



Lisbon School
of Economics
& Management
Universidade de Lisboa

MESTRADO EM ECONOMIA INTERNACIONAL E ESTUDOS
EUROPEUS

TRABALHO FINAL DE MESTRADO
DISSERTAÇÃO

UMA ANÁLISE DOS DETERMINANTES DO COMÉRCIO
INTERNACIONAL ATRAVÉS DO MODELO GRAVITACIONAL:
EVIDÊNCIA PARA UMA AMOSTRA DE 30 PAÍSES

JOANA MARIA HENRIQUES RIBEIRO SALGADO

TRABALHO ORIENTADO POR:
PEDRO LEÃO

JULHO 2026

RESUMO

A presente dissertação analisa os determinantes do comércio internacional através da estimação do modelo gravitacional, comparando duas abordagens distintas: uma especificação direcional, que distingue explicitamente o país exportador do país importador, e uma especificação simétrica, baseada no comércio agregado entre pares de países.

A análise utiliza dados da base CEPII para uma amostra de 30 países, referentes ao ano de 2019, resultando em 861 observações direcionais e 433 pares bilaterais simétricos. Para cada abordagem foram estimadas três especificações econométricas sucessivas, incorporando progressivamente variáveis económicas (PIB e distância bilateral), institucionais (acordos comerciais regionais, língua oficial comum, fronteira terrestre, relações coloniais e colonizador comum) e estruturais (exportação de recursos naturais e um índice de *remoteness*, utilizado como aproximação à resistência multilateral ao comércio).

Os resultados confirmam as principais previsões da teoria gravitacional: o efeito positivo da dimensão económica e o efeito negativo da distância geográfica, sobre os fluxos comerciais. Os acordos comerciais regionais e as relações coloniais contribuem positivamente para o comércio bilateral, enquanto as restantes variáveis institucionais não são estatisticamente significativas. A exportação de recursos naturais está associada a um menor volume de comércio, e o índice de *remoteness* tem um efeito positivo e significativo em ambas as especificações. Os testes de diagnóstico confirmam a adequação da especificação e a ausência de multicolinearidade relevante, mitigando-se a heterocedasticidade detetada através de erros-padrão robustos.

A comparação entre as duas especificações revela elevada consistência de resultados, sendo isto o principal contributo desta dissertação: mostrar que os determinantes do comércio internacional são robustos à forma como os fluxos comerciais são representados.

Palavras-chave: modelo gravitacional; comércio internacional; resistência multilateral; *remoteness*; acordos comerciais regionais.

ABSTRACT

International trade plays a fundamental role in economic growth and in the integration between economies, making it relevant to identify the factors that determine the intensity of trade flows between countries. In this context, the present dissertation analyses the determinants of international trade through the estimation of the gravity model, comparing two distinct approaches: a directional specification, which explicitly distinguishes the exporting country from the importing country, and a symmetric specification, based on aggregated trade between country pairs.

The analysis uses data from the CEPII database for a sample of 30 countries, referring to 2019, resulting in 861 directional observations and 433 symmetric country pairs. For each approach, three successive econometric specifications were estimated, progressively incorporating economic variables (GDP and bilateral distance), institutional variables (regional trade agreements, common official language, land border, colonial relationship and common colonizer) and structural variables (natural resource exports and a remoteness index, used as a proxy for multilateral resistance to trade).

The results confirm the main predictions of gravity theory, showing a positive effect of economic size and a negative effect of geographic distance on trade flows. Regional trade agreements and colonial ties contribute positively to bilateral trade, while the remaining institutional variables are not statistically significant. Natural resource exports are associated with lower trade volumes, and the remoteness index shows a positive and significant effect in both specifications. Diagnostic tests confirm the adequacy of the specification and the absence of relevant multicollinearity, with heteroskedasticity mitigated through robust standard errors.

The comparison between the two specifications reveals high consistency of results, representing the main contribution of this dissertation: showing that the determinants of international trade are robust to how trade flows are represented.

Keywords: gravity model; international trade; multilateral resistance; remoteness; regional trade agreements.

ÍNDICE

RESUMO	1
ABSTRACT	2
1. INTRODUÇÃO	6
2. REVISÃO DA LITERATURA	7
3. DADOS E METODOLOGIA	9
3.1. DADOS E CONSTRUÇÃO DA AMOSTRA	9
3.2. VARIÁVEIS DO MODELO	10
3.2.1. VARIÁVEL DEPENDENTE	12
3.2.2. VARIÁVEIS ECONÓMICAS	13
3.2.3. VARIÁVEIS INSTITUCIONAIS	13
3.2.4. VARIÁVEIS ADICIONAIS	14
3.3. CÁLCULO DO ÍNDICE DE REMOTENESS	15
3.4. CONSTRUÇÃO DA BASE DE DADOS DO MODELO SIMÉTRICO	16
3.5. ESPECIFICAÇÃO ECONOMÉTRICA	17
3.5.1. MODELO GRAVITACIONAL DIRECIONAL	17
1. MODELO 1 - ESPECIFICAÇÃO BÁSICA DO MODELO GRAVITACIONAL	17
2. MODELO 2 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COM VARIÁVEIS INSTITUCIONAIS	18
3. MODELO 3 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COMPLETA	18
3.5.2. MODELO GRAVITACIONAL SIMÉTRICO	19
1. MODELO 1 - ESPECIFICAÇÃO BÁSICA DO MODELO GRAVITACIONAL	19
2. MODELO 2 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COM VARIÁVEIS INSTITUCIONAIS	19
3. MODELO 3 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COMPLETA	19
4. RESULTADOS EMPÍRICOS	20
4.1. ESTATÍSTICA DESCRITIVA	20
4.2. MATRIZ DE CORRELAÇÕES	21
4.3. MODELO GRAVITACIONAL DIRECIONAL	23
4.4. MODELO GRAVITACIONAL SIMÉTRICO	28
4.5. COMPARAÇÃO ENTRE MODELOS	33
5. TESTES DE VALIDADE	35
5.1. TESTE DE HETEROCEDASTICIDADE	35
5.2. TESTE RESET	36
5.3. TESTE DE MULTICOLINEARIDADE	37
6. CONCLUSÕES	38
BIBLIOGRAFIA	41
ANEXOS	42
ANEXO A - LISTA DOS PAÍSES DA AMOSTRA	42

ANEXO B - PRINCIPAIS COMANDOS UTILIZADOS NO STATA	44
ANEXO C - OUTPUT DO CÁLCULO DE REMOTENESS: LISTA DE PAÍSES POR NÍVEL DE REMOTENESS ASCENDENTE	47
ANEXO D - OUTPUTS COMPLETOS DAS ESTIMAÇÕES E DOS TESTES DE VALIDADE	48

AGRADECIMENTOS

Dirijo o meu agradecimento ao Professor Pedro Leão, pelo seu acompanhamento e orientação ao longo deste percurso.

1. INTRODUÇÃO

O comércio internacional constitui um dos principais motores do crescimento económico. Compreender os fatores que determinam a intensidade das trocas comerciais tornou-se, por isso, um tema de particular relevância para a formulação de políticas públicas.

Entre os diversos modelos desenvolvidos para esse efeito, o modelo gravitacional destacou-se pela sua elevada capacidade explicativa. Inspirado na Lei da Gravitação Universal de Newton, este modelo estabelece que o comércio entre dois países tende a aumentar com a sua dimensão económica e a diminuir com a distância geográfica que os separa. Apesar de a sua estrutura assentar em determinantes relativamente intuitivos, como a dimensão económica dos países e a distância geográfica entre eles, a literatura tem vindo a mostrar que fatores institucionais e características estruturais desempenham igualmente um papel relevante na explicação do comércio bilateral.

Neste contexto, a presente dissertação procura responder à seguinte questão de investigação:

Em que medida os determinantes tradicionais do modelo gravitacional explicam os fluxos comerciais internacionais, e de que forma os resultados diferem quando o comércio é analisado através de uma especificação direcional ou de uma especificação simétrica?

Para responder a esta questão, procede-se à estimação de dois modelos gravitacionais distintos, utilizando informação relativa a uma amostra de 30 países para o ano de 2019. O primeiro modelo considera cada fluxo comercial de forma direcional, distinguindo explicitamente o país exportador do país importador. O segundo agrega os fluxos comerciais dos dois sentidos, permitindo analisar a intensidade total do comércio entre cada par de países. Em ambas as especificações são estimados três modelos sucessivos, incorporando progressivamente variáveis económicas, institucionais e estruturais. Entre estas destaca-se um índice de *remoteness*, utilizado como aproximação empírica ao conceito de resistência multilateral ao comércio.

A comparação entre estas duas abordagens constitui um dos principais contributos da presente dissertação, permitindo avaliar a robustez dos determinantes tradicionalmente

identificados pela literatura e analisar até que ponto a forma de construção da base de dados influencia os resultados obtidos.

A dissertação encontra-se organizada em seis capítulos. Após a presente introdução, o capítulo 2 apresenta a revisão da literatura relativa ao modelo gravitacional e aos seus principais desenvolvimentos teóricos e econométricos. O capítulo 3 descreve os dados utilizados, a construção da amostra, as variáveis consideradas e a metodologia adotada. O capítulo 4 apresenta e interpreta os resultados empíricos obtidos para as especificações direcional e simétrica. No capítulo 5 são apresentados os testes de validade realizados aos modelos estimados. Finalmente, o capítulo 6 sintetiza as principais conclusões, identifica as limitações do estudo e sugere possíveis linhas de investigação futura.

2. REVISÃO DA LITERATURA

A primeira formulação do modelo gravitacional aplicada ao comércio internacional foi apresentada por Tinbergen (1962), que mostrou empiricamente que o volume de comércio entre dois países é diretamente proporcional ao Produto Interno Bruto (PIB) dos países envolvidos e inversamente proporcional à distância entre eles. Apesar do elevado sucesso empírico desta primeira aplicação, o modelo permaneceu, durante vários anos, essencialmente como uma regularidade estatística, sem uma fundamentação teórica suficientemente robusta.

Um dos desenvolvimentos mais relevantes ocorreu com Anderson e van Wincoop (2003), que introduziram o conceito de *multilateral resistance*. Os autores mostraram que o comércio entre dois países não depende apenas da distância bilateral ou da dimensão das respectivas economias, mas também da posição relativa de cada país face ao conjunto dos restantes parceiros comerciais: a decisão de comerciar entre dois países é assim influenciada não só pelos custos bilaterais, mas igualmente pelos custos de comércio que cada um enfrenta relativamente ao resto do mundo. Este contributo permitiu ultrapassar uma das principais limitações das especificações gravitacionais tradicionais e constitui, atualmente, uma referência incontornável na literatura. A operacionalização empírica deste conceito tem sido feita de duas formas principais: através de índices de *remoteness*, como o utilizado nesta dissertação, ou através da inclusão de efeitos fixos de exportador e de importador. Esta segunda abordagem é hoje mais comum na literatura por

captar de forma mais completa a resistência multilateral, embora à custa de não permitir identificar separadamente o efeito de variáveis específicas de cada país.

