a balanced panel, we have fulfilled the matrix with zeros to get all the possible pairs origin-destination (ij) obtaining 195,026 observations.

non-linearity between trade flows and transport cost at very short distances as shown by the kernel regressions. Thanks to this quadratic specification of transport costs we examine whether there are increasing returns in transportation; that is, distance has a negative effect, but marginally decreasing. In that case, we would expect a negative sign in the first term of the transport cost variable but a positive sign in the quadratic one (Combes et al., 2005). Finally, *year* corresponds to each dummy-year variable in the period and X_{ijt} stands for each one of the trade decomposition variables already mentioned.

Table 3 shows estimates for the first level of trade decomposition variables (the extensive and the intensive margin) taking into account the additional case of trade flows in quantities (tons), to compare it with trade flows in monetary units (total value of trade). Starting with the border or *home bias effect* while controlling for transport costs, the total value of trade taking place within the same municipality (*Nuts_5*) is much greater than inter-municipal trade flows, specifically it is 42 times larger.[18] In addition, the higher *Nuts_5* coefficient in the regression of the extensive margin (number of shipments) is indicating that this margin drives intra-municipal trade to a larger extent than the intensive margin (average value per shipment).

Additionally, if we consider other administrative levels, provincial boundaries (*Nuts_3*) have a much lower effect on trade (5.52 times) than the *Nuts_5* level, while regions (*Nuts_2*) loose importance as administrative boundary (1.4 times), with negligible effects (not even statistically significant) for trade in quantities. It turns out that the border effects for relatively long shipments between regions is almost nonexistent, supporting the idea that the geographical reach of trade is mainly driven by local markets; i.e., the existence of *natural trade areas* in terms of transport costs that we study in the next section. Also, by using PPML and fixed effects by origin and by destination we are correcting for potential gravity problems, thus, we are able to reduce possible biased coefficient estimations than those obtained by Hillberry and Hummels (2008) and Borraz et al. (2012), indicating a higher impact of the municipal boundary on trade flows and their margins.

For the regressions of the aggregate extensive margin decomposition, the number of shipments per commodity (frequency) and the number of different commodities shipped explain approximately the same proportion of this margin for all administrative boundaries; particularly for the municipal border (*Nuts_5*) that shows a greater agglomeration of trade flows. The *Nuts_3* and *Nuts_2* variables reflect the same pattern as for the same

[18]The PPML estimation can be sensitive to extreme observations in the dependent variable causing this high value of the *Nuts_5* level, Van der Marel (2012). However, it can correctly capture the amount of trade observed in short distances, besides additional econometric advantages, over alternative specifications such as OLS (Martin and Pham, 2008; Santos Silva and Tenreyro, 2006, 2010, 2011), as previously mentioned.

first level decomposition; i.e., provinces (*Nuts_3*) reduce its importance as trade border while regions (*Nuts_2*) have a very reduced impact on trade flows.

For the aggregate intensive margin, the coefficient associated to average tons (physical quantities) is the most relevant explaining it. Average tons shipped within the municipality are higher than inter-municipalities ones, although showing a decreasing trend between borders, and even exhibiting a negative impact at the regional level (*Nuts_2*); i.e., the average tons of inter-regional flows are higher than intra-regional ones. Also, thanks to the inclusion of total trade in physical quantities, we confirm the robustness of the coefficients when considering monetary measures as in the intensive margin; especially for the regional dummy (*Nuts_2*), which shows a non-significant sign.

Table 2.3: Fixed effects estimation with GTC.

	First Trade Decomposition Variables			Extensive Margin		Intensive Margin		
Variables	Total Value of Trade (T_{ij})	Extensive Margin (N_{ij})	Intensive Margin ($\overline{PQ}_{ij}$)	N. of Commodities (N^K_{ij})	Trading Pairs (N^F_{ij})	Price ($\overline{P}_{ij}$)	Tons ($\overline{Q}_{ij}$)	Trade in Quantity (Q_{ij})
GTC	-0.266***	-0.290***	-0.0423***	-0.154***	-0.182***	-0.00772	-0.114***	-0.447***
	(0.0322)	(0.0142)	(0.00537)	(0.00467)	(0.00566)	(0.00469)	(0.00420)	(0.0292)
GTC Square	9.06e-11***	9.99e-11***	1.27e-11***	4.17e-11***	6.47e-11***	-1.09e-11***	4.41e-11***	1.88e-10***
	(1.70e-11)	(7.40e-12)	(3.12e-12)	(2.50e-12)	(2.89e-12)	(2.77e-12)	(2.28e-12)	(1.50e-11)
Contiguity	1.365***	1.126***	0.458***	0.659***	0.827***	0.123***	0.665***	1.262***
	(0.0708)	(0.0330)	(0.0311)	(0.0141)	(0.0264)	(0.0232)	(0.0222)	(0.0518)
Nuts_5	3.764***	3.236***	0.983***	1.507***	1.933***	0.504***	1.122***	3.495***
	(0.101)	(0.0496)	(0.0594)	(0.0333)	(0.0371)	(0.0435)	(0.0388)	(0.0792)
Nuts_3	1.717***	1.406***	0.320***	0.845***	0.781***	0.432***	0.438***	1.415***
	(0.0840)	(0.0392)	(0.0255)	(0.0169)	(0.0206)	(0.0208)	(0.0178)	(0.0655)
Nuts_2	0.346***	0.338***	0.0375	0.239***	0.139***	0.188***	-0.0321*	0.0767
	(0.0777)	(0.0354)	(0.0259)	(0.0168)	(0.0207)	(0.0207)	(0.0166)	(0.0554)
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destination F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	195,026	195,026	195,026	195,026	195,026	195,026	195,026	195,026
R^2	0.708	0.879	0.087	0.076	0.146	0.512	0.344	0.624

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05,*p<0.1.

Looking at the coefficients corresponding to the GTCs, they present the expected signs in all trade decomposition variables, indicating the existence of increasing returns in transport. The *Contiguity* variable is significant in all regressions; that is, contiguous municipalities trade more among themselves than with more distant municipalities, reinforcing the market area interpretation of the distribution of trade flows. Finally, there is a good fit achieved by the model R^2 in explaining total trade flows and the extensive margin, although it is less clear for the intensive margin. Such goodness of fit is in contrast to the low values obtained by Hillberry and Hummels (2008).

With this first set of estimations we conclude that the use of very detailed measures of trade flows and transport costs reduces the impact of higher regional boundaries showing the existence of a weak internal border effect at the regional level that is only relevant when the municipal level is considered. It confirms the existence of a "illusory effect" at high regional boundaries once low-level administrative limits are controlled for. Also, it indicates a non-disruption of the market for large administrative levels, in contrast to other studies for the Spanish case: Garmendia et al. (2012); Requena and Llano (2010); or the Chinese case: Poncet (2005). These findings shed light on the idea that trade flows tend to concentrate in short distances, with their density reducing as distance increases, and drawing a trade pattern that motivates the use of structural tests to explicitly determine the point at which trade concentration finishes. Also, this stresses the importance of regionally disaggregating the trade flows and the transport costs if we want to measure the "real" border effect between areas.

In a further step, we are interested in studying the dynamics of the internal home bias effect. To achieve this goal, we estimate the same models resorting to a cross-section analysis instead of pooling the data in a whole panel database. Table 2.4 presents the results of regressing Eq.(2.6) with data corresponding to 2003 and 2007 (initial and final

Table 2.4: Cross-section regressions for 2003 and 2007 (First Level Decomposition).

GTC	Boundary	2003	2007
Total Value of Trade	Nuts_5	3.795***	2.835***
	Nuts_3	1.581***	1.079***
	Nuts_2	0.179*	-0.0313
Extensive Margin	Nuts_5	3.151***	2.555***
	Nuts_3	1.184***	0.906***
	Nuts_2	0.174*	-0.00365
Intensive Margin	Nuts_5	0.958***	0.841***
	Nuts_3	0.363***	0.291***
	Nuts_2	0.00161	0.0201

years, respectively) but considering only the first level decomposition, Eq.(2.2); that is, the total value of trade, and the extensive and the intensive margins. To summarize the output table presents only the results for the three administrative boundaries, as we are interested on the dynamics of the effect of borders on trade margins.

All administrative boundaries reflect the same pattern for the total value of trade; i.e., there exists a slowdown trend between 2003 and 2007. This finding reflects that the internal home bias effect is not constant along the years.[19] Indeed, administrative levels do not concentrate the same amount of trade within themselves throughout the whole period. In 2003, the *Nuts_5* is the one with the highest impact on trade flows, even larger than in the panel data regression. Meanwhile, it is confirmed that the *Nuts_2* level exhibits a non-significant, or even negative, effect on trade flows, indicating again that regional borders are not as important as trade literature customarily remarks.[20] The extensive margin shows the same pattern as the total value of trade, whereas the intensive margin does not change to a greater extent during the period. We conclude that the reduction in the effect of administrative boundaries on trade flows arises mainly from the extensive margin.

[19] We have performed a set of mean tests to analyze whether each administrative boundary is statistically different between 2003 and 2007. Thus, we reject the null hypotheses of equal coefficients.

[20] We stress that, in the Spanish case, the jurisdictional change associated to the regional administrative border (Nuts 2) is larger than to the provincial (Nuts 3) and municipal (Nuts 2) ones. For example, in all cases, the Spanish regions (Nuts 2) run their own budget (around 40% of the overall public expenditure), with a great autonomy in the field of economic and fiscal policy, as well as in the design of their own transport infrastructures. Moreover, in some cases, such administrative borders coincide with former kingdoms and historical territories, and the use of co-official languages. As a consequence, if the internal home bias were real and driven by exogenous barriers to trade, it should be expected to be significant at this administrative level.

2.4 Structural trade patterns: breakpoints analysis

We regard the standard regressions estimated in the previous section 2.3 as *naïve gravity equations* because they simply provide an average estimate of the effect of transport costs on trade flows, qualified by the customary second order effect. In this section we systematically undertake a series of structural tests to determine the transportation cost thresholds that define relevant discontinuities in their effect on trade flows. Based on them, we split the database to regress the same gravity specification but allowing for the different distance (GTC) thresholds. Additionally, the empirical evidence on the border effect phenomenon usually gives a "general" value for the border effect, which is corrected by different misspecifications such as Anderson and van Wincoop (2003) or Santos Silva and Tenreyro (2006, 2011), among others. Nevertheless, it does not consider the possibility of having different border effects once we control for multiple distance thresholds. In this sense, Eaton and Kortum (2002) propose different distance intervals that act as trade barriers. Nevertheless, these intervals are arbitrary as there is not an objective criterion to divide the transport cost data.

The existence of at least a relevant threshold in the transport cost-trade flows relationship for the whole sample can be easily established by simple visual inspection of the kernel regressions at an approximate value of 150 euros (Figure 2.2); that is, over very short distances trade flows are radically affected by relative low transport costs (extensive margin), while for larger values are less dependent and becoming relatively "flat" (intensive margin). But further thresholds against GTCs can exist at different values. Indeed, the previous regressions confirm the idea according to which trade flows tend to concentrate at the municipal level and are mainly predetermined by the extensive margin, which follows the same pattern on transport costs.

It is our understanding that one should control for this non-linear relationship when studying this trade concentration pattern on short distances. Thus, we argue that one should test whether trade flows change radically once they have reached a given transport cost threshold or *breakpoint*. In other words, we contend that trade flows present different structural relationships (structural stability) with transport costs, resulting in relevant changes in their negative effect on trade.

To determine these structural breakpoints we rely on Berthelemy and Varoudakis (1996), who present an endogenous Chow structural test for cross-section studies to check the stability of the parameters.[21] The test is carried out by establishing first

[21]This test divides the sample in two sub-samples. Afterwards, it begins to perform several iterations with the number of observations in each sub-sample changing continuously until it finds a breakpoint, in terms of the threshold variable(GTC), in which the Chow test rejects the null hypotheses of the non-existence of a structural change in our model (at the 5% significance level). This test presents the advantage that it is not necessary to provide an exogenous point in which the researcher suspects the possibility of a structural change, as in time-series models. By contrast, it endogenously determines the

a model that poses the existence of the structural change and the threshold variable responsible for this breakpoint (non-linear relation in our endogenous variable). In our case, Eq.(2.6) represents the *baseline* model to develop such analysis identifying the GTC as the threshold variable (but without including them in quadratic form). We have performed the Chow test several times to check the existence of successive GTC thresholds.

To summarize the information, instead of performing the test for each year in the sample, we have chosen the mean of each variable in Eq.(2.6) so as to use single points for the whole period (2003-2007). Table 5 reports the breakpoints for the total value of trade and the extensive margin.[22] Unsurprisingly, the test fails in detecting breakpoints for the intensive margin as its components do not drive reductions in trade as shown in the kernel regressions (Fig. 2.2c),[23] in sharp contrast with the total value of trade the extensive margin. Finally, all breakpoints obtained are significant at the 5%, except otherwise indicated.

Table 5 corroborates Hillberry and Hummels (2008) findings but qualifies them as we consider the full spectrum of administrative borders, whose presence may not exist at each range of trade values (e.g., intra-municipal trade at a Nuts-5 level cannot be normally observed for shipments with transport costs over the first breakpoint corresponding to 189 euros, since no municipality is so wide geographically). These results provide strong evidence that the trade flows are caused by the extensive margin, specially for short-medium GTCs as the first and the second breakpoints are mostly the same as those obtained for the total trade flows. Additionally, trade flows are highly concentrated at low transport cost values (around 189 and 233 euros). But after these two thresholds the difference between breakpoints becomes larger, indicating a declining tendency of trade flows on transport costs.[24]

To show further evidence in favor of the breakpoints obtained by way of the Chow test and stress their consistency with our trade database and the real distribution of cities, Figures 2.5a and 2.5b shows the specific GTC breakpoints obtained for the two largest Spanish cities: Barcelona and Madrid.[25] This map presents a first indication of

exact GTC breakpoint at which the endogenous variable dependence changes significantly. This test has been recently made available in STATA by Diallo (2012).

[22]The number of breakpoints obtained is six because, for GTC values higher than 706€, the test starts to fail obtaining very high GTC values with no economic sense.

[23]Nevertheless, the breakpoints obtained for the total value of trade mainly correspond to the particular "peak" points observed in the kernel regression for the intensive margin (Fig. 2.2c).

[24] In Díaz-Lanchas et al. (2013) the two distance and time proxies of transport costs present the same pattern as those of the GTCs, as they are highly correlated in cross-section estimations (0.91 and 0.95, respectively). Moreover, from the transport cost database it is possible to relate travel times with distance by considering the legal speed limit: 90 km/hour, while the relation between GTCs and distance is about 1.1€/km. With these equivalences we could approximate the relation between transport costs, obtaining similar distance and time values to those shown by the breakpoints.

[25]The GTC presented are in terms of distance equivalency as to correctly measure the market area

natural trade areas in terms of the transport reach along the existing road network. The Arc/GIS Network Toolbox allows us to calculate the exact coordinates corresponding to the maximum distance given the type of road (arc) and its specific attributes as capacity, gradient, congestion level, etc. It can be observed that Barcelona and Madrid are linked to a further extent with those cities with a high volume of surrounding highways; i.e., each breakpoint represents an *isocost line* in terms of GTCs that is longer, the larger the high capacity road network. We term these market areas as natural trade areas, because they have been obtained by way of the objective procedure represented by the Chow's structural test.

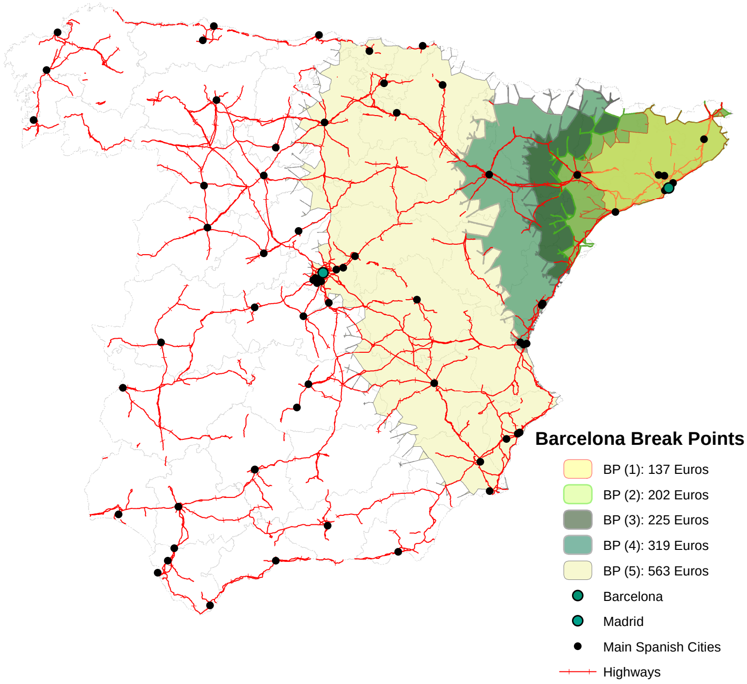
Observing these two maps multiple aspects can be remarked: i) the first breakpoint refers to the supply center between the city and its metropolitan area and other small cities not far from it; ii) the second point reaches some important cities (provincial capitals), especially from Madrid; iii) the third point presents the same pattern as the previous one, although it links to higher level and richer cities such as Valladolid, Burgos and Salamanca for the case of Madrid; iv) the fourth point appears as very relevant as it joins Madrid and Barcelona with the richest cities in Spain in terms of trade, these are mainly Valencia and Zaragoza; v) the last point directly links Madrid and Barcelona, indicating that trade flows overlap for the two largest Spanish cities.

An additional conclusion is that the road network centrality is highly important as Madrid's trade area is always larger than Barcelona's for all breakpoints: within the boundaries of Spain, Madrid almost reaches all regions within the last breakpoint, while Barcelona leaves half of the Iberian Peninsula out of its direct reach. It is important to note, at this point, that the map is only showing the points at which trade flows change structurally as given by the Chow Test; i.e., the geographical reach of each city given the GTC values obtained, and therefore does no reflect trade volumes. Indeed, Madrid and Barcelona present different trade patterns as we will show later. In this case, the map shows that the breakpoints obtained have an empirical correspondence with the actual geographical distribution of Spanish cities.

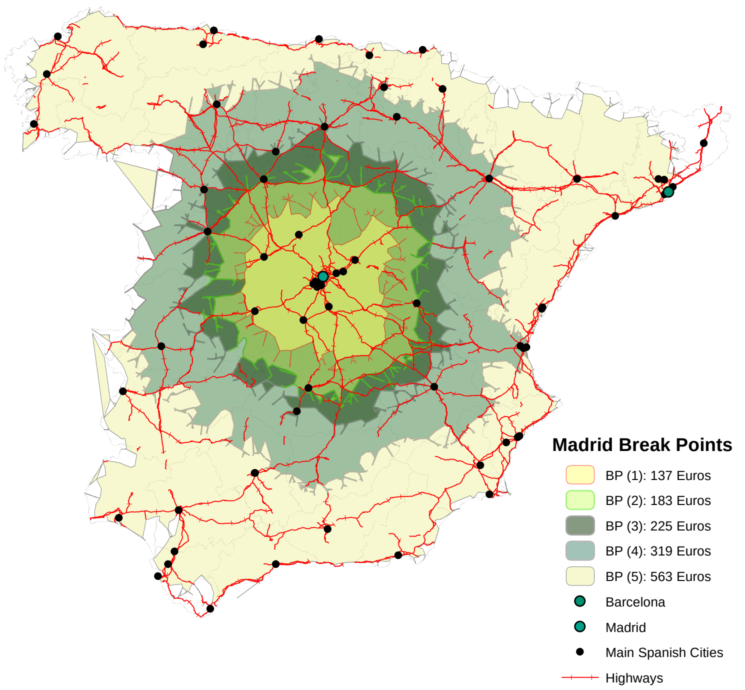
With these findings we argue that the methodology based on the Chow structural test represents a promising procedure to determine trade areas because, as far as we know from the literature, there have not been direct empirical measurements of cities' market areas, Löffler (1998). Specifically, these empirical regularities are in line with the predictions of the Lösch and Christaller's model, whose analytical framework constitutes the core of the central place theory (CPT), and gives rise to the so-called urban hierarchical systems.[26] Although these ideas are undertaken in the next section, we now anticipate

on the Arc/GIS software.

[26] CPT establishes that cities have market areas that are decreasing on transport costs, and where the largest cities producing diversified goods under increasing returns to scale can reach the furthest locations, meanwhile smaller cities have a reduced influence because they provide more standardized



(a) *Barcelona*



(b) *Madrid*

Figure 2.5: Natural trade areas using GTC Breakpoints for the whole period 2003-2007.

that, as predicted by these two models, these market areas represent geographical locations where cities compete with each other, with market areas reaching longer radii the higher the city-ranks.

Once we have determined the exact breakpoints where the effect of each transport cost on trade flows change, we split our database according to these thresholds. We perform a set of regressions using Eq.(2.6) but, as anticipated, we have eliminated the administrative boundary for the interval in which, on average, no observation is recorded. As a clear example, for a transport cost interval of 285-513 euros (breakpoints *3* and *4*) there are no shipments either at the intra-municipal level or between contiguous municipalities; and therefore these variables must be dropped in the regressions for this interval. We follow the same rationale for the rest of transport costs intervals. In fact, we contend that this is the correct methodology for measuring the effect of transport costs and administrative boundaries on trade flows when accounting for the precise transport cost-trade value (non-linear) relationship; otherwise we would obtain an overall home bias effect which is not *real* in the sense that it would not attend to the specific characteristic of each shipment. To demonstrate this idea, we include in the same table a "general specification" (a naïve gravity equation) using again equation Eq.(2.6) but without controlling for distance thresholds, and resorting to the quadratic GTC as way to control for the non-linear effect of transport costs on trade flows. As we show in what follows, administrative boundaries overestimate the effect of the border effect in short distances, meanwhile it has a different effect when we split the distance by thresholds.

goods normally characterized by constant returns to scale.

Table 2.6: Total Value of Trade by GTC thresholds.

	GTC Thresholds					*General*
Variables	(0€-189€]	(189€-285€]	(285€-513€]	(513€-706€]	(More than 706€)	*Specification*
GTC	-0.0145***	0.00262**	-0.00243***	-0.00329***	-0.00142***	-0.00525***
	(0.000339)	(0.00127)	(0.000185)	(0.000309)	(0.000125)	(0.000339)
GTC Square	—	—	—	—	—	3.69e-06***
						(2.47e-07)
Contiguity	0.911***	—	—	—	—	1.232***
	(0.0335)					(0.0702)
Nuts_5	3.022***	—	—	—	—	3.269***
	(0.0421)					(0.0836)
Nuts_3	1.059***	2.233***	—	—	—	1.247***
	(0.0348)	(0.318)				(0.0645)
Nuts_2	0.427***	0.0184	0.189***	1.347***	—	0.0945
	(0.0374)	(0.117)	(0.0708)	(0.102)		(0.0692)
F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,532	4,493	7,914	4,720	4,386	39,045
R^2	0.957	0.739	0.675	0.825	0.638	0.939

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

Table 2.7: Extensive Margin (Number of Shipments) by GTC thresholds.

	GTC Thresholds						*General*
Variables	(0€-185€]	(185€-246€]	(246€-321€]	(321€-582€]	(582€-655€]	(More than 655€)	*Specification*
GTC	-0.0140***	-0.00357***	-0.00528***	-0.00175***	-0.00231***	-0.00133***	-0.00725***
	(0.000199)	(0.000572)	(0.000558)	(0.000143)	(0.000858)	(7.74e-05)	(0.000233)
GTC Square	—	—	—	—	—	—	4.48e-06***
							(1.64e-07)
Contiguity	0.767***	—	—	—	—	—	1.037***
	(0.0171)						(0.0359)
Nuts_5	2.556***	—	—	—	—	—	2.870***
	(0.0232)						(0.0518)
Nuts_3	0.808***	1.276***	1.553***	—	—	—	1.060***
	(0.0198)	(0.0789)	(0.101)				(0.0436)
Nuts_2	0.293***	-0.0164	0.161***	0.553***	—	—	0.118***
	(0.0231)	(0.0363)	(0.0487)	(0.0578)			(0.0423)
F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,316	3,172	2,860	8,607	1,834	5,256	39,045
R^2	0.959	0.880	0.809	0.613	0.896	0.736	0.946

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

Table 2.8: Intensive Margin (Average Value per Shipment) by GTC thresholds.

	GTC Thresholds					*General*
Variables	(0€-189€]	(189€-285€]	(285€-513€]	(513€-706€]	(More than 706€)	*Specification*
GTC	-0.00422***	8.78e-05	-0.000507***	-0.000861***	-0.000811***	-0.000485***
	(0.000149)	(0.000372)	(9.97e-05)	(0.000194)	(6.84e-05)	(0.000108)
GTC Square	—	—	—	—	—	2.15e-07**
						(8.41e-08)
Contiguity	0.286***	—	—	—	—	0.456***
	(0.0153)					(0.0315)
Nuts_5	0.847***	—	—	—	—	0.978***
	(0.0254)					(0.0607)
Nuts_3	0.257***	0.608***	—	—	—	0.314***
	(0.0161)	(0.0759)				(0.0293)
Nuts_2	0.138***	0.150***	0.0466	0.112*	—	0.0384
	(0.0165)	(0.0349)	(0.0467)	(0.0670)		(0.0271)
F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,532	4,493	7,914	4,720	4,386	39,045
R^2	0.406	0.575	0.387	0.445	0.444	0.288

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

Tables 2.6, 2.7 and 2.8 show the results for the fixed effects estimation by GTCs thresholds for the total value of trade (table 2.6), the extensive margin (table 2.7) and the intensive margin (table 2.8). The GTC is highly penalizing in short distances, while it reduces its negative impact on trade flows as distance increases. It confirms that transport costs do not have the same negative linear effect on trade flows along the spectrum of distances, but it also indicates that transport costs are not as detrimental to trade as normally thought. Administrative boundaries are very penalizing on the first threshold, especially the municipal border, as in the naïve regression for the whole sample in Table 2.3. The fact of having a high impact of the *Nuts_3* and *Nuts_2* levels in this GTC-interval, is reflecting those municipalities that are geo near each other graphically, although in different provinces or regions; but at the same time they are concentrating trade within their own boundaries (the *Nuts_5* mainly). *Nuts_2* increases in value and significance as GTCs increase; meanwhile it is non-significant for intermediate distances as the *Nuts_3* level captures the border effect. We have performed several analyses to confirm this idea. Finally, if we compare the analysis with the general specification, all coefficients are distorted or even overestimated if we do not control for the GTC thresholds. Indeed, the effect of transport costs in quadratic terms results underestimated for shorter distances whereas it is overestimated for intermediate and longer distances.

The extensive margin (Table 2.7) and the intensive margin (Table 8) present the same decreasing pattern on distance as the total value of trade. *Nuts_3* presents an increasing impact on trade flows, while *Nuts_2* has the same evolution but for the third and the fourth thresholds; i.e., higher order boundaries present higher impacts on trade as long as the GTC increases. It could be indicating that the impediments to trade traditionally obtained in the literature is not correctly capturing this effect according to which, higher regional borders present higher impacts on trade when the distance is large. Thanks to the use of the Chow's structural test, we obtain a specific geographical threshold where the main concentration of trade takes place and changes structurally in term of transport costs (GTC), qualifying the results obtained by Hillberry and Hummels (2008). That is why we conclude that the border effect calculated in the trade literature is biased in the sense that it does not control for the specific (and non-linear) relationship between trade flows and transport costs. Also, this high density of trade areas within short transport cost values corroborate the existence of trade areas defined by these value thresholds, whose existence is explained in the literature by agglomeration economies, either external or internal to the firm.

