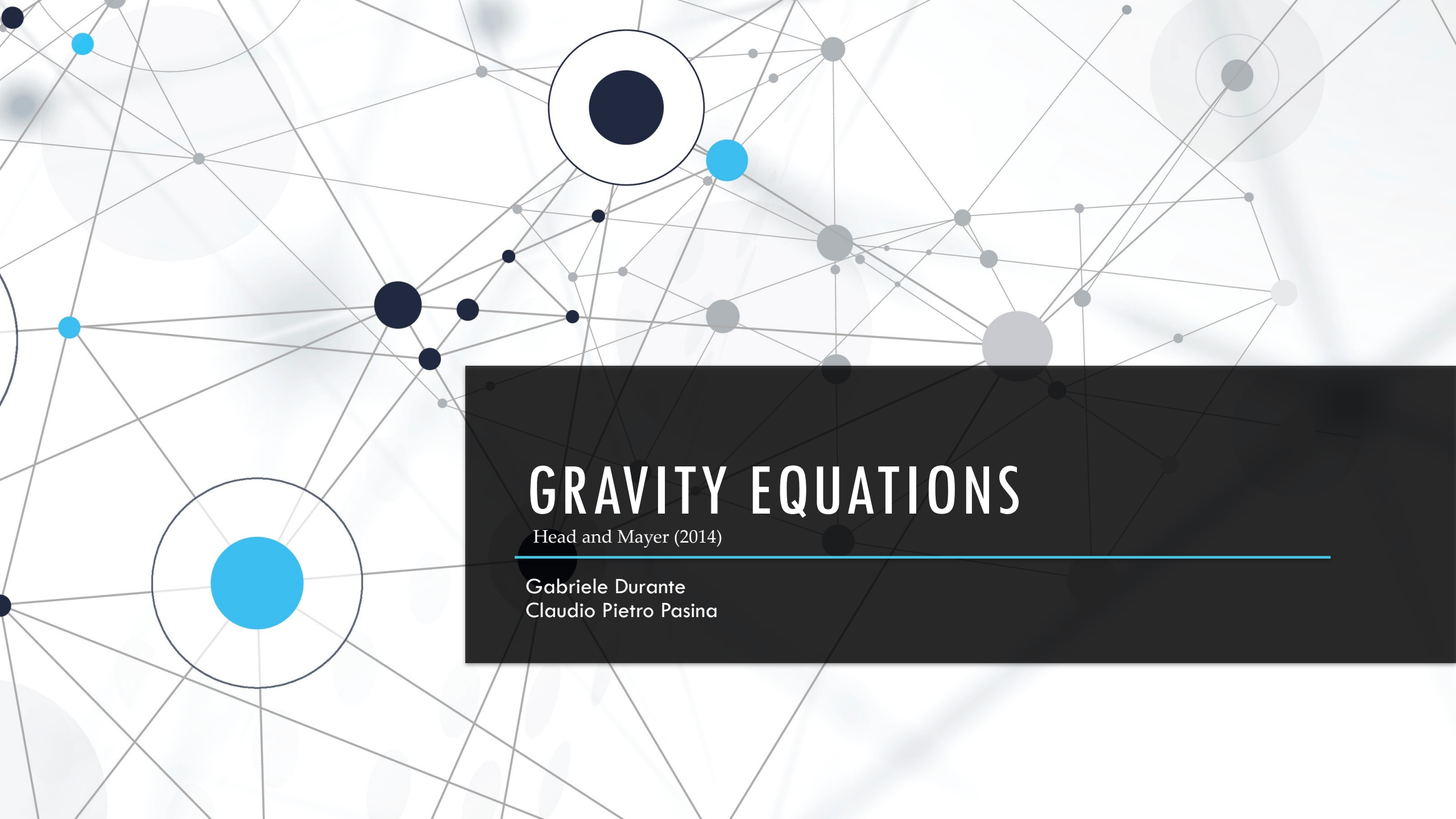


GRAVITY EQUATIONS

Head and Mayer (2014)