Em paralelo com estes desenvolvimentos teóricos, a literatura identificou também problemas de natureza econométrica na estimação tradicional do modelo por Mínimos Quadrados Ordinários (OLS) aplicados à forma logarítmica. Santos Silva e Tenreyro (2006) mostraram que esta abordagem apresenta dois problemas relevantes: por um lado, a exclusão das observações de comércio nulo, frequentes em amostras com países heterogêneos, pode enviesar as estimativas; por outro, a heterocedasticidade dos erros, decorrente da desigualdade de Jensen ao aplicar logaritmos a uma variável cuja variância cresce com o nível, compromete a eficiência e a consistência dos estimadores OLS. Os autores propuseram como alternativa o método de *Pseudo-Maximum Likelihood Poisson* (PPML), atualmente amplamente utilizado na literatura gravitacional. Esta discussão é particularmente relevante para o presente estudo, uma vez que os testes de diagnóstico apresentados no capítulo 5 confirmam a presença de heterocedasticidade nos resíduos, problema que é mitigado através do uso de erros-padrão robustos, mas que reforça a pertinência desta linha crítica da literatura.

Em consequência destes desenvolvimentos, o modelo gravitacional deixou de incluir apenas as variáveis económicas clássicas, passando a incorporar um conjunto mais vasto de fatores institucionais, como a existência de uma língua oficial comum, fronteiras partilhadas, relações coloniais históricas ou acordos comerciais regionais. Relativamente a estes últimos, Baier e Bergstrand (2007) alertaram para o facto de a participação em acordos comerciais regionais (RTA) não ser aleatória, mas antes resultar de decisões de política comercial potencialmente correlacionadas com fatores não observados que também afetam o comércio, levantando problemas na estimação do seu efeito, causados pela captação de variáveis explicativas relevantes que estão omitidas. Embora a presente dissertação não implemente correções específicas para este problema, esta ressalva deve ser tida em conta na interpretação do coeficiente estimado para a variável RTA.

Com base na formulação tradicional do modelo gravitacional e nos desenvolvimentos introduzidos pela literatura mais recente, procede-se nesta dissertação à estimação de duas especificações distintas, direcional e simétrica, para uma amostra de 30 países, com dados referentes a 2019. O estudo incorpora ainda variáveis institucionais e um índice de *remoteness*, construído com base na abordagem de Anderson e van Wincoop (2003). O principal contributo

para a literatura é a comparação entre especificações e a avaliação da robustez dos principais determinantes do comércio internacional.

3. DADOS E METODOLOGIA

Neste capítulo, será apresentada a base de dados utilizada, o processo de construção da amostra, as variáveis incluídas no modelo e a estratégia econométrica adotada. A análise empírica assenta numa abordagem gravitacional do comércio internacional, aplicada a uma amostra de países selecionados para o ano de 2019.

O estudo combina uma especificação direcional, em que cada observação representa um fluxo comercial de um país exportador para um país importador, com uma especificação simétrica, em que os fluxos nos dois sentidos são agregados num único par bilateral. Esta dupla abordagem permite analisar os determinantes dos fluxos comerciais tanto do ponto de vista das exportações como da intensidade comercial total entre pares de países.

3.1. DADOS E CONSTRUÇÃO DA AMOSTRA

A base de dados utilizada nesta dissertação foi construída a partir do Gravity Dataset do CEPII, versão V202211. A análise foi restringida ao ano de 2019, por ser o período mais recente anterior às perturbações económicas associadas à pandemia de COVID-19, permitindo assim observar os fluxos comerciais internacionais num contexto relativamente estável, sem os efeitos atípicos que marcaram os anos seguintes.

A amostra é composta por 30 países, selecionados com o objetivo de garantir heterogeneidade em várias dimensões: economias desenvolvidas e emergentes, diferentes regiões geográficas, e países com relevância na exportação de recursos naturais. Esta diversidade permite captar diferenças em termos de dimensão económica, localização geográfica, grau de integração institucional e estrutura produtiva, características que são centrais para a análise dos determinantes do comércio internacional proposta nesta dissertação. A lista completa dos países incluídos na amostra encontra-se no Anexo A.

A partir desta seleção de países, a base foi tratada em várias etapas. Em primeiro lugar, restringiu-se a base aos pares em que tanto o país de origem como o país de destino pertenciam à

amostra selecionada, e removeram-se as observações em que o país exportador coincidia com o país importador, uma vez que o objetivo do estudo é analisar comércio internacional e não fluxos comerciais domésticos.

A estrutura principal da base resultante é direcional: cada observação corresponde a um fluxo comercial específico do país exportador i para o país importador j . Por exemplo, o fluxo de Portugal para Espanha é tratado como uma observação distinta do fluxo de Espanha para Portugal. Esta opção foi deliberada, uma vez que permite captar eventuais assimetrias nos fluxos comerciais bilaterais e analisar separadamente o efeito de características do país exportador e do país importador, algo que seria perdido caso os dois sentidos do comércio fossem agregados à partida.

No decurso do tratamento da base, foram identificadas observações duplicadas para alguns pares exportador-importador. Uma análise mais detalhada destes casos revelou que as duplicações resultavam sobretudo de diferenças na disponibilidade de informação relativa aos fluxos comerciais entre as observações repetidas. Optou-se, por isso, por ordenar os dados de forma a privilegiar as observações com valores efetivamente disponíveis para a variável de comércio, retendo-se posteriormente apenas uma observação por par direcional.

A variável de comércio foi construída com base na variável *tradeflow_baci*, disponibilizada pelo *CEPII*, que corresponde a fluxos comerciais harmonizados e é expressa em milhares de dólares correntes. Após a remoção das observações sem informação disponível para o comércio bilateral, a amostra final utilizada na estimação do modelo direcional ficou composta por 861 observações.

3.2. VARIÁVEIS DO MODELO

Este capítulo apresenta as variáveis utilizadas na análise empírica, agrupadas em quatro categorias: a variável dependente, as variáveis económicas, as variáveis institucionais e as variáveis adicionais de natureza estrutural. Como o estudo compreende duas especificações, algumas variáveis assumem formas distintas consoante o modelo em que são utilizadas, correspondência que é identificada na Tabela 3.2.

Tabela 3.2. - identificação e descrição das variáveis do modelo

Variável	Descrição	Tipo	Sinal esperado	Modelo
X_{ij}	Fluxo comercial do país exportador para o país importador	Contínua	N/A	Direcional
$Trade$	Comércio bilateral total entre o par de países	Contínua	N/A	Simétrico
GDP_i	PIB do país exportador	Contínua	+	Direcional
GDP_j	PIB do país importador	Contínua	+	Direcional
GDP_{pair}	Produto dos PIB dos dois países	Contínua	+	Simétrico
$Dist_{ij}$	Distância geográfica bilateral (km)	Contínua	-	Ambos
$Comlang_{ij}$	Língua oficial comum	Dummy	+	Ambos
$Contig_{ij}$	Fronteira terrestre comum	Dummy	+	Ambos
$ColDep_{ij}$	Relação colonial	Dummy	+	Ambos
$ComCol_{ij}$	Colonizador comum	Dummy	+	Ambos
RTA_{ij}	Existência de acordo comercial	Dummy	+	Ambos
$Resource_i$	País exportador de recursos naturais	Dummy	Indeterminado	Direcional
$Remoteness_i$	Índice de <i>remoteness</i> do exportador	Contínua	-	Direcional

$Remoteness_j$	Índice de <i>remoteness</i> do importador	Contínua	-	Direcional
$Remoteness_{pair}$	<i>Remoteness</i> médio do par	Contínua	-	Simétrico

Todas as variáveis contínuas apresentadas na tabela foram transformadas em logaritmo natural antes da estimação dos modelos econométricos. Esta opção segue a prática habitual na literatura sobre modelos gravitacionais e tem três vantagens principais: permite linearizar relações que, na sua forma original, são multiplicativas; reduz a assimetria das distribuições, atenuando a influência de observações extremas; e possibilita interpretar os coeficientes estimados diretamente como elasticidades, facilitando a leitura económica dos resultados. Os capítulos seguintes detalham a construção e a fundamentação teórica de cada uma destas variáveis.

3.2.1. VARIÁVEL DEPENDENTE

A variável dependente do modelo direcional corresponde ao fluxo comercial do país exportador para o país importador, X_{ij} , e representa o valor total das exportações do país i para o país j , expresso em milhares de dólares correntes.

No modelo simétrico, a variável dependente corresponde ao comércio bilateral total, *Trade*, obtido através da agregação dos fluxos comerciais nos dois sentidos entre cada par de países. Assim, o comércio entre Portugal e Espanha, por exemplo, corresponde à soma das exportações de Portugal para Espanha e de Espanha para Portugal.

Em ambas as especificações, a variável dependente foi transformada em logaritmo natural, de acordo com o procedimento descrito no capítulo anterior.

3.2.2. VARIÁVEIS ECONÓMICAS

As variáveis económicas constituem o núcleo do modelo gravitacional e refletem os dois principais determinantes do comércio internacional identificados pela teoria: a dimensão económica dos países e a distância geográfica entre eles.

A dimensão económica é representada pelo Produto Interno Bruto (PIB). No modelo direcional, são utilizadas duas variáveis distintas: o PIB do país exportador, GDP_i e o PIB do país importador, GDP_j . Estas permitem captar, respetivamente, a capacidade produtiva do país de origem e a capacidade de absorção da procura do país de destino. Espera-se que ambos os coeficientes apresentem sinal positivo, uma vez que economias de maior dimensão tendem a produzir mais bens e serviços e, simultaneamente, a gerar maior procura por importações.

No modelo simétrico, a distinção entre país exportador e país importador deixa de existir. Em vez de considerar separadamente os dois PIB, foi construída uma variável correspondente ao produto dos PIB dos dois países, GDP_{pair} , refletindo a dimensão económica conjunta do par bilateral. Esta abordagem é consistente com a formulação tradicional do modelo gravitacional e permite avaliar de que forma a dimensão agregada das duas economias influencia o volume total de comércio entre elas.

A distância geográfica bilateral, $Dist_{ij}$, medida em quilómetros e disponibilizada pela base de dados do CEPIL, foi utilizada como proxy dos custos associados ao comércio internacional, incluindo custos de transporte, logística, informação e coordenação entre mercados. A teoria gravitacional prevê uma relação negativa entre a distância e os fluxos comerciais, uma vez que distâncias maiores tendem a aumentar os custos das trocas internacionais e a reduzir a intensidade das relações comerciais entre países.

3.2.3. VARIÁVEIS INSTITUCIONAIS

Para além das variáveis económicas tradicionalmente utilizadas nos modelos gravitacionais, foram incluídas variáveis institucionais e históricas que procuram captar fatores suscetíveis de facilitar ou dificultar as relações comerciais entre países.

A variável $Comlang_{ij}$ identifica os pares de países que partilham uma língua oficial comum. A existência de uma língua comum tende a reduzir os custos de comunicação, facilitar as negociações comerciais e diminuir a incerteza nas transações internacionais, pelo que se prevê um efeito positivo sobre os fluxos comerciais.

