

# Grado en Estadística

---

**Título:** Modelo de gravedad para el comercio internacional de 1990 a 2020.

**Autor:** Marta Moragues Mateo

**Director:** Raul Ramos Lobo

**Departamento:** Econometría, Estadística y Economía Aplicada

**Convocatoria:** Junio 2023



UNIVERSITAT DE  
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

Facultat de Matemàtiques i Estadística

## **Resumen**

El presente estudio se enmarca en el análisis de datos multidimensionales, dentro del ámbito de la economía internacional, poniendo el foco en los flujos comerciales bilaterales entre países y su medición. Con el principal objetivo de analizar los determinantes de estos flujos durante un periodo determinado, y ver cómo estos factores tienen mayor o menor relevancia en función del momento en el que se evalúe. Para ello, en primer lugar se ha revisado la literatura sobre el tema identificando la aproximación teórica y metodológica más adecuada al contexto analizado. Tras la recopilación de la información estadística necesaria para poder realizar un análisis detallado de los flujos comerciales entre países se ha implementado un modelo econométrico, en concreto el modelo de gravedad con datos de panel, para un conjunto de 252 países durante el periodo 1990-2020. A partir de esta base de datos se han obtenido estimaciones que sugieren diferencias a lo largo del tiempo en las elasticidades de los factores que determinan los flujos bilaterales del comercio internacional.

## **Palabras clave**

Comercio internacional, modelo de gravedad, datos de panel, Mínimos Cuadrados Ordinarios, resistencia multilateral, Poisson estimado por Pseudo Máxima Verosimilitud.

## **Clasificación AMS**

El código principal para este trabajo de fin de grado de acuerdo con la Sociedad Americana de Matemáticas (AMS) debido a los métodos estadísticos que se han usado son **62-J05 Linear Regression**, **62-J12 Generalized lineal models** y **91-B60 Trade models**.

## **Summary**

The present study is part of the analysis of multidimensional data, within the scope of the international economy, focusing on bilateral trade flows between countries and their measurement. With the main objective of analyzing the determinants of these flows during a given period and seeing how relevant these factors are, depending on the moment in which they are evaluated. To do this, firstly, the literature on the subject has been reviewed, identifying the most appropriate theoretical and methodological approach to the context analyzed. After collecting the necessary statistical information to carry out a detailed analysis of trade flows between countries, an econometric model has been implemented, specifically the gravity model with panel data, for a set of 252 countries during the period 1990-2020. From this database, estimates have been obtained that suggest differences over time in the elasticities of the factors that determine the bilateral flows of international trade.

## **Keywords**

International trade, gravity model, panel data, Ordinary Least Squares, multilateral resistance, Poisson Pseudo Maximum Likelihood.

## Índice

1. Introducción.....	2
2. Antecedentes del comercio internacional y el modelo de gravedad .....	4
2.1. Discusiones sobre la configuración teórica del modelo de gravedad .....	8
3. Metodología .....	12
3.1. Datos de panel .....	12
3.2. Descripción de la base de datos .....	13
3.3. Modelo de Poisson estimado por Pseudo Máxima Verosimilitud (PPML).....	15
4. Análisis exploratorio de los datos .....	17
5. Modelización .....	21
6. Conclusiones.....	26
7. Bibliografía .....	28
8. Anexos .....	30

# 1. Introducción

El comercio a nivel global se encuentra en constante cambio. La orientación y la estructura del comercio mundial en cada generación difiere considerablemente de las anteriores.

Revisando las teorías del comercio internacional, podemos ver cómo han ido experimentando constantes cambios a lo largo del tiempo, debido a las evoluciones que se van dando en la economía mundial. No obstante, la interconexión de sucesos e ideas ha dado lugar a nuevas hipótesis, entre ellas la percepción generalizada de la globalización comercial. Teniendo en cuenta estos antecedentes, el objetivo principal de este trabajo consiste en examinar algunas de estas hipótesis y principales tendencias en el análisis inter temporal de los flujos comerciales internacionales y evaluar su evolución.

Con el fin de comprender tanto los diversos factores involucrados en el intercambio de bienes entre países, como los cambios que se han ido generando en su importancia dentro del proceso comercial a lo largo de la historia, hemos recogido tanto datos unilaterales como bilaterales sobre diferentes países exportadores e importadores e intentaremos arrojar luz sobre los obstáculos que limitan o impulsan las relaciones comerciales internacionales.

Llevaremos a cabo nuestro objetivo a través de la especificación y estimación de modelos de gravedad, los cuales nos ofrecen un marco intuitivo para los factores que intervienen en este intercambio entre países. Este modelo fue introducido en el marco de la economía internacional por Jan Tinbergen (1962) siendo el primer economista que utilizó dicha ecuación relacionando los flujos comerciales bilaterales de manera directa con el tamaño de sus economías y de manera inversa con la distancia que los separa (de aquí, la relación el modelo gravitacional de Newton). Desde entonces, diferentes investigadores han sugerido modificaciones y extensiones de dicha ecuación.

Analizaremos cuantitativamente fenómenos económicos reales que incluiremos en los modelos para valorar la capacidad explicativa de distintas teorías sobre la influencia relativa de los distintos factores. Llevando a cabo así un análisis estructural siguiendo un enfoque probabilístico de la realidad, estimando los efectos de los determinantes en las relaciones económicas a través de la modelización matemática y técnicas basadas en los métodos inferenciales de la estadística.

En concreto, a partir de la revisión de la amplia literatura existente sobre el tema de estudio, especificaremos y estimaremos diferentes modelos de gravedad para el comercio internacional desde 1990 hasta 2020. Se pondrán en relieve discusiones y conceptos esenciales que han ido surgiendo a lo largo de la historia de las bases teóricas del modelo de gravedad presentadas por diferentes economistas. Las aportaciones que han sido fundamentales para la especificación del modelo de gravedad se atribuyen a trabajos como (Anderson J. , 1979), (Feenestra, 2002), (Anderson & van Wincoop, 2003), (Silva & Tenreyro, 2006) entre muchos otros.

La principal meta consiste en sacar conclusiones sobre la influencia e importancia de los factores que impactan en los flujos comerciales, introducidos en los modelos, e identificar si han ido evolucionando a lo largo del tiempo.

Ningún estudio basado simplemente en el modelo de gravedad tradicional del comercio internacional a través de Mínimos Cuadrados Ordinarios ha llegado a la conclusión de que la globalización a lo largo de los últimos años haya generado una dispersión en la actividad económica, mediante una disminución de los costos de transacción, como los costos de transporte. Es más, en (Brun, Guillaum, & Melo, 1999) y (Learner & Levinsohn, 1995) entre otros, obtenían el paradójico resultado de que el efecto de la distancia en los patrones comercial no disminuía con el tiempo, totalmente contrario a la impresión que se tiene con la globalización.

Con la revisión de la literatura actual se explican estos resultados debidos a la exclusión de los flujos comerciales nulos para poder llevar a cabo la modelización con MCO log-linealizados, sesgando así la estimación, además de la omisión de lo que se denomina “resistencias multilaterales del comercio”.

Por lo tanto, para conseguir la consolidación del objetivo propuesto y teniendo todo lo expuesto anteriormente en cuenta, se ha optado por utilizar un modelo de Poisson estimado a través del método de Pseudo Máxima Verosimilitud (PPML) teniendo en cuenta efectos fijos de pares de países.

El resto del trabajo se estructura de la siguiente forma, en primer lugar, se lleva a cabo una contextualización sobre el comercio internacional, la ecuación del modelo de gravedad y algunas de sus diferentes configuraciones teóricas. Después de esto, se inicia el apartado de metodología, donde se introducen los datos de panel, la base de datos que usaremos y la explicación del modelo de gravedad a partir de un modelo de Poisson estimado a través del método de Pseudo Máxima Verosimilitud. Seguidamente se lleva a cabo un análisis exploratorio de los datos, con el fin de ver cómo se relacionan y se distribuyen nuestras variables y los resultados obtenidos en la modelización con sus correspondientes interpretaciones. Por último, se presentan las conclusiones generales, la bibliografía y los anexos.

## 2. Antecedentes del comercio internacional y el modelo de gravedad

Este apartado proporciona una base general sobre el modelo de gravedad acompañada de una definición simple del comercio internacional. Se establecerán las ideas básicas que constituyen el modelo de gravedad y algunas de las diferentes configuraciones teóricas que han ido surgiendo a lo largo del tiempo. Existen varias opciones de modelado de la ecuación de gravedad, las más usadas se explican más detalladamente en el apartado 3.3. *Métodos de estimación del modelo de gravedad.*

Definimos el comercio internacional como el intercambio de bienes y servicios que se produce entre dos naciones, dando origen a salidas de mercancías de un territorio, denominadas exportaciones y a entradas de mercancías en otro territorio, es decir importaciones.

El modelo de gravedad del comercio internacional se presentó en un inicio como una analogía simple e intuitiva entre la Ley de Gravitación Universal y los factores que determinan los flujos comerciales bilaterales.

Según la ley de gravitación de Newton la fuerza con la que se atraen dos cuerpos es proporcional al producto de sus masas e inversamente proporcional a la distancia que los separa al cuadrado.

$$F_{ij} = G \frac{M_i \cdot M_j}{D_{ij}^2} \quad (1)$$

El modelo de gravedad para el comercio internacional en su forma tradicional se inspiró en la ley de gravitación y originalmente se postuló que el flujo de comercio entre dos países es proporcional a la masa económica de ambos (medida por sus respectivos PIB) e inversamente proporcional con la distancia que existe entre ambos (Tinbergen, 1962).

$$X_{ij} = A \frac{Y_i \cdot Y_j}{D_{ij}} \quad (2)$$

Donde:

$X_{ij}$  es el flujo del comercio del país exportador  $i$  hacia el país importador  $j$

$A$  es una constante del comercio

$Y_i, Y_j$  son respectivamente el PIB del país exportador  $i$  y del país importador  $j$

$D_{ij}$  es la distancia entre el país exportador  $i$  y el país importador  $j$

Desde el punto de vista empírico la explicación de los patrones del comercio que nos aporta el modelo de gravedad es indudable. Aunque en un inicio carecía de fundamentos teóricos sólidos, en el trabajo realizado en (Anderson, 1979) se establecieron las primeras bases teóricas para respaldar la evidencia empírica de este modelo. A partir de ese momento la literatura existente sobre la fundamentación teórica ha ido creciendo, centrándose en la importancia de una correcta especificación para la estimación de la ecuación para la obtención de resultados no sesgados, así como en ampliar la lista de covariables utilizadas para identificar otras variables naturales, históricas, culturales y relacionadas con las políticas que afectan al comercio bilateral.

Head (2000) señaló lo siguiente en su estudio: 'La ecuación de gravedad es bastante idónea para comprender el comercio, con tan solo el tamaño de las economías y sus distancias. Sin embargo, hay una gran parte de la variación del comercio que no se puede explicar'.

