Empirical Project: The Impact of EU Accession on Eastern European Exports
Course: International Trade Economics

Group: Mardan Kydyrbek s2504509, Giovanni Paoli s2504441

(1.1) Research Question and Data Sourcing

Research Question: The objective of this empirical project is to analyze the impact of European Union (EU) accession on the export performance of Eastern European countries. Specifically, we investigate whether becoming a member of the EU has a positive causal effect on bilateral trade flows to other EU members.

Data Sources: To answer this question, we constructed a panel dataset by combining two primary sources from the CEPII (Centre d'Études Prospectives et d'Informations Internationales) database:

1. BACI Database: This provides bilateral trade flows (export values), which we aggregated to the country level.
2. Gravity Database: This provides country-specific macroeconomic variables (GDP) and bilateral fixed factors (distance, contiguity, common language).

We filtered the data to focus on specific Eastern European exporters (e.g., Poland, Hungary, Romania) and EU importers over the period 1990–2020.

(1.2) Descriptive Statistics

Using Stata, we analyzed the properties of our sample to understand the data structure before estimation.

a) Number of observations: The final dataset used for the regression contains 6,064 observations.

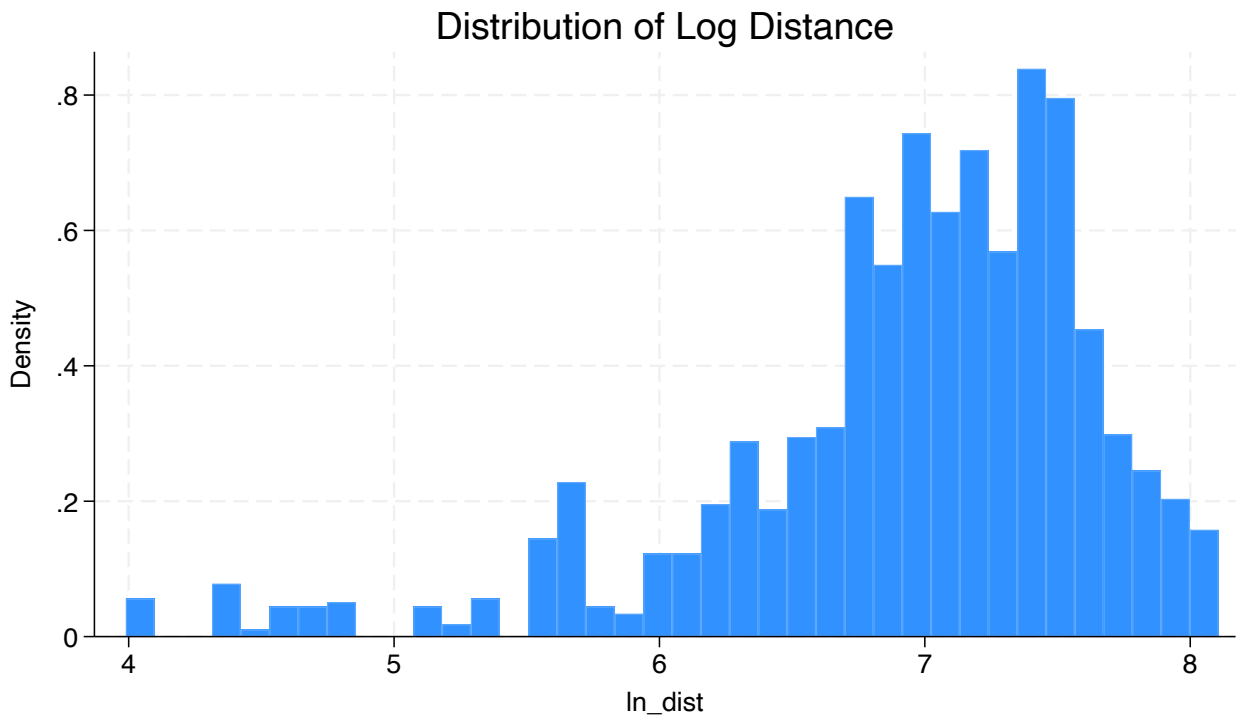
b) Measures of Central Tendency:

- Ln(Trade): The mean of our dependent variable (log of trade flows) is 12.54.
- Post-Accession Dummy: The mean is 0.65, indicating that approximately 65% of the observations in our sample occur in the years after the exporter joined the EU.

c) Measures of Variability:

- Ln(Trade): The standard deviation is 2.03, showing significant variation in trade volumes across country pairs and time.
- Ln(Distance): The standard deviation is 0.74, reflecting the geographic spread of the trading partners.

d) Distribution of Data: The graph below displays the distribution of the logarithmic distance between trading partners. The distribution clusters around a mean of approximately 6.9, indicating that most trade partners in our sample are located at a medium range within the European continent.



Graph 1. Distribution of Log Distance

In the graph 1 above, we can see that this histogram displays the density and spread of the geographic distance (in logs) between the Eastern European exporters and their EU trading partners. The data is not perfectly symmetrical; it is *negatively skewed* (skewed to the left), with a high concentration of observations between the values of 7 and 7.5 on the log scale.

As we know, in international trade, the sample distribution of distance tells us about the "geographic reach" of the trade agreements. The large cluster around ln_dist ≈ 7.3 represents the "natural" trading neighborhood. This corresponds to the distance between major Eastern European hubs (like Warsaw or Prague) and core EU economies (like Germany, France, or Italy).

The Left Tail means, that the smaller bars on the left (lower values of ln_dist) represent trade with immediate neighbors (contiguity).

While The Right Tail means, that the bars on the far right represent trade with more peripheral EU members (e.g., Portugal, Ireland, or Cyprus).

What it means for our Regression:

- For a Gravity Model to work well, we need significant variation in distance to accurately estimate its negative impact on trade. This graph confirms that our dataset has a wide enough spread—from very close neighbors to distant partners—to provide a robust estimate for the β3 coefficient.
- While the distribution has a long left tail, the central mass of the data follows a relatively bell-shaped pattern in logs. This is ideal for OLS estimation, as it reduces the influence of extreme outliers and helps satisfy the normality assumptions of the error term.

(1.3) Variables and Methodology

To answer the research question, we utilize a Structural Gravity Model. This econometric framework explains bilateral trade flows based on the economic size of the trading partners and the friction (costs) between them.

- Dependent Variable (y): ln_trade (). The natural logarithm of real bilateral export flows from Eastern European country i to EU member j at time t.
- Independent Variables (x's):
    1. ln_gdp_o: Log of Exporter's GDP (Economic Mass).
    2. ln_gdp_d: Log of Importer's GDP (Economic Mass).
    3. ln_dist: Log of Bilateral Distance (Trade Costs).
    4. contig: Contiguity (Dummy variable = 1 if countries share a border).
    5. post_accession: Variable of Interest (Dummy variable = 1 if year accession year).