A variável $Contig_{ij}$ assume o valor 1 quando os dois países partilham uma fronteira terrestre e 0 nos restantes casos. A proximidade geográfica entre países vizinhos pode favorecer o comércio através da redução dos custos de transporte e da maior integração das infraestruturas logísticas, esperando-se igualmente um coeficiente positivo.

A variável $ColDep_{ij}$ identifica pares de países que mantiveram uma relação colonial no passado. Apesar de estas relações históricas terem terminado, a literatura sugere que podem continuar a influenciar os padrões atuais de comércio, através da manutenção de ligações económicas, institucionais, culturais e linguísticas estabelecidas ao longo do tempo. Consequentemente, espera-se um efeito positivo sobre os fluxos comerciais.

De forma complementar, a variável $comcol$ identifica pares de países que partilharam um mesmo colonizador. A existência de um passado institucional comum pode traduzir-se em sistemas jurídicos semelhantes e afinidades administrativas, fatores que poderão contribuir para uma maior intensidade das trocas comerciais.

Por último, a variável RTA_{ij} identifica a existência de um acordo comercial regional entre os dois países, construída a partir da informação disponibilizada pelo CEP II, assumindo o valor 1 quando existe um acordo em vigor entre o par de países e 0 caso contrário. A inclusão desta variável permite captar o efeito da integração económica sobre os fluxos comerciais, esperando-se um impacto positivo, uma vez que os acordos comerciais tendem a reduzir barreiras tarifárias e não tarifárias.

3.2.4. VARIÁVEIS ADICIONAIS

Para complementar os determinantes tradicionalmente considerados no modelo gravitacional, foram incluídas duas variáveis adicionais: uma variável representativa da especialização em recursos naturais e um índice de *remoteness*. A inclusão destas variáveis

procura captar características estruturais dos países que poderão influenciar os fluxos comerciais, para além da dimensão económica, distância geográfica ou fatores institucionais.

A variável $Resource_i$ identifica os países exportadores com economias baseadas na produção e exportação de recursos naturais, nomeadamente petróleo, gás natural e outros recursos minerais, assumindo o valor 1 para os países classificados como exportadores de recursos naturais e 0 para os restantes. A sua inclusão permite avaliar se a especialização neste tipo de recursos influencia a intensidade dos fluxos comerciais. Atendendo à evidência empírica pouco consensual na literatura, não é possível estabelecer, à partida, um sinal esperado inequívoco para este coeficiente. Esta variável foi construída especificamente para o modelo direcional: dado que caracteriza o país exportador, a sua utilização no modelo simétrico comprometeria a interpretação económica do coeficiente, uma vez que a agregação dos fluxos comerciais elimina a distinção entre exportador e importador.

Foi igualmente considerado um índice de *remoteness*, utilizado como proxy da resistência multilateral ao comércio. Procura captar a distância média de cada país face ao conjunto dos *restantes* parceiros comerciais, permitindo controlar o facto de países mais periféricos poderem enfrentar condições de acesso ao comércio internacional que não são explicadas apenas pela distância bilateral. A metodologia de cálculo é apresentada no capítulo 3.3.

No modelo direcional, foram calculados índices distintos para o país exportador, $Remoteness_i$, e para o país importador, $Remoteness_j$. No modelo simétrico, em que deixa de existir a distinção entre exportador e importador, foi utilizado um índice médio, $Remoteness_{pair}$, obtido a partir da média dos índices individuais dos dois países que compõem cada par bilateral.

3.3. CÁLCULO DO ÍNDICE DE REMOTENESS

A literatura sobre modelos gravitacionais reconhece que a distância bilateral, por si só, poderá não captar integralmente os custos associados ao comércio entre países. Como referido na revisão de literatura, Anderson e van Wincoop (2003) mostram que os fluxos comerciais dependem também da distância relativa de cada economia face ao conjunto dos *restantes* parceiros comerciais: a resistência multilateral ao comércio (*multilateral resistance*).

Para incorporar este efeito na análise empírica, foi construído um índice de *remoteness* para cada país da amostra, calculado como uma média das distâncias bilaterais ponderada pelo PIB dos países parceiros. Desta forma, países com maior peso económico exercem uma influência proporcionalmente superior no cálculo do índice. O índice foi obtido através da seguinte expressão:

$$Remoteness_i = \frac{\sum_{j=1}^N Dist_{ij} \times GDP_j}{\sum_{j=1}^N GDP_j}$$

,onde $Dist_{ij}$ representa a distância geográfica entre o país i e o país j , GDP_j corresponde ao Produto Interno Bruto do país parceiro, e N representa o número de países considerados na amostra.

O índice foi calculado em duas versões: uma primeira considerando cada país enquanto exportador, $Remoteness_i$, e uma segunda considerando cada país enquanto importador, $Remoteness_j$. Ambas as variáveis foram posteriormente transformadas em logaritmo natural, seguindo o procedimento já descrito no capítulo 3.2.

3.4. CONSTRUÇÃO DA BASE DE DADOS DO MODELO SIMÉTRICO

Para além da base direcional, utilizada como especificação principal desta dissertação, foi desenvolvida uma segunda base de dados correspondente a uma especificação simétrica do modelo gravitacional. Esta abordagem permite analisar o comércio bilateral ao nível do par de países, eliminando a distinção entre país exportador e país importador, e constitui um exercício adicional de robustez dos resultados obtidos com a especificação direcional.

Enquanto a especificação direcional considera separadamente os fluxos de i para j e de j para i , o modelo simétrico trata ambos os sentidos como pertencentes ao mesmo par bilateral, permitindo analisar o comércio total entre dois países independentemente da direção das trocas.

A construção da base simétrica seguiu três etapas. Em primeiro lugar, criou-se uma identificação única para cada par de países (*pair_id*), obtida ordenando alfabeticamente os

códigos *ISO3* dos dois países e combinando-os numa única variável. Este procedimento garante que pares como Portugal–Espanha e Espanha–Portugal sejam tratados como uma única observação bilateral.

Em segundo lugar, os fluxos comerciais registados nos dois sentidos foram agregados, dando origem à variável *Trade*, que representa o comércio bilateral total entre cada par de países. A base foi então reduzida para uma única observação por par bilateral, eliminando a duplicação inerente à estrutura direcional. A base simétrica final ficou constituída por 433 pares únicos de países.

Por fim, uma vez eliminada a distinção entre exportador e importador, foi necessário adaptar as variáveis explicativas correspondentes: a dimensão económica do par passou a ser representada pelo produto dos dois PIB, GDP_{pair} , e o índice de *remoteness* foi condensado numa média aritmética dos dois índices do par, $Remoteness_{pair}$. Ambas as variáveis, tal como a variável dependente *Trade*, foram transformadas em logaritmo natural antes da estimação, garantindo consistência com a especificação direcional e permitindo interpretar os coeficientes estimados como elasticidades.

3.5. ESPECIFICAÇÃO ECONOMETRICA

A análise empírica desta dissertação assenta na estimação de modelos gravitacionais através do método dos Mínimos Quadrados Ordinários (OLS). Para cada uma das duas bases construídas (direcional e simétrica) foram estimadas três especificações sucessivas, permitindo avaliar de forma progressiva o contributo das variáveis económicas, institucionais e estruturais para a explicação dos fluxos comerciais bilaterais: um modelo básico, com apenas dimensão económica e distância; um segundo modelo, alargado às variáveis institucionais; e um terceiro modelo, completo, que acrescenta a variável de recursos naturais e o índice de *remoteness*.

3.5.1. MODELO GRAVITACIONAL DIRECIONAL

1. MODELO 1 - ESPECIFICAÇÃO BÁSICA DO MODELO GRAVITACIONAL

$$\ln(X_{ij}) = \beta_0 + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) + \beta_3 \ln(Dist_{ij}) + \varepsilon_{ij}$$

Onde:

X_{ij} representa o fluxo comercial do país exportador i para o país importador j ;

GDP_i corresponde ao PIB do país exportador;

GDP_j corresponde ao PIB do país importador;

$Dist_{ij}$ representa a distância bilateral.

2. MODELO 2 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COM VARIÁVEIS INSTITUCIONAIS

$$\ln(X_{ij}) = \beta_0 + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) + \beta_3 \ln(Dist_{ij}) + \beta_4 Comlang_{ij} + \beta_5 Contig_{ij} + \beta_6 ColDep_{ij} + \beta_7 ComCol_{ij} + \beta_8 RTA_{ij} + \varepsilon_{ij}$$

Mantêm-se as variáveis incluídas no Modelo 1, acrescentando-se as seguintes variáveis institucionais:

$Comlang_{ij}$ identifica países que partilham uma língua oficial comum;

$Contig_{ij}$ identifica países que partilham fronteira terrestre;

$ColDep_{ij}$ identifica países que mantiveram uma relação colonial;

$ComCol_{ij}$ identifica países com um colonizador comum;

RTA_{ij} identifica a existência de um acordo comercial regional entre o par de países.

3. MODELO 3 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COMPLETA

$$\ln(X_{ij}) = \beta_0 + \beta_1 \ln(GDP_i) + \beta_2 \ln(GDP_j) + \beta_3 \ln(Dist_{ij}) + \beta_4 Comlang_{ij} + \beta_5 Contig_{ij} + \beta_6 ColDep_{ij} + \beta_7 ComCol_{ij} + \beta_8 RTA_{ij} + \beta_9 Resource_i + \beta_{10} \ln(Remoteness_i) + \beta_{11} \ln(Remoteness_j) + \varepsilon_{ij}$$

Mantêm-se as variáveis incluídas no Modelo 2, acrescentando-se as seguintes variáveis:

$Resource_i$ identifica países exportadores de recursos naturais;

$Remoteness_i$ e $Remoteness_j$ representam o índice de resistência multilateral ao comércio, utilizado como proxy da posição relativa de cada país face aos restantes parceiros comerciais.

3.5.2. MODELO GRAVITACIONAL SIMÉTRICO

1. MODELO 1 - ESPECIFICAÇÃO BÁSICA DO MODELO GRAVITACIONAL

$$\ln(Trade) = \beta_0 + \beta_1 \ln(GDP_{pair}) + \beta_2 \ln(Dist_{ij}) + \varepsilon_{pair}$$

Mantêm-se as variáveis incluídas no Modelo Direcional, e $pair$ representa o par bilateral formado pelos países i e j .

2. MODELO 2 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COM VARIÁVEIS INSTITUCIONAIS

$$\begin{aligned} \ln(Trade) = & \beta_0 + \beta_1 \ln(GDP_{pair}) + \beta_2 \ln(Dist_{ij}) + \beta_3 Comlang_{ij} + \beta_4 Contig_{ij} \\ & + \beta_5 ColDep_{ij} + \beta_6 ComCol_{ij} + \beta_7 RTA_{ij} + \varepsilon_{pair} \end{aligned}$$

Mantêm-se as variáveis incluídas no Modelo Direcional, e $pair$ representa o par bilateral formado pelos países i e j .

