TRABAJO FINAL DE MÁSTER

EU trade and economic integration: Revisiting border effects inside the European Union

ALUMNO: Guillem Rochina Aguas

TUTOR: Prof. Francisco Requena Silvente

Máster en Internacionalización Económica: Gestión del Comercio Internacional

Instituto de Economía Internacional

Curso: 2022-2023

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

I. INTRODUCCIÓN

En el presente documento se incluyen los dos componentes necesarios para la entrega de un trabajo de final de máster completo.

Primeramente, se detalla una descripción de la empresa donde he realizado las “prácticas profesionales” como Junior Data Scientist en el equipo de Advanced Analytics (y a partir del 1 de septiembre me incorporo con contrato indefinido como Data Scientist), cómo encontré dichas prácticas y cuál ha sido mi camino hasta llegar donde me encuentro ahora mismo en mi carrera profesional, así como las tareas, responsabilidades y otras actividades adicionales que he realizado dentro de la compañía. En segundo lugar, presento una valoración de mis prácticas, así como todas las cosas que me ha proporcionado el Máster en Internacionalización Económica y el Instituto de Economía Internacional de la Universitat de València en mi vida profesional.

A continuación, una vez presentado el resumen de mi carrera tras el primer año del máster, dado el carácter eminentemente académico de mi trabajo, se incluye un Working Paper, previo envío para su evaluación, producto de todo el trabajo de investigación realizado con la ayuda y bajo la supervisión del profesor Francisco Requena.

En concreto, el artículo presentado consiste en un análisis del proceso de integración europea desde el punto de vista del comercio internacional a nivel de industria. Haciendo uso del concepto de “efecto frontera” (entendida a grosso modo como la magnitud relativa en la que un país comercia más consigo mismo que con sus socios comerciales), la ecuación de gravedad estructural del comercio y los avances experimentados estos últimos años tanto a nivel conceptual como econométrico, se analiza cómo ha ido evolucionando la magnitud del efecto frontera desde varios niveles (a nivel de la Unión Europea como entidad única, a nivel de países miembros del Mercado Único, y por último a nivel de industria) y se estudia el impacto de posibles fuentes que influyen en dicho efecto a nivel de producto/sector.

II. ACTIVIDAD PROFESIONAL

1. ¿Cómo he llegado hasta aquí?

Para contextualizar correctamente cómo he llegado a este punto en mi vida profesional nos debemos remontar al año en qué estalló la pandemia del Covid-19. Unos meses antes del inicio del confinamiento había terminado mis prácticas de grado como Analista de Mercados en el Instituto Tecnológico del Mueble (AIDIMME). Durante el transcurso de estas prácticas me había encargado de realizar informes de análisis cualitativo y cuantitativo para distintos organismos públicos y privados, como la Junta de Extremadura o FEVAMA (Federación Empresarial de la Madera y mueble de la Comunidad Valenciana).

Fue en este punto donde me di cuenta de que la investigación de mercados y, sobre todo, la vertiente del análisis cuantitativo era el tipo de carrera profesional que se adaptaba perfectamente a lo que buscaba dedicarme en el futuro. Por desgracia, no había suficiente presupuesto para poder quedarme en el departamento y los puestos de investigación de mercados en las empresas estaban siendo reemplazados por nuevas posiciones mucho más técnicas llamadas "Data Analyst" o "Data Scientist", por lo que empecé a repasar y ampliar toda la estadística que había estudiado durante la carrera, aprendí a programar en varios lenguajes y decidí apuntarme a un posgrado de análisis de datos que me permitiera compaginarlo con la búsqueda de trabajo.

Pocos meses después, estalló la crisis del Covid-19, empezaron los periodos de confinamiento y con ello mi búsqueda de empleo se vio completamente truncada. Tras quedarme a las puertas de varios procesos de selección y observar que todas las posiciones a las que podía optar dentro del mundo del análisis de datos implicaban mudarse a Madrid o Barcelona en época de COVID, finalmente decidí aceptar una posición como asesor de banca personal en Caixa Popular con la intención de ir saltando poco a poco a puestos más enfocados en Inteligencia de Negocio y Análisis.

Es durante estos meses trabajando en el banco, mientras estudiaba los contenidos del examen del MIFID, investigaba sobre la certificación CFA y continuaba haciendo proyectos personales relacionados con el análisis de datos, donde realmente entiendo que la parte que más disfruto de un análisis es cuantificar y entender las implicaciones de esos valores dentro de un contexto concreto, no el proceso en sí, es decir, para mí los algoritmos, la estadística y los métodos numéricos son sólo herramientas para llegar a un resultado que me interesa, no un objeto de estudio per se. A partir de ese momento, me planteo la idea de realizar un doctorado en el campo de la economía con el objetivo de acceder como analista en alguna institución como el BCE o la OMC.

Tras investigar varias opciones, encontré el Máster en Internacionalización Económica del IEI y, viendo las actividades profesionales de algunos de los miembros egresados dentro de la Universidad y de otras instituciones públicas, me animé a apostar por el IEI como punto de partida para mi camino hacia el mundo de la investigación, con la suerte de que pocas semanas después recibí la notificación por parte del profesor Vicente Pallardó sobre mi admisión en el máster.

Tras finalizar el año académico, por vicisitudes de la vida, si quería dedicarme a la investigación debería compaginarlo con un trabajo al menos a tiempo parcial y, con suerte, a largo plazo encontrar una beca o contrato temporal con la que poder enfocarme completamente al mundo académico.

Después de hablar con el profesor Vicente Pallardó, se me permitió realizar un TFM con un enfoque más académico, aunque estuviera trabajando en una empresa y no necesariamente dedicándome a la investigación. En primer lugar, traté de encontrar alguna posición de prácticas en instituciones como Caixabank Research, pero con el tiempo empecé a aplicar a otras posiciones más relacionadas con Supply Chain y Operaciones para tener un abanico más amplio de posibilidades.

Por una serie de casualidades acabé aplicando a una posición de prácticas en Supply Chain en Sanofi-Aventis. En la primera entrevista con la técnico de selección de Addeco me indicaron que ese puesto estaba dirigido a personas con formación más relacionada con ingeniería industrial que con el ámbito económico pero, dado que tenía un posgrado en análisis de datos y mi TFM en el máster del IEI estaba relacionado con la investigación, poseía un perfil interesante para una vacante que se había quedado sin cubrir dentro del Programa de Talento de este laboratorio farmacéutico, por lo que acepté continuar con las siguientes entrevistas para este nuevo puesto.

Tres días antes de realizar la entrevista con los que serían mis “tutores” dentro de la empresa, me llegó un correo indicándome sobre qué temas a grandes rasgos me iban a preguntar durante la entrevista. Entre varios temas, se me indica que se me iban a hacer preguntas sobre algoritmos de ML básicos (Árboles de decisión, Modelos lineales, etc.), Estadística y Programación, en concreto un tipo de paradigma de programación que no había utilizado demasiado antes, la programación orientada a objetos.

La cuestión es que durante esa semana estaba en el norte de Cataluña visitando a unos familiares, pero, por suerte me había traído todos los apuntes relacionados para poder hacer una entrevista decente en caso de que me preguntaran cosas específicas del puesto. Durante esos tres días, dado que la entrevista iba a ser más técnica de lo que en un principio esperaba, buscaba cualquier momento muerto para leerme de arriba abajo todos los apuntes que había traído, llegando incluso a leerme un libro entero sobre programación orientada a objetos en Python para poder responder a esas preguntas con confianza.

Llegó el día de la entrevista y, a pesar de estar nervioso por ser mi primera entrevista técnica en inglés, tuve la suerte de poder manejar y responder todas las preguntas e incluso acordarme de conceptos que no había si quiera considerado cuando me preparaba las cosas que me podían preguntar. En conclusión, las sensaciones tras la entrevista fueron buenas y, tras varias semanas de incertidumbre y nervios finalmente llegó la noticia de que había sido seleccionado para el puesto, con unas condiciones que se me hacían difíciles de rechazar.

Tras los primeros momentos de euforia, aparece un elemento común entre muchos trabajadores de los sectores IT, el síndrome del impostor. Un rápido vistazo a los perfiles de LinkedIn de los que serían mis compañeros me indica que voy a compartir equipo con doctores en biotecnología, computación, matemáticos, estadísticos o físicos y entonces aparece la pregunta: ¿Voy a poder estar a la altura?, ¿Estoy apuntando demasiado alto?

Tras varios días de luchas internas (y externas) y de conversaciones con muchas personas de mi entorno, decido acallar todas las inquietudes al respecto y aprovechar esta oportunidad.

Finalmente, me mudé a Barcelona y empecé mis primeros días en Sanofi.

Como añadido a esta introducción, aprovecho para agradecer a los profesores Francisco Requena y Vicente Pallardó sus ánimos y sus palabras de confianza para empujarme a no perder esta oportunidad, pues también tuvieron peso en mi decisión.

2. Descripción de la Empresa

Dado que mis prácticas empiezan en Sanofi-Aventis, pero a lo largo del año mi contrato se traspasa a una spin-off de la compañía especializada en Consumer Healthcare, a continuación, se presenta un breve resumen de ambas compañías y los productos que comercializan ambas entidades.

2.1 Sanofi-Aventis

2.1.1 Historia y descripción de la empresa

Sanofi-Aventis es un laboratorio farmacéutico multinacional francés con sede en Chentilly (Francia), resultado de una serie de fusiones que finaliza con la unión de dos empresas de gran envergadura, Sanofi-Synthélabo y Aventis.

Su origen se remonta a la década de los 70, momento en el que Sanofi se establece como la filial de la petrolera francesa Elf Aquitaine. En sus orígenes, Sanofi se centra en la producción de productos químicos y agrícolas para más adelante incursionarse en el sector farmacéutico gracias a su potencial crecimiento y su capacidad para desarrollar productos relacionados con la salud de las personas.

A su vez, en la misma década se funda Synthélabo en Francia, como una empresa especializada desde sus inicios en productos farmacéuticos. Enfocada en la investigación y desarrollo de medicamentos innovadores para abordar diversas enfermedades y condiciones médicas, en poco tiempo logra establecer una sólida reputación dentro de la industria farmacéutica gracias a priorizar los esfuerzos científicos y la calidad de sus productos.

En 1999, dentro de un contexto de consolidación del mercado farmacéutico, Sanofi y Synthélabo deciden fusionarse para crear Sanofi-Synthélabo, logrando situarse como una de las empresas más grandes y diversificadas de su sector. Esta fusión permite a la nueva compañía aprovechar las fortalezas y capacidades de ambas entidades, combinando la experiencia en productos químicos y agrícolas de Sanofi con el enfoque científico en investigación y desarrollo de medicamentos de Synthélabo.

Paralelamente, ese mismo año ocurre otra importante fusión dentro de la industria, con la unión de la empresa alemana Hoeschst AG y la francesa Rhône-Poulenc S.A., las cuales forman Aventis, a fin de lograr establecer en el mercado otra compañía de gran envergadura, una amplia cartera de productos y una relevante presencia mundial.

Finalmente, en 2004 se produce la fusión de Sanofi-Synthélabo y Aventis, creando la compañía que ahora se conoce como Sanofi-Aventis, o Sanofi para abreviar. La fusión de estas dos compañías líderes en el sector fue un hito en la historia de la industria farmacéutica y logró que Sanofi se posicionara entre los 10 laboratorios farmacéuticos más grandes del mundo.

Desde su unión, la nueva entidad ha estado enfocada en fortalecer su presencia global, expandir su cartera de productos y apostar por el desarrollo de medicamentos innovadores que supongan un breaktrhough en el sector farmacéutico.

Con el tiempo, la compañía ha seguido evolucionando y adaptándose al dinamismo de la industria en la que compite, estableciendo asociaciones estratégicas, adquiriendo empresas de menor dimensión y enfocándose en soluciones terapéuticas avanzadas, con especial foco en "Specialty Care" o Enfermedades Raras, donde es y ha sido pionera desde hace décadas. No obstante, uno de los hitos más relevantes relacionados con el puesto que ocupo en la empresa está relacionado con la creación de Global Innovation Center en 2016, un hub tecnológico que se encarga de llevar a cabo la transformación digital de la compañía a nivel mundial, el cual describo brevemente en la subsección **2.2.3**.

2.1.2 Divisiones y productos

Tras todas las fusiones, adquisiciones y ampliaciones de cartera de producto, Sanofi se encuentra presente en una gran variedad de categorías terapéuticas. A continuación, destaco las más relevantes:

- General Medicines - Diabetes: La división de diabetes de Sanofi se enfoca en desarrollar medicamentos y soluciones para el tratamiento de la diabetes mellitus, una enfermedad crónica caracterizada por niveles elevados de azúcar en sangre. En concreto, los productos más relevantes en esta línea incluyen insulinas de acción prolongada y combinaciones de insulina con agonistas del receptor de GLP-1. En esta categoría destacan:
  - Lantus, insulina de acción prolongada para la diabetes tipo 1 y tipo 2.
  - Toujeo, similar a Lantus, pero con una concentración más alta.
  - Soliqua, combinación de insulina y agonista del receptor GLP-1 para la diabetes tipo 2.
- General Medicines - Enfermedades cardiovasculares: Esta división se enfoca en medicamentos que reduzcan el riesgo de eventos cardíacos y mejoren el perfil lipídico. Dentro del portafolio se incluyen medicamentos en más de 20 áreas terapéuticas y 40 de los fármacos se encuentran incluídos en la lista de medicamentos esenciales de la Organización Mundial de la Salud. Entre sus productos descatan:
  - Plavix, antiagregante plaquetario enfocado en reducir el riesgo de eventos cardiovasculares en pacientes con enfermedad arterial coronaria, infarto de miocardio o accidentes cerebrovasculares.
  - Praluent, utilizado para reducir el colesterol LDL en pacientes con hipercolesterolemia familiar o enfermedades cardiovasculares derivadas.
- Vacunas: Destacando Sanofi-Pasteur, la división de vacunas del laboratorio, Sanofi es uno de los líderes en desarrollo de vacunas para la prevención de enfermedades infecciosas. Entre la cartera de productos se encuentran las primeras vacunas de la infancia y de refuerzo, así

como las de la gripe, la COVID-19 y otras endémicas o para viajes. Entre sus desarrollos encontramos:

  - Fluozone y derivados, es decir, vacunas para proteger contra la Influenza o gripe estacional.
  - Menactra, vacuna para proteger contra la meningitis meningocócica.
  - Finalmente, cabe destacar que Sanofi presenta el primer anticuerpo monoclonal, autorizado por la Comisión Europea, para la protección de los lactantes durante su primera temporada de virus respiratorio sincitial. Este hecho permitirá proteger a los neonatos de la gripe desde los primeros compases de su vida en los que no pueden ser vacunados todavía.

- Specialty Care: La unidad especializada en medicamentos focalizados en inmunología, enfermedades raras, oncología y esclerosis múltiple. Dentro de esta área destacan las enfermedades raras y enfermedades inflamatorias tipo 2, donde destaca el producto estrella de Sanofi:

  - Dupixent, medicamento biológico aprobado para el tratamiento de la dermatitis atópica moderada y grave en adultos y adolescentes, así como otras enfermedades inflamatorias. Este producto, si nos basamos en la matriz de la Boston Consulting Group (BCG), puede considerarse como un producto estrella que ya ha iniciado su proceso hacia un producto vaca, pues ya supone una gran fuente de ingresos para la compañía, pero a su vez, dicho producto puede ser utilizado para tratar otro tipo de enfermedades para las cuales todavía no se receta (se prevé que en un futuro próximo lo sea).

- Consumer Healthcare: División en la que trabajo yo, y que actualmente se ha constituido como entidad legalmente independiente dentro del grupo Sanofi. En la siguiente subsección se presenta un breve resumen de esta división y sus productos.

2.2 Opella Healthcare / Sanofi CHC

2.2.1 Origen de la Spin Off

Sanofi Consumer Healthcare es la división de Sanofi enfocada en productos de salud del consumidor, es decir, productos que se venden sin necesidad de receta médica, llamados productos over-the-counter (OTC) o coloquialmente productos de autocuidado.

A partir de la entrada del nuevo CEO de Sanofi, el británico Paul Hudson, un nuevo rumbo se decide para la división de CHC. Siguiendo la dinámica del resto de biofarmacéuticas, se decide llevar a cabo la escisión del negocio OTC en una compañía independiente, bajo el pretexto de que este tipo de divisiones compiten en un mercado completamente distinto a los productos propios de un laboratorio farmacéutico.

En este contexto, se acuña el concepto de "Fast Moving Healthcare Goods" (FMHG) a raíz del reconocimiento de la consolidación de un mercado, dentro del sector de productos farmacéuticos, con unas características mucho más similares a los ya bien conocidos mercados de "Fast Moving Consumer Goods", más enfocados en la relación directa con el cliente final, que a aquellos mercados en los que los productos están sometidos a legislaciones mucho más rigurosas y a procesos de tenders donde las estrategias de comunicación, distribución y pricing, por nombrar algunas de ellas, están mucho más enfocadas a las instituciones públicas, médicos y asociaciones de pacientes.

A fin de ofrecer mayor autonomía al negocio de Consumer Healthcare y ser capaces de competir en un mercado mucho más dinámico que al que están acostumbrados en la industria biofarmacéutica, se crea Opella Healthcare, Spin off con identidad jurídica propia que actualmente se mantiene dentro del grupo Sanofi.

2.2.2 Productos

Dentro de Opella Healthcare encontramos gran variedad de productos exitosos a nivel local. Por ejemplo, Lizipaina (adquirida al laboratorio Boeringher Ingelheim) es una marca muy reconocida en España y Francia (Lysopainne en Francia). No obstante, más allá de las fronteras de estos países no son productos reconocidos por los consumidores, actuando estas como soporte para potenciar las marcas globales de Opella.

Así, en línea con la estrategia actual de potenciación de las marcas globales, a continuación, se presentan los productos más representativos a nivel mundial de Sanofi CHC:

- Allegra/Telfast: Antihistamínico utilizado para aliviar los síntomas de alergias estacionales, como la rinitis alérgica, incluyendo estornudos, picazón de ojos y nariz o secreción nasal.
- Dulcolax/Enterogermina: Laxante suave que ayuda a aliviar el estreñimiento ocasional y a facilitar la evacuación intestinal.
- Dorflex/IcyHot: Productos enfocados en aliviar los síntomas de dolores corporales y de cabeza derivados de la migraña, fiebre, dolor menstrual etc.
- Buscopan: Antiespasmódico que busca relajar los músculos del tracto gastrointestinal y otros órganos internos en busca de aliviar espasmos y calambres que puedan surgir en estas zonas del cuerpo. Principalmente se enfoca en tratar cólicos, colitis y otros problemas digestivos que causen dolor abdominal.
- Mucosan/Bisolvon: Utilizado como expectorante y mucolítico para el tratamiento de afecciones respiratorias como la bronquitis, neumonía y otras enfermedades pulmonares obstructivas.
- Pharmaton y derivados: Gama de productos enfocados en el bienestar físico y mental, aportando energía, mejorando la calidad del sueño y la salud de las articulaciones mediante compuestos vitamínicos y suplementación alimentaria.

2.2.3 ¿Dónde se sitúa mi departamento?

Mas allá de las divisiones por producto, existen ciertas funciones que son transversales para toda la compañía, pues la estructura de Sanofi debe entenderse que se asemeja a una matricial (sin ajustarse completamente a esta pues la dirección de un departamento tiene una única jerarquía y no existe ningún “choque” de decisiones provenientes de jefes con misma autoridad).

Dentro de este abanico de funciones transversales encontramos el Global Innovation Center, compuesto de equipos internacionales que ofrecen servicios para fomentar la innovación y la transformación digital en Sanofi y Opella por igual.

Entre estos servicios encontramos programas como Ishift o SCCore, encargado de optimizar los procesos financieros, de producción y de Supply Chain para conseguir una minimización de costes y una mayor eficiencia de las cadenas de suministro. No obstante, mis actividades se encuentran dentro del área llamada Digital, la cual engloba los departamentos de Business Analytics, Digital Manufacturing o Digital Customer Experience. En concreto, dentro de Business Analytics, me encuentro en el equipo de Advanced Analytics and AI, encargado de solucionar problemas de negocio mediante el uso de algoritmos de Machine Learning y otros métodos relacionados con la estadística y la investigación operativa.

3. Responsabilidades y Actividades

En el presente apartado se presentará un resumen a alto nivel de los proyectos en los que he participado dentro del departamento de Advanced Analytics, así como los roles que he ido asumiendo a lo largo de mi primer año en el equipo y otras actividades complementarias que han aportado valor a mis primeros pasos dentro de la compañía. Finalmente, se presenta una valoración de mi evolución profesional durante este año, enlazándolo con elementos del máster que considero me han proporcionado un plus que de otra forma no hubiera podido aprovechar.

3.1 Resumen ejecutivo de los proyectos de Advanced Analytics

A continuación, se presenta un resumen simplificado de los proyectos en los que he estado en mayor o menor medida involucrado a lo largo de este año, contextualizando el objetivo y las razones de su existencia, así como aportado un resumen de las funcionalidades que ofrecen.

3.1.1 Allocations and Promotions (A&P)

El proyecto de Allocations and Promotions (A&P) responde a una pregunta clave para cualquier negocio dentro de mercados de tipo FMCG: ¿Dónde debo gastar mi presupuesto de marketing para maximizar el impacto en mis ventas?

En un contexto donde los consumidores centran cada vez más su atención a los medios digitales, el rango de touchpoints (puntos de contacto con potenciales clientes) se expande de forma masiva, provocando que una combinación efectiva y eficiente dentro de un ecosistema multicanal sea cada vez más complejo.

Dentro de este contexto, cobra sentido una herramienta que, mediante un modelo fundamentado en el paradigma estadístico bayesiano (actualmente accesible gracias a los avances en computación y la opción de realizar simulaciones con métodos de Monte Carlo) y la programación entera, proporcione recomendaciones a los equipos de marketing con respecto a cómo deberían repartir su presupuesto entre productos, touchpoints y mercados, a fin de "liberar" inversiones en marcas saturadas y redirigirlos a innovaciones o marcas que tienen mayor potencial estratégico. En pocas palabras, este proyecto entra dentro de las técnicas aplicadas de modelado estadístico conocidos como Marketing Mix Modelling.

A grandes rasgos la herramienta desarrollada funciona de la siguiente forma:

1. Se recolectan datos incluyendo 3 años de histórico, con granularidad semanal, sobre la evolución de ventas de producto, junto con el gasto y las métricas de impacto de cada uno de los touchpoints utilizados por marca y producto.

2. Se computan regresores que se introducirán en el ajuste del modelo final, incluyendo el cálculo de la elasticidad precio por producto, la estacionalidad del histórico o el "efecto del Adstock" (el impacto residual de arrastre que tiene una campaña de promoción o publicidad incluso después de que haya finalizado).

3. Basado en estos regresores y el conjunto de datos originales, un modelo de marketing mix (MMM) basado en modelos bayesianos jerárquicos proporciona un output basado en curvas

de respuesta por producto y touchpoint, las cuales representan el volumen incremental de ventas por niveles de gasto en dicho punto de contacto con el cliente.

4. Las curvas resultantes son introducidas en un solucionador basado en programación entera para encontrar oportunidades de optimización. En este paso también se incluye, en forma de restricción o como coeficiente multiplicativo, información relativa a la importancia estratégica de los productos y los touchpoints, así como información de los planes estratégicos de cada división de negocio por país y producto.

5. En el momento la solución termina por ser implementada, los usuarios directos del proyecto (personal de Marketing, Media y otros departamentos comerciales) pueden llevar a cabo simulaciones utilizando una interfaz de usuario, lo que permite a los equipos locales crear y comparar escenarios con distintas distribuciones presupuestarias. En este caso, los usuarios tienen total libertad para influir en el ajuste del modelo mediante la introducción de objetivos específicos de negocio (por ejemplo, una unidad de negocio puede tener como objetivo maximizar el CAAP en un producto y maximizar el volumen de sell-out en otro) y restricciones que observan en su unidad local (Por ejemplo, si una unidad de negocio local ya ha pactado cierto presupuesto a largo plazo para realizar campañas publicitarias en televisión, esta restricción puede ser incluida para ofrecer una visión más fidedigna del futuro simulado).

6. Más allá del simulador, la contribución de cada touchpoint se muestra en un gráfico de cascada, indicando el porcentaje de ventas o volumen que es consecuencia directa de invertir en un touchpoint concreto. En este también se incluye un baseline, es decir, el volumen que no puede ser explicado por ningún dato utilizado como regresor externo y se asume es debido a la “Brand Equity” u otras características no tomadas en consideración en el ajuste de la serie temporal.

7. Por último, el modelo final también presenta un informe de las curvas de respuesta para cada touchpoint por producto. En este caso, dadas las características de los datos ajustados, se proporcionan curvas que indican el volumen incremental de ventas por nivel de gasto de

marketing (también llamado ROS o Return on Sales) o bien curvas que representan el ROI (Return on Investment) con respecto al gasto en las campañas de marketing.

En conclusión, el proyecto de Allocations and Promotions consiste en un pipeline analítico que pretende ofrecer recomendaciones a equipos de marketing locales con respecto a cómo deben redistribuir sus presupuestos a fin de optimizar esfuerzos y maximizar de forma más ágil los principales KPIs de cada unidad de negocio local.

3.1.1 Trigger Models / Precision Marketing Forecast (TM-PM)

El proyecto de Trigger Models nace dentro de un contexto donde la publicidad tradicional está perdiendo peso con respecto a la publicidad dentro de los nuevos canales digitales, mucho más dinámicos y que permiten ajustar de forma mucho más rápida cambios en la intensidad y duración de las campañas de promoción y comunicación. En este contexto, la publicidad entra en un nuevo paradigma, llamado Precision Marketing, donde:

1. No se busca alcanzar grandes audiencias, sino que se busca enfocarse en targets concretos mediante un previo análisis de datos relevantes.
2. No se busca generar un contenido que pueda atraer al mayor número de consumidores posible, sino que se busca ofrecer un contenido personalizado en base a una segmentación mucho más granular de grupos de consumidores.
3. No se busca mantener la misma intensidad a lo largo de la campaña, sino que lo que más interesa es orientar el gasto de las campañas para los momentos exactos donde mayor impacto tendrán.