2.5 Trade areas and the hierarchy of Spanish cities

The breakpoints previously calculated and the accumulation of the trade flows in short distances convey relevant empirical findings in terms of the central place theory and its associated hierarchical urban system (Hsu, 2012; McCann, 2001; Mulligan et al., 2012; Parr, 2002; Tabuchi and Thisse, 2011). Following this literature based on the Lösch and Christaller's model, we would expect areas of influence whose geographical reach is driven by transport costs; that is, consumers and firms will locate in places where they can be supplied by different cities (municipalities in our case) and taking into account the transport costs in which they incur because of their consumption or production processes. In the model, cities cover all locations as long as consumers and firms are willing to cover the transports costs of having the goods shipped to their particular location. As a result of this demand schedule, cities' market areas show a decreasing trend in shipments as transport costs increase, eventually coming into spatial competition with each other for the geographical space where their areas of influence overlap.[27]

This competition between locations as predicted by the CPT can be shown resorting to our data—Figures 2.8 and 2.9 in Appendix 2.A. Both figures show the kernel distributions of trade for the two largest Spanish municipalities (Nuts-5) and provinces (Nuts-3) corresponding to Barcelona and Madrid. It is clear that both market areas overlap in space, although they present interesting differences in their trade patterns depending on the geographical location (coastal and land-lock, respectively), but also on the level of data aggregation. Looking at municipal data in Figure 2.8, while Barcelona spreads along the geography, i.e., it presents a larger density for medium range GTC values, Madrid shows a higher density of trade on short distances, which is related to its larger

[27] The reduction in the level of trade and the number of shipments (extensive margin) because of the geographical increment in the market area, has been clearly shown by the GTC breakpoints and the corresponding regressions in the Tables 2.6 and 2.7.

metropolitan size concentrating more economic activity than Barcelona, as it distributes larger quantities within itself to meet intermediate production processes (vertical linkages) and final demand (population). This is due to the fact that Madrid incorporated many of its surrounding (industrial) municipalities in 1948; an annexation process that did not take place in Barcelona, and that conditions the density of the trade flows within both municipalities (along with the sea transport mode in Barcelona as a coastal city supplying neighborhood locations).[28]

In a stylized and simple version of the CPT model, these market areas are spatially represented by nested hexagons within a geographical lattice, where the city is in the middle of each hexagon and its radii of influence are increasing in city's size. It means that, when the radii reach a specific point (breakpoint in our previous analyses), the market area changes its pattern radically until it eventually disappears for the case of smaller cities. Specifically and following the theoretical foundations presented by Behrens (2005); Cavailhes et al. (2007); Tabuchi and Thisse (2011), an urban hierarchy à la Christaller emerges because, considering a range of different goods, few cities can serve high-order items (specialized goods) which results in larger GTC thresholds. Meanwhile, as the order-scale of items reduces (low-order items), smaller cities provide more standardized products. In this sense, *"the most central location in the entire system provides all of its goods and services (...). But, moving down through this functional continuum (of goods), other locations on the landscape are sufficiently well located to provide some, but not all, goods that are provided at the most central location"* (voir Mulligan et al., 2012, pp. 404). So, according to Christaller's framework we should expect that, within a country, a hierarchical system of cities emerges where few cities (*Rank 1 cities* or *Dominant cities*) present the largest market areas supplying the full range of products; a second group of big cities (*Rank 2 cities*) serves a huge variety of commodities with a large market area too; while in the next levels other medium size and small cities (*Rank 3 cities* and *Rank 4 cities*) are scattered geographically between these two previous city-ranks, presenting more standardized products with a lower or even insignificant market areas only supplying the most homogeneous goods needed for consumption.

Although the theoretical predictions about the existence of an urban hierarchy system is related to the magnitude of trade flows and its associated product mix, empirical studies have only focused on a proxy corresponding to cities populations, showing that

[28]Although it is beyond the scope of this paper, the relative position of the cities within a given rank level (e.g., Madrid and Barcelona as dominant first level cities) is prone to the modifiable areal unit problem. Nevertheless, to show the robustness of the results as regards to a fairly broad level of aggregation, Figure 2.9 portraits the kernel density functions aggregating trade flows at the provincial Nuts-3 classification (two levels of aggregation above Nuts-5 municipalities). With a similar extension around 8,000 km^2 results show that the Barcelona province now exhibits higher trade densities at higher GTC values, but the general classification of these two locations within the dominant category does not change.

the larger the population, the higher will be its rank within the country. The underlying assumption is that population size is not only a good proxy of city's size, but also implies a more diversified demand that supports the production of a wider range of products. Despite this attempt to relate the city's size with its market area, we consider that there is lack of empirical evidence based on relevant trade flows. That is why, so as to study the concentration of trade flows in shorter distances, and to complement and reinforce our breakpoint thresholds and home bias (border) effect analyses, we rely on our empirical data to provide evidence that supports the existence of an urban system based on the volume and sectoral characteristics of the distribution of trade flows, i.e., the number of commodities and trading partners that each city shows, and its relation with the city's market area.

Table 2.9: Shipments distribution by municipalities and products.

2003	*Number of Municipalities*					
Number of commodities	1	(1-5]	(5-10]	(10-50]	(50-100]	More than 100
1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
(1-5]	0.00%	0.16%	1.12%	3.19%	0.00%	0.00%
(5-10]	0.00%	0.32%	1.44%	5.42%	0.16%	0.00%
(10-50]	0.00%	0.00%	0.32%	38.6%	11.0%	0.16%
(50-100]	0.00%	0.00%	0.00%	1.91%	17.3%	2.71%
More than 100	0.00%	0.00%	0.00%	0.00%	1.91%	14.19%
2007	*Number of Municipalities*					
Number of commodities	1	(1-5]	(5-10]	(10-50]	(50-100]	More than 100
1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
(1-5]	0.00%	0.16%	1.12%	1.60%	0.00%	0.00%
(5-10]	0.00%	0.00%	0.64%	3.83%	0.00%	0.00%
(10-50]	0.00%	0.00%	1.12%	42.1%	8.31%	0.00%
(50-100]	0.00%	0.00%	0.00%	1.76%	19.1%	1.60%
More than 100	0.00%	0.00%	0.00%	0.00%	3.04%	15.5%

Source: Own elaboration from the Road Freight Transportation Survey data.

Table 2.9 shows these distributions in 2003 and 2007, differentiating by intervals the number of commodities shipped and the numbers of cities with which shipments take place. Data are given as percentage over the total number of municipalities, indicating the amount of municipalities that trade a determined number of commodities and the number of different municipalities with which they trade.[29] It can be observed that the largest number of municipalities trade between 10 to 50 commodities with a set of 10 to 50 municipalities interval, although this percentage has shifted in 2007 as the number of commodities shipped to the same number of municipalities increases, showing

[29]The intervals considered are in line with Mayer and Ottaviano (2008) for the case of exporting firms, but adding some intervals to better disentangle the regional and sectorial shipments distributions.

a diversification in the shipments' product composition, as well as a wider array of destinations. In fact the percentage of shipments in the upper intervals, greater than 10 commodities and 10 municipalities, increase from 87.86% in 2003 to 91.55% in 2007. Finally, there has been an increase in the number of municipalities that trade more than 100 commodities to more than 100 counterparts; from 14.19% to 15.50%, respectively.

Table 2.9 sheds light on the urban hierarchy system. The main diagonal characterizes different types of cities; that is, a huge proportion of cities represents medium-small cities (Rank 4 and 3 cities in CPT terminology), other group of cities trade more commodities with more cities (Rank 2 cities); meanwhile only a few proportion of cities trade a huge number of varieties with a huge number of other cities (Rank 1 cities). That is, large cities supply more goods to more different destinations than small cities, Tabuchi and Thisse (2011). Specifically, we have clustered the municipalities according to these two variables, the number of different commodities that they ship and the number of trading partners, in order to split the sample attending to an objective criteria and using the two variables predicted by the CPT as indicators for varying market areas. We have used cluster techniques obtaining four different groups corresponding to the Christaller's idea about product diversity and geographical reach/rank. Concretely, we have performed the analysis using the raw data from the RFTS. Table 2.10 shows the city-type clusters obtained after applying the cluster analysis (K-means based on centroid distances).[30] This city-type distribution remarkably matches the real economic distribution of Spanish cities.

Table 2.10: Number of Cities by Cluster-Type.

City Rank	Number of Cities
Rank 1	4
Rank 2	21
Rank 3	101
Rank 4	507

According to the theory, each of these city-groups should present different market areas that shrink with the rank of the city. Figure 2.6 depicts the specific relationship by showing different kernel regressions for each cluster of cities.[31] As expected, cities with a higher rank present higher densities for all distances, thereby enveloping the distributions of lower rank cities–both at any trade cost value and threshold. Moreover, the elasticity of trade flows to GTCs is lower for all transport costs; i.e., over short distances *Rank*

[30]Based on the dendogram, we decided to form four groups of cities for, in a subsequent step, using the K-means analysis to obtained the four groups after 25 iterations

[31]As in the first kernel regressions, we regress the total value of trade against the GTCs in euros but accounting for each type of city.

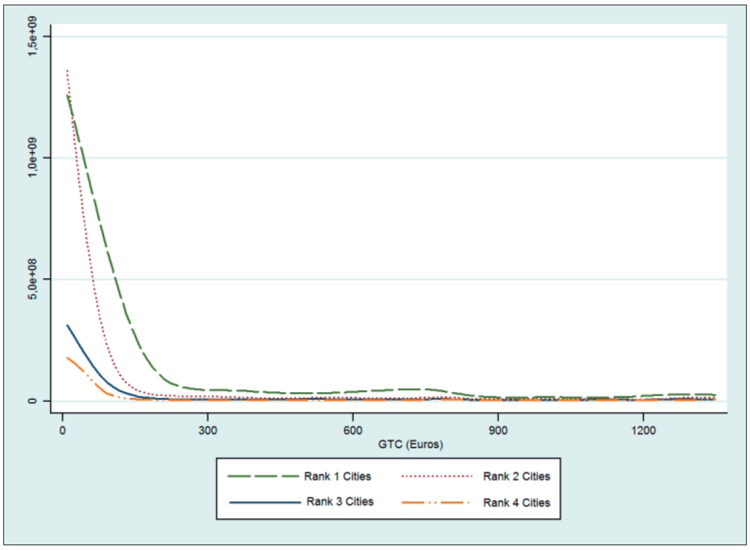
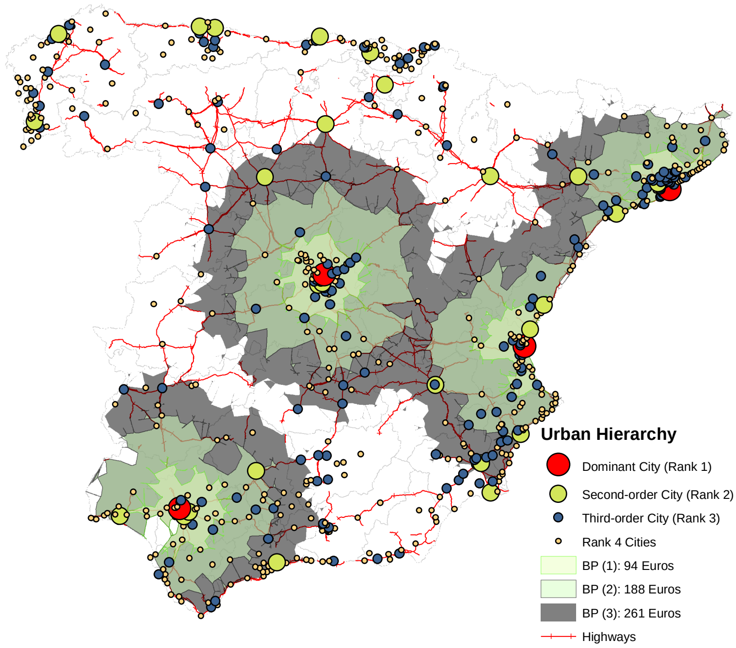


Figure 2.6: Kernel regressions for different city-cluster.

3 and *Rank 4* cities are not as sensitive to GTCs as higher rank cities. Indeed, the concentration of trade flows in short distances observed in Figs. 2.2a-c is clearly driven by these two city-clusters.

We map this information to show a detail representation of the Spanish urban hierarchical system. As far as we know Figure 2.7 is the first illustration of an actual hierarchical system based on trade flows. It illustrates cities in relation to the four city-type already calculated. Additionally, we map the breakpoint thresholds for the dominant (*Rank 1*) cities after applying again the Chow test only for this group of four cities (Madrid, Barcelona, Valencia and Seville). It can be observed that the first threshold covers the metropolitan area of these big cities, while the second and the third breakpoints reach *Rank 2* cities or even intermediate (*Rank 3*) cities as predicted by the central place theory. Indeed, it seems that the map is a remarkable representation of the geometry proposed by Christaller (Parr, 2002); specially, if we consider that his theoretical representation was simply based on a homogeneous space with constant distance-decay transport costs function. Moreover, it is clear that the geographical reach of these breakpoints exceeds all administrative levels of the spatial units; particularly, Nuts-5 (municipal) and Nuts-3 (provincial) territorial units, i.e, again trading activity spreads over administrative levels. Also, there exists an additional breakpoint at the value of 665€, although we have not shown it to summarize the map information. This last breakpoint represents again the distance between the main *Rank 1* cities, particularly Madrid and Barcelona, which emphasizes the results obtained in the previous section.[32] This pattern explains the high concentration (home bias effect) found at the

[32]Additionally, calculating the breakpoints for Rank 2 cities we obtain a third point that exactly matches the same GTC value (261€), that is obtained for *Rank 1* or *Dominant cities*. This finding emphasizes the idea of competing market areas—Cavailhes et al. (2007); Tabuchi and Thisse (2011).



Note: The fourth breakpoint is equal to 665€.

Figure 2.7: Urban Hierarchy and Natural Trade Areas for Dominant Cities. First, Second and Third GTCs Breakpoints.

municipal level in Table 2.3, indicating that it is precisely the trade flows between first rank (*Dominant*) cities what shapes the structural trade distribution of the data, thereby resulting in high market areas, i.e., large breakpoints. In fact, and taking into account that each successive breakpoint entails lower levels of trade flows, this result points out that trade flows within a country tend to concentrate because of the existence of market areas that surround the largest cities. More concretely, these market areas are creating an urban hierarchical system where only few cities perform the main part of the trade flows.

2.6 Conclusions

This study aims to develop a research strategy that validates an improved methodology to assess and explain intra an inter-municipal trade flows based on real transport costs rather than on its distance and time proxies. It establishes the non-linear relationship between trade flows and transport cost in terms of suitable thresholds that are identified by way of Chow's test for structural breaks. These breakpoints are later used to: i) qualify the existing results that place an (overstated) explanatory premium on administrative (border) effects within countries; i.e., the administrative border effect hampering trade is unlikely to hold, and ii) define trade areas that convincingly portrait an urban hierarchy system based on the geographical reach and diversity of trade.

Specifically, making use of two novel micro databases in the literature on international and interregional trade flows we analyze the agglomerating patterns of trade flows around specific areas, i.e., cities. The first database compiles information about freight shipments transported by trucks between Spanish municipalities for the period 2003-2007. The second one involves the calculation of an alternative and very precise monetary measure of transport cost between cities (the generalized transport cost, GTC). We show that this GTC measure is the most suitable measure of transport cost when explaining trade flows within a panel data structure–in contrast to distance and travel time normally used in the literature. This is because GTC is the only measure capturing the real dynamics and effects on trade flows of changes in operating costs over the years.

Thanks to the detailed information on shipments, we decompose aggregate trade flows into their extensive and intensive margin, so as to determine, in a first step, the effects of the geographical frictions corresponding to the three internal territorial boundaries existing in Spain (border effects) on each one of them, while controlling for the precise generalize transport cost measure. In a first naïve analysis, we conclude that the effects of the internal administrative levels on trade flows are higher than the ones reported by Hillberry and Hummels (2008) for low transport cost values. However, allowing for a precise trade flows-transport cost relationship, we show that regional borders have a much reduced, or even negative, impact on the trade flows taking place within them; while intra-municipal trade (or surrounding areas) tends to concentrate the largest share of trade flows. In contrast to Hillberry and Hummels (2008), which suggest that this city's border effect is a *reductio ad absurdum* on the "home bias" literature, we argue that this higher density has nothing to do with border impediments, but it arises as a result of transport cost thresholds that define the geographical scope of agglomeration economies, which mass shipments around *Rank 1* or *dominant cities* resulting in large market areas.

To show this idea we introduce the endogenous Chow Test into the trade literature, allowing us to determine the transport costs thresholds at which the trade flows–transport cost relationship changes structurally. Thanks to this methodology, we run equivalent regressions for each range of shipments within subsequent breakpoints, so as to correctly measure the impact of administrative boundaries on trade flows within a country defining a series of cities' market areas. In this respect we confirm the high density of shipments at GTC values below 189€ (170km or 150 minutes), presenting strong evidence on how agglomeration economies shape market areas and an urban hierarchy based on trade flows. In this respect we present empirical evidence by which few large cities dominate geographically and account for the largest trade share, while small cities spread along the geography without significant influence areas in terms of market size. We argue that these results are in line with the empirical predictions emanating from the Central Place Theory, producing the first map of an urban hierarchy system based on actual trade flows. Moreover, we provide evidence corroborating the hypothesis that associates the ranking of the cities not only to trade volumes, but most importantly to the diversity of the goods being shipped (production mix) and their multiple destination geographical reach.

All these results call for future studies based on how this very sharp picture of trade flows and market areas, emerges in an urban hierarchy. In this sense, it is necessary to expand the analysis focusing on the intensive margin (average value of each shipment) at the sectorial level, so as to understand in what sense it leads to the specialization of the economic structure of Spanish cities. Finally, focusing on sectorial analyses, it is worth explaining the existence of large trade flows between far away (*rank 1* or *dominant*) cities, which in turn implies the study of the value transport cost relative to the origin (mill) and destination process subject to high transport costs; i.e., a challenging analysis in terms of relative prices, *Alchian-Allen* effects and related demand analysis. Particularly, establishing whether goods that are shipped far away are highly differentiated (heterogeneous) as opposed to goods that are traded over very short distances and would present close substitutes (homogenous goods).

Acknowledgments

I acknowledge support and advise from Jose Luís Zofío (UAM) and Carlos Llano Verduras (UAM) to develop this chapter. It started when I was as a visiting researcher at the Kiel Institute for World Economy and Kiel University (Kiel, 2011), under the supervision of professor PhD. Horst Raff to whom I express my honest gratitude for helping me during this stay. By that time, I was working on the Master Thesis as the seminal study for this chapter. This Master Thesis was awarded in 2012 with the National Prize for Young Researchers ("IX Certámen Arquímedes para Jóvenes Investigadores") from the Ministerio de Educación.

I also thank comments Thierry Mayer, Johannes Bröcker at the Kiel Institute for World Economy, Geoffrey Hewings at R.E.A.L. (University of Illinois at Urbana-Champaign) and John Parr at the NARSC-2013 Conference, for their useful comments. I also express my gratitude to Francisco Requena-Silvente (Universitàt de Valencia and Sheffield University), Jesús Fernández- Huertas Moraga (FEDEA) and Alvaro Martín Remesal (CEMFI) for suggesting better improvements on this chapter.

Finally, I acknowledge support from research projects such as ECO2010-21643 financed by the Ministerio de Economía, the TRANSPORTRADE (S2007-HUM-0467) project financed by the Comunidad Autónoma de Madrid and funds received from the FPU-2010 program.

2.7 Bibliography

1. Anas, A., Xiong, K., (2003). Intercity trade and the industrial diversification of cities. *Journal of Urban Economics*, 54 (2), 258–276.

2. Anderson, J., van Wincoop, E., (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93 (1), 170–192.

3. Baldwin, R., Taglioni, D., (2006). Gravity for dummies and dummies for gravity equations. *Working Paper 12516*, National Bureau of Economic Research.

4. Behrens, K., (2005). Market size and industry location: traded vs non-traded goods. *Journal of Urban Economics*, 58 (1), 24–44.

5. Behrens, K., Mion, G., Murata, Y., Südekum, J., (2013). Spatial frictions. Tech. rep., *Discussion Paper Series*, Forschungsinstitut zur Zukunft der Arbeit.

6. Benedictis, L., Taglioni, D., (2011). The gravity model in international trade. In *De Benedictis, L., Salvatici, L. (Eds.), The Trade Impact of European Union Preferential Policies.* Springer Berlin Heidelberg, pp. 55–89.

7. Berthelemy, J., Varoudakis, A., (1996). Economic growth, convergence clubs, and the role of financial development. *Oxford Economic Papers-New Series*, 48 (2), 300–328.

8. Borraz, F., Cavallo, A., Rigobon, R., Zipitría, L., (2012). Distance and political boundaries: Estimating border effects under inequality constraints. Tech. rep., *National Bureau of Economic Research.*

9. Cavailhes, J., Gaigne, C., Tabuchi, T., Thisse, J.-F., (2007). Trade and the structure of cities. *Journal of Urban Economics*, 62 (3), 383–404.

10. Chaney, T., (2008). Distorted gravity: the intensive and extensive margins of international trade. *The American Economic Review*, 98 (4), 1707–1721.

11. Chen, N., (2004). Intra-national versus international trade in the european union: why do national borders matter?, *Journal of International Economics*, 63 (1), 93–118.

12. Combes, P.-P., Lafourcade, M., Mayer, T., (2005). The trade-creating effects of business and social networks: evidence from france. *Journal of International Economics*, 66 (1), 1 – 29.

13. Diallo, I. A., (2012). Suchowtest: Stata module to calculate successive chow tests on cross section data. https://ideas.repec.org/c/boc/bocode/s457536.html

14. Díaz-Lanchas, J., Llano-Verduras, C., Zofío, J. L., (2013). Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy. *Working Papers in Economic Theory 2013/10*, Universidad Autónoma de Madrid (Spain), Department of Economic Analysis (Economic Theory and Economic History).

15. Duranton, G., Morrow, P., Turner, M. A., (2013). Roads and Trade: Evidence from the US. *The Review of Economic Studies*, 81 (2), 681-724.

16. Eaton, J., Kortum, S., (2002). Technology, geography, and trade. *Econometrica*, 70 (5), 1741–1779.

17. Felbermayr, G. J., Kohler, W., (2006). Exploring the intensive and extensive margins of world trade. *Review of World Economics*, 142 (4), 642–674.

18. Garmendia, A., Llano, C., Minondo, A., Requena, F., (2012). Networks and the disappearance of the intranational home bias. *Economics Letters*, 116 (2), 178–182.

19. Hillberry, R. H., (2002a). Aggregation bias, compositional change, and the border effect. *Canadian Journal of Economics*, 35 (3), 517–530.

20. Hillberry, R., (2002b). Commodity flow survey data documentation. Tech. rep., *U.S. International Trade Commission*, Forum for Research in Empirical International Trade (F.R.E.I.T.).

21. Hillberry, R., Hummels, D., (2003). Intranational home bias: Some explanations. *Review of Economics and Statistics*, 85 (4), 1089–1092.

22. Hillberry, R., Hummels, D., (2008). Trade responses to geographic frictions: A decomposition using micro-data. *European Economic Review*, 52 (3), 527–550.

23. Hsu, W.-T., (2012). Central place theory and city size distribution. *The Economic Journal*, 122 (563), 903–932.

24. Llano, C., Esteban, A., Pérez, J., Pulido, A., (2010). Opening the interregional trade black box: The c-intereg database for the spanish economy (1995-2005). *International Regional Science Review*, 33 (3), 302–337.

25. Llano-Verduras, C., Minondo, A., Requena-Silvente, F., (2011). Is the border effect an artefact of geographical aggregation? *The World Economy*, 34 (10), 1771–1787.

26. Löffler, G., (1998). Market areas–a methodological reflection on their boundaries. *GeoJournal*, 45 (4), 265–272.

27. Martin, W., Pham, C., (2008). Estimating the gravity model when zero trade flows are important. *World Bank*. Mimeo.

28. Mayer, T., Ottaviano, G. I., (2008). The happy few: The internationalisation of european firms. *Intereconomics*, 43 (3), 135–148.

29. McCann, P., (2001). Urban and regional economics. *Oxford University Press*,Vol. 15, Oxford.

30. McCann, P., (2005). Transport costs and new economic geography. *Journal of Economic Geography*, 5 (3), 305–318.

31. Melitz, M. J., (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71 (6), 1695–1725.

32. Mulligan, G. F., Partridge, M. D., Carruthers, J. I., (2012). Central place theory and its reemergence in regional science. *The Annals of Regional Science*, 48 (2), 405–431.

33. O'sullivan, A., (2003). Urban economics. *McGraw-Hill/Irwin*.

34. Parr, J. B., (2002). The location of economic activity: central place theory and the wider urban system. *Edward Elgar* Cheltenham.

35. Poncet, S., (2005). A fragmented china: measure and determinants of chinese domestic market disintegration. *Review of International Economics*, 13 (3), 409–430.

36. Puga, D., (2010). The magnitude and causes of agglomeration economies. J *ournal of Regional Science*, 50 (1), 203–219.

37. Requena, F., Llano, C., (2010). The border effects in spain: an industry-level analysis. *Empirica*, 37 (4), 455–476.

38. Santos Silva, J. M., Tenreyro, S., (2006). The log of gravity. *The Review of Economics and Statistics*, 88 (4), 641–658.

39. Santos Silva, J. M., Tenreyro, S., (2010). On the existence of the maximum likelihood estimates in poisson regression. *Economics Letters*, 107 (2), 310–312.

40. Santos Silva, J. M., Tenreyro, S., (2011). Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112 (2), 220–222.

41. Tabuchi, T., Thisse, J.-F., (2011). A new economic geography model of central places. *Journal of Urban Economics*, 69 (2), 240–252.

42. Van der Marel, E., (2012). Determinants of comparative advantage in services. Tech. Rep. 87, *FIW Working Paper.*

43. Wolf, H. C., (2000). Intranational home bias in trade. *Review of economics and statistics,* 82 (4), 555–563.

44. Yi, K.-M., 2010. Can multistage production explain the home bias in trade? The American Economic Review 100 (1), 364–393.

45. Yilmazkuday, H., 2012. Understanding interstate trade pattern. *Journal of International Economics*, 86, 158–166.

46. Zofío, J. L., Condeço-Melhorado, A. M., Maroto-Sánchez, A., Gutiérrez, J., (2014). Generalized transport costs and index numbers: A geographical analysis of economic and infrastructure fundamentals. *Transportation Research Part A: Policy and Practice*, 67, 141-157.

2.8 Appendix 2.A: Kernel Regressions

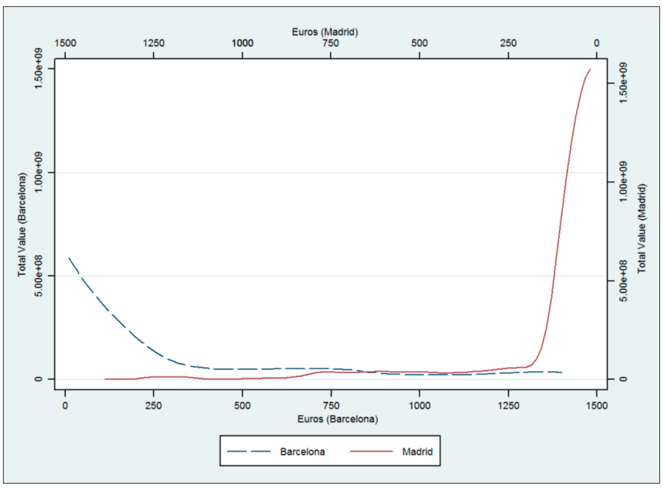


Figure 2.8: Trade areas competition for Madrid and Barcelona municipalities (NUTS-5).

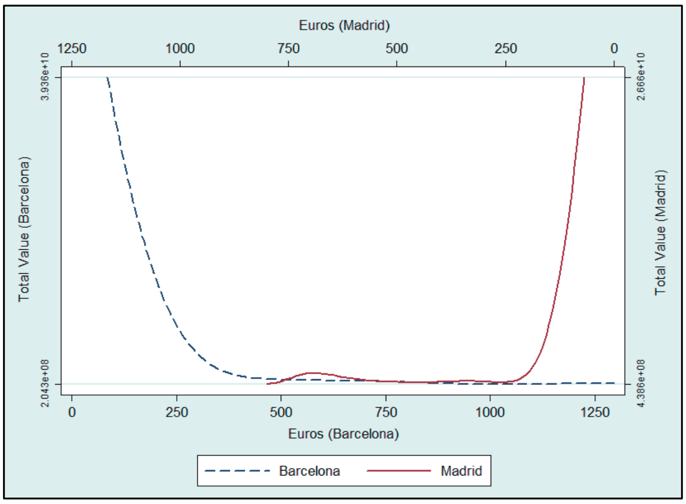


Figure 2.9: Trade areas competition for Madrid and Barcelona provinces (NUTS-3).

Chapter 3

Cities Export Specialization

"Cities are the absence of physical space between people and companies. They are proximity, density, closeness. They enable us to work and play together, and their success depend on the demand for physical connections."
(Edward Glaeser, *Triumph of the City*, 2011).

3.1 Introduction

Cities are one of the most salient examples of agglomeration economies. Despite the reduction in communication costs, the value of proximity does not seem to decline and an increasing share of people live in cities (Gaspar and Glaeser, 1998; Glaeser, 2011).

Cities clearly differ in their population size.[1] In particular, cities' population in most countries tend to follow a simple rank-size distribution: there are few big cities and a much larger number of cities with a smaller magnitude (Gabaix, 1999; Soo, 2005). For example, in the year 2013, the most populated urban area of Brazil, Sao Paulo, was 100 times larger than the urban area of Chapec, which had a population close to the median.[2]

The literature shows that urban areas population is correlated with other economic variables such as human capital, firm and labour productivity, and wage. Concretely, large cities present a higher share of skilled workers, host more productive firms and pay higher wages than small cities (Glaeser and Resseger, 2010; Behrens and Robert-Nicoud, 2015). In this paper we analyze whether export specialization also varies across cities.