3. MODELO 3 - ESPECIFICAÇÃO DO MODELO GRAVITACIONAL COMPLETA

$$\begin{aligned} \ln(Trade) = & \beta_0 + \beta_1 \ln(GDP_{pair}) + \beta_2 \ln(Dist_{ij}) + \beta_3 Comlang_{ij} + \beta_4 Contig_{ij} \\ & + \beta_5 ColDep_{ij} + \beta_6 ComCol_{ij} + \beta_7 RTA_{ij} + \beta_8 \ln(Remoteness_{pair}) + \varepsilon_{pair} \end{aligned}$$

Mantêm-se as variáveis incluídas no Modelo Direcional, e $pair$ representa o par bilateral formado pelos países i e j .

As diferentes especificações foram estimadas de forma sequencial, permitindo avaliar o contributo incremental de cada grupo de variáveis para a explicação dos fluxos comerciais internacionais. A estimação progressiva dos modelos facilita ainda a comparação dos coeficientes e da capacidade explicativa de cada especificação, tanto na abordagem direcional como na abordagem simétrica.

4. RESULTADOS EMPÍRICOS

Este capítulo apresenta os resultados da estimação dos modelos gravitacionais descritos no capítulo 3. A análise inicia-se com a caracterização estatística das variáveis (4.1) e com a análise de correlações entre as principais variáveis explicativas (4.2). Seguem-se os resultados do modelo direcional (4.3) e do modelo simétrico (4.4), terminando com uma síntese comparativa entre as duas especificações (4.5).

4.1. ESTATÍSTICA DESCRITIVA

Tabela 4.1. - estatística descritiva

Variável	Média	Desvio padrão	Mínimo	Máximo
Logaritmo do comércio do exportador para o importador	13,959	2,732	1,426	19,872
Logaritmo do PIB do exportador	20,546	1,500	16,549	23,785
Logaritmo do PIB do importador	20,547	1,505	16,549	23,785
Logaritmo da distância bilateral	8,699	0,814	5,159	9,885
Logaritmo do <i>remoteness</i> do exportador	8,985	0,231	8,690	9,446
Logaritmo do <i>remoteness</i> do importador	8,986	0,231	8,690	9,446

A Tabela 4.1 apresenta a estatística descritiva das principais variáveis do modelo direcional, numa amostra de 861 observações, todas expressas em logaritmo natural.

O comércio de exportador para importador apresenta uma dispersão elevada (desvio-padrão de 2,732, intervalo entre 1,426 e 19,872), evidenciando a coexistência de relações comerciais muito intensas com fluxos relativamente reduzidos entre os pares de países da amostra.

Quanto à dimensão económica, o PIB do exportador e o do importador apresentam médias praticamente idênticas (20,546 e 20,547, respetivamente), o que era expectável, já que ambos os conjuntos correspondem aos mesmos 30 países, apenas em papéis diferentes. O desvio-padrão de aproximadamente 1,5 confirma diferenças consideráveis na dimensão económica dos países da amostra, um resultado coerente com a heterogeneidade procurada na sua seleção.

A distância bilateral tem uma média de 8,699 e desvio-padrão de 0,814, refletindo a diversidade geográfica dos pares considerados, desde países vizinhos a países separados por grandes distâncias.

Por fim, os índices de *remoteness* do exportador e do importador apresentam estatísticas praticamente idênticas entre si (médias próximas de 8,99) e uma dispersão bastante reduzida (0,231), sensivelmente inferior à do comércio bilateral ou da dimensão económica. Isto é consistente com a natureza mais estrutural desta variável, que capta a posição relativa de um país face ao conjunto dos parceiros comerciais. Esta característica tende a variar menos entre países do que o volume de comércio ou o PIB.

4.2. MATRIZ DE CORRELAÇÕES

Tabela 4.2. - matriz de correlações de Pearson

Variável	Log. do comércio bilateral	Log. do PIB do país exportador	Log. do PIB do país importador	Log. da distância bilateral	Log. do <i>remoteness</i> do país exportador	Log. do <i>remoteness</i> do país importador
Log. do comércio exp - imp	1,000					

Log. do PIB do país exportador	0,551 (0,000)	1,000				
Log. do PIB do país importador	0,472 (0,000)	-0,039 (0,253)	1,000			
Log. da distância bilateral	-0,197 (0,000)	0,004 (0,915)	0,002 (0,958)	1,000		
Log. do <i>remoteness</i> do país exportador	-0,168 (0,000)	-0,243 (0,000)	0,009 (0,804)	0,353 (0,000)	1,000	
Log. do <i>remoteness</i> do país importador	-0,189 (0,000)	0,006 (0,868)	-0,247 (0,000)	0,352 (0,000)	-0,034 (0,315)	1,000

A Tabela 4.2 apresenta a matriz de correlações de Pearson entre as principais variáveis contínuas do modelo, com o respectivo nível de significância. As correlações mais relevantes são estatisticamente significativas ($p < 0,001$).

O comércio de exportador para importador correlaciona-se positivamente com o PIB do exportador (0,551) e do importador (0,472), em linha com a teoria gravitacional: economias maiores têm mais capacidade produtiva e maior capacidade de absorção da procura, gerando fluxos comerciais mais elevados. Em sentido oposto, a distância bilateral correlaciona-se negativamente com o comércio (-0,197); embora a relação seja moderada, o sinal confirma o papel dos custos de transporte e transação na intensidade do comércio internacional. Os índices de *remoteness* apresentam igualmente correlação negativa com o comércio (-0,168 para o exportador, -0,189 para o importador), sugerindo que países mais periféricos face aos principais

mercados tendem a comerciar menos, resultado consistente com o conceito de resistência multilateral de Anderson e van Wincoop (2003).

As correlações entre variáveis explicativas apresentam, de um modo geral, magnitudes reduzidas, não sugerindo à partida problemas de multicolinearidade, evidência confirmada à frente através dos fatores de inflação da variância (VIF), no capítulo 5.

4.3. MODELO GRAVITACIONAL DIRECIONAL

Tabela 4.3. - resultados das 3 regressões do modelo gravitacional direcional

Variável	Modelo 1	Modelo 2	Modelo 3
Logaritmo do PIB do país exportador	1,039*** (0,043)	1,023*** (0,042)	0,986*** (0,041)
Logaritmo do PIB do país importador	0,898*** (0,040)	0,883*** (0,038)	0,895*** (0,038)
Logaritmo da distância bilateral	-0,671*** (0,101)	-0,446*** (0,091)	-0,584*** (0,111)
Língua oficial comum	-	-0,022 (0,199)	-0,034 (0,195)
Fronteira comum	-	0,475 (0,464)	0,252 (0,458)
Relação colonial	-	0,604*** (0,229)	0,617*** (0,228)
Colonizador comum	-	0,412 (0,493)	0,328 (0,502)
Acordo comercial regional (RTA)	-	0,874*** (0,104)	0,809*** (0,108)

Exportador de recursos naturais	-	-	-0,516*** (0,150)
Logaritmo do <i>remoteness</i> do exportador	-	-	0,976*** (0,240)
Logaritmo do <i>remoteness</i> do importador	-	-	0,263 (0,263)
Constante	-20,013*** (1,755)	-21,677*** (1,643)	-30,840*** (3,391)
Observações	861	861	861
R^2	0,587	0,612	0,621

A Tabela 4.3 apresenta os resultados das três especificações do modelo gravitacional direcional, estimadas de forma sequencial. Esta abordagem permite observar não apenas os coeficientes finais, mas também a forma como cada grupo de variáveis absorve ou reforça o efeito das restantes, um aspecto particularmente relevante na relação entre a distância e as variáveis institucionais, como se discute à frente. Todas as estimações utilizam erros-padrão robustos à heterocedasticidade, dada a evidência encontrada nos testes de diagnóstico apresentados no capítulo 5. As três especificações são globalmente significativas ($p < 0,001$), com o coeficiente de determinação a aumentar de 0,587 no Modelo 1 para 0,621 no Modelo 3: confirma que as variáveis institucionais e estruturais acrescentam poder explicativo sem alterar a estrutura fundamental do modelo.

Dimensão económica

O PIB do exportador e o PIB do importador apresentam, nas três especificações, coeficientes positivos e altamente significativos ($p < 0,001$), confirmando uma das previsões centrais da teoria gravitacional: economias maiores comerciam mais, tanto pelo lado da oferta como da procura. No Modelo 3, o coeficiente do PIB do exportador é de 0,986, o que significa

que um aumento de 1% no PIB do país exportador está associado, em média, a um aumento de aproximadamente 0,99% no comércio do país exportador para o país importador, mantendo constantes as restantes variáveis. Para o PIB do importador, o coeficiente de 0,895 implica um efeito ligeiramente inferior: um aumento de 1% no PIB do importador traduz-se num aumento médio de cerca de 0,90% o comércio do país exportador para o país importador.

A diferença entre os dois coeficientes (ainda que pequena, cerca de 9 pontos percentuais) é consistente ao longo das três especificações e sugere que, nesta amostra, a capacidade produtiva do país exportador tem um peso ligeiramente superior ao da capacidade de absorção do país importador. A diferença é suficientemente pequena para não permitir conclusões fortes sobre uma assimetria estrutural entre exportação e importação.

É também relevante notar que ambos os coeficientes se mantêm notavelmente estáveis entre o Modelo 1 e o Modelo 3: o PIB do exportador desce apenas de 1,039 para 0,986, e o do importador varia entre 0,883 e 0,898. Esta estabilidade é, em si, um resultado importante pois mostra que o efeito da dimensão económica sobre o comércio não é uma correlação falsa, que desaparece quando se controla por fatores institucionais, coloniais ou estruturais, mas sim uma relação robusta que resiste à introdução de oito variáveis adicionais.

Distância geográfica

A distância bilateral apresenta, nas três especificações, um coeficiente negativo e estatisticamente significativo ($p < 0,01$), confirmando que o afastamento geográfico entre dois países reduz a intensidade do seu comércio. No Modelo 1, o coeficiente de -0,671 indica que um aumento de 1% na distância está associado a uma redução média de cerca de 0,67% no comércio do país exportador para o país importador: este efeito é superior ao de qualquer outra variável do modelo básico.

Após a introdução das variáveis institucionais no Modelo 2, o coeficiente reduz-se, em valor absoluto, para -0,446. Esta redução sugere que uma parte do que inicialmente parecia ser "efeito da distância" era, na realidade, um efeito de variáveis correlacionadas com a proximidade geográfica: países vizinhos tendem também a partilhar acordos comerciais e, nalguns casos, relações históricas. Quando estas variáveis são explicitamente controladas, deixam de estar "escondidas" dentro do coeficiente da distância, que passa a captar apenas o seu efeito próprio.