Es en este último punto donde entra en juego el proyecto de Trigger Models. Si A&P responde a la pregunta "¿Dónde he de gastar mi dinero en la campaña?", TM responde a "¿Cuándo he de gastar mi dinero en la campaña?": En líneas generales, los Trigger Models consisten en forecasts semanales con especial foco en el corto plazo, que permiten a los equipos de Medios Digitales adaptar su gasto en publicidad y promoción en base a las predicciones de demanda de los consumidores.

La herramienta permite diseñar un forecast completamente personalizable a distintos niveles de granularidad para todas las dimensiones consideradas, ya sea a nivel geográfico (por País, Región, Ciudad…), a nivel de producto (Categoría, Marca, SKU…) o a nivel temporal (de 2 a 36 semanas de predicción).

A grandes rasgos la herramienta desarrollada funciona de la siguiente forma:

1. Se recopilan todos los históricos de ventas por región y producto y se analiza mediante test estadísticos el potencial de predicción (o "forecastibility") de las series temporales dependiendo de la presencia de estacionalidad, tendencia y nivel de volatilidad.

2. Dado que se trata de un ajuste multivariante, se analiza la potencial influencia de otras series temporales mediante correlaciones lineales, el coeficiente de información máxima (para capturar relaciones no lineales) y variantes de la prueba de causalidad de Granger. Dado que tratamos con productos farmacéuticos, por norma general hacemos uso de históricos de datos meteorológicos, índices de polen o de gripe, aunque en algunos casos utilizamos también datos macroeconómicos como la evolución semanal del PIB (obtenida a través de los resultados de un modelo desarrollado por analistas de la OCDE). Asimismo, capturamos la intención de compra de ciertos productos a través de extraer información sobre las búsquedas en Google de ciertas palabras clave (i.e. Google Trends).

3. Una vez obtenida toda la información relevante, se lleva a cabo el ajuste de uno o varios modelos de Machine Learning que, mediante un método de optimización basado de nuevo en un enfoque bayesiano, minimiza una función de coste propia de los modelos de regresión. En concreto, se suele hacer uso del MAPE (Error porcentual absoluto medio), común en las herramientas de forecasting, o el RMSE (Raíz del error cuadrático medio).

4. Tras obtener un modelo ajustado, se lleva a cabo la simulación de una campaña de predicción

de ventas con el conjunto de datos del último año, el cual ha sido previamente separado del resto de datos que el modelo ha observado para realizar el cálculo de sus parámetros. Este paso se realiza para comprobar cuál sería el rendimiento real del algoritmo en caso de que hubiera sido preparado para una campaña de forecasting ese año en concreto, a modo de una suerte de "prueba de robustez" que nos permite asegurarnos que el modelo que hemos desarrollado ha capturado un "patrón general" en los datos y nos dará una imagen fidedigna de lo que ocurrirá en el futuro. Por supuesto, esta prueba no exime al algoritmo de fallar en sus predicciones para el próximo año, pues estamos asumiendo, como cualquier proyecto de forecasting, que el comportamiento de los datos de la serie temporal seguirá siendo el mismo el próximo año y, generalmente esto es habitual en productos estacionales como las marcas de Alergia, pero no tan común en productos con menor dependencia estacional como pueden ser los productos Antiinflamatorios, Digestivos etc.

5. En caso de que las simulaciones nos den unas métricas suficientemente satisfactorias, teniendo en cuenta la variabilidad de la serie, el modelo se programa para que las predicciones para el próximo período sean realizadas semanalmente el día y hora que haya establecido el equipo de marketing local o los usuarios que hayan solicitado recibir la predicción.

6. Una vez programado, los usuarios reciben un correo semanal con las predicciones para los distintos horizontes solicitados (2 semanas, 4 semanas, 3 meses...) hasta que se alcanza la fecha acordada, momento en el que el modelo se pausa y se elimina de nuestros servidores. A partir de este punto se produce una disyuntiva en cuanto a qué camino tomar. Un proyecto completo de Precision Marketing implicaría que las predicciones fueran utilizadas como input, junto con las restricciones presupuestarias y los canales de comunicación escogidos, en un optimizador similar al que encontramos en el proyecto de Allocations and Promotions, a fin de proporcionar recomendaciones semanales relacionadas con la repartición presupuestaria de los gastos en comunicación digital. No obstante, dado que al fin y al cabo nos encontramos con un proyecto de forecasting bastante estandarizado y otras áreas de negocio están demandando este servicio, el proyecto actualmente se encuentra en un proceso de

mutación donde el principal objetivo de desarrollo implica implementar mejoras sustanciales a su parte matemática, a fin de mejorar al máximo la precisión de los pronósticos de ventas proporcionados y aprovechar la versatilidad que esta herramienta ofrece para muchos departamentos de la compañía.

3.2 Roles asumidos dentro de los proyectos

Si bien es cierto que mi posición en el equipo era de Junior Data Scientist, los acontecimientos que hemos vivido a lo largo del año con la reestructuración del equipo (incluyendo varios abandonos) debido a la escisión de Sanofi CHC y su constitución como Opella, han implicado que asuma gran variedad de roles y tome las riendas del proyecto de forecasting presentado. Seguidamente, presento un resumen de lo que ha implicado cada uno de roles que he debido tomar.

3.2.1 Analytics translator y punto de contacto con negocio

Como analista (o cualquiera de sus acepciones: Data Analyst, Data Scientists, ML Engineer, BI Analyst) puedes haber desarrollado una combinación de algoritmos, fuentes de datos e ingeniería de características infalible que sea capaz de proporcionar predicciones muy precisas pero, si "negocio" (entiéndase los clientes internos que deben usar tu solución) no son capaces de entender tu herramienta o no llegan a entender la utilidad o impacto positivo que puede tener sobre alguno de sus objetivos, ni siguiera la etiqueta "Inteligencia Artificial" les va a convencer para que la decidan implementar en sus operaciones.

Dado que el mundo de la "Inteligencia Artificial" y la "Ciencia de Datos" como la entendemos ahora tiene todavía muy poco recorrido, existen ciertos rasgos en el ámbito profesional son completamente ignorados a pesar de su notable importancia. Tan importante es que un analista entienda los entresijos de la solución que se está implementando, como que sea capaz de explicar a personas no técnicas su funcionamiento. Aquí es donde entra en juego el rol de "Analytics translator".

Asumiendo este rol en el proyecto de los Trigger Models, me he encargado de que las personas que hacían uso de esta herramienta entendieran su funcionamiento, sus ventajas, sus limitaciones y resolver cualquier duda técnica que tuvieran. Asimismo, durante dichas conversaciones, cualquier comentario o sugerencia que aflorara la anotaba para incluirlo dentro de la lista de mejoras a desarrollar, siendo estas conversaciones una fuente de feedback constante para optimizar la herramienta.

3.2.2 Delivery lead

En mi equipo se conoce como "Delivery lead" a aquellos analistas que se encargan de utilizar el código desarrollado para crear una solución personalizada para cada usuario final (cada unidad de negocio o usuario que vaya a utilizar la herramienta) con las mínimas modificaciones de código posibles.

La personalización de la solución puede variar en dimensiones distintas. Un caso muy habitual son los tipos de datos a utilizar como regresores externos. Por ejemplo, en México es el único país en el que se utiliza el número de remesas recibidas cada mes como variable exógena. Por otro lado, dada la influencia del turismo en ciertas marcas, en Japón se hace uso de la evolución del turismo asiático y no asiático en el país a lo largo del año como regresor adicional.

En el caso del proyecto de los Trigger Models la personalización va incluso un paso más allá. Cada campaña de cada país tiene sus requisitos específicos: En los países de Europa Occiental las series temporales a modelar se deben dividir por región (Por ejemplo, en Bélgica los datos se modelan a nivel nacional y a nivel de la región de Bruselas, la región flamenca y la región valona), a nivel de marca y con un horizonte de predicción de corto plazo a 2 semanas y a largo plazo a 32 semanas. Sin embargo, en México y Brasil, las series temporales se dividen por producto, pero a nivel nacional, en este caso con un horizonte de predicción a corto plazo de 3 semanas, a medio plazo de 6 y a largo plazo de 32. Estas especificaciones implican que la fase de modelado sea muy distinta y requiera un enfoque particular en cada caso.

La función del "Delivery lead" que he asumido se ha basado en mantener un contacto constante con los usuarios finales y tomar en consideración todas estas particularidades para adaptar el código estándar de la herramienta a fin de que se proporcionaran los resultados de la forma deseada.

3.2.3 Developer e Investigador de nuevas propuestas

Aunque este rol está reservado para miembros con más experiencia profesional en el equipo, dadas las circunstancias en las que nos encontrábamos, he podido asumir el rol de desarrollador para el proyecto de los Trigger Models.

Entre las funciones principales de un desarrollador se encuentra la creación de nuevas funcionalidades en la herramienta o la mejora de las ya existentes. Dada mi falta de experiencia, tan sólo me he encargado de poner en producción pequeñas mejoras y funcionalidades a la herramienta actual.

Por ejemplo, el anterior código tan sólo permitía hacer predicciones a corto y largo plazo, por lo que, dada la creciente necesidad de un tercer escenario intermedio, modifiqué el código para poder ofrecer forecasts con estos tres horizontes temporales. Por otro lado, me he encargado de mejorar toda la parte de análisis previo y la búsqueda de series temporales que fueran influyentes en la predicción, específicamente incluyendo dentro de dicho proceso los tests de causalidad de Granger, el coeficiente de máxima información demorada y otros tratamientos a los datos que revelan más información para adaptar de forma más efectiva el proceso de ajuste del modelo.

Asimismo, en momentos en que la carga de trabajo de los “deliveries” previamente mencionados no era muy elevada, se me ha dado libertad de investigar y proponer nuevas mejoras a implementar en el momento exista mayor estabilidad de personal/cargas de trabajo en otros proyectos dentro del equipo. Esto me ha permitido empezar a dar pasos en dirección a asumir posiciones más enfocadas a desarrollo en un futuro próximo.

Dentro de todas las pruebas y propuestas que he prototipado destaco en primer lugar el desarrollo de una combinación de algoritmos que se encarga de agrupar las series temporales por su “similitud”: En concreto he hecho uso de algoritmos de clustering jerárquico utilizando como métrica de distancia un algoritmo de programación dinámica conocido como “Deformación dinámica del tiempo” (o “Dynamic time warping). Generalmente este algoritmo se usa en intérpretes de voz como el Google Assistant y parte de la premisa de que dos series temporales pueden tener un patrón muy similar, pero la duración (o puntos de observaciones) en los que ese ciclo se repite puede variar.

En otras palabras, pongamos que una persona dice la palabra “Hola” cinco veces, pero con distinta velocidad en cada intento. El patrón de sonido del primer “Hola” debería ser muy similar a cualquiera de los subsiguientes “Holas” pero con longitudes distintas. Sin entrar en detalles técnicos, la deformación dinámica del tiempo permite ignorar dichas diferencias en longitud y centrase en el patrón subyacente del sonido “Hola”. Esto permite que las distancias calculadas entre las ondas de voz de todas las grabaciones de la palabra “Hola” sean menores entre ellas y mayores entre grabaciones en las que se digan otras palabras.

Una vez calculada la distancia (o medida de similitud) entre las distintas series temporales, el algoritmo de clustering aglomerativo tan solo se encarga de crear de forma iterativa grupos de series en función de las distancias calculadas, empezando con un grupo de las dos series con mayor similitud entre sí, para finalizar con un único grupo donde se incluyan todas.

Tras crear los grupos, se observa su distribución en un gráfico conocido como dendrograma para determinar a partir de qué distancia los grupos creados tienen suficiente distancia (o “disimilitud”) entre sí. Finalmente, se crean distintas agrupaciones de series que serán tratadas en las siguientes fases de modelado en función de un punto de corte, es decir, se agrupan las series temporales en función de aquellas que se presentan una distancia entre ellas menor a un valor específico.

El objetivo final de este proyecto se basa en la idea de que series con patrones muy similares, independientemente de que el patrón dure más periodos o menos, presentan unas características también muy parecidas (estacionariedad o no estacionariedad, estacionalidad o no estacionalidad, tendencia…), por lo que todas ellas se beneficiarán aproximadamente de la misma forma con un enfoque de modelado concreto y estarán influenciados de forma similar por unos regresores exógenos concretos. Este hecho agiliza mucho el proceso de preparación de los datos y selección de los algoritmos a utilizar para la predicción de valores futuros. Por tanto, en un escenario donde las unidades de negocio locales están empezando a demandar predicciones de venta por producto (a razón de 30 productos distintos por marca), ser capaz de agilizar la modelación de los forecast implica poder atender de forma satisfactoria el incremento de la demanda sin aumentar el número de horas trabajadas.

Otro caso de uso de negocio en el que he realizado las primeras propuestas se basa en modelos "interpretativos". Específicamente, el equipo de Brasil estaba interesado en saber qué variables afectaban más a las predicciones de venta para saber que acciones deberían tomar para maximizar las ventas, es decir, querían que el equipo de Advanced Analytics desarrollara un estudio de causalidad.

Dado que hacer un análisis de causalidad es un ejercicio muy complejo, sobre todo si la variable que pretendes predecir está influenciada por muchas variables de confusión desconocidas, y tomar enfoques como el desarrollo de variables instrumentales implicaría tener datos muy concretos y un amplio conocimiento de las dinámicas del mercado brasileño (y un método basado en Efectos Fijos o Diferencias en diferencias fue descartado por el equipo), se le propuso al equipo de negocio local utilizar un enfoque en el que los modelos de predicción son capaces de "explicar" qué variables han sido más relevantes en la predicción y en qué medida afectan al valor predicho. Evidentemente, esto puede lograrse con un modelo lineal típico y el estudio de sus coeficientes. Sin embargo, a medida que aumentas la complejidad del modelo utilizado en busca de aumentar la precisión de las predicciones, dicha interpretabilidad se pierde completamente.

Tras mostrarse interesados también por este nuevo enfoque propuesto, durante algunas semanas me he encargado de desarrollar una pequeña herramienta donde se ejecuta una predicción del volumen de ventas a un día y, mediante una simulación que incluya varias semanas no observadas por el modelo, realiza varias predicciones para finalmente determinar las variables más importantes y su impacto en el valor predicho.

Para ello he hecho uso de dos enfoques distintos:

El primero se basa en los valores Shapley que encontramos en la teoría de juegos, los cuales básicamente representan el valor esperado (o la media) de la contribución marginal de "un jugador" (en este caso una de las variables que utiliza el modelo escogido para hacer la predicción).

La segunda propuesta se basa en un algoritmo llamado “RuleFit”, el cual se basa en generar reglas basadas en las variables que van siendo añadidas como regresores adicionales en un modelo basado en árboles con una regularización que se encarga de eliminar la multicolinealidad existente entre variables independientes. A grandes rasgos, si tratamos de predecir las ventas de un producto en base al número de farmacias abiertas y el diferencial de precios con respecto a un producto competidor, este algoritmo se encargaría de generar nuevas variables dummy o indicador en base a condiciones booleanas (Por ejemplo, el número de farmacias en esta ciudad es mayor a X y el diferencial de precio es menor a Y), para seguidamente introducirlas en un nuevo modelo de predicción y comprobar si mejora la precisión de ese modelo. Cuando el modelo finaliza, se obtiene un conjunto de reglas con un coeficiente de interpretación similar a los de una Regresión Lineal Múltiple y un soporte (número de veces que se cumple con la regla). En el caso que nos ocupa, yo tan sólo selecciono y presento como resultados las que tienen un soporte mayor a 0.5 (que dichas reglas ocurran al menos en un 50% de los casos).

Ambos enfoques están siendo actualmente introducidos en el desarrollo de un dashboard interactivo con el que el equipo de negocio podrá empezar a trabajar en el futuro.

3.2.4 Apagador de fuegos

Tal y como comentaba en subapartados anteriores, las circunstancias en mi equipo derivaron en la obligación de tomar las riendas del proyecto de forecasting. Por tanto, a lo largo de este año han ido surgiendo problemas con la herramienta de los que me he tenido que encargar de solucionar en la mayoría de los casos por mi cuenta.

Destacando un caso concreto, durante unos meses las conexiones con la API (un tipo de servicio que permite comunicarte con otros servidores y, en el caso que nos ocupa, extraer datos de interés) de Google Trends fallaron constantemente, impidiendo poder obtener datos semanales sobre las búsquedas de palabras clave que utilizamos en nuestro proyecto de predicción de ventas. Dado que no podíamos mantener estable la herramienta durante demasiado tiempo sin los datos de Google trends, desarrollé un pequeño bot conocido como “Arañas Web” o “Web Scrapper”, que se encargara de simular el comportamiento de un usuario que entra a la web de Google Trends y descarga los datos manualmente. Aunque tosca, la solución funcionó y hasta el momento se ha ido utilizando en momentos en que la API dejaba de funcionar otra vez.

En retrospectiva puedo afirmar que, aunque las soluciones no eran perfectas, y en muchos casos han sido bastante rudimentarias, con ellas he logrado sortear todos los problemas que han surgido y he conseguido mantener vivo un proyecto que bien podría haber muerto si no hubiera estado alguien pendiente de él.

3.3 Actividades adicionales realizadas en la empresa

Dentro del programa de jóvenes talentos con el que me incorporé, debíamos desarrollar otra serie de actividades dentro de la empresa más allá de las específicas de nuestro puesto en la compañía. A continuación, resumo dos de los principales proyectos en los que me vi involucrado.

3.3.1 Programa de inserción laboral para estudiantes paralímpicos

Aprovechando que Sanofi es patrocinador de los juegos olímpicos y paralímpicos de París, varios miembros del programa de jóvenes talentos formamos un grupo para desarrollar un proyecto que facilitara la inserción laboral de personas con diversidad funcional que, por su dedicación al deporte profesional, encuentran dificultades adicionales, más allá de las que pueden aparecer relativas a su situación, a la hora de insertarse en el mundo profesional una vez finaliza su etapa de deportista de élite.

Para ello, en primer lugar, tuvimos una entrevista con un medallista paralímpico de tenis de mesa, cuyo nombre no indicaré en deferencia a su privacidad, en la que aprendimos sobre los puntos donde más dificultades se puede encontrar un deportista con diversidad funcional dentro del mercado laboral, así como su experiencia en compaginando su vida deportiva, personal y profesional (fuera del deporte).

Seguidamente, nos encargamos de contactar con varias personas clave dentro de Sanofi para obtener apoyos y posibles ideas para desarrollar nuestro programa, así como otras asociaciones relevantes como la Fundación Universia, encargada de poner en contacto estudiantes con diversidad funcional y empresas, y el Comité Paralímpico Español, los cuales desde un principio fueron muy colaborativos y partidarios de llevar el proyecto adelante.

En paralelo con la celebración de dichas reuniones, se desarrolló el esquema del proyecto, el cual estaría basado en: La celebración de varios eventos patrocinados por el Comité Paralímpico en los que se recaudaría dinero para apoyar a estudiantes que estuvieran preparándose la clasificación a los juegos paralímpicos, mientras se ofrecía a los participantes poder experimentar el punto de vista de un deportista con diversidad funcional (con la celebración de partidos de baloncesto en silla de ruedas, carreras populares con los ojos vendados, etc.). Una vez finalizada la recaudación se daría paso a un programa de mentorías con trabajadores de Sanofi, y otro de soporte a la inserción laboral por parte de Fundación Universia para los deportistas paralímpicos, junto con un programa de apoyo psicológico para facilitar la transición de la vida de deportista de élite a una vida dentro del mundo laboral.

No obstante, a pesar del apoyo y colaboración del Comité Paralímpico Español y la Fundación Universia, así como el apoyo de muchas personas relevantes dentro de Sanofi, el proyecto se mantiene por ahora congelado.

3.3.2 Sesiones sobre Ética dentro de la IA y la Ciencia de datos

Al margen del programa de jóvenes talentos, uno de mis compañeros me invitó a formar parte de un grupo encargado de poner de manifiesto y educar sobre la importancia de la ética dentro de los nuevos avances en Inteligencia Artificial y Ciencia de Datos.

Hasta el momento, nos hemos encargado de preparar y presentar charlas dentro de la empresa sobre los principales aspectos que unen la ética y la ciencia de datos. Por destacar un ejemplo, una de las últimas charlas consistía en saber identificar sesgos y los mecanismos que tenemos para paliarlos en nuestras predicciones o inferencias.

En concreto, en esta charla se destacaron varias fuentes de posibles sesgos, como la paradoja de Berkson, la cual se basa en un fenómeno estadístico en el cual dos variables que, estando en cualquier otra circunstancia correlacionadas de una forma determinada, parecen estar inversamente relacionadas cuando se analizan de forma conjunta debido a un proceso de selección que favorece que ciertos patrones estén sobrerrepresentados en relación con otros más comunes en el conjunto de la "población".

Por otro lado, se hizo hincapié en la importancia de no eliminar variables que capturen sesgos producidos en el "mundo real". Por ejemplo, si existía un comportamiento misógino en las admisiones a una universidad, la variable sexo no debería ser eliminada de un modelo de predicción de probabilidades de acceso a una carrera universitaria, ya que otras variables serán capaces de construir un "proxy" que capture la influencia de esta variable, sino que se deberían introducir otros elementos en el análisis que controlen dicho sesgo.

Por último, se destacó la necesidad de tener datos balanceados si el objetivo del análisis es sensible (como es el caso de los ensayos clínicos) y posibles mecanismos para paliar el desbalanceo como las técnicas SMOTE o la modificación de pesos en las funciones de coste.

En conclusión, el objetivo de estas charlas es concienciar de la creciente relevancia de las prácticas éticas dentro del mundo de la inteligencia artificial, así como la necesidad de empezar a implementar protocolos que aborden está materia antes de que la ley empiece a regularla, pues ya encontramos casos en los que el trato no discriminatorio por parte de algoritmos de toma de decisiones empieza a tomarse en consideración en el conjunto de leyes españolas tras las recomendaciones presentadas por la UE. Ejemplo de ello es la implementación del artículo 23 de la Ley 15/2022, el cual indica que: "Las administraciones públicas y las empresas promoverán el uso de una Inteligencia Artificial ética, confiable y respetuosa con los derechos fundamentales, siguiendo especialmente las recomendaciones de la Unión Europea en este sentido" (BOE, 2022).

3.4 Valoración personal de mis prácticas y aportaciones del máster

Para finalizar el apartado relacionado con mi año de prácticas profesionales, presento una valoración personal de mi experiencia, indicando algunas de las vivencias y aprendizajes del máster que me han aportado valor en mi desempeño profesional.

A nivel general, la valoración personal de mi primer año dentro del mundo de la ciencia de datos es más que positiva. Si bien es cierto que mi objetivo tras entrar en el máster era tratar de encontrar un hueco en el mundo académico en el largo plazo (y sigo sin descartar esta opción para un futuro), en esta nueva profesión he encontrado un híbrido que me ha permitido hasta cierto punto encontrarme en un mundo más relacionado con la investigación, pues he tenido margen para leer artículos académicos e implementar nuevos métodos y herramientas recién salidos de la universidad y centros de investigación, a la vez que ganaba experiencia en la empresa privada, por lo que me considero muy afortunado de haber logrado encontrar una ocupación que me permite satisfacer mi inquietud investigadora a la vez que me asegura estabilidad económica.

A pesar de entrar como becario, he asumido mucha responsabilidad desde el principio dentro del equipo y esto me ha ayudado a aprender mucho más rápido, tanto a nivel técnico como a nivel de gestión de stakeholders y manejo del tiempo. Ha sido un año de mucha carga de trabajo y esfuerzo, pero también de gran avance personal y profesional. En conclusión, este año me ha proporcionado unas experiencias que han tenido un gran impacto en mi vida personal y profesional, lo cual debo agradecer en especial a mi jefe, mi tutora de empresa, así como todos los compañeros que me han acompañado en este viaje.

Aun así, en esta valoración personal me gustaría añadir una breve crítica, destacando un hecho que he ido corroborando con el tiempo. A pesar de que es un campo profesional que ha explotado en popularidad los últimos años, mi percepción tanto dentro de Sanofi como en otras empresas de tamaño similar (trasladada por compañeros de profesión) es que estamos experimentando un distanciamiento entre la ciencia de datos y el objetivo que en realidad esta debería perseguir en el mundo privado: solucionar problemas de negocio y agilizar la toma de decisiones del resto de departamentos. Esta situación, a mi parecer, parte de dos factores principales: el abandono de la lógica empresarial en pos de paliar la típica dolencia del siglo XXI, el FOMO o “Fear of Missing Out”, y el relativo desapego de la universidad con el mundo empresarial. Es más, considero que el segundo factor es catalizador del primero.

Con la aparición y desarrollo espectacular de la Inteligencia Artificial Generativa, todas las unidades de negocio desesperan por introducir dentro de sus procesos proyectos que incorporen esta nueva herramienta a fin de mejorar su eficiencia. No obstante, ¿qué problema se trata de solucionar?, ¿qué proceso tratar de automatizar o agilizar? Probablemente ninguno, pues el único objetivo que se persigue con su implementación, en la mayoría de los casos, será la de alzarse como compañía que "incorpora la IA en sus actividades y procesos", derivando en millones de euros y horas invertidas en proyectos que terminarán en el ostracismo, pues desde el principio no estaban enfocados en solucionar un problema empresarial real.

Pero ¿quién es el culpable de esta situación, el directivo que solicita o el científico de datos que desarrolla y proporciona soluciones? A nivel superficial podríamos decir que el cliente es el culpable de que el proyecto que ha demandado no se utilice, pues al fin y al cabo ha sido él quién lo ha demandado.

Sin embargo, al igual que un médico no receta morfina por el mero hecho de que un paciente la solicite, ¿no debería un científico de datos redireccionar las demandas de un cliente hacia una solución que realmente sea útil para negocio? Aquí es donde realmente está el problema, ante un cliente obnubilado por su miedo de no estar tan actualizado como sus competidores, encontramos un científico de datos que no es capaz de aportar la lógica de negocio que realmente necesita un proyecto de analítica avanzada.