We argue that since cities differ in their endowments and factors of production, we should expect cities to specialize and export some products and not others. Using very detailed product-level international export data for Brazilian urban areas, we find that

[1] We use the term "city" and "urban area" interchangeably.

[2] Data obtained from the Census of the Brazilian Statistical Institute. For a deeper analysis of the Zip's Law for the Brazilian case, revise Moura and Ribeiro (2006).

more populated Brazilian urban areas export proportionately more complex and skill-intensive goods than less populated Brazilian urban areas. Moreover, we show that cities that have increased more in population, have also augmented more than proportionately their exports of complex and skill-intensive goods.[3]

Complexity and skill-intensity are two alternative measures of the knowledge contended in goods. Whereas complexity focuses on the set of capabilities, or non-tradable inputs, needed to manufacture products, the skill-intensity emphasizes the qualification required to produce them. Although both dimensions seem to be similar, there are cases in which a large skill-intensity does not imply a high level of complexity. For instance, some goods such as pesticides, electrical equipment, medical diagnoses or software writing could be characterized by a high degree of skill-intensity, but at the same time they require very few capabilities to be produced (low complexity), Minondo and Requena-Silvente (2013).

Different models explain why large cities export proportionately more complex and skill-intensive goods. Regarding complexity, Hausmann and Hidalgo (2011) argue that as larger cities are endowed with a wider set of capabilities than smaller cities, they can produce, and export, more complex goods. Davis and Dingel (2014) develop a model to explain why more populated urban areas specialize in the production of more skill-intensive goods. In their model, skilled workers are more productive if they are surrounded by other skilled workers.[4] This generates an incentive for skilled workers to concentrate geographically, leading to a hierarchy of cities where the most skilled workers are concentrated in the most populated areas, and the least skilled workers in the least populated areas. If the productivity advantage of skilled workers raises with the skill-intensity of the good, larger cities will have a comparative advantage in producing skill-intensive products, and will export them.[5]

Our paper makes empirical contributions to the fields of urban and international trade economics. In particular, we show that population size leads to differences in export specialization across cities. Since trade data is characterized by a much larger product disaggregation than industry data, we can test predictions on cities' productive specialization with a much higher detail than previous studies. The paper also adds to the trade literature showing that variation in population might lead to differences in export specialization within a country.

[3]In the Appendix 3.A, we analyze and discuss the export specialization of the Spanish cities.

[4]The idea that big cities present advantages over small cities is also shared by other authors. In particular, Kok and Weel (2014) argue that big cities have the ability to geographically connect people that perform multiple tasks with different levels of skills. In the same spirit, Bacolod et al. (2009) show that social and cognitive abilities are more rewarded in big cities.

[5]Davis and Dingel (2014) measure the comparative advantage of cities resorting to production data for US. Contrary to theses authors, we argue that production data is not a good measure of cities comparative advantage. Export specialization through trade flows seems a better indicator than production or employment indicators, to show the differences in comparative advantage, Hausmann et al., (2007).

The remainder of the paper is organized as follows. Section 3.2 describes the data sources used in the empirical analyses. Section 3.3 presents patterns on cities' trading structure and shows the results of the empirical analyses. The last section concludes.

3.2 Data

To undertake our empirical analyses we combine three data sources for the year 2013: the distribution of Brazilian municipalities in urban areas, exports at the product and urban area level, and indicators on goods' complexity and skill-intensity. A more detailed description of the data follows below.

3.2.1 Urban areas

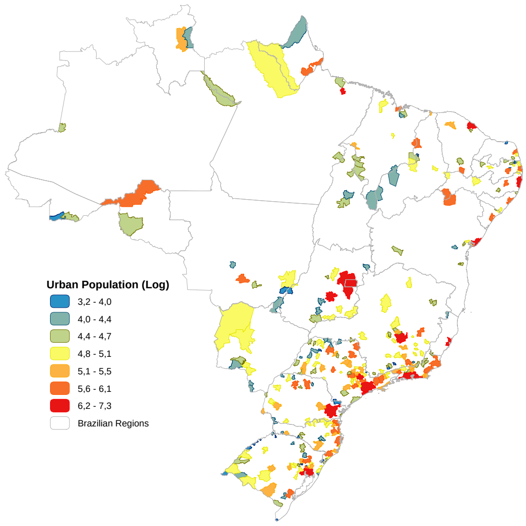
The Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística, IBGE) identifies urban areas (*arranjos*) attending to commuting labor movements. For this study, we select urban areas with a minimum of 50,000 inhabitants. This leaves us with 156 urban areas which are shown in Figure 3.1. The most populated urban areas are Sao Paulo (11.3 million), Rio de Janeiro (6.3 million), Salvador (2.7 million), Brasilia (2.6 million) and Fortaleza (2.5 million), whereas the less populated urban areas are Barra and Viana, both with 50.000 inhabitants.[6] Brazil is characterized by an unequal population distribution across cities. The Gini coefficient for population takes the value 0.73 in 2013; a previous estimate of the Zip's coefficient of around 2.26 for 2000 by Moura and Ribeiro (2006) confirms this unequal pattern. Attending to the spatial distribution, we can see that most urban areas are located in the south and near the coast, while areas in the north-west part of the country or the Amazon River tend to be much less populated.

3.2.2 Trade flows

Data on cities exports is obtained from the *DataViva* project.[7] *DataViva* uses firm-level exports data from the Brazilian Ministry of Development, Industry and Foreign Trade (MDIC), and assigns exports to a municipality according to the location of the firm's headquarter. Data is then collapsed at the municipal level and 4-digit HS product classification. We aggregate municipalities exports data into the 156 urban areas identified above. We restrict the sample to narrowly defined manufacturing industries, removing

[6] Urban area population is obtained from the Census of the Brazilian Statistical Institute and calculated using the *arranjos* definition.

[7] Available at http://en.dataviva.info/



Source: Own elaboration from Dataviva.

Figure 3.1: Urban areas in Brazil (2013).

those where natural resources play a major role in determining comparative advantage.[8] This leaves a sample of 780 manufactures. We have export data for the years 2000, 2005 and 2013; we use year 2013 for the baseline cross-section analyses and add the other two years for robustness checks and panel data analyses.

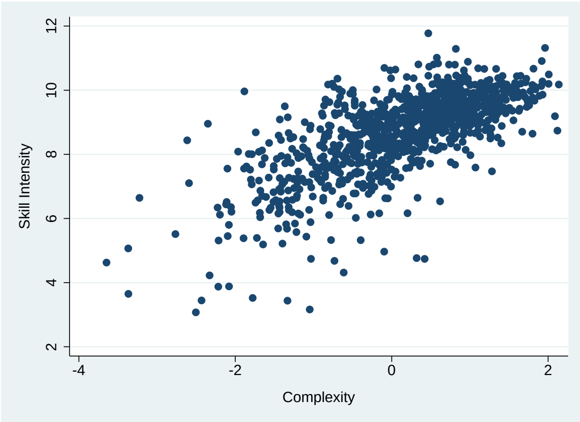
Brazil's exports grew above 300% between 2000 and 2013, with remarkable changes in the export patterns of the Brazilian cities during this period. Table 3.1 shows the average exports per city (intensive margin) and the number of different products exported per city (product extensive margin) in the years 2000 and 2013. As it can be observed, the intensive margin almost triples its value. In addition, cities exported more products in 2013 than in 2000 (272 vs. 297 at the HS4 level). The third column indicates that not only trade volumes, but also the number of exporting cities increased over the period of analysis.

Table 3.1: Descriptive Statistics.

Year	*Average Exports per City* (in millions)	*Number of different products per City (mean)*	*Number of Exporting Cities*
2000	2,33	272	147
2013	6,84	297	156

Source: Own elaboration from Dataviva.

[8]The narrow definition of manufacturing industries excludes agricultural and mineral raw materials, food and beverages, wood products, and non-metallic minerals.



Source: Own elaboration from Dataviva.

Figure 3.2: Complexity Vs. Skill-Intensity Indicators (2013).

3.2.3 The complexity and skill-intensity indicators

Data on products' complexity is obtained from the Observatory of Economic Complexity.[9] Complexity is calculated through an iterative process between countries diversification level, the number of products that countries export with a revealed comparative advantage, and products ubiquity, the number of countries that have a revealed comparative advantage in the product (Hausmann and Hidalgo, 2014). According to these authors, the complexity indicator measures the non-observable capabilities (know-how) required in the production of goods.[10]

Products' skill-intensity is proxy by the UNCTAD's Revealed Human Capital Intensity Index (Shirotori et al., 2010).[11] The index is a weighted average of schooling years of the countries that have a revealed comparative advantage in the product. The data covers the period 1980-2007 with annual frequency. We use the latest available year (2007) for our empirical tests.

Both indicators capture the products' skill-content taking two different but complementary perspectives. While complexity focuses on the diversity of skills used in the production process, the skill-intensity pays attention to the level of qualification required to manufacture a product. Overall, there is a strong positive correlation between the complexity and the skill-intensity index: 0.71. However, as shown in Figure 3.2, for a given level of skill-intensity there is ample variation in complexity. Note that for lower levels of complexity, the skill intensity indicator presents much more variability than for higher levels of complexity.

[9] Available at http://atlas.cid.harvard.edu/rankings/

[10] Complexity is a time-varying variable as it is calculated for each year.

[11] Available at http://www.unctad.info

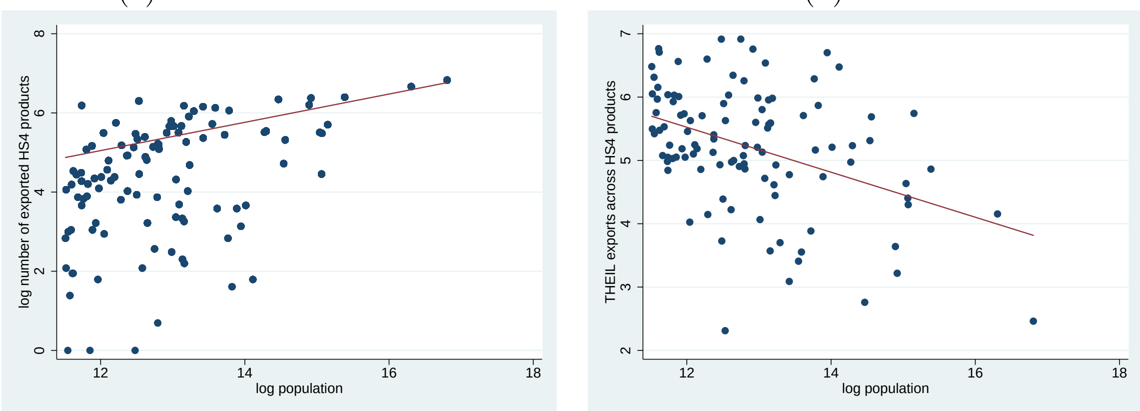
3.3 Empirical analyses

After describing the data set, we now analyze whether larger cities export disproportionately more complex and skill-intensive goods. For that purpose, we divide our analysis into two parts. First, we provide a descriptive analysis on the export specialization pattern of the Brazilian cities and, then, we perform the non-parametric and parametric tests to validate our hypotheses.

3.3.1 Descriptive analysis: Brazilian cities trade pattern

According to Hausmann and Hidalgo (2011), since large cities command a larger set of capabilities, they can produce a wider set of goods than smaller cities. Figure 3.3 gives evidence on these differences between small and large cities. Panel (a) shows the relationship between city's population and the number of exported products. Clearly, big cities trade more products than small ones. Attending to the extent of diversification in the export product portfolio, panel (b) plots the Theil index against the urban population.[12] The Theil index is inversely related with a city's export diversification level: the larger the index, the lower the diversification level. This sub-figure confirms that exports are less concentrated in large urban areas in Brazil.[13]

(a) Number of Varieties (b) Theil



Source: Own elaboration from Dataviva.

Figure 3.3: Cities trading-patterns (2013).

[12]The Theil index is defined as:

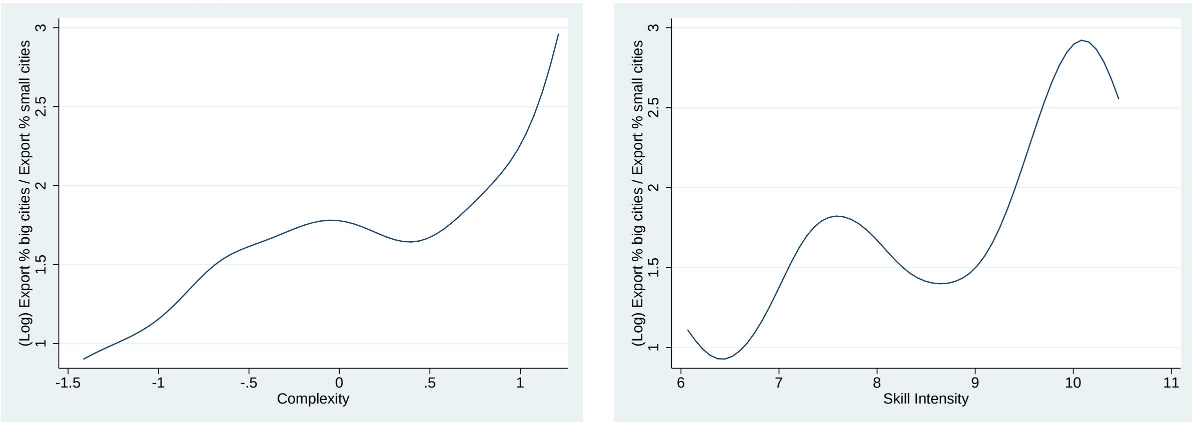
$$T = \frac{1}{J}\sum_{i=1}^{J}\frac{x_i}{\mu}\ln(\frac{x_i}{\mu}), \tag{3.1}$$

where $\mu = \frac{\sum_{x_i=1}^{J} x_i}{J}$, J is the total number of products and x_i denotes the amount of urban area exports of product i.

[13]Similar results are obtained using the Herfindahl and Gini indexes as measures of exports concentration.

Next, we divide cities in two groups according to their population: cities with population above the national median (large cities) and cities with population below the national median (small cities). Then, for each product, we calculate the ratio of the share of exports in large cities over the share of exports in small cities. If the theoretical predictions are correct, we would expect the ratio to become larger as the complexity (and the skill-intensity) of the good increases. Figure 3.4 presents the non-parametric relationship between the ratio of export shares and the complexity (panel a) and skill-intensity (panel b) indicators.[14] In both indicators, we observe that the share of exports in big cities relative to the share of exports in small cities rises as we increase the complexity and skill-intensity of the good. These two figures give the first insights on the pattern between the nature of the product and cities' exports: big cities specialize in more complex and skill-intensive goods. In the next section, we test the validity of these visual findings using alternative techniques.

(a) Complexity (b) Skill-intensity



Source: Own elaboration from Dataviva.

Figure 3.4: Ratio of exports shares big cities/small cities vs complexity and skill-intensity (2013).

3.3.2 A non-parametric analysis: The elasticity test

We present here the so-called elasticity-test, recently proposed by Davis and Dingel (2014) to analyze whether the population elasticity of exports is increasing with the level of complexity (skill-intensity) of the products. If we compare a complex (skill-intensive) product with a non-complex (unskilled-intensive) product, we expect exports to rise more with population in the former than in the latter. That is, if large cities export proportionately more complex goods, then the population elasticity of exports

[14]We estimate kernel regressions to obtain these non-parametric relationships.

should increase with products complexity. To test this hypothesis, first, we estimate the population elasticity of exports for each product in the following way:

$$\ln x_i^k = \alpha^k + \beta_k' \alpha^k \ln pop_i + \epsilon_i^k \tag{3.2}$$

where x_i^k are exports of product k by urban area i, α^k is a dummy variable if exports correspond to industry k and zero otherwise, pop_i is the population of urban area i and ϵ_i^k is the error term. Our coefficient of interest is β_k', which is a vector that comprises all the population elasticities for the k-products.

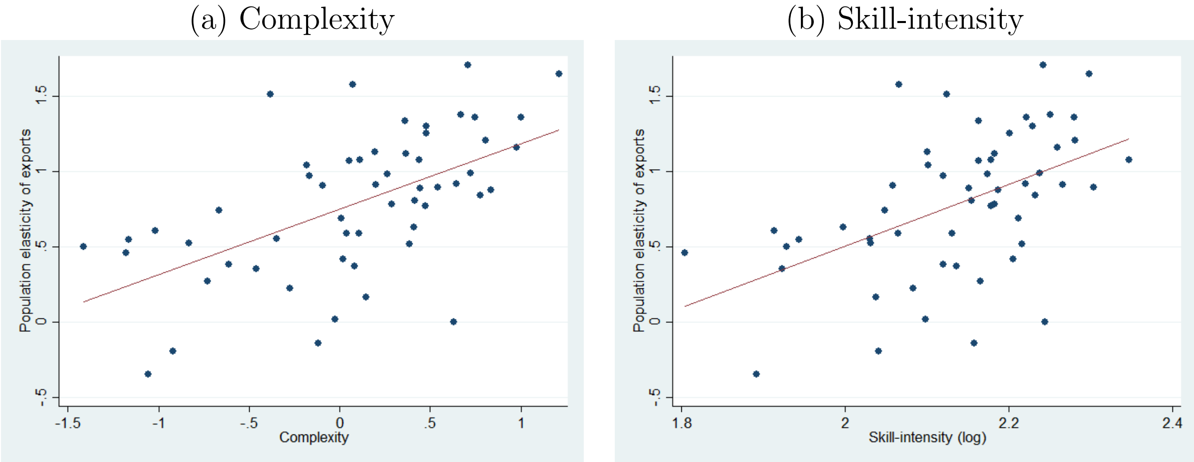
If we use our 4-digit product HS classification, we can distinguish 780 different manufactured products; unfortunately, the number of urban areas exporting product k will be, in many cases, very small. In those cases, the regression cannot be estimated; or, even when the estimation is feasible, the elasticity coefficient will have a large standard error. To overcome this problem, we collapse exports at the 2-digit HS classification, which distinguishes 59 manufactured groups. To ensure the precision of the estimation, product k should be exported, at least, by 10 urban areas. This condition reduces the sample of products to 55, but they still represent 98% of exports.

Figure 3.5-panel (a) presents the scatter diagram between products' complexity and the population elasticity of exports; panel (b) presents a similar scatter diagram for products' skill-intensity. We can see a positive relationship between both indicators and the population elasticity of exports. It is important to note that there are 4 products that have negative elasticities. These negative signs are awkward as we expect more populated cities to export more of any good. Nevertheless, the 4 negative coefficients are not statistically significant, so we cannot reject the hypothesis that their true values are zero. More important, their inclusion in the sample does not alter the outcome of the elasticity test. We have also run the test including only products whose elasticities are statistically significant. In that case the sample is reduced to 44 products and the conclusions are not altered.

To confirm the visual appreciation from Figure 3.5, the elasticity test compares whether a more complex (skill-intensive) product has a higher population elasticity of exports than a less complex (skill-intensive) product. More concretely, for any given pair of products k and j, the elasticity test is defined as:

$$Complexity_k > Complexity_j \iff \beta_k' > \beta_j' \tag{3.3}$$

That is, if k is a more complex product than j, we should expect the population elasticity of k to be higher than j. According to the previous expression, we define an indicator (I_{kj}) that takes the value 1 if equation 3.3 holds, i.e. $\beta_k' - \beta_j' > 0$, and 0 otherwise. The elasticity test calculates the number of cases such as I_{kj}=1 among all



Source: Own elaboration from Dataviva.

Figure 3.5: The population elasticity of exports vs. complexity and skill intensity (2013).

possible (k,j) pairs. More concretely, it takes the following expression:

$$Elasticity\_Test_{kj} = \frac{[\sum_{kj}(I_{kj})]}{N_{kj}} \tag{3.4}$$

Where N_{kj} represents the total number of pairs (k,j). If all the pairs (k,j) fulfill equation 3.3, the final indicator will reach the value 1 (100% of the cases), fully confirming the hypothesis. Notice that if equation 3.4 takes a value of 0.5 (50% of matches are correct), the performance of the test is quite poor. In this case, similar to the sign test used to validate the Heckscher-Ohlin-Vanek equation, our theory would predict as well as flipping a coin (Trefler, 1995).

As shown in Table 3.2, the hypothesis is confirmed for 71% of cases when products are classified by complexity, and in 69% of cases when products are classified by skill-intensity.[15] The table also reports the success rate for weighted comparisons. We argue that it is more damaging for the test when a low complexity product has a higher population elasticity than a high complexity product, than if a low complexity product has a higher population elasticity than a moderately more complex product. To take that into account we weight the success events by the difference in complexity (skill-intensity) of the products that are compared. The weighting (W) version of the elasticity test takes the form:

$$I^W = \sum_{kj} I * \left(\frac{\mid complexity_k - complexity_j \mid}{\sum_{kj} \mid complexity_k - complexity_j \mid} \right), \forall (\beta'_k - \beta'_j) > 0 \tag{3.5}$$

Where I reflects the indicator from equation 3.4. Then, for the cases such as $\beta'_k - \beta'_j >$

[15]Concretely, we perform (55 coefficients x 54 coefficients)/2= 1,485 comparisons.

0, we weight by the differences in complexity (skill-intensity) between the products k and j. With weighting, the success rate improves to 81% for complexity and 78% for skill-intensity. These percentages confirm the hypothesis that large cities specialize in more complex and skill-intensive products. This result is in line with those obtained by Davis and Dingel (2014) for the US who find that the population elasticity of sectoral employment is increasing in skill-intensity.

Table 3.2: Results for the elasticity test (% of successful comparisons)

Variable	Unweighted (%)	Weighted (%)
Complexity	71	81
Skill-intensity	69	78

3.3.3 A parametric analysis: The log-supermodularity regression

The elasticity test gives the first confirmation of our hypothesis; now we turn to the parametric test to corroborate it.[16] This test is based on a log-supermodularity regression. As explained by Costinot (2009a), two variables have a log-supermodular relationship when increasing one variable is relatively more important when the other variable is high. In our analysis, if large cities export proportionally more complex (skill-intensive) goods, then exports should be log-supermodular in population and complexity (skill-intensity). To test the log-supermodularity hypothesis, we estimate the following regression equation adapted from previous studies (Costinot, 2009b; Chor, 2010):

$$\ln x_{ik} = \beta(\ln pop_i * complexity_k) + \mu_i + \mu_k + \epsilon_{ik} \tag{3.6}$$

where x_{ik} are product k exports from urban area i, pop_i is urban area i population, $complexity_k$ is the complexity (skill-intensity) of product k, μ_i is an urban area fixed effect, μ_k is a product fixed effect, and ϵ_{ik} is the error term. If exports are log-supermodular in population and complexity, we should expect a positive value for the coefficient β. We estimate a similar equation substituting complexity with skill-intensity as product-specific explanatory variable.

We estimate equation 3.6 by OLS. However, due to the large number of zeros in our sample, the estimation might be biased. To test the robustness of our estimations, we also estimate equation 3.6 using the pseudo-maximun likelihood (PPML) model proposed by Silva and Tenreyro (2006). In Table 3.3, columns (1) to (4) present the results for 2013

[16] In the Appendix 3.B we present a further discussion on the empirical tests.

of the OLS estimation. In columns (2) and (4) we include zero exports transforming all zeros into ones, so logarithms become zero. Columns (5) and (6) estimate the equation with the PPML estimator.

Attending to the results, we can see that coefficients for the interaction variables are positive and statistically significant in the OLS estimations, both for the complexity and the skill-intensity indicators, giving support to our hypotheses on log-supermodularity. As expected, the OLS coefficients are slightly higher than those obtained with the PPML estimation, specially for the complexity variable. Note also that results are robust in the PPML estimations. Hence, we can conclude that exports are log-supermodular in complexity (skill-intensity) and population.

Table 3.3: Log-supermodularity regressions. Regression results for 2013.

	OLS				PPML	
	(1)	(2)	(3)	(4)	(5)	(6)
Pop*complex	0.178***	0.171***			0.114**	
	(0.035)	(0.039)			(0.053)	
Pop*skills			0.111***	0.0839***		0.101***
			(0.022)	(0.023)		(0.038)
Zero exports	No	Yes	No	Yes	Yes	Yes
Observations	12142	114816	12142	114816	114816	114816

Note: All regressions include city-product and year fixed effects.
Standard errors in parentheses. *** statistically significant at 1%, ** statistically significant at 5%.

We test whether exports' log-supermodularity also holds in the long run. We expect exports of complex goods to increase more in urban areas where population has grown more. To test this hypothesis, we extend the database for the years 2000, 2005 and 2013 to estimate the following expression:

$$\ln x_{ikt} = \beta(\ln pop_{it} * complexity_{kt}) + \mu_{ik} + \mu_t + \epsilon_{ikt} \tag{3.7}$$

This test is similar to the quasi-Rybczynski effect analyzed, among others, by Ventura (1997) and Romalis (2004). Note that we resort to panel data techniques to estimate equation 3.7. To control for zeros, we use Poisson fixed-effects regression techniques including year (μ_t) and city-product (μ_{ik}) fixed effects.

Table 3.4 shows the results for the two indicators using the within-groups estimation in columns (1) to (4) and the Poisson-fixed effects estimation in columns (5) and (6).[17] All the interaction coefficients are positive and significant for complexity and skill-intensity,

[17]The Hausman test indicates that the model should be estimated with fixed effects. Results for the Poisson estimation are the same either using random or fixed effects; thus, we select the fixed effects estimation because its equivalency with the PPML estimator when resorting to a panel database.

in line with those obtained in Table 3.3. They confirm that exports of complex goods have grown more in cities that have experimented larger increases in population. All these findings confirm the export specialization patterns observed in previous sections.

Table 3.4: Log-supermodularity regressions. Panel regression results for 2000, 2005 and 2013.

	PANEL				POISSON	
	(1)	(2)	(3)	(4)	(5)	(6)
Pop*complex	2.826***	1.414***			2.658***	
	(0.130)	(0.051)			(0.000)	
Pop*skills			0.530***	0.247***		0.528***
			(0.016)	(0.006)		(0.000)
Zero exports	No	Yes	No	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	31660	342868	31660	342868	51151	51151

Note: All regressions include city-product and year fixed effects.
Standard errors in parentheses. *** statistically significant at 1%, ** statistically significant at 5%.

3.4 Conclusions

Previous research shows that large cities have more productive firms, pay higher wages and have a superior share of educated workers than small cities. In this paper we test whether cities also differ in another economic aspect: export specialization. We analyze whether large cities export proportionately more complex and skill-intensive goods than small cities. To carry out the empirical analyses we use very detailed product-level export data for Brazilian urban areas. We find that cities differ in their trading-structure: large cities have a more diversified export product portfolio and tend to specialize in more knowledge-intensive products.

The elasticity test and the log-supermodularity regressions provide strong empirical support to the hypothesis that large cities specialize in complex and skill-intensive goods. We also provide evidence that as cities grow in population they tend to specialize in more complex goods. These results suggest that a large population facilitates the development of a comparative advantage in complex and skill-intensive goods. This first analysis opens the door for a future research on the role that the demand from the destination has on product specialization.

Acknowledgments

I deeply acknowledge support and advise from Asier Minondo (Deusto Business School) and Franciso Requena-Silvente (University of Valencia) to develop this chapter. It started when I was a visiting researcher at Deusto Business School (San Sebástian, Autumn 2014).

I thank comments received from José Luis Zofío, Jonathan Dingel, Jesús Fernández Huertas and participants at the XVIII Encuentro de Economía Aplicada, the EUREKA Seminar (Vrije Universiteit Amsterdam), the XLI International Conference on Regional Science, the Third International Workshop on Inter-Industry Relatedness, the 18th EBES Conference, the 62nd Annual North American Meeting of the RSAI, the XII INTECO Workshop, and participants at the Research Seminar at the Universidad Autónoma de Madrid and the University of Sheffield.

Also I gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (MINECO ECO2013-46980-P, co-financed with FEDER), the Basque Government Department of Education, Language policy and Culture (IT629-13) and the FPU-2010 Scholarship for pre-doctoral Education Training by the Spanish Ministry of Education.

3.5 Bibliography

1. Bacolod, M., Blum, B. S., and Strange, W. C. (2009). Skills in the city. *Journal of Urban Economics*, 65 (2), 136–153.

2. Behrens, K. and Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In *G. Duranton, J.V. Henderson, and W. Strange (eds.) Handbook of Regional and Urban Economics*, Volume 5. Amsterdam: Elsevier.

3. Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics*, 82 (2), 152–167.

4. Costinot, A. (2009a). An elementary theory of comparative advantage. *Econometrica*, 77 (4), 1165–1192.

5. Costinot, A. (2009b). On the origins of comparative advantage. *Journal of International Economics*, 77 (2), 255–264.

6. Davis, D. R. and Dingel, J. I. (2014). The Comparative Advantage of Cities. *Working Paper 20602*, National Bureau of Economic Research.

7. Gabaix, X. (1999). Zipf's law and the growth of cities. *The American Economic Review*, 89 (2), 129–132.

8. Gaspar, J. and Glaeser, E. L. (1998). Information technology and the future of cities. *Journal of Urban Economics*, 43 (1), 136–156.