No Modelo 3, o coeficiente volta a aumentar em valor absoluto, para -0,584, depois de se controlar adicionalmente pela dependência de recursos naturais e pelo *remoteness*. Este movimento indica que a distância continua a ter um efeito genuíno sobre o comércio, distinto dos custos institucionais e da posição relativa dos países face aos restantes mercados. Ou seja, mesmo depois de se ter em conta todo o conjunto de fatores institucionais e estruturais considerados nesta dissertação, os custos de transporte, de coordenação e de informação associados à distância continuam a dificultar o comércio internacional.

Variáveis institucionais

Ao contrário das variáveis contínuas, as variáveis institucionais são dummies, e a sua interpretação não pode seguir a lógica direta de elasticidade usada até aqui. Os coeficientes não representam variações percentuais diretamente, uma vez que a variável dependente está em logaritmo mas a variável explicativa não. Para converter estes coeficientes em efeitos percentuais interpretáveis, aplica-se a transformação $(e^{\beta} - 1) \times 100$, seguindo a prática padrão na literatura de modelos log-lineares com regressores binários.

Entre as cinco variáveis institucionais incluídas no modelo, apenas duas mantêm significância estatística de forma consistente: a relação colonial e o acordo comercial regional (RTA).

A existência de uma relação colonial passada apresenta um coeficiente de 0,617 no Modelo 3 ($p = 0,007$), o que corresponde a um efeito de $(e^{0,617} - 1) \times 100 \approx 85\%$. Isto significa que, mantendo tudo o resto constante, pares de países que mantiveram uma relação colonial no passado apresentam, em média, um volume de comércio do exportador para o importador cerca de 85% superior ao dos restantes pares. Trata-se de um efeito economicamente muito expressivo, sobretudo tendo em conta que estas relações terminaram há várias décadas. A sua persistência sugere que os laços institucionais, jurídicos, etc., criados durante o período colonial continuam a reduzir os custos de transação entre estes países, seja através de sistemas legais semelhantes, seja através de redes empresariais e diplomáticas historicamente estabelecidas.

O acordo comercial regional (RTA) apresenta um efeito ainda mais forte: o coeficiente de 0,809 ($p < 0,001$) corresponde a um aumento médio de $(e^{0,809} - 1) \times 100 \approx 125\%$ no

comércio do exportador para o importador, entre países que partilham um acordo comercial, face a países sem qualquer acordo. Este é o efeito mais expressivo entre todas as variáveis institucionais estimadas, e está em linha com a literatura, que atribui aos RTA um papel central na eliminação de barreiras tarifárias e não tarifárias. Este mecanismo, ao contrário da relação colonial, opera de forma direta e deliberada sobre os custos do comércio.

Em contraste, a língua oficial comum ($\beta = -0,034$; $p = 0,861$), a fronteira terrestre comum ($\beta = 0,252$; $p = 0,582$) e o colonizador comum ($\beta = 0,328$; $p = 0,513$) não apresentam significância estatística em nenhuma das especificações. Isto não significa necessariamente que estes fatores sejam irrelevantes na prática, mas sim que, nesta amostra específica e uma vez controlados os restantes determinantes, nomeadamente a dimensão económica, a distância e as relações coloniais (que podem já captar parte do efeito de proximidade cultural e geográfica), o seu efeito autónomo não é estatisticamente distinguível de zero.

Recursos naturais e remoteness

O Modelo 3 introduz duas variáveis estruturais adicionais. A variável que identifica países exportadores de recursos naturais apresenta um coeficiente de $-0,516$ ($p = 0,001$), correspondente a um efeito de $(e^{-0,516} - 1) \times 100 \approx -40\%$: mantendo constantes as restantes variáveis, os países classificados como exportadores de recursos naturais apresentam, em média, um volume de comércio do exportador para o importador, cerca de 40% inferior ao dos restantes países da amostra.

Este resultado não significa que estes países exportem pouco em termos absolutos. Muitos deles, como a Arábia Saudita ou a Rússia, têm volumes de exportação elevados em termos globais. O que a variável capta é antes um padrão de concentração: países cuja pauta exportadora depende fortemente de um número reduzido de produtos (petróleo, gás, minérios) tendem a concentrar o seu comércio num número também reduzido de parceiros e mercados de destino, em vez de o distribuir de forma ampla pelos 29 restantes países da amostra. Assim, o coeficiente negativo pode refletir mais uma questão de concentração e falta de diversificação geográfica do que uma fraqueza exportadora genuína. Esta distinção importante deve ser tida em conta na interpretação deste resultado. Esta variável, tal como construída, não permite discriminar de forma direta.

O índice de *remoteness* do exportador apresenta um coeficiente positivo de 0,976 ($p < 0,001$), o que implica que um aumento de 1% no afastamento relativo de um país face aos restantes mercados da amostra está associado a um aumento médio de cerca de 0,98% no seu comércio de exportador para importador. Note-se porém que este coeficiente está a ser estimado controlando pela distância bilateral direta, ou seja, o efeito do *remoteness* aqui captado é distinto do efeito puro da distância entre dois países, representando antes a posição estrutural do exportador no conjunto da rede de comércio da amostra.

Já o *remoteness* do importador não é estatisticamente significativo ($\beta = 0,263$; $p = 0,318$), pelo que não há evidência, nesta amostra, de que a posição relativa do país importador face aos restantes mercados influencie de forma sistemática o volume de comércio que recebe. Este resultado assimétrico entre exportador e importador, significativo num caso, não significativo no outro, é interessante em si mesmo e sugere que a resistência multilateral ao comércio pode operar de forma diferente consoante o país ocupa a posição de origem ou de destino do fluxo, um aspeto que a formulação teórica original de Anderson e van Wincoop (2003) não distinguem explicitamente, mas que a estrutura direcional desta base de dados permite investigar.

Em conjunto, a introdução destas duas variáveis estruturais eleva o R^2 de 0,612 para 0,621, um contributo frágil em termos de poder explicativo agregado, mas que revela determinantes que o modelo gravitacional não capta.

4.4. MODELO GRAVITACIONAL SIMÉTRICO

Tabela 4.4. - resultados das 3 regressões do modelo gravitacional simétrico

Variável	Modelo 1	Modelo 2	Modelo 3
Logaritmo do PIB do par de países	0,920*** (0,038)	0,903*** (0,037)	0,927*** (0,036)
Logaritmo da distância bilateral	-0,592*** (0,148)	-0,366** (0,149)	-0,518*** (0,184)
Língua oficial comum	-	0,022 (0,258)	-0,019 (0,254)

Fronteira comum	-	0,622 (0,643)	0,358 (0,655)
Relação colonial	-	0,528* (0,291)	0,571* (0,291)
Colonizador comum	-	0,279 (0,734)	0,203 (0,735)
Acordo comercial regional (RTA)	-	0,806*** (0,149)	0,859*** (0,147)
Logaritmo do <i>remoteness</i> do par de países	-	-	1,170** (0,476)
Constante	-17,868*** (2,338)	-19,460*** (2,263)	-29,635*** (4,013)
Observações	433	433	433
R^2	0,586	0,611	0,615

A Tabela 4.4 apresenta os resultados das três especificações do modelo gravitacional simétrico, estimadas sobre os 433 pares bilaterais de países que resultam da agregação dos fluxos comerciais em ambos os sentidos. A lógica de estimação sequencial é idêntica à seguida no modelo direcional, o que permite não só analisar os determinantes do comércio bilateral total, mas também avaliar até que ponto os resultados obtidos anteriormente dependem da forma como os fluxos comerciais são representados. Todas as especificações são globalmente significativas ($p < 0,001$), com o R^2 a evoluir de 0,586 no Modelo 1 para 0,615 no Modelo 3, uma trajetória e magnitude muito próximas das observadas no modelo direcional (0,587 a 0,621), o que constitui já, por si só, um primeiro indício de robustez.

Dimensão económica

O PIB conjunto do par de países apresenta, nas três especificações, um coeficiente positivo, estável e altamente significativo ($p < 0,001$). No Modelo 3, o coeficiente de 0,927 indica que um aumento de 1% na dimensão económica conjunta do par, isto é, no produto dos PIB dos dois países, está associado, em média, a um aumento de aproximadamente 0,93% no comércio bilateral total entre eles, mantendo constantes as restantes variáveis.

Este valor situa-se muito próximo da média dos dois coeficientes de PIB estimados no modelo direcional (0,986 para o exportador e 0,895 para o importador): quando dois países comerciam nos dois sentidos, o comércio total do par tende a refletir o efeito combinado da dimensão económica de ambos, de forma semelhante à que se obtém tratando exportador e importador em separado. A principal diferença é que o modelo simétrico já não permite distinguir qual dos dois países contribui mais para essa relação; o coeficiente único de 0,927 é, nesse sentido, uma média do efeito conjunto, e não uma soma direta dos efeitos individuais. Ainda assim, a proximidade dos valores é um resultado que confirma que a forma como a dimensão económica dos países se relaciona com o comércio bilateral não é motivada pela estrutura direcional da base de dados, mas sim uma característica robusta dos dados que persiste independentemente da unidade de análise escolhida.

Distância geográfica

A distância bilateral mantém, nas três especificações, um coeficiente negativo e estatisticamente significativo, replicando o padrão de "quebra e recuperação" já identificado no modelo direcional. No Modelo 1, o coeficiente de -0,592 indica que um aumento de 1% na distância está associado a uma redução média de cerca de 0,59% no comércio bilateral total, um valor semelhante ao obtido no Modelo 1 direcional (-0,671, ainda que ligeiramente inferior).

Após a introdução das variáveis institucionais no Modelo 2, o coeficiente reduz-se substancialmente, em valor absoluto, para -0,366 ($p = 0,015$). Esta é uma queda proporcionalmente ainda mais acentuada do que a observada no modelo direcional, que sugere que, quando o comércio é agregado ao nível do par bilateral, uma parcela ainda maior do efeito aparente da distância está de facto a captar a distribuição geográfica dos acordos comerciais entre os pares da amostra. No Modelo 3, com a introdução do *remoteness*, o coeficiente volta a aumentar em valor absoluto, para -0,518 ($p = 0,005$), recuperando grande parte da magnitude

perdida no Modelo 2. Este comportamento confirma, tal como no modelo direcional, que a distância geográfica mantém um efeito negativo genuíno e economicamente relevante sobre o comércio, distinto do efeito das variáveis institucionais e da posição relativa dos países no conjunto da amostra.

Variáveis institucionais

Segue-se a mesma lógica de interpretação aplicada no capítulo 4.3: por se tratar de variáveis dummy num modelo log-linear, os coeficientes não podem ser lidos como elasticidades diretas, sendo necessário aplicar a transformação $(e^{\beta} - 1) \times 100$ para obter o efeito percentual sobre o comércio.

O acordo comercial regional confirma-se, também nesta especificação, como a variável institucional com o efeito mais forte e mais robusto. No Modelo 3, o coeficiente de 0,859 ($p < 0,001$) corresponde a um efeito de $(e^{0,859} - 1) \times 100 \approx 136\%$: pares de países abrangidos por um acordo comercial regional apresentam, em média, um volume de comércio bilateral total cerca de 136% superior ao dos pares sem qualquer acordo, mantendo tudo o resto constante. Este valor é não só consistente com o resultado direcional (125%), como ligeiramente superior, reforçando a ideia de que a integração económica formal tem um impacto muito substancial e bem identificado sobre o comércio, independentemente da forma como este é representado.