Gabriele Durante
Claudio Pietro Pasina

GRAVITY EQUATION: CHOICE OF TOPIC AND SIMILAR RESEARCH

- Our research focuses on the application of econometrics to the gravity equation in international trade, a widely used model for understanding trade flows between countries.
- This work is interesting because it provides an empirical perspective on the effect of economic and geographic variables on international trade, allowing us to assess the importance of factors such as economic size, geographic distance and other influences on trade relations.
- Through the use of econometric models, we will analyze the effect of independent variables on international trade and evaluate the adequacy of the model in explicating the observed trade flows.
- Some reference articles that inspired my work include: *(1) Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. (2) Santos Silva, J. M., & Tenreyro, S. (2006). The log of gravity.*
- These studies provided a solid theoretical and empirical basis forour analyses, but our work stands out because of the specificity of the variables/period considered and its unique implications.

DATA INFORMATIONS AND WHY THEY ARE APPROPRIATE FOR OUR RESEARCH

- The database in analysis is the one provided by the CEPII (Centre d'Études Prospectives et d'Informations Internationales), who is an economic research center based in Paris, France. CEPII focuses on the analysis of international economic issues and the study of economic relations between countries. (LINK)
- It is a panel data containing informations for each country in different years.
- Recent years may have variables with many missing data.
- The database compares each country with each other to check for relationships
- We used a simplified version of the database and analyzed the informations provided in 2020
- the use of information from a CEPII database is appropriate because it offers comprehensive, detailed and reliable data covering a wide range of countries and provides key information for analyzing international trade flows.

THE GRAVITY EQUATION

The gravity equation is an empirical model used to explain bilateral trade flows between countries. It is based on the intuition that trade flows are positively related to the size (usually measured by GDP or population) of the trading partners and negatively related to the distance between them. The gravity equation is commonly specified as follows:

$$trade_{iJ} = \beta_0 \times \frac{(GDP_i \times GDP_j)}{distance_{ij}{}^{\alpha}} \times (other\ factors)^{\gamma}$$

- Trade_ij represents the bilateral trade flow between country i and country j.
- GDP_i and GDP_j are the GDPs of country i and country j, respectively.
- Distance_ij is the geographical distance between country i and country j.
- β_0, α, and γ are the parameters to be estimated.
- Other Factors represents additional factors that influence trade flows, such as cultural, political, or institutional variables.

THE CORE OF THE ANALYSIS

The **objective** of our analysis was to determine whether the **Gravity equation** could be somehow improved using different **control variables.**

Such variables were grouped by:

1. **Trade facilitation variables**
2. **Deeper distance-related variables**
3. **Cultural distance variables**

To carry out the analysis, we used both a statistical and econometric approach

For the econometric analysis, we used the Ordinary Least Squares (OLS) model

ORDINARY LEAST SQUARES MODEL (OLS)

The OLS model is one of the most common methods used in econometrics to estimate the parameters of a linear model

we chose the OLS model for several reasons:

1. Linearity: This makes the OLS model particularly suitable for estimating parameters and obtaining consistent estimates.
2. Simplicity: The OLS model is relatively simple to implement and interpret. It requires only the assumption of linearity between the independent and dependent variables and the assumption of homoschedasticity of errors.
3. Efficiency: In the case where the errors are homoschedastic and uncorrelated, the OLS model provides the most efficient parameter estimators among all unbiased linear estimators.
4. Interpretation: The coefficients estimated by the OLS model provide estimates of marginal ratios, that is, the marginal effect of a change in an independent variable on changes in the dependent variable, holding the other independent variables constant. This allows a direct interpretation of the impact of variables on international trade.

The use of the OLS model presupposes the fulfillment of certain basic assumptions, such as the absence of autocorrelation of the errors and the absence of correlation between the errors and the independent variables. Therefore, it was essential to conduct residual analysis and verify the adequacy of the model before drawing valid conclusions from the estimated results.