9. Glaeser, E. L. (2011). *Triumph of the city: How our greatest invention makes US richer, smarter, greener, healthier and happier*. Pan Macmillan.

10. Glaeser, E. L. and Resseger, M. G. (2010). The complementarity between cities and skills. *Journal of Regional Science*, 50 (1), 221–244.

11. Hausmann, R. and Hwang, J. and Rodrik, D. (2007). What you export matters. *Journal of Economic Growth*, 12 (1), 1–25.

12. Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16 (4), 309–342.

13. Hausmann, R. and Hidalgo, C. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press.

14. Kok, S. and Weel, B. t. (2014). Cities, tasks, and skills. *Journal of Regional Science*, 54 (5), 856–892.

15. Minondo, Asier and Requena-Silvente, Francisco (2013). Does complexity explain the structure of trade? *Canadian Journal of Economics*, 46 (3), 928–955.

16. Moura, N. J. and Ribeiro, M. B. (2006). Zipf law for brazilian cities. *Physica A: Statistical Mechanics and its Applications*, 367, 441–448.

17. Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review*, 94 (1), 67–97.

18. Shirotori, M., Tumurchudur, B., Cadot, O., et al. (2010). Revealed factor intensity indices at the product level. *United Nations Conference on Trade and Development.*

19. Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88 (4), 641–658.

20. Soo, K. T. (2005). Zipf's law for cities: a cross-country investigation. *Regional science and urban Economics*, 35 (3), 239–263.

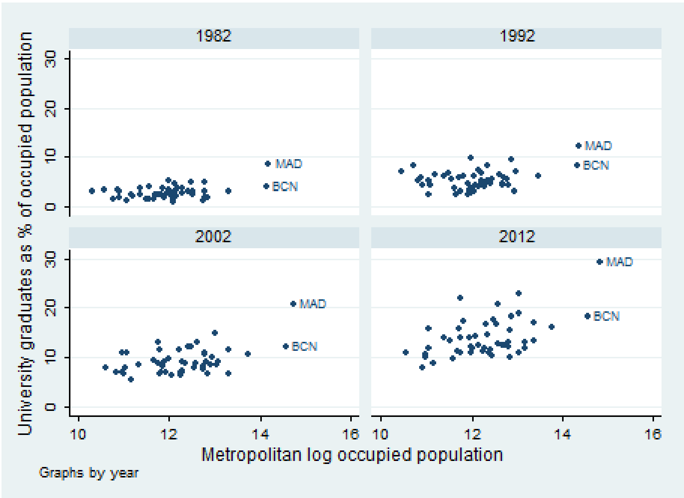
21. Trefler, D. (1995).The case of the missing trade and other mysteries. *The American Economic Review*, 85 (5), 1029–1046.

22. Ventura, J. (1997). Growth and interdependence. *Quarterly Journal of Economics*, 112 (1), 57–84.

3.6 Appendix 3.A: Spanish trading-cities

The chapter focuses on the Brazilian cities, but similar evidence on the export specialization of cities has been obtained for Spain in two stages of development: 1988 and 2012. This Appendix summarizes and highlights the most important results for the Spanish cities not included in the main paper.[18]

Spain proves to be a relevant case of study as its cities have experimented remarkable changes in recent years. As show in Figure 3.6, Spanish cities present a dynamic urban pattern as a result of attracting high-educated workers, specially in the biggest cities Madrid (MAD) and Barcelona (BCN). As argue along this paper, because big cities host more educated workers, they should specialize in skill-intensive products.[19]



Source: Own elaboration from INE Census.

Figure 3.6: Share of educated people over occupied urban population (1982-2012).

3.6.1 Appendix 3.A: Data

For this Appendix, we use the functional urban areas identified by the OECD for Spain. The OECD follows a three step approach to define functional urban areas (Brezzi et al., 2012).[20] First, they identify densely populated municipalities. Second, they aggregate densely populated municipalities into an urban area if more than 15% of the population of one municipality commutes to work in the other municipality. Finally, municipalities

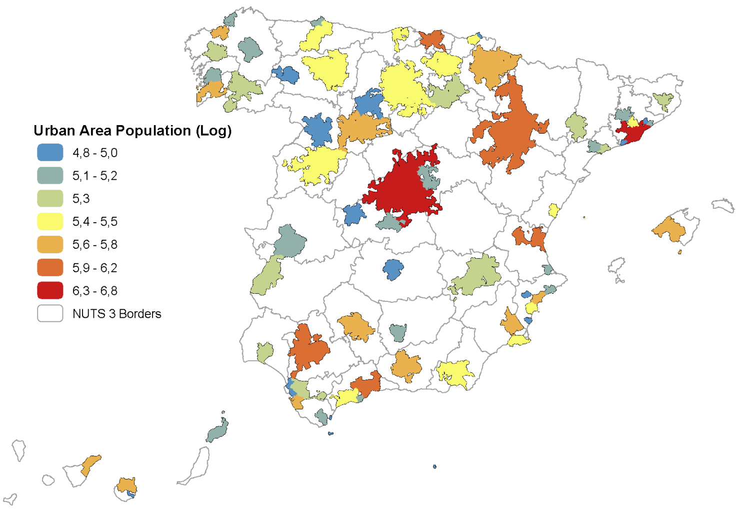
[18] As explained below, databases for Spain are not consistent and robust across years, leading us to focus the paper only in the Brazilian case as its panel database is robust and comparable in time.

[19] Data on urban area population and education levels are obtained from the Spanish Statistical Institute (INE) Census.

[20] Brezzi, M., Piacentini, M., Rosina, K., and Sanchez-Serra, D. (2012). Redefining urban areas in OECD countries. *Redefining "Urban"*, pages 19–58.

that have a low population density are assigned to an urban area if at least 15% of their employed population work in that urban area.

Figure 3.7 shows the 76 urban areas identified by the OECD in Spain. They account for 67% of the Spanish population in 2012. Spain has 2 large metropolitan areas: Barcelona and Madrid (with a population of 1.5 million of more), 6 metropolitan areas (with a population between 500,000 and 1.5 million), 22 medium-size urban areas (with a population between 200,000 and 500,000), and 46 small urban areas (with a population below 200,000 people). All the large metropolitan and metropolitan areas, and most of the medium-size urban areas, are located around a province capital.



Source: Own elaboration from OECD Urban Areas Classification.

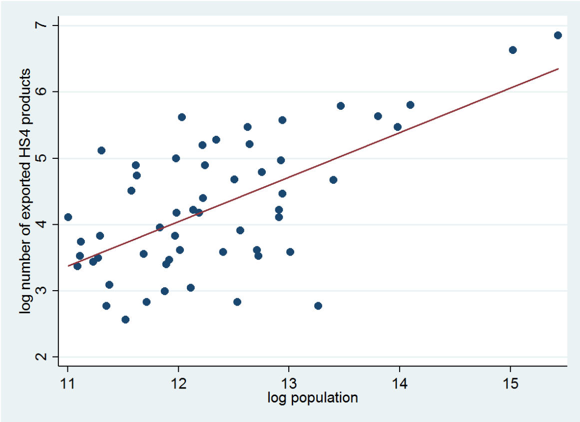
Figure 3.7: Spanish Urban Areas (2012).

We calculate cities exports for 1988 and 2012 resorting to two firm-level databases: CamerData and the Directory of the Spanish Exporting and Importing firms.

- *CamerData (1988).*

  This database corresponds to the Census of manufacturing firms with more than 10 employees that exported products in 1988. With this database we get information on about 7,300 firms where we can identify their corresponding municipal-ZIP code and the products (HS4 classification) that they export. In a further step, we aggregate at the urban level according to the previous OECD classification.

  Figure 3.8 plots the relationship between the number (varieties) of exported products and the urban population in 1988 using CamerData. As can be observed, more populated urban areas export a higher number of varieties. This evidence is similar to the one for the Brazilian case in Figure 3.3 (a).



Source: Own elaboration from CamerData.

Figure 3.8: The relationship between the number of exported products and population.

- *Directory of the Spanish Exporting and Importing firms (2012).*

  Data for 2012 comes from the Directory of the Spanish Exporting and Importing firms, which is maintained by the Spanish Chamber and the Inland Revenue Agency (Cámara de Comercio and Agencia Tributaria de España). Firms included in the Directory report the annual value of exports per product (at the 8-digit level of the European Common Nomenclature classification) and destination. The database also reports the location of the firm, so we can add-up exports by product and destination at the urban area level. The export transactions that can be allocated to an urban area represent 73% of all export transactions included in the Directory, representing 81% of the total export value accounted in this database. The Directory includes exports data for 5,896 firms, which represent approximately 7% of Spanish exporters in 2012. These firms account for 27% of Spanish exports in 2012.[21]

3.6.2 Appendix 3.A: Empirics

In this section we present the results for the two empirical tests analyzed through the paper: the elasticity test and the log-supermodularity test. Both have been performed

[21] While CamerData consider all exporting firms in 1988, the Directory fails in representativeness for 2012. The largest effort to include firms in the Directory took place during the years 2003 and 2004. In that time, almost 30,000 exporters were approached by mail by the Inland Revenue Agency and the Chamber of Spain asking them permission to include their trade data in the Directory. In these large group of firms they included the most important exporters, which explains the highest representativeness of the Directory in terms of value than in terms of exporters. Unfortunately, the Directory has not been revised and maintained since the years 2003-2004, causing its loss of representativeness as the main part of the exporting firms have disappeared from the Directory. We resort to 2012 to show and analyze the trends of the Spanish exporting cities in comparison to those their Brazilian counterparts, although having in mind that the comparisons are not as robust as one should expect.

using the complexity and the skill-intensive measures previously explained. We only show the results alternating the complexity and the skill-intensity measure to not overextend the Appendix.

- *The elasticity test.*

  Figure 3.9 presents the correlation between industries population elasticities for Spain in 1988 (a) and 2012 (b), and the two indicators alternatively. In all cases, there is a positive relationship between complexity/skill-intensity measures and population elasticities. The slope is stepper for 2012 as the Directory is biased towards products that heavily depend on the distribution processes that take place in big cities to be exported.[22]

Spain (1988)-Complexity Indicator   Spain (2012)-Skill Intensity Indicator

Source: Own elaboration from CamerData (1988) and the Directory of exporting firms (2012).

Figure 3.9: Industries' population elasticities and skill intensities.

Once the population elasticities have been calculated, we carry out the elasticity test following the previous expression 3.4 in the main text. Table 3.5 shows the results for the elasticity test using the weighted version of the test as, above, Table 3.2 points out to more accurate results when differences in complexities/skill-intensity between products are accounted for. Either for Spain in 1988 or in 2012, the empirical test concludes that more complex/skill-intensive goods have higher population elasticities. These results are in line to those obtained with the elasticity test for Brazil.[23]

[22] Panel (b) plots the products' nomenclature to show the predominance of those belonging to the groups 32 and 33 (chemical and cosmetic products) from the 2-digit HS-classification.

[23] To reinforce the results in 2012, we also resort to an alternative database. Concretely, we use 8-digit product exports at the Province levels (NUTS-3) from the Inland Revenue Agency. When the main city of a Province accounts for different Province's population thresholds, we assume that exports from this

Table 3.5: Results for the elasticity test (% of weighted successful comparisons)

Variable	Spain 1988 (%)	Spain 2012 (%)
Complexity	66	94
Skill-intensity	65	97

- *The log-supermodularity test.*

  Finally, Table 3.6 relates to the log-supermodularity hypothesis using equation 3.6 to validate the preliminary findings from the elasticity test. It shows the results from the OLS estimations for the two years attending to the complexity and the skill-intensity dimensions.

Table 3.6: Log-supermodularity regressions. OLS Regression results for 1988 and 2012.

	SPAIN (1988)		SPAIN (2012)	
Pop*complex	0.129***		0.0916***	
	(0.021)		(0.010)	
Pop*skills		0.122***		0.0121***
		(0.023)		(0.001)
Zero exports	No	No	No	No
Observations	6820	6820	30657	30657

Note: All regressions include city-product fixed effects.
Standard errors in parentheses. *** statistically significant at 1%.

The complexity and the skill-intensity indicators are both positive and significant in all specifications. Table 3.6 accepts the hypothesis of log-supermodularity between the urban population and the complexity/skill-intensity of the product either in 1988 or 2012. Hence, we can conclude that exports are log-supermodular in complexity (skill-intensity) and population. In particular, we confirm that large Spanish cities export proportionally more complex (skill-intensive) goods than smaller ones. These results are in line those obtained for Brazil. Even more important, we find that cities from two very different countries (Spain and Brazil) in separate moments of time and development stages (1988 and 2012 for Spain; 2000 and 2013 for Brazil), follow similar patterns in export specialization just because they host high-educated workers.

Province are representative of those at the city level. In particular, we argue that if a city represents about 66%, 75% or 90% of total Province's population, Province's exports are representative of those from its main city. Following this criteria, the elasticity test gives the same results as those in Table 3.5.

3.7 Appendix 3.B: The pairwise comparisons test

Apart from the elasticity test, Davis and Dingel (2014) propose another additional non-parametric test for analyzing cities comparative advantage, called the *pairwise comparisons test*. According to their framework, this test can be directly obtained from the properties belonging to the elasticity test, that is, if the conditions for the elasticity test are accomplished, they argue the pairwise comparison test should also holds.

In the context of our analysis, for any pair of products (k and m) and cities (i and j), the share in exports (x) of the more complex/skill-intensive product relative to the share in exports of the less complex/skill-intensive product should be equal or larger in the more populated city than in the less populated city. Algebraically, the pairwise comparisons test takes the form:

$$\frac{x_i^k}{x_i^m} \geqslant \frac{x_j^k}{x_j^m} \tag{3.8}$$

In the spirit of the elasticity test, the pairwise comparisons test consists in determining the number of cases in which Equation 3.8 holds. Following these authors, to perform this test we aggregate urban areas into bins according to population size. For example, in the first pairwise comparison urban areas are aggregated in two bins: the large-population bin and the small-population bin. Afterwards, we repeat the test for an increasing number of bins, until we reach the highest disaggregation, where the number of bins equals the number of urban areas in the sample. Obviously, as the number of bins increases the test performs a larger number of comparisons.

Although Davis and Dingel (2014) show that this test fits well for production-city data. Through this paper we check that this test does not work properly for export-city data when the share of zero-export flows becomes enormous. As explained above, our sample comprises 156 urban areas and 780 manufactures. With this disaggregation, the maximum number of comparisons is 3,7 billion.[24] However, for many combinations of urban areas and manufactures the value of exports is zero.[25] As the test compares two ratios, if the product in the denominator has zero exports, the ratio cannot be calculated. We perform two sets of calculations. In the first set we calculate the ratios only when all exports are positive. In the second set of calculations we allow a zero exports value for the product in the numerator, providing that the denominator is different from zero.[26]

Considering only the positive-export flows, the test is confirmed in the 53% of the

[24]Number of comparisons=[(156x155)/2]x[(780x779)/2]

[25]Concretely, only the 10% of all possible export combinations of urban areas and manufactures have positive values.

[26]In the case that both the left-hand side and the right-hand side ratios are zero, we exclude them.

cases, when we do not weight for the differences in complexity between products, and in 55% of the cases when weights are included. We check that differences among cities reduce in this sub-sample as it only encompasses trade for medium and big cities where their shares of positive exports are bigger.[27] When accounting for zero export-flows, although the results improved to 77% in the weighted version, the asymmetry between both sides of Equation 3.8 increases because of incorporating more smaller cities with no trade in the numerator in one of these sides. Thus, the test gets bias in favor of medium and big cities. Taking into account all these issues, we reject the pairwise comparisons test as a valid empirical test when the share of zero-export flows is remarkably high.

[27] As a result of restricting the sample to only medium and big cities (i.e. positive trade flows), we consider that the test fails in accomplish its target, first, because the differences in trade patterns reduces across cities, making harder to analyze our hypotheses, and second, because we are dropping a remarkable number of cities from the sample.

Chapter 4

On the Emergence of Diversified and Specialized Cities

"Every cheapening of the means of communication, every new facility for the free interchange of ideas between distant places alters the action of the forces which tend to localize industries. Speaking generally we must say that a lowering of tariff, or of freights for the transport of goods, tend to make each locality buy more largely from a distance what it requires; and thus tends to concentrate particular industries in special localities." (Alfred Marshall, *Principles of Economics*, 1890)

4.1 Introduction

What determines the emergence of specialized or diversified cities? Many economists have dealt with this fundamental question. Determining why this is, has so far been approached mainly from the standpoint of solving a theoretical puzzle. Leading contributions in this field develop models that explain size and composition of cities in an urban system in the spirit of Christaller (1933); Jacobs (1969); Henderson (1974); Fujita et al. (1999). We contribute to this literature with an empirical study that aims to assess the key mechanisms and assumptions proposed in the theoretical literature. Most of the existing empirical literature about the composition, size and location of cities presents evidence from the US (Duranton and Puga, 2000). In contrast, we focus on Spain. In our analysis we consider the possibility that cities cannot be only either specialized or diversified, but also either a combination of both (co-agglomeration economies) or none of them (cf. Ellison and Glaeser (1997); Duranton and Puga (2000); Ellison et al. (2010)). We not only consider internal cities characteristics but also the role of cities interactions as potential determinant of sector composition differences across the urban system. To this aim, we propose an innovative empirical approach, where we develop bivariate probit models to data from three micro-datasets: a firm-level database that permits us to calculate levels of specialization and diversification at the city level; a trade database between Spanish cities; and a unique database on economic transport cost of moving between Spanish cities.[1]

For our period of analysis (1995-2007) Spain is interesting in its own because it exhibits above average dynamics in terms of GDP and population growth, city formation, transport costs and mobility of goods and households as a result of reforming its economy in the context of EU membership. After decades of very limited infrastructure improvements, the vast amounts of resources from the European regional funds (EU-FEDER) enabled Spain to significantly reduce the economic costs of moving goods and people, both within the country (Zofío et al., 2014) and within Europe. At the same time Spain experienced a housing bubble that modified its internal economic structure, while relatively re-orienting trading activity from international to internal markets (Llano et al., 2010). We show in this paper that, over time, Spanish cities change their typology more than expected: Cities dynamics instead of cities stability is more the rule than the exception in our sample.

Different strands of theoretical literature have analyzed to what extent city formation responds to (changes in) the production and sector structure (Fujita, 1988; Rivera-Batiz, 1988; Abdel-Rahman, 1988, 1990a,b, 1996; Abdel-Rahman and Fujita, 1993; Henderson, 2000, 2003; Helsley and Strange, 2014), trading costs and interaction with other cities and regions (Eaton and Lipsey, 1982; Abdel-Rahman, 1996; Duranton and Puga, 2001; Fujita et al., 1999; Anas and Xiong, 2003, 2005; Tabuchi and Thisse, 2011; Fujita, 2012), and technological change, product life-cycles and learning processes (Brezis and Krugman, 1997; Henderson, 1997; Duranton and Puga, 2000). The range of theoretical models presented in these and other contributions differ in their assumptions about the (micro foundations of) agglomeration and dispersion forces, mobility of goods and households and city formation mechanisms.[2] In our empirical search for the determinants of diversification vs. specialization patterns across Spanish cities, we study a wide range of characteristics including the role of city size, transport costs, skills, sector composition, product standardization, natural advantage, trade intensity, trade variety, typology of intercity-trading partners, and the degree of specialization in the past.

We find, in line with previous evidence for the US, a higher probability for a city to become specialized if its labor force is relatively low skilled, its production structure is biased towards manufacturing and standardized products, and its size relatively small. The opposite characteristics increase the probability that a city becomes diversified, while a mix of these results help explaining the emergence of so-called co-agglomeration

[1]Through this paper and in contrast to Chapter 3, trade is considered as internal trade within a country (Spain). The original sources for the transport costs and the trade databases are the same as in Chapter 2.

[2]We refer to Duranton and Puga (2000) for an in-depth review.

cities. In addition, we find that history matters: the degree of specialization in the recent past proves to be a strong predictor of current specialization patterns. Finally, and most interestingly, we find that the probability of becoming a specialized (diversified) city depends positively on low (high) transport costs in the past (from 1980 and subsequent years), while the opposite holds for more recent and present-day transport costs. We hypothesize that these findings suggest that over time the internal structure of cities has become a more important driver of specialization and diversification patterns across the urban system, but the former location of the city plays even a more crucial role in city development.

The remainder of the paper is organized as follows. In section 4.2 we present our empirical approach, including definitions of indices and variables as well as the econometric specification. Section 4.3 describes our data sources as well as a number of stylized facts about the Spanish urban system. In section 4.4 we present the results of our analysis. Section 4.5 discusses the results and 4.6 concludes.

4.2 Methodology

4.2.1 Indices for Diversification and Specialization in cities

Following Henderson (2003) and Duranton and Puga (2000), we define the diversification and specialization of cities resorting to the share of employment x of sector k in city i at time t as s_{ikt}:

$$s_{ikt} = \frac{x_{it}^{k}}{\sum_k x_{it}} \tag{4.1}$$

To control for the fact that some sectors account for larger shares than others, we calculate specialization in city i in relative terms by dividing it by the share of each sector at the national level s_{kt} such as:

$$s_{kt} = \frac{x_{t}^{k}}{\sum_k x_{t}} \tag{4.2}$$

With s_{ikt} and s_{kt}, we consider the *relative-specialization index* (RZI_{it}) in city i as:

$$RZI_{it} = max_k(\frac{s_{ikt}}{s_{kt}}) \tag{4.3}$$

Attending to diversification, the *relative-diversification index* (RDI_{it}) defines as:

$$RDI_{it} = \frac{1}{\sum_k |s_{ikt} - s_{kt}|} \tag{4.4}$$

RZI_{it} and RDI_{it} are continuous variables. We split cities by considering the median of each index per year to categorize them according to their specialization or diversification patterns. We create two discrete variables, S_{it} (specialization) and D_{it} (diversification), where S_{it} takes the value 1 if the RZI of city i is above the median in year t and 0 otherwise, and D_{it} does the same for values of RDI above the median in year t. Our empirical strategy lies on combining the two previous discrete variables. Specifically, we classify cities in four types resorting to the S_{it} and D_{it} variables in the following way:

Table 4.1: Categorization of Cities.

Type of City	*Diversified* (D_{it}=1)	*Non-Diversified* (D_{it}=0)
Specialized (S_{it}=1)	(1,1)	(1,0)
Non-Specialized (S_{it}=0)	(0,1)	(0,0)

As argued above, cities are not only specialized or diversified, but co-agglomeration economies may emerge. Table 4.1 shows these three type of cities: only diversified (0,1) cities; specialized (1,0) cities; but also (1,1) cities which are specialized and diversified both at the same time (co-agglomeration economies). Moreover, it accounts for one additional city that is not usually considered in the literature, the (0,0) cities or non-specified cities. These are cities that are not specified in terms of specialization and diversification. Thanks to the approach followed in Table 4.1, we can control for these four types of cities.

4.2.2 Indices for city size, agglomeration economies and cities' interactions

We measure city size in terms of its *population size* as well as its *firm size*. Bigger cities tend to harbor bigger firms (Glaeser et al., 2010),[3] and we define the average firm size in city i as:

$$Size_{it} = \frac{Number\_Establishments_{it}}{Number\_Workers_{it}} \tag{4.5}$$

In our analysis we control for different dimensions of agglomeration economies, including education level, natural advantages, sectoral composition and degree of product standardization. Education level reflects the fact that larger cities tend to attract relatively more high-skill workers (Glaeser et al., 2010). Ellison and Glaeser (1997) showed that industries locate where there are other industries to benefit from external economies of scale and natural advantages. In a further step, Ellison et al. (2010) argue that

[3]Along the text we use the term "size" to account for the scale effect of the city, basically *population* or the *average firm size*

natural advantages and the type of specialization are key drivers of co-agglomeration economies.

We define the *share of high educated people* in city i as:

$$Sh\_Educ_{it} = \frac{High\_Education\_People_{it}}{Population_{it}} \tag{4.6}$$

Natural advantages are defined as the total tonnes produced in the k sectors in relation to *mining* in city i:

$$Mining_i = \sum_k Tonnes\_Mining_i \tag{4.7}$$

The *type of specialization* is defined by the ratio of manufacturing workers over service workers, both in city i:

$$Ratio\_MS_i = \frac{\sum_k Manufacturing\_Workers_i}{\sum_k Services\_Workers_i} \tag{4.8}$$

Cities internal structure can be hard to change in time. To deal with the idea that cities structure can be affected by previous levels of specialization and diversification, we control for the *RZI* and *RDI* indices in 1995 (first year in our sample).

Diversified and specialized cities also differ in product standardization. Duranton and Puga (2001) argue that firms move to specialized cities from diversified ones when they reach such efficiency gains from internal scales of production that it allows them to produce standard products. Clark and Stanley (1999) propose that standardization can be measured by plant-level scale economies, such as the minimum efficient scale (MES). According to them, there exist a positive correlation between the MES and the product standardization; i.e., standardized products are the result of plant-scale economies whose costs decline as plant size increases. Following the proposed measures for MES at the industry level (Clark and Stanley, 1999; Cilasun and Günalp, 2012), we adapt them to the city level. We define *MES_i* as the average sales per firm (p) corresponding to the first P largest firms out of the total number of firms F located in city i, that account for at least 50% of city total sales:

$$MES_i{=}\frac{\sum_{p=1}^{P}(Sales\_per\_firm_i^p)}{P_i} \mid \left(\frac{\sum_{p=1}^{P} Sales\_per\_firm_i^p}{\sum_{f=1}^{F} Sales\_per\_firm_i^f}\right) \geqq 50\% \tag{4.9}$$

Next, we consider the role of interactions between cities as determinant of whether cities become diversified or specialized (Abdel-Rahman, 1996). We define interactions in terms of trade flows and transport costs. As regards trade flows, we measure them in three ways: *exports per capita* from city i; *number of varieties* exported from city i; and *trade flows with other type of cities* from city i. We account for exports per capita as a measure of exports intensity. Specifically, exports per capita ($Exports\_pc$) are considered as total inter-city trade flows departing from city i divided by its population.[4]

$$Exports\_pc_{it} = \frac{\sum_j Inter\_city\_trade_{ijt}}{Population_i} \ \forall i \neq j \tag{4.10}$$

Because diversified and specialized cities differ in the number of industries that they host (Duranton and Puga, 2000; Tabuchi and Thisse, 2011), we should expect that the number of different varieties exported changes accordingly.[5] Hence, we calculate the number of varieties z exported from city i to all possible destinations j as:

$$Number\_Varieties_{it} = \sum_j z_{ijt} \ \forall i \neq j \tag{4.11}$$

To account for the role of trading partners as potential drivers for diversification or specialization, we consider for each type of destination city c, the trade flows with it as:[6]

$$Trade^c_{it} = \sum_j Inter\_city\_trade^c_{ijt} \ \forall i \neq j \tag{4.12}$$

Focusing on transport costs, we calculate the relevance of each city within a system of cities considering all possible destinations departing from city i. In our database, transport costs are defined as the minimum economic cost (per shipment) of moving from one origin i to any other destination j. Thus, we define transport costs as the average of the minimum economic costs ($GTCs$) of moving from city i to all destinations j in year t in the system N such as:

$$\overline{GTC}_{it} = \frac{\sum_j GTC_{ijt}}{N} \tag{4.13}$$

Note that equation 4.13 also includes city i as possible destination to control for intra-city transport costs. For measuring trade cost, we rely on a monetary measure of transport costs. This variable corresponds to a Generalized Transport Cost (GTC) definition belonging to the least cost itinerary between an origin and a destination. The GTC is calculated using GIS software (Arc/GIS) with the digitalized Spanish road network, as presented in Zofío et al. (2014) and discussed by Díaz-Lanchas et al. (2013), in comparison with other transport cost proxies (distance and time). The GTCs differentiate

[4] Inter-city trade between i and j is defined as in Hillberry and Hummels (2008), that is, $T_{ij} = S_{ij}\overline{PQ}_{ij}$, where S_{ij} represents the total number of shipments between the origin i and the destination j, $\overline{PQ}_{ij}$ is the average value per shipment measured as the average price between i and j ($\overline{P}_{ij}$) multiplied by the average quantity between i and j ($\overline{Q}_{ij}$).

[5] We consider *varieties* as those goods with different classifications in our trade database.

[6] Different cities at the destination are indicated by c which takes the values 1 to 4. Along the text and to help the reader, we refer to 1 as the (0,0) cities, 2 for the (1,0) cities, 3 for (0,1) cities and finally, 4 for (1,1) cities.

the economic costs related to both distance and time in the following way:

$$GTC_{ijt} = min(DistCost_{ijt} + TimeCost_{ijt}) \tag{4.14}$$

The *distance economic cost* (euros per kilometer) includes the following variables: Fuel costs (fuel price); Toll costs (unit cost per km, multiplied by the length of the road); Accommodation and allowance costs; Tire costs; and Vehicle maintenance and repairing operating costs. By contrast, the *time economic cost* (euros per hour) includes the following variables: Labor costs (gross salaries); Financial costs associated to the amortization; Insurance costs; Taxes; Financing of the truck (assuming that it remains operative only for a certain number of hours/year); and indirect costs associated to other operating expenses including administration and commercial costs.