Este hecho, en parte, tiene su origen en la educación del grueso de científicos de datos. Provenientes de carreras con un fuerte componente técnico o científico, el enfoque de su educación se centra en su totalidad en la teoría para, una vez salidos de la universidad, toparse con un mundo donde gran parte de sus conocimientos no se aplican ni se aplicarán de forma directa en su vida profesional. No niego la importancia de ninguna de las materias que se imparten en la universidad, pero ¿no sería mucho más útil que estas se impartieran enfocándose en la parte práctica más que en la pura teoría? Y no me refiero a tener sesiones prácticas donde los alumnos se sienten a calcular los coeficientes de una regresión, se pongan a programar cualquier ejercicio que se les indique o redacten algún tipo de informe para un caso práctico genérico. Me refiero a lecciones realmente prácticas donde el alumno se enfrente a problemas, se encargue de hacerse las preguntas adecuadas y proponga una solución (sea mejor o peor, pero propia), todo ello sin un guion estipulado a seguir.

No obstante, esta situación en buena parte beneficia a personas con perfiles como el mío y, precisamente, considero que el Máster en Internacionalización Económica, incluso aunque las materias no acaben de casar con los conocimientos de un científico de datos, ha reforzado mi capacidad de entender las necesidades de los distintos departamentos de negocio, de poder interactuar con ellos de forma más fluida y encontrar un lenguaje común al que poder traducir conceptos en ambas direcciones (de un departamento técnico, como en el que trabajo, a uno de negocio, como el de marketing, y viceversa).

Por ejemplo, la asignatura de Marketing Internacional ha reforzado los conocimientos para poder hablar de forma mucho más fluida con los departamentos que consumían ambos de los proyectos en los que estaba involucrado, pues A&P está estrechamente relacionado con el Marketing Mix y los Trigger Models muy relacionados con el marketing digital. Los conocimientos que la asignatura de logística me ha aportado me van a permitir entender de forma mucho más sencilla a qué problemas se enfrentan los equipos de Supply Chain y Manufacturing de cara a los nuevos proyectos que se nos presentan en los próximos meses de la mano de estos departamentos.

Las asignaturas de dirección y estrategia de empresas me han permitido, una vez dentro de la empresa, entender las razones detrás de ciertos comportamientos y dinámicas dentro de una empresa de estructura de corte matricial y con una cultura que se asemeja a una multilocal en transición a una empresa global. Asimismo, estas asignaturas me han permitido entender desde otro punto de vista las implicaciones que tendrán ciertas decisiones estratégicas que la compañía ha tomado para los próximos años, así como el impacto que tendrán en mi departamento.

Trabajar en un equipo global me ha permitido colaborar con personas de todas las partes del mundo, por lo que he podido experimentar de primera mano los conceptos aprendidos en la asignatura de Management Intercultural, facilitándome en muchas ocasiones la compresión de ciertos comportamientos por parte de mis compañeros y permitiéndome ser capaz de manejar las situaciones de forma mucho más sencilla.

Finalmente, las asignaturas de Integración Económica, Coyuntura Económica, Finanzas Internacionales y Comercio Internacional, así como el enfoque que he tomado con mi trabajo de final de máster, han afianzado mi interés por dedicar mi actividad profesional al análisis, cuantificación y estudio empírico de variables dentro del campo de la economía, pues son las asignaturas que más he disfrutado estudiando y las que me ofrecen una base de refuerzo y añaden un factor diferencial en mi perfil profesional para poder en un futuro encontrar un puesto de trabajo más ligado al análisis económico que a las actividades empresariales que analizo actualmente, ya sea en el sector público, en una Universidad o en una empresa privada.

4. Referencias para la sección de actividad profesional

Atance, C. (2021), "Opella: spin off para Sanofi, OTC". PharmaBiz.NET. Disponible a través del link: https://www.pharmabiz.net/opella-spin-off-para-sanofi-otc/

Blake, H. (2013), "A history of Sanofi", Pharmaphorum. Disponible a través del link: https://pharmaphorum.com/articles/a_history_of-_sanofi

Sanofi (2023). "Our Strategy". Disponible a través del link: https://www.sanofi.com/en/our-company/our-strategy

Sanofi (2023). "Áreas de actuación". Disponible a través del link: https://www.sanofi.es/es/sobre-nosotros/sanofi-espana/dosier/dosier-areas

BOE (2022), "Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación", BOE-A-2022-11589. Disponible a través del link: https://www.boe.es/eli/es/l/2022/07/12/15/con

III. ARTÍCULO ACADÉMICO

A continuación, presento los contenidos completos de mi artículo académico (todavía Working Paper sujeto a mejoras).

Este paper es el resultado de horas de estudio y trabajo bajo la supervisión de mi tutor, el profesor Francisco Requena.

Aunque el documento original está escrito en un archivo Latex aparte de este, he ajustado el formato para que se adapte al documento general de forma homogénea y se ciña las directrices a seguir para el TFM.

Asimismo, a partir de este punto, el idioma presentado es el inglés.

Espero que lo disfruten tanto como yo (la mayor parte del tiempo) escribiéndolo.

EU trade and economic integration: Revisiting border effects inside the European Union

Guillem Rochina Aguas†

† *Institute of International Economics, University of Valencia*

Abstract

This study delves into the intricate dynamics of trade barriers and their implications within the context of European Union's (EU) economic integration. Employing the structural gravity equation framework, we examine the multifaceted factors that influence border effects in international trade. Our analysis encompasses various levels of granularity, including country groups based on year of EU accesion, countries, and industries, over distinct time periods. By disentangling the intricate interplay between exogenous and endogenous forces, we shed light on the magnitude of border effects and their evolution within the EU.

Our findings reveal a decreasing and heterogeneous trend in border effects at the European Union industry level, driven by several exogenous and endogeous factors such as industrial concentration or elasticity of substitution, but above all, we find that technical harmonization efforts have contributed to reducing border barriers when accounting for trade within the Single Market.

While our findings provide valuable insights into the complexities of trade barriers, there remain unanswered questions that warrant further exploration, including the role of consumer preferences, intermediate products or cultural factors.

Keywords: *Border effect, European Economic Integration, Structural Gravity Equation, Technical Harmonization*

1. Introduction

The quest for economic integration has been a cornerstone of the European Union's (EU) aspirations, forging a path towards closer collaboration and increased prosperity among member states. Since the establishment of the Treaty of Rome in 1957 and the enactment of the Single European Act in 1987, the EU has actively worked to reduce various barriers among its member states. These efforts have consistently focused on the central goal of the EU's endeavors: promoting the unhindered movement of goods, services, capital, and people.

A significant milestone in the EU integration process was achieved with the signing of the Maastricht Treaty in 1993. This treaty laid the groundwork for the completion of the Single Market and, not only advanced economic integration, but also fostered shared political values among member states. Subsequent developments further propelled such integration: The creation of the European Economic Area in 1994, together with Schengen Agreement's on the abolition of border controls in 1995, and the establishment of the Monetary Union in 1999 accelerated the deepening integration beyond the economic dimension.

Although most of those agreements where of different nature, among the diverse ream of pursued objectives, the study of one of them has been of particular interest in the academy: The reduction of economic and political borders among its members, and its impact on welfare.

As a matter of fact, national borders have long been recognized as one of the major nemeses of international trade, hindering the flow of goods and services across different geographical regions. The detrimental impact of these borders on trade volume is widely acknowledged and captured by the concept of "border effect". Coined by McCallum (1995), this effect represents the disparity between domestic and international trade volumes caused by a wide variety of geographic, political and cultural borders, as well as interactions between countries or regions such as FTAs. In brief, when we speak about "border effects" we are responding to the question "To what extent does a country trade with itself more than with the rest of its partners?".

While it is intuitive to understand that national borders hinder international trade flows, the intriguing aspect of this empirical phenomenon lies in the fact that the economic theory has yet to fully comprehend the underlying reasons for its substantial magnitude. Since its discovery, a line of investigation emerged to understand the factors that produce this well-known effect in international trade. Extensive literature has identified numerous factors that contribute to a decline in trade volume once it crosses national borders, including geographical distance, tariffs, varying regulatory frameworks, and even cultural differences such as language, all of which contribute to an increment of transportation and communication costs which, consequently, may be deter firms from engaging in exporting or importing activities, eventually resulting in a decline in international trade flows.

Nevertheless, even if those factors explained a portion of the border effect, no study has still minimized the unintelligible part of its magnitude to a nil value, which means that there could be several sources contributing to its magnitude yet to be discovered.

In view of the above, the European Single Market, comprising 27 countries (i.e. 27 borders), over 500 million citizens and accounting for a significant portion of global GDP and international trade volume, provides an intriguing subject for studying the existence of border effects. Analyzing the EU's integration project allows us to explore the expected reduction of the border effect among member states at different stages of time, as well as the possible reasons behind such reduction.

Thus, in an effort to shed some light in the European Union integration process, this paper has four main objectives:

First, we estimate the border effects at the European Union level, providing an initial overview of the overall border effect for the EU27, in comparison to non-EU countries. To offer a more comprehensive understanding, we further dissect the border effect into three distinct EU blocks, depending on their year of accession. We examine the evolution of this effect from 1999 to 2014, with 2006 as the midpoint, for the EU15 countries, the Central and Eastern European countries (CEECs) that joined in the 2004 enlargement, and the late joiners of 2007, namely Bulgaria and Romania. Due to the lack of available intra-national data, such comparison had not been possible until the recent years. Although previous studies have already analyzed this intra-national vs international trade at country level, we provide an estimation at industry level using a new and comprehensive international trade database. To our knowledge, no previous research has been able to present those results at this level of granularity and taking time into consideration.

Second, we further investigate the changes in the border effect size at sector level. This analysis includes a comparison between the border effects of European and non-European industries, as well as a comparison of EU industries' border effects in 2014, reflecting in this case the differences between border coefficients when we take into consideration whether trade partners belong to the European Union or not. This approach is taken under the hypothesis that, beyond bilateral idiosyncrasies, the estimated border effect is heavily dependent on the partners a country trades with, so does the estimated results of a trade policy developed under the umbrella of a bigger agreement such as the Single Market is. Magnified once we reach the industry level, we believe that accounting for the partners involved allows us to obtain a more reliable quantification of the results coming from a policy decision.

Third, we aim to assess the potential influence of exogenous and endogenous factors on the border effect of each industry. Based on the approaches of Chen (2004) and Evans (2003), we reutilize the Rauch's classification of industries to determine the significance of either the search costs or the elasticity of substitution as a deterrent or catalyst of cross-border trade within the European Union. Moreover, we include the Ellison-Glaeser index in our estimations to assess the importance of economies of agglomeration in the determination of the border effects' magnitude per industry.

As our fourth goal, we conclude by studying the impact of the harmonization process on trade flows within the EU, specifically the reduction of the border effect in industries affected by technical harmonization efforts. To add an additional perspective to the harmonization question, we also include a new variable that indicates the regulatory intensity of each industry so as to determine whether those industries where regularization has been more severe have experienced a greater reduction of border effects or not.

Our empirical results show a general decreasing and heterogeneous trend of border effects at all levels of granularity.

First, in line with the conclusions of Spornberger (2021), we find diminishing barriers to international trade both inside and outside the Single Market. Namely, EU membership has a decreasing relevance in such reduction for the "Old Europe" while an increasing or still highly relevance for subsequent joiners. This results lead us to hypothesize that the integration process exhibits a shape similar to a logistic or logarithmic curve, with a threshold that European countries seem to have a hard time crossing.

Second, the fact that EU membership is still relevant for all groups in the cross-section corresponding to our last available year estimations confirms our hypothesis that the concept of border effect is heavily dependent on the partners considered. The very same pattern, even if far more heterogeneous, is found at the industry level's estimations.

Third, our results show that economies of agglomeration and the elasticity of substitution play a key role in explaining border effects inside the European Union. Specifically, our results show that industries that are less tied to a particular location and industries that produce homogeneous products exhibit bigger border effects.

Fourth, our estimations conclude that the technical harmonization efforts exerted by the European Commission have in fact had a critical impact on the further reduction of border effects across the industries involved successfully in the process.

The rest of the paper proceeds as follows: Section 2 includes a brief summary of the literature review regarding the structural gravity equation and the border effect estimation. Focusing mainly the attention on the EU case, Section 3 provides an explanation of the theoretical framework and the econometric specification we use in the present paper. Section 4 describes the data sources and how we have proceeded with data preparation. Section 5 presents the results of the national pooled estimations and its evolution over time. In Section 6 we delve into the details of the evolution of the border effect at industry level. Section 7 provides an explanation of the possible sources of the current size of the border effect in the EU industries and provides an answer to the impact of the harmonization process on this effect. Finally, Section 8 presents the conclusions of the investigation and Section 9 includes some areas where we can improve or redefine our research.

2. Literature Review

As presented in the introduction, the investigation of the border effect originated from the seminal paper by McCallum (1995). In this study, analyzing state-level trade between Canada and the USA in 1988, McCallum found that even after accounting for various trade cost factors, the resulting border effect revealed that Canadian provinces' trade was approximately 22 times larger between them compared to their trade volume with any state of the neighboring country. Cataloged as one of the major puzzles in international economics by Obsteld and Rogoff (2000), the complexity of this effect sparked a now extensive line of research.

Building on McCallum's work, subsequent influential papers by Engel and Rogers (1996, 2000, 2001) highlighted that significant price discrepancies persisted across the US-Canadian border, concluding that price discrimination, in combination of transportation costs, are two of the main causes of border effects. One important conjecture presented in this paper, indicates that price discrimination behavior involved in cross-border price movements goes beyond the standard conception of it, proposing that if sticky nominal prices account for the importance of the border effect, this could mean that differences between markets influence prices due to the fact that price setters tend to take into account prices of nearby competitors. This implies that accounting for the "isolation" of the market for the product of each price setter could influence the estimations of border effects.

Other studies, such as Helliwell and Verdier (2001), further documented a substantial decline in trade flows between Canadian provinces and US states compared to trade flows within states and provinces, even after controlling for a more robust distance metric based on similarity of economic structure between regions, highlighting that the impact of borders varied depending on the type of goods being traded and the economic characteristics of the regions involved. However, Wolf

(2000) does not only encounter the same result, but he also reports that trade within individual US states is significantly larger than trade between US states, implying that not only national but also regional borders are relevant trade deterrents. Correspondingly, although not including formal international trade barriers associated with the single market, Nitsch (2000) concluded that intra-national trade within the average European Union member was accounted to be 10 times larger than their trade with other fellow member country, although its border effect was found to be declining overtime, from 15 in the 80s until 10 in the 90s.

Nevertheless, not long after, a highly influential paper from Anderson and Van Wincoop (2003) demonstrated that a significant reduction in McCallum's estimated border effect, from 22 to 5, could be achieved when controlling for differences in price levels. Their proposal, further summarized in the following section, was taken and simplified in Feenstra (2003), which has since then become the standard approach in the contemporary state-of-the-art estimations of the gravity equation.

Such reduction, even being quite dramatic, still left a significantly high border effect whose causes where yet to be fully understood. Thus, since then, numerous authors have endeavored to shrink further McCallum's estimations and unravel the complexities surrounding the magnitude of the border effect while putting the European Union economic integration in the spotlight, parting from the aforementioned contribution of Nitsch (2000).

Chen (2004) adopted an industry-level perspective to investigate this phenomenon and proposed a two-group classification of factors that may elucidate the impact of national borders on trade, differentiating between exogenous and endogenous factors. From the exogenous perspective, the study highlighted the pivotal role of tariffs, non-tariff barriers, and disparities in product information in explaining the size of the border effect. Additionally, it underscored the significant influence of endogenous factors, such as the degree of substitutability between local and foreign products (referred to as "home bias") and the optimal location decisions of producers (referred to as "economies of agglomeration"). However, due to lack of data that fulfilled her requirements, the sample of study only included seven EU countries in the year 1996 and thus not accounted for trade outside the European Union. Similarly, Evans (2000) found that the border effect tends to be lower for products with lower elasticity of substitution, emphasizing the importance of substitutability between domestic and foreign products as a key driver behind the significance of borders.

In an alternative line of investigation, several authors have pursued a different approach for analyzing border effects, focusing on the examination of their geographical granularity. For instance, Gil-Pareja et al. (2004) found that, even after controlling for market size and distance, Spanish regions traded approximately 21 times more with other regions within Spain than with other OECD members, not being this effect uniform across regions. Expanding on these findings, Requena and Llano (2010) conducted a similar analysis of intra-regional, inter-regional, and international trade within Spain, introducing an additional industry layer. Their results demonstrated that Spanish regions traded approximately 30 times more with themselves than with other regions in Spain, and the volume of intra-national Spanish trade was around 10 times greater than international trade, even after accounting for size, distance, adjacency, and industry-specific characteristics. Moreover, they emphasized that product differentiation played an even more crucial role in explaining inter-industry differences in external border effects compared to internal ones.

More recent studies have delved into the analysis of border effects at different geographical levels. Cheptea (2013), establishing that the study of border effects (considered in this paper as home bias) is exactly the same as the analysis of European integration, incorporates average weighted measures of intra-national distances and trade data between the 27 EU members. Among its conclusions, the main findings reveal that economic integration among CEEC countries and the "Old Europe" has increased in the form of a reduction of border effects, meaning that the reduction of the magnitude of those effects has been particularly significant for countries that joined the Single Market, although there is still a huge room of improvement in terms of further EU integration.

Taking a country-specific approach, Gallego and Llano (2015) estimate both external and internal border effects in Spain using a novel inter-regional trade dataset including eight of its main European partners. Their findings suggest that considering region-to-region international flows leads to smaller estimations of border effects compared to previous studies that did not account for these flows. These conclusions are of particular interest, as they revive and reinforce one of the ideas present in Wolf (2000), that Hillberry and Hummels (2008) further developed using ZIP-code-level domestic US trade flows, that is to say, border effects do not only exist in international trade, but they also are present within a country (i.e. domestic border effects), as trade seems to be still heavily concentrated at local level. What is more, not only domestic borders at regional granularity impose a notable barrier to trade among regions even if they belong to the same country, but national border effects have been reported both in Hillberry and Hummels (2008) and Llano, C., Minondo, A., and Requena, F. (2011). to be overestimated due to aggregation bias. That is, the more broader the geographic dimensions we are grouping the data to estimate border effects, the more we are ignoring the localization pattern existing in intra-national trade flows (intra-national trade tend to concentrate in small geographic areas) and thus, the more we are overestimating the real national border effect, as we are wrongly attributing the role of local trade concentration to national borders. Such conclusions are also found in Coughlin and Novy (2021), in which their results show that domestic border effects are in fact bigger than international border effects, that is, the first kilometers associated with leaving the local province/regions are way harder to overcome than crossing the national border once the region of origin is left behind. Such conclusions, lead us to consider the relevance of accounting for geographical concentration to offer proper estimations of national border effects.

Pursuing different answers, Spornberger (2021) provides a comprehensive analysis of EU effects proposing a new methodology that allows to capture valuable information absorbed by fixed effects through a two-step estimation, while preserving the structural foundation of the equation, which allows to interpret the results in a causal way. Beyond this new econometric approach, the paper concludes that deep trade integration among EU-15 has already been realized and has not changed since 1995, while CEEC countries have experienced significant integration effects by joining the Single Market, concluding that the EU has not reached its full integration potential yet. Moreover, another key conclusion highlighted in this paper underlines that the velocity at which a country reduces its border effect, in this case reflecting the speed at which an EU member integrates inside the European Union, is heavily dependent on its initial state. That is to say, the reason that Poland shows stronger economic integration efforts than the ones showed by Germany, or the Netherlands arises mainly in the fact that they come from an scenario where Poland was far more isolated in the international trade system than Germany and the Netherlands.

Finally, and in line with the previous conclusions, Santamaría, Ventura and Yesilbayraktar (2021) go even further, proposing a completely new framework based on causal inference as an identification strategy to estimate country border effects on trade flows. Their results show that European Union is still far from being a Single Market, as the border coefficients of its members remain still high nowadays, concluding that some of the reasons behind this situation are the development of national bias in preferences and a national cost advantage that penalize international trade in benefit for intra-national trade.

3. Research design

In the present section we start providing a brief explanation of the structural gravity model and how its configuration has evolved since its conception back before the 2000s, focusing on the two elements that give its structural form and the new structural element introduced in the new wave of gravity models. On the other hand, a general explanation of the standard econometric approach used along the paper is presented in the second subsection.

3.1. Theoretical framework: The structural gravity model

The empirical analysis of the present paper is based on the structural gravity model of trade. In a nutshell, the gravity equation in international economics is an adaptation of Newton's Law of Universal Gravitation which, in its naïve form, predicts that the bilateral trade flows between two countries are directly proportional to the product of their respective economic sizes (i.e. the bigger their GDPs, the bigger the trade volume) and inversely proportional to the trade frictions (or the "distance") between them (i.e. the longer the "distance", the smaller the trade volume).

Specifically, we will employ the structural form of the gravity equation, popularized by Anderson and Van Wincoop (2003), which takes the following form:

$$X_{ij,k} = \frac{Y_{i,k} E_{j,k}}{Y} \left(\frac{t_{ij,k}}{\Pi_{i,k} P_{j,k}} \right)^{1-\sigma} \tag{1-1}$$

Where i and j refers to the exporting and importing country, respectively, and k the industry. $X_{ij,k}$ is the bilateral trade volume for industry k, $Y_{i,k}$ and $E_{j,k}$ denote the production and expenditure of origin and destination countries on industry k (normalized by world's GDP, Y) and $\sigma > 1$ is the elasticity of substitution among goods from different countries. On the other hand, $t_{ij,k}$ refers to trade costs, and $\Pi_{i,k}$ and $P_{j,k}$ are the structural terms of the gravity system, coined by Anderson and Van Wincoop (2003) as **Inward and Outward Multilateral Resistance terms**.

In essence, the structural gravity equation can be understood as the composition of two terms:

A size term, $\frac{Y_{i,k} E_{j,k}}{Y}$, which can be interpreted as the hypothetical level of "frictionless" ($t_{ij,k} = 1$) trade between two partners. All in all, this term transmits the idea that large producers will always export more to all destinations and rich economies will import more from all sources.

To avoid this naïve interpretation, the second term, $\left(\frac{t_{ij,k}}{\Pi_{i,k} P_{j,k}} \right)^{1-\sigma}$, captures all the effects of trade costs that may cause realized trade to deviate from the frictionless trade. Being $t_{ij,k}$ tipically approximated in the literature by various geographic and policy variables, the key components of this term are the ones that make the model completely structural.

Namely, the outward multilateral resistance (OMR), Π_i, and the inward multilateral resistance (IMR) P_j, fully itemized in equation 1-2 and 1-3, are two interrelated terms which point out the fact that any variation in bilateral trade costs in any dyadic relationship not only will affect the two countries involved in such relationship (Partial Equilibrium effects) but it will also affect all other countries in the world, with possible feedback effects on the two originator countries (General Equilibrium effects).

$$\Pi_{i,k}^{1-\sigma} = \sum_j \left(\frac{t_{ij,k}}{P_{j,k}} \right)^{1-\sigma} \frac{E_j}{Y} \tag{1-2}$$

$$P_{j,k}^{1-\sigma} = \sum_i \left(\frac{t_{ij,k}}{\Pi_{i,k}} \right)^{1-\sigma} \frac{Y_i}{Y} \tag{1-3}$$

As it can be interpreted from equation 1-1, a reduction in trade costs between two trading partners would cause the increase in the bilateral trade flow between them but, at the same time, it would decrease its multilateral resistance terms (equation 1-2 and 1-3), leading to a reduction of imports and exports from any other source, becoming in the final analysis more isolated from any other partner. Thus, a reduction of trade costs between two countries (e.g. due to the signature of an FTA) will not only cause an increase in both imports and exports from one partner to another due to a reduction of trade costs, but it will be also translated, through the MR terms, into additional indirect effects at country level, which will take into account changes in relative prices, incomes and expenditures induced by trade costs changes, affecting all other participants in the international trade system.

Furthermore, these terms do not only provide us a mechanism to capture the indirect effects of a FTA (or any other agreement) on the rest of the world, but also act as a "normalization factor" that allows us to estimate the "real" effect of a reduction in bilateral costs and not an over or underestimation of them. In plain words, if a country that, due to its characteristics, is already highly isolated from the rest of the world (i.e. it has high MR terms) reaches an agreement that reduces bilateral trade costs with another partner that is way more integrated, by logical reasoning, would this translate in a overwhelming influx of trade on goods and services or, conversely, it would be translated into a timid increment of trade flows between those two countries? If we take the naïve form we would probably face the first scenario, where trade creation and diversion are inflated in our estimations. However, when introducing the MR terms such bias would we eliminated, resulting in the second scenario, where trade creation and diversion remains as low as it should be.

In fact, this example reflects the essence of what we should interpret by multilateral resistance terms, being the "easiness" of a country either to export (Outward MR) or import (Inward MR).

That is to say, if a country is already isolated in the international trade system, a simple reduction of bilateral trade cost, even if it increases the bilateral flow of goods with the liberalizing partner, it will not do much to alleviate its situation.

In conclusion, the two features provided by the MR terms allow us to make robust estimates in a scenario that takes into account all the consequences of trade policy decisions and the circumstances in which a country finds itself with respect of the rest of the world.

Nevertheless, there is still one more structural term left to include, as the equation explained until this moment was not design to account for one of the key elements in our analysis, the inclusion of intra-national trade in the estimations (i.e. $X_{i,j}$ where $i = j$). Once we include trade happening inside the borders of a country, a new structural term endogenously arises in the equation, which captures the extent to which a country trades with itself more that it does with any other partner in the international market. The inclusion of this term allows us to understand and quantify the impact of geographical and political borders on international trade flows between countries through the estimation of the so-called "border effects", just by adding one more parameter in our estimations, as it is presented in the following section.