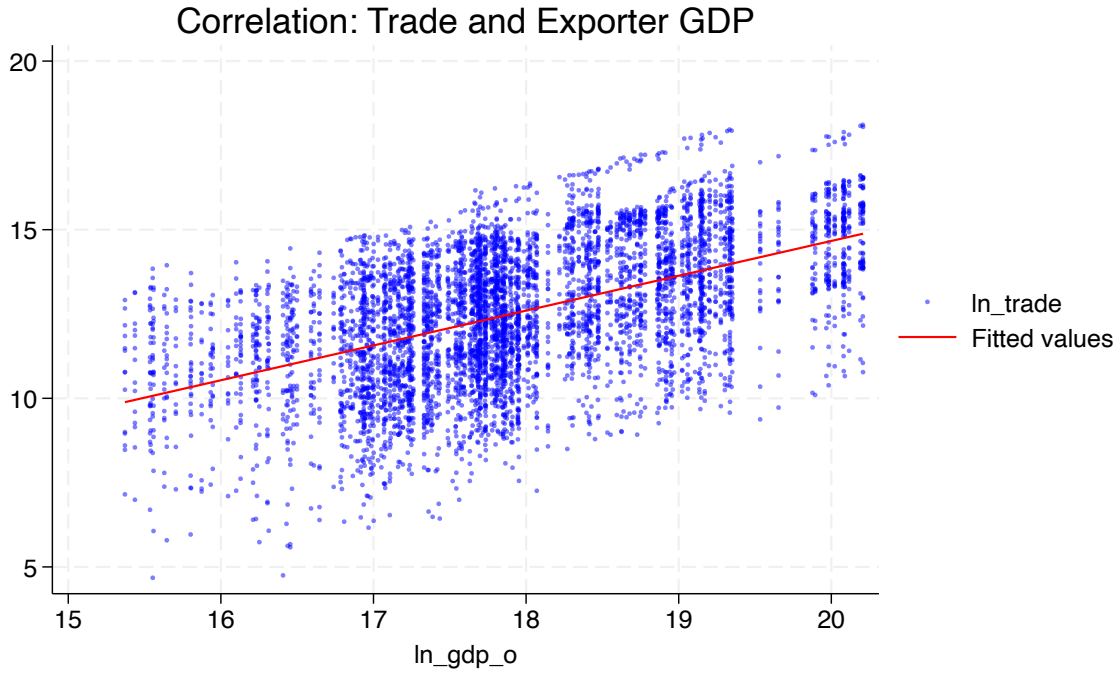
(1.4) Variable Justification and Expected Signs

a) Justification (Literature):

While classical theories like Ricardo or Heckscher-Ohlin explain patterns of specialization based on technology or factor endowments, our empirical strategy relies on the **Gravity Model framework** to estimate aggregate trade flows based on economic mass and trade costs.

- GDP (Origin and Destination): Following the foundational work of Tinbergen (1962) and the theoretical derivation by Anderson and van Wincoop (2003), trade is proportional to the economic mass of the trading partners. Larger economies produce more varieties of goods (supply) and have higher purchasing power (demand).
- Distance: Distance serves as a proxy for transport costs, time costs, and information barriers. The "gravity" concept dictates that trade diminishes as distance increases.
- Post-Accession: Based on Viner's (1950) theory of Customs Unions, joining a union leads to "Trade Creation" by removing tariff and non-tariff barriers. We include this to capture the specific effect of EU integration.