4.2.3 Econometric Specification

When discussing the emergence of different cities, we are referring to the probability of getting specialized (S_i) or diversified (D_i) cities. In an empirical setup, it hinges upon two independent probit models for city i, where 1 and 2 identify each equation (t subscripts have been removed):

$$S^*_{i1} = X_{i1}\beta_{i1} + \epsilon_{i1}, \quad S_{i1} = 1 \; if \quad S^*_{i1} > 0, \; 0 \; otherwise, \tag{4.15a}$$

$$D^*_{i2} = X_{i2}\beta_{i2} + \epsilon_{i2}, \quad D_{i2} = 1 \; if \quad D^*_{i2} > 0, \; 0 \; otherwise, \tag{4.15b}$$

$$\begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{pmatrix} \Big| X_1, X_2 \backsim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$

S^*_{i1} and D^*_{i2} indicate each type of city. The key point in this setting is ρ, the *tetrachoric correlation* between ϵ_{i1} and ϵ_{i2}. If $\rho = 0$, both expressions 4.15a and 4.15b are independent so we can simply estimate the probability of getting each type of cities by using two independent probit models. If $\rho \neq 0, \epsilon_{i1}$ and ϵ_{i2} are correlated, and the joint probability of getting each type of city must be estimated by resorting to a bivariate probit model (Greene, 2012). It is expressed as:

$$logL = \sum_{i=1}^{2} log\Phi_2 \begin{bmatrix} (2S_{i1}-1)\beta_1 X_{i1} \\ (2D_{i2}-1)\beta_2 X_{i2} \\ (2S_{i1}-1)(2D_{i2}-1)\rho \end{bmatrix} = \sum_{i=1}^{2} log\Phi_2[q_{i1}\beta_1 X_{i1}, q_{i2}\beta_2 X_{i2}, q_{i1}q_{i2}\rho] \tag{4.16}$$

Where *logL* refers to the log-likelihood function, $q_{i1} = (2S_{i1} - 1) = -1$ if $S^*_{i1} = 0$ and

$q_{i1} = 1$ if $S^*_{i1} = 1$; $q_{i2} = (2D_{i2} - 1) = -1)$ if $D^*_i = 0$ and $q_{i2} = 1$ if $D^*_i = 1$.

Now let, $w_{i1} = \rho_{i1} X_{i1} \beta_1$ and $w_{i2} = \rho_{i2} X_{i2} \beta_2$. Thus, the probabilities entering the previous log-likelihood function are:

$$Prob(S^*_{i1} = S_{i1}, D^*_{i2} = D_{i2} | X_1, X_2) = \Phi_2(w_{i1}, w_{i2}, q_{i1} q_{i2} \rho) \tag{4.17}$$

The possible outcomes in terms of probabilities vary as long as $\rho \neq 0$. To the extent that expressions 4.15a and 4.15b are dependent, equation 4.11 would estimate the probability of observing one city with the two patterns considered (specialization and diversification) at the same time. It is based on a 2x2 probability matrix that takes into account the four different probabilities that can be obtained in terms of specialization and diversification. These probabilities exactly correspond to the four different cities previously presented in Table 4.1, that now are reproduced in Table 4.2 in terms of probabilities.

Table 4.2: Probability for each type of city.

Probability of each city	D_{it}=1	D_{it}=0
S_{it}=1	$P(1,1)$	$P(1,0)$
S_{it}=0	$P(0,1)$	$P(0,0)$

To estimate these probabilities, we take logarithms in 4.15a and 4.15b and include time (γ_i) and spatial fixed effects (μ_i),[7] obtaining the final expressions to be estimated:

$$S^*_{i1t} = \alpha_{1t} + lnX_{i1}\beta_1 + \mu_{i1} + \gamma_{t1} + \epsilon_{i1t} \tag{4.18}$$

$$D^*_{i2t} = \alpha_{2t} + lnX_{i2}\beta_2 + \mu_{i2} + \gamma_{t2} + \epsilon_{i2t} \tag{4.19}$$

According to the literature, cities can be diversified or specialized depending almost on the same mechanisms. If that is the case, our assumption is that the regressors for the probability of being a diversified city (X_{i1}) should not be independent from being a specialized one (X_{i2}); in other words, almost the same regressors are determining the joint probabilities in Table 4.2. It translates into two X-vectors in the following way:[8]

[7]Spatial effects are included at the NUTS-2 level (*Autonomous Communities*). Low degrees of freedom force us to consider them at this level instead of as city fixed effects. Moreover, the use of NUTS-2 spatial effects makes sense in the way administrative levels are organized in Spain. They comprise NUTS-3 and urban levels, with NUTS-2 representing fiscal-autonomous regions with remarkable differences across them in terms of infrastructure, and regional and tax policy. A NUTS-3 (provincial) level would be inadequate as all of them are captured in any of the urban areas considered.

[8]Note that in each vector, X_1 (for specialization) and X_2(for diversification), trade with the same type of city is not included. From the theory we do not have a clear identification of the expected results when cities trade with other cities with the same typology. Co-agglomeration economies and non-specified cities are, by contrast, included in an attempt to let the data speak to get new empirical evidence for future theoretical research. Also, by not including trade with the same type of city, we avoid a collinear problem with the other trade variables.

$$X_{i1} = (Size_i, \overline{GTC}_i, Sh\_Educ_i, Ratio\_MS_i, MES_i, Mining_i, Number\_Varieties_i, Exports\_pc_i, Trade_i^1, Trade_i^3, Trade_i^4, RZI1995_i) \quad (4.20)$$

$$X_{i2} = (Size_i, \overline{GTC}_i, Sh\_Educ_i, Ratio\_MS_i, MES_i, Mining_i, Number\_Varieties_i, Exports\_pc_i, Trade_i^1, Trade_i^2, Trade_i^4, RDI1995_i) \quad (4.21)$$

4.3 Data

This section presents the micro-databases used in this study for the period 1995-2007. These are: A firm-level database on industry characteristics; the considered classification of the Spanish urban areas and their population; an origin-destination database on the economic transport cost of moving from one city to the others (GTCs); and a database on inter-city trade flows (shipments) between Spanish cities.

4.3.1 Service and Manufacturing Data

To calculate the diversification and specialization of cities, we rely on an unique firm-level database on services and manufacturing at the municipality level called the *SABI database*. This database, produced by Bureau van Dijk according to the information registered in the Spanish Registry of Commerce (*Registro Mercantil*), provides economic and financial information for about 1,3 million Spanish firms for the period 1995-2007. SABI identifies the municipality in which the firm is located, the sector in which the firm operates and its number of workers. With it, we can obtain reliable information on agricultural, manufacturing and service sectors at the municipality level that, in a further step, we aggregate at the urban area level. To avoid outliers at the sectoral level, we aggregate the employment information up to 38 different sectors according to the NACE classification. It includes all relevant sectors such as agriculture, manufacturing production, banking activities, services and activities related to the public sector.

The SABI database is similar to that used by Henderson (2003) for measuring Marshallian economies. To give support for the use of the SABI database as a source of employment at the metropolitan level, we include in the Appendix 4.A a set of tables summarizing the representativeness of this database with respect to two alternative databases from the Spanish Statistical Institute and the Spanish Ministry of Labour.[9] These

[9] Apart from the *RZI* and *RDI* indexes, we use the SABI database to calculate the *MES* and the

two official database do not provide information at the urban level, but at a higher regional level (NUTS-3 level). Due to this restriction we only use the official databases to assess SABI's reliability and check its representativeness by aggregating its data according to the following dimensions: the *national level* by year and size of the firm; the *sectorial level* (38 NACE Classification), and the *regional level* (Provinces, NUTS-3 level). As it can be observed, its representativeness increases as it includes 73% of the official labour force by 2007. It covers very accurately the employment of medium and big firms, as official registration is compulsory for these firms but not for small firms (less than 50 employees), and the employment in the main sectors of the Spanish economy. Also, the main provinces are well captured even if with a slight bias in favour of the richest Spanish provinces (Madrid, Barcelona and Bizkaia), as firms' HQs are mainly localized in these regions.

Despite apparently opposite, specialization and diversification are not always contrary dimensions of cities's performance. While the former focuses on whether a sector is more concentrated in one city in comparison to the national level, the latter attends to the number of different sectors in the city. Indeed, we can find cases in which specialization and diversification do not play a significant role, or even they coexist, within the same city.[10] Figure 4.1 plots the RDI and RZI indexes for 1995 and 2007 at the city level. As it can be observed, there are cities that are very specialized (high RZI) or diversified (high RDI), but there are cases in which both characteristics are very small, or cities with high values in both indexes. Even more, it is remarkable the shift of the Spanish cities toward the diversification dimension, as almost all cities have improved their level of diversification in 2007 in comparison to 1995.

4.3.2 Urban Areas

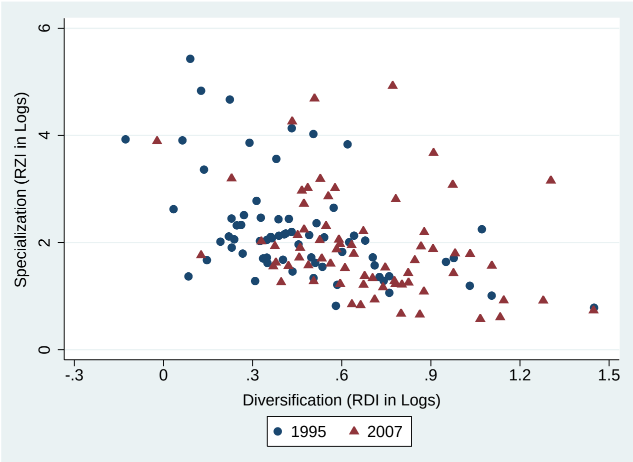
To define cities we use the functional urban areas identified by the OECD for Spain. The OECD follows a three step approach to define functional urban areas (Brezzi et al., 2012). First, they identify densely populated municipalities. Second, they aggregate densely populated municipalities into an urban area if more than 15% of the population of one municipality commutes to work in the most dense populated municipality. Finally, municipalities that have a low population density are assigned to an urban area if at least 15% of their employed population work in that urban area.

In this study we consider 69 Spanish Urban Areas according to the OECD classification.[11] They account for 46% of the Spanish population on average for the period

ratio between manufacturing and service workers (*Ratio_MS*), both at the city level.

[10]Indeed, the (Pearson) correlation coefficient between the RZI and RDI for the whole sample 1995-2007, takes only the value of -0.26. That is, these two measures present an inverse relation but the are are far from being completely opposite.

[11]Concretely, the OECD calculates 76 Urban Areas for Spain, but we have only considered those for



Source: Own elaboration from GTC database.

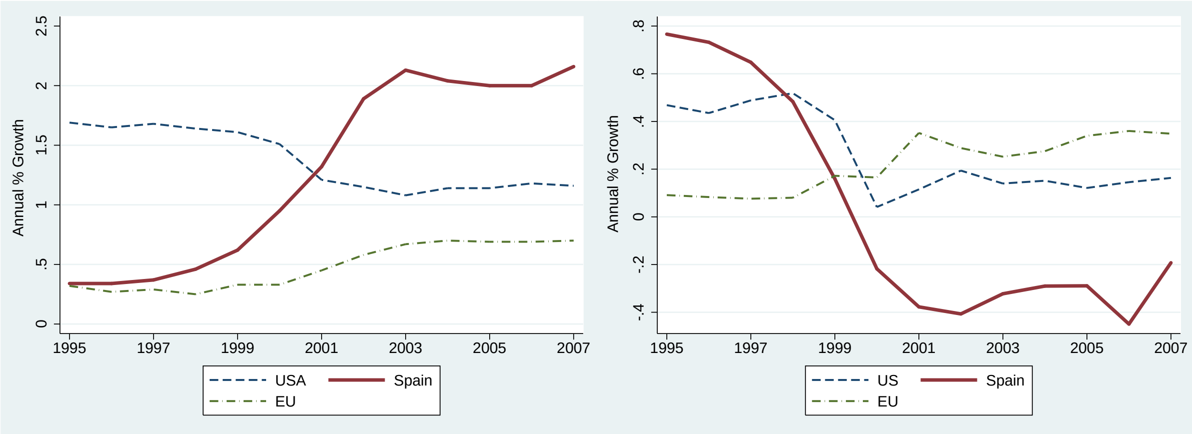
Figure 4.1: Diversification and Specialization by Urban Area (1995 and 2007).

1995-2007. Spain has 2 large metropolitan areas: Barcelona and Madrid (with a population of 1.5 million of more), 6 metropolitan areas (with a population between 500,000 and 1.5 million), 22 medium-size urban areas (with a population between 200,000 and 500,000) and 39 small urban areas (with a population below 200,000 people). The larger metropolitan areas, and most of the medium-size ones, are located around the provinces' capitals (NUTS-3). Data on urban area population and education levels are obtained from the Spanish Statistical Institute (*INE*) Census.

Finally, to measure mining activities at the urban level we resort to a database on tonnes produced in mining-related activities at the municipality level. This database comes from the Spanish Geological Survey (*Instituto Geológico y Minero de España*) that provides data on the number of effective tonnes produced in raw materials such as sand, marble, clay and limestone. Hence, we aggregate tonnes from all raw materials to the urban area level. This database is only available for 2005.

During 1995-2007 Spain proves to be a dynamic country in terms of population. Figure 4.2 plots the growth rates for the overall urban population (a) and the urban population of the biggest cities (b) in comparison to those in US and the European Union. Spain shows a dichotomic pattern in relation to the other economic areas as urban growth, as a whole, experimented a huge increase from 1995 to 2002, whereas the share of population located in the big cities steadily decrease until being stagnant since 2001.

which there exist employment data at the municipality level according to the SABI database. Indeed, the remaining 7 Urban Areas only represent either urban areas located in the Spanish islands or those with little relevance in terms of population.



(a) Urban Population Growth

(b) % of Total population in urban agglomerations of more than 1 million (Annual growth)

Source: Own elaboration from World Bank database.

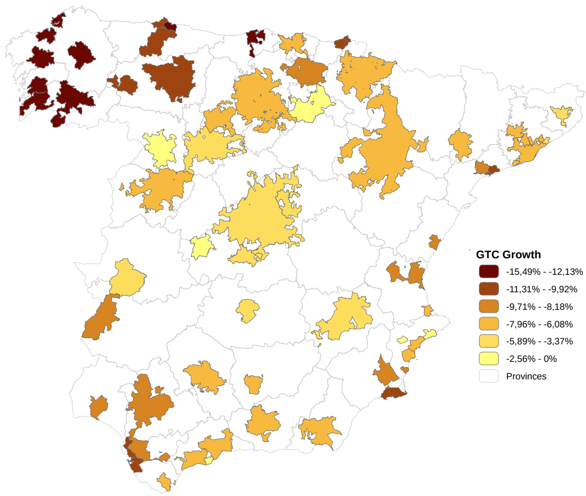
Figure 4.2: Population patterns between 1995 and 2007.

4.3.3 The Generalized Transport Cost (GTC)

Our monetary measure of Generalized Transport Cost (GTC) is calculated considering prices at the provincial level (NUTS-3 level), specifically those observed at the province where the shipments originates. Due to all these economic costs, the minimum GTCs route calculated with Arc/GIS could differ from the respective measures of distance and travel time. Indeed, its main advantage over other proxies for transport costs (i.e. physical distance and travel time) is that the GTC is a time-varying trade cost. As it is composed by prices (mainly fuel prices, salaries and tolls) the GTC changes annually, so we can get a precise economic measure for transport costs that varies along the years even though the distance components of the GTC are almost time-invariant proxies for transport costs.

As mentioned, the GTC is calculated on a digitalized road network. This network is available every five years from 1980 to 2005 and 2007. Given the annual availability of data in the SABI and trade databases covering the years 1995-2007, we only consider the road network from 1995 onwards and linearly interpolate the GTC database. Finally, in the empirical part, we select the years 1980, 1985 and 1990 to check the robustness of our results and to analyze the existence of path dependence patterns.

Figure 4.3 represents the reduction rates of GTCs per urban area between 1995-2007 using equation 4.13. Cities further located from the Spanish center (Madrid), specially those in the north-east part of the country, are those that experience the highest reductions in transport costs. Indeed, the biggest cities have not seen a remarkable reduction in transport costs, while smaller cities increased their accessibility during this



Source: Own elaboration from GTC database.

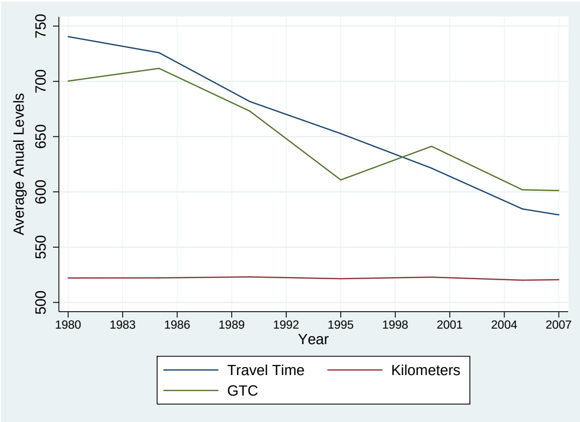
Figure 4.3: GTC reduction rates by Urban Area (1995-2007).

period. This pattern is consistent with the core-periphery structure obtained by previous studies for the Spanish case (Zofío et al., 2014).

The GTC decreases by -16,30% between 1980-2007.[12] Prior to 1980 the road network in Spain was underdeveloped. For instance, the number of highway-kilometers (high-speed roads) in 1970 was 185 km, in total of 139,406 km (IETC, 1979),[13] barely increasing to 335 km in 1980, but reaching 9,557 km by 2007 (Zofío et al., 2014). Figure 4.4 plots from 1980 (the first available year for transport costs) to 2007, the trend in the average value per year of moving from each city to another (equation 4.13) resorting to the three main transport costs: the GTC (in euros); the physical distance (in kilometers); and the travel time (in minutes). Urban areas do not experience remarkable declines in accessibility through physical distance, even after the huge investment in infrastructure during this period, whereas the GTC mainly decreased as a result of the reductions in travel time. By contrast, as the GTC entails prices from petrol, tolls and salaries, it shows increments between 1980-2007, not seen in the travel time measure. For this reason, we contend that the GTC is the only reliable and true measure of transport cots as it is the only one that can accurately capture these dynamics within a panel database.

[12]Based on the economic approach to index numbers, (Zofío et al., 2014) decomposes this reduction into an economic (price) component (EC) and an infrastructure (quantity) component (IC). From the overall -16.30% reduction in the GTCs between 1980-2007, price reductions EC contribute -6.93% while infrastructure improvements IC contribute the remaining -10.04%. Dividing the period into the two sub-periods 1980-1995 and 1995-2007, we can find that, between 1980-1995 the GTC reduces -11.04%, having a reduction of -6.37% in its economic component and -4.99% in the infrastructure component. For the period 1995-2007, the GTC shows a reduction of -5.26%. This reduction comes from a -0.59% in the EC and -5.05% in IC.

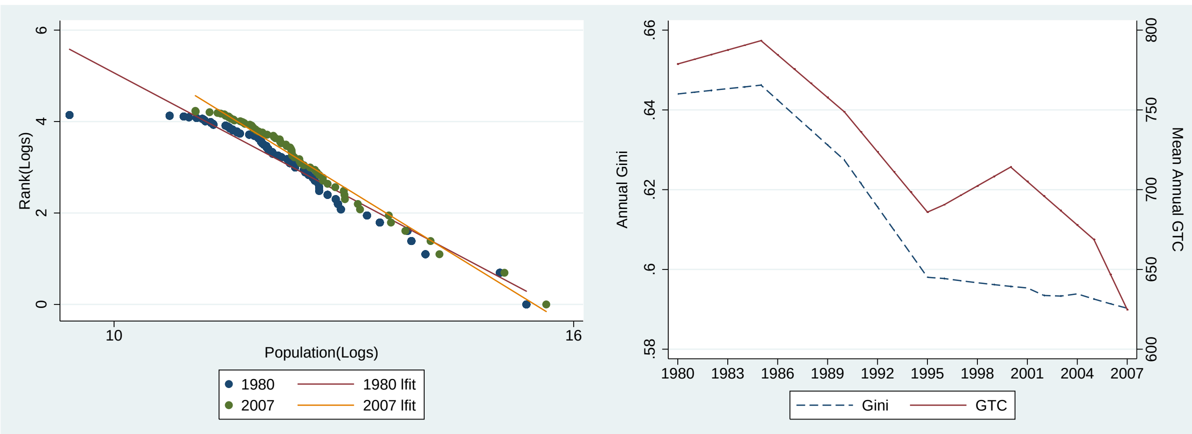
[13]To give a comparison, by that time France had 1,542 km of high-speed roads over a total of 1,679,549 km, and Italy had 3,913 km over 285,318 km (IETC, 1979).



Source: Own elaboration from GTC database.

Figure 4.4: Average Annual Transport Cost Changes (1980-2007).

These changes in transport costs come hand-in-hand with structural changes in the Spanish urban population, even deeper than those evidenced by Figure 4.2. Figure 4.5 plots for 1980 and 2007 the Zip's Law (a), i.e. the relation between the (log) rank of the city and its (log) population, and the distribution of cities population (b) measured through the Gini coefficient for population. Whereas transport costs remarkably decrease, Spanish cities present a more equal distribution of population in those 30 years.[14] This joint pattern between the GTC and the Gini coefficient tells that the reduction in transport costs has not been followed by a more unequal distribution of population across cities, as proposed by Tabuchi and Thisse (2011).



(a) Zip's Law 1980 and 2007 (b) Gini(Population) and GTC (1980-2007)

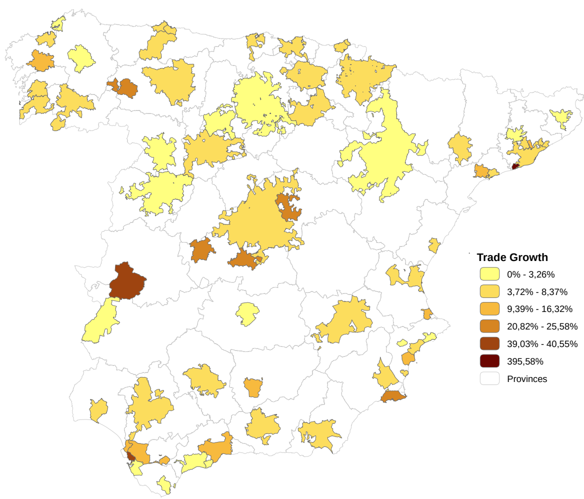
Figure 4.5: Population patterns between 1980 and 2007.

[14]The more equal distribution of the Spanish cities is also confirmed by running simple OLS regressions between the (log) rank of the city and its (log) population. These regressions take the form: $lnRank_{it} = \alpha_i + \beta_i lnPop_{it} + \epsilon_{it}$, where t=1980, 2007. The β coefficient equals 0.88 for 1980 and 1.03 for 2007.

4.3.4 Trade value data: The road freight transportation survey

To account for trade between cities we rely on a micro-database on shipments by road within Spain in the period 1995-2007 elaborated within the research project C-intereg. As explained in Díaz-Lanchas et al. (2013), this database is based on the annual Road Freight Transportation Survey (RFTS) compiled by the transport division of the Ministry of Public Works (*Ministerio de Fomento*), which randomly surveys a sample of freight companies and independent truckers, with vehicles over 3.5 tons operating within the national territory.[15] It includes information about the characteristics of the vehicle and shipments such as the number of tons transported by the truck, the number of shipments between the origin i and the destination j, the type of product,[16] the operations performed by the truck in each shipment, as well as the actual travel distance in kilometers between the geographical origin and the destination of each shipment. With this database we can calculate a very precise measure of the volume of trade developed by each urban area when interacting with other.

In Figure 4.6 we present the inter-city growth rates between 1995-2007. Middle cities (Guadalajara and Toledo) surrounding big cities (Madrid) have experienced significant increases in inter-city trade flows. Moreover, cities in the north-east exhibiting important reductions in transport costs (Figure 4.3) are those with higher increments in trade flows.



Source: Own elaboration from RFTS database.

Figure 4.6: Inter-city trade growth rates by Urban Area (1995-2007).

[15]This database accounts for almost the 85% of all Spanish interregional trade flows; the remaining 15% corresponding to rail, maritime and air modes.

[16]In the database, commodities (varieties) are recorded attending to the "NST-R Classification" compiled by Eurostat which differentiates between 180 products.

4.4 Empirics

4.4.1 Descriptive Analysis

In this section we describe the main features of the urban system in Spain, as regards to the evolution of specialization and diversification patterns over time. Table 4.3 summarizes the characteristics of the four types of cities that we consider by resorting to the mean and the standard deviation between 1995-2007 for each one of the indexes introduced in section 4.2.1. The value 1 represents non-specified cities, 2 and 3 refer only to specialized or diversified cities, respectively, while 4 is used for co-agglomeration economies. Finally, total values for the whole sample are also included.

As can be observed, non-specified cities (1) are mainly small-medium cities in terms of population and trade per capita. They export very few varieties and show middle levels of educated population and transport costs, as well as the lowest firm size. Specialized cities (2) are the smallest in population but carry out the highest amount of trade flows per capita. Both their share of educated people and the number of different varieties exported are very low. By contrast and expectedly, diversified cities are the most populated cities with the lowest trade intensity and accessibility in terms of transport costs. They have the highest educated population and the biggest firms. Co-agglomeration economies correspond to medium-big cities with middle-high levels of trade flows, but the worst accessibility through transport costs. These cities also host medium-big firms and important levels of educated people. Finally, note that specialized cities and non-specified cities show the highest ratio of manufacturing workers, while the MES shows important levels either in specialized and diversified cities.

Next, maps (a) and (b) in Figure 4.7 visualize cities according to their categorization. In 1995, diversified cities (0,1) were mainly province capitals (Sevilla, Albacete, Alicante or Badajoz, between others) or the richest cities in Spain (Madrid, Barcelona, Bizkaia and Valencia). These cities were located in the south-east and center of the country. Co-agglomeration economies (1,1) were cities further located from the center (Lugo, Pontevedra, Ciudad Real) or very small cities surrounding diversified ones (Toledo, Guadalajara), with the exception of Zaragoza which is one of the most populated cities in Spain. Specialized cities (1,0) were located in the north-west part of the country but close to cities with some sort of diversification pattern. Meanwhile, non-specified cities (0,0) were spread along the geography but located near to diversified cities.

By 2007, many cities had evolved into a diversified structure. Cities in the north-west part of the country that were only non-specified or specialized cities, became either diversified (Valladolid, Salamanca, Pontevedra, Burgos, Álava, between others) or presented co-agglomeration economies (León, Oviedo, La Coruña). Indeed, these cities were the

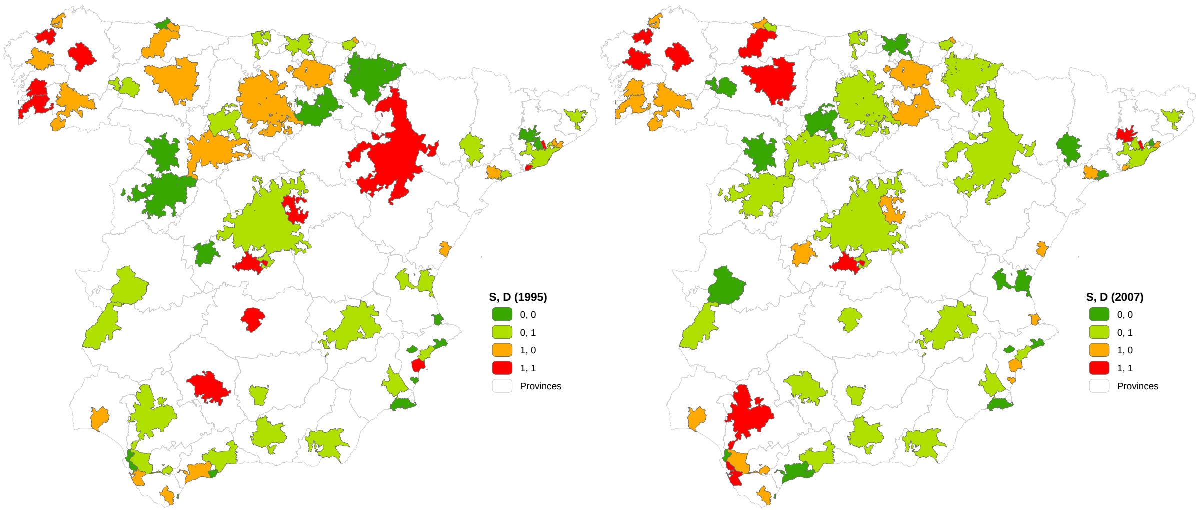
Table 4.3: Descriptive statistics.