3.2. Econometric specification

From the structural gravity equation presented in the previous section, an econometric specification can be derived for estimation purposes. Following Santos Silva and Tenreyro (2006) proposal, the basic form of our gravity equation will be the following:

$$X_{ij,k} = e^{\left(\pi_{i,k}+\chi_{j,k}+\phi D\right)} \times \epsilon_{ij,k} \tag{1-4}$$

for $i, j = 1, \ldots, N$ countries and $k = 1, \ldots, N$ industries. Where $X_{ij,k}$ is the bilateral trade flows for a particular industry, $\pi_{i,k}$ and $\chi_{j,k}$ denote exporter-industry and importer-industry fixed effects which, according to the Feenstra's (2003) proposal, capture the inward and outward multilateral resistance terms, and $\epsilon_{ij,k}$ refers to the multiplicative error term. Finally, focusing on the most relevant component of our estimations, the D term should be conceived as a vector of both time varying and time invariant trade costs, taking the following form:

$$\phi D = \phi_1 ln(distance_{ij}) + \phi_2 adjacency_{ij} + \phi_3 language_{ij} + \phi_4 fta_{ij} + \phi_5 border_{ij} \tag{1-5}$$

Where $ln(distance_{ij})$ denotes the geographical distance in logarithmic scale between importer and exporter (or viceversa), $adjacency_{ij}$ and $language_{ij}$ indicates whether importer and exporter share a physical border or the same language, respectively, and fta_{ij} is a dynamic indicator variable that specify if there is an agreed free trade agreement between the exporter and the importer in the year considered. This last variable has been set to 0 when both importer and exporter correspond to EU members in the respective year corresponding to the observation, as we will take into consideration the differences between trading inside and outside the Single Market.

Finally, as we mentioned in the previous subsection, given the fact that we are comparing the values of intra-versus international trade, $X_{ij,t}$ includes not only international but also domestic trade ($X_{ii,t}$). Thus, the key coefficient to estimate in this case is ϕ_5, which absorbs the influence of $border$, a variable that indicates when trade occurs inside the borders of a country ($border_{ij} = 0 \quad \forall\ i = j$, where i denotes country of origin, and j country of destiny) and, therefore, indicates the size of the border effect. However, one deeper appreciation must be given to the border vector, as its conception varies along our research:

Even though we keep our observations at industry-level granularity, in our initial regressions, we attempt to estimate a pooled border effect, considering various EU groups and country levels. Therefore, when we refer to the border effect of the EU, denoted as $border_{EU}$, we are actually discussing the average border effect from EU member countries. This explanation also holds true for any other group of countries analyzed in our research. However, when we conduct regressions on a country-specific basis, the estimated border effects are uniquely associated with each corresponding country.

As we dig deeper in our analysis a new dimension arises. Once we have reached the industry level, our border variables need to be estimated as $border_{ij,k}$, where k correspond to the industry we are considering in our estimations. Thus, from this point we are no longer estimating border effects per EU member or group, but we are estimating the average border effect of each industry inside the European Union.

Bearing in mind that we are playing with this different dimensions in our estimations, which could lead to confusion to the reader, a more detailed explanation of the particularities of each specification of the border variables, and the associated interaction terms, we estimate is provided in the following parts of this article, since we also include different interactions along the following chapters.

Returning to the technical characteristics of our economic specification, the multiplicative form of the present model allows us to make use of the Poisson Pseudo-Maximum Likelihood (PPML) estimator, which helps us overcome the heteroskedasticity problem that plagues trade data (Santos Silva and Tenreyro, 2006). Moreover, thanks to the PPML approach we are also able to take advantage of the information that could be contained in zero trade flows, as we are not in the need of log-linearizing the target variable. These two features are particularly interesting in our research case at industry level, as the more granular the data is, the higher the risk of finding heteroskedasticity and zero values in the observation set.[1]

Finally, the inclusion of $\pi_{i,k}$ and $\chi_{j,k}$ into our estimations, which denote exporter-industry and importer-industry fixed effects, enables us to control for unobservable multilateral resistances or any other characteristic that may differ between the industries of each exporter or importer, respectively, such as the differences in prices for each product or any other idiosyncrasy we could not otherwise take into account. On the other hand, even though they should be included in any gravity estimation so as to account for any unobserved characteristic between country pairs, we will not include pair fixed effects, as we are estimating border effect dummies using cross-sectional

[1] Moreover, thanks to the additive property of this estimator we ensure that the estimated fixed effects are equivalent to their corresponding structural terms (Fally, 2015) which, although they will not be used in our present work, ensures that our estimations are as robust as possible when using the reduced-form gravity proposed in Feenstra (2003).

data and the inclusion of such effects would cause our dummies of interest to be dropped due to multicollinearity problems.

4. Data

This section summarizes the sources used for data collection and the decisions taken during the preparation of the dataset.

The data used in the present paper comes from different sources. Bilateral trade flows come from the International Trade and Production Database for Estimation V.2 (Borchert, I., Larch, M., Shikher, S., and Yotov, Y., 2022). This dataset is constructed at the industry level, containing 170 industries from the agriculture, mining and energy, manufacturing and service sectors, over the period from 1995 until 2017. However, we will focus exclusively on the manufacturing sector data.

In terms of the scope of countries, the present dataset covers all world trade flows, including small islands and city states. However, due to a process of elimination caused by missing values and perceived lack of consistency, the relevant subjects in our study are relegated to Ireland, the Netherlands, Germany, Austria, Denmark, Great Britain, Finland, Spain, Portugal and Sweden, representing the EU15 members, the Czech Republic, Slovakia, Slovenia, Estonia, Hungary, Latvia and Poland, as representatives of the CEEC expansion and, finally, Romania and Bulgaria as joiners of the last expansion analyzed.

Despite the obligation to eliminate some of the study subjects, using the ITPD-E V.2 has several advantages: (1) We are able to estimate the border effect over time, using 1999, 2006 and 2014 as reference years, which allows us to analyze the evolution of the European integration process through this variable along time. (2) The EU is not an independent entity, but is part of a much more complex system, so using this database allows us to analyze European integration at the same time we are accounting for the effects of the globalization process, which greatly enriches the resulting analysis. (3) The inclusion of both (real, not estimated) inter- and intranational trade ensures consistency with the gravity theory (Yotov, 2012), where consumers can choose between consumption of domestically produced goods and goods imported from a foreign country. Moreover, it solves the so-called “distance puzzle” in trade, by allowing us to measure the effects of distance in international trade in relation to the effects of distance at the level of intra-national trade. But, above all, it allows us to capture the effects of globalization and to correct for any possible bias in the estimation of the impact of an RTA, leading to a theoretically consistent identification of the effects of multilateral “discriminatory” trade policies, as it is the one analyzed in this paper.

On another note, although the ITPD-E database has 120 industries in the manufacturing sector, we have chosen to group the data into industries at the three-digit NACE code level, in order to be consistent with the harmonization analysis presented in 2014 by the European Commission, offering a more trustworthy explanation of the general picture of the European Integration process in terms of trade flows. In addition, the sectors related to TV transmitters, Electronic Valves, Coke Oven Products, Publishing of recorded media and other printing services have been eliminated, due to the fact that either no suitable match with NACE codes has been found, they were not considered a clear manufacturing sector, or their estimates resulted in a border effect equal to 0, contaminating our analysis. Therefore, the resulting dataset consists of information from 69 NACE industries at the three-digit level.

In order to enrich the information of our dataset, we have obtained the rest of the variables of interest from the CEPII distance dataset (Mayer, T. and Zignago, S., 2011), as well as from the dynamic gravity dataset (Borchert, I., Larch, M., Shikher, S., and Yotov, Y., 2022), also provided by the International Trade Commission of the United States. From these datasets we obtain factors that we consider relevant to use as a proxy of time varying and time invariant trade costs, such as language or agreed FTAs and, above all, membership of the European Union, used to construct the border dummy for the European Union countries and the different enlargement groups, as well as some of the interaction terms. It is important to mention that the FTAs dummy variable has been set to 0 between EU members of the corresponding years in order to avoid possible multicollinearity issues and capture effectively the effects of the EU membership.

Finally, for the purpose of finding the forces that determine border effects at industry level, we extracted information from the recently updated Rauch's classification and conducted a matching between the NACE classification and the SITC rev.2 codes. Rauch utilized the SITC rev.2 codes as a sector identifier, while we used the Harmonized System codes as an intermediary identifier during the process.

Ellison-Glaeser index's gammas have been obtained from Ellison and Glaeser (1997), performing a matching between US industries and NACE codes.

Ultimately, our harmonized dummy variable was derived from the findings presented in the "Study on the costs and benefits of the revision of the Mutual Recognition Regulation," which was provided by the European Commission. Moreover, we have extracted the calculations found in Gasiorek and Tamberi (2023) for the number of regulations per industry, understood in our resulting regressors as the Single Market's regulatory intensity per industry.

The final dataset contains 341,476 observations, although the number of observations varies depending on the year.

5. Main results: Analysis of the geographical dimension

5.1. EU and Non-EU Analysis

In this section, we present the main results obtained for EU border effects. A comparison of two different scenarios and different year cross-sectional data is offered to provide a comprehensive picture of the evolution of the European integration process in comparison to the globalization progress. Starting from the most general picture, we proceed step by step to fully granulate the border effects until reaching the country level estimations.

First of all, we have created different indicator variables using as a basis the border dummy explained in section 3, $border_{ij} = 0 \quad \forall\ i = j$, interacting it with a time-dynamic indicator of EU membership, resulting in the following border indicator variables: $border_{NoEU}$ for all those countries that are not members of the EU by 2014, $border_{EU}$ for those who are by 2014, $border_{EU15}$ for grouping the countries that represent the "Old Europe", $border_{EU_{CEEC}}$ for the group of countries that become members in the "CEEC enlargement" and finally $border_{EU_{LastJoiners}}$ for Romania and Bulgaria, the last members of the EU to be considered in the present study.

In order to further capture the deeper integration effects of becoming a member of the EU, we also include in the equation the interaction terms with EU trade binary variables, which take a value of 1 when trade happens between two EU member countries and 0 otherwise. The resulting regressors are summarized as: $border_{EU \times EU_{trade}}$, $border_{EU15 \times EU_{trade}}$, $border_{EU_{CEEC} \times EU_{trade}}$ and $border_{EULJoiners \times EU_{trade}}$. Thanks to this approach, we are not only able to estimate the difference between the reduction of the border effect due to globalization and its reduction due to the further integration efforts fostered by European Union integration, but we are also able to compare the different levels of economic integration in the European Union depending on the year of accession.

Thus, recalling the model econometric specification presented in subsection 3.2, in this finer estimation we can disaggregate the formula as follows:

$$X_{ij,k} = e^{\left(\pi_{i,k}+\chi_{j,k}+\phi_1 ln(distance_{ij})+\phi_2 adjacency_{ij}+\phi_3 language_{ij}+\phi_4 fta_{ij}+\beta border\right)} \times \epsilon_{ij,k} \tag{1-6}$$

Where now, as a result of the approach we have taken to compute our border variable, our finer estimation of beta coefficients has been converted into a vector with the following composition:

$$\begin{aligned} \beta border_{ij} = {} & \beta_1 border_{NoEU_{ij}} + \beta_2 border_{EU15_{ij}} + \beta_3 border_{EUCEEC_{ij}} + \beta_4 border_{EULJoiners_{ij}} \\ & + \beta_5 border_{EU15_{ij}} \times EUtrade_{ij} + \beta_6 border_{EUCEEC_{ij}} \times EUtrade_{ij} \\ & + \beta_7 border_{EULJoiners_{ij}} \times EUtrade_{ij} \end{aligned} \tag{1-7}$$

Albeit it might initially appear confusing, the results of the estimations are actually straightforward to analyze. If we did not include the interaction terms, we would interpret the betas as the quantification of the average border effect of the corresponding group. For instance, if we interpreted β_2 in a regression with no interactions, its magnitude would correspond to the average size of the international border effect corresponding to the EU15 countries we are considering in our estimations (Ireland, the Netherlands, Germany, Austria, Denmark, Great Britain, Finland, Spain, Portugal and Sweden).

However, when introducing those interaction terms, we are actually "doubling" the border effects we estimate. Namely, the role of the interaction terms is to capture the influence that trading within the Single Market has on the border effects of European Members. Therefore, when we account for the influence of trade within the Single Market, the estimated β_2 now represent the quantification of the border effect for EU15 countries when they trade outside the European Union. On the other hand, the sum of β_2 and β_5 provides the quantification of the same border effect when they trade within the Single Market. This additional information enables us to better understand the impact of trading both inside and outside the European Union, which we hypothesize it is crucial to correctly assess the impact of policy decisions that directly affect the bilateral relationships between the member countries.[2]

The results of this first approximation are summarized in Table 1, which shows the estimations for the same world scenario with and without controlling for Intra-European Union trade. Readers should note that, while the coefficients will be presented in negative terms, we will make the interpretation of their absolute value. The negative signs of the coefficients are due to the method of construction of the indicator variables ($border = 0 \;\; \forall i = j$ instead of $border = 1 \;\; \forall i = j$). As we are working with dummy variables, the interpretation remains consistent if we focus on the absolute values of the βs. In fact, we could reinterpret the coefficients as "how much less does a country trade with their partners compared to how much it trades with itself", that is, the inverse of the border effect concept. However, in order to continue the narrative of previous works, we will keep working with the original concept using this "workaround" with absolute values.

[2]From this point, whenever we mention that we are "controlling or accounting for intra-European trade" we would be referring to the inclusion of the presented interaction term in our estimations. Moreover, if our explanations involve β explanation, we would be also referring to the sum of βs of the border variable and intra-eu trade interaction term.

Table 1: Average EU and NonEU border effects; Cross-section estimation for 2014 (With and without controlling for Intra-EU trade)

	(1)	(2)	(3)	(4)
($\hat{\phi}_1$) ln $distance_{ij}$	-0.782***	-0.680***	-0.777***	-0.700***
	(0.094)	(0.113)	(0.094)	(0.111)
($\hat{\phi}_2$) $adjacency_{ij}$	0.384***	0.394***	0.393***	0.422***
	(0.115)	(0.119)	(0.117)	(0.123)
($\hat{\phi}_3$) $language_{ij}$	0.346***	0.345***	0.330***	0.337***
	(0.105)	(0.110)	(0.107)	(0.104)
($\hat{\phi}_4$) fta_{ij}	0.498***	0.520***	0.503***	0.510***
	(0.139)	(0.135)	(0.139)	(0.136)
($\hat{\beta}_1$) $border_{NonEU}$	-2.838***	-2.954***	-2.861***	-2.947***
	(0.267)	(0.278)	(0.269)	(0.268)
$border_{EU}$	-1.778***	-2.406***	–	–
	(0.266)	(0.392)	–	–
$border_{EU \times EU_{trade}}$	–	0.568***	–	–
	–	(0.181)	–	–
($\hat{\beta}_2$) $border_{EU15}$	–	–	-1.704***	-2.082***
	–	–	(0.283)	(0.372)
($\hat{\beta}_5$) $border_{EU15 \times EU_{trade}}$	–	–	–	0.276*
	–	–	–	(0.147)
($\hat{\beta}_3$) $border_{EU_{CEEC}}$	–	–	-2.397***	-3.324***
	–	–	(0.189)	(0.409)
($\hat{\beta}_6$) $border_{EU_{CEEC} \times EU_{trade}}$	–	–	–	0.498***
	–	–	–	(0.187)
($\hat{\beta}_4$) $border_{EU_{LJoiners}}$	–	–	-3.232***	-4.677***
	–	–	(0.253)	(0.532)
($\hat{\beta}_7$) $border_{EU_{LJoiners} \times EU_{trade}}$	–	–	–	0.852***
	–	–	–	(0.237)
Origin-Industry FE	Yes	Yes	Yes	Yes
Destiny-Industry FE	Yes	Yes	Yes	Yes
Observations	113698	113698	113698	113698
Pseudo R2	0.968	0.968	0.968	0.968

Note 1: Coefficients estimated using Poisson Pseudo-Maximum Likelihood Estimator; The cluster-robust standard error is reported in parenthesis; * $p < .1$, ** $p < .05$, *** $p < .01$. Both Origin and Destiny-Industry Fixed Effects indicate whether it has been included country-industry fixed-effects in the estimations.

Note 2: Abbreviations include: NonEU, non-members of the European Union by 2014, EU, members of the European Union by 2014, EU_{15}, members of the Europe of fifteen, EU_{CEEC}, new members corresponding to the enlargements until 2004, $EU_{LJoiners}$, new members corresponding to the 2007 enlargement (Bulgaria and Romania). The $\times EU_{trade}$ indicates the interaction of the border with the dummy that absorbs the effects of trading between EU members (E.G. EU_{CEEC} with the rest of EU members).

From the first column (1) we already obtain highly statistically significant border coefficients for both EU and Non-European countries. Namely, $border_{EU}$ and $border_{NonEU}$ suggest us that, on average, EU countries trade almost $e^{1.778} \approx 5.9$ times more with themselves than with the rest of the countries, while the countries of the rest of the world trade on average up to $e^{2.838} \approx 17.1$ times more with themselves than with the rest of the countries, almost triple the size of the effect in European countries. However, when we control for intra-European union trade in column (2), it becomes evident that EU countries are not per se far more integrated in the international trade system than the rest of the world, as in this estimation we see that European members trade 11 times more with themselves than with partners outside the EU, while RoW countries are expected to trade 19 times more with themselves than with other partners (from almost triple to less than double the size of this effect).

Combining $border_{EU}$ and $border_{EU \times EU_{trade}}$, i.e. $border_{EU} + border_{EU \times EU_{trade}}$, results in the quantification of the real border effect inside the European Union. With this approach we demonstrate that, in reality, EU countries trade approximately 6 times more with themselves than with their fellow members, not with their partners outside the EU. It is now, that we control for intra-European trade, when we see the first demonstration that estimating the effects of a trade policy inside a group of countries without accounting which trade partners are involved in such agreement (and which ones remain outside) is not a proper approach, as the baseline scenarios differ tremendously.

If we look at the rest of the control variables, whose coefficients are denoted by letter ϕ, we observe that the distance coefficient is always between -0.68 and -0.78 despite controlling for intra-European trade, which is consistent with what the structural gravity equation literature tells us, although it is within a range slightly above the values reported in many previous studies (around -0.6), but still far from unity. Likewise, adjacency always presents a positive coefficient between 0.38 and 0.42, values that we also consider consistent with what we would expect to obtain, as well as language dummy coefficients do, with values around 0.34.

For its part, FTA coefficients seem to be constant around 0.5 even when we include the control for intra-European union trade, which makes sense as we recoded this dummy so as to ignore the fta agreed among EU members.

Returning to the effect we are analyzing, in column (3) and column (4) we find the border effects of Non-European countries together with a set of border effects corresponding to the "Old Europe", the countries of the CEECs expansion and the last countries to enter the EU (Bulgaria and Romania, excluding Croatia). In this case we observe that the idea we are trying to highlight becomes even more evident, above all if we focus on the comparison of CEEC and RoW countries.

Apparently, if we ignored the impact that the Single Market has on trade among EU members, one could state that CEEC countries are in fact more economically integrated than the rest of the world, as in these estimations they are expected to trade $e^{2.397} \approx 11$ times more with themselves than with other partners while RoW countries do it by a magnitude of $e^{2.838} \approx 17.4$ times. Nonetheless, once intra-EU trade is controlled, CEEC countries are in fact less integrated in the international trade system than RoW ones, as their border effects, once we remove the impact of trading inside the EU, indicates an intra-national trade $e^{3.324} \approx 28$ times bigger than their trade that crosses borders, while the very same effect for RoW countries remains around $e^{2.947} \approx 19$.

Counterarguments could state that using a complete FTA regressor would capture the effects of being a member of the European Union, resulting in more trustworthy estimations of the border effect from EU members. While we do not disagree with this idea (in fact, working with the original FTA variable leads to similar results until we broke down the estimations by EU member groups), once again this approach runs the risk of making such effects to dilute, since they are mixed with the impact of other FTAs, just as it would happen if we analyse the impact of a policy that affects just a group of countries without accounting with which partner a country is trading.

Now, that we have offered a first snapshot of the border effect for both European and Non-European Union countries, we now introduce the time dimension to our estimations. As we do not want to overcomplicate further our analysis, the variable of time is introduced via the comparison of three different regressions using cross-sectional data for 1999, 2006 and 2014 (2014 estimations are also included in this table for a more convenient comparison). In Table 2 we find the results of the aforementioned regressions, both accounting and not accounting for intra-European trade.

Table 2: Evolution of the average border effects for EU members; Cross-section estimation for 1999, 2006, 2014 (With and without controlling for Intra-EU trade)

	(1) 1999	(2) 2006	(3) 2014	(4) 1999	(5) 2006	(6) 2014
$(\hat{\phi}_1)$ ln $distance_{ij}$	-0.656***	-0.851***	-0.777***	-0.420**	-0.768***	-0.700***
	(0.149)	(0.099)	(0.094)	(0.211)	(0.120)	(0.111)
$(\hat{\phi}_2)$ $adjacency_{ij}$	0.586***	0.417***	0.393***	0.580***	0.443***	0.422***
	(0.146)	(0.093)	(0.117)	(0.128)	(0.097)	(0.123)
$(\hat{\phi}_3)$ $language_{ij}$	0.186***	0.205***	0.330***	0.197**	0.216***	0.337***
	(0.064)	(0.071)	(0.107)	(0.079)	(0.072)	(0.104)
$(\hat{\phi}_4)$ fta_{ij}	0.695***	0.413**	0.503***	1.036***	0.442**	0.510***
	(0.205)	(0.206)	(0.139)	(0.308)	(0.214)	(0.136)
$(\hat{\beta}_1)$ $border_{NonEU}$	-2.893***	-2.557***	-2.861***	-3.298***	-2.658***	-2.947***
	(0.276)	(0.211)	(0.269)	(0.383)	(0.216)	(0.268)
$(\hat{\beta}_2)$ $border_{EU15}$	-2.539***	-2.004***	-1.704***	-3.774***	-2.418***	-2.082***
	(0.288)	(0.245)	(0.283)	(0.637)	(0.364)	(0.372)
$(\hat{\beta}_5)$ $border_{EU15\times EU_{trade}}$	–	–	–	1.100***	0.329*	0.276*
	–	–	–	(0.410)	(0.174)	(0.147)
$(\hat{\beta}_3)$ $border_{EU_{CEEC}}$	-5.081***	-2.914***	-2.397***	-5.847***	-3.920***	-3.324***
	(0.479)	(0.237)	(0.189)	(0.748)	(0.423)	(0.409)
$(\hat{\beta}_6)$ $border_{EU_{CEEC}\times EU_{trade}}$	–	–	–	0.325	0.481***	0.498***
	–	–	–	(0.274)	(0.160)	(0.187)
$(\hat{\beta}_4)$ $border_{EU_{LJoiners}}$	-6.992***	-4.322***	-3.232***	-8.265***	-5.023***	-4.677***
	(0.462)	(0.435)	(0.253)	(0.761)	(0.560)	(0.532)
$(\hat{\beta}_7)$ $border_{EU_{LJoiners}\times EU_{trade}}$	–	–	–	0.616**	0.551*	0.852***
	–	–	–	(0.307)	(0.303)	(0.237)
Origin-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Destiny-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	113757	114021	113698	113757	114021	113698
Pseudo R2	0.963	0.956	0.968	0.964	0.956	0.968

Note 1: Coefficients estimated using Poisson Pseudo-Maximum Likelihood Estimator; The cluster-robust standard error is reported in parenthesis; * $p < .1$, ** $p < .05$, *** $p < .01$. Both Origin and Destiny-Industry Fixed Effects indicate whether it has been included country-industry fixed-effects in the estimations.
Note 2: Abbreviations include: NonEU, non-members of the European Union by 2014, EU, members of the European Union by 2014, EU_{15}, members of the Europe of fifteen, EU_{CEEC}, new members corresponding to the enlargements until 2004, $EU_{LJoiners}$, new members corresponding to the 2007 enlargement (Bulgaria and Romania). The $\times EU_{trade}$ indicates the interaction of the border with the dummy that absorbs the effects of trading between EU members (E.G. EU_{CEEC} with the rest of EU members).

At first sight we already observe that for the countries of the European Union the globalization process does not seem to be slowing down despite the 2008 crisis and the trigger of the so-called "de-globalization era". However, in the rest of the world group we do observe a certain reversal of this process, as its border effect decreases from 1999 to 2006 (albeit slightly) to grow again in 2014.

Thus, an important point to highlight from Table 2 is the idea that despite the fact that the process of European economic integration has had an observable impact in all member countries regardless of their year of entry into the Single Market, their average border effect with respect to world trade has not been adversely affected, at least in such a magnitude that would have reversed the globalization force.

In other words, despite entering a single market and activating all the mechanisms that cause a variation in multilateral terms of resistance, we do not observe that the countries of the European Union have become more isolated from the rest of the world in terms of trade flows. Therefore, we can say that either the creation of a single market does not in itself imply an increase in the isolation of its member countries if it is carried out correctly (that is to say, if further agreements are developed with external countries under the umbrella of the Single Market, acting as an instrument for increasing negotiating power), or else the process of globalization is impossible to stop even if different types of FTAs are developed, since the integration force brought about by the former can never be counteracted. This hypothesis is strengthened once we make a comparison between $border_{NonEU}$ and the rest of border dummies corresponding to the three EU member groups. While, as stated, we see a decreasing trend in the border effects of EU members, the rest of the world seems to be depicting a reverse trend.

Focusing on columns (4), (5), and (6), in which we account for trade between EU member partners, RoW border coefficient lowers from 3.298 to 2.658, which implies that countries outside the current European Union went from trading 27 times more with themselves than with other partners to trading approximately 14 times more. Nevertheless, a trend reversal of this effect is observed, since the same countries exhibit that they trade 19 times more with themselves than with any other partner by 2014. Determining whether this phenomenon is due to a “de-globalization” process, to the trade diversion caused by the Single Market, or to other forces that we are not taking into consideration in our estimations is beyond the scope of this paper’s objectives, although it leaves room for further investigation using this dataset.

Moreover, although we do not observe this reverse trend in the European groups, the decline in the non-EU border effect seems to have slowed down regardless of whether or not the impact of intra-EU trade is accounted for.