A relação colonial apresenta um coeficiente positivo de 0,571 no Modelo 3, correspondente a um efeito de $(e^{0,571} - 1) \times 100 \approx 77\%$, mas com um nível de significância consideravelmente mais fraco do que no modelo direcional: $p = 0,051$ contra $p = 0,007$. Esta perda de robustez estatística é um dos resultados mais interessantes desta comparação. Uma possível explicação reside na redução do número de observações: ao agregar os fluxos nos dois sentidos, a amostra passa de 861 para 433 pares, o que torna mais difícil distinguir estatisticamente um efeito de magnitude moderada. Outra explicação possível, de natureza mais substantiva, é que a agregação dos dois sentidos do comércio pode diluir uma assimetria genuína: é possível que o efeito das relações coloniais seja mais forte num sentido do fluxo (por exemplo, das antigas colónias para as antigas metrópoles, ou vice-versa) do que no outro, uma nuance que a estrutura direcional consegue captar e que a estrutura simétrica, por construção, elimina ao somar os dois sentidos numa única observação.

As restantes variáveis institucionais, língua oficial comum ($\beta = -0,019$; $p = 0,940$), fronteira terrestre comum ($\beta = 0,358$; $p = 0,585$) e colonizador comum ($\beta = 0,204$; $p = 0,782$), não apresentam significância estatística em nenhuma das especificações, replicando de forma consistente o padrão já observado no modelo direcional. A convergência destes resultados entre as duas especificações reforça a conclusão de que, nesta amostra, estes três fatores não exercem um efeito autónomo sobre o comércio bilateral uma vez controlados os restantes determinantes económicos, institucionais e estruturais do modelo.

Remoteness

O índice de *remoteness* médio do par de países, incluído apenas no Modelo 3, apresenta um coeficiente positivo de 1,170 ($p = 0,014$), superior em magnitude ao do modelo direcional (0,976 para o *remoteness* do exportador). Este coeficiente implica que um aumento de 1% no afastamento médio do par de países face aos restantes mercados da amostra está associado a um aumento médio de aproximadamente 1,17% no comércio bilateral total entre os dois países.

O facto de este efeito não só se manter significativo, como aumentar ligeiramente de magnitude ao passar da especificação direcional para a simétrica sugere que a distância média dos países face ao conjunto das economias da amostra, e não apenas a distância bilateral direta entre dois parceiros específicos, constitui um determinante robusto e não negligenciável do comércio bilateral, resistente tanto à forma como as variáveis são construídas (individual versus par) como à unidade de análise utilizada (fluxo direcional versus par agregado). A interpretação económica mantém-se semelhante à discutida no capítulo 4.3: pares de países mais periféricos face aos restantes países da amostra tendem a compensar esse afastamento estrutural através de relações comerciais relativamente mais intensas com os parceiros disponíveis.

A inclusão desta variável eleva o R^2 de 0,611 para 0,615, um ganho semelhante, em magnitude, ao observado na passagem equivalente do modelo direcional (de 0,612 para 0,621), confirmando que o contributo do *remoteness* para a capacidade explicativa do modelo é modesto, mas consistente entre as duas especificações.

4.5. COMPARAÇÃO ENTRE MODELOS

Os capítulos 4.3 e 4.4 apresentaram os resultados de duas formas distintas de representar o mesmo fenómeno económico: o comércio bilateral entre os 30 países da amostra. Este capítulo pretende responder a uma pergunta mais específica: em que medida as conclusões desta dissertação dependem da unidade de análise escolhida - o fluxo direcional ($i \rightarrow j$) ou o par bilateral agregado?

A resposta, em traços gerais, é que dependem pouco. Isso é, em si, um resultado relevante: as conclusões sobre os determinantes do comércio internacional obtidas nesta dissertação refletem relações económicas suficientemente robustas para persistir sob duas representações distintas dos mesmos fluxos comerciais.

A dimensão económica mantém-se como o determinante mais forte e mais estável do comércio bilateral em ambas as especificações, com elasticidades muito próximas, 0,986 e 0,895 (exportador e importador, no modelo direcional) face a 0,927 (PIB conjunto, no modelo simétrico). A distância geográfica exhibe, nos dois modelos, o mesmo padrão de comportamento ao longo das especificações sucessivas: um efeito negativo forte no modelo básico, uma atenuação parcial quando se introduzem as variáveis institucionais, e uma recuperação quando se controla adicionalmente pelo *remoteness*, padrão que sugere que o mecanismo subjacente (parte do efeito da distância é absorvido por variáveis institucionais correlacionadas com a proximidade geográfica) é genuíno e não um acaso da estrutura direcional dos dados.

Entre as variáveis institucionais, o acordo comercial regional confirma-se como o determinante mais robusto em qualquer das especificações, com efeitos percentuais da mesma ordem de grandeza (+125% no modelo direcional, +136% no simétrico) e significância estatística ao nível de 1% em ambos os casos. Do lado oposto, a língua oficial comum, a fronteira terrestre comum e o colonizador comum mantêm-se sistematicamente não significativos nas duas especificações, um resultado negativo, mas consistente, que reforça a conclusão de que estes três fatores não têm, nesta amostra, um efeito autónomo sobre o comércio quando controlados os restantes determinantes.

Por fim, o índice de *remoteness* mantém sinal positivo e significância estatística em ambas as especificações (0,976 para o exportador no modelo direcional, e 1,170 para o par no

modelo simétrico), confirmando que este resultado não é sensível à forma como a variável é construída ou agregada.

As diferenças que se observam entre as duas especificações resultam quase exclusivamente da forma como a agregação dos fluxos comerciais afeta a construção das variáveis, e não de uma discrepância genuína nos resultados económicos.

A variável de recursos naturais é o exemplo mais claro: foi construída para caracterizar especificamente o país exportador (*Resource_i*), pelo que, uma vez que o modelo simétrico agrega os dois sentidos do comércio num único par, deixa de existir uma forma coerente de a incluir sem perder a sua interpretação económica. A sua ausência do modelo simétrico não é, portanto, um resultado substantivo, mas uma limitação estrutural da agregação por par, limitação esta que só a especificação direcional permite ultrapassar, e que constitui um dos argumentos a favor de manter esta última como especificação principal da dissertação.

O caso da relação colonial é mais interessante, precisamente porque revela uma diferença real de robustez, e não apenas uma impossibilidade de construção. No modelo direcional, esta variável é significativa ao nível de 1% ($p = 0,007$); no modelo simétrico, a significância cai para o limiar entre 5% e 10% ($p = 0,051$). Como discutido no capítulo 4.4, esta perda de robustez pode dever-se à redução do número de observações (de 861 fluxos direcionais para 433 pares) ou a uma possível assimetria no efeito colonial consoante o sentido do fluxo comercial: uma assimetria que a estrutura direcional consegue captar e que a agregação por par, por construção, elimina. Este é, na prática, o único resultado desta dissertação cuja robustez depende, de forma não trivial, da unidade de análise escolhida.

Em termos de ajustamento global, os dois modelos apresentam desempenhos muito semelhantes: o modelo direcional atinge um R^2 máximo de 0,621, o modelo simétrico atinge 0,615. Esta diferença é suficientemente pequena para concluir que a agregação dos fluxos comerciais em pares bilaterais não compromete, de forma relevante, a capacidade do modelo gravitacional para explicar a variação do comércio internacional nesta amostra.

A comparação entre as duas especificações confirma que os principais determinantes do comércio identificados nesta dissertação (a dimensão económica, a distância geográfica, os acordos comerciais regionais e, com menor intensidade, os laços coloniais e a distância média dos países face aos restantes da amostra) não dependem da forma como os fluxos comerciais são representados. O modelo simétrico funciona, assim, como um exercício de robustez

bem-sucedido: as diferenças que surgem entre as duas abordagens são explicáveis pela estrutura dos dados (no caso dos recursos naturais) ou pela perda de potência estatística associada à redução do número de observações (no caso da relação colonial), e não por uma contradição substantiva nos resultados económicos obtidos.

5. TESTES DE VALIDADE

Após a estimação dos modelos gravitacionais, procedeu-se à realização de um conjunto de testes, com o objetivo de avaliar a validade dos pressupostos do modelo de regressão linear e a robustez dos resultados obtidos. Em particular, analisou-se a presença de heterocedasticidade, a correta especificação do modelo e a eventual existência de problemas de multicolinearidade entre as variáveis explicativas.

5.1. TESTE DE HETEROCEDASTICIDADE

Um dos pressupostos fundamentais do modelo de regressão linear clássico consiste na homocedasticidade dos resíduos, isto é, na existência de uma variância constante do termo de erro para todas as observações da amostra. Quando este pressuposto não se verifica, ocorre heterocedasticidade, podendo os erros-padrão estimados por mínimos quadrados ordinários tornar-se inconsistentes e comprometer a validade dos testes de significância estatística.

Tabela 5.1. - resultado do teste da heterocedasticidade

Teste	Estatística	P-value
Breusch-Pagan/Cook-Weisberg	$\chi^2(1)=156,73$	<0,001

Para avaliar a presença de heterocedasticidade foi realizado o teste de Breusch-Pagan/Cook-Weisberg. Neste teste, a hipótese nula estabelece que a variância dos resíduos é constante (homocedasticidade), enquanto a hipótese alternativa admite a existência de heterocedasticidade.

Os resultados obtidos encontram-se apresentados na Tabela 5.1. O teste apresenta um valor de $\chi^2(1)=156,73$, correspondente a um p-value inferior a 0,001. Assim, rejeita-se a hipótese nula de homocedasticidade, concluindo-se que existe evidência estatística da presença de heterocedasticidade nos resíduos do modelo.

Face a este resultado, todas as estimações apresentadas nesta dissertação foram realizadas utilizando erros-padrão robustos à heterocedasticidade (Huber-White), garantindo a validade das inferências estatísticas efetuadas.

5.2. TESTE RESET

A correta especificação funcional constitui outro pressuposto importante dos modelos de regressão. Um modelo incorretamente especificado poderá resultar da omissão de variáveis relevantes, da inclusão de formas funcionais inadequadas ou da existência de relações não lineares entre as variáveis.

Tabela 5.2. - resultado do teste de reset

Teste	Estatística	P-value
Ramsey RESET	$F(3,847)=1,81$	0,143

Para avaliar este aspeto foi aplicado o teste de Ramsey RESET (Regression Specification Error Test). A hipótese nula estabelece que o modelo se encontra corretamente especificado, enquanto a hipótese alternativa sugere a existência de erros de especificação.

Os resultados obtidos indicam um valor de $F(3,847)=1,81$, correspondente a um p-value de 0,143. Dado que este valor é superior ao nível de significância de 5%, não se rejeita a hipótese nula.