Type of City	*Stats*	*RDI* Index	*RZI* Index	*Population* (Thousands)	*GTC* (Mean)	*City Exports* Per Capita (Logs)
1	*Mean*	1.49	5.00	163.80	651.94	6.06
	Std	0.20	1.34	222.31	162.16	1.00
2	*Mean*	1.45	24.58	125.12	699.67	6.03
	Std	0.25	34.82	85.89	172.69	0.92
3	*Mean*	2.33	3.78	575.05	639.24	5.30
	Std	0.63	1.34	949.34	167.72	0.82
4	*Mean*	2.11	20.83	196.78	715.08	5.90
	Std	0.49	25.61	182.11	195.00	0.91
Total	*Mean*	1.88	13.29	301.46	673.11	5.77
	Std	0.60	23.84	607.09	174.94	0.96

Type of City	*Stats*	*Number of Varieties*	*High Education* (% of People)	*Firm Size* (Av. Employees)	*MES* (Logs)	*Ratio* M/S
1	*Mean*	31.27	8.55	26.83	14.88	0.55
	Std	19.84	3.13	32.01	0.51	0.50
2	*Mean*	31.12	7.88	30.41	14.96	0.60
	Std	15.54	2.71	40.38	0.59	0.54
3	*Mean*	46.03	10.3	81.18	14.97	0.34
	Std	24.60	2.88	127.32	0.53	0.19
4	*Mean*	39.14	9.64	45.83	14.91	0.39
	Std	18.08	3.12	52.40	0.33	0.22
Total	*Mean*	37.39	9.08	49.10	14.94	0.47
	Std	21.12	3.08	84.06	0.52	0.41

Note:(1): S=0, D=0 ; (2): S=1, D=0 ; (3): S=0, D=1; (4): S=1, D=1.

ones that experimented the highest reduction in transport costs as shown in Figure 4.3. In this line, large and diversified cities such as Seville transitioned into co-agglomerated economies but, at the same time, the opposite trend was observed (even if to a lesser extent), as some cities like Zaragoza (a co-agglomeration economy in 1995) lost their specialization pattern to become diversified only. Cities near the biggest ones, also changed their patterns by becoming only specialized cities (Guadalajara), or shifted to a diversified economic structure as Pamplona, Cádiz, and Manresa.



(a) 1995 (b) 2007

Figure 4.7: Four type of cities in Spain for 1995 and 2007.

As the previous maps reflect the existence of all kind of dynamic pattern between cities, Table 4.4 shows the transition matrix for the four different types of cities between the years 1995 and 2007. The absolute numbers represent the number of cities that change their structure in 2007 in comparison to 1995, while the percentages report the same information in relative terms over the total number of cities. Cities that are only specialized or diversified are the most prominent and stable cities (main diagonal). By contrast, the most interesting features from Table 4.4 come from the patterns off the diagonal. For the four type of cities, the most probable result is the change in their status to one of the other scenarios. Non-specified cities are more likely to become either specialized or diversified. Specialized cities tend to diversified their economic structure to become diversified or co-agglomerated economies. Although diversified cities are the most stable cities, when compared to the other cities they are more capable of losing their diversification in favor of non-specified cities or specialized cities. Co-agglomeration economies are very likely to change their status in any of the two dimensions to become either diversified or specialized cities. Within the 69 cities considered, only 31 (44%) maintain their original status from 1995. Table 4.4 indicates that changing cities' structures is more the rule than the exception.

Table 4.4: Transition Matrix between type of cities (1995-2007).

	City (2007)				
City (1995)	(1)	(2)	(3)	(4)	**Total**
(1)	2	5	4	2	13
Share (%)	*15.38*	*38.46*	*30.77*	*15.38*	*100*
(2)	3	12	3	4	22
Share (%)	*13.64*	*54.55*	*13.64*	*18.18*	*100*
(3)	7	1	13	1	22
Share (%)	*31.82*	*4.55*	*59.09*	*4.55*	*100*
(4)	0	5	3	4	12
Share (%)	*0*	*41.67*	*25*	*33.33*	*100*
Total	12	23	23	11	69
%	*17.39*	*33.33*	*33.33*	*15.94*	*100*

Note:(1): S=0, D=0 ; (2): S=1, D=0 ; (3): S=0, D=1; (4): S=1, D=1.

4.4.2 Results: Baseline model

Once the main insights on the cities' typology have been presented, we proceed to the multivariate analysis to determine its magnitude and statistical significance. Table 4.5 presents the estimated coefficients for the two probit models using the specifications in equations 4.18 (for specialization) and 4.19 (for diversification). $Population_i$ and Sh_Educ_i control for agglomeration economies and the size of the city. According to the theory these factors should have a higher impact on diversified cities. $Ratio\_MS_i$, MES_i, $Mining_i$, reflect the type of specialization of the city. Following Duranton and Puga (2000) specialized cities are more likely to host workers within the manufacturing sector while diversified cities tend to attract services-related workers, so a positive sign for specialized cities should arise.[17] MES_i, as the measure for product standardization, is expected to have a positive sign in specialized cities (Duranton and Puga, 2001). Finally, $Mining_i$, as proxy for natural advantages, should be higher in specialized cities as firms depending on natural resources are more likely to move to these areas (Ellison and Glaeser, 1997).[18]

Table 4.5 summarizes results as follows: columns (1) to (3) indicate the probability of being specialized, while columns (4) to (6) do the same for diversified. Also, columns (1) and (4) only consider agglomeration economies and transport costs variables, while columns (2) and (5) add the trade variables, and columns (3) and (6) represent our benchmark including the level of specialization (column (3)) and diversification (column

[17] Viladecans-Marsal (2004) gives evidence on the effect that agglomeration economies have on localization economies for the Spanish case.

[18] Previous works for the Spanish case show evidence on the positive relationship between the use of mining activities and the level of specialization of the region (Betran, 2012).

(6)) in 1995. Table 4.5 includes the results for spatial and time fixed effects using the population as a proxy for city size, while in the Appendix 4.B we include robustness checks with the average firm size.

Results show that population, the share of educated people and the MES have the expected effects on the probabilities of classifying these two type of cities. The number of varieties and inter-city exports per capita are crucial for specialized ones. By contrast, trade with other type of cities is mainly important for diversified cities but not for specialized ones. The type of specialization through natural advantages (*Mining*) and the ratio of manufacturing workers (*Ratio_MS*) also show the expected positive signs for specialized cities. Also, it is worth highlighting the effect that transport costs (*GTC*) have on diversified cities. Surprisingly, GTCs are highly positive and significant in all specifications. The GTC positive impact on diversified cities confirms the hypothesis proposed by Abdel-Rahman (1996); Anas and Xiong (2005) by which high transport costs lead to (inefficient) diversified cities. Previous levels of specialization (RZI1995) and diversification (RDI1995) seem to play a role for both cities; that is, they are dependent of past levels of specialization and diversification. These final results are in line with previous evidence for the Spanish case (Viladecans-Marsal, 2002).

Table 4.5 shows also that the emergence of specialized cities heavily depends on city size, trade intensity, the type of specialization and the composition of the labor force. Meanwhile the existence of diversified cities is basically affected by city size, transport costs and the amount of trade with other cities (specialized and co-agglomeration economies). Finally, note that the ρ coefficients are negative and significant in all the specifications, indicating that expressions in 4.15a and 4.15b are interdependent and have an inverse relation. The ρ coefficient motivates the estimation of the bivariate probit model rather of using two independent probit models. At the same time, as the relation is negative for almost the same regressors, our regressions support the theoretical result by Abdel-Rahman (1996); Anas and Xiong (2003, 2005) as an approach to explain puzzle corresponding to the opposite, and simultaneous, effects that the same regressors have on the emergence of cities.

Table 4.6 presents results accounting for the joint average marginal effects obtained from the bivariate procedure using our benchmark in columns (3) and (6) from Table 4.5, and classified in accordance to the four possible probability outcomes detailed in Table 4.2. Subscripts indicate the type of city where the first number refers to the specialization, and the second number, to the diversification dimension.

Leaving the discussion for the next section, we now remark that Table 4.6 disentangles the previous results obtained for specialized (P_{10}) and diversified (P_{01}) cities. Indeed, it shows that both types of cities heavily depend on their size, although the effects of transport costs seem to vanish in favour of non-specified (P_{00}) and co-agglomeration

Table 4.5: Regression Results from Bivariate Probit Model (Baseline Model, 1995-2007).

	Pr(S=1| X)			Pr(D=1| X)		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\text{Pop}_i)$	-0.240***	-0.450**	-0.128	0.836***	0.947***	0.908***
	(0.089)	(0.185)	(0.219)	(0.123)	(0.222)	(0.222)
$\ln\overline{GTC_i}$	0.988**	1.606***	0.975	1.858***	1.755***	1.710***
	(0.480)	(0.557)	(0.662)	(0.565)	(0.629)	(0.619)
$\ln(\%\text{ High Education}_i)$	-0.939***	-1.455***	-1.092**	3.369***	3.893***	3.889***
	(0.301)	(0.394)	(0.463)	(0.482)	(0.562)	(0.556)
$\ln(\text{Ratio\_MS}_i)$	0.411***	0.556***	0.587***	0.132	0.170	0.194
	(0.086)	(0.123)	(0.137)	(0.096)	(0.129)	(0.132)
$\ln(\text{MES}_i)$	0.319**	0.602***	0.639***	-0.692***	-0.842***	-0.859***
	(0.140)	(0.166)	(0.169)	(0.158)	(0.196)	(0.198)
$\ln(\text{Mining}_i)$	-0.070	0.048	0.221**	0.120	0.140	0.079
	(0.063)	(0.080)	(0.087)	(0.078)	(0.088)	(0.092)
Number of Varieties$_i$		-0.166	-0.488*		-0.745**	-0.737**
		(0.256)	(0.291)		(0.311)	(0.318)
$\ln(\text{Intercity\_Tradepc}_i)$		0.232	0.438**		-0.332**	-0.317**
		(0.161)	(0.172)		(0.153)	(0.155)
lnTrade with (0,0)$_i$		0.021	-0.037		0.053	0.062
		(0.044)	(0.050)		(0.045)	(0.045)
lnTrade with (1,0)$_i$		—	—		0.138**	0.120*
					(0.062)	(0.064)
lnTrade with (0,1)$_i$		0.099	-0.069		—	—
		(0.105)	(0.116)			
lnTrade with (1,1)$_i$		-0.034	-0.016		0.121**	0.118*
		(0.051)	(0.057)		(0.062)	(0.062)
lnRZI1995$_i$			0.912***			—
			(0.103)			
lnRDI1995$_i$			—			0.545**
			—			(0.269)
Spatial FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	842	723	722	842	723	722
ρ	-0.312***	-0.301***	-0.343***	-0.312***	-0.301***	-0.343***

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

Table 4.6: Marginal Effects from Bivariate-Probit Model (Baseline Model, 1995-2007).

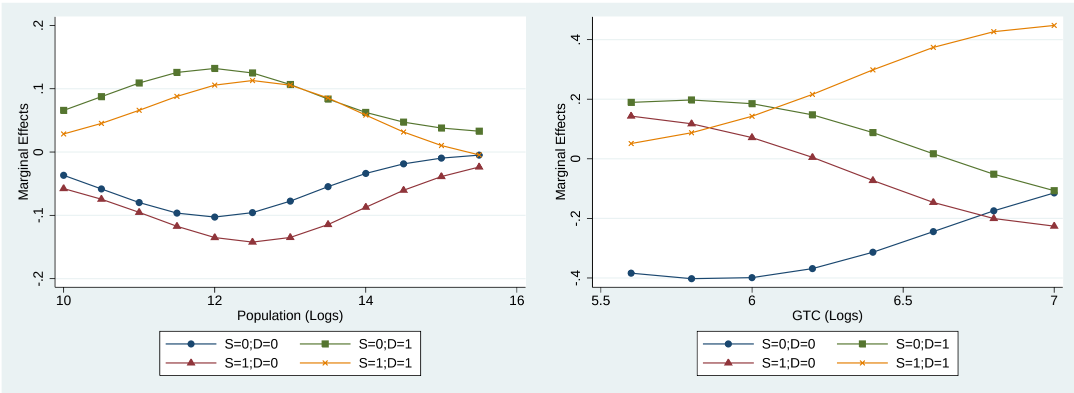
	Average Marginal Effects			
Variables	(P_{00})	(P_{10})	(P_{01})	(P_{11})
ln(Pop$_i$)	-0.151***	-0.183***	0.202***	0.132**
	(0.048)	(0.054)	(0.074)	(0.066)
ln$\overline{GTC}_i$	-0.456***	-0.174	0.071	0.558***
	(0.136)	(0.164)	(0.225)	(0.178)
ln(% High Education$_i$)	-0.572***	-0.859***	1.003***	0.429***
	(0.121)	(0.141)	(0.177)	(0.146)
ln(Ratio_MS$_i$)	-0.118***	0.047	-0.113**	0.184***
	(0.029)	(0.034)	(0.047)	(0.038)
ln(MES$_i$)	0.071*	0.246***	-0.323***	0.006
	(0.040)	(0.050)	(0.062)	(0.051)
ln(Mining$_i$)	-0.046**	0.017	-0.042	0.071***
	(0.020)	(0.022)	(0.030)	(0.026)
Number of Varieties$_i$	0.206***	0.066	-0.014	-0.258***
	(0.067)	(0.077)	(0.103)	(0.087)
ln(Intercity_Tradepc$_i$)	-0.002	0.119***	-0.171***	0.054
	(0.036)	(0.039)	(0.056)	(0.049)
lnTrade with (0,0)$_i$	-0.006	-0.016	0.021	0.002
	(0.010)	(0.011)	(0.016)	(0.014)
lnTrade with (1,0)$_i$	-0.022*	-0.022*	0.022*	0.022*
	(0.012)	(0.012)	(0.012)	(0.012)
lnTrade with (0,1)$_i$	0.010	-0.010	0.018	-0.018
	(0.016)	(0.016)	(0.029)	(0.029)
lnTrade with (1,1)$_i$	-0.020	-0.024	0.026	0.017
	(0.013)	(0.015)	(0.020)	(0.017)
ln(RZI1995$_i$)	-0.128***	0.128***	-0.232***	0.232***
	(0.017)	(0.017)	(0.028)	(0.028)
ln(RDI1995$_i$)	-0.102**	-0.099**	0.102**	0.099**
	(0.050)	(0.050)	(0.050)	(0.050)
Spatial FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	717	717	717	717

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

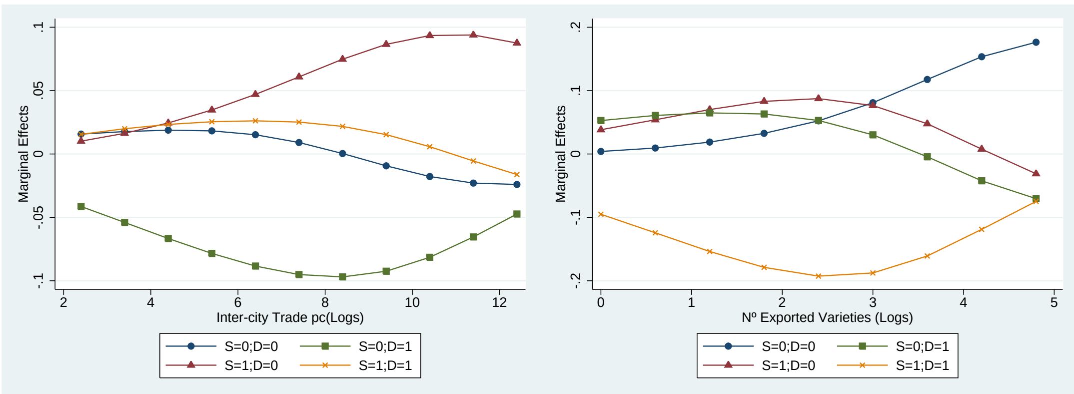
economies (P_{11}). Even so, specialized cities are very affected by natural advantages, but also by the type of specialization and products that they produce (MES). Trade, although relevant, is not significant when specialized cities trade with diversified ones. This last type of city follows the same pattern than specialized cities but in the opposite direction. They positively depend on city size and labor composition but trade is not that crucial in driving these cities. Focusing on non-specified cities (P_{00}), they show negative and significant signs in almost all the variables. The more they trade with other cities, the less likely for these cities to appear. Finally, co-agglomeration economies (P_{11}) follow a peculiar pattern in comparison to the other cities. They show a mix behavior between specialized and diversified cities as they positively depend on city size, type of specialization and labor composition, but at the same time, natural advantages and trade with other cities play a role in the emergence of these cities. Finally and as evidenced in Figure 4.3, GTCs are highly positive and significant for these cities as co-agglomeration economies are those further located from other cities or the center of Spain.

Although Table 4.6 reports average marginal effects, the bivariate estimation allows to capture non-linear marginal effects of each independent variable over each city type. Further exploring these results, Figures 4.8 and 4.9 split previous findings by showing the non-linear marginal effects of classifying each city as a function of the main variables independently. Figure 4.8 focuses on city size, GTC and trade variables. As can be observed, diversified cities and co-agglomeration economies are more likely to appear the higher their *city size* (a), although for very high levels of population, only diversified cities tend to emerge. At a lower levels of *GTC* (b) diversified and specialized cities arise, but rapidly they disappear in favour of co-agglomeration economies; i.e., the higher the transport costs, the more likely to observe co-agglomeration economies but the less likely to get diversified and specialized cities. These non-linear trends allows us to disentangle the non-significant marginal effect obtained in Table 4.6 for (P_{10}) and (P_{01}), but not for (P_{11}). For *inter-city trade* (c), only specialized cities are more probable to appear the more they intensively trade, except for diversified cities whose probability to emerge is even more negative the more they trade. Finally and attending to *exported varieties* (d), only non-specified cities follow an upward trend the more varieties they trade, while specialized and diversified cities rapidly disappear (they get zero or negative impacts). We hypothesize that this pattern exhibited by non-specified cities responds to cities that do not produce enough for themselves to satisfied locally demanded goods so they continuously need exports from other cities. Although the theoretical literature for this type of cities is almost non-existent, we find similarities between them and the propositions made by Tabuchi and Thisse (2011) for small cities to emerge.[19]

[19]In the Appendix 4.B, Figure 4.11 plots the rest of the variables from the baseline model. In short, diversified cities tend to emerge when they host more educated people. Specialized cities prevail the more manufacturing workers they host and the more standard products they produce. Co-agglomeration



(a) City Size (b) GTC



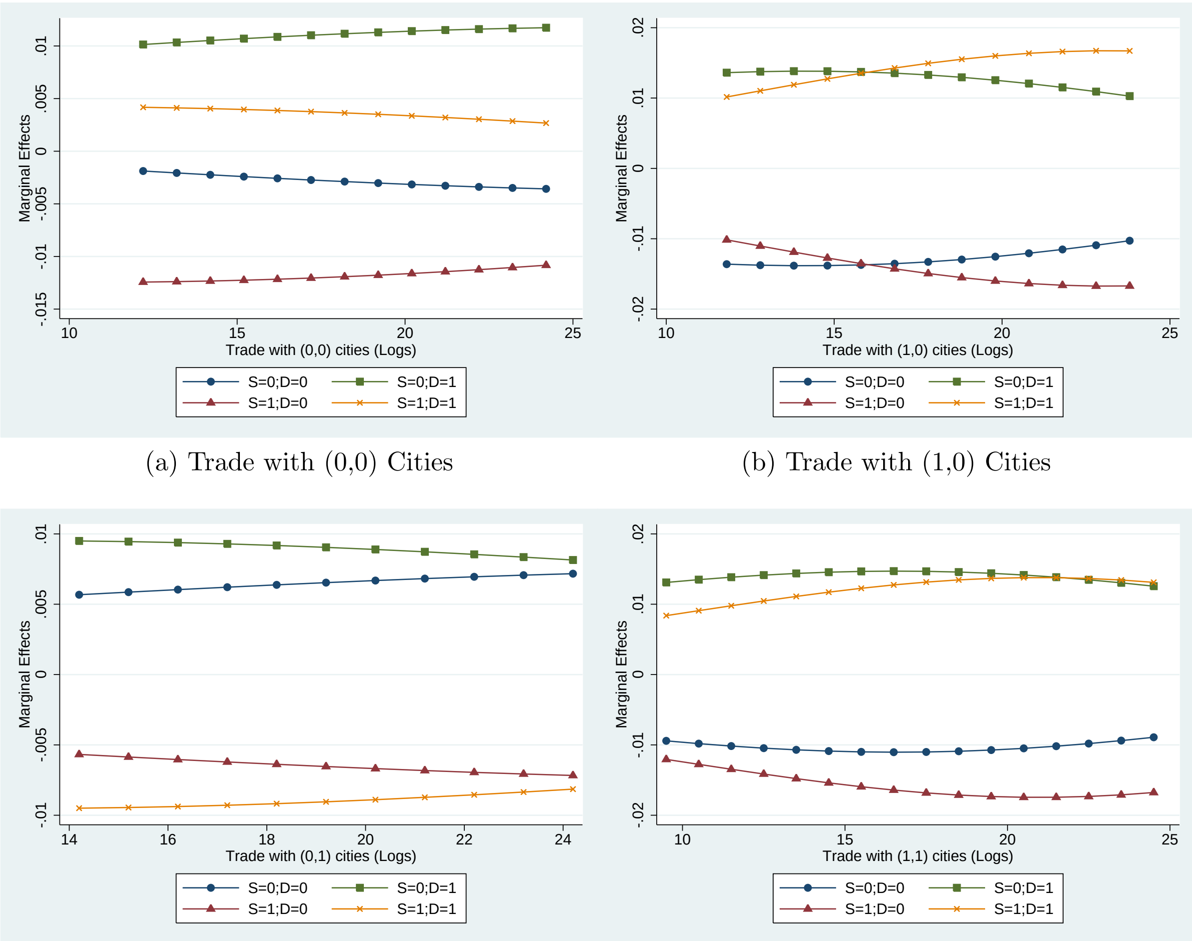
(c) Inter-city trade per capita (d) Number of varieties

Figure 4.8: Marginal effects for each city by regressors, 1995-2007 (Baseline Model).

Figure 4.9 only portraits the marginal effects for trade with other type of cities. The pattern is clear for diversified cities and, at a lower scale, for co-agglomeration economies: The more they trade with other type of cities, the more likely for them to arise; particularly if they trade with non-specified and specialized cities. For diversified cities, this result is in line with the theoretical propositions by Anas and Xiong (2005). But in contrast to these authors findings, specialized cities do not tend to emerge when they trade with diversified cities. Only diversified cities (and non-specified ones) are more likely to appear when they trade with other diversified cities. Focusing on specialized and non-specified cities, they show similar effects when trading. At lower levels of trade, specialized cities are more probable to appear, but in all the cases the impacts are very small and decreasing as they increase trade with other cities. Finally, when the trading partner is another co-agglomeration economy, cities with the same typology and

economies appearance heavily depends on workers involved in services, the levels of high-educated people and, surprisingly, the mining activities they harbor, giving support to previous findings from Ellison et al. (2010). Last, the emergence of non-specified cities only depends on the production of standard products. Levels of specialization and diversification follows robust patterns, being the effect of the former even higher for co-agglomeration economies than for specialized cities, whereas the impact of the latter is only positive and mainly constant for diversified and co-agglomeration economies.

diversified cities are the ones that tend to arise.



(a) Trade with (0,0) Cities

(b) Trade with (1,0) Cities

(c) Trade with (0,1) Cities

(d) Trade with (1,1) Cities

Figure 4.9: Marginal effects for trading partners, 1995-2007 (Baseline Model).

At the light of these results, it is important to remark that co-agglomeration economies are the most very peculiar type of cities. By mixing specialized and diversified structures, these cities tend to show effects belonging to their (1,0) and (0,1) counterparts. In our understanding, these kind of cities are those that show pretty similar qualitative patterns as those for diversified cities, but with lower quantitative impacts as a result of hosting specialized sectors. These sectors allow co-agglomeration economies to take advantage of the elements attached to specialized cities and complement their diversified structure.

4.4.3 Path Dependency

Today cities are result of their past. As evidenced by Eaton and Eckstein (1997), past levels of urban population are a key determinant of long-run paths of cities' development. The model by Anas and Xiong (2005) proposes that not only population, but also previous levels of transport costs, determine future patterns of diversification and specialization. This idea is also shared by Tabuchi and Thisse (2011) although in an opposite way: whereas the former argue that high transport costs leads to diversified cities, the latter refer to reductions in transport costs as the explanation for actually diversified cities. Tables 4.5 and 4.6 give the first insights on this issue as the levels of specialization and diversification in 1995 are crucial determinants of the future structure of the city. But even these levels in 1995 can be affected by past trend in population and transport costs.

With this aim, Table 4.7 sheds light on the causal effects that the past has on the future. It accounts for lagged levels of GTCs, particularly those from 1980 (first available year), 1985 and 1990, and population (in 1980 so as to be consistent with the transport costs). It considers our baseline model from Table 4.5 with variations according to the new variables. Columns (1) and (4) only use lagged levels of GTC; columns (2) and (5) do the same with the population in 1980 as a proxy for city size; and columns (3) and (6) represent our new benchmark, controlling either for lagged GTCs and population. Again, Table 4.7 shows the probit models for the probability of being specialized or diversified, while Table 4.8 focuses on the average marginal effects obtained from the bivariate procedure.

The new variables are highly significant for specialized cities (Column (3)). Population in 1980 has a negative impact while the GTC starts with a very negative impact that intensifies in 1985 as a result of the increase in transport costs in Spain in that year (as reflected in Figure 4.4).[20] By contrast, transport costs in 1990 seem to have a positive impact on the probability of being specialized. In other words, transport costs seem to have a non-linear effect on the emergence of specialized cities, being more negative the higher they were in the past, but becoming positive 5 years before the city gets finally specialized. The rest of the variables present the same impacts than in the benchmark from Table 4.5, except trade with other diversified cities that gains significance. For diversified cities, among the lagged variables only population in 1980 presents positive and significant impacts, whereas transport costs do not show clear effects. That is, for these cities their previous size more than their accessibility through transport costs seems to play the most important role. Note again that the ρ coefficients are significant in all specifications, supporting the bivariate estimation in the next step.[21]

Table 4.8 splits the previous results into the four type of cities. Now we can identify that population in 1980 have a direct negative impact on specialized and non-specified

[20] Between 1980-1985 Spain experimented a huge increase in petrol prices leading transport costs to increase, but after 1985 these prices suddenly reduced, INE (1988).

[21] We have made several regressions to confirm these results. Concretely, they are robust either (i) including only GTC in the past and not current GTCs levels (those for 1995-2007); (ii) dropping the level of specialization and diversification in 1995; or (iii) including separately GTCs for 1980, 1985 and 1990.