For the “Old Europe” we only observe a drastic drop in the border effect in 1999 estimations, some years after the creation of the Single Market, as the border coefficient of this group indicates that, by that year, EU15 countries tended to trade $e^{3.774} \approx 43$ times more inside their borders than with countries outside the European Union, while they just traded $e^{3.774-1.1} \approx 14$ times more with themselves than with their fellow partners of the Union. Despite this initial and strongly significant effect of the interaction term that captures trade among EU member countries, the corresponding βs are almost not statistically significant in 2006 and 2014 estimations. Aligned with the pattern of the border effect for EU15, $border_{EU15\times EU_{trade}}$ experiments a significant reduction in its magnitude from 1999 to 2006, and follows a downward trend in the following year, which indicates that the impact of entering the Single Market in integration terms reaches a plateau rapidly to slowly dilute as years pass by and country members become more economically integrated inside the international trade system. Hence, if we analyze the evolution of the sum $\beta_2 border_{EU15} + \beta_5 border_{EU15\times EU_{trade}}$ we see that the border effect of EU15 members once trade among them is accounted starts from 2.64 in 1999, lowers to 2.089 in 2006 and finally reaches 1.806 in 2014, meaning that EU15 members went from trading 14 times more with themselves than with their fellow partners to trade 6 times more inside their borders than with other EU members. Furthermore, the fact that, by 2014, $border_{EU15\times EU_{trade}}$ still exhibits a statistically significant β_5 is the first demonstration that, even though entering the European Union has a diminishing effect in integration terms along the years, the efforts of the European Commission and the rest of organisms of this community still favor a further economic integration of the eldest members among the European community.

For its part, the estimation of the border coefficients for the CEECs shows exactly the behavior we would expect, since the interaction term $border_{EU_{CEEC} \times EU_{trade}}$ is not statistically significant in 1999, result that is aligned with our hypothesis, as those countries were still not members of the European Union and thus, were not benefiting from the goods of being inside the Single Market. Conversely, results in 2006 and 2014 show both a statistically significant interaction term, indicating that becoming a member in the European Union has actually helped those countries to experiment a deeper economic integration through the reduction of the magnitude of the intra-national trade with respect to the trade flows with a particular group of countries with which it has strengthen their ties, both politically and economically.

In this case, in contrast to the pattern we saw with EU15 countries, the interaction term and the border coefficient present opposite trends. While $border_{EU_{CEEC}}$ coefficient exhibits a lower value for the most recent years, the value of the interaction term appears to be higher in 2014 than in the previous periods. Although we do not see much difference between the values of same β_5 in 2006 and 2014, it should be taken into account that interaction terms are relative to other coefficients. Thus, although an increase from 0.481 to 0.498 might not seem relevant, if we analyze it in relative terms, the border effect reduction of this interaction terms goes from 12.2% in 2006 to 14.7% in 2014. However, this 2.5% increase might still not seem such an increment, nor there is a significant increase if we analyze the relation of it in real terms. Namely, in 2006 the relation was $e^{3.920} \approx 50$ to $e^{3.920-0.481} \approx 31$, meaning that EU membership reduced border effect by a factor of 1.61, while in 2014 the relation was $e^{3.324} \approx 27$ to $e^{3.324-0.498} \approx 17$, a reduction by a factor of 1.64. Thus, from these second results we can conclude that, whereas becoming a member of the European Union has indeed had a significant impact in a deeper reduction of the border coefficients (at least when we only consider trade among other country members), such impact seems to have an upper threshold that does not allow to further reduce the border effects once the members have reached a certain level of integration in the Single Market (the aforementioned plateau). This conclusion does not come into conflict with the previous hypothesis that EU institution efforts act a as a catalyst of deeper economic integration among the country members, but highlights that those efforts could have a rather limited impact inside the Single Market, meaning that there could be forces pushing up the border coefficients that could not be tackled by politics nor by legislation. Some of these forces are analyzed in the following industrial level analysis.

All in all, we must note that even if we perceive a steady importance of EU membership among this group of countries, this steadiness has occurred in an scenario where globalization has pushed down its general border effect. Hence, once again we see signs indicating that even if the effects of becoming an EU member dilute overtime, they are still significant thanks to the efforts of the European Commission and the rest of institutions inside the Union.

Finally, the most notable and contradictory variations are to be found in the group of Bulgaria and Romania. Between 1999 and 2006, we observe the most drastic fall in both scenarios, result of their transition from a communist to a democratic government and the consequent international openness to the majority of countries considered in our dataset, but, in contrast with the CEEC countries, this fall does not seem to slow down too much for these two countries in Europe while it slows down considerably in the international environment between 2006 and 2014. In this case we detect that in all three adjustments the interaction terms are statistically significant, although the most significant one and the one with highest value (both in relative and absolute terms) correspond to the year 2014. Focusing on the combination of the border and the interaction term, we find that the border effect for this group once we account for intra-European trade indicates that in 2006 Romania and Bulgaria would have an intra-national trade $e^{5.023-0.551} \approx 87$ times higher than their trade with other EU members and $e^{5.023} \approx 157$ times with respect to their trade with the rest of the world, while in 2014 the very same effects accounted for $e^{4.667-0.852} \approx 45$ and $e^{4.667} \approx 100$ times, respectively.

Two key ideas must be taken away from this subsection. First of all, we see the first signs confirming an hypothesis that we could already observe in previous papers such as Spornberger (2021), that is to say, if our estimations point to the right direction the speed at which a country member integrates in the Single Market would be determined on the time period we are considering, as this process is heavily dependent on the point of start a country is found. In our case study, the fact that the EU15 group shows a slowing integration trend compared to the rest of joiners is mainly because the period we are considering starts when its integration process had already reached its plateau, while in case of the rest, to greater or lesser extent, it is still in its early stages.

The second idea to bear in mind is what we have been highlighting so far: The European Union has been successful promoting the free movement of goods. However, there is a threshold where further economic integration becomes significantly harder to achieve, meaning that the EU has yet to reach its full potential and it is still far from being a Single Market.

5.2. Country level analysis

We now turn to the analysis of the coefficients for each of the members of the EU. To do so, we adopt a more granular approach by replacing the dummy variables corresponding to the different EU member groups with a country-specific dummy (the NoEU dummy is kept in the estimations but removed in the following plots).

So as to offer a comparison of the very same effects, we show the estimations of the βs corresponding to the $Border_{country}$ and the combination of βs corresponding to $Border_{country}$ + $Border_{country \times EU_{trade}}$. Namely, the econometric specification in these estimations takes in the following form[3]:

$$X_{ij,k} = e^{\left(\pi_{i,k} + \chi_{j,k} + \phi D' + \beta_{NoEU} border_{NoEU_{ij}} + \sum^{c \in \{country\}} \beta border_c\right)} \times \epsilon_{ij,k} \tag{1-8}$$

Where $border_c$ indicates a vector containing the border effect variable and the interaction term capturing the effect that trading with EU member partners has on the border effect of the corresponding country. For instance, the variables of the vector corresponding to Germany will be:

$$\beta_{DEU} border_{DEU} = \beta_{DEU_0} border_{DEU} + \beta_{DEU_1} border_{DEU \times EU_{trade}} \tag{1-9}$$

[3]Equation (1-8) has been reduced for the sake of simplicity. At this point readers should note that D' corresponds to a vector including all the exogenous variables but the ones corresponding to national borders (distance, adjacency, language and fta).

Once again, the interpretation is exactly the same as the one presented in section 4. In this case, the $\beta_{country_0}$ corresponding to the border of an EU country has to be interpreted as the quantification of the border effect outside the European Union, i.e. how much does a European member trade with itself more than it does with another partner outside the Single Market. On the other hand, the sum of $\beta_{country_0} + \beta_{country_1}$ quantifies the border effect inside the EU, i.e. how much does a European member trade with itself more that it does with a fellow member of the Single Market.

Figure 1 shows the evolution of the border effect between 1999 and 2014 for the different member countries of the European Union, both accounting and ignoring the effect of intra-European trade in the respective coefficients. For better reference, all the coefficients and their cluster-robust standard errors are available in the table 9 of Appendix B.

Although with subtle differences, the two subplots offer us the same general picture we saw in the previous tables, depicting a funnel structure in both cases. Going from top to bottom, the two highest border coefficients by 1999 correspond to the two the last joiners in the EU, Romania and Bulgaria, which at the same time exhibit some of the most notorious reductions of the very same coefficient between the last 15 years. Close to these two countries we find the ones corresponding to the CEEC enlargement, depicting also a less drastic reduction in the border effects between 1999 and 2014. Finally, EU15 countries come together from the mid positions until the bottom of the plots, showing more insignificant reduction of the border effect between years as the level of the starting point decreases. Closing this group we find Great Britain, Ireland, Germany and the Netherlands, which show border coefficients between 2 and 1. Particularly interesting is the case of the Netherlands, which in 2014 shows a nonexistent border effect, meaning that the volume of trade flows inside the borders of this country are expected to be roughly the same as the volumes of international trade flows. However, given the size of the country, the estimations are not completely reliable, and we must be cautious and not rush into any conclusion without double-checking, as we were when we drop Belgium or Luxembourg given the fact that we found even more far-fetched (negative and statistically significant) border effect estimations.

Focusing on the differences between accounting for intra-European trade or not, these become less subtle once we dissect the plot by parts. In general terms, we can see that trade between EU partners does indeed reduce the border coefficient for almost all of the EU members, with the exception (highlighted in bold in the Table 9 of the Appendix B) of Germany, Ireland and the Netherlands, which have negative or nil $\beta_{country_1}$ coefficients in 2014 adjustment. However, for the rest of cases we find that all countries had in fact integrated way more inside the EU trade system than in the international trade system.

In case of Romania, the most relevant subject of the last joiners, we see that the general border effect lowers from 8.197 to 3.616, while when we account for intra-European trade, the very same coefficient decreases from 7.388 to 2.526. The same results are found for Bulgaria and all the CEEC countries, as it can be identified if we examine the relatively wider separation between 1999 and 2014 coefficients present in subplot (b) for the members of this group. In fact, as demonstrated in Table 9 from Appendix B, even though we observe a decrease in the border coefficients of these countries, from β_{1999} to β_{2014} ($\beta_{country_0}$ for years 1999 and 2014 respectively), the interaction terms, $\beta_{1999 \times EUtrade}$ and $\beta_{2014 \times EUtrade}$ (in this case $\beta_{country_1}$ for 1999 and 2014), exhibit a notable increment.

Another insight to point out from both plots and the corresponding table is the fact that, while these border coefficients are distributed around the value of $border_{EU_{CEEC}}$ showed in Table 1 and Table 2, all interaction terms are considerably higher than the $border_{EU_{CEEC} \times EU_{rade}}$ obtained in the aforementioned tables. These results lead us to believe that, despite the fact that we are accounting for any idiosyncrasy present in any industry or country, when calculating the average border effect of European Members and the influence of EU intra-trade policies, we are in fact infraestimating the real impact that such policies have for both inside and outside the single market.

A completely different picture is revealed when we focus our analysis on the EU15 countries. In this case the differences between coefficients are much smaller for the case in which intra-European trade is considered. This fact indicates, in line with what we concluded when analyzing the previous tables, that the countries that were already members in 1999 have already been satisfactorily integrated into the Single Market and, although they present smaller border coefficients within the EU, their reduction between 1999 and 2014 is more due to the globalization process than to the fact of being a member of the Union. Once again, the idea of a hypothetical threshold for the economic integration potential of creating a Single Market comes to light.

From both plots, we can come up with different conclusions:

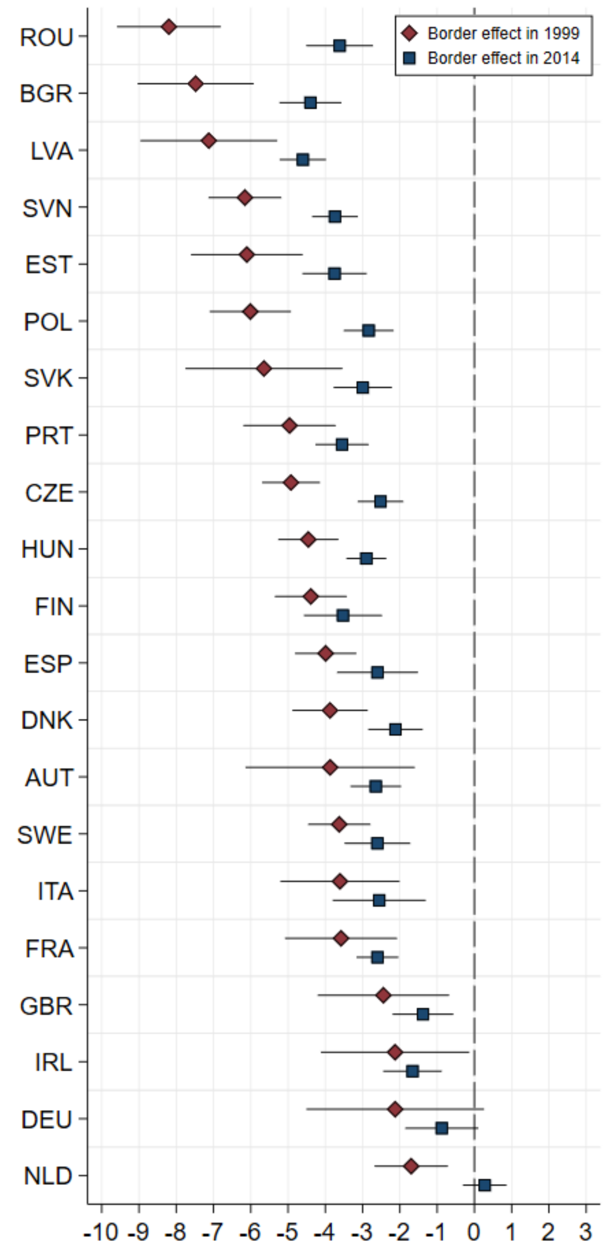
First of all, the fact that our estimations of the Single Markets' impact on the EU countries' border effect is bigger when we analyze it at country level makes us consider that it could be a better option to estimate country-specific, or even industry-specific, effects and average those results rather than estimating a common border dummy for those groups, otherwise we could be introducing a aggregation bias that led to a underestimation of the real impact of entering the Single Market.

Furthermore, now it becomes more than evident the hypothesis that a country exhibits a faster integration both inside the Single Market and in the international trade system depending on their starting point. That is, the economic integration process does not show a linear evolution but a logarithmic or logistic one, meaning that the impact of either opening the economy or entering an economic union such as the European Union has its biggest magnitude during the first years and starts losing weight as years pass by and a country becomes more integrated inside the trade system. However, this hypothesis is based on the evidence we found in our estimations, which is supported by different cross-sectional estimations. Hence, further studies should include a panel data approach in order to offer a more robust confirmation of these conclusions we have presented.

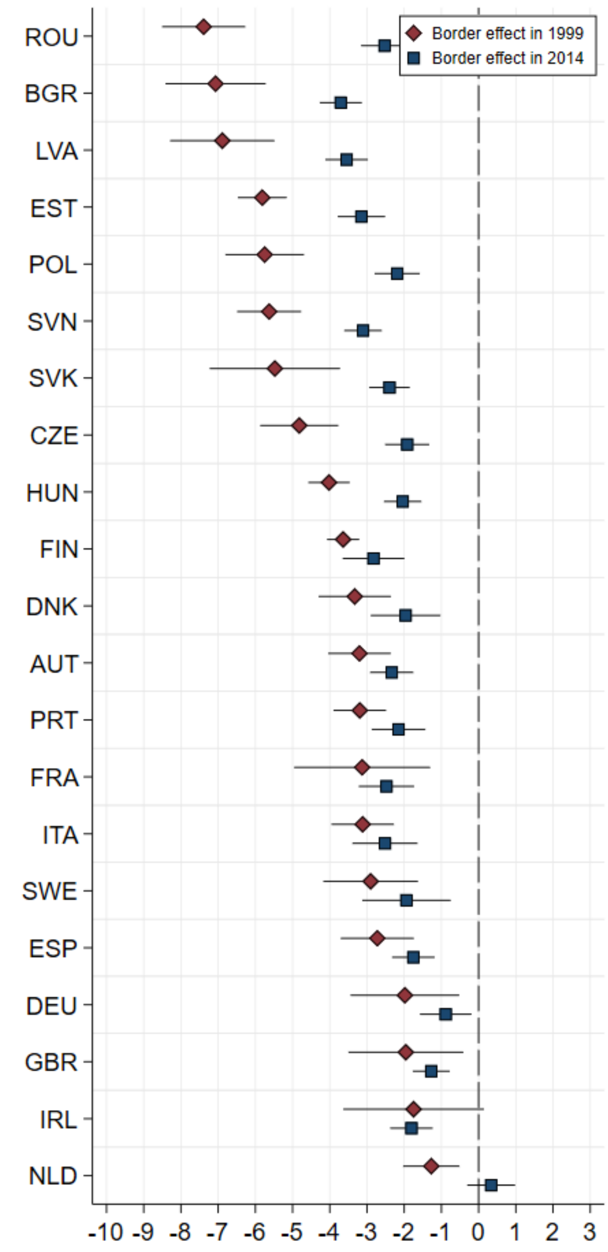
Finally, our estimations show that the progress of economic integration for EU15 countries has been quite timid between those 15 years if we compare it with the countries that joined later the EU. In line with the previous conclusion, we can support the aforementioned idea that the integration process has a threshold that, once crossed, slows down and reduces the impact of any effort for achieving further integration.

Figure 1: Border effect per EU Country; 1999 - 2014 comparison; World and EU integration scenarios (i.e. Including $Border_{Country} \times EU_{trade}$ interaction or not)

(a) World Border Effects evolution



(b) European Union Border Effects evolution



Note 1: This figure shows the comparison between 1999's and 2014's border effect of the representative European Union Members. In this plot we consider the evolution of $Border_{Country}$ and $Border_{Country} \times EU_{trade}$ between 15 years as a proxy to analyze EU members' integration evolution in the international trade system and the European Union itself.
Note 2: The coefficients have been estimated using the Poisson Pseudo-Maximum Likelihood Estimator, using cluster-robust errors, as well as exporter-industry and importer-industry fixed effects.
Note 3: The coefficients and their cluster-robust standard errors can be seen in the Table of the Appendix A.

6. Main results: Analysis at industry level

Turning to the analysis at the industry level, this section offers a comparative view of the estimated border effect for industries outside the European Union with respect to the same average effect for industries operating inside the Single Market in 2014. It also analyzes the differences of this effect inside the EU when we do and do not account for intra-European trade, offering a comparison based on the estimations for the cross section of 2014.

In this case, we compute iteratively 69 different coefficient groups, using only the observations of each industry per fit (approximately 1600/1700 records per regression). Hence, each of the regressions is constructed using the following specification:[4]

$$X_{ij,k} = e^{(\pi_i + \chi_j + \phi D' + \beta_1 border_{NoEU ij} + \beta_2 border_{EU ij} + \beta_3 border_{EU ij \times EU_{trade_{ij}}})} \times \epsilon_{ij,k} \tag{1-10}$$

Table 3 provides a summary of the basic descriptive statistics of the estimated border βs calculated for European and non-European industries when we remove the effect of intra-EU trade, as well as the result of a T-test for comparison of means to determine whether the differences between the average industry border effects are indeed statistically significant or not.

As we observe, the results depict one of the main ideas we have been observing in the previous estimations since, on average, the industries present a mean of $\hat{\beta}$s notably higher than the estimated $\hat{\beta}$s corresponding to the $border_{EU}$ and $border_{NoEU}$ in Table 1. In real terms this implies that, on average, when we analyze the effect from an industry level, countries outside the EU tend to trade not $e^{2.954} \approx 13.5$ but $e^{3.39} \approx 29$ times more with themselves than with the rest of the world, and European Members intra-national trade is as well no 11 but 15 times bigger.

So, by focusing just on the geographical granularity, we could in reality be ignoring certain casuistics that may lead to an underestimation of the actual border effect and the actual impact that

[4]Equation (1-10) has been reduced for the sake of simplicity. At this point readers should note that D' corresponds to a vector including all the exogenous variables but the ones corresponding to national borders (distance, adjacency, language and fta). On another note, fixed effects are not country-industry but just country specific, as we are running one regression per industry.

a Single Market has on reducing it.

Table 3: T-Test for mean comparison between average EU and World industries-level border

	Obs	Mean	Std. Err.	Std.Dev	[95% Conf. Interval]
(β_2) $Border_{EU}$	69	-2.70	0.17	1.44	-3.05 -2.35
(β_1) $Border_{RoW}$	69	-3.39	0.15	1.29	-3.70 -3.08
diff	69	0.69	0.13	1.09	0.42 0.95
mean(diff)	=	mean($Border_{EU}$ -	$Border_{RoW}$)		t = 5.228
H_0 mean(diff)	=	0			
H_a mean(diff)	>	0			Pr(T > t) = 0

Note 1: T-test for mean comparison using the 69 coefficients estimated for both EU and Non-EU industries when accounting for Intra-EU trade. This table is used as a proxy to analyze the difference between border coefficients once we have sustracted the influence of trading inside the European Union.
Note 2: The test has been applied assuming unknown and equal standard deviations/variance.
Note 3: The interpretation of the results have to be reversed. That is to say, the fact that $Border_{EU}$ is bigger than $Border_{RoW}$ implies that the border effect of European Members is still lower than the countries of the Rest of the World even when we substract the effect of intra-EU trade.

That is to say, contrary to the conclusions of Hillberry and Hummels (2008), Llano, C., Minondo, A., and Requena, F. (2011)., Coughlin and Novy (2021), as well as the results we have previously presented, we show that the larger the group of industries aggregated, the smaller the border effect is, so the pattern seems to be the opposite when industry granularity is the variable we take into account. Therefore, while geographical granularity is negatively related to the value of the border effect, it seems that industrial granularity could be positively related to the magnitude of the very same (i.e. the finer our samples, the greater the estimated border effect).

However, the main conclusion is identical, the use of disaggregated trade information plays a role in the estimation of the border effect and, by ignoring this fact we are running the risk of ignoring a problem of aggregation bias. Ideally, we would overcome such bias by using industry level trade among European regions (e.g. trade flows between Valencian industries with Venetian ones). Nevertheless, a complete and reliable data set with those characteristics is still to be developed.

Focusing on the results of the T-test we observe that the differences between the average border effect of EU and NonEU countries is actually statistically significant, as both their 95% confidence intervals have a nearly nil overlapping and the pvalue of the contrast is significant at 0.001 confidence level. This results are in line with what we have observed in Table 1: European industries, on average, are more economically integrated than the rest of the world in the international trade system, even if we control the impact of trading inside the Single Market.

If we further inquire into the details of the individual border effects by sector, found in Table 10 of Appendix C, columns $border_{NoEU_{2014}}$ and $border_{EU_{2014}}$, our results show a similar picture of Chen (2004) findings, as the NACE sectors with highest border coefficients still are the ones related with concrete and concrete products (235 and 236), metal structures (NACE 301 and 251), and products directly related to food and its derivatives (101, 105, 107 and 109). This first shallow finding makes logical sense, since we are speaking of products of either perishable nature such as the cement or food, or products with a relatively higher transport costs, which makes local consumers tend to prefer locally produced goods of this kind rather than foreign ones.

On the other hand, we must also point out the relatively higher coefficients for the **aircraft and spacecraft** (303) industry in EU with respect to the rest of the world, sign of the high protectionism that still persists in the European Union in this sector for both intra-regional and international trade, as the interaction term is also inappreciable and not statistically significant, implying that EU countries' trade in this sector is approximately 20 times bigger inside the borders of a country than outside independently of the partners we are comparing with.

In general terms, we see that the trend of EU border coefficients (without the effect of intra-European trade) being smaller than Non-EU's is consistent in most of the estimations but 15 (a 20% of the total estimations), which are put aside in Table 11 of the Appendix C for better reference. Although in some of the cases the differences are petty, we still find some notable dissimilarities, being the aforementioned aircraft and spacecraft the one with the most significant difference, followed by **other electrical equipment** (279) and **sawmilling and planning of wood** (161) and all food related industries.

As a matter of fact, in the manner of the general border coefficients of food manufacturing sectors are in most cases higher in EU countries, the corresponding values of the interaction terms also are among the biggest ones, meaning that borders do not exert as much opposition to international trade of manufactured food when the trading partners belong to the European Union, as they do when countries belong to the rest of the world.

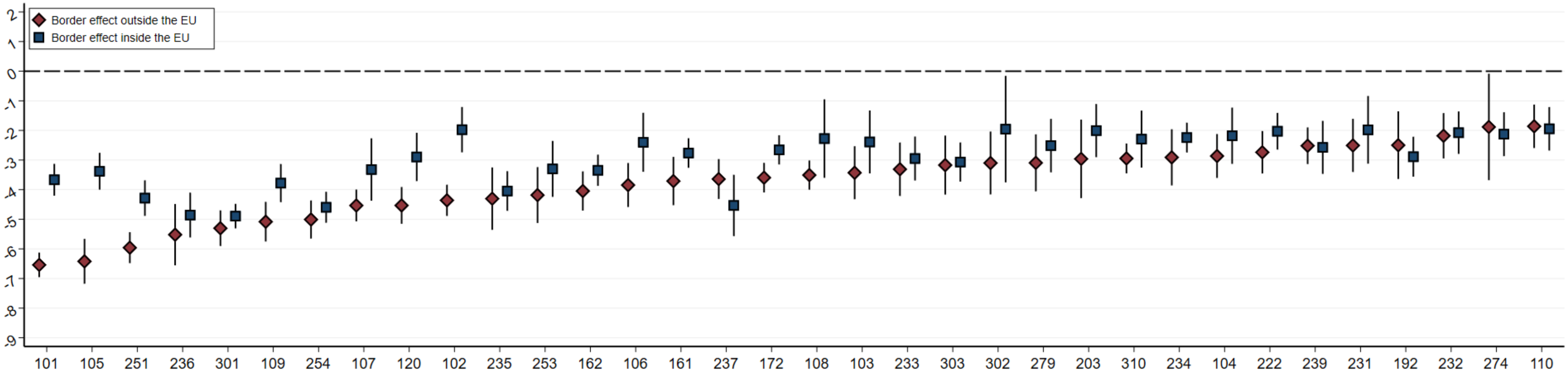
This fact can be easily seen in Figure 2, which shows the comparison between $Border_{EU_{2014}}$ (β_2) and $Border_{EU_{2014}}+Border_{EU_{2014}} \times EU_{trade}$ ($\beta_2 + \beta_3$) for the 69 industries subject to study. In the subplot (a) we observe that the food industries are in fact the ones with the greatest separation between both coefficients, specifically industry Processing of meat (101), Dairy products (105) and Processing of fish (102) are the ones that present the most significant differences between considering or not that the trading partners with which they trade are members of the EU.

Nonetheless, as we were expecting, the subplots show that, even after accounting for the idiosyncrasies of all industries and the relationships between the fellow members of the EU, in most industries trade between partners inside the European Union is systematically less deterred by national borders than trade occurring between member and non-member countries.