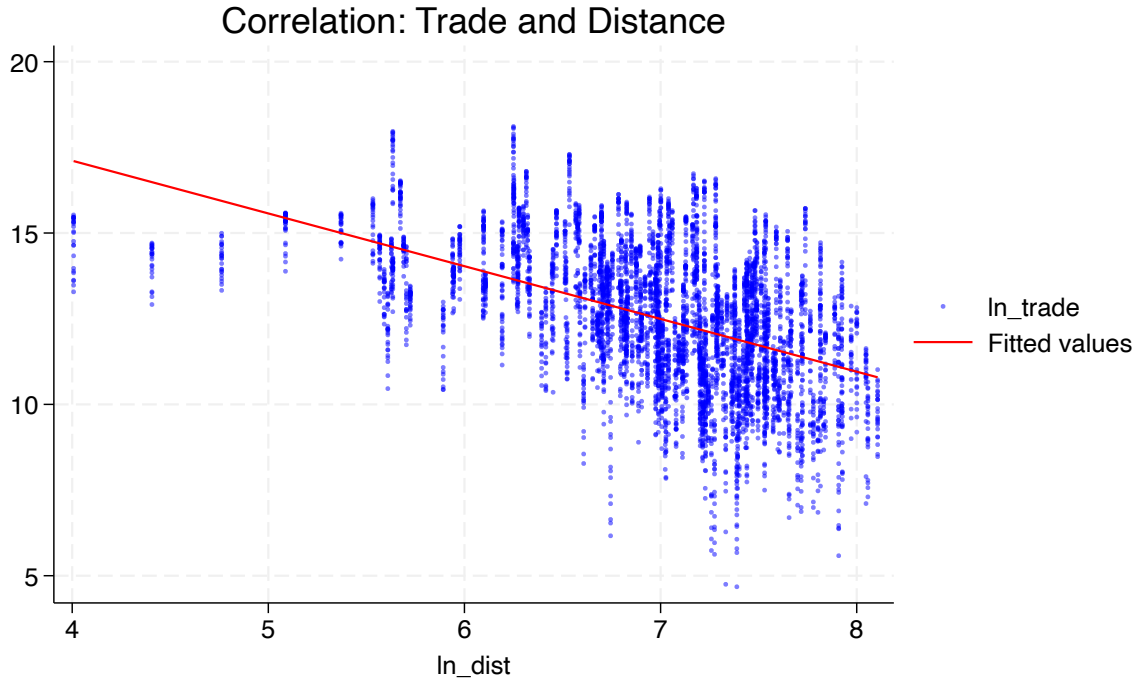
b) Graphical Relationships: Below, we plot the relationship between the dependent variable and our key independent variables.



Graph 2. Relationship between Trade and Exporter GDP

In the graph 2 above, we can see that, the scatter plots for both ln(gdp_o) and ln(gdp_d) display a strong, positive linear trend. The data points are concentrated along an upward-sloping path. According to the Gravity Model, larger economies produce more goods to export (supply side) and have higher income to spend on imports (demand side). Which means that, this confirms that

"economic mass" is a primary driver of trade in our sample. Larger Eastern European economies (like Poland) and larger EU partners (like Germany) naturally trade more with each other.



Graph 3. Relationship between Trade and Distance

In the graph 3, the scatter plot for ln(dist) shows a clear downward-sloping relationship with trade. If explain it from economic point of view, distance is a proxy for variable trade costs, including transportation, fuel, time, and even "cultural distance." Higher distance increases the cost of delivered goods, making them less competitive. Which means, the steep negative slope indicates that trade between Eastern Europe and the EU is highly sensitive to distance. Countries closer to the EU core (like the Czech Republic or Hungary) exhibit much higher trade volumes than those further away.

c) Before conducting the regression, we establish our expectations for the coefficients based on the core tenets of International Trade theory and existing empirical research.

1. Exporter and Importer GDP ($\beta1>0$, $\beta2>0$). Expected Sign - Positive.

According to the Foundations of the Gravity Model (Tinbergen, 1962), trade is proportional to the economic size of the trading partners. A higher GDP in the exporting country (i) signifies a greater productive capacity and a wider variety of goods available for export. Conversely, a higher GDP in the importing country (j) represents higher aggregate demand and purchasing power. Empirical studies, such as those by Helpman and Krugman (1985), justify this positive sign through the lens of increasing returns to scale and consumer preference for variety.

2. Bilateral Distance ($\beta3<0$). Expected Sign - Negative.

Distance is the classic proxy for trade frictions. We expect a negative sign because distance increases transport costs, insurance, and time-sensitive delivery risks. Anderson and van Wincoop (2003) demonstrate that bilateral trade is heavily restricted by these "iceberg" trade costs. In the context of Eastern Europe, further distance from core EU markets (like the "Blue Banana" industrial corridor) is expected to significantly diminish trade volumes.

3. Contiguity / Common Border ($\beta4>0$). Expected Sign - Positive.

Sharing a land border reduces transport complexity and often implies shared infrastructure (railways, pipelines). Eichengreen and Irwin (1998) argue that neighboring countries often have

long-standing historical trade networks and integrated supply chains. Therefore, we expect a positive coefficient for contig, as it significantly lowers the "border effect."