Conclui-se, assim, que não existe evidência estatística de erros de especificação ou da omissão de variáveis relevantes, reforçando a adequação da especificação econométrica utilizada ao longo da análise.

5.3. TESTE DE MULTICOLINEARIDADE

A multicolinearidade corresponde à existência de correlações elevadas entre as variáveis explicativas do modelo. Embora não provoque enviesamento dos coeficientes estimados, pode aumentar os erros-padrão e reduzir a precisão das estimativas, dificultando a correta identificação do efeito individual de cada variável.

Tabela 5.3. - resultado do teste de multicolinearidade

Variável	VIF
Logaritmo da distância bilateral	2,06
Fronteira comum	1,51
Logaritmo do <i>remoteness</i> do exportador	1,45
Exportador de recursos naturais	1,43
Logaritmo do <i>remoteness</i> do importador	1,39
Logaritmo do PIB do exportador	1,31
Língua oficial comum	1,29
Acordo comercial regional	1,25
Relação colonial	1,17
Colonizador comum	1,12
Logaritmo do PIB do importador	1,10
VIF médio	1,37

Para avaliar a existência de multicolinearidade foi calculado o Variance Inflation Factor (VIF), indicador que mede o grau em que a variância de cada coeficiente é inflacionada pela

correlação com as restantes variáveis explicativas. De acordo com a literatura, valores de VIF superiores a 5 ou 10 podem indicar problemas relevantes de multicolinearidade.

Os resultados apresentados na Tabela 5.3 evidenciam valores de VIF compreendidos entre 1,10 e 2,06, sendo o valor médio igual a 1,37. Todos os valores observados se encontram significativamente abaixo dos limites habitualmente considerados problemáticos.

Deste modo, conclui-se que não existe evidência de problemas significativos de multicolinearidade entre as variáveis explicativas, permitindo considerar que as inferências estatísticas realizadas ao longo da análise são fiáveis.

6. CONCLUSÕES

O objetivo desta dissertação consistiu em analisar os determinantes dos fluxos comerciais internacionais através da estimação de um modelo gravitacional aplicado a uma amostra de 30 países, utilizando dados referentes ao ano de 2019. Para esse efeito foram estimadas duas especificações distintas, uma direcional e outra simétrica, permitindo comparar a influência dos principais determinantes do comércio consoante a forma de representação dos fluxos comerciais.

Os resultados obtidos confirmam, de forma consistente, as previsões fundamentais da teoria gravitacional. Em ambas as especificações, a dimensão económica revelou uma influência positiva e estatisticamente significativa sobre o comércio internacional, enquanto a distância geográfica apresentou um efeito negativo, evidenciando que economias maiores tendem a comerciar mais e que o aumento da distância continua a representar um fator limitador das trocas internacionais.

A inclusão das variáveis institucionais permitiu igualmente identificar fatores que contribuem para explicar o comércio bilateral para além das características puramente económicas. Em particular, a existência de uma relação colonial passada e de acordos comerciais regionais revelou um impacto positivo e estatisticamente significativo nos fluxos comerciais, sugerindo que fatores históricos e institucionais continuam a desempenhar um papel relevante na integração económica entre países.

No modelo direcional, a introdução das variáveis estruturais revelou resultados adicionais de interesse. A variável representativa dos países exportadores de recursos naturais apresentou um coeficiente negativo, sugerindo que, controlando pelos restantes determinantes do modelo,

estes países tendem a apresentar menores fluxos comerciais bilaterais. Paralelamente, o índice de *remoteness* do país exportador revelou um efeito positivo e estatisticamente significativo, evidenciando que a distância de um país face aos restantes da amostra constitui um elemento adicional na explicação dos padrões de comércio internacional. Pelo contrário, o índice de *remoteness* do país importador não apresentou significância estatística.

A comparação entre as duas especificações constitui um dos principais contributos desta dissertação. Apesar das diferenças na construção das bases de dados, os resultados mostram uma notável estabilidade dos principais determinantes do comércio internacional. As variáveis económicas mantêm o sinal esperado e magnitudes semelhantes, a distância conserva o seu efeito negativo e as variáveis institucionais mais relevantes permanecem estatisticamente significativas. Esta consistência sugere que as conclusões fundamentais do modelo gravitacional são robustas à forma como os fluxos comerciais são representados.

Ainda assim, verificam-se algumas diferenças relevantes entre as duas abordagens. A especificação direcional permite distinguir explicitamente o papel desempenhado pelo país exportador e pelo país importador, possibilitando analisar separadamente características específicas de cada economia, como a disponibilidade de recursos naturais ou diferentes níveis de resistência multilateral ao comércio. Por sua vez, a especificação simétrica simplifica a estrutura da informação ao nível do par bilateral, produzindo resultados muito semelhantes relativamente aos determinantes tradicionais do comércio, mas sacrificando parte dessa informação individual. Assim, a escolha entre ambas as abordagens deverá depender dos objetivos específicos da investigação.

Os testes de validade realizados indicam que os modelos estimados não apresentam problemas relevantes de especificação funcional nem de multicolinearidade. Embora tenha sido detetada heterocedasticidade, este problema foi mitigado através da utilização de erros-padrão robustos, permitindo preservar a validade da inferência estatística.

Em síntese, esta dissertação confirma a elevada capacidade explicativa do modelo gravitacional na análise do comércio internacional e mostra que os seus principais determinantes permanecem consistentes quando os fluxos comerciais são analisados através de especificações direcionais ou simétricas. A comparação entre ambas as abordagens constitui, assim, um contributo adicional para a compreensão das potencialidades e limitações das diferentes formas de operacionalizar empiricamente o modelo gravitacional.

Apesar dos resultados obtidos, importa reconhecer algumas limitações do estudo. Em primeiro lugar, a análise foi realizada apenas para um único ano, não permitindo captar alterações na evolução temporal dos fluxos comerciais. Em segundo lugar, a estimação foi efetuada através de Mínimos Quadrados Ordinários, não recorrendo ao estimador *PPML* atualmente recomendado por parte da literatura para lidar com problemas de heterocedasticidade e fluxos comerciais nulos. Adicionalmente, a variável relativa aos acordos comerciais regionais poderá estar sujeita a problemas de endogeneidade, conforme salientado por Baier e Bergstrand (2007), aspeto que não foi objeto de tratamento específico nesta investigação.

Como linhas de investigação futura, seria interessante alargar a análise a dados em painel, incorporar efeitos fixos de exportador e importador, estimar o modelo através do método *PPML* e comparar os resultados obtidos com os da presente dissertação. Poderia igualmente explorar-se a utilização de outras medidas de resistência multilateral ao comércio e avaliar a robustez dos resultados para diferentes conjuntos de países e períodos temporais.

BIBLIOGRAFIA

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. Periodicals Service Company

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, 69(1), 106–116.

Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *Review of Economics and Statistics*, 67(3), 474–481.
<https://doi.org/10.2307/1925976>

Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1), 170–192.
<https://doi.org/10.1257/000282803321455214>

Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4), 641–658. <https://doi.org/10.1162/rest.88.4.641>

Baier, S. L., & Bergstrand, J. H. (2007). Do free trade agreements actually increase members' international trade? *Journal of International Economics*, 71(1), 72–95.
<https://doi.org/10.1016/j.jinteco.2006.02.005>

Conte, M., Cotterlaz, P., & Mayer, T. (2022). *The CEPII gravity database* (CEPII Working Paper N° 2022-05). CEPII.

ANEXOS

ANEXO A - LISTA DOS PAÍSES DA AMOSTRA

País	Código ISO3
AGO	Angola
ARG	Argentina
BEL	Bélgica
BRA	Brasil
CAN	Canadá
CHL	Chile
CHN	China
DEU	Alemanha
ESP	Espanha
FRA	França
GBR	Reino Unido
IND	Índia
IRN	Irão
ISR	Israel
ITA	Itália
JPN	Japão
KOR	Coreia do Sul

MAR	Marrocos
MEX	México
MOZ	Moçambique
NGA	Nigéria
NLD	Países Baixos
PRT	Portugal
QAT	Catar
RUS	Rússia
SAU	Arábia Saudita
SGP	Singapura
USA	Estados Unidos da América
ARE	Emirados Árabes Unidos
ZAF	África do Sul

Fonte: Elaboração própria com base na seleção efetuada a partir da base de dados CEPIL.

ANEXO B - PRINCIPAIS COMANDOS UTILIZADOS NO STATA

Seleção da amostra:

```
keep if year == 2019
```

```
keep      if      inlist(iso3_o,"PRT","ESP","FRA","DEU","ITA","NLD","BEL","GBR","RUS",  
"USA","CAN","MEX","BRA","ARG","CHL","AGO","MOZ","ZAF","NGA","MAR","CHN",  
"JPN","KOR","IND","SGP","SAU","ARE","QAT","ISR","IRN")
```

```
keep      if      inlist(iso3_d,"PRT","ESP","FRA","DEU","ITA","NLD","BEL","GBR","RUS",  
"USA","CAN","MEX","BRA","ARG","CHL","AGO","MOZ","ZAF","NGA","MAR","CHN",  
"JPN","KOR","IND","SGP","SAU","ARE","QAT","ISR","IRN")
```

```
drop if iso3_o == iso3_d
```

Construção das variáveis:

```
gen ln_trade = ln(trade)
```

```
gen ln_gdp_o = ln(gdp_o)
```

```
gen ln_gdp_d = ln(gdp_d)
```

```
gen ln_dist = ln(dist)
```

```
gen ln_remoteness_i = ln(remoteness_i)
```

```
gen ln_remoteness_j = ln(remoteness_j)
```

Construção do modelo simétrico:

```
gen pair_id = cond(iso3_o < iso3_d, iso3_o + "_" + iso3_d, iso3_d + "_" + iso3_o)
```

```
bysort pair_id: egen trade_sym = total(trade)
```

```
bysort pair_id: keep if _n == 1
```

```
gen gdp_pair = gdp_o * gdp_d
```

```
gen ln_gdp_pair = ln(gdp_pair)
```

```
gen remoteness_pair = (remoteness_i + remoteness_j)/2
```

```
gen ln_remoteness_pair = ln(remoteness_pair)
```

```
gen ln_trade_sym = ln(trade_sym)
```

Estimação do modelo gravitacional direcional:

Modelo 1

```
reg ln_trade ln_gdp_o ln_gdp_d ln_dist, robust
```

Modelo 2

```
reg ln_trade ln_gdp_o ln_gdp_d ln_dist comlang_off contig col_dep_ever comcol rta, robust
```

Modelo 3

```
reg ln_trade ln_gdp_o ln_gdp_d ln_dist comlang_off contig col_dep_ever comcol rta resource_i  
ln_remoteness_i ln_remoteness_j, robust
```

Estimação do modelo gravitacional simétrico:

Modelo 1

```
reg ln_trade_sym ln_gdp_pair ln_dist, robust
```

Modelo 2

```
reg ln_trade_sym ln_gdp_pair ln_dist comlang_off contig col_dep_ever comcol rta, robust
```