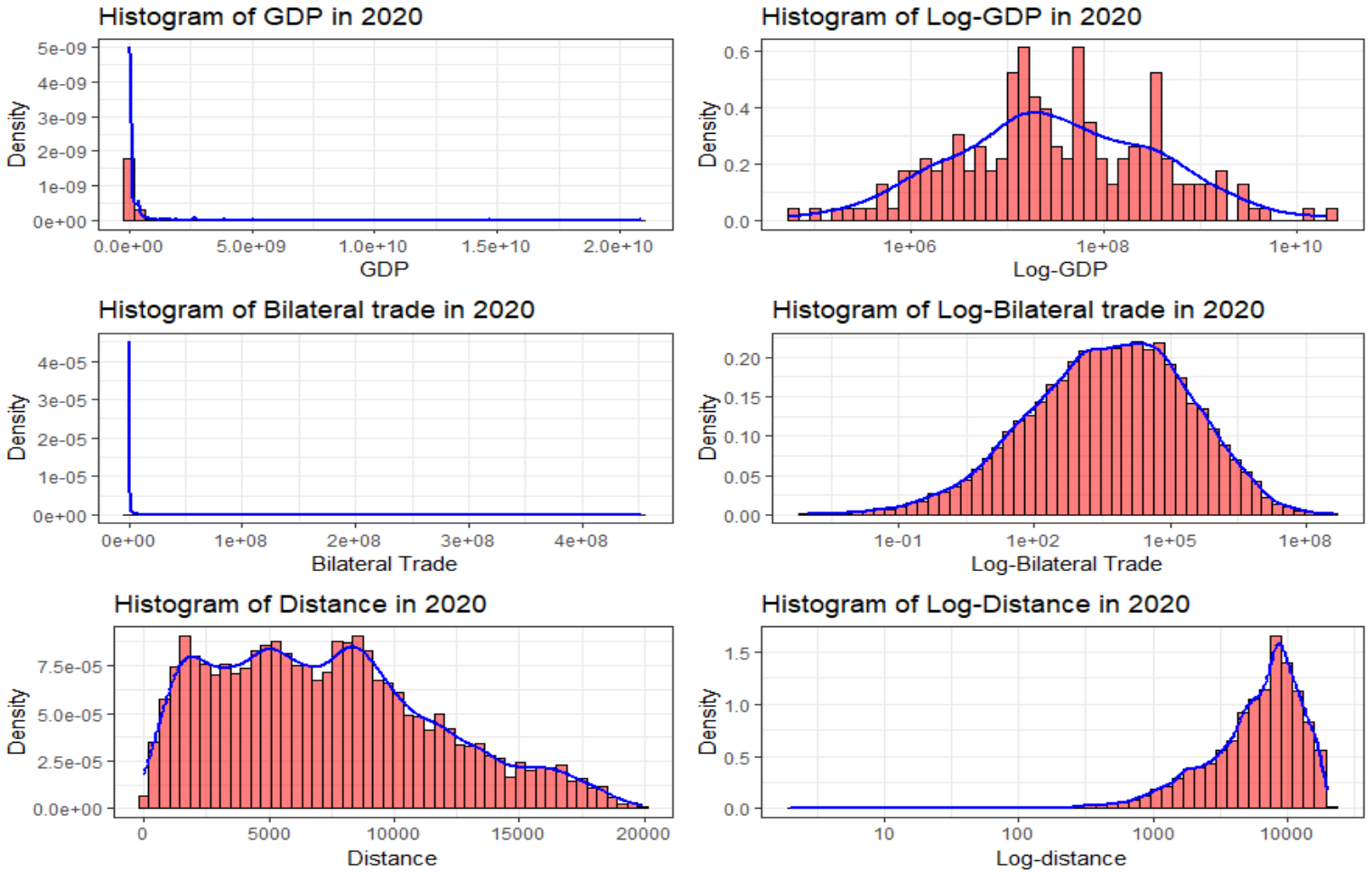
DESCRIPTIVE STATS: VARIABLES

- **distance**: distance between the origin and destination countries.
- **gdp:** GDP (Gross Domestic Product) of the country.
- **population:** population of the country.
- **gdp_percapita:** GDP per capita of the country.
- **diplo_disagreement:** Indicator variable for diplomatic disagreement between the origin and destination countries.
- **sib_conflict:** Indicator variable for sibling conflict between the origin and destination countries.
- **distcap:** Distance between the capital cities of the origin and destination countries
- **religion_sim:** Indicator variable for religious similarity between the origin and destination countries.
- **common_language:** Indicator variable for a common language between the origin and destination countries.
- **gatt:** Indicator variable for General Agreement on Tariffs and Trade (GATT) membership of the country.
- **eu_member:** Indicator variable for European Union (EU) membership of the country.
- **wto_member:** Indicator variable for World Trade Organization (WTO) membership of the country.
- **contiguity:** Indicator variable for geographical contiguity between the origin and destination countries.
- **gmt_offset_2020:** Time zone difference (GMT offset) of the country in the year 2020.