4. Common Official Language (β5>0). Expected Sign - Positive.

A common language reduces information costs and communication barriers. Melitz (2008) found that linguistic ties significantly facilitate trade by lowering the costs of negotiating contracts and understanding foreign legal systems. Although this variable might be omitted in our specific sample due to collinearity with geography, the theoretical expectation remains strongly positive.

5. EU Accession / post_accession (β6>0). Expected Sign - Positive.

This is our key variable of interest. We expect a positive sign based on the Customs Union Theory (Viner, 1950), which suggests that the removal of tariffs and non-tariff barriers leads to Trade Creation. Furthermore, Baier and Bergstrand (2007) have shown that Deep Integration Agreements (like the EU Single Market) significantly increase bilateral trade by harmonizing regulations and providing institutional stability, which encourages long-term export investments from Eastern European firms.

(1.5) Econometric Model

We estimate the following log-linearized Gravity equation:

$$\ln(Trade_{ijt}) = \beta_0 + \beta_1 \ln(GDP\_o_{it}) + \beta_2 \ln(GDP\_d_{jt}) + \beta_3 \ln(Dist_{ij}) + \beta_4 Contig_{ij} + \beta_5 Comlang\_off_{ij} + \beta_6 PostAccession_{ijt} + \varepsilon_{ijt}$$

Where represents the error term.

(1.6) Estimation

The model was estimated using the Ordinary Least Squares (OLS) method in Stata.

(1.7) Results and Discussion

Table 1. Gravity Model Estimation Results

Independent Variable	Coefficient (β)	Std. Error	t-statistic	P-value
ln_gdp_o (Exporter GDP)	0.9148***	0.0104	87.73	0.000
ln_gdp_d (Importer GDP)	0.8007***	0.0116	68.51	0.000
ln_dist (Distance)	-1.2392***	0.0278	-44.47	0.000
contig (Common Border)	0.6985***	0.0526	13.28	0.000
post_accession (EU Impact)	0.2872*	0.0354	8.11	0.000
comlang_off (Common Lang)	(omitted)	-	-	-
Constant (β0)	-30.4560***	0.3524	-86.41	0.000
*Note: *** denotes statistical significance at the 1% level*				

As shown in Table 1, all core gravity variables and our variable of interest (post_accession) are statistically significant at the 1% level. The high R-squared of 0.824 indicates that the model has a very strong explanatory power, capturing the majority of the variation in trade flows between Eastern Europe and the EU.

Note that the variable comlang_off was omitted from the regression due to collinearity. In our specific sample of Eastern European exporters and EU importers, there is insufficient variation in

official language commonality to estimate a separate effect, or it is perfectly correlated with other geographic factors.

Model Diagnostics:

Number of Observations: 6,064
R-squared: 0.8242
Adjusted R-squared: 0.8240
F-statistic (6, 6057): 5674.38 ($p < 0.000$)

(1.8) Limitations

- Zero Trade Flows: Our model uses logarithms (ln_trade), which automatically drops observations where trade is zero. If zero trade flows are not random, this could create selection bias.
- Endogeneity: While distance is exogenous, GDP might be endogenous (trade increases GDP). However, in gravity models, this is often considered a minor issue compared to omitted variables.
- Multilateral Resistance: We used a basic OLS model. Advanced literature (Anderson & van Wincoop) suggests using fixed effects (Importer-Year and Exporter-Year) to control for relative price levels ("Multilateral Resistance Terms"), which were not fully included in this simplified model.

References

- Anderson, J. E., & van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. American Economic Review.
- Tinbergen, J. (1962). Shaping the World Economy: Suggestions for an International Economic Policy. The Twentieth Century Fund.
- Viner, J. (1950). The Customs Union Issue. Carnegie Endowment for International Peace.