Como factores determinantes del comercio tenemos tanto obstáculos como impulsores (*push and pull factors*). Por ejemplo, barreras comerciales a las que se enfrentan los países exportadores al intentar acceder a los mercados extranjeros, que pueden ser tanto aranceles, como cuotas de importación, o regulaciones técnicas no arancelarias de tipo cultural, políticas, etc. Los países imponen estas barreras comerciales a las importaciones de otros países con el objetivo de proteger sus industrias nacionales o para promover sus intereses comerciales, influyendo así estas resistencias en el volumen de los flujos comerciales internacionales.

La formulación general actual del modelo de gravedad tiene la forma multiplicativa que se puede resumir en:

$$X_{ij} = A S_i M_j \phi_{ij} \quad (3)$$



Donde:

$X_{ij}$  es el flujo del comercio del país exportador  $i$  hacia el país importador  $j$

$A$  es una constante que no depende ni de  $i$  ni de  $j$

$S_i$  término que abarca los factores específicos de  $j$  (constituye la cantidad total de exportadores que están dispuestos a suministrar)

$M_j$  término que abarca los factores específicos de  $i$  (constituye la demanda total del importador)

$\phi_{ij}$  representa la facilidad de  $i$  para acceder al mercado  $j$

Es usual el uso de logaritmos para transformar la ecuación del modelo de gravedad a lineal. Pasando así de la ecuación (2) a la (4) y finalmente obteniendo una regresión log-log en la que ya no existen denominadores gracias a las propiedades de los logaritmos, que se muestra en la ecuación (5).

$$\ln(X_{ij}) = \ln\left(A^1 \frac{Y_i^{\beta_1} \cdot Y_j^{\beta_2}}{D_{ij}^{\beta_3}}\right) \quad (4)$$

$$\ln(X_{ij}) = \beta_0 + \beta_1 \ln(Y_i) + \beta_2 \ln(Y_j) - \beta_3 \ln(D_{ij}) + \epsilon_{ij} \quad (5)$$

---

<sup>1</sup> la constante  $A$  se convierte en  $\beta_0$

Para concluir con este apartado, presentamos los diferentes factores que se han tenido en cuenta al definir nuestro modelo de gravedad y sus efectos esperados, relacionados con sus correspondientes hipótesis planteadas.

Tabla 2: Hipótesis planteadas y resultados esperados.

<b>Hipótesis</b>	<b>Variables</b>	<b>Efecto previsto coeficiente</b>
El flujo comercial entre dos naciones está directamente relacionado con el tamaño económico de estas mismas.	<i>gdp</i>	positivo
El flujo comercial entre dos naciones está inversamente relacionado con la distancia que las separa.	<i>distw_arithmetic</i>	negativo
El flujo comercial entre dos naciones está inversamente relacionado con la diferencia lingüística de estas.	<i>comlang_ethno</i>	positivo
El flujo comercial entre dos naciones está directamente relacionado con su contigüidad.	<i>contig</i>	positivo
El flujo comercial entre dos naciones está directamente relacionado con su compatibilidad religiosa	<i>comrelig</i>	positivo
El flujo comercial entre dos naciones está directamente relacionado con su enlace colonial	<i>col45</i>	positivo
El flujo comercial entre dos naciones está directamente relacionado con la presencia de un acuerdo comercial	<i>rta</i>	positivo

## 2.1. Discusiones sobre la configuración teórica del modelo de gravedad

A lo largo de la historia del modelo de gravedad se han propuesto varias configuraciones teóricas para el modelo gravitacional del comercio internacional. En este apartado haremos un repaso por algunas de ellas, poniendo en escena las discusiones detrás de cada una.

Empezamos con la primera propuesta de Anderson (1979) a través del método de Mínimos Cuadrados Ordinarios, y comentaremos los errores que se cometen en el enfoque tradicional de este modelo, el cual produce estimadores sesgados por diferentes motivos.

El primer método de modelización para llevar a cabo la estimación del modelo de gravedad es a través del uso de Mínimos Cuadrados Ordinarios, ya que es el procedimiento habitual en los modelos de regresión lineal múltiple. Aun así, este método que permite estimar modelos de gravedad en su forma tradicional sin tener en cuenta los términos de resistencia multilateral, y como su mismo nombre indica lo hace a través de la minimización de la suma de los errores al cuadrado, asume que la relación entre la variable dependiente y las variables independientes es lineal, aplicando logaritmos a las variables dependientes y las variables independientes, y teniendo en cuenta ciertas hipótesis para el término del error.

Las hipótesis que plantea son:

- Los errores tienen media cero y no están correlacionados con las variables explicativas.
- Homocedasticidad, los errores se distribuyen según una Normal con un valor dado (fijo) varianza.
- Ninguna de las variables explicativas es una combinación lineal de otras variables explicativas.

Los coeficientes estimados en esta regresión log-log representan los cambios porcentuales en la variable dependiente asociados con un cambio unitario en las variables independientes.

En (Baldwin & Taglioni, 2006) los errores que se cometen al realizar la estimación a través de MCO los dividen en *gold*, *silver* and *bronze medal errors* y son respectivamente:

- o Omisión a lo que en (Anderson & van Wincoop, 2003) se denominan términos de resistencia multilateral o en (Head K. , 2000) *remoteness*, dichos términos están correlacionados con los costos comerciales, por consecuencia este error resulta en obtener estimaciones sesgadas.

- Promediar los flujos comerciales recíprocos y no tratarlos por separado en cada sentido, es decir, debemos tratar por separado las exportaciones de  $i$  a  $j$  y las de  $j$  a  $i$  en el momento  $t$ .
- Deflación inapropiada de los flujos, creando sesgos a través de falsas correlaciones, ya que la gravedad es una función de cuánto del PIB nominal se asigna a las importaciones nominales. Sin embargo, si tenemos en cuenta el primer error no cometeremos este otro, porque las variables artificiales o los efectos fijos de país lo tienen en cuenta.

En cuanto al posterior esfuerzo de los investigadores tras la primera base teórica presentada por Anderson (1979) al derivar una especificación adecuada cabe destacar la contribución de Anderson & van Wincoop (2003), los cuales determinaron que el control de los costes del comercio debe ser relativo, ya que no es observable. Determinaron que es esencial el ajuste de los costes relativos para especificar un modelo de gravedad adecuado, a través de la inclusión de los términos de “resistencia multilateral”. En concreto, la idea que hay detrás de este concepto es la siguiente: la tendencia del país importador  $j$  a importar del país  $i$  está determinada por el coste comercial asociado al intercambio de  $j$  hacia  $i$  en relación con su resistencia general a las importaciones (promedio ponderado de los costes comerciales) y a la resistencia media a la que se enfrentan los exportadores en el país  $y$ , no simplemente por los costes comerciales absolutos entre el país  $i$  y  $j$  (Anderson & van Wincoop, 2003).

La literatura existente propone tres diferentes maneras de abordar este problema:

- Mínimos Cuadrados Iterados Estructuralmente proporcionado por Anderson & van Wincoop (2003): busca estimar los coeficientes del modelo de gravedad al tener en cuenta las estructuras económicas y geográficas de los países. La idea principal es que los tamaños económicos y las distancias geográficas pueden estar correlacionados con otros factores no observados que afectan el comercio, como las políticas comerciales, las barreras no arancelarias, la cultura, entre otros. Y se aborda esta correlación introduciendo variables instrumentales que capturan la influencia de estos factores no observados, denominadas resistencias multilaterales.
- Head (2003) presenta el uso de una variable *proxy*<sup>2</sup> llamada *remoteness* tanto para importadores como exportadores, definida de la siguiente manera:

---

<sup>2</sup> Variable que representa una medida sin interés por sí sola, pero que permite la obtención de otras de mayor interés, debido a la correlación existente con el valor inferido.

$$\ln(R\_EXP_{it}) = \ln\left(\frac{\sum_j D_{ij}}{\frac{E_{jt}}{Y_t}}\right) \quad (6)$$

$$\ln(R\_IMP_{jt}) = \ln\left(\frac{\sum_i D_{ij}}{\frac{Y_{jt}}{Y_t}}\right) \quad (7)$$

Estas variables miden la distancia ponderada promedio de un país con sus socios comerciales, siendo las ponderaciones las participaciones de los socios en el comercio internacional ( $Y_t$ ).

- Controlar por efectos fijos de exportador e importador propuesto por Rose and van Wincoop (2001) y Fenestra (2002): simplemente se trata de una estimación a través de MCO teniendo en cuenta efectos fijos para los países de origen y destino, el objetivo es capturar la relación entre las variables exógenas y la endógena, controlando y eliminando los efectos individuales. Teniendo en cuenta que debido al uso de estos efectos las variables unilaterales como el PIB no se pueden incluir como variables independientes en la estimación.

Las técnicas de modelización presentadas hasta el momento no han tenido en cuenta la heterocedasticidad de los parámetros de los modelos. Este hecho es fundamental para la obtención de estimaciones no sesgadas de la verdadera elasticidad. El problema se otorga principalmente a la log-linealización, ya que el valor esperado del logaritmo de una variable aleatoria depende de los momentos de orden superior de su distribución, por lo tanto, siendo con la presencia de heterocedasticidad de los errores su transformación lineal también la tendrá.

En (Silva & Tenreyro, 2006) se demuestra como en los resultados de las estimaciones de MCO y MCO controlando por efectos fijos, la presencia de heterocedasticidad genera estimaciones notablemente diferentes, comparando las ecuaciones de gravedad log-linealizadas con la estimación en niveles Pseudo verosimilitud de Poisson, concluyendo que los coeficientes resultantes de una estimación a través de regresiones logarítmicas linealizadas son sesgados pudiendo producir conclusiones erróneas. En cambio, la utilización de un modelo de Poisson estimado por Pseudo Máxima Verosimilitud (PPML) es robusto frente patrones de heterocedasticidad.

Otro de los retos con el cual se encuentran los investigadores en el momento de derivar modelos de gravedad en su forma lineal con el uso de logaritmos es el tratamiento de los numerosos valores faltantes o ceros de los flujos comerciales bilaterales.

Existen diferentes métodos para solventar esta casuística:

- Truncar la muestra eliminando los valores faltantes o ceros.
- Asignar un valor pequeño (1, por ejemplo) a los valores faltantes o ceros, asegurando que el logaritmo está definido.
- Usar un modelo de Poisson estimado a través del método de Pseudo Máxima Verosimilitud para resolver el problema utilizando así una estimación del modelo en su forma original multiplicativa en lugar de logarítmica, además de corregir los problemas de heterocedasticidad (Silva & Tenreiro, 2006).

Cada una de las posibles soluciones presenta algún pequeño inconveniente, en el primer caso perdemos información relevante sobre los países que no comercian entre ellos, en segundo lugar, añadir pequeños números positivos aleatoriamente provoca grandes variaciones cuando se realiza una pequeña variación en el número seleccionado, y por último la PPML elimina la necesidad del uso de logaritmos naturales, pero tiende a aportar más peso a valores altos.

Después de la revisión del estado del arte que se ha realizado se ha creído conveniente el uso del modelo de Poisson debido a sus propiedades deseables en el modelo de gravedad. Se explica más detalladamente en el apartado 3.3. *Modelo de Poisson estimado por Pseudo Máxima Verosimilitud*.