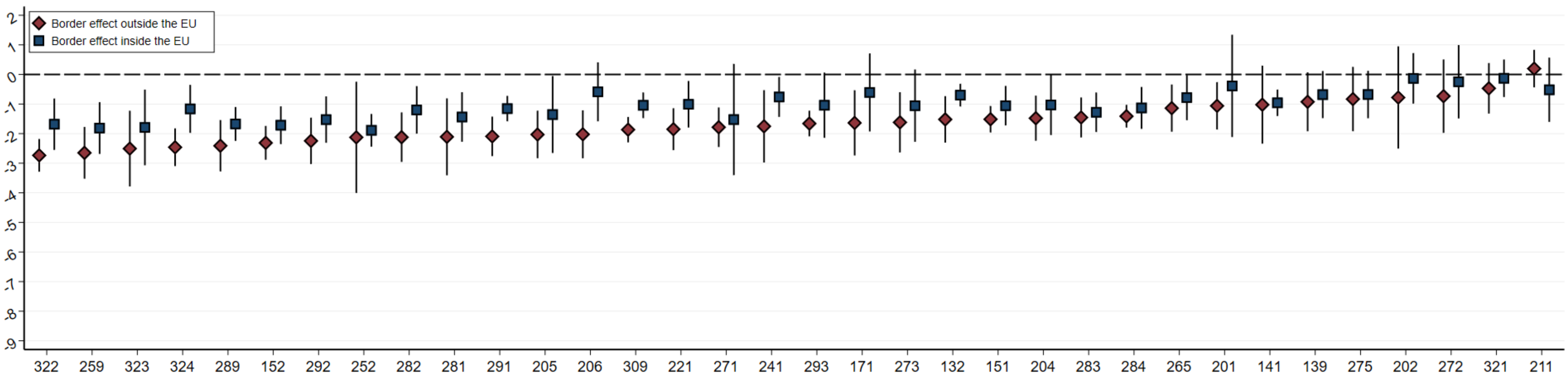
However, we also find some industries in which the interaction term is negative (highlighted in bold in Table 10 from Appendix C) and thus plays the other way round. Namely, the most relevant industries where we find that border coefficients are higher when we account for EU trade are Cutting shaping & finishing of stone (237), Refined petroleum products (192) and Basic pharmaceutical products (211). This behavior can be understood by pure logic for particular industries. For instance, it makes sense that refined petroleum products exhibit higher border effects inside the European Union, due to the energetic dependence of the country members. The Pharmaceutical sector also reflects a coefficient in line with the characteristics of the industry, as most of the production is spread across the globe so as to minimize costs. However, it is intriguing to see the highly significant and negative coefficient of Stone related industries.

Figure 2: Border effect at Industry level in the European Union; European Union Scenario (i.e. including the coefficient of $Border_{industry_{EU} \times EU_{trade}}$ interaction term); Intra EU and Outer EU comparison

(a) Industries with border effect above the mean



(b) Industries with border effect below the mean



Note 1: This figure shows the comparison between European Union border effect including and excluding the effect of trading inside the "border" of the European Union for each of the NACE industries analyzed, ordered by the size of the aforementioned effect outside the EU.
Note 2: The coefficients have been estimated using the Poisson Pseudo-Maximum Likelihood Estimator, using cluster-robust errors, as well as exporter-industry and importer-industry fixed effects. The coefficients and their cluster-robust standard errors can be seen in the Tables of the Appendix B.

In conclusion, from both Table 10 and Figure 2 we can observe the evident heterogeneous effects that trading inside the Single Market has on the border effect magnitude corresponding to each of the industries analyzed.

In general terms, although border barriers seem to be less an obstacle for trade inside the European Union, we observe that the most benefited industries from operating in an EU member are by far the ones corresponding to the food manufacturing sectors, while there seem to be other particular sectors where national borders exert the same resistance force to international trade regardless of the country where the trade partner belongs. Although we could attribute the heterogeneous effect of trading inside the Single Market to a stochastic effect, the fact that industries with similar characteristics exhibit similar patterns of trade inside and outside the European Union makes us believe that this is a systematic behavior rather than the result of pure randomness.

Being such heterogeneity of particular interest for our research, we now proceed to analyze different possible sources of it in the following section.

7. Exploring the sources of border effect heterogeneity at industry level

7.1. Explanatory variables

Now that we have seen the heterogeneous intensity and evolution of border effects at industry level, both in the rest of the world and, more importantly, inside and outside the European Union, it is time to find the reasons behind such variability among industries.

Several previous attempts aimed at finding and explaining the forces behind border effects have found significant explanatory variables for this phenomenon. In this section, we take over the helm of these works and try to shed some more light on this matter, putting at test some previously used variables, presented and explained in the following lines, as well as new ones never used in an analysis of this nature.

The first force considered is connected to the idea of economies of agglomeration and the spatial clustering of firms acting as endogenous factor that makes border effects arise. Our hypothesis is based on the rational of Home Market Effect (HME), described by Krugman (1980) and Helpman and Krugman (1985) as the fact that a country with the highest consumption of a product will in the final analysis incur in a trade surplus of the very same good, as it will host a more than proportionate share of firms with respect to domestic consumption, that is to say, countries with higher demand for some products locally tend to have higher sales of the very same product in foreign markets due to the fact that the inherent nature of companies make them tend to cluster in locations with higher demand so as to minimize transportation costs and increase returns to scale.

On the other hand, beyond the Home Market Effect logic, industries with low geographic dependency, either because they do not rely on a specific raw material or do not benefit from agglomeration externalities, are expected to be more prone to relocate in different locations so as to minimize border related costs.

Thus, we would expect that the spatial clustering of firms should contract border effects, showing a clear negative relationship between industrial concentration levels and border coefficients. Chen (2004) empirically demonstrates, using the gammas of Ellison and Glaeser index (1997), that industries with smaller values (i.e. less industrial concentration) showed on average larger border effects. Posterior studies, such as Llano and Requena (2010) further support the thesis of geographic concentration of industries being a driver of border effects reduction, as they concluded that a small value of the Ellison and Glaeser index was directly associated to larger border effects.

In this paper, we recover the gammas calculated in the original paper of Ellison and Glaeser (1997), which measures the extend to which firm's production is deeply rooted in a particular location, and perform a matching between US industries and the EU NACE codes we use in our study. In order to focus on the effect of economies of agglomeration among EU industries, we interact the obtained gammas with the $border_{EU \times EU_{trade}}$ dummy, resulting in the triple interaction variable $border_{EU \times EU_{trade} \times \gamma}$. If our hypothesis is ought to be true, we would expect that this variable depicted a negative relation with the raw border effect, confirming in our empirical analysis that a higher industrial concentration tends to reduce the border effect of an industry due to the presented reasoning.

Other factor that could influence on the variability of border effects across industries is the product differentiation. We find certain controversy when studying previous answers to this question. Chen (2004), based on the "search costs" logic of Rauch (1999), and supported by the reported estimations, concludes that differentiated goods would exhibit a larger border effect due to the fact that there is less information available and the search costs would be consequently higher, deterring consumers from buying foreign products. Hence, if these assumptions are correct, in our estimates we would expect to obtain larger border effects for differentiated products whilst the smallest ones correspond to referenced priced and goods traded on an organized exchange.

Conversely, Evans (2000) presents a completely opposite explanation. Based on the argument of elasticity of substitution, the empirical results show that in fact less elastic products, per se more differentiated, exhibit smaller border effects independently on the search costs that it carries importing them. Supporting this thesis, Llano and Requena (2010), focused on Spain as its subject of study, also demonstrate that a higher product differentiation is associated with a lower border effect, being the commodities the ones that show both a larger border effect and lower values in the three proxies they used as differentiation regressors (Intra-industry trade, relative Research and development expenditure and Advertising expenditure).

Parting from this controversy, we try to offer a new vision of this either catalyst or deterrent of border effects using the updated Rauch's classification. So as to adapt it to our estimations, we have matched the NACE codes of our industries with the ones corresponding to SITC rev.2, finally opting for the "liberal" classification's of w (homogenous goods or products on "organised exchange" markets), r (reference priced) and n (differentiated products) categories as our indicator variables. Thus, we have created two new interaction terms with the dummy $border_{EU \times EU_{trade}}$, i.e. $home_{EUx \times EU_{trade} \times w}$ and $home_{EU \times EU_{trade} \times n}$, to capture any possible differences among this three groups of industries.

Namely, we find that 45 industries correspond to the "n" classification, 17 correspond to the "r" and only 7 to the "w". Due to our approach in our estimations, the three variables cannot be used in the same adjustment as we would incur in multicollinearity problems. Moreover, our first estimations showed no statistically significant differences between the border effects of r and w categories. Thus, the referenced price industries (categorized as "r") will be used as the reference level in the following estimations.

In line with the impact of the reduction of non-tariff barriers, we propose a new variable to analyze the impact that the elimination or attenuation of technical barriers to trade (TBT) has had through the harmonization of technical standards pursued by the European Commission.

Taking into account the premise that the higher and more difficult it is to comply with the technical standards of a product in a foreign country, the higher the export costs will be, if the goods are regulated by different technical standards, these costs will increase, which will consequently lead to a decrease in the sale and purchase of foreign products and, therefore, magnify the border effects of certain industries. In summary, the imposition of different technical standards is considered one of the main non-tariff barriers to trade, and is therefore a force that increases border effects.

Following this reasoning, one would expect that sectors that, according to the European Commission, have harmonized their technical standards would register a greater difference when we compare the EU industries' border coefficient with and without removing the effect of intra-EU trade, as well as a lower border coefficient compared to sectors that have not yet undergone such harmonization (i.e. the βs corresponding to $border_{EU_{ij,k} \times EU_{trade_{ij}}}$, would be higher when estimating it for an industry that has successfully completed the technical harmonization process).

As reference, we make use of the document "The Study on the Costs and Benefits of the Revision of the Mutual Recognition Regulation" of the European Commision (European Comission, 2017), which is aimed at examining the need for revising the MRR so as to facilitate the free movement of goods within the EU. This study focuses on the economic value of non-harmonized products, the costs associated with the current implementation and potential policy options for improving.

Along the analysis provided in this report, a defined list of NACE industries determines whether a sector is considered to have achieved the technical harmonization or it is still in the ongoing process. Hence, the summary table, available in the Appendix C of that document, provides our source of information for our harmonization indicator variable.

Hence, based on the premise of a causal relation between harmonization and a reduction of exporting costs, in the present paper we interact the $border_{EU \times EU_{trade}}$ dummy with the indicator variable harmonized, resulting in $border_{EU \times EU_{trade} \times harmonized}$, which captures the impact that trading inside the Single Market has on the border effect of an industry once it has successfully undergone the harmonization process.

Finally, and relatively in line with the previous variable, we make use of the Single Market regulatory intensity metric created by Gasiorek and Tamberi (2023), which stands for the number of regulations and directives firms outside the Single Market have to conform (and prove that they conform) in order to make their exports eligible to cross the EU border. By, once again, interacting this count variable with $border_{EU \times EU_{trade}}$, we obtain our last variable put at test, $border_{EU \times EU_{trade} \times reg.intensity}$.

If our hypothesis holds, we would expect to see a negative relation between this new regressor and the border coefficient for a simple reason: the Mutual Recognition principle of the Single Market. That is to say, once a product is lawfully placed inside the market of a country member, no border barrier can stop it from being sold in the market of other fellow member. Therefore, no matter the regulatory burden existent for one specific industry, if a product is manufactured within the EU, those regulations will be directly complied with by simply abiding by national laws, supposing this an advantage with respect to those products fabricated outside the borders of the European Union, which are the ones that will not escape from conforming them before their products enter the Single Market. In this scenario, EU products are relieved from an extra layer of regulatory burden, making the EU industries specially affected by such regulations more willing to trade outside their national borders when it comes to trade within the European Union than when it comes to trade with a third partner outside the Single Market (i.e. the border effect on these industries should be even lower when intra-European trade is taken into account).

In conclusion, the final specification of our econometric model can be summarized with the following formula[5]:

$$X_{ij,k} = e^{(\pi_{i,k} + \chi_{j,k} + \phi D' + \beta_1 border_{NoEU_{ij}} + \beta_2 border_{EU_{ij}} + \beta_3 border_{EU_{ij} \times EU_{trade}} + \beta_4 border_{interactions})} \times \epsilon_{ij,k} \tag{1-11}$$

Where border interactions is a vector composed by all the triple interaction variables we presented:

$$\begin{aligned} \beta_4 border_{interactions} = {} & \beta_{4a} border_{EU_{ij} \times EU_{trade} \times w} + \beta_{4b} border_{EU_{ij} \times EU_{trade} \times n} \\ & + \beta_{4c} border_{EU_{ij} \times EU_{trade} \times \gamma_{ij,k}} + \beta_{4d} border_{EU_{ij} \times EU_{trade} \times harmonized} \\ & + \beta_{4e} border_{EU_{ij} \times EU_{trade} \times reg.intensity} \end{aligned} \tag{1-12}$$

7.2. Analysis of results

Turning to the analysis of our estimations, Table 3 offers a summary of all the regressions performed with the aforemetioned variables.

Columns (1) and (2) display the β coefficients for w and n Rauch categories (β_{4a} and β_{4b}). At first sight, we find that the β corresponding to the industries with homogeneous products is negative and statistically significant. This result indicates that, contrary to the conclusions of Chen (2004) and supporting the ones obtained in Evans (2000) and Llano and Requena (2010), products that are homogeneous show higher border coefficients than the rest of products. In other words, column (1) is indicating us that the border effect corresponding to non-homogeneous products (either differentiated or referenced price goods, absorbed by variable $border_{EU \times EU_{trade}}$ in this case) is of the magnitude of 1.74 ($abs(\hat{\beta}_2 + \hat{\beta}_3) = abs(-2.393 + 0.652) = 1.74$), while homogeneous products hold a border coefficient equal to 3.17 ($abs(\hat{\beta}_2 + \hat{\beta}_4 b) = abs(-2.393 - 0.783) = 3.17$). In real terms this implies that European Members trade differentiated products $e^{1.74} \approx 5.7$ times

[5] Equation (1-11) has been reduced for the sake of simplicity. At this point readers should note that D' corresponds to a vector including all the exogenous variables but the ones corresponding to national borders (distance, adjacency, language and fta).

more with themselves than with other members of the Single Market, while trading homogeneous products almost $e^{3.17} \approx 24$ times more with themselves than with the other fellow members. From our standpoint, the reasoning behind these estimations is the degree of substitutability between domestic and foreign products:

From a consumer-centric perspective, it is logical to assume that when a product with high elasticity (i.e., a homogeneous product) is being purchased, transport and search costs will play a more important role in the purchase decision making. That is, since the purchased product will be approximately the same regardless of the location where it is bought, minimizing the acquisition costs will be the primary focus of any consumer (i.e. the cost of the product plus the cost of searching for that product and transporting it, which includes border related costs). Therefore, an industry that produces homogeneous goods will tend to exhibit greater border effects, as consumers will tend to rely on domestic products for the sake of avoiding transport and border related costs.

Exactly the opposite happens with differentiated products. In the matter that buying a German car does not equal buying a Japanse one, the same applies for every industry that produces differentiated products. In this sense, we encounter that consumer preferences act as a reducer of border coefficients, meaning that not only legislative actions but also culture spreading efforts can act as a catalyst of integration inside the trade system. In other words, diversity promotes economic integration.

Results in column (2) reaffirm this conclusions, as the interaction term corresponding to w is still negative and statistically significant, whereas the one corresponding to n does not reflect any significant difference with respect to the base case (referenced price products), although it seems that this type of products present slightly lower border effects.

The fact that β_3, corresponding to $border_{EU\times EU_{trade}}$, reduces significance and value once we introduce $border_{EU\times EU_{trade}\times n}$ makes us consider that the latter might be acting as a confounding variable, which further supports the idea that both differentiated and referenced priced products do not show a significantly different border effect level inside the Single Market.

In column (3) we explore the influence of spatial clustering in the border coefficients. Aligned with previous findings, the β_{4c} coefficient for γ shows a positive and statistically significant value, implying that a higher value of Ellison and Glaeser index are related to smaller border effects. Hence, our results still support the thesis that firms belonging to industries more strongly tied to a specific location are less prone to relocate and, conversely, firms from less "geographically-binded" industries are more easily relocated to minimize trade costs. Such re-locations imply that intra-national trade will be the main focus for those industries, leaving international trade at marginal levels and thus making border effects arise endogenously.

Reaching the most important part of the present section, in column (4) we identify the impact of the technical harmonization process inside the European Union trade flows. As we were expecting, the β_{4d} coefficient of the triple interaction $border_{EU} \times EU_{trade} \times harmonized$ is positive and statistically significant at 0.05 confidence level. Thus, we confirm empirically that the efforts performed by the European institutions to achieve homogeneous technical standards for manufacturing products has in fact contributed to a reduction of the border effects inside the European Union, reducing the non-tariff barriers derived from having different technical standards among members of the Single Market.

At the same time, as we were also expecting, the coefficient β_{4e} for $border_{EU \times EU_{trade} \times reg.intensity}$ in column (5) is positive and statistically significant at 0.001 level. This implies that industries with high regulatory intensity present lower border effects inside the European Union. Although this might seem contradictory, our hypothesis lies on the fact that Mutual Recognition has completely wiped out the regulatory burden that firms of those industries must face inside the Single Market. As a consequence, firms outside EU tend to face higher barriers to trade in industries with a high number of EU directives they must abide, making them less prone to increase the flow of goods with member countries. On the other hand, the very same occurs with EU firms, which have to face regulatory burdens outside the EU that they do not encounter inside the Single Market, making European firms more prone to trade with fellow members rather than outside the EU. All things considered, the regulatory intensity in certain industries could be acting as a subtle mechanism of trade diversion.

Table 4: Exploring sources of border effects; Cross-section data for 2014

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
($\hat{\phi}_1$) ln $distance_{ij}$	-0.677***	-0.677***	-0.677***	-0.679***	-0.678***	-0.674***	-0.673***
	(0.113)	(0.113)	(0.102)	(0.102)	(0.102)	(0.102)	(0.102)
($\hat{\phi}_2$) $adjacency_{ij}$	0.397***	0.397***	0.398***	0.393***	0.394***	0.400***	0.400***
	(0.120)	(0.120)	(0.123)	(0.122)	(0.122)	(0.123)	(0.123)
($\hat{\phi}_3$) $language_{ij}$	0.344***	0.343***	0.349***	0.344***	0.344***	0.348***	0.348***
	(0.111)	(0.111)	(0.113)	(0.115)	(0.114)	(0.114)	(0.114)
($\hat{\phi}_4$) fta_{ij}	0.521***	0.522***	0.517***	0.521***	0.521***	0.520***	0.520***
	(0.135)	(0.136)	(0.119)	(0.120)	(0.120)	(0.120)	(0.120)
($\hat{\beta}_1$) $border_{NonEU}$	-2.962***	-2.964***	-2.957***	-2.956***	-2.957***	-2.967***	-2.967***
	(0.281)	(0.282)	(0.283)	(0.284)	(0.283)	(0.284)	(0.283)
($\hat{\beta}_2$) $border_{EU}$	-2.393***	-2.396***	-2.403***	-2.420***	-2.430***	-2.413***	-2.418***
	(0.390)	(0.388)	(0.355)	(0.358)	(0.355)	(0.350)	(0.351)
($\hat{\beta}_3$) $border_{EU\times EU_{trade}}$	0.652***	0.564**	0.315*	0.432***	0.313*	0.190	0.187
	(0.196)	(0.276)	(0.177)	(0.166)	(0.187)	(0.207)	(0.204)
($\hat{\beta}_{4a}$) $border_{EU\times EU_{trade}\times n}$	–	0.134	–	–	–	0.003	-0.067
	–	(0.270)	–	–	–	(0.135)	(0.122)
($\hat{\beta}_{4b}$) $border_{EU\times EU_{trade}\times w}$	-0.783***	-0.694***	–	–	–	-0.697***	-0.742***
	(0.204)	(0.247)	–	–	–	(0.131)	(0.148)
($\hat{\beta}_{4c}$) $border_{EU\times EU_{trade}\times \gamma}$	–	–	6.028***	–	–	6.080***	5.419***
	–	–	(1.466)	–	–	(1.433)	(1.523)
($\hat{\beta}_{4d}$) $border_{EU\times EU_{trade}\times harmonized}$	–	–	–	0.225**	–	0.322***	0.232
	–		–	(0.102)	–	(0.117)	(0.143)
($\hat{\beta}_{4e}$) $border_{EU\times EU_{trade}\times reg.intensity}$	–	–	–	–	0.013***	–	0.007
	–	–	–	–	(0.005)	–	(0.005)
Origin-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Destiny-Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	113698	113698	113698	113698	113698	113698	113698
Pseudo R2	0.968	0.968	0.968	0.968	0.968	0.969	0.969

Note 1: Estimated using Poisson Pseudo-Maximum Likelihood Estimator; The cluster-robust standard error is reported in parenthesis; * $p < .1$, ** $p < .05$, *** $p < .01$.
Note 2: Both Origin and Destiny Fixed Effects indicate whether it has been included country-industry fixed-effects in the estimations.
Note 3: Abbreviations include: NonEU, non-Members of the EU by 2014, EU, Members of the EU by 2014, EU_{trade} indicates wether intra-EU trade is being accounted, $\times n$, $\times w$, $\times r$, correspond to Rauch's classification categories, $\times\gamma$ is Ellison-Glaeser index value corresponding to the industry of the observation, $\times reg.intensity$, indicates the regulatory intensity per industry, $\times harmonized$, indicates whether the European Commission claims that the sector has completed the harmonization process or not.

Finally, in columns (6) and (7), all interactions are included in the same adjustment. In column (6) all variables (excluding regulatory intensity) remain significant at the same level, and we do not see a change in signs. Now, the β corresponding to $border_{EU} \times EU_{trade} \times n$ is almost 0, indicating that, at least in our estimations, product-specific information costs are irrelevant when explaining border effects if we use this variables as proxy, but our hypothesis of product substitutability remains consistent, as the coefficient of $border_{EU} \times EU_{trade} \times w$ remains negative and highly significant. On the other hand, while γ variable maintains the same coefficient, the value of β corresponding to the triple interaction with the harmonization variable increases in value and becomes statistically significant at 0.001 level. In this regard, we could conclude that, once we account for substitutability of products and industrial concentration, the effects of technical harmonization become more than evident in their pursue to reduce border effects among EU members and foster economic integration inside the Single Market.

However, when we add the regulatory intensity to the equation, both this new variable and the harmonization's coefficients become not statistically significant. This behavior makes sense, since if we take a look to the Table 12 in Appendix D, the industries that have successfully harmonized their technical standards exhibit higher number of European directives and regulations. As a matter of fact, the mean number of directives for harmonized industries is 20 while the corresponding to non-harmonized industries is below 10.

7.3. Robustness checks

Concluding this section, we perform some robustness checks so as to reexamine the results we previously obtained. Namely, we fit different regressions using the βs (and the combination of them) corresponding to the industry-specific border coefficients, available in Appendix C, as dependent variable and all the ones used to create the interaction terms as the independent ones. Hence the final regressions take the following form:

$$\beta_2 \ (border_{ind_{EU}}) = c + \theta_1\gamma + \theta_2 w + \theta_3 r + \theta_4 n + \theta_5 reg.intensity + \theta_6 harmonized + \epsilon \tag{1-13}$$

$$\beta_2 + \beta_3 \ (border_{ind_{EU}} + border_{ind_{EU} \times EU_{trade}}) = c + \theta_1\gamma + \theta_2 w + \theta_3 r + \phi_4 n + \theta_5 reg.intensity + \theta_6 harmonized + \epsilon \tag{1-14}$$

$$\beta_3 \left(border_{ind_{EU} \times EU_{trade}}\right) = c + \theta_1 \gamma + \theta_2 w + \theta_3 r + \theta_4 n + \theta_5 reg.intensity + \theta_6 harmonized + \epsilon \quad (1\text{-}15)$$

As our data presents heteroskedasticity due to the fact that comes from coefficients with different level of significance, following Head and Mayer (2000) and Chen (2004), we adjust a weighted least squares regression to account for this lack of homogeneity of variance, where the weights are calculated as the inverse of the standard error of the industry-specific border coefficients.

Readers must note and recall that, although our interpretation were based on the absolute values of the βs, when we use them as dependent variables in these last regressions, their sign remains negative. Hence, if the θ coefficients exhibit a negative sign this will imply that the variable contributes to a bigger border effect, while the opposite will be concluded if the sign is positive.

Table 5 offers the results of the adjustment with equation (1-12). Although the results are consistent in terms of sign with the results we previously shared, there is a clear lack of significance in the majority of coefficients.

The $\hat{\theta}_1$ of γ from column (1) exhibits a positive coefficient, meaning that spatial agglomeration of firms is associated with larger border effects also when we do not account for intra-EU trade, which it is reasonable as a priori it should not be an exogenous variable that could be influenced by the trade partner selection in our estimations.

In terms of search costs or elasticity of substitution, corresponding to columns (2) and (3), not n nor w-type industries terms are statistically significant, despite the coherent results in terms of sign and magnitude of the coefficients. Despite the lack of signification, the signs of $\hat{\theta}_2$ and $\hat{\theta}_3$ still support our thesis that industries that produce homogeneous products are more prone to exhibit higher border effects due to the higher product substitutability and bigger role played by border costs for this kind of products.

Regularization intensity, column (5), does not exhibit a statistically significant coefficient either. Moreover, the fact that it shows a positive sign enters in conflict with the logic we were presenting in previous sections, putting regulatory intensity as a mechanism of trade diversion in favor of those country who benefit from mutual recognition.

Finally, the harmonized indicator variable does not show statistical significance when used in isolation in column (6) nor in conjunction with the rest of variables in column (7).

It is particularly interesting to see that when we use the general border of European Union industries as dependent variable (i.e. the one that EU manufacturing sector present when we consider trade that do not occur within the Single Market), the harmonization variable presents negative coefficients. This result indicates that border effects when trading with non-members might be greater for industries which have been correctly harmonized. Therefore, it could be that, in fact, the harmonization process, and not regulatory intensity, acts as the lever of trade diversion in favor of EU members. However, due to the lack of statistical significance, we cannot draw robust conclusions from these estimations.