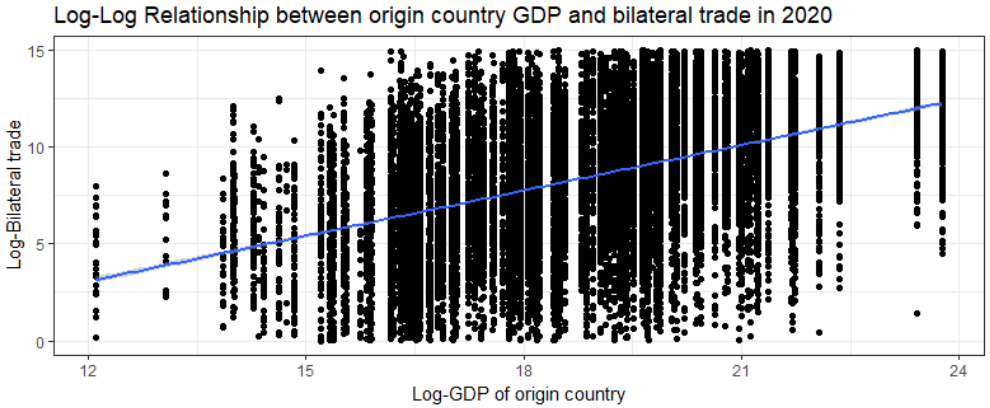
ANALYSIS

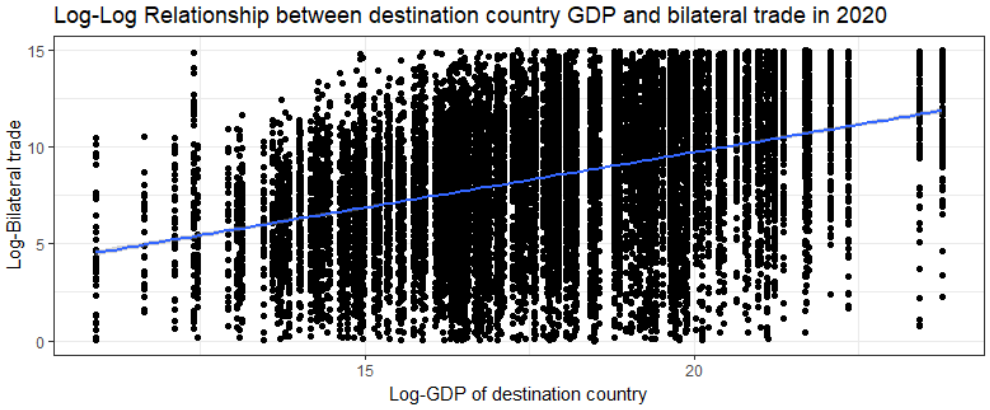
MAIN VARIABLES - DENSITY ANALYSIS AND RESIDUALS

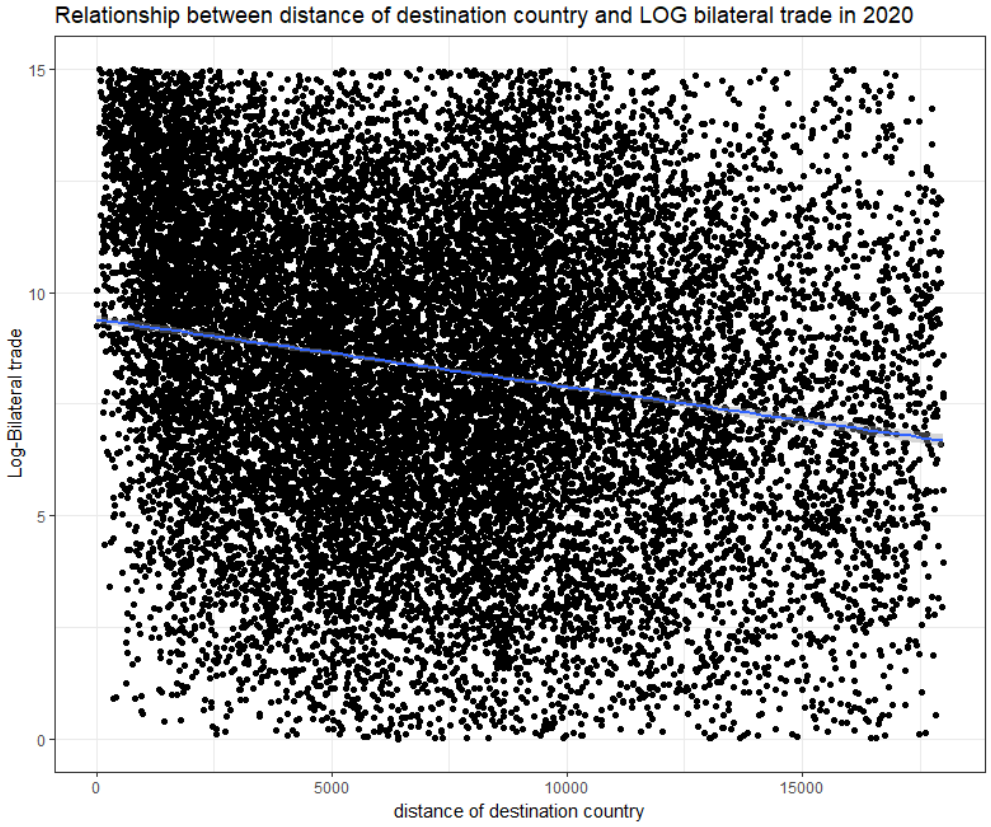
1. We selected the variables we wanted to study and studied the **correlation**.
2. we represented the main variables graphically to choose the best **scale** basing on the distribution of the variables.
3. We also did a **residual** analysis and removed the outliers that would adversely affect the results of the regressions.
4. Finally, we plotted the main variables to analyze the **relationship** between them, observable by a positive regression line between the first two, and inverse in the last.



MAIN VARIABLES RELATIONSHIPS







OLS REGRESSION ANALYSIS

Original gravity equation:

$$log(bilateral\_trade) = \beta 0 + \beta 1 log(distance) + \beta 2 log(gdp_{origin}) + \beta 3 log(gdp\_dest) + \varepsilon$$

Summary:

- **R2 of 0.605** indicates that the model is able to explain about 60.5% of the total variance in the dependent variable.
- **high F-value** suggests that the model's independent variables, taken together, have a strong relationship with the dependent variable.
- An increase of one unit