### 3. Metodología

En esta sección introduciremos la tipología de los datos con los que vamos a tratar, los cuales son datos de panel, además de una descripción general de la base de datos utilizada y sus fuentes. Finalizando con las limitaciones y discusiones de la teoría del modelo de gravedad tradicional, a continuación de los métodos de modelaje que se van a realizar en el presente estudio.

#### 3.1. Datos de panel

Cuando hablamos de datos de panel, también conocidos como datos longitudinales, nos referimos a la presencia de observaciones repetidas a lo largo del tiempo para una determinada muestra. En la cual se tiene una variable  $y_{ijt}$  donde  $i = 1 \dots N$  y  $j = 1 \dots N$  son los individuos, en nuestro caso cada individuo  $ij$  es el par de países donde  $i$  representa el país importador y  $j$  el país exportador, observados a lo largo de  $t = 1 \dots T$  periodos de tiempo.

Adicionalmente, los datos longitudinales se clasifican dependiendo de si son balanceados o no. Cuando cada individuo  $ij$  se contemple en el total de los  $T$  periodos, se trata de datos balanceados, en caso contrario estaríamos hablando de datos no balanceados.

En el presente estudio trabajaremos con datos longitudinales balanceados, siendo nuestros individuos los pares de países, como se ha especificado anteriormente. De los cuales seleccionamos diferentes variables de interés para el posterior análisis que han sido recogidas en intervalos de tiempo regulares, concretamente en años, y ordenadas en orden cronológico.

A continuación, una ejemplificación de lo explicado hasta el momento con las 15 primeras observaciones de nuestra base de datos:

Tabla 3.1: Vista previa de las primeras 15 observaciones de la base de datos usada.

year	country_id_o	country_id_d	rta_coverage	distw_harmonic	contig	comlang_ethno	col45
1990	ESP	ALB	NA	1793	0	0	0
1990	ESP	CHN	NA	9550	0	0	0
1990	ESP	NPL	NA	7832	0	0	0
1990	ESP	PRK	NA	9740	0	0	0
1990	ESP	SOM	NA	6203	0	0	0
1990	ALB	ESP	NA	1793	0	0	0
1990	CHN	ESP	NA	9550	0	0	0
1990	MNG	ESP	NA	7991	0	0	0
1990	NPL	ESP	NA	7832	0	0	0
1990	PRK	ESP	NA	9740	0	0	0
1990	SOM	ESP	NA	6203	0	0	0
1990	ESP	YEM.2	NA	5317	0	0	0
1990	YEM.2	ESP	NA	5317	0	0	0
1990	ESP	IRN	NA	4668	0	0	0

Una de las principales ventajas de trabajar con datos de panel es que, a diferencia de otros, la dimensión temporal de estos enriquece la información de la base de datos y nos permite controlar diferencias inobservables en un corte transversal, es decir que varían a lo largo del tiempo, pero no entre los individuos (efectos aleatorios) y las diferencias entre los individuos, pero no a lo largo del tiempo (efectos fijos).

Debido a las características de este tipo de datos, existen técnicas específicas de regresión que permiten capturar la heterogeneidad entre distintos individuos además de su evolución en los diferentes periodos de tiempo. Estas técnicas de regresión son modelos de efectos fijos y modelos de efectos aleatorios.

### **3.2. Descripción de la base de datos**

Para llevar a cabo nuestro estudio vamos a trabajar con la base de datos *Gravity* de la web de CEPII (CEPII Database Gravity, 2022), la cual contiene una colección de variables útiles para estimar ecuaciones de gravedad. Cada observación de la base de datos representa una combinación de un país exportador, un país importador y un año, es decir trataremos con datos longitudinales, y proporciona información sobre flujos comerciales, variables geográficas, culturales, de facilitación del comercio y macroeconómicas.

La base abarca desde 1948 hasta 2020 e incluye 252 países, algunos de los cuales sólo existen durante un período de tiempo determinado, teniendo en cuenta todos los cambios que han ido experimentando los países a lo largo del tiempo.

Es "cuadrada", es decir, cada par de países aparece cada año, incluso cuando uno de los países no existe en ese año concreto. Cuando esto sucede, las variables se establecen como faltantes y se dispone de variables ficticias para identificar fácilmente estos casos.

El identificador de la base es la variable "country\_id", se refiere al país y combina el código alfabético ISO3 del país con un número que identifica posibles transformaciones territoriales.

Las variables que encontramos pueden ser tanto unilaterales, como el PIB o la población, como bilaterales, distancias o flujos comerciales. Para las variables unilaterales, se utiliza un sufijo "\_o" cuando se refiere al país de origen y "\_d" cuando se refiere al país de destino.

Esta base de datos se ha creado a partir de muchas fuentes diferentes, tanto institucionales como de una variedad de investigadores. Cuenta con un total de 4699296 observaciones y 87 variables.

El período que abordaremos en nuestra investigación va del año 1990 al 2020. La razón por la cual hemos seleccionado este marco temporal es debido a que, a partir de 1990, con motivo del fin de los regímenes socialistas el número de países existentes presenta un gran cambio



por la recomposición geográfica de bastantes países. Y cerramos nuestro marco temporal en 2020, debido a la pandemia sufrida mundialmente del Covid-19, la cual provocó que el comercio se viera afectado por sus consecuencias.

Nuestra base de datos final para la estimación de nuestro modelo de gravedad constará finalmente de los registros de los flujos comerciales de 252 países desde 1990 a 2020. La lista de países se encuentra en la Tabla 8.1 del anexo.

Podemos diferenciar las variables seleccionadas para nuestro estudio en cinco grandes grupos: geográficas, culturales, indicadores macroeconómicos, de facilitación del comercio y de flujo comercial. Las variables escogidas se muestran en la Tabla 3.2.

Tabla 3.2: Variables seleccionadas de la base de datos de (CEPII Database Gravity, 2022).

<b>Categoría</b>	<b>Variable</b>	<b>Descripción</b>	<b>Nivel</b>	<b>Fuente</b>
<b>Identificadores</b>	<i>country_id</i>	Identificador del país	Unilateral	CEPII
	<i>year</i>	Año	Bilateral	
<b>Geográficas</b>	<i>distw_arithmetic</i>	Media armónica <sup>3</sup> de la distancia ponderada por la población media de las ciudades más pobladas de cada país, en kilómetros	Bilateral	CEPII's GeoDist
	<i>contig</i>	1 si los países son contiguos (vecinos)	Bilateral	
<b>Culturales</b>	<i>comlang_ethno</i>	1 si los países comparten un idioma común hablado por al menos el 9% de la población	Bilateral	CEPII's GeoDist
	<i>col45</i>	1 si los países tuvieron o tienen una relación colonial posterior al 1945	Bilateral	
	<i>comrelig</i>	Índice de proximidad religiosa <sup>4</sup>	Bilateral	LaPorta et al. (1999)

<sup>3</sup> Según (Head, Keith, & Mayer, 2010) la media armónica es una medida más apropiada que la aritmética, ya que las estimaciones empíricas de la elasticidad de la distancia de los flujos comerciales están inequívocamente más cerca de -1 que de 1.

<sup>4</sup> Índice (Anne-Célia Disdier, 2006): obtenido sumando los productos de las participaciones de católicos, protestantes y musulmanes en los países de origen y de destino. Varía de 0 a 1, aumenta cuando los países comparten una misma religión practicada por gran parte de la población.

<b>Indicadores macroeconómicos</b>	<i>pop</i>	Población (en miles)	Unilateral	World Bank's Development Indicators
	<i>gdp</i>	PIB (en miles de \$)	Unilateral	
<b>Facilitación del comercio</b>	<i>rta</i>	1 si los países tienen un acuerdo comercial <sup>5</sup>	Bilateral	WTO's Regional Trade Agreements Information System
<b>Flujo comercial</b>	<i>tradeflow_imf</i>	Flujo comercial (en miles de \$)	Unilateral	IMF's Direction of Trade Statistics

### 3.3. Modelo de Poisson estimado por Pseudo Máxima Verosimilitud (PPML)

En este apartado desarrollaremos y explicaremos el método de estimación que se ha usado en el presente estudio.

Como se ha introducido anteriormente, el modelo de Poisson estimado por Pseudo Máxima Verosimilitud es un modelo lineal generalizado, el cual usa la distribución de probabilidad exponencial lineal de cuasi-poisson para la variable respuesta y la función de enlace<sup>6</sup> log.

La modelización a través del modelo de Poisson estimado por Pseudo Máxima Verosimilitud propuesto por Silva y Tenreyro (2006) en la búsqueda de ofrecer una estimación del modelo de gravedad en su forma multiplicativa, con la inclusión de flujos nulos y teniendo en cuenta la heterocedasticidad, enuncia que bajo la sencilla suposición de que el modelo de gravedad contiene las variables explicativas adecuadas, este procedimiento permite obtener estimaciones consistentes del modelo no lineal original, es decir en su forma multiplicativa.

El modelo presenta propiedades deseables para la estimación de modelos de gravedad para el comercio internacional como:

- Consistente con la presencia de efectos fijos ingresados como variables ficticias.
- Soluciona la presencia de heterocedasticidad en los modelos de flujos comerciales bilaterales. Consiguiendo así estimaciones consistentes y errores eficientes.

<sup>5</sup> Acuerdos de Integración Regional (*Regional Trade Agreements*).

<sup>6</sup> Función del valor esperado de  $Y$ ,  $E(Y)$ , como una combinación lineal de las variables predictoras.

- No requiere la eliminación de los valores faltantes, evitando así el sesgo en la selección de la muestra.
- Estima el modelo de gravedad en su forma multiplicativa, otorgando a todas las observaciones la misma importancia.
- Admite tanto variables explicativas continuas como dicotómicas, y aunque la variable endógena esté especificada en niveles de flujo comercial y no en logaritmos, los coeficientes de las variables exógenas continuas, aunque estén con logaritmos, se interpretan como elasticidades simples. En cambio, las dicotómicas se interpretan en términos de porcentaje.
- No es necesario que los datos se distribuyan según una Poisson, ya que es un estimador de probabilidad pseudo máxima. La única premisa necesaria para derivar un estimador consistente es que la variable exógena esté especificada de manera correcta. Para la comprobación de este supuesto se puede utilizar la prueba de Ramsey.

La especificación del modelo de gravedad en su forma multiplicativa es la siguiente:

$$X_{ij} = \exp [\beta_0 + \beta_1 \ln(Y_i) + \beta_2 \ln(Y_j) - \beta_3 \ln(D_{ij})] \quad (8)$$

La función *ppml* del paquete de *Gravity* de R estima modelos de gravedad en su forma multiplicativa, es un método de estimación para modelos de gravedad que pertenece a los modelos lineales generalizados y se estima a través de *glm* utilizando la distribución cuasi-poisson y un enlace *log-link*.