Table 4.7: Regression Results from Probit Model (with Lagged Variables)

	Pr(S=1| X)			Pr(D=1| X)		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Pop$_i$)	-0.265	—	—	0.845***	—	—
	(0.221)			(0.228)		
ln$\overline{GTC}_i$	3.763	0.334	4.600	6.019*	2.079***	5.374
	(3.629)	(0.712)	(3.805)	(3.319)	(0.670)	(3.388)
ln$\overline{GTC}1980_i$	-16.391*	—	-21.370**	7.713	—	8.635
	(8.408)		(8.781)	(6.969)		(7.051)
ln$\overline{GTC}1985_i$	-45.532***	—	-39.254***	-1.029	—	4.117
	(10.886)		(11.839)	(10.437)		(11.624)
ln$\overline{GTC}1990_i$	56.600***	—	54.046***	-10.922	—	-15.740
	(10.060)		(10.796)	(8.979)		(10.544)
ln(% High Education$_i$)	-0.946**	-1.316***	-1.412***	3.857***	4.672***	4.610***
	(0.457)	(0.495)	(0.514)	(0.593)	(0.655)	(0.689)
ln(Ratio_MS$_i$)	0.664***	0.582***	0.732***	0.183	0.163	0.138
	(0.135)	(0.154)	(0.168)	(0.131)	(0.148)	(0.149)
ln(MES$_i$)	0.604***	0.562***	0.499***	-0.875***	-0.776***	-0.771***
	(0.166)	(0.158)	(0.166)	(0.197)	(0.206)	(0.204)
ln(Mining$_i$)	0.216**	0.287**	0.349**	0.074	-0.010	0.008
	(0.099)	(0.139)	(0.156)	(0.095)	(0.126)	(0.140)
Number of Varieties$_i$	-0.399	-0.528	-0.316	-0.687**	-0.976***	-0.924***
	(0.304)	(0.331)	(0.352)	(0.322)	(0.352)	(0.357)
ln(Intercity_Tradepc$_i$)	0.278	-0.025	-0.138	-0.336**	-0.152	-0.166
	(0.177)	(0.238)	(0.227)	(0.160)	(0.172)	(0.176)
lnTrade with (0,0)$_i$	-0.047	0.009	-0.013	0.074	0.050	0.065
	(0.052)	(0.058)	(0.059)	(0.046)	(0.048)	(0.049)
lnTrade with (1,0)$_i$	—	—	—	0.130**	0.067	0.068
				(0.066)	(0.070)	(0.072)
lnTrade with (0,1)$_i$	0.027	0.369**	0.421***	—	—	—
	(0.119)	(0.165)	(0.161)			
lnTrade with (1,1)$_i$	-0.039	-0.045	-0.076	0.122**	0.085	0.093
	(0.058)	(0.059)	(0.061)	(0.062)	(0.066)	(0.067)
ln(RZI1995$_i$)	0.900***	1.022***	1.011***	—	—	—
	(0.103)	(0.127)	(0.128)			
ln(RDI1995$_i$)	—	—	—	0.612**	0.200	0.183
				(0.292)	(0.308)	(0.355)
ln(Pop1980$_i$)	—	-0.457*	-0.612***	—	1.209***	1.152***
		(0.248)	(0.237)		(0.245)	(0.249)
Spatial FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	722	678	678	722	678	678
ρ	-0.315***	-0.316***	-0.272***	-0.315***	-0.316***	-0.272***

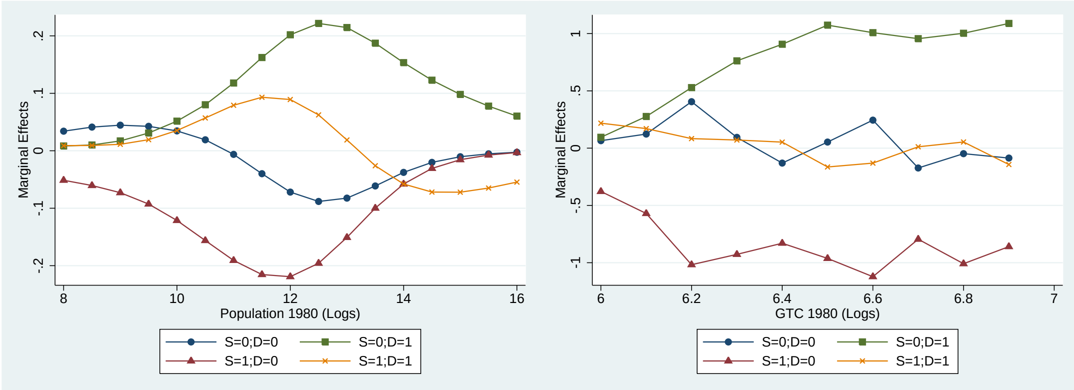
cities, while it is positive for diversified ones. Thanks to the bivariate estimation we can correctly disentangle the effects of GTCs on diversified cities. GTCs from 1980 and 1985 have a positive impact on these cities as proposed by Abdel-Rahman (1996); Anas and Xiong (2005), but negative 5 years before the city diversifies (in 1990). For these cities actual agglomeration economies variables, and not inter-city trade flows, are the most important determinants. The opposite holds for specialized cities (P_{10}), being GTC in 1990 the most explanatory variable. Co-agglomeration economies (P_{11}) again show a mix composition effect between specialized and diversified cities. Urban population in 1980 is not significant for co-agglomeration economies whereas GTCs from previous years follow the same pattern than for specialized cities. Finally, agglomeration economies, and, partially, trade flows (number of varieties and trade with diversified cities) have a remarkable impact on these cities. This Table confirms the non-linear effect that transport costs have on cities appearance. Co-agglomeration economies and specialized cities follow the same trend according to which GTC in 1990 is positive and highly significant, but previous levels of transport costs present a negative impact on these cities. For diversified cities the non-linear effect of GTC holds but in the opposite direction; i.e., transport costs from 1980 and 1985 have a positive impact but suddenly change in 1990.

Finally, to further explore the non-linear effects obtained in Tables 4.7 and 4.8, Figure 4.10 plots the marginal effects for the lagged variable over each city type. Population in 1980 (a) confirms the previous results: Only diversified cities, and to a lower scale co-agglomeration economies, tend to arise the higher their population was in the past. GTCs from 1980 (b) and 1985 (c) present and support the idea by which high levels of transport costs in the past result in a higher probability of observing diversified cities in the future. This pattern drastically changes for 1990 (d) when the situation turns around as worse accessibility through transport costs leads to lower impact on getting diversified cities but higher on obtaining specialized cities. These effects of transport costs were mainly focused in these two type of cities, but not in non-specified or co-agglomeration economies. For their part, transport costs in the past seem to have an effect of almost null (0) impact or 0.5 at best.

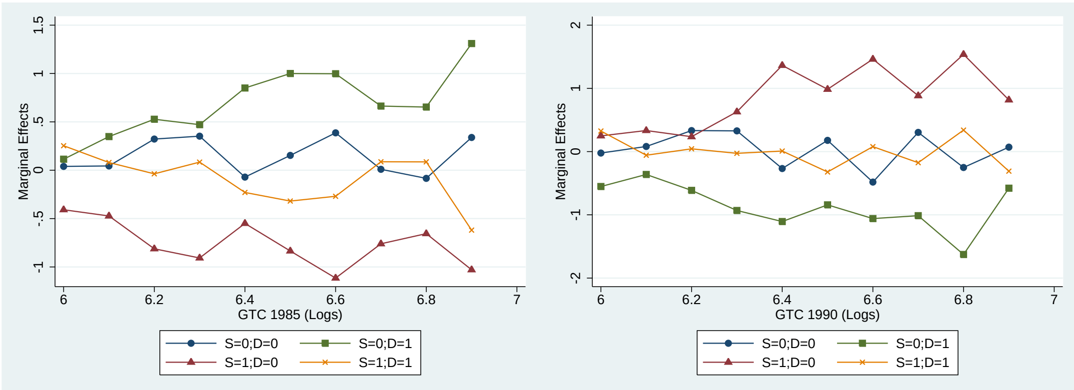
Table 4.8: Marginal Effects from Bivariate-Probit Model (with Lagged Variables).

	Average Marginal Effects			
Variables	(P_{00})	(P_{10})	(P_{01})	(P_{11})
$\ln\overline{GTC}_i$	-1.763**	-0.187	0.020	1.930*
	(0.832)	(0.779)	(1.234)	(0.999)
$\ln\overline{GTC}1980_i$	1.079	-4.212**	7.020**	-3.886*
	(1.797)	(1.733)	(2.767)	(2.280)
$\ln\overline{GTC}1985_i$	4.467	-5.961**	10.409***	-8.915***
	(3.029)	(2.498)	(3.801)	(3.377)
$\ln\overline{GTC}1990_i$	-4.018	9.730***	-16.463***	10.751***
	(2.540)	(2.424)	(3.675)	(2.959)
ln(% High Education$_i$)	-0.784***	-0.889***	1.319***	0.354**
	(0.166)	(0.139)	(0.211)	(0.175)
ln(Ratio_MS$_i$)	-0.129***	0.079**	-0.148***	0.199***
	(0.037)	(0.034)	(0.056)	(0.044)
ln(MES$_i$)	0.095**	0.184***	-0.284***	0.005
	(0.044)	(0.045)	(0.067)	(0.046)
ln(Mining$_i$)	-0.049	0.046	-0.083*	0.086**
	(0.036)	(0.030)	(0.049)	(0.043)
Number of Varieties$_i$	0.239***	0.097	-0.119	-0.217**
	(0.086)	(0.075)	(0.121)	(0.100)
ln(Intercity_Tradepc$_i$)	0.054	0.006	-0.002	-0.059
	(0.046)	(0.042)	(0.069)	(0.060)
lnTrade with (0,0)$_i$	-0.012	-0.012	0.017	0.007
	(0.013)	(0.011)	(0.018)	(0.016)
lnTrade with (1,0)$_i$	-0.014	-0.010	0.014	0.010
	(0.015)	(0.011)	(0.015)	(0.011)
lnTrade with (0,1)$_i$	-0.057***	0.057***	-0.102**	0.102**
	(0.022)	(0.022)	(0.040)	(0.040)
lnTrade with (1,1)$_i$	-0.009	-0.024*	0.038*	-0.004
	(0.016)	(0.014)	(0.021)	(0.017)
ln(RZI1995$_i$)	-0.137***	0.137***	-0.246***	0.246***
	(0.020)	(0.020)	(0.034)	(0.034)
ln(RDI1995$_i$)	-0.039	-0.028	0.039	0.028
	(0.075)	(0.054)	(0.075)	(0.054)
ln(Pop1980$_i$)	-0.161***	-0.257***	0.392***	0.026
	(0.060)	(0.052)	(0.081)	(0.070)
Spatial FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	678	678	678	678

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.



(a) Population in 1980 (b) GTC in 1980



(c) GTC in 1985 (d) GTC in 1990

Figure 4.10: Marginal effects for each city by lagged variables.

4.5 Discussion

Several relevant results have been obtained in Section 4.4. From the baseline model, city size and variables related to agglomeration economies are key drivers in the emergence of cities. Although these results concur with previous studies, among others, by Henderson (2003), we elaborate the analysis by splitting the impact of agglomeration economies considering four type of cities. Diversified and specialized cities present a dichotomic pattern by city size. Diversified cities are more likely to appear the larger their size, whereas the specialized cities are affected by population in the opposite way. This conclusion supports the theoretical propositions by Anas and Xiong (2003, 2005). Moreover, the effect of the city size becomes huge when we control for previous levels of population, reflecting the existence of long-run urban growth trends (Eaton and Eckstein, 1997). Non-specified cities and co-agglomeration economies also follow this similar dichotomic pattern by city size. Helsley and Strange (2014) give insights of this result for co-agglomeration economies explaining that cities can (inefficiently) attract or maintain levels of population higher that the ones we should expect. For non-specified cities we do not have a clear identification apart from our own hypothesis by which any increase in the population of these small cities can be directly related to any change in their economic structure, causing these cities to disappear in favour of other type of cities.

Agglomeration economies show the expected results. Natural advantages and localization economies are important for specialized cities and co-agglomeration economies but not for diversified ones, supporting in this case previous findings by Abdel-Rahman and Fujita (1993); Ellison et al. (2010). The type of specialization by products is particularly important, being standard products by manufacturing workers the case of specialized cities (Duranton and Puga, 2001), while diversified cities are characterized by workers in services-related sectors exploiting *economies of scope* at the city level (Abdel-Rahman and Fujita, 1993). As we contend that co-agglomeration economies are the result of a mix structure between specialized and diversified cities, we find that either localization economies (specialization advantages) and economies of scope (diversification advantages) play a significant role in this type of cities.

One of the novel aspects of this paper is the use of monetary transport costs. Diversified and specialized cities do not seem to be, on average, directly affected by current GTC levels in the baseline model. This goes against the main assumptions of Abdel-Rahman (1996); Anas and Xiong (2005); Tabuchi and Thisse (2011). By contrast, when disentangling the effects of GTCs, we show that they are clearly non-linear, as for lower levels of GTC diversified and specialized present a higher chance to appear, but as long as accessibility (transport costs) worsen, co-agglomeration economies are more likely to emerge. Indeed, when we resort to past levels of GTC, we see that transport costs are highly significant on the probability of getting either diversified or specialized cities. This effect is even bigger than those from city size the farther back (lagged) we move in time. On this regard and according to our database, transport costs with a 15 years lag seem to have a very important impact on cities, being negative for specialized and positive for diversified cities. Moreover, transport costs closer to recent years seem to have, not a negative, but a positive impact on the urban structure; that is, there is a non-linear time effect of transport costs on cities.

Even if these results relating transport costs with city typology are puzzling, they might have an explanation recalling the theoretical framework developed within the New Economic Geography literature, which yield non-linear relations. The bifurcation diagrams, known as tomahawks in Fujita et al.(1999), illustrate that transports costs (GTCs) variations may change the spatial configuration of economic activity. In our case, the initial increase in transport costs goes against the specialization (agglomeration) of cities by spreading economic activity across the landscape (the so-called sustain point). In terms of our analysis, the spread of economic activity results in more diversified cities (a spatial dispersion that would eventually characterize a flat-earth equilibrium). Thus, the increase in transport costs between 1980 and 1985 as a result of higher fuel prices, triggers the emergence of diversified cities and the corresponding reduction in specialized ones. By contrast, following again NEGs framework, the subsequent reduction in transport costs after 1985 results in the opposite agglomeration of economic activity in specialized cities by way of the existing increasing returns to scale (MES), so a core-periphery pattern is favored as the break point is reached. A conclusion that is observed in our results as the number of specialized and diversified cities increases and reduces, respectively. Indeed, our findings constitute an empirical validation of probably the most important NEG proposition, i.e., the non-linear relation between transports costs and the spatial distribution of economic activity; without the theoretical (instantaneous) changes between short run and lung run equilibria associated to "catastrophic" agglomerations and dispersions. The Anas and Xiong (2005) model also gives similar insights to those summarized by Fujita et al.(1999) for this puzzling effect of transport costs, giving an additional theoretical explanation for our results on the emergence of different types of cities.

Further studying the determinants of city characteristics, trade flows partially exhibit the expected results in the baseline model. Diversified cities are more probable to appear when they trade with cities with some sort of specialization. Specialized cities do the opposite, although their chance to arise reduces the more they intensively trade regardless of destination (inter-city trade per capita). When lagging the variables, the effects of trade flows vanish with the exception of trade flows with diversified cities, showing a positive effect for specialized and co-agglomerated cities, but a negative one for diversified cities. This last result goes in line with the propositions by Anas and Xiong (2003, 2005) but, in general, we can conclude that interactions through trade flows are not as important as the literature has emphasized.

Finally, we find additional sources of path-dependency patterns not only coming from past city size and transport costs, but through previous levels of specialization. With diversification we only get significant effects when transport costs and city size in the past are not controlled for. The theoretical benchmark is very limited on this regard impeding us to have a clear explanation for this result. Nevertheless, although cities suffer from path-dependency elements, there is a (restricted) role for other factors, mainly agglomeration economies and actual levels of GTC, to modify cities structure.

4.6 Conclusions

Diversification and specialization are normally the two sides of the same coin discussed in he literature, while in turn out that they may coexist within a system of cities. In this paper we have analyzed under what factors or conditions different cities emerge. Therefore, we have gone beyond the classical dichotomy between diversified and specialized cities cities characterized by both dimensions through co-agglomeration economies, or none of this. As the literature lacks having a clear unified framework on cities typology, we take elements from different strands of the literature to develop an empirical strategy that allows us to asses the validity of the theoretical assumptions proposed.

Out econometric approach is based on bivariate-probit techniques, which are implemented using three unique databases on Spanish urban areas. We find that agglomeration economies and transport costs are the key drivers on the emergence of cities. Interactions through trade flows, although relevant, are not as important as the literature has emphasized. We find that cities tend to change their internal structure (transition to a different category) much more than expected. Within this dynamic, path-dependence factors, past levels of population and transport costs, emerge as key determinants of future levels of cities' diversification and/or specialization. Indeed and rather surprisingly, we obtain a non-linear effect of previous transport costs on cities emergency for which we do not have a clear explanation.

The main part of the results obtained goes hand-by-hand with the propositions reached by the theoretical literature. Nevertheless, further research is needed to clearly understand the role that agglomeration economies, transport costs and trade flows, have on the emergence of cities, specially in the case of co-agglomeration economies and non-specified cities. Our findings leave a limited space for policy interventions in cities economic structure. Thus, this research opens the door for a future empirical research on the role that urban planners and transport policy makers may have on cities formation.

Acknowledgments

I deeply acknowledge support and advise from Raymond Flórax (Purdue University) and Peter Mulder (Vrije Universiteit Amsterdam) to develop this chapter. The seminal idea for this collaboration arises when I was as visiting researcher at Purdue University (Krannert Bussiness School) in the Spring Semester 2014. Later, the chapter was developed while I was visiting the Department of Spatial Economics at the Vrije Universiteit Amsterdam during the Summers of 2014 and 2015.

I thank the comments received at the SERC 2016 Annual Conference (LSE) from John Parr and other participants. The authors gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness (MINECO ECO2013-46980-P, co-financed with FEDER) and the FPU-2010 Scholarship for pre-doctoral Education Training by the Spanish Ministry of Education.

4.7 Bibliography

1. Abdel-Rahman, H. M. (1988). Product differentiation, monopolistic competition and city size. *Regional Science and Urban Economics*, 18 (1), 69–86.

2. Abdel-Rahman, H. M. (1990a). Agglomeration economies, types, and sizes of cities. *Journal of Urban Economics*, 27 (1), 25–45.

3. Abdel-Rahman, H. M. (1990b). Sharable inputs, product variety and city sizes. *Journal of Regional Science*, 30 (3), 359–374.

4. Abdel-Rahman, H. M. and Fujita, M., (1993). Specialization and diversification in a system of cities. *Journal of Urban Economics*, 33 (2), 189–222.

5. Abdel-Rahman, Hesham M. (1996). When do cities specialize in production? *Regional Science and Urban Economics*, 26 (1), 1–22.

6. Anas, A. and Xiong, K. (2003). Intercity trade and the industrial diversification of cities. *Journal of Urban Economics*, 54 (2), 258–276.

7. Anas, A. and Xiong, K. (2005). The formation and growth of specialized cities: efficiency without developers or malthusian traps. *Regional Science and Urban Economics*, 35 (4), 445–470.

8. Betran, Concha (2011). Regional specialisation and industry location in the long run: Spain in the US mirror (1856–2002). *Cliometrica*, 5 (3), 259–290.

9. Brezis, Elise S. and Krugman, Paul R. (1997). Technology and the life cycle of cities. *Journal of Economic Growth*, 2 (4), 369–383.

10. Brezzi, M., Piacentini, M., Rosina, K., and Sanchez-Serra, D. (2012). Redefining urban areas in OECD countries. *Redefining "Urban"*, pages 19–58.

11. Cilasun, S. Mümin and Günalp, Burak (2012). An Empirical Analysis of the Average Plant Start-Up Size in Turkish Manufacturing Industries. *Review of Industrial Organization*, 40 (4), 273–289.

12. Christaller, Walter (1933). *Central places in Southern Germany.* Translated Edition. Prentice-Hall.

13. Clark, Don P. and Stanley, Denise L. (1999). Determinants of intra-industry trade between developing countries and the United States. *Journal of Economic Development*, 24 (2), 79–95.

14. Davis, D. R. and Dingel, J. I. (2014). The comparative advantage of cities. *National Bureau of Economic Research*, Working Paper 20602.

15. Desmet, K. and Rossi-Hansberg, E. (2014). Spatial development. *The American Economic Review*, 104 (4), 1211–1243.

16. Díaz-Lanchas, J., Llano-Verduras, C., and Zofío, J. L. (2013). Trade margins, transport cost thresholds and market areas: Municipal freight flows and urban hierarchy. *Working Papers in Economic Theory 2013/10*, Universidad Autónoma de Madrid (Spain), Department of Economic Analysis (Economic Theory and Economic History).

17. Duranton, G. and Puga, D. (2000). Diversity and specialisation in cities: why, where and when does it matter?, *Urban studies*, 37 (3), 533–555.

18. Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91 (5), 1454–1477.

19. Eaton, B. Curtis and Lipsey, Richard G. (1982). An economic theory of central places. *The Economic Journal*, 92 (365), 56–72.

20. Eaton, J. and Eckstein, Z. (1997). Cities and growth: Theory and evidence from France and Japan. *Regional science and urban Economics*, 27 (4), 889–927.

21. Ellison, G. and Glaeser, E. L. (1997). Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy*, 105 (5), 889–927.

22. Ellison, G. and Glaeser, E. L. (1999). The geographic concentration of industry: Does natural advantage explain agglomeration?. *The American Economic Review*, 89 (2), 311–316.

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

24. Fujita, M. (1988). A monopolistic competition model of spatial agglomeration: Differentiated product approach. *Regional Science and Urban Economics*, 18 (1), 87–124.

25. Fujita, M. (2012). Thünen and the new economic geography. *Regional Science and Urban Economics*, 42 (6), 907–912.

26. Fujita, M., Krugman, P. R., and Venables, A. (1999). *The spatial economy: Cities, regions, and international trade.* MIT press.

27. Glaeser, E. L., Ponzetto, G. A. M. and Tobio, K. (2014). Cities, skills and regional change. *Regional Studies*, 48 (1), 7–43.

28. Greene, W. H. (2012). *Econometric Analysis.* Pearson Education, seventh edition.

29. Helsley, R. W. and Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, 122 (5), 1064–1093.

30. Henderson, J. V. (1974). The sizes and types of cities. *The American Economic Review*, 64 (4), 640–656.

31. Henderson, J. V. (1997). Externalities and industrial development. *Journal of Urban Eonomics*, 42 (3), 449–470.

32. Henderson, V. (2000). Political economy of city sizes and formation. *Journal of Urban Economics*, 48, 453–484.

33. Henderson, J. V. (2003). Marshallian scale economies. *Journal of Urban Economics*, 53, 1–28.

34. Hillberry, R. and Hummels, D. (2008). Trade responses to geographic frictions: A decomposition using micro-data. *European Economic Review*, 52 (3), 527 – 550.

35. Instituto de Estudios de Transporte y Comunicaciones (1979). Estadísticas de transporte: Series cronológicas 1950-1977. *Ministerio de Transporte y Comunicaciones*, Madrid.

36. Instituto Nacional de Estadística (1988). Transporte regular de viajeros por carretera. *Instituto Nacional de Estadística*, Madrid.

37. Jacobs, J. (1969). *The economy of cities.* Random House, New York.

38. Llano, C., Esteban, A., Pérez, J., and Pulido, A. (2010). Opening the interregional trade “black box”: The c-intereg database for the spanish economy (1995-2005). *International Regional Science Review*, 33 (3), 302–337.

39. Rivera-Batiz, Francisco L. (1988). Increasing returns, monopolistic competition, and agglomeration economies in consumption and production. *Regional Science and Urban Economics*, 18 (1), 125–153.

40. Tabuchi, T. and Thisse, J.-F. (2011). A new economic geography model of central places. *Journal of Urban Economics*, 6 (4), 240–252.

41. Viladecans-Marsal, Elisabet (2002). The growth of cities: Does agglomeration matter. *Institut d'Economia de Barcelona (IEB)*, Document de treball 2002/3.

42. Viladecans-Marsal, Elisabet (2004). Agglomeration economies and industrial location: city-level evidence. *Journal of Economic Geography*, 4 (5), 565–582.

43. Zofío, J. L., Condeço-Melhorado, A. M., Maroto-Sánchez, A., and Gutiérrez, J. (2014). Generalized transport costs and index numbers: A geographical analysis of economic and infrastructure fundamentals. *Transportation Research Part A: Policy and Practice*, 67 (C), 141–157.

4.8 Appendix 4.A: Data Representativeness

To check the representativeness of our data on employment at the firm level (the SABI database), we take official data on employment at the industry-level (38 different sectors) and at province level both collected by the Spanish National Institute and the Spanish Ministry of Labor to get official levels of employment (registered workers) by year and by size of the firm. Using these sources, we calculate how representative is the SABI database attending to different dimensions, these are: the size of the firm by year, the industry and the province levels both for all years in the panel database.

We resort to different ratios, where x represents the number of workers in city c in time t according to both databases on employment, the SABI and the Official ("Off") database. For each dimension we calculate two indexes, a *simple ratio* and a *ratio of ratio*, both define as:

Table 4.9: Ratios used for representativeness of SABI Database.

Equations	*Dimension*	*Simple Ratio*	*Ratio of Ratio*
(Eq.1) ; (Eq.2)	Size of Firm (s)	$Ratio_t^s = \frac{\sum_i x_t^s{}_{SABI}}{\sum_i x_t^s{}_{Off}}$	$(Ratio\text{-}Ratio)_t^s = \frac{\frac{\sum_c x_t^s{}_{SABI}}{\sum_c x_{t\,SABI}}}{\frac{\sum_c x_t^s{}_{Off}}{\sum_c x_{t\,Off}}}$
(Eq.3) ; (Eq.4)	Industry (k)	$Ratio_k = \frac{\sum_t x_t^k{}_{SABI}}{\sum_t x_t^k{}_{Off}}$	$(Ratio\text{-}Ratio)_k = \frac{\frac{\sum_t x_t^k{}_{SABI}}{\sum_t x_{t\,SABI}}}{\frac{\sum_t x_t^k{}_{Off}}{\sum_t x_{t\,Off}}}$
(Eq.5) ; (Eq.6)	Provinces (r)	$Ratio_r = \frac{\sum_t x_t^r{}_{SABI}}{\sum_t x_t^r{}_{Off}}$	$(Ratio\text{-}Ratio)_r = \frac{\frac{\sum_t x_t^r{}_{SABI}}{\sum_t x_{t\,SABI}}}{\frac{\sum_t x_t^r{}_{Off}}{\sum_t x_{t\,Off}}}$

Table 4.10: Firm Level: SABI Vs.Official Database (by year)

			Ratio SABI/Official Database				Ratio of Ratio		
Year	Workers SABI	Workers Official Data	(1) Total	(2) Small	(3) Medium	(4) Big	(5) Small	(6) Medium	(7) Big
1995	1,357,964	6,006,100	.22	.17	.26	.3	.74	1.14	1.33
1996	3,660,271	7,886,500	.46	.28	.59	.88	.59	1.25	1.87
1997	3,860,721	8,364,800	.46	.3	.6	.75	.65	1.29	1.61
1998	4,271,373	8,988,200	.47	.3	.57	.82	.64	1.19	1.74
1999	4,401,396	9,659,100	.45	.29	.54	.8	.63	1.18	1.75
2000	4,193,074	10,234,600	.41	.23	.5	.76	.56	1.24	1.85
2001	4,367,842	10,652,300	.41	.22	.49	.79	.53	1.19	1.93
2002	5,448,871	10,977,300	.5	.3	.58	.88	.61	1.16	1.78
2003	6,929,670	11,278,000	.61	.45	.66	.95	.74	1.09	1.54
2004	8,081,766	11,772,100	.69	.56	.68	.99	.81	.99	1.44
2005	9,786,354	12,499,500	.78	.63	.69	1.23	.8	.88	1.57
2006	9,507,272	13,100,300	.73	.64	.68	.96	.88	.94	1.33
2007	9,928,389	13,557,300	.73	.67	.7	.88	.92	.96	1.2

Columns (1)-(4): Ratio of the number of workers on SABI over the Official Database by type of firm. *Columns (5)-(7)*: Ratio of Ratio of SABI over the Official Database by type of firm. *Small Firm*: Lower than 50 employees; *Medium Firm*: 50-250 employees; *Big Firm*: Higher than 250 employees. *Source*: SABI Database and Spanish Ministry of Employment.

Table 4.11: Industry Level: SABI Vs.Official Database (all years)

NACE	SABI/Official	Ratio of Ratio
01-03	.18	.42
05-09	.18	.43
10-12	.04	.09
13-15	.7	1.78
16-18	.42	.96
19	3.12	7.15
20	.42	.96
21	.97	2.23
22-23	.3	.7
24-25	.47	1.1
26	1.76	4.1
27	.59	1.35
28	1.24	2.85
29-30	.36	.85
31-33	.85	1.96
35	.64	1.46
36-39	.75	1.72
41-43	.03	.07
45-47	.32	.73
49-53	1.97	4.52
55-56	.32	.75
58-60	1.5	3.46
61	.11	.25
62-63	.85	1.96
64-66	.32	.74
69-71	.5	1.14
72	2.75	6.33
73-75	4.47	10.27
77-82	.07	.15
85	.09	.20
90-93	.6	1.36
94-96	.00	.00
97-98	.00	.00

SABI/Official: Ratio of the number of workers on SABI over the Official Database by Industry.
Ratio of Ratio: Ratio of Ratio of SABI over the Official Database by Industry.
Source: SABI Database and Spanish National Statistic Institute.

Table 4.12: Province Level, SABI Vs.Official Database (all years).