Modelo 3

```
reg ln_trade_sym ln_gdp_pair ln_dist comlang_off contig col_dep_ever comcol rta ln_remoteness_pair,  
robust
```

Estatísticas descritivas e correlações:

```
summarize
```

```
pwcorr ln_trade ln_gdp_o ln_gdp_d ln_dist ln_remoteness_i ln_remoteness_j, sig
```

Testes de diagnóstico:

estat hettest

estat ovtest

vif

ANEXO C - OUTPUT DO CÁLCULO DE *REMOTENESS*: LISTA DE PAÍSES POR NÍVEL
DE *REMOTENESS* ASCENDENTE

	iso3_o	remote~i
1.	NLD	5945.879
2.	BEL	5953.447
3.	GBR	6138.23
4.	FRA	6211.466
5.	RUS	6241.078
6.	DEU	6248.641
7.	CAN	6480.537
8.	ESP	6531.79
9.	ITA	6532.439
10.	PRT	6550.044
11.	MAR	6844.156
12.	IRN	7154.13
13.	ISR	7198.637
14.	KOR	7539.251
15.	QAT	7747.426

16.	ARE	7814.895
17.	SAU	7834.16
18.	IND	8024.958
19.	JPN	8278.086
20.	MEX	8683.136
21.	NGA	8736.14
22.	USA	8964.779
23.	CHN	9483.946
24.	AGO	9778.348
25.	SGP	10102.23
26.	MOZ	11049.6
27.	ZAF	11070.23
28.	BRA	11872.99
29.	CHL	12528.87
30.	ARG	12662.57

ANEXO D - OUTPUTS COMPLETOS DAS ESTIMAÇÕES E DOS TESTES DE VALIDADE

4.1. ESTATÍSTICA DESCRITIVA

```
. summarize trade gdp_o gdp_d dist remoteness_i remoteness_j
```

Variable	Obs	Mean	Std. dev.	Min	Max
trade	861	1.01e+07	3.04e+07	4.162	4.27e+08
gdp_o	861	2.37e+09	4.43e+09	1.54e+07	2.14e+10
gdp_d	861	2.38e+09	4.43e+09	1.54e+07	2.14e+10
dist	861	7562.935	4130.687	174	19629
remoteness_i	861	8208.3	2003.74	5945.879	12662.57
remoteness_j	861	8215.306	2004.04	5945.879	12662.57

```
. summarize ln_trade ln_gdp_o ln_gdp_d ln_dist ln_remoteness_i ln_remoteness_j
```

Variable	Obs	Mean	Std. dev.	Min	Max
ln_trade	861	13.95925	2.732162	1.425996	19.8723
ln_gdp_o	861	20.54629	1.50001	16.54923	23.78537
ln_gdp_d	861	20.54664	1.504725	16.54923	23.78537
ln_dist	861	8.699498	.8138588	5.159055	9.884764
ln_remoten~i	861	8.985364	.2306174	8.690454	9.446406
ln_remoten~j	861	8.986221	.230676	8.690454	9.446406

```
.
```

4.2. MATRIZ DE CORRELAÇÕES

```
. pwcorr ln_trade ln_gdp_o ln_gdp_d ln_dist ln_remoteness_i ln_remoteness_j, sig
```

	ln_trade	ln_gdp_o	ln_gdp_d	ln_dist	ln_rem~i	ln_rem~j
ln_trade	1.0000					
ln_gdp_o	0.5505 0.0000	1.0000				
ln_gdp_d	0.4722 0.0000	-0.0390 0.2533	1.0000			
ln_dist	-0.1970 0.0000	0.0037 0.9148	0.0018 0.9582	1.0000		
ln_remoten~i	-0.1675 0.0000	-0.2427 0.0000	0.0085 0.8041	0.3525 0.0000	1.0000	
ln_remoten~j	-0.1888 0.0000	0.0057 0.8682	-0.2471 0.0000	0.3522 0.0000	-0.0343 0.3152	1.0000

4.3. MODELO GRAVITACIONAL DIRECIONAL

MODELO 1:

```
. reg ln_trade ln_gdp_o ln_gdp_d ln_dist, robust
```

Linear regression	Number of obs	=	861
	F(3, 857)	=	378.91
	Prob > F	=	0.0000
	R-squared	=	0.5871
	Root MSE	=	1.7586

ln_trade	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_gdp_o	1.039143	.0428546	24.25	0.000	.9550305	1.123255
ln_gdp_d	.8984673	.0387645	23.18	0.000	.8223829	.9745518
ln_dist	-.6711818	.1013506	-6.62	0.000	-.8701063	-.4722572
_cons	-20.01283	1.754653	-11.41	0.000	-23.45675	-16.56891

MODELO 2:

```
. reg ln_trade ln_gdp_o ln_gdp_d ln_dist comlang_off contig col_dep_ever comcol rta, robust
```

Linear regression	Number of obs	=	861
	F(8, 852)	=	166.03
	Prob > F	=	0.0000
	R-squared	=	0.6122
	Root MSE	=	1.7093

ln_trade	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_gdp_o	1.023288	.0415318	24.64	0.000	.9417714	1.104805
ln_gdp_d	.8829667	.0383023	23.05	0.000	.8077887	.9581447
ln_dist	-.4458898	.091377	-4.88	0.000	-.6252402	-.2665394
comlang_off	-.0222691	.1994475	-0.11	0.911	-.4137352	.369197
contig	.4754479	.4635126	1.03	0.305	-.4343124	1.385208
col_dep_ever	.6040199	.229139	2.64	0.009	.1542768	1.053763
comcol	.4119525	.4927534	0.84	0.403	-.5552003	1.379105
rta	.8738228	.1044294	8.37	0.000	.6688538	1.078792
_cons	-21.67691	1.642908	-13.19	0.000	-24.90153	-18.45229

MODELO 3:

```
. reg ln_trade ln_gdp_o ln_gdp_d ln_dist comlang_off contig col_dep_ever comcol rta resource_i ln_
> remoteness_i ln_remoteness_j, robust
```

```
Linear regression               Number of obs   =       861
                               F(11, 849)      =      126.77
                               Prob > F        =      0.0000
                               R-squared        =      0.6210
                               Root MSE     =      1.6928
```

ln_trade	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_gdp_o	.9856087	.0410607	24.00	0.000	.9050164	1.066201
ln_gdp_d	.894938	.0377722	23.69	0.000	.8208001	.9690759
ln_dist	-.5836821	.1107896	-5.27	0.000	-.8011358	-.3662284
comlang_off	-.0341263	.194623	-0.18	0.861	-.416125	.3478723
contig	.2524298	.4581346	0.55	0.582	-.6467795	1.151639
col_dep_ever	.6171005	.2279457	2.71	0.007	.1696973	1.064504
comcol	.3280261	.5015243	0.65	0.513	-.6563467	1.312399
rta	.8094825	.1080339	7.49	0.000	.5974377	1.021527
resource_i	-.5159892	.1499372	-3.44	0.001	-.8102803	-.2216981
ln_remoteness_i	.975674	.2402988	4.06	0.000	.5040246	1.447323
ln_remoteness_j	.2629794	.2633616	1.00	0.318	-.2539368	.7798956
_cons	-30.84034	3.391442	-9.09	0.000	-37.49693	-24.18374

4.4. MODELO GRAVITACIONAL SIMÉTRICO

MODELO 1

```
. reg ln_trade_sym ln_gdp_pair ln_dist, robust
```

```
Linear regression               Number of obs   =       433
                               F(2, 430)      =      379.87
                               Prob > F        =      0.0000
                               R-squared        =      0.5863
                               Root MSE     =      1.6647
```

ln_trade_sym	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_gdp_pair	.9203381	.0372724	24.69	0.000	.8470793	.9935968
ln_dist	-.5923841	.1476837	-4.01	0.000	-.8826558	-.3021124
_cons	-17.86759	2.337717	-7.64	0.000	-22.46237	-13.27282

MODELO 2

```
. reg ln_trade_sym ln_gdp_pair ln_dist comlang_off contig col_dep_ever comcol rta, robust
```

Linear regression	Number of obs	=	433
	F(7, 425)	=	132.52
	Prob > F	=	0.0000
	R-squared	=	0.6114
	Root MSE	=	1.6229

ln_trade_sym	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_gdp_pair	.9031123	.0368818	24.49	0.000	.830619	.9756057
ln_dist	-.3656797	.1492215	-2.45	0.015	-.6589838	-.0723757
comlang_off	.0223803	.2577971	0.09	0.931	-.4843358	.5290964
contig	.622211	.6430573	0.97	0.334	-.6417576	1.88618
col_dep_ever	.5284688	.2910008	1.82	0.070	-.0435111	1.100449
comcol	.2785107	.7336072	0.38	0.704	-1.163439	1.720461
rta	.8063812	.1491723	5.41	0.000	.5131739	1.099588
_cons	-19.45989	2.262582	-8.60	0.000	-23.90713	-15.01264

MODELO 3

```
. reg ln_trade_sym ln_gdp_pair ln_dist comlang_off contig col_dep_ever comcol rta ln_remoteness_pa  
> ir, robust
```

Linear regression	Number of obs	=	433
	F(8, 424)	=	116.37
	Prob > F	=	0.0000
	R-squared	=	0.6147
	Root MSE	=	1.618

ln_trade_sym	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
ln_gdp_pair	.9266491	.0364372	25.43	0.000	.8550291	.9982692
ln_dist	-.5177632	.1841227	-2.81	0.005	-.8796701	-.1558563
comlang_off	-.019148	.2544405	-0.08	0.940	-.5192697	.4809738
contig	.35823	.6553406	0.55	0.585	-.9298908	1.646351
col_dep_ever	.5705357	.2909304	1.96	0.051	-.0013097	1.142381
comcol	.2034609	.7347312	0.28	0.782	-1.240708	1.64763
rta	.8593355	.1467606	5.86	0.000	.5708666	1.147804
ln_remoteness_pair	1.170169	.4762169	2.46	0.014	.234129	2.106209
_cons	-29.63538	4.01301	-7.38	0.000	-37.52325	-21.74751

5. TESTES DE DIAGNÓSTICO

TESTE DE HETEROCEDASTICIDADE

```
. estat hettest

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of ln_trade

H0: Constant variance

      chi2(1) = 156.73
Prob > chi2 = 0.0000

.
```

TESTE RESET

```
. estat ovtest

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of ln_trade

H0: Model has no omitted variables

F(3, 847) = 1.81
Prob > F = 0.1431
```

TESTE DE MULTICOLINEARIDADE

```
. vif
```

Variable	VIF	1/VIF
ln_dist	2.06	0.485907
contig	1.51	0.661026
ln_remoten~i	1.45	0.688179
resource_i	1.43	0.698281
ln_remoten~j	1.39	0.721409
ln_gdp_o	1.31	0.761475
comlang_off	1.29	0.776330
rta	1.25	0.801743
col_dep_ever	1.17	0.852703
comcol	1.12	0.891338
ln_gdp_d	1.10	0.908420
Mean VIF	1.37	