Table 5: Robustness check (β_2 (ind_{EU}) as dependent variable)

	(1)	(2)	(3)	(4)	(5)	(6)
($\hat{\theta}_1$) γ	2.520	–	–	–	–	1.751
	(3.645)	–	–	–	–	(4.146)
($\hat{\theta}_2$) w	–	-1.178	–	–	–	-0.933
	–	(0.830)	–	–	–	(0.844)
($\hat{\theta}_3$) n	–	–	0.522	–	–	0.137
	–	–	(0.444)	–	–	(0.380)
($\hat{\theta}_4$) *reg.intensity*	–	–	–	0.006	–	0.011
	–	–	–	(0.013)	–	(0.014)
($\hat{\theta}_5$) *harmonized*	–	–	–	–	-0.130	-0.100
	–	–	–	–	(0.342)	(0.367)
Observations	69	69	69	69	69	69
Pseudo R2	0.007	0.029	0.000	0.003	0.002	0.039

Note 1: Estimated using Weighted-Least-Squares; The standard error is reported in parenthesis; * $p < .1$, ** $p < .05$, *** $p < .01$.
Note 2: Abbreviations include: n, w, and r correspond to Rauch's classification categories, γ is Ellison-Glaeser index value of the corresponding to the industry of the observation, *reg.intensity*, indicates the regulatory intensity per industry, *harmonized*, indicates whether the European Commission claims that the sector has completed the harmonization process or not.

In Table 6 we find the results of the adjustments with equation (1-13). In this case, we find similar results as the ones found in Table 5, as all coefficients lack of statistical significance but the one corresponding to the regularization intensity in column (5).

The industries that produce homogeneous products, once we consider intra-European trade, seem to not have a significantly different border effect with respect to the rest of industries, as the $\hat{\theta}_2$

of the corresponding variable (w), although still negative, has lowered its magnitude drastically, alongside the indicator variables for the rest of industry categories.

In this case, regularization intensity does seem to be statistically significant at 0.1 level, which supports our idea that higher regulatory burdens deter EU members from trading outside the EU in favor of trade inside the Single Market. Nevertheless, due to the results presented in Table 5, this results still need to be double checked in further studies. Moreover, it loses its significance once we use all variables in our estimations, partly due to the strong linear relationship with the harmonized variable that we highlighted in the previous subsection.

Finally, the harmonization's $\hat{\theta}_5$, as we were expecting, has changed its sign, which indicates that technical harmonization could be acting as a reducer of industries border effect. However, as we do not observe significance at least at a level of 10%, we cannot conclude confidently that industries with harmonized technical requirements do show smaller border effects.

Table 6: Robustness check ($\beta_2 + \beta_3$ $(ind_{EU} + ind_{EU} \times EU_{trade})$ as dependent variable)

	(1)	(2)	(3)	(4)	(5)	(6)
$(\hat{\theta}_1)$ γ	3.511	–	–	–	–	3.869
	(3.252)	–	–	–	–	(3.513)
$(\hat{\theta}_2)$ w	–	-0.304	–	–	–	-0.294
	–	(0.542)	–	–	–	(0.579)
$(\hat{\theta}_3)$ n	–	–	0.180	–	–	0.079
	–	–	(0.334)	–	–	(0.347)
$(\hat{\theta}_4)$ *reg.intensity*	–	–	–	0.013*	–	0.013
	–	–	–	(0.008)	–	(0.012)
$(\hat{\theta}_5)$ *harmonized*	–	–	–	–	0.096	0.043
	–	–	–	–	(0.302)	(0.339)
Observations	69	69	69	69	69	69
Pseudo R2	0.017	0.005	0.001	0.043	0.002	0.045

Note 1: Estimated using Weighted-Least-Squares; The standard error is reported in parenthesis; * $p < .1$, ** $p < .05$, *** $p < .01$.
Note 2: Abbreviations include: n, w, and r correspond to Rauch's classification categories, γ is Ellison-Glaeser index value of the corresponding to the industry of the observation, *reg.intensity*, indicates the regulatory intensity per industry, *harmonized*, indicates whether the European Commission claims that the sector has completed the harmonization process or not.

As a final check, the last table presented discloses the results of the last adjustment, using just the value of the interaction term as the dependent variable.

Once again, the conclusions previously reached still hold in the current adjustment. However, we still do not see significance at any level in most of the coefficients. In this case, the sign of the coefficients is reversed but, far from not being consistent, it is due to the fact that the dependent variable now is related to the difference between border coefficients when we account for intra-EU trade.

Specifically, we observe that the coefficient for the w-type industry exhibits a positive and significant value at the 0.05 significance level. This finding is in line with the consistent results presented in both Table 5 and Table 6. It suggests that the disparities in border effects between homogeneous products produced within the borders of the European Union and those produced outside of it are pronounced. This result becomes comprehensible when we reflect on the hypothesis underpinning the coefficient's significance: transportation and border-related costs exert a more substantial influence on homogeneous products compared to differentiated ones.

In an economic context where border-related costs are diminishing or in some cases eliminated, such as within the framework of the Single Market, it follows logically that the border effect on homogeneous products within this economic zone would be significantly smaller than the same effect observed outside of it.

On the other hand, n-type industries exhibit a coefficient with a negative value that is statistically significant at the 0.05 level. This implies that industries producing differentiated products show fewer differences between border effects when accounting for intra-European Union trade compared to not accounting for it. Once again, these results align with the previous reasoning, as the trade of differentiated products appears to be less influenced by border-related costs. This can be attributed to the fact that less substitutable products will experience a smaller reduction in trade even if border costs increase, simply because there are no perfect substitutes for them.

Finally, not the regulatory intensity nor the harmonized's indicator variable are statistically significant in any of the adjustments.

Table 7: Robustness check ($\beta_{Border_{EU_{2014}} \times EU_{trade}}$ as dependent variable)

	(1)	(2)	(3)	(4)	(5)	(6)
($\hat{\theta}_1$) γ	-0.299	–	–	–	–	0.091
	(2.080)	–	–	–	–	(1.991)
($\hat{\theta}_2$) w	–	0.875*	–	–	–	1.045***
	–	(0.442)	–	–	–	(0.370)
($\hat{\theta}_3$) n	–	–	-0.342*	–	–	-0.130
	–	–	(0.189)	–	–	(0.152)
($\hat{\theta}_4$) *reg.intensity*	–	–	–	0.007	–	0.004
	–	–	–	(0.006)	–	(0.005)
($\hat{\theta}_5$) *harmonized*	–	–	–	–	0.226	0.206
	–	–	–	–	(0.164)	(0.163)
Observations	69	69	69	69	69	69
Pseudo R2	0.000	0.055	0.047	0.017	0.028	0.173

Note 1: Estimated using Weighted-Least-Squares; The standard error is reported in parenthesis; * $p < .1$, ** $p < .05$, *** $p < .01$.
Note 2: Abbreviations include: n, w, and r correspond to Rauch's classification categories, γ is Ellison-Glaeser index value of the corresponding to the industry of the observation, *reg.intensity*, indicates the regulatory intensity per industry, *harmonized*, indicates whether the European Commission claims that the sector has completed the harmonization process or not.

In conclusion, our empirical findings do not withstand the robustness tests. Despite the consistent results we have derived in the preceding sections, the absence of statistical significance in several coefficient estimates poses challenges in asserting the actual relevance of any of the examined variables as explanatory factors for the diversity in industry border effects. Consequently, additional endeavors and alternative forms of analysis are imperative to arrive at dependable conclusions.

8. Concluding remarks

National and regional borders wield considerable influence over international trade by diminishing trade flows once products traverse them. Nevertheless, scrutinizing this empirical occurrence and its potential origins through the lens of the structural gravity equation continues to yield contentious responses and unveil various gaps in our comprehension and quantification of the underlying reasons.

Furthermore, when dissecting and analyzing economic integration within a single market, such as the European Union, through border effects, the assessment must extend beyond the impact only on EU member states, but also on the rest of the world, i.e., it must be analyzed from a general equilibrium perspective to properly consider both trade creation and trade deviation. Furthermore, since certain policy decisions may affect various groups of countries differently, it is also crucial to be able to monitor how and in what way a certain interest group is affected by a trade policy decision.

Therefore, using a structural gravity equation in combination with an interaction variable that accounts for trade inside the Single Market becomes imperative to faithfully capture the repercussions of policies within economic unions like the EU.

Through this approach, our study endeavors to scrutinize and elucidate the magnitude of border effects across varying levels of granularity – groups of countries belonging to the same RTA, individual countries, and distinct industries – over different time junctures (1999, 2006, and 2014). This multifaceted examination has been directed so as to address a pivotal question in its final stage: Have the technical harmonization endeavors undertaken by the European Commission indeed played a role in diminishing barriers within the EU's internal borders among member states, considering all relevant explanatory variables established in previous research papers?

Following such reasoning, firstly, we have tried to shed some light on the idea that we cannot analyze the true effects of a trade policy inside an economic union if we do not account for the deeper influence of such policy among the members it impacts. In our results, it is demonstrated that estimating one border coefficient, ignoring the implication of having one trade partner or another, does not lead to trustworthy conclusions even if we account for any idiosyncrasy of a bilateral relationship via fixed effects, as the magnitude of the border effect still severely differs depending on the trade partners we consider.

Secondly, once border effects are estimated at country level, we find that we could be incurring in an aggregation bias when estimating coefficients by group of countries. Supporting the conclusions reached in previous works, we show that the border effect estimations could not only be biased due to geographic aggregation when we estimate at national level but also when we perform pooled estimations, reflecting in our case an underestimation of the real impact of becoming a member of the EU. Moreover, the same pattern is found when the industrial level is taken into account, that is, pooled estimations yield lower border effects than the ones that are estimated industry by industry. Thus, we could conclude that the estimation of the real average border effect in the European Union could require to estimate several and then aggregate results, instead on relying on only one coefficient from a pooled estimation.

Thirdly, beyond the conclusions found regarding our methodology, our results reveal that, as we would expect, the speed at which a country member integrates economically in the European Union is heavily dependent on the point where it starts. As a matter of fact, we find some evidence that this process could be of logarithmic or logistic form, implying that the biggest waves of a country's economic integration occur at the beginning of the process and start losing momentum as years pass by. This pattern is found at different stages in our cross-sectional estimations, where the Bulgaria and Romania seem to be on the early stages, CEEC countries appear to have reached the plateau and are starting to lose momentum and finally the EU15 members have reached a threshold where it becomes exponentially difficult to further reduce the magnitude of the border effect.

In the second part of this paper, we focus our study on the industry level of economic integration. Our estimations show the notable heterogeneity of border coefficients once we include this level of aggregation. Interestingly, trading inside the European Union has a significantly heterogeneous effect depending on the industry or group of industries. The fact that similar industries show similar patterns of trade inside and outside the Single Market lead us to believe that this heterogeneity is not due to pure chance but it reflects an underlying pattern fueled by different determinants.

The analysis of the possible sources of the border effects' magnitude reveals that the elasticity of substitution hypothesis seems to make more sense than the search cost theory presented by Chen (2004), implying that the more elasticity of substitution, the smaller the resulting border coefficient. On the contrary, estimations including economies of agglomeration indicate that they might serve as an endogenous source of border effects. In this context, industries less constrained by a particular location tend to display more pronounced border coefficients.

Regulatory intensity a priori seems to be playing a role on the border effect magnitude inside and outside the Single Market, increasing the border coefficients outside the European Union while reducing them when restricting the study to the intra-EU trade due to the Mutual Recognition principle.

Finally, our results confirm that technical harmonization efforts to foster economic integration within the Single Market have actually reduced the border effects of the fellow members when we account for trade inside the Single Market.

Nevertheless, the robustness check through our sensitivity analysis casts doubts on our conclusions. Consequently, an additional endeavor to validate the results or a shift in our approach becomes imperative in order to provide a definitive resolution to the inquiries and issues raised throughout the paper.

9. Limitations and areas of improvement

Despite the efforts behind the writing of this paper, there is still many room for improvement. In the following lines, we present feasible improvements of the current work:

First, although the influence of time has been considered throughout the whole research, the primary focus on the industry dimension has forced us to simplify our estimations reducing the relevance of the time dimension in our study. Thus, further efforts to dig deeper in our findings could imply the retrieval of time relevance via the transformation of the raw dataset from cross-sectional to panel data. In combination with Sporberger's (2021) approach, adding a temporal layer to our estimations could be the last piece we need to reach a clearer conclusion of the influence of the different sources of border effect put a test, as some of them might have a dynamic influence along time, such as the harmonization variable. Moreover, with this extended approach we could depict the real shape of the economic integration process and determine in a more robust way if it approximates to a logarithmic or logistic form.

Second, data quality has been an issue in our research, since we have limited our estimations to no further than 2014 due to the inconsistencies and missing values found for more recent years. However, an improved version of the ITPD, the ITPD-S, is expected to be released in the following months. Being notably better adapted to the needs of our econometric approach and allowing general equilibrium modelling, the results of new estimations without changing the current approach could provide us more reliable, and robust results than the ones we find right now.

With the objective of being consistent with the European Commission report on the harmonization of technical standards, we have aggregated all the ITPD industries in NACE code groups (either geographical or industrial). However, throughout the whole paper we have been stressing the risk of incurring in an aggregation bias and obtaining unreliable results as a consequence of grouping data or pooling estimations without taking into consideration that there may be certain characteristics that would be obscured once data is aggregated in a more general group. Hence, maintaining intact the industry groups from the ITPD could be an option to ensure reliability of our results, while at the same time obtaining more signal for our robustness checks (more than 100 observations instead of just 69). As we have been in contact with the responsible of the creation of the ITPD-E database, we are in the know that a new version, the ITPD-S database, will be released in December. Hence, a new a quick adjustment will suffice to put into practice this refinement idea.

In this paper we have centered our attention on exogenous and endogenous forces that act as catalyst and suppressor of border coefficients. However, we have been neglecting a dimension of the border effect that plays a key role in its magnitude, the "home bias". Beyond the theoretical world, we know that in reality consumer preferences are not homothetic and identical across countries. Therefore, using proxy variables to capture consumer preferences, beyond the proxy we use for elasticity of substitution, is needed to calibrate the real effect of the variables we have already analyzed. Furthermore, not only private consumers, but also the public sector sins of "home bias" and the public tenders favoritism is still an issue in the international market. Therefore, capturing the influence of more cultural and social-related variables would be key to improve the conclusions of our present study.

Finally, in terms of endogenous sources of border effects related to geography, we only have used the Ellison-Glaeser index as proxy. However, we are still neglecting one powerful driver of geographic location of firms, intermediate goods, as industries focused on producing intermediates are more prone to locate in the surroundings of the focal points of concentrated industrial demands in order to minimize transport related costs. Including another variable that accounts for geographical concentration of firms would allow us to obtain more robust and trustworthy estimations by accounting for the effect of locality trade, which otherwise we cannot remove without having more granular data in terms of geographical units (provinces or regions inside EU countries).

In summary, our estimations have been successful in contributing to the pursue of untangling the border effect puzzle. However, there are still many questions unsolved behind this phenomenon, which could be partly addressed by the proposals presented. Furthermore, the aim of this working paper is to compose the first part of a thesis chapter. As it was mentioned, a new ITPD database will be released soon, which will overcome one of the limitations of the current ITPD-E, it only allows us to do Partial Equilibrium analysis. The next step thus is to continue the current efforts in a General Equilibrium analysis of the previous findings.

10. Acknowledgements

I am specially indebted to professor Francisco Requena for their invaluable mentoring, support and dedication throughout the process of writing this article.

I would also like to thank professor Vicente J. Pallardó and the entire team of professors from the International Economic Institute of the Universitat de València for a very enriching and insightful academic course.

Lastly, and most importantly, special thanks to my dear parents, who tried to teach how to be a curious, persistent, and punctual person. At least they succeeded in the first two. Without them, I would not have achieved even half of what I have accomplished up to now.

11. References

Anderson, J.E. and Van Wicoop, E. (2003). "Gravity with Gravitas: A solution to the Border Puzzle". American Economic Review, Vol.93, No.1 pp. 170-192.

Borchert, I., Larch, M., Shikher, S., and Yotov, Y. (2022). “The International Trade and Production Database for Estimation - Release 2 (ITPD-E-R02)”, USITC Working Paper 2022–07–A.

Borchert, I., Larch, M., Shikher, S., and Yotov, Y. (2021), “The International Trade and Production Database for Estimation (ITPD-E)”, International Economics, 166, 140–166.

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.

Cheptea, A. (2013). "Border effects and European integration". CESifo Economic Studies. Vol. 59. pp. 277-305

Coughlin, C.C. and Novy, D. (2021). "Estimating Border Effects: The Impact of Spatial Aggregation. International Economic Review Vol. 63(4) pp. 1453-1487.

Engel, C., and Rogers, J.H. (1996). "How Wide is the Border?". The American Economic Review 86(5), 1112-1125.

Ellison, G., Glaeser, E.L. (1997). "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach". Journal of Political Economy. Vol. 105,No.5. pp. 889-927.

Evans, C.L. (2000). "The Economic Significance of National Border Effects". The American Economic Review 93(4), 1291-1312.

European Commission, (2017). "Study on the costs and benefits of the revision of Mutual Recognition Regulation (EC) No 764/2008". Directoreate-General for Internal Market, Industry, Entrepreneurship and SMEs.

Feenstra, R.C. (2003). "Border effects and the gravity equation: Consistent methods for estimation". Scottish Journal of Political Economy Vol 49(5) pp.491-506

Fally, T. (2015). "Structural gravity and fixed effects". Journal of International Economics, Vol. 97, pp. 76-85.

Gallego, N., and Llano, C. (2015). "Thick and thin borders in the European union: How deep internal integration is within countries, and how shallow between them". The World Economy 38(12), 1850-1879.

Gasiorek, M., Tamberi, N. (2023). "The effects of leaving the EU on the geography of UK trade". Economic Policy.

Gil-Pareja, S., Lloca-Vivero, R., Martínez-Serrano J.A., and Oliver-Alonso, J. (2005). "The border effect in Spain". World Economy 28(11), 1617-1631.

Head, K., Mayer, T., (2000). "Non-Europe: the magnitude and causes of market fragmentation in the EU". Weltwirtschaftliches Archive 136 (2), 285-314.

Helliwell, J.F., V, G. (2001). "Measuring internal trade distances: A new method applied to estimate provincial border effects in Canada". The Canadian Journal of Economics. Vol.34, No.4 pp. 1025-1041.

Helpman, E. and Krugman, P. (1985). "Market structure and foreign trade: Increasing returns, imperfect competition and the international economy". Journal of International Economics. Vol.21 pp. 183-187

Hillberry, R.H. (2002). “Aggregation Bias, Compositional Change, and the Border Effect”. The Canadian Journal of Economics. Vol. 35, No.3. pp. 517-530.

Hillberry, R. and Hummels, D. (2008). "Trade response to geographic frictions: A decomposition using micro-data". European Economic Review. 52(2008) pp. 527-550.

Krugman, P. (1980). "Scale Economies, Product Differentiation, and the Pattern of Trade". The American Economic Review. Vol.70, No.5, pp. 527-550.

Requena, F. and Llano, C. (2010). "The border effects in Spain: an industry-level analysis". Empirica 37:455-476.

Llano, C., Minondo, A., and Requena, F. (2011). "Is the border effect an artefact of geographic aggregation?". The World Economy 34(10), 1771-1787.

Mayer, T. and Zignago, S. (2011). "Notes on CEPII's distances measures : the GeoDist Database", CEPII Working Paper 2011-25

McCallum, J. (1995). "National borders matter: Canada-US regional patterns". The America Economic Review 85(3), 615-623.

Nitsch, V. (2000). "National Borders and International Trade: Evidence from the European Union". Canadian Journal of Economics 33, 1091-1105.

Obstelf, M. and Rogoff, M. (2000). "The Six Major Puzzles in International Macroeconomics. Is there a common cause". NBER Macroeconomics Annual 2000, B.S. Bernanke and K.Rogoff, eds. Cambridge, MA: MIT Press.

Rauch, J. (1999). "Networks versus markets in international trade". Journal of International Economics Vol.48 pp. 7-35

Santamaría, M.A., Ventura, J. and Yesilbayraktar, U. (2021). "Borders whithin Europe". NBER Working Paper Series. Working Paper 28301.

Silva, J. S. and S. Tenreyro, (2006). "The log of gravity". The Review of Economics and Statistics 88(4), pp. 641-658.

Spornberger, J., (2021). "EU integration and structrual gravity: A comprehensive quanitfication of the border effect on trade". Review of International Economics 30(4), 915-938.

Wolf, H., (2000). "Intranational Home Bias in Trade". Review of Economics and Statistics 82(4):555-563.

Yotov, Y. V., (2012). "A simple solution to the distance puzzle in international trade". Economic Letters, 117(3), pp. 794-798.

Yotov, Y. V., Permaritini, R., Monterio and J.A., Larch, M. (2017). "An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model". UN, New York.

APPENDIX A (Correspondence tables)

Table 8a: ITPD-NACE codes correspondence

ITPD	ITPD Industry name	NACE	NACE ITPD Industry name
36	Processing/preserving of meat	101	Processing/preserving of meat
37	Processing/preserving of fish	102	Processing/preserving of fish
38	Processing/preserving of fruit & vegetables	103	Processing/preserving of fruit & vegetables
39	Vegetable and animal oils and fats	104	Vegetable and animal oils and fats
40	Dairy products	105	Dairy products
41	Grain mill products	106	Grain mill products, starches and starch products
42	Starches and starch products	106	Grain mill products, starches and starch products
43	Prepared animal feeds	109	Prepared animal feeds
44	Bakery products	107	Bakery and farinaceous
45	Sugar	108	Other food products
46	Cocoa chocolate and sugar confectionery	108	Other food products
47	Macaroni noodles & similar products	107	Bakery and farinaceous
48	Other food products n.e.c.	108	Other food products
49	Distilling rectifying & blending of spirits	110	Beverages
50	Wines	110	Beverages
51	Malt liquors and malt	110	Beverages
52	Soft drinks; mineral waters	110	Beverages
53	Tobacco products	120	Tobacco products
54	Textile fibre preparation; textile weaving	132	Weawing of textiles
55	Made-up textile articles except apparel	139	Other textiles
56	Carpets and rugs	139	Other textiles
57	Cordage rope twine and netting	139	Other textiles
58	Other textiles n.e.c.	139	Other textiles
59	Knitted and crocheted fabrics and articles	139	Other textiles
60	Wearing apparel except fur apparel	141	Wearing apparel, except fur
61	Dressing & dyeing of fur; processing of fur	151	Tanning and dressing of fur and leather
62	Tanning and dressing of leather	151	Tanning and dressing of fur and leather
63	Luggage handbags etc.; saddlery & harness	151	Tanning and dressing of fur and leather
64	Footwear	152	Footwear
65	Sawmilling and planing of wood	161	Sawmmilling and planning of wood
66	Veneer sheets plywood particle board etc.	162	Products of wood, cork, straw and plaiting materials
67	Builders' carpentry and joinery	162	Products of wood, cork, straw and plaiting materials
68	Wooden containers	162	Products of wood, cork, straw and plaiting materials
69	Other wood products; articles of cork/straw	162	Products of wood, cork, straw and plaiting materials
70	Pulp paper and paperboard	171	Pulp, paper, and paperboard
71	Corrugated paper and paperboard	172	Articles of paper and paperboard
72	Other articles of paper and paperboard	172	Articles of paper and paperboard
80	Refined petroleum products	192	Refined petroleum products
82	Basic chemicals except fertilizers	201	Basic chemicals, fertilisers, etc
83	Fertilizers and nitrogen compounds	201	Basic chemicals, fertilisers, etc
84	Plastics in primary forms; synthetic rubber	201	Basic chemicals, fertilisers, etc
85	Pesticides and other agro-chemical products	202	Pesticides and other agrochemical products
86	Paints varnishes printing ink and mastics	203	Paints, varnishes and coatings
87	Pharmaceuticals medicinal chemicals etc.	211	Basic pharmaceutical products
88	Soap cleaning & cosmetic preparations	204	Soap cleaning & cosmetic preparations
89	Other chemical products n.e.c.	205	Other chemical products
90	Man-made fibres	206	Man-made fibres
91	Rubber tyres and tubes	221	Rubber products
92	Other rubber products	221	Rubber products
93	Plastic products	222	Plastic products
94	Glass and glass products	231	Glass and glass products
95	Pottery china and earthenware	234	Porcelain and ceramic products
96	Refractory ceramic products	232	Refractory products
97	Struct.non-refractory clay; ceramic products	233	Clay building materials
98	Cement lime and plaster	235	Cement, lime and plaster
99	Articles of concrete cement and plaster	236	Articles of concrete, cement and plaster

Note 1: The present table shows all the codes from ITPD manufacturing sector and the corresponding NACE codes with which we have matched each of the industries.