Province	Ratio SABI/Official	Ratio of Ratio	Province	Ratio SABI/Official	Ratio of Ratio
Álava	.32	1	—	—	—
Albacete	.21	.66	León	.15	.46
Alicante	.25	.79	Lleida	.21	.65
Almería	.18	.57	La Rioja	.23	.71
Ávila	.10	.32	Lugo	.14	.44
Badajoz	.14	.43	Madrid	.58	1.83
Barcelona	.41	1.3	Málaga	.19	.59
Burgos	.25	.77	Murcia	.27	.86
Cáceres	.13	.42	Navarra	.3	.95
Cádiz	.16	.49	Ourense	.19	.6
Castellón	.35	1.1	Asturias	.26	.8
Ciudad Real	.17	.52	Palencia	.14	.44
Córdoba	.18	.57	Pontevedra	.28	.89
La Coruña	.26	.81	Salamanca	.15	.47
Cuenca	.12	.37	Cantabria	.18	.57
Girona	.29	.9	Segovia	.14	.43
Granada	.15	.48	Sevilla	.21	.66
Guadalajara	.15	.49	Soria	.15	.46
Guipuzkoa	.27	.86	Tarragona	.23	.73
Huelva	.18	.55	Teruel	.17	.54
Huesca	.17	.52	Toledo	.19	.6
Jaén	.12	.52	Valencia	.3	.96
Valladolid	.3	.93	Bizkaia	.39	1.21
Zamora	.12	.37	Zaragoza	.31	.98

SABI/Official: Ratio of the number of workers on SABI over the Official Database by Province.
Ratio of Ratio: Ratio of Ratio of SABI over the Official Database by Industry.
Source:SABI Database and Spanish National Statistic Institute

4.9 Appendix 4.B: Robustness Analysis

Table 4.13: Regression Results from Probit Models 1995-2007 (using Av. Firm Size)

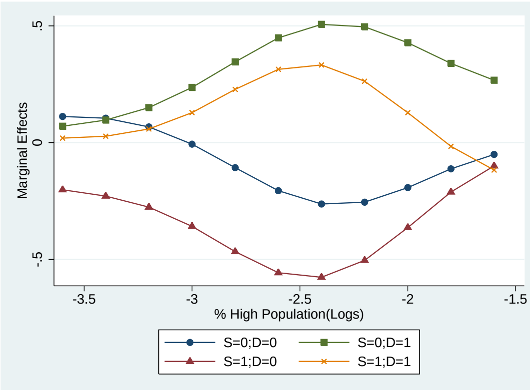
	Pr(S=1| X)			**Pr(D=1| X)**		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Average Firm Size$_i$)	-0.204**	-0.354***	-0.224**	0.770***	0.675***	0.673***
	(0.082)	(0.117)	(0.106)	(0.111)	(0.138)	(0.145)
ln$\overline{GTC_i}$	0.791*	1.474***	1.011	2.394***	2.228***	2.231***
	(0.471)	(0.569)	(0.647)	(0.595)	(0.652)	(0.655)
ln(% High Education$_i$)	-1.229***	-1.398***	-0.962**	3.844***	4.070***	4.129***
	(0.319)	(0.455)	(0.482)	(0.550)	(0.574)	(0.577)
ln(Ratio_MS$_i$)	0.407***	0.599***	0.600***	0.013	0.038	0.042
	(0.090)	(0.130)	(0.139)	(0.096)	(0.128)	(0.131)
ln(MES$_i$)	0.226	0.397**	0.527***	-0.236*	-0.434**	-0.439**
	(0.140)	(0.159)	(0.167)	(0.134)	(0.172)	(0.174)
ln(Mining$_i$)	-0.067	0.071	0.255***	0.038	0.029	0.006
	(0.061)	(0.081)	(0.082)	(0.079)	(0.091)	(0.094)
Number of Varieties$_i$		-0.435*	-0.587**		-0.326	-0.343
		(0.254)	(0.272)		(0.274)	(0.273)
ln(Intercity_Tradepc$_i$)		0.367***	0.451***		-0.632***	-0.620***
		(0.131)	(0.141)		(0.131)	(0.141)
lnTrade with (0,0)$_i$		0.021	-0.029		0.093**	0.095**
		(0.043)	(0.049)		(0.047)	(0.047)
lnTrade with (1,0)$_i$		—	—		0.184***	0.179***
					(0.065)	(0.067)
lnTrade with (0,1)$_i$		0.056	-0.040		—	—
		(0.095)	(0.104)			
lnTrade with (1,1)$_i$		-0.049	-0.012		0.168***	0.165***
		(0.048)	(0.052)		(0.060)	(0.060)
lnRZI1995$_i$			0.890***			—
			(0.105)			
lnRDI1995$_i$			—			0.154
						(0.291)
Spatial FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	873	718	717	873	718	717
ρ	-0.236***	-0.216***	-0.282***	-0.236***	-0.216***	-0.282***

Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.

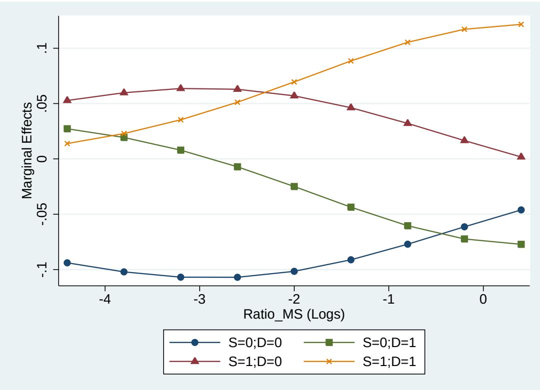
Table 4.14: Marginal Effects from Bivariate-Probit Model, 1995-2007 (Av.Firm Size).

	Average Marginal Effects			
Variables	(P_{00})	(P_{10})	(P_{01})	(P_{11})
ln(Average Firm Size$_i$)	-0.097***	-0.149***	0.186***	0.061*
	(0.031)	(0.032)	(0.040)	(0.036)
$\ln\overline{GTC}_i$	-0.565***	-0.252	0.167	0.649***
	(0.147)	(0.160)	(0.221)	(0.186)
ln(% High Education$_i$)	-0.655***	-0.857***	1.033***	0.479***
	(0.129)	(0.144)	(0.183)	(0.152)
ln(Ratio_MS$_i$)	-0.091***	0.076**	-0.145***	0.160***
	(0.029)	(0.033)	(0.047)	(0.039)
ln(MES$_i$)	0.011	0.150***	-0.218***	0.058
	(0.038)	(0.042)	(0.058)	(0.049)
ln(Mining$_i$)	-0.036*	0.034*	-0.064**	0.066**
	(0.021)	(0.020)	(0.028)	(0.026)
Number of Varieties$_i$	0.147**	-0.021	0.085	-0.210***
	(0.061)	(0.065)	(0.091)	(0.080)
ln(Intercity_Tradepc$_i$)	0.056*	0.171***	-0.233***	0.006
	(0.030)	(0.032)	(0.049)	(0.041)
lnTrade with (0,0)$_i$	-0.014	-0.021*	0.026	0.009
	(0.010)	(0.011)	(0.016)	(0.014)
lnTrade with (1,0)$_i$	-0.034***	-0.031***	0.034***	0.031***
	(0.013)	(0.012)	(0.013)	(0.012)
lnTrade with (0,1)$_i$	0.006	-0.006	0.010	-0.010
	(0.014)	(0.014)	(0.027)	(0.027)
lnTrade with (1,1)$_i$	-0.030**	-0.030**	0.034*	0.026
	(0.013)	(0.013)	(0.019)	(0.016)
ln(RZI1995$_i$)	-0.123***	0.123***	-0.227***	0.227***
	(0.017)	(0.017)	(0.028)	(0.028)
ln(RDI1995$_i$)	-0.029	-0.027	0.029	0.027
	(0.056)	(0.051)	(0.056)	(0.051)
Spatial FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	717	717	717	717

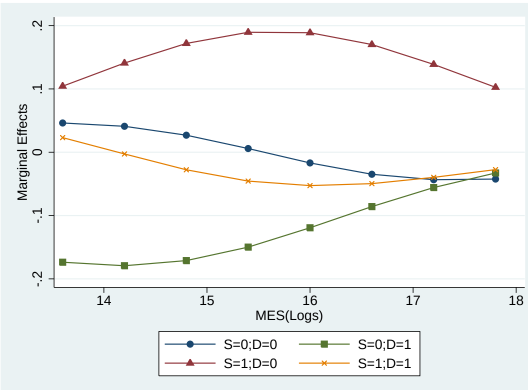
Robust standard errors. Standard errors in parentheses. Significance level: ***p<0.01, **p<0.05, *p<0.1.



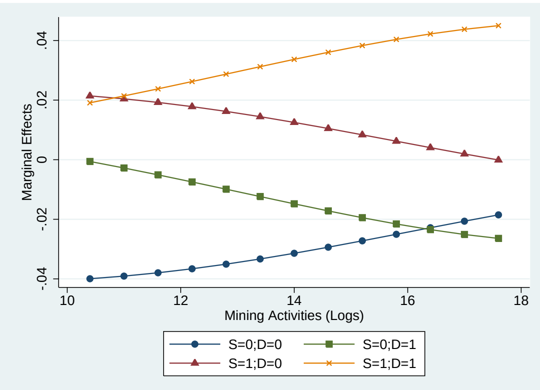
(a) % of High Educated Population



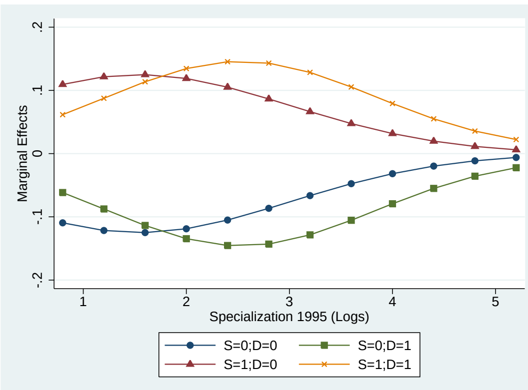
(b) Manufacturing Workers/Services Workers



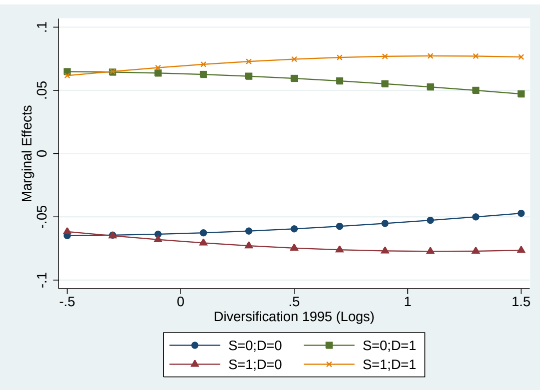
(c) MES (Product Standardization)



(d) Mining Activities



(e) Level of Specialization in 1995



(f) Level of Diversification in 1995

Figure 4.11: Marginal effects for each city by regressors, 1995-2007 (Bivariate Baseline Model).

Chapter 5

Conclusions and Policy Implications

Given the empirical results on trade patterns and city typologies obtained in this Dissertation, several conclusions arise, but even more further explorations are required. This section translates the main findings to a set of policy recommendations that can be achieved, and proposes promising ideas for future thinking.

The first paper focuses on the role that transport costs have on trade flows at very short distances (around 150km/185€), and how trade flows shape market areas surrounding big cities. Policymakers should be aware of this concentration of trade flows when proposing regional growth strategies. On this regard, because some cities are more predominant (exhibit a higher hierarchy) than others in the Spanish urban system, these cities would benefit relatively more than the smaller ones when promoting regional development through public infrastructure. It may leads to a higher agglomeration of the trading activity in these big cities. This issue becomes more crucial as these big cities present advantages within the network of cities by which, for the same distance thresholds (market areas), Madrid can reach further distances within Spain than Barcelona (Figure 2.5, Maps *a* and *b*). These results are in line with theoretical propositions from previous studies for the Spanish case (Barbero and Zofío, 2012).

Going beyond this concentration in short distances, road infrastructure investment during the Spanish housing bubble has become a disputed issue in recent times. Although there is evidence that road investment in Spain has got inefficiency problems (De La Fuente, 2010), i.e. the road investment shows a negative marginal rate of return, the positive effects of this investment must be considered too. In this sense, the reductions in transport costs through road infrastructure during the period 2003-2007 seem to promote trade (Table 2.2). Indeed, they have also helped to reduce the accumulation of trade flows at the municipal level (a lower coefficient in 2007 for the NUTS-5 variable in Table 2.4).[1] Moreover, connecting more municipalities through the road network has allowed

[1]Truly, we have not studied to what extent this lower concentration at the municipal level has been compensated with a higher agglomeration at the urban level (Metropolitan areas) in detriment to

more municipalities to trade more commodities to more destinations (Table 2.9), that is, municipalities have improved their diversification in two dimensions: commodities and destinations.

The findings obtained on this first paper open the door for a product level analysis to understand why some products are only traded in short distances while others travel over very long distances creating the observed hierarchical urban system, where some cities are more predominant than others. The so-called *Alchian-Allen effect* gives the theoretical (Borcherding and Silberberg, 1978) but also the empirical (Hummels and Skiba, 2004) underpinnings for these trade-patterns, which at the end allows the specialization on cities in very specific products.[2]

The second paper studies the degree of cities' specialization and how, because high-skill workers are attracted by some (big) cities and not for others, some cities specialize in complex/skill-intensive products. If complex goods are those that finally generate economic development (Hausmann and Hidalgo, 2011), we should expect some cities to grow more than others, leading to regional divergence within a country. In terms of policy implications the main (daunting) insight that arises is the inability of policy markers to promote specialization in small cities toward complex goods, just because high-skill workers will not have incentives to move to these cities in their search for better jobs. Once an economic crisis comes, big cities endure it in a better way than small ones as a result of their product specialization and more productive workers (capabilities).

To avoid this long-run pattern, policymakers should center their action in creating capabilities in smaller cities that will attract high-skill workers, and will allow for cities' specialization in skill-intensive products. On this respect, capabilities do not only arise through the investment on very important research centers but from promoting connections between individuals, firms and institutions. Indeed, this is a much cheaper way of creating capabilities. Given the actual endowments that each city has, if policymakers at the local level help by adding new activities to the city and by establishing connections across economic agents through information channels, cities will foster the capabilities needed to enhance diversity and, finally, economic growth.

This paper belongs to a prominent area of research; that is, to what extent cities specialization depending on their stage of development. Previous studies (Imbs and Wacziarg, 2003) find a non-linear relationship between countries' development (GDP) and how they diversify. At a very first stage of development, countries specialize in agricultural products, then they diversify by promoting new sectors and, finally, they specialize again but in more skill-intensive products. If this pattern is driven by those at

non-cities areas; an analysis that is worth performing in the future.

[2]By the time this Dissertation is being written, this idea is a work-in progress through a theoretical model by which cities specialize in high-quality goods depending on the demand-side effects that could arise at the destination when shipping goods.

the city level, future trends of countries' specialization could be predicted attending to their cities' performance. Hence, research urban projects within the biggest cities should be promoted to understand future trends of countries development.

The final paper sheds light on cities emergence and the factor behind their appearance. Probably the most important implication of this paper is the one corresponding to the research agenda. It warns about the need for future theoretical and empirical research on cities emergence, specially in those related to co-agglomeration economies. It is not very well analyzed in the literature why some cities present advantages in terms on diversification and specialization (co-agglomeration economies). Further comprehension of the role that agglomeration economies and cities interactions have on cities' emergence, would help policymakers to find cities' *comparative advantages*.

In the spirit of the second paper, the implications from the third paper are challenging in terms of policy interventions. On one side, cities' emergence is driven by the internal agglomeration economies, specially those related to the skills and the type of workers at the city level. On the other side, path dependence patterns seem to be crucial on the performance of cities. In particular, previous levels of population and transport costs (first nature advantages) are those that mainly determine the future typology of cities, leaving small space for policy interventions. As low transport costs in the past positively affect future cities' specialization, policymakers face a continuous trade-off. If they want to reduce transport costs to increase cities' trade flows (implication from the first paper), they would do it at the expense of fostering more specialized cities, resulting in inequalities across cities and core-periphery patterns *à la* NEG. Even more, if diversified cities are those that promote economic performance in the long run whereas specialize cities enhance efficiency gains by producing standard products (Duranton and Puga, 2001), the fact of reducing transport costs would generate that some areas (specialized cities) would not grow at the same extent than others (diversified cities).

The lines of research that arise from this paper probably are amongst the most promising. First, it points out to a future research on the role that government interventions, through regional policies and federal decentralization, have on cities' development. The searching of natural-experiments, such as the emergency of the regional governments (Comunidades Autónomas) in Spain at the beginning of the 80s, could be an appropriate strategy to disentangle the effects that agglomeration economies and regional governments have on cities. Last, the unexpected result between the decline in transport costs and a more equal population distribution of the Spanish cities (Figure 4.5) proves to be a puzzling result opposite to the theoretical propositions by Tabuchi and Thisse (2011). Again, further research is needed to analyze to what extent this pattern is robust in other countries or, by contrast, whether the theoretical model does not consider that it can differ between or within countries.

5.1 Bibliography

1. De La Fuente, A. (2010). Infrastructures and Productivity: an Updated Survey. *Barcelona Graduate School of Economics*, 475.

2. Barbero, J. and Zofío, J. L. (2012). The multiregional core-periphery model: The role of the spatial topology. *Networks and Spatial Economics*, 1-28.

3. Borcherding, T.E. and Silberberg, E. (1978) "Shipping The Good Apples Out: The Alchian And Allen Theorem Reconsidered." *The Journal of Political Economy*, 86 (1), 131-138.

4. Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91 (5), 1454–1477.

5. Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16 (4), 309–342.

6. Hummels, D. and Skiba, A. (2004). Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture. *Journal of political economy*, 112 (6), 1384-1402.

7. Imbs, J. and Wacziarg, R. (2003). Stages of diversification. *American Economic Review*, 63-86.

8. Tabuchi, T. and Thisse, J.-F. (2011). A new economic geography model of central places. *Journal of Urban Economics*, 6 (4), 240–252.

Conclusiones (In Spanish)

A lo largo de esta Tesis se han obtenido numerósos resultados empíricos acerca de la de tipología de las ciudades y los patrones de comercio entre las mismas. En esta sección se plasman las principales recomendaciones de política económica que se derivan de estos resultados, así como las futuras líneas de investigación que se desarrollarán a tenor de esta Tesis.

Tras un primer capítulo introductorio, el primer de los artículos de la Tesis se centra en el papel que tienen los costes de transporte sobre los flujos de comercio en distancias muy cortas (alrededor de 150 kilómetros–185€) y cómo, alrededor de las grandes ciudades españolas, surgen áreas de mercado que delimitan sus zonas de influencia geográfica. A la hora de plantear estrategias de crecimiento regional, los responsables políticos deberían ser conscientes de esta concentración del comercio. En este sentido, debido a que algunas ciudades son más predominantes (poseen mayor influencia comercial y jerarquía) que otras dentro del sistema de ciudades españolas, éstas consiguen beneficiarse en mayor medida que el resto cuando se llevan a cabo políticas de infraestructuras públicas. De hecho, mejoras en la dotación de carreteras pueden generar una mayor aglomeración de la actividad comercial en torno a estas grandes ciudades. Este problema se vuelve aún más acuciante si consideramos que algunas ciudades poseen mayores ventajas de localización que otras dentro de la red de carreteras. A modo de ejemplo, para los mismos umbrales de distancia (áreas de mercado), Madrid puede alcanzar distancias más largas, dentro de España, que Barcelona (Figura 2.5, Mapas *a* y *b*), lo que le permite beneficiarse en mayor medida de mejoras en las infraestructuras.[3]

Yendo un paso más allá de esta concentración de la actividad en cortas distancias, la inversión en infraestructuras durante la burbuja inmobiliaria española ha sido fuertemente criticada tras la aparición de estudios que arrojan evidencia sobre las ineficiencias en la asignación de recursos que tal inversión ha conllevado. En este sentido, trabajos como el de De La Fuente, (2010) muestran que la inversión en carreteras ha llegado a alcanzar tasas marginales de retorno negativas. No obstante, bien es cierto que, aunque estos efectos sobre la eficiencia de recursos sean bastante robustos, los efectos positivos

[3]Estos resultados empíricos estarían en línea con previos modelos teóricos para el caso español (Barbero and Zofío, 2012).

de la inversión en infraestructuras no llegan a considerarse correctamente dentro de la literatura. De este modo, la reducción de los costes de transporte a través de las infraestructuras de carreteras en el período 2003-2007, parece impulsar el comercio entre ciudades (Tabla 2.2). Es más, estas infraestructuras llegan incluso a reducir la acumulación de flujos comerciales a nivel municipal (medido a través de un menor coeficiente en el año 2007 para el límite administrativo municipal en la Tabla 2.4).[4] Así mismo, conectar más municipios a través de la red de carreteras ha permitido que más municipios puedan comerciar más productos (variedades) con más destinos (Tabla 2.9), es decir, los municipios han experimentado mejoras en dos dimensiones de la diversificación: la de productos y la de destinos.

Los resultados obtenidos en este primer capítulo abren la puerta a futuros análisis centrados en la dimensión de producto a nivel municipal. En concreto, resulta de interés entender por qué sólo algunos productos son comerciados en cortas distancias, mientras que otros llegan a alcanzar distancias muy largas, creando con ello una jerarquía de ciudades en las que algunas muestran un impacto enorme en el comercio internacional debido a los productos que consiguen comerciar, mientras que otras ciudades únicamente se limitan a abastecer mercados locales. El denominado como *efecto Alchian-Allen* proporciona un marco tanto teórico (Borcherding y Silberberg, 1978) como empírico (Hummels y Skiba, 2004) a través de cuál se puede analizar los motivos que llevan a que algunas ciudades puedan especializarse en productos de alta calidad, permitiéndoles mejorar dentro de la estructura jerárquica de ciudades.[5]

El segundo capítulo estudia el grado de especialización en productos complejos e intensivos en conocimiento de la ciudades, y cómo esta especialización surge gracias a que las (grandes) ciudades consiguen atraer trabajadores de alta cualificación. Si los productos complejos son los que permiten generar crecimiento económico (Hausmann and Hidalgo, 2011), deberíamos de esperar que algunas ciudades crezcan mucho más que otras, lo que provoca la aparición de divergencias entre unas regiones y otras dentro de un mismo país. En términos de política económica, la principal conclusión que se deriva de este análisis es la incapacidad de los representantes políticos para promocionar la especialización en bienes complejos y de alto conocimiento de las ciudades pequeñas. Esta incapacidad surge tan sólo por el hecho de que los trabajadores más cualificados tienen el incentivo a emigrar a las ciudades más grandes en su búsqueda por mejores oportunidades laborales. De ahí que, al llegar una crisis, sean las ciudades más grandes

[4]Ciertamente, en este primer capítulo no llegamos a estudiar hasta qué punto la menor acumulación de comercio a nivel municipal ha venido acompañada por una mayor concentración a nivel de áreas urbanas (áreas metropolitanas), en detrimento del comercio desarrollado por zonas rurales (fuera de las ciudades).

[5]En el momento en el que esta Tesis Doctoral está siendo escrita, esta idea ya está siendo tratada a través de un modelo teórico por el cuál las ciudades se especializan en productos de alta gama (calidad) dependiendo de cómo son los efectos de demanda en el destino cuando se envían productos.

las que mejor pueden responder a la misma tan sólo por el mero hecho de poseer una mayor dotación de capacidades derivadas de estos trabajadores con alto capital humano.

Para evitar estas divergencias de largo plazo, los responsables políticos (a nivel nacional y local) tendrían que centrar sus esfuerzos en generar más capacidades en las ciudades pequeñas, de modo que puedan atraer a trabajadores de alta formación y, con ello, favorecer la especialización de estas ciudades en productos complejos. Estas capacidades pueden generarse no sólo gracias de la creación de grandes complejos de investigación (que requieren de una fuerte inversión inicial y sus efectos positivos son de muy largo plazo), sino favoreciendo las conexiones entre individuos, empresas e instituciones. Dadas las dotaciones actuales de factores en las ciudades, un uso más intenso de las redes de información entre estos agentes, de modo que la interacción entre los mismos sea más intensa, puede resultar una manera mucho más barata y eficiente de que los responsables políticos a nivel local creen nuevas actividades en las ciudades, permitiendo con ello la generación de más capacidades y, por último, el crecimiento económico.

Este capítulo pertenece a un área de investigación muy relevante, esta es, hasta qué punto las ciudades se especializan en función de su "estadio de desarrollo". Estudios previos (Imbs and Wacziarg, 2003) han encontrado evidencia de la existencia de relaciones no-lineales entre el nivel de desarrollo de un país (PIB) y su grado de diversificación. En primeras fases del desarrollo, los países de especializan en productos eminentemente agrícolas, después diversifican su estructura productiva mediante la creación de nuevos sectores para, finalmente, volver a especializarse en productos de alto conocimiento. Esta pauta que se observa a nivel nacional resulta de especial interés analizarla para el caso de las ciudades. Si a nivel urbano se observa un patrón similar que en el conjunto del país, podríamos inducir que la actual especialización de las ciudades serviría para predecir patrones futuros de especialización de los países. De ahí que se deberían impulsar proyectos de investigación que analicen con detalle la estructura productiva de las ciudades.

El último capítulo arroja luz sobre la aparición de ciudades y cuáles son los factores que llevan a este fenómeno. Quizás la implicación más importante de este capítulo sea la de hacer una llamada de atención a la agenda investigadora en su conjunto y, más concretamente, a la necesidad por realizar futuras investigaciones, tanto teóricas como empíricas, relacionadas con la aparición de economías de co-aglomeración. Dentro de la literatura no está claro por qué algunas ciudades presentan ventajas tanto en diversificación como en especialización (economías de co-aglomeración). Una mayor y mejor compresión de este tipo de fenómenos serviría para entender el rol que juegan tanto las economías de aglomeración como las interacciones entre ciudades en la aparición de este tipo de ciudades. Todo ello con el intento de ayudar a los responsables políticos a encontrar las *ventajas comparativas* a nivel local.

En línea con el segundo capítulo, las implicaciones derivadas de éste tercero suponen ser un reto para las intervenciones de política económica. Por un lado, la aparición de ciudades está dirigida por las economías de aglomeración internas, especialmente aquellas relacionadas con el tipo de trabajadores y el conocimientos (formación) que éstos poseen. Por otra parte, los patrones de dependencia del pasado (path dependence) parecen jugar un papel fundamental en el desarrollo urbano. En concreto, previos niveles de población y costes de transporte ("ventajas de primera naturaleza") suponen ser aquellos que en mayor medida determinan la tipología de las ciudades en el futuro. Al cumplirse estos efectos, el espacio que se deja a la intervención de políticas resulta ser muy reducido a la hora de impulsar el crecimiento en las ciudades. De hecho, puesto que bajos costes de transporte en el pasado parecen afectar positivamente a la especialización futura de las ciudades, los responsables políticos continuamente se estarían enfrentando a disyuntivas (trade-offs). Si las intervenciones políticas van encaminadas a reducir los costes de transporte con el objetivo de impulsar el comercio de las ciudades (implicación del primer capítulo), éstas tendrían que hacerlo a expensas de saber que con ello favorecerán la aparición de ciudades especializadas, generando así un patrón de centro-periferia *à la NEG* dentro del país, y enfatizando de este modo las desigualdades entre ciudades. Siguiendo con esto, si las ciudades diversificadas son aquellas que promueven el crecimiento económico en el largo plazo, mientras que las ciudades especializadas únicamente se centran en la producción de productos estandarizados (Duranton y Puga, 2001), el hecho de reducir los costes de transporte generaría que algunas áreas (ciudades especializadas) no puediesen crecer de la misma manera que otras (ciudades diversificadas).

Las líneas de investigación que surgen a raíz de este artículo posiblemente se encuentran entre las más interesantes a explorar a futuro. En primer lugar, la agenda investigadora tendría que profundizar en el rol que la intervención gubernamental puede tener en el desarrollo de las ciudades a través de políticas de crecimiento regional o mediante el propio federalismo fiscal. Así, la búsqueda de cuasi-experimentos naturales, como la creación del Estado de las Autonomías en España al principio de los años 80, podría servir como estrategia empírica con la que desentrañar y diferenciar los efectos que tanto la aglomeración económica como los pertenecientes a la acción de los gobiernos regionales tienen sobre la estructura productiva de las ciudades. Por último, los resultados obtenidos entre la caída de los costes de transporte y la reducción de las desigualdades en la distribución de la población en España (Figura 4.5), abren la puerta a futuras investigaciones que analizen este inesperado y contradictorio patrón, pues va en contra de proposiciones teóricas planteadas en trabajos de referencia como el de Tabuchi and Thisse (2011). En este sentido, la investigación tendría que centrarse en estudiar si dicha pauta tiene lugar sólo en España, o por el contrario, se observa en más países. De ser así, la literatura teórica estaría fallando, lo que requeriría de una nueva formulación de modelos acerca de la evolución jerárquica de las ciudades.

5.2 Bibliografía

1. De La Fuente, A. (2010). Infrastructures and Productivity: an Updated Survey. *Barcelona Graduate School of Economics*, 475.

2. Barbero, J. and Zofío, J. L. (2012). The multiregional core-periphery model: The role of the spatial topology. *Networks and Spatial Economics*, 1-28.

3. Borcherding, T.E. and Silberberg, E. (1978) "Shipping The Good Apples Out: The Alchian And Allen Theorem Reconsidered." *The Journal of Political Economy*, 86 (1), 131-138.

4. Duranton, G. and Puga, D. (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91 (5), 1454–1477.

5. Hausmann, R. and Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, 16 (4), 309–342.

6. Hummels, D. and Skiba, A. (2004). Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture. *Journal of political economy*, 112 (6), 1384-1402.

7. Imbs, J. and Wacziarg, R. (2003). Stages of diversification. *American Economic Review*, 63-86.

8. Tabuchi, T. and Thisse, J.-F. (2011). A new economic geography model of central places. *Journal of Urban Economics*, 6 (4), 240–252.