## 4. Análisis exploratorio de los datos

En este apartado se ha llevado a cabo una exploración inicial de los datos que vamos a tratar, antes de dar paso a los resultados de nuestra estimación del modelo de gravedad, realizando así un primer paso para examinar la lógica indudable detrás de dicha ecuación. Se presentan entonces las relaciones y los efectos de los determinantes del comercio internacional que el modelo de gravedad tiene en cuenta a través de diversos gráficos, sentando así las bases para las futuras interpretaciones y conclusiones.

Primeramente, se han calculado las correlaciones entre las variables del flujo comercial, la distancia y el PIB con la función  $cor()$ <sup>7</sup> de R. Las correlaciones de interés se encuentran fuera de la diagonal de las matrices presentadas a continuación, a parte se han calculado las correlaciones diferenciando por los diferentes periodos de estudio.

Tabla 4.1: Matriz de correlaciones variables numéricas 1990-1999.

<b>Periodo</b> <b>1990 - 1999</b>	<b>trade_flow_imf_o</b>	<b>gdp_o</b>	<b>gdp_d</b>	<b>distw_harmonic</b>
<b>trade_flow_imf_o</b>	1	0,4983	0,3751	-0,1929
<b>gdp_o</b>	0,4983	1	-0,2383	0,1466
<b>gdp_d</b>	0,3751	-0,2383	1	0,0625
<b>distw_harmonic</b>	-0,1929	0,1466	0,0625	1

Tabla 4.2: Matriz de correlaciones variables numéricas 2000-2009.

<b>Periodo</b> <b>2000 - 2009</b>	<b>trade_flow_imf_o</b>	<b>gdp_o</b>	<b>gdp_d</b>	<b>distw_harmonic</b>
<b>trade_flow_imf_o</b>	1	0,5577	0,3539	-0,2429
<b>gdp_o</b>	0,5577	1	-0,1711	0,0744
<b>gdp_d</b>	0,3539	-0,1711	1	0,0370
<b>distw_harmonic</b>	-0,2429	0,0744	0,0370	1

Tabla 4.3: Matriz de correlaciones variables numéricas 2010-2019.

<b>Periodo</b> <b>2010 - 2019</b>	<b>trade_flow_imf_o</b>	<b>gdp_o</b>	<b>gdp_d</b>	<b>distw_harmonic</b>
<b>trade_flow_imf_o</b>	1	0.5706	0.3514	-0.2466
<b>gdp_o</b>	0.5706	1	-0.1556	0.0573
<b>gdp_d</b>	0.3514	-0.1556	1	0.0360
<b>distw_harmonic</b>	-0.2466	0.0573	0.0360	1

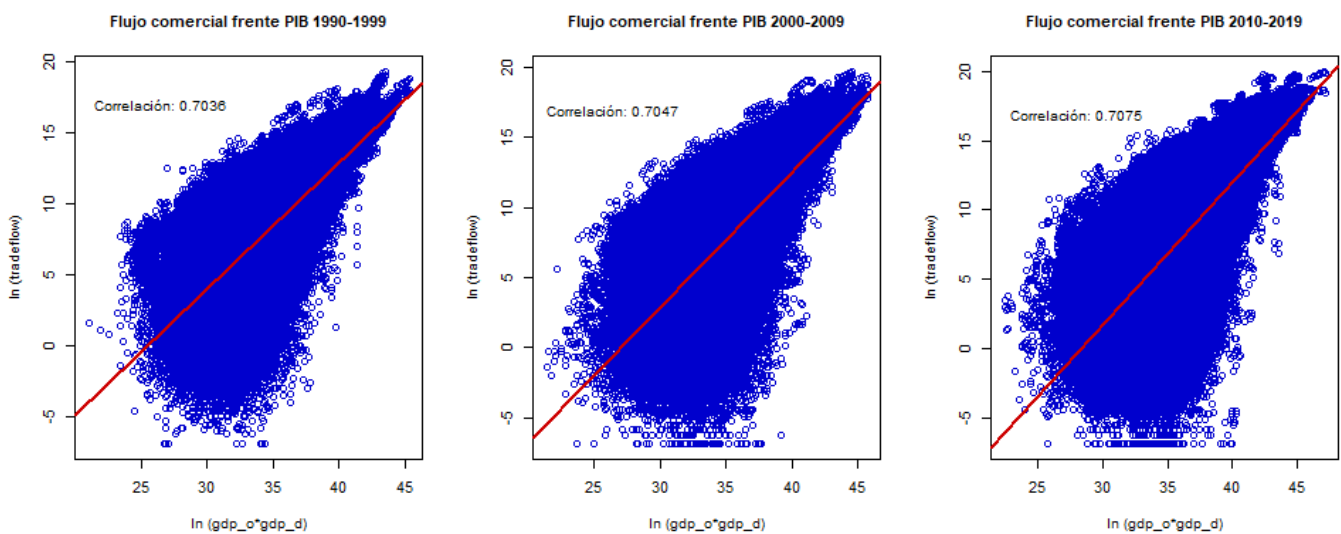
<sup>7</sup> Calcula el coeficiente de correlación lineal de Pearson sin tener en cuenta valores faltantes

Tras la examinación de los coeficientes de correlación obtenidos para nuestras variables numéricas en los diferentes intervalos de años resaltaremos dos hechos. Uno de ellos es que se confirma que el flujo comercial y la distancia entre ambos países están correlacionados negativamente, y la correlación positiva entre el flujo comercial y el PIB, tanto del país de origen como del de destino, siendo más fuerte la correlación con el PIB del país de origen, que ha presentado en los tres periodos un valor alrededor del 0'5, en cambio el PIB del país de destino alrededor del 0'3. En segundo lugar se observa un aumento en el valor del coeficiente de correlación del flujo comercial y el PIB del país de origen a través de los años, y una disminución en el caso del flujo comercial y el PIB con el país de destino y con la distancia

Estos hallazgos están en línea con la idea básica que presenta el modelo de gravedad tradicional. Los países que están más alejados tienden a comerciar menos y los países económicamente grandes tienden a comerciar más.

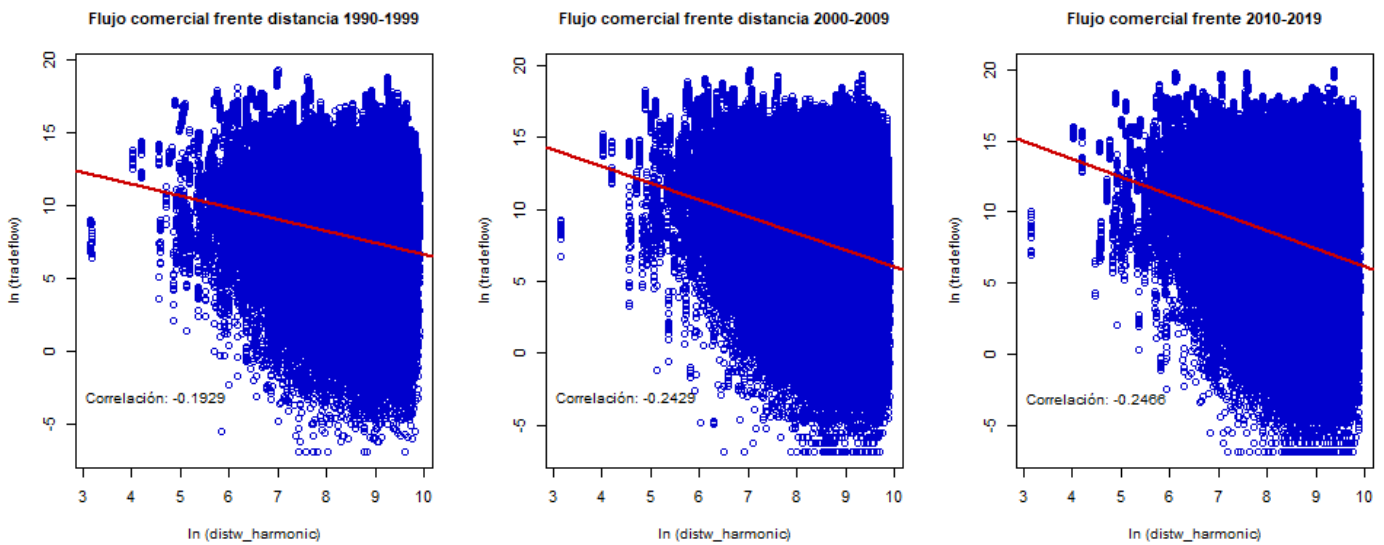
Después del cálculo de los coeficientes de la correlación de Pearson se ha usado la función *plot()* de R para presentar la misma información gráficamente, a través de diagramas de dispersión, los cuales representan visualmente las relaciones entre dos variables cuantitativas.

Figura 4.1: Diagramas de dispersión flujo comercial frente PIB para los tres periodos de estudio.



La figura 4.1 muestra tres diagramas de dispersión para el flujo comercial correspondiente a un par de países frente el tamaño económico (PIB) de ambos, cada uno de los diagramas corresponde a uno de los periodos de estudio. Estos diagramas muestran claramente la asociación positiva existente entre estas variables, de acuerdo con el análisis de correlación. Superpuesta a la nube de puntos encontramos la línea de mejor ajuste a los datos, la cual tiene una pendiente positiva considerable.

Figura 4.2: Diagramas de dispersión flujo comercial frente distancia para los tres periodos de estudio.

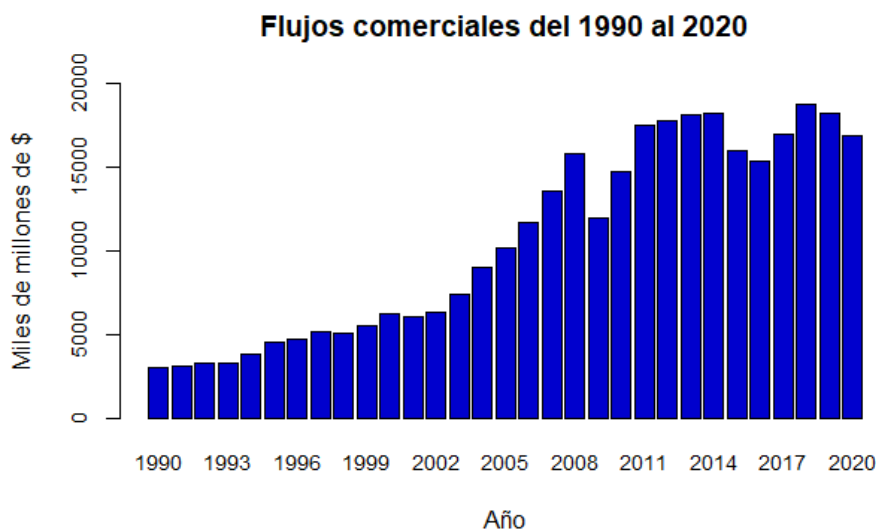


La figura 4.2 muestra tres diagramas de dispersión para el flujo comercial correspondiente a un par de países frente la distancia entre ambos, cada uno de los diagramas corresponde a uno de los periodos de estudio. Estos diagramas sugieren una asociación negativa existente entre estas variables, de acuerdo con el análisis de correlación. Superpuesta a la nube de puntos encontramos la línea de mejor ajuste a los datos, la cual tiene una pendiente negativa.

Acabando con esta primera parte del análisis exploratorio de los datos podemos concluir que las evidencias gráficas también respaldan las bases del modelo de gravedad tradicional juntamente con los coeficientes de correlación calculados.

Por lo que se refiere al volumen del comercio internacional a lo largo de los años que vamos a tener en cuenta, se presenta el siguiente gráfico de barras con dichos volúmenes.

Figura 4.3: Diagramas de barras flujo comercial por años.



De la figura 4.3 se pueden hacer las siguientes observaciones:

- Observando los datos en orden cronológico, se puede observar que el volumen total de flujos bilaterales internacionales ha experimentado un aumento general a lo largo de los años, con algunas fluctuaciones y variaciones en el ritmo de crecimiento, pero se ve una clara tendencia general creciente
- Durante la década de 1990, el volumen de comercio internacional muestra un crecimiento constante, pasando de alrededor de 2998.003 en el año 1990 a 6267.284 en el 2000. Este período de crecimiento constante puede estar asociado con la apertura de mercados y la globalización económica.
- A partir de principios de la década del 2000, se observa la continuidad del crecimiento constante de la anterior década, pero con una pendiente creciente aún más acentuada, y acompañado de cierta variabilidad y fluctuaciones en el volumen de comercio internacional, como se ve en el año 2008. Estos cambios pueden estar relacionados con eventos económicos globales, como la crisis financiera de 2008 y la recesión resultante. Después de la recesión de 2008 se muestra una notable recuperación económica.
- En los últimos 10 años, se observa cierta estabilización y fluctuación en el volumen de comercio internacional, cabe mencionar que la disminución del volumen comercial internacional en 2020 se debe a la pandemia de COVID-19.

Después de lo evaluado visualmente en el Gráfico 4.3, dividiremos nuestros modelos en tres periodos diferentes, estos serán: 1990 a 1999, 2000 a 2009 y 2010 a 2019. Cada uno de ellos cuenta con un total de 10 años, y se ha considerado adecuado no incluir el año 2020, debido al impacto de la pandemia de la COVID-19 y que no puede ser capturado en su totalidad hasta poder disponer de una muestra que incluya información estadística más reciente.

A continuación, se presentan los resultados obtenidos a través del estimador PPML teniendo en cuenta los tres periodos mencionados anteriormente, con el fin de ver si existen diferencias en las elasticidades de los factores que determinan el comercio internacional dependiendo del periodo.

## 5. Modelización

Se presentan las diferentes tablas con los resultados de la modelización de la ecuación de gravedad para el comercio internacional a través de PPML en tres periodos diferentes, del 1990 al 1999, del 2000 al 2009 y del 2010 al 2019.

Se ha realizado además la modelización a través de MCO y MCO controlando por efectos fijos, y se comprueba lo enunciado en (Silva & Tenreyro, 2006) sobre las diferencias en las elasticidades comparando con el modelo PPML, tienden a sobredimensionar el efecto sobre los flujos de comercio, los resultados de estos modelos se encuentran en el anexo.

En cada tabla se especifican las variables predictoras del modelo con sus correspondientes coeficientes, errores estándar, y el valor del estadístico T y el p valor asociado a esta prueba de significación de los coeficientes.

Es importante tener en cuenta que los coeficientes asociados a las variables *dummy* (como idioma común, colonizador y frontera comunes no pueden ser interpretados de la misma manera que los coeficientes de las variables en logaritmos. Estos coeficientes necesitan ser transformados para poder interpretarse como elasticidades<sup>8</sup>.

Por otra parte, la interpretación de los coeficientes en el modelo PPML teniendo en cuenta que a la variable dependiente y a las *dummy* no se les aplica el logaritmo y que a las variables independientes continuas si, los efectos marginales de cada parámetro se interpretan como:

- Variables continuas: una variación de un 1% en dicha variable supone un cambio de  $\beta_i/100$  unidades en  $Y$ .
- Variables *dummy*: en caso de que esa variable se de (es decir cuando toma el valor 1, ya que estas variables únicamente presentan los valores 0 y 1) supone un cambio de  $(e^{\beta_i} - 1)$  unidades en  $Y$ .

---

<sup>8</sup> Transformación  $(e^{\beta_i} - 1)$



Tabla 5.1: Resultados modelo PPML 1990 - 1999

**Estimación del modelo de Poisson por Pseudo Máxima Verosimilitud (1990 – 1999)**

<i>Predictors</i>	<b>Dependent variable</b>			
	<i>Incidence Rate Ratios</i>	<i>std. Error</i>	<i>T value</i>	<i>p</i>
(Intercept)	12.36046	0.21558	57.337	< 2e-16 ***
dist log	-0.20035	0.02002	-10.006	<2e-16***
contig	1.92526	0.05963	32.289	< 2e-16 ***
comlangethno	0.35918	0.04501	7.980	1.48e-15 ***
col45	0.26339	0.10461	2.518	0.0118 *
comrelig	1.07170	0.07252	-16.433	<2e-16***
rta	1.77258	0.04787	37.028	< 2e-16 ***
Observations	252160			

El modelo muestra que todas las variables son estadísticamente significativas, ya que todas presentan un p-valor inferior a 0.05

Una variación de un 1% en la variable referente a la distancia supone un cambio de 0,0020035 unidades menos en el flujo comercial, es decir de -2.003,5\$.

El hecho de que los países sean contiguos supone un incremento de 5,8569 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países compartan una lengua común hablada al menos por un 9% de la población supone un incremento de 0,432154 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países tengan en el presente o hayan tenido en el pasado una relación colonial supone un incremento de 0,301334 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países presenten afinidad religiosa supone un incremento de 1,9203398 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países tengan algún tipo de acuerdo comercial supone un incremento de 5,886019 miles de \$ en el flujo comercial entre ambos.

Tabla 5.2: Resultados modelo PPML 2000 - 2009

**Estimación del modelo de Poisson por Pseudo Máxima Verosimilitud (2000 – 2009)**

<i>Predictors</i>	<b>Dependent variable</b>			
	<i>Incidence Rate Ratios</i>	<i>std. Error</i>	<i>T value</i>	<i>p</i>
(Intercept)	14.06325	0.18333	76.711	<2e-16***
dist log	-0.17548	0.02081	-8.435	<2e-16***
contig	1.72540	0.05046	34.195	<2e-16***
comlangethno	0.31671	0.03869	8.185	<2e-16***
col45	0.38135	0.09861	3.867	0.00011***
comrelig	1.07838	0.08115	-13.289	< 2e-16 ***
rta	1.27435	0.04201	30.336	<2e-16***
Observations	291810			

El modelo muestra que todas las variables son estadísticamente significativas, ya que todas presentan un p-valor inferior a 0.05

Una variación de un 1% en la variable referente a la distancia supone un cambio de 0,0017548 unidades menos en el flujo comercial, es decir de -1.754,8\$.

El hecho de que los países sean contiguos supone un incremento de 5,61476 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países compartan una lengua común hablada al menos por un 9% de la población supone un incremento de 0,372604 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países tengan en el presente o hayan tenido en el pasado una relación colonial supone un incremento de 0,464260 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países presenten afinidad religiosa supone un incremento de 1,939913 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países tengan algún tipo de acuerdo comercial supone un incremento de 3,576376 miles de \$ en el flujo comercial entre ambos.

Tabla 5.3: Resultados modelo PPML 2010 - 2019

**Estimación del modelo de Poisson por Pseudo Máxima Verosimilitud (2010 – 2019)**

<i>Predictors</i>	<b>Dependent variable</b>			
	<i>Incidence Rate Ratios</i>	<i>std. Error</i>	<i>T value</i>	<i>p</i>
(Intercept)	14.75841	0.17776	83.022	<2e-16***
dist log	-0.03745	0.02456	-1.525	<2e-16***
contig	1.86613	0.05246	35.572	<2e-16***
Comlangethno	0.39083	0.03955	9.883	<2e-16***
col45	0.15647	0.10846	1.443	0.149
comrelig	0.84014	0.06680	-12.577	<2e-16***
rta	1.04416	0.03872	26.967	<2e-16***
Observations	306866			

El modelo muestra que todas las variables son estadísticamente significativas, ya que todas presentan un p-valor inferior a 0.05. Menos la variable *col45* con un p valor de 0.149.

Una variación de un 1% en la variable referente a la distancia supone un cambio de 0,0003745 unidades menos en el flujo comercial, es decir de -374,5\$.

El hecho de que los países sean contiguos supone un incremento de 6,463235 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países compartan una lengua común hablada al menos por un 9% de la población supone un incremento de 0,478207miles de \$ en el flujo comercial entre ambos.

El hecho de que los países tengan en el presente o hayan tenido en el pasado una relación colonial supone un incremento de 0,169375 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países presenten afinidad religiosa supone un incremento de 1,316691 miles de \$ en el flujo comercial entre ambos.

El hecho de que los países tengan algún tipo de acuerdo comercial supone un incremento de 2,841011 miles de \$ en el flujo comercial entre ambos

La relación comercial entre países no puede ser completamente explicada por la distancia, ya que los costos de transacción pueden ser altos. Por ejemplo, la incapacidad de comunicarse entre países puede afectar el comercio, lo que significa que países que comparten el mismo idioma tienden a tener un mayor intercambio comercial. Para tener en cuenta este factor, se incluye una variable *dummy* para representar el idioma común.

Además, parte del efecto del idioma común puede ser atribuido al pasado colonial, ya que países que fueron colonias de otros tienden a compartir el mismo idioma. Estos lazos coloniales tienen una correlación positiva con el comercio entre estos países, por lo tanto, se incluye una variable *dummy* para representar el colonizador común.

Por último, el efecto de tener una frontera común entre países tiene un impacto positivo en el comercio, pero este efecto ha ido disminuyendo con el avance de las comunicaciones y la reducción de los costos de transporte.

Resumiendo, hemos obtenido que la variable referente a la distancia presenta el signo esperado (negativo) en los tres periodos, con una disminución en su efecto, representaría un obstáculo para el comercio menor a lo largo del tiempo. En cuanto a las estimaciones de las demás correspondientes variables *dummy* todas presentan signo positivo, siendo todas significativas, cabe mencionar que *col45* (la cual era 1 si se tenía o se había tenido una relación colonial) ha sido la que ha presentado un nivel de significación mas bajo entre todas las variables para cada periodo.

## 6. Conclusiones

El papel del analista a la hora de tratar con modelos de gravedad es en primera estancia elegir la mejor técnica para llevar a cabo su estimación, dependiendo de las limitaciones y naturaleza de los datos y de la pregunta en cuestión, para poder responder sólidamente y de una manera significativa.

Después de la revisión de las posibles versiones del modelo de gravedad viendo que la literatura sobre la gravedad ha ido cambiando en las últimas dos décadas, y comprendiendo que la ecuación de gravedad para el comercio internacional no es solo una forma intuitiva de resumir la relación entre el flujo comercial y factores determinantes si no que se necesita una técnica de modelización que presente una base sólida teórica para poder realizar interpretaciones y conclusiones significativas, se ha destacado la importancia de la inclusión de efectos fijos, o la corrección de algún otro modo de la resistencia multilateral y el uso del método de estimación a través de Pseudo verosimilitud de Poisson.

Se ha comprobado como las estimaciones de los coeficientes a través de MCO o MCO controlando por efectos fijos son claramente distintas de las obtenidas con un modelo de Poisson estimado por Pseudo Máxima Verosimilitud, debido a la presencia de heterocedasticidad, añadiendo también la diferencia con la que se tratan los valores faltantes de los flujos comerciales a través de MCO en el cual se eliminan.

Recordamos el objetivo central del presente estudio, el cual era comprender tanto los diversos factores involucrados en el intercambio de bienes entre países, como los cambios que se han ido generando en su importancia dentro del proceso comercial a lo largo de la historia con el uso del modelo de gravedad del comercio internacional.

Los modelos de gravedad han demostrado ser herramientas fundamentales para evaluar el comercio internacional en flujos bilaterales a lo largo del tiempo y analizar los factores determinantes de dichos flujos, su aplicación en esta investigación ha permitido capturar y cuantificar la importancia de diversos factores en el comercio internacional a través del tiempo. Además, los modelos de gravedad permiten analizar otros factores determinantes del comercio, aparte de la distancia o el tamaño de la economía, así como variables culturales y políticas, como el idioma común o los lazos coloniales.

En (Brun, Carrère, Guillaumont, & de Melo, 2005) se investiga el hecho de que el coeficiente estimado para la distancia aumente en lugar de disminuir a través del tiempo utilizando el modelo de gravedad tradicional. Y se asocia la obtención de estos resultados a la omisión del término de *remotness*, al incluir dicho término se elimina la paradoja del aumento de la elasticidad de la distancia. Además, en este estudio se descubre que la disminución del efecto de la distancia se limita en gran medida al comercio entre países ricos, quedando marginado este suceso del comercio entre países pobres.

Una de las principales ideas que se han desarrollado en el trabajo con la ayuda de la revisión de la literatura sobre el modelo de gravedad, es que es necesaria una adecuada especificación para obtener resultados y conclusiones no erróneas. El método de MCO resultara en estimaciones sesgadas y difieren de los valores obtenidos con PPML como hemos podido comprobar. Gracias a esto, hemos podido obtener que las elasticidades de los diferentes factores estudiados han evolucionado en el tiempo, no podríamos a ver sacado estas conclusiones si no fuese por la adecuada modelización con PPML.

Hemos obtenido resultados esperados y relevantes en cuanto a la influencia e importancia en términos de elasticidad de los factores que determinan los flujos comerciales bilaterales, e hemos cuantificado su evolución a lo largo del tiempo, diferenciando en tres periodos, del 1990 al 1999, del 2000 al 2009 y del 2010 al 2019.

Por ejemplo, se ha demostrado que la distancia geográfica entre países afecta negativamente el comercio, ya que implica mayores costos de transporte y barreras logísticas, pero con el paso del tiempo la distancia influye menos en el volumen del flujo comercial. Asimismo, se ha observado que la presencia de acuerdos comerciales, la contigüidad entre dos países, la afinidad religiosa, compartir una lengua común por un mínimo del 9% de la población o presentar un alto índice de afinidad religiosa facilita el intercambio comercial entre países, presentando los coeficientes de estos parámetros signo positivo, y observándose que en los diferentes periodos de estudio cada uno de ellos ha ido disminuyendo su efecto. Todos los efectos son similares en los tres periodos, pero sus magnitudes difieren.

Gracias a que los modelos de gravedad pueden ser aplicados en diferentes períodos de tiempo mediante el uso de datos de panel, se posibilita el análisis de la evolución del comercio bilateral a lo largo de los años. Teniendo todo esto en cuenta los modelos de gravedad son herramientas muy útiles y fundamentales para la evaluación del comercio internacional, lo cual ha sido así durante el último medio siglo y seguirá siéndolo gracias a su capacidad para incorporar una amplia gama de factores determinantes del comercio y su aplicabilidad a lo largo del tiempo que los convierten en una herramienta imprescindible para los investigadores del comercio internacional.

## 7. Bibliografía

- Anderson, J. (1979). A Theoretical Foundation for Gravity Equation. *American Economic Review*.
- Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 170-192.
- Anne-Célia Disdier, T. M. (2006). Je t'aime, moi non plus: Bilateral opinions and international trade. *European Journal of Political Economy*, 1140-1159.
- Baldwin, R., & Taglioni, D. (2006). *Gravity for dummies and dummies for gravity equations*. National Bureau of Economic Research Working Paper 12516.
- Brun, J. F., Guillaum, P., & Melo, J. d. (1999). La distan mondialisation du commerce extérieur. *In A. Bouet and J. politiques économiques: les marges*.
- Brun, J.-F., Carrère, C., Guillaumont, P., & de Melo, J. (2005). Has Distance Died? Evidence from a Panel Gravity Model. *The Wrold Bank Economic Review*, págs. 99-120.
- CEPII Database Gravity. (2022). Obtenido de [http://www.cepii.fr/CEPII/en/bdd\\_modele/bdd\\_modele\\_item.asp?id=8](http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=8)
- Estimation Methods for Gravity Models*. (2023). Obtenido de R-Project: <https://cran.r-project.org/web/packages/gravity/gravity.pdf>
- Feenestra, R. C. (2002). Border effects and the gravity equation: consistent methods for estimation. *Scottish Journal of Political Economy*, págs. 491-506.
- From Keith Head and Thierry Mayer, Gravity Equations: Workhorse, Toolkit, and Cookbook. (2014). En K. R. Elhanan Helpman, *Handbooks of International Economics*. North-Holland.
- Glenn Magerman, Z. S. (2016). Distance and Border Effects in International Trade: A Comparison of Estimation Methods. *The Open-Access, Open Assessment E-Journal*.
- Head, K. (2000). *Gravity for Beginners*.
- Head, K., T, M., & J, R. (2010). The erosion of colonial trade linkages after independence. *Journal of International Economics*, 1-14.
- Head, Keith, & Mayer, T. (2010). Illusory Border Effects : Distance Mismeasurement Inflates Estimates of home Bias in Trade. En *The Gravity Model in International Trade*.

Krugman, P. R., & Obstfeld, M. (2006). *Economía Internacional Teoría y política*. PEARSON EDUCACIÓN, S.A.

LaPorta, Lopez-de-Silanes, F., Shleifer, A., & Vishny, R. (1999). The quality of Government. *Journal of Law, Economics and Organization* 15.1, 222-227.

Lerner, E., & Levinsohn, J. (1995). International Trade: The Evidence. En *Handbook of International Economics*, vol.3. New York: Elsevier, North-Holland.

Maddalena Conte, P. C. (2022). *The CEPII Gravity Database*.

Silva, S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 641-658.

Tinbergen, J. (1962). *Shaping the World Economy*. New York: The Twentieth Century Fund.



## 8. Anexos

### 8.1. Países presentes en la base de datos

Tabla 8.1: Países que contiene la base de datos utilizada para la modelización.

Abreviatura	País	Abreviatura	País	Abreviatura	País
ABW	Aruba	CHL	Chile	GAB	Gabón
AFG	Afganistán	CHN	China	GBR	Reino Unido
AGO	Angola	CIV	Costa de Marfil	GEO	Georgia
AIA	Anguila	CMR	Camerún	GHA	Ghana
ALB	Albania	COD	República Democrática del Congo	GIB	Gibraltar
AND	Andorra	COG	Congo	GIN	Guinea
ANT	Antillas Neerlandesas	COK	Islas Cook	GLP	Guadalupe
ARE	Emiratos Árabes Unidos	COL	Colombia	GMB	Gambia
ARG	Argentina	COM	Comoras	GNB	Guinea-Bissau
ARM	Armenia	CPV	Cabo Verde	GNQ	Guinea Ecuatorial
ASM	Samoa Americana	CRI	Costa Rica	GRC	Grecia
ATG	Antigua y Barbuda	CSK	Checoslovaquia	GRD	Granada
AUS	Australia	CUB	Cuba	GRL	Groenlandia
AUT	Austria	CUW	Curazao	GTM	Guatemala
AZE	Azerbaiyán	CXR	Isla de Navidad	GUF	Guayana Francesa
BDI	Burundi	CYM	Islas Caimán	GUM	Guam
BEL	Bélgica	CYP	Chipre	GUY	Guyana
BEN	Benin	CZE	República Checa	HKG	Hong Kong
BES	Bonaire, San Eustaquio y Saba	DDR	Alemania Oriental	HND	Honduras
BFA	Burkina Faso	DEU	Alemania	HRV	Croacia
BGD	Bangladesh	DJI	Yibuti	HTI	Haití
BGR	Bulgaria	DMA	Dominica	HUN	Hungría
BHR	Baréin	DNK	Dinamarca	IDN	Indonesia
BHS	Bahamas	DOM	República Dominicana	IND	India
BIH	Bosnia y Herzegovina	DZA	Argelia	IOT	Territorio Británico del Océano Índico
BLR	Bielorrusia	ECU	Ecuador	IRL	Irlanda
BLZ	Belice	EGY	Egipto	IRN	Irán
BMU	Bermudas	ERI	Eritrea	IRQ	Iraq
BOL	Bolivia	ESH	Sáhara Occidental	ISL	Islandia
BRA	Brasil	ESP	España	ISR	Israel
BRB	Barbados	EST	Estonia	ITA	Italia
BRN	Brunéi	ETH	Etiopía	JAM	Jamaica
BTN	Bután	FIN	Finlandia	JOR	Jordania
BWA	Botsuana	FJI	Fiyi	JPN	Japón
CAF	República Centroafricana	FLK	Islas Malvinas	KAZ	Kazajistán
CAN	Canadá	FRA	Francia	KEN	Kenia
CCK	Islas Cocos	FRO	Islas Feroe	KGZ	Kirguistán
CHE	Suiza	FSM	Estados Federados de Micronesia	KHM	Camboya

Abreviatura	País	Abreviatura	País	Abreviatura	País
KIR	Kiribati	NIC	Nicaragua	SUR	Surinam
KNA	San Cristóbal y Nieves	NIU	Niue	SVK	Eslovaquia
KOR	Corea del Sur	NLD	Países Bajos	SVN	Eslovenia
KWT	Kuwait	NOR	Noruega	SWE	Suecia
LAO	Laos	NPL	Nepal	SWZ	Esuatini (Suazilandia)
LBN	Líbano	NRU	Nauru	SXM	Sint Maarten
LBR	Liberia	NZL	Nueva Zelanda	SYC	Seychelles
LBY	Libia	OMN	Omán	SYR	Siria
LCA	Santa Lucía	PAK	Pakistán	TCA	Islas Turcas y Caicos
LIE	Liechtenstein	PAN	Panamá	TCD	Chad
LKA	Sri Lanka	PCN	Islas Pitcairn	TGO	Togo
LSO	Lesoto	PER	Perú	THA	Tailandia
LTU	Lituania	PHL	Filipinas	TJK	Tayikistán
LUX	Luxemburgo	PLW	Palaos	TKL	Tokelau
LVA	Letonia	PNG	Papúa Nueva Guinea	TKM	Turkmenistán
MAC	Macao	POL	Polonia	TLS	Timor Oriental
MAR	Marruecos	PRI	Puerto Rico	TON	Tonga
MCO	Mónaco	PRK	Corea del Norte	TTO	Trinidad y Tobago
MDA	Moldavia	PRT	Portugal	TUN	Túnez
MDG	Madagascar	PRY	Paraguay	TUR	Turquía
MDV	Maldivas	PSE	Palestina	TUV	Tuvalu
MEX	México	PYF	Polinesia Francesa	TWN	Taiwán
MHL	Islas Marshall	QAT	Catar	TZA	Tanzania
MKD	Macedonia del Norte	REU	Reunión	UGA	Uganda
MLI	Malí	ROU	Rumania	UKR	Ucrania
MLT	Malta	RUS	Rusia	URY	Uruguay
MMR	Myanmar	RWA	Ruanda	USA	Estados Unidos
MNE	Montenegro	SAU	Arabia Saudita	UZB	Uzbekistán
MNG	Mongolia	SCG	Serbia y Montenegro	VAT	Ciudad del Vaticano
MNP	Islas Marianas del Norte	SDN	Sudán	VCT	San Vicente y las Granadinas
MOZ	Mozambique	SEN	Senegal	VEN	Venezuela
MRT	Mauritania	SGP	Singapur	VGB	Islas Vírgenes Británicas
MSR	Montserrat	SHN	Santa Elena	VIR	Islas Vírgenes de los Estados Unidos
MTQ	Martinica	SLB	Islas Salomón	VNM	Vietnam
MUS	Mauricio	SLE	Sierra Leona	VUT	Vanuatu
MWI	Malawi	SLV	El Salvador	WLF	Wallis y Futuna
MYS	Malasia	SMR	San Marino	WSM	Samoa
MYT	Mayotte	SOM	Somalia	YEM	Yemen
NAM	Namibia	SPM	San Pedro y Miquelón	YUG	Yugoslavia
NCL	Nueva Caledonia	SRB	Serbia	ZAF	Sudáfrica
NER	Níger	SSD	Sudán del Sur	ZAR	Zaire
NFK	Isla Norfolk	STP	Santo Tomé y Príncipe	ZMB	Zambia
NGA	Nigeria	SUN	Unión Soviética	ZWE	Zimbabue

## 8.2. Resumen estadístico de las variables de la base de datos

Tabla 8.2: Países que contiene la base de datos utilizada para la modelización.

<b>year</b>	<b>country_id_o</b>	<b>country_id_d</b>	<b>rta_coverage</b>
Min. :1990	Length:1968624	Length:1968624	Length:1968624
1st Qu.:1997	Class :character	Class :character	Class :character
Median :2005	Mode :character	Mode :character	Mode :character
Mean :2005			
3rd Qu.:2013			
Max. :2020			
<b>distw_harmonic</b>	<b>contig</b>	<b>comlang_ethno</b>	<b>col45</b>
Min. : 0	Min. :0.00	Min. :0.0	Min. :0
1st Qu.: 4752	1st Qu.:0.00	1st Qu.:0.0	1st Qu.:0
Median : 8106	Median :0.00	Median :0.0	Median :0
Mean : 8481	Mean :0.01	Mean :0.2	Mean :0
3rd Qu.:12020	3rd Qu.:0.00	3rd Qu.:0.0	3rd Qu.:0
Max. :19904	Max. :1.00	Max. :1.0	Max. :1
NA's :287562	NA's :287562	NA's :419936	NA's :419936
<b>comrelig</b>	<b>pop_o</b>	<b>pop_d</b>	<b>gdp_o</b>
Min. :0.0	Min. : 3.4	Min. : 3.4	Min. :8.825e+03
1st Qu.:0.0	1st Qu.: 529.8	1st Qu.: 529.8	1st Qu.:2.948e+06
Median :0.1	Median : 5262.0	Median : 5262.0	Median :1.381e+07
Mean :0.2	Mean : 29982.6	Mean : 29982.6	Mean :2.650e+08
3rd Qu.:0.2	3rd Qu.: 18336.0	3rd Qu.: 18336.0	3rd Qu.:9.929e+07
Max. :1.0	Max. :1411100.0	Max. :1411100.0	Max. :2.137e+10
NA's :711081	NA's :279720	NA's :279720	NA's :457380
<b>gdp_d</b>	<b>wto_o</b>	<b>wto_d</b>	<b>eu_o</b>
Min. :8.825e+03	Min. :0.00	Min. :0.00	Min. :0.00
1st Qu.:2.948e+06	1st Qu.:0.00	1st Qu.:0.00	1st Qu.:0.00
Median :1.381e+07	Median :1.00	Median :1.00	Median :0.00
Mean :2.650e+08	Mean :0.53	Mean :0.53	Mean :0.09
3rd Qu.:9.929e+07	3rd Qu.:1.00	3rd Qu.:1.00	3rd Qu.:0.00
Max. :2.137e+10	Max. :1.00	Max. :1.00	Max. :1.00
NA's :457380	NA's :149688	NA's :149688	NA's :149688
<b>eu_d</b>	<b>tradeflow_imf_o</b>	<b>tradeflow_imf_d</b>	<b>rta</b>
Min. :0.00	Min. : 0	Min. : 0	Min. :0.00000
1st Qu.:0.00	1st Qu.: 121	1st Qu.: 78	1st Qu.:0.00000
Median :0.00	Median : 2574	Median : 2021	Median :0.00000
Mean :0.09	Mean : 511938	Mean : 470514	Mean :0.07011
3rd Qu.:0.00	3rd Qu.: 40129	3rd Qu.: 34341	3rd Qu.:0.00000
Max. :1.00	Max. :480688672	Max. :539503424	Max. :1.00000
NA's :149688	NA's :1326881	NA's :1260404	

### 8.3. Resultados modelización con MCO

Tabla 8.3.1: Modelización con MCO periodo 1990 - 1999.  
**MCO (1990-1999)**

Dependent variable				
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	-15.39549	0.09494	-15.58157 – -15.20940	<b>&lt;0.001</b>
dist log	-1.19262	0.00865	-1.20958 – -1.17566	<b>&lt;0.001</b>
inc o log	1.09282	0.00284	1.08725 – 1.09839	<b>&lt;0.001</b>
inc d log	0.84437	0.00264	0.83919 – 0.84955	<b>&lt;0.001</b>
contig	0.89596	0.03864	0.82022 – 0.97170	<b>&lt;0.001</b>
comlang ethno	0.77373	0.01663	0.74113 – 0.80633	<b>&lt;0.001</b>
col45	1.71785	0.05143	1.61705 – 1.81864	<b>&lt;0.001</b>
comrelig	-0.16975	0.02463	-0.21803 – -0.12148	<b>&lt;0.001</b>
rta	0.37154	0.02443	0.32366 – 0.41942	<b>&lt;0.001</b>
Observations	135221			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.624 / 0.624			

Tabla 8.3.2: Modelización con MCO periodo 2000 - 2009.  
**MCO (2000-2009)**

<i>Predictors</i>	<b>Dependent variable</b>			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	-17.76417	0.09335	-17.94714 – -17.58120	<b>&lt;0.001</b>
dist log	-1.22945	0.00850	-1.24610 – -1.21279	<b>&lt;0.001</b>
inc o log	1.20398	0.00262	1.19884 – 1.20912	<b>&lt;0.001</b>
inc d log	0.83724	0.00248	0.83237 – 0.84210	<b>&lt;0.001</b>
contig	1.16346	0.03880	1.08740 – 1.23951	<b>&lt;0.001</b>
comlang ethno	0.89688	0.01618	0.86517 – 0.92859	<b>&lt;0.001</b>
col45	1.42909	0.05402	1.32321 – 1.53496	<b>&lt;0.001</b>
comrelig	0.00420	0.02341	-0.04167 – 0.05008	0.857
rta	0.60529	0.01981	0.56645 – 0.64412	<b>&lt;0.001</b>
Observations	191166			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.628 / 0.628			

Tabla 8.3.3: Modelización con MCO periodo 2010 - 2019.  
**MCO (2010-2019)**

<i>Predictors</i>	<b>Dependent variable</b>			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	-21.08459	0.09514	-21.27105 – -20.89813	<b>&lt;0.001</b>
dist log	-1.25881	0.00808	-1.27464 – -1.24298	<b>&lt;0.001</b>
inc o log	1.30102	0.00266	1.29581 – 1.30624	<b>&lt;0.001</b>
inc d log	0.88842	0.00256	0.88341 – 0.89343	<b>&lt;0.001</b>
contig	0.99804	0.04032	0.91901 – 1.07707	<b>&lt;0.001</b>
comlang ethno	0.91212	0.01634	0.88009 – 0.94414	<b>&lt;0.001</b>
col45	1.08585	0.05617	0.97577 – 1.19594	<b>&lt;0.001</b>
comrelig	-0.05562	0.02367	-0.10201 – -0.00924	<b>0.019</b>
rta	0.84352	0.01530	0.81353 – 0.87350	<b>&lt;0.001</b>
Observations	211340			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.633 / 0.633			

#### 8.4. Resultados de modelización con MCO controlando por efectos fijos

Tabla 8.4.1: Modelización con MCO con efectos fijos periodo 1990 - 1999.

##### MCO - Efectos fijos (1990-1999)

Dependent variable				
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	10.88049	0.20212	10.48434 – 11.27665	<0.001
dist log	-1.45411	0.00909	-1.47193 – -1.43630	<0.001
contig	0.63967	0.03566	0.56977 – 0.70957	<0.001
comlang ethno	0.59793	0.01833	0.56200 – 0.63386	<0.001
col45	1.71129	0.05064	1.61204 – 1.81055	<0.001
comrelig	0.47648	0.02742	0.42273 – 0.53023	<0.001
rta	0.22475	0.02406	0.17760 – 0.27191	<0.001
Observations	135221			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.707 / 0.706			

Tabla 8.4.2: Modelización con MCO con efectos fijos periodo 2000 - 2009.

**MCO - Efectos fijos (2000-2009)**

<b>Dependent variable</b>				
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	12.46852	0.14311	12.18802 – 12.74901	<b>&lt;0.001</b>
dist log	-1.57420	0.00863	-1.59112 – -1.55728	<b>&lt;0.001</b>
contig	0.63859	0.03490	0.57019 – 0.70698	<b>&lt;0.001</b>
comlang ethno	0.65695	0.01738	0.62289 – 0.69101	<b>&lt;0.001</b>
col45	1.52954	0.05145	1.42870 – 1.63038	<b>&lt;0.001</b>
comrelig	0.54584	0.02511	0.49664 – 0.59505	<b>&lt;0.001</b>
rta	0.60106	0.01912	0.56358 – 0.63854	<b>&lt;0.001</b>
Observations	191166			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.722 / 0.722			



Tabla 8.4.3: Modelización con MCO con efectos fijos periodo 2010 - 2019.  
**MCO - Efectos fijos (2010-2019)**

<b>Dependent variable</b>				
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	12.37297	0.17316	12.03359 – 12.71235	<b>&lt;0.001</b>
dist log	-1.65226	0.00814	-1.66821 – -1.63631	<b>&lt;0.001</b>
contig	0.70722	0.03539	0.63785 – 0.77658	<b>&lt;0.001</b>
comlang ethno	0.70005	0.01715	0.66644 – 0.73366	<b>&lt;0.001</b>
col45	1.37639	0.05254	1.27342 – 1.47935	<b>&lt;0.001</b>
comrelig	0.43070	0.02456	0.38256 – 0.47883	<b>&lt;0.001</b>
rta	0.45670	0.01551	0.42630 – 0.48710	<b>&lt;0.001</b>
Observations	211340			
R <sup>2</sup> / R <sup>2</sup> adjusted	0.736 / 0.735			

## 8.5. Enlaces web de las fuentes de datos

[http://www.cepii.fr/CEPII/en/bdd\\_modele/bdd\\_modele\\_item.asp?id=6](http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=6)

<https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators>

<https://rtais.wto.org/UI/PublicMaintainRTAHome.aspx>

<https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&slid=1479329132316>

## 8.6. Código

```
# ----- DADES -----
# Obtener los datos bilaterales
gravdata_cepil <-
readRDS("C:/Users/marti/OneDrive/Documents/TFG/codis/Gravity_rds_V20
2211/Gravity_V202211.rds")
gravdata_cepil <- subset(gravdata_cepil, year %in% c(1990:2020))
gravdata_cepil <- gravdata_cepil[order(gravdata_cepil$country_id_o,
gravdata_cepil$country_id_d, gravdata_cepil$year), ]

#Variables que usaremos
library(readr)
data <- read_csv("TFG/data.csv")

data <- gravdata_cepil[ ,
c("year","country_id_o","country_id_d","rta","distw_harmonic","conti
g","comlang_ethno","col45","comrelig","pop_o","pop_d","gdp_o","gdp_d
","wto_o","wto_d","eu_o","eu_d","tradeflow_imf_o","tradeflow_imf_d")
]

# ----- DESCRIPTIVA -----
library(stats)

correlation1 <- cor(log(data_1[,c('tradeflow_imf_o','gdp_o','gdp_d',
'distw_harmonic')]),
use="na.or.complete")
correlation1

correlation2 <- cor(log(data_2[,c('tradeflow_imf_o','gdp_o','gdp_d',
'distw_harmonic')]),
use="na.or.complete")
correlation2

correlation3 <- cor(log(data_3[,c('tradeflow_imf_o','gdp_o','gdp_d',
'distw_harmonic')]),
use="na.or.complete")
correlation3

#gdp vs trade
par(mfrow=c(1,3))
plot(log(data_1$gdp_o * data_1$gdp_d), log(data_1$tradeflow_imf_o),
pch=1, xlab = "ln (gdp_o*gdp_d)", ylab = "ln (tradeflow)",
col="blue3",
main="Flujo comercial frente PIB 1990-2000",
cex.lab=0.9,cex.main=1)

abline(lm(log(data_1$tradeflow_imf_o)~log(data_1$gdp_o *
data_1$gdp_d)), col="red3", lwd=2)
```

```

text(paste("Correlación:", round(cor(log(data_1$gdp_o *
data_1$gdp_d),
      log(data_1$tradesflow_imf_o)), 4)), x = 26.5, y =
17,cex=0.9)

plot(log(data_2$gdp_o * data_2$gdp_d), log(data_2$tradesflow_imf_o),
      pch=1, xlab = "ln (gdp_o*gdp_d)", ylab = "ln (tradesflow)",
col="blue3",
      main="Flujo comercial frente PIB 2001-2010",
cex.lab=0.9,cex.main=1)

abline(lm(log(data_2$tradesflow_imf_o)~log(data_2$gdp_o *
      data_2$gdp_d)),
col="red3", lwd=2)

text(paste("Correlación:", round(cor(log(data_2$gdp_o *
data_2$gdp_d),
      log(data_2$tradesflow_imf_o)),
4)), x = 26.5, y = 17,cex=0.9)

plot(log(data_3$gdp_o * data_3$gdp_d), log(data_3$tradesflow_imf_o),
      pch=1, xlab = "ln (gdp_o*gdp_d)", ylab = "ln (tradesflow)",
col="blue3",
      main="Flujo comercial frente PIB 2011-2019",
cex.lab=0.9,cex.main=1)

abline(lm(log(data_3$tradesflow_imf_o)~log(data_3$gdp_o *
      data_3$gdp_d)),
col="red3", lwd=2)

text(paste("Correlación:", round(cor(log(data_3$gdp_o *
data_3$gdp_d),
      log(data_3$tradesflow_imf_o)),
4)), x = 28, y = 17,cex=0.9)

#trade vs dist
plot(log(data_1$distw_harmonic), log(data_1$tradesflow_imf_o),
      pch=1, xlab = "ln (distw_harmonic)", ylab = "ln (tradesflow)",
col="blue3",
      main="Flujo comercial frente distancia 1990-2000",
cex.lab=0.9,cex.main=1)

abline(lm(log(data_1$tradesflow_imf_o)~log(data_1$distw_harmonic)),
col="red3", lwd=2)

text(paste("Correlación:", round(cor(log(data_1$distw_harmonic),
      log(data_1$tradesflow_imf_o)),
4)), x = 4.5, y = -3,cex=0.9)

plot(log(data_2$distw_harmonic), log(data_2$tradesflow_imf_o),
      pch=1, xlab = "ln (distw_harmonic)", ylab = "ln (tradesflow)",
col="blue3",

```

```

    main="Flujo comercial frente distancia 2001-2010",
    cex.lab=0.9,cex.main=1)

abline(lm(log(data_2$tradeflow_imf_o)~log(data_2$distw_harmonic)),
col="red3", lwd=2)

text(paste("Correlación:", round(cor(log(data_2$distw_harmonic),
                                     log(data_2$tradeflow_imf_o)),
4)), x = 4.5, y = -3,cex=0.9)

plot(log(data_3$distw_harmonic), log(data_3$tradeflow_imf_o),
      pch=1, xlab = "ln (distw_harmonic)", ylab = "ln (tradeflow)",
col="blue3",
      main="Flujo comercial frente 2011-2019",
cex.lab=0.9,cex.main=1)

abline(lm(log(data_3$tradeflow_imf_o)~log(data_3$distw_harmonic)),
col="red3", lwd=2)

text(paste("Correlación:", round(cor(log(data_3$distw_harmonic),
                                     log(data_3$tradeflow_imf_o)),
4)), x = 4.5, y = -3,cex=0.9)

summary(data)

library(dplyr)
total <- aggregate(data$tradeflow_imf_o/1000000 ~ data$year,
                   FUN = sum)
colnames(total) <- c('year', 'flujo')

barplot(total$flujo,names.arg=total$year, col='blue3',
        xlab = "Año", ylim=c(0,20000), cex.axis=0.9,
        ylab = "Miles de millones de $", cex=0.9, cex.lab=1,
        main = "Flujos comerciales del 1990 al 2020")

# ----- MODELS -----
library(sjPlot)
library(gravity)
MCO_EF_1 <- fixed_effects(dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors =
c("contig","comlang_ethno","col45","comrelig","rta"),
code_origin="country_id_o",
code_destination="country_id_d",
robust = FALSE,
data=data_1)

MCO_EF_2 <- fixed_effects(dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors =
c("contig","comlang_ethno","col45","comrelig","rta"),
code_origin="country_id_o",

```

```

code_destination="country_id_d",
robust = FALSE,
data=data_2)

MCO_EF_3 <- fixed_effects(dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors =
c("contig","comlang_ethno","col45","comrelig","rta"),
code_origin="country_id_o",
code_destination="country_id_d",
robust = FALSE,
data=data_3)

tab_model(MCO_EF_1, show.se = TRUE, digits = 5, title =
'MCO - Efectos fijos (1990-1999)')
tab_model(MCO_EF_2, show.se = TRUE, digits = 5, title =
'MCO - Efectos fijos (2000-2009)')
tab_model(MCO_EF_3, show.se = TRUE, digits = 5, title =
'MCO - Efectos fijos (2010-2019)')
tab_model(MCO_EF_1,MCO_EF_2,MCO_EF_3, show.se = TRUE, digits = 5,
title =
'MCO - Efectos fijos')

ols_1 <- ols(
dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors =
c("contig","comlang_ethno","col45","comrelig","rta"),
income_origin="gdp_o",
income_destination="gdp_d",
code_origin="country_id_o",
code_destination="country_id_d",
uie = FALSE,
robust = FALSE,
data=data_1)

ols_2 <- ols(
dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors =
c("contig","comlang_ethno","col45","comrelig","rta"),
income_origin="gdp_o",
income_destination="gdp_d",
code_origin="country_id_o",
code_destination="country_id_d",
uie = FALSE,
robust = FALSE,
data=data_2)

ols_3 <- ols(
dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",

```

```

additional_regressors =
c("contig","comlang_ethno","col45","comrelig","rta"),
income_origin="gdp_o",
income_destination="gdp_d",
code_origin="country_id_o",
code_destination="country_id_d",
uie = FALSE,
robust = FALSE,
data=data_3)

tab_model(ols_1, show.se = TRUE, digits = 5, title ='MCO (1990-
1999)')
tab_model(ols_2, show.se = TRUE, digits = 5, title ='MCO (2000-
2009)')
tab_model(ols_3, show.se = TRUE, digits = 5, title ='MCO (2010-
2019)')
tab_model(ols_1,ols_2,ols_3, show.se = TRUE, digits = 5, title
='MCO')

ppml_1 <- ppml(
dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors=
c("contig","comlang_ethno","col45","comrelig","rta"),
robust = T, method = "whitel", data=data_1_p
)

ppml_2 <- ppml(
dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors=
c("contig","comlang_ethno","col45","comrelig","rta"),
robust = T, method = "whitel", cluster =
c("country_id_d","country_id_o"),
data=data_2_p
)

ppml_3 <- ppml(
dependent_variable="tradeflow_imf_o",
distance="distw_harmonic",
additional_regressors=
c("contig","comlang_ethno","col45","comrelig","rta"),
robust = T, method = "whitel", cluster =
c("country_id_d","country_id_o"),
data=data_3_p
)

library(lmtest)
r1<-resettest(ppml_1)
r2<-resettest(ppml_2)
r3<-resettest(ppml_3)

tab_model(ppml_1, show.se = TRUE, digits = 2, title =

```

```
'Pseudo verosimilitud Poisson (1990-1999)')
tab_model(ppml_2, show.se = TRUE, digits = 5, title =
'Pseudo verosimilitud Poisson (2001-2010)')
tab_model(ppml_3, show.se = TRUE, digits = 5, title =
'Pseudo verosimilitud Poisson (2011-2020)')
tab_model(ppml_1, ppml_2, ppml_3, show.se = TRUE, digits = 5, title =
'Pseudo verosimilitud Poisson')
```