Table 8b: (Cont.) ITPD-NACE codes correspondence

ITPD	ITPD Industry name	NACE	NACE Industry name
100	Cutting shaping & finishing of stone	237	Cutting shaping & finishing of stone
101	Other non-metallic mineral products n.e.c.	239	Other non-metallic
102	Basic iron and steel	241	Basic iron and steel
104	Structural metal products	251	Structural metal products
105	Tanks reservoirs and containers of metal	252	Tanks reservoirs and containers of metal
106	Steam generators	253	Steam generators
107	Cutlery hand tools and general hardware	257	Cutlery, tools and general hardware
108	Other fabricated metal products n.e.c.	259	Other fabricated metal products
109	Engines & turbines (not for transport equipment)	281	General-purpose machinery
110	Pumps compressors taps and valves	281	General-purpose machinery
111	Bearings gears gearing & driving elements	281	General-purpose machinery
112	Ovens furnaces and furnace burners	282	Other General-purpose machinery
113	Lifting and handling equipment	282	Other General-purpose machinery
114	Other general purpose machinery	282	Other General-purpose machinery
115	Agricultural and forestry machinery	283	Agricultural and forestry machinery
116	Machine tools	284	Metal forming machinery and machine tools
117	Machinery for metallurgy	289	Other special-purpose machinery
118	Machinery for mining & construction	289	Other special-purpose machinery
119	Food/beverage/tobacco processing machinery	289	Other special-purpose machinery
120	Machinery for textile apparel and leather	289	Other special-purpose machinery
121	Weapons and ammunition	254	Weapons and ammunition
122	Other special purpose machinery	289	Other special-purpose machinery
123	Domestic appliances n.e.c.	275	Domestic appliances
124	Office accounting and computing machinery	282	Other General-purpose machinery
125	Electric motors generators and transformers	271	Electric motors generators and transformers
126	Electricity distribution & control apparatus	271	Electric motors generators and transformers
127	Insulated wire and cable	273	Wiring and wiring devices
128	Accumulators primary cells and batteries	272	Batteries and Accumulators
129	Lighting equipment and electric lamps	274	Electric lighting equipment
130	Other electrical equipment n.e.c.	279	Other electrical equipment
135	Measuring/testing/navigating appliances etc.	265	Measuring/testing/navigating appliances etc.
137	Watches and clocks	265	Measuring/testing/navigating appliances etc.
138	Motor vehicles	291	Motor vehicles
139	Automobile bodies trailers & semi-trailers	292	Bodies for motor vehicles
140	Parts/accessories for automobiles	293	Parts/accessories for motor vehicles
141	Building and repairing of ships	301	Building and repairing of ships
142	Building/repairing of pleasure/sport. boats	301	Building and repairing of ships
143	Railway/tramway locomotives & rolling stock	302	Railway/tramway locomotives & rolling stock
144	Aircraft and spacecraft	303	Aircraft and spacecraft
145	Motorcycles	309	Transport equipment n.e.c.
146	Bicycles and invalid carriages	309	Transport equipment n.e.c.
147	Other transport equipment n.e.c.	309	Transport equipment n.e.c.
148	Furniture	310	Furniture
149	Jewellery and related articles	321	Jewellery
150	Musical instruments	322	Musical instruments
151	Sports goods	323	Sports goods
152	Games and toys	324	Games and toys

Note 1: The present table shows all the codes from ITPD manufacturing sector and the corresponding NACE codes with which we have matched each of the industries.

APPENDIX B (Country estimation table)

Table 9: Coefficients and Standard Errors of Country Border Effects; EU1999 vs EU2014

Country	β_{1999}	$\beta_{1999\times EU_{trade}}$	$\beta_{1999}+\beta_{1999\times EU_{trade}}$	$SE(\beta_{1999}+\beta_{1999\times EU_{trade}})$	β_{2014}	$\beta_{2014\times EU_{trade}}$	$\beta_{2014}+\beta_{2014\times EU_{trade}}$	$SE(\beta_{2014}+\beta_{2014\times EU_{trade}})$
ROU	-8.197	.809	-7.389	.746	-3.616	1.09	-2.526	.292
BGR	-7.479	.411	-7.068	.687	-4.399	.695	-3.704	.291
LVA	-7.126	.2349	-6.888	.932	-4.603	1.056	-3.548	.478
SVN	-6.157	.5286	-5.63	.965	-3.74	.634	-3.107	.353
EST	-6.106	.292	-5.814	.894	-3.752	.602	-3.149	.447
POL	-6.012	.263	-5.749	.649	-2.837	.648	-2.189	.303
SVK	-5.647	.171	-5.476	.788	-2.996	.601	-2.395	.254
PRT	-4.961	1.767	-3.194	.502	-3.552	1.4	-2.152	.291
CZE	-4.921	.101	-4.82	.716	-2.519	.599	-1.92	.325
HUN	-4.455	.434	-4.021	.496	-2.899	.857	-2.041	.31
FIN	-4.39	.749	-3.641	.536	-3.526	.702	-2.824	.38
ESP	-3.991	1.267	-2.724	.439	-2.6	.847	-1.754	.323
DNK	-3.873	.544	-3.329	.539	-2.118	.156	-1.963	.423
AUT	-3.871	.669	-3.202	.57	-2.646	.31	-2.336	.289
SWE	-3.626	.724	-2.901	.386	-2.603	.664	-1.939	.328
ITA	-3.607	.491	-3.116	.361	-2.554	.036	-2.517	.257
FRA	-3.579	.449	-3.13	.284	-2.601	.123	-2.477	.278
GBR	-2.442	.486	-1.955	.221	-1.381	.108	-1.273	.297
IRL	**-2.129**	**.381**	**-1.747**	**.429**	**-1.661**	**-.145**	**-1.805**	**.368**
DEU	**-2.125**	**.142**	**-1.982**	**.335**	**-.875**	**-0.009**	**-.884**	**.25**
NLD	**-1.696**	**.424**	**-1.27**	**.429**	**.277**	**.6**	**.337**	**.607**

Note 1: The present table shows the beta coefficients of $Border_{Country}$ and $Border_{Country\times EU_{trade}}$ for the years 1999 and 2014, together with the calculated standard error.

Note 2: These coefficients correspond to the ones plotted in Figure 1 from subsection 5.

Note 3: β_{1999} corresponds to the border country for the year 1999 removing the effect of trading in the single market, $\beta_{1999\times EU_{trade}}$ is the quantification of the interaction term and $\beta_{1999}+\beta_{1999\times EU_{trade}}$ corresponds to the border country for the year 1999 taking into account the effect of trading in the single market. As we are including the standard error of a combination of two coefficients, its calculations takes the following form (taking into account the notation presented in this table): $SE_{\beta_{1999}+\beta_{1999\times EU_{trade}}} = \sqrt{SE(\beta_{1999})^2+SE(\beta_{1999\times EU_{trade}})^2+COV(\beta_{1999}\beta_{1999\times EU_{trade}})}$

Note 4: We have used a reduced notation for the sake of simplicity. β_{1999} in reality corresponds to $\beta_{border_{country}}$ for the cross-section of year 1999 and $\beta_{1999\times EU_{trade}}$ corresponds to $\beta_{border_{country}}+\beta_{border_{country}\times EU_{trade}}$ for the cross-section of year 1999.

APPENDIX C (Industry estimation table)

Table 10a: Coefficients and Standard Errors of Industry Border Effects; Intra-EU border scenario

NACE	Name	$border_{NoEU_{2014}}$	$border_{EU_{2014}}$	$border_{EU_{2014} \times EU_{trade}}$	$border_{EU} + border_{EU \times EU_{trade}}$	$SE_{interaction}$	Obs
301	Building and repairing of ships	-5.77	-5.3	.41	-4.89	.58	1638
236	Articles of concrete, cement and plaster	-5.73	-5.52	.66	-4.86	.49	1625
254	Weapons and ammunition	-4.97	-5.01	.41	-4.59	.37	1565
237	**Cutting shaping & finishing of stone**	**-3.87**	**-3.64**	**-.89**	**-4.53**	**.68**	**1622**
251	Structural metal products	-5.14	-5.96	1.67	-4.28	.48	1671
235	Cement, lime and plaster	-6.1	-4.3	.26	-4.05	.54	1482
109	Prepared animal feeds	-5.78	-5.08	1.3	-3.78	.32	1627
101	Processing/preserving of meat	-5.3	-6.54	2.87	-3.67	.21	1645
105	Dairy products	-6.34	-6.42	3.04	-3.38	.21	1642
162	Products of wood, cork, straw and plaiting materials	-4.68	-4.05	.7	-3.35	.34	1672
107	Bakery and farinaceous	-4.54	-4.53	1.21	-3.32	.33	1658
253	Steam generators	-4.72	-4.18	.88	-3.3	.38	1589
303	Aircraft and spacecraft	-1.17	-3.17	.1	-3.07	.92	1654
233	Clay building materials	-4.51	-3.31	.36	-2.95	.46	1585
120	Tobacco products	-6.06	-4.53	1.64	-2.9	.54	1588
192	**Refined petroleum products**	**-3.57**	**-2.5**	**-.39**	**-2.89**	**.42**	**1648**
161	Sawmmilling and planning of wood	-2.25	-3.57	-.39	-2.76	.48	1655
172	Articles of paper and paperboard	-3.71	-3.6	.94	-2.66	.38	1673
239	**Other non-metallic**	**-4.13**	**-2.52**	**-.06**	**-2.57**	**.26**	**1663**
279	Other electrical equipment	-1.43	-3.09	.58	-2.51	.46	1650
106	Grain mill products, starches and starch products	-4.16	-3.84	1.44	-2.4	.34	1655
103	Processing/preserving of fruit & vegetables	-3.56	-3.43	1.04	-2.39	.27	1665
310	Furniture	-3.47	-2.95	.65	-2.29	.37	1674
108	Other food products	-3.64	-3.51	1.23	-2.28	.27	1668
234	Porcelain and ceramic products	-3.41	-2.91	.67	-2.24	.51	1654
104	Vegetable and animal oils and fats	-4.02	-2.86	.68	-2.18	.53	1651
274	**Electric lighting equipment**	**-2.51**	**-1.88**	**-.24**	**-2.13**	**.32**	**1664**
232	Refractory products	-4.24	-2.18	.1	-2.08	.46	1612
222	Plastic products	-2.56	-2.74	.71	-2.03	.25	1675
203	Paints, varnishes and coatings	-4.32	-2.96	.95	-2.01	.34	1670
231	Glass and glass products	-2.5	-2.51	.53	-1.98	.25	1672
102	Processing/preserving of fish	-2.47	-4.36	2.38	-1.97	.39	1648
302	Railway/tramway locomotives & rolling stock	-3.34	-3.1	1.14	-1.96	.39	1580
110	**Beverages**	**-3.48**	**-1.86**	**-0.08**	**-1.95**	**.27**	**1669**

Note 1: The present table shows the beta coefficients of $Border_{Industry_{NoEU}}$, $Border_{Industry_{EU}}$ and $Border_{Industry_{EU} \times EU_{trade}}$ for 2014, together with the calculated standard error for the EU coefficients. Order goes from lowest to highest based on $Border_{Industry_{EU} \times EU_{trade}}$ column. **Observations in bold indicate the coefficients where interaction terms are negative, meaning that border effect inside the EU-trade is bigger than outside-EU trade.**

Note 2: These coefficients correspond to the ones plotted in Figure 2 from subsection 6.

Table 10b: (Cont.) Coefficients and Standard Errors of Industry Border Effects; Intra-EU border scenario

NACE	Name	$border_{NoEU_{2014}}$	$border_{EU_{2014}}$	$border_{EU_{2014} \times EU_{trade}}$	$border_{EU} + border_{EU \times EU_{trade}}$	$SE_{interaction}$	Obs
257	Cutlery, tools and general hardware	-1.46	-2.46	.52	-1.94	.36	1660
252	Tanks reservoirs and containers of metal	-3.24	-2.13	.24	-1.89	.36	1636
259	Other fabricated metal products	-3.16	-2.65	.84	-1.81	.34	1668
323	Sports goods	-2.03	-2.5	.71	-1.79	.44	1661
152	Footwear	-4.09	-2.31	.6	-1.72	.44	1656
322	Musical instruments	-2.46	-2.73	1.05	-1.68	.63	1653
289	Other special-purpose machinery	-2.79	-2.41	.74	-1.67	.41	1670
292	Bodies for motor vehicles	-3.08	-2.24	.72	-1.53	.67	1650
271	Electric motors generators and transformers	-1.89	-1.78	.26	-1.52	.22	1666
281	General-purpose machinery	-1.78	-2.11	.67	-1.44	.23	1666
205	Other chemical products	-2.41	-2.03	.67	-1.35	.66	1660
283	Agricultural and forestry machinery	-2.94	-1.45	.18	-1.28	.34	1656
282	Other General-purpose machinery	-1.76	-2.12	.92	-1.2	.39	1671
324	Games and toys	-5.83	-2.46	1.3	-1.16	.32	1649
291	Motor vehicles	-3.29	-2.09	.94	-1.15	.41	1654
284	Metal forming machinery and machine tools	-2.78	-1.41	.28	-1.13	.2	1655
151	Tanning and dressing of fur and leather	-2.16	-1.51	.46	-1.06	.33	1662
273	Wiring and wiring devices	-2.48	-1.62	.56	-1.05	.52	1660
309	Transport equipment n.e.c.	-4.5	-1.87	.82	-1.04	.55	1658
293	Parts/accessories for motor vehicles	-2.91	-1.66	.62	-1.04	.51	1666
204	Soap cleaning & cosmetic preparations	-3.1	-1.48	.45	-1.03	.43	1662
221	Rubber products	-2.54	-1.85	.84	-1.01	.41	1665
141	Wearing apparel, except fur	-3.74	-1.02	.06	-.96	.65	1655
265	Measuring/testing/navigating appliances etc.	-1.79	-1.14	.36	-.78	.62	1653
241	Basic iron and steel	-2.95	-1.75	.99	-.76	.22	1664
132	Weawing of textiles	-3.23	-1.52	.82	-.7	.28	1658
139	Other textiles	-1.27	-.92	.24	-.68	.45	1671
275	Domestic appliances	-3.07	-.83	.15	-.68	.4	1658
171	Pulp, paper, and paperboard	-2.44	-1.64	1.03	-.61	.29	1660
206	Man-made fibres	-3.59	-2.02	1.44	-.59	.34	1629
211	**Basic pharmaceutical products**	**-2.62**	**.2**	**-.72**	**-.52**	**.41**	**1663**
201	Basic chemicals, fertilisers, plastics and rubber	-1.93	-1.06	.67	-.39	.4	1668
272	Batteries and Accumulators	-2.26	-.73	.49	-.25	.56	1652
202	Pesticides and other agrochemical products	-3.17	-.78	.65	-.13	.96	1622
321	Jewellery	-2.06	-.47	.34	-.13	.88	1632

Note 1: The present table shows the beta coefficients of $Border_{Industry_{NoEU}}$, $Border_{Industry_{EU}}$ and $Border_{Industry_{EU} \times EU_{trade}}$ for 2014, together with the calculated standard error for the EU coefficients. Order goes from lowest to highest based on $Border_{Industry_{EU} \times EU_{trade}}$ column. **Observations in bold indicate the coefficients where interaction terms are negative, meaning that border effect inside the EU-trade is bigger than outside-EU trade.**

Note 2: These coefficients correspond to the ones plotted in Figure 2 from subsection 6.

Table 11: Coefficients and Standard Errors of Industry Border Effects; Intra-EU border scenario

NACE	Name	$border_{EU_{2014} \times EU_{trade}}$	$SE_{border_{EU}}$	$border_{NoEU_{2014}}$	$border_{EU_{2014}}$	$border_{EU} + border_{EU \times EU_{trade}}$	Obs
254	Weapons and ammunition	.41	.52	-4.97	-5.01	-4.59	1565
251	Structural metal products	1.67	1	-5.14	-5.96	-4.28	1671
101	Processing/preserving of meat	2.87	.41	-5.3	-6.54	-3.67	1645
105	Dairy products	3.04	.72	-6.34	-6.42	-3.38	1642
303	Aircraft and spacecraft	.1	.95	-1.17	-3.17	-3.07	1654
161	Sawmmilling and planning of wood	.95	.73	-2.25	-3.71	-2.76	1655
279	Other electrical equipment	.58	.66	-1.43	-3.09	-2.51	1650
222	Plastic products	.71	.51	-2.56	-2.74	-2.03	1675
231	Glass and glass products	.53	.31	-2.5	-2.51	-1.98	1672
102	Processing/preserving of fish	2.38	.69	-2.47	-4.36	-1.97	1648
257	Cutlery, tools and general hardware	.52	.55	-1.46	-2.46	-1.94	1660
323	Sports goods	.71	.82	-2.03	-2.5	-1.79	1661
322	Musical instruments	1.05	.75	-2.46	-2.73	-1.68	1653
281	General-purpose machinery	.67	.4	-1.78	-2.11	-1.44	1666
282	Other General-purpose machinery	.92	.59	-1.76	-2.12	-1.2	1671

Note 1: The present table shows a selection of the beta coefficients of $Border_{Industry_{NoEU}}$ and $Border_{Industry_{EU} \times EU_{trade}}$ for 2014, corresponding to the industries where $|Border_{EU}| > |Border_{NoEU}|$, that is to say, the industries where the border effect with respect to the rest of the world (removing the effect of intra-EU trade) is greater for EU members than for the rest of countries. Order goes from lowest to highest based on $Border_{Industry_{EU} \times EU_{trade}}$ column.

Note 2: These coefficients correspond to the ones plotted in Figure 2 from subsection 6.

APPENDIX D (Endogenous and exogenous sources table)

Table 12a: NACE Industries Summary and Characteristics

NACE	Name	Rauch's Class	Harmonization	Reg. Intensity	Gamma
101	Processing/preserving of meat	w	1	27	.042
102	Processing/preserving of fish	r	1	27	.061
103	Processing/preserving of fruit/vegetables	r	1	21	.044
104	Vegetable and animal oils and fats	r	1	25	.049
105	Dairy products	w	1	24	.002
106	Grain mill starches & starch products	w	1	11	.019
107	Bakery and farinaceous	n	1	14	.001
108	Other food products	w	0	14	.013
109	Prepared animal feeds	r	0	20	.011
110	Beverages	n	0	19	.079
120	Tobacco products	r	0	6	.168
132	Weawing of textiles	w	0	6	.17
139	Other textiles	n	1	5	.021
141	Wearing apparel, except fur	n	1	6	.053
151	Tanning and dressing of fur and leather	r	0	7	.025
152	Footwear	n	1	5	.088
161	Sawmmilling and planning of wood	n	0	12	.039
162	Products of wood, cork and straw	w	0	3	.029
171	Pulp, paper, and paperboard	n	0	3	.047
172	Articles of paper and paperboard	r	0	5	.039
192	Refined petroleum products	r	0	9	.088
201	Basic chemicals, fertilisers, etc	r	1	7	.02
202	Pesticides and other agrochemical products	r	1	7	.031
203	Paints, varnishes and coatings	n	1	8	.007
204	Soap cleaning & cosmetic preparations	n	1	9	.004
205	Other chemical products	n	1	21	.006
206	Man-made fibres	n	0	2	.159
211	Basic pharmaceutical products	r	0	12	.023
221	Rubber products	n	1	19	.022
222	Plastic products	r	1	19	.005
231	Glass and glass products	n	1	16	.011
232	Refractory products	r	0	7	.078
233	Clay building materials	n	1	7	.16
234	Porcelain and ceramic products	n	0	7	.044
235	Cement, lime and plaster	r	1	6	.01
236	Articles of concrete, cement and plaster	r	1	6	.012
237	Cutting shaping & finishing of stone	n	0	6	.036

Note 1: The present table includes a short description of the industries as well as the characteristics analyzed in section 7.
Note 2: Abbreviations include: n, w and r in Rauch's classification column indicate whether industries are classified as differentiated products (n), homogeneous products (w) and reference price products (r), the values in Gamma column correspond to the Ellison-Glaeser index value of the NACE industry, Reg.Intensity column indicates the regulatory intensity per industry (number of EU regulations), Harmonization column indicates whether the European Commission claims that the sector has completed the harmonization process or not.

Table 12b: (Cont.) NACE Industries Summary and Characteristics

NACE	Name	Rauch's Class	Harmonized	Reg. Intensity	Gamma
239	Other non-metallic	n	0	4	.004
241	Basic iron and steel	r	0	3	.059
251	Structural metal products	n	1	13	.004
252	Tanks reservoirs and containers of metal	n	1	13	.014
253	Steam generators	n	1	46	.023
254	Weapons and ammunition	n	0	5	.08
257	Cutlery, tools and general hardware	n	0	5	.056
259	Other fabricated metal products	n	1	13	.018
265	Measuring/testing/navigating appliances	n	1	13	.039
271	Electric generators and transformers	n	1	27	.054
272	Batteries and Accumulators	r	1	27	.01
273	Wiring and wiring devices	n	1	27	.009
274	Electric lighting equipment	n	1	16	.01
275	Domestic appliances	n	1	16	.03
279	Other electrical equipment	n	1	27	.015
281	General-purpose machinery	n	1	46	.02
282	Other General-purpose machinery	n	1	46	.004
283	Agricultural and forestry machinery	n	1	46	.064
284	Metal forming machinery and machine tools	n	1	46	.071
289	Other special-purpose machinery	n	1	46	.007
291	Motor vehicles	n	1	46	.127
292	Bodies for motor vehicles	n	0	26	.008
293	Parts/accessories for motor vehicles	n	0	26	.089
301	Building and repairing of ships	n	1	6	.014
302	Railway/tramway locomotives & rolling stock	n	1	5	.123
303	Aircraft and spacecraft	n	0	3	.023
309	Transport equipment n.e.c.	n	0	26	.01
310	Furniture	n	0	16	.007
321	Jewellery	w	0	7	.095
322	Musical instruments	n	0	2	.014
323	Sports goods	n	1	6	.003
324	Games and toys	n	1	6	.011

Note 1: The present table includes a short description of the industries as well as the characteristics analyzed in section 7.
Note 2: Abbreviations include: n, w and r in Rauch's classification column indicate whether industries are classified as differentiated products (n), homogeneous products (w) and reference price products (r), the values in Gamma column correspond to the Ellison-Glaeser index value of the NACE industry, Reg.Intensity column indicates the regulatory intensity per industry (number of EU regulations), Harmonization column indicates whether the European Commission claims that the sector has completed the harmonization process or not.

APPENDIX E (Structural gravity equation derivation)

In the present Appendix I will present the derivation of the Structural Gravity Framework. Namely, I will present the derivation from the demand side assuming the Armington setting ((1969), i.e. goods are differentiated by place of origin) with CES preferences, as this is the one I present in section 3.

In this case we fix the supply of each good to Q_i and the factory-gate price for each variety to p_i, defining the value of domestic production (or nominal income of a country) as $Y_i = p_i Q_i$. On the other hand, country's aggregate expenditure is expressed in terms of the presented nominal income, $E_i = \phi_i Y_i$, where $\phi_i > 1$ indicates trade deficit while $0 < \phi_i < 1$ reflects trade surplus (trade deficit or surplus is taken as given).

Focusing on the demand side, we consider consumer preferences to be homothetic and identical across countries, given by a CES-utility function for importing country j:

$$\left\{ \sum_i \alpha_i^{\frac{1-\sigma}{\sigma}} c_{ij}^{\frac{\sigma-1}{\sigma}} \right\}^{\frac{\sigma-1}{\sigma}} \tag{D-1}$$

Where $\sigma > 1$ is the elasticity of substition among goods from different countries, $\alpha_i > 0$ is the CES preference parameter (treated as an exogenous taste parameter), and c_{ij} reflects the consumption of varieties from exporter i to importer j.

Consumers maximize the equation from above subject to the following budget constraint:

$$\sum_i p_{ij} c_{ij} = E_j \tag{D-2}$$

This constraint ensures that local expenditure in the importer country, E_j is exactly equal to the total spending on varieties from all countries, including itself. at delivered prices $p_{ij} = p_i t_{ij}$, result of marking up factory-gate prices by bilateral trade costs ($t_{ij} > 1$).

By solving this optimization problem, we obtain the expenditures on goods shipped from exporter i to importer j in the following specification:

$$X_{ij} = \left(\frac{\alpha_i p_i t_{ij}}{P_j} \right) E_j \tag{D-3}$$

where X_{ij} denotes trade flow from exporter to importer and, in this step, P_j can be interpreted as a CES consumer price index:

$$P_j = \left[\sum_i (\alpha_i p_i t_{ij})^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \quad \text{(D-4)}$$

Once reached this point we just need to impose market clearance for goods from each origin:

$$Y_i = \sum_j \left(\frac{\alpha_i p_i t_{ij}}{P_j}\right)^{1-\sigma} E_j \quad \text{(D-5)}$$

That is to say, with the previous equation we force that, at delivered prices (factory-gate prices multiplied by trade costs), the domestic production of an exporter, Y_i, should be equal to the total expenditure on the exporter's variety in all countries in the world, including also its own consumption (All in all we are just summing equation (D-3), that is to say, $\sum_j X_{ij} \forall j$).

If we divide the equation by $Y \equiv \sum_i Y_i$ and rearrange terms, we come up with the following specification:

$$(\alpha_i p_i)^{1-\sigma} = \frac{\frac{Y_i}{Y}}{\sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{E_j}{Y}} \quad \text{(D-7)}$$

It is now that we find in the denominator, the first of the structural terms presented in section 3, the inward multilateral resistance term, $\Pi_i^{1-\sigma} \equiv \sum_j \left(\frac{t_{ij}}{P_j}\right) \frac{E_j}{Y}$, which can be substituted in equation D-7 to obtain:

$$(\alpha_i p_i)^{1-\sigma} = \frac{\frac{Y_i}{Y}}{\Pi_i^{1-\sigma}} \quad \text{(D-8)}$$

If we substitute this final power transform $(\alpha_i p_i)^{1-\sigma}$ in equation D-3 and D-4, and take the definition of $\Pi_i^{1-\sigma}$ that we already have, we finally obtain the equations that compose the structural gravity system:

$$X_{ij} = \frac{Y_i E_j}{Y} \left(\frac{t_{ij}}{\Pi_i P_j}\right)^{1-\sigma} \quad \text{(D-9)}$$

$$\Pi_i^{1-\sigma} = \sum_j \left(\frac{t_{ij}}{P_j}\right)^{1-\sigma} \frac{E_j}{Y} \quad \text{(D-10)}$$

$$P_j^{1-\sigma} = \sum_i \left(\frac{t_{ij}}{\Pi_i}\right)^{1-\sigma} \frac{Y_i}{Y} \quad \text{(D-11)}$$

APPENDIX F (Considerations in the construction of the dataset)

We have found missing data in several years both for countries relevant to our study, such as Croatia or Greece, so in the preparation of the dataset, for modeling purposes, we have been forced to eliminate these countries from the final dataset.

Other countries relevant to the study have been eliminated due to the resulting lack of consistency in estimating their border effects, including Belgium, Luxembourg, Lithuania, Malta and Cyprus.

On the other hand, due to the large number of missing records and the low weight in trade volumes found in a significant portion of the dataset, we have limited the RoW countries list to: China, Japan, Norway, New Zealand, Australia, Russia, USA, Mexico, Hong Kong, Canada, India, Brazil, South Korea, Singapur, Thailand, Turkey, Taiwan and Ukraine.

Some arguments could be raised about avoiding the use of the whole world's trade information, as this would give a more robust and reliable scenario, however, when using all countries data, the estimations remained almost the same in our analysis. Therefore, we consider that using those subjects as representants of the rest of the world gives a wide enough picture of the whole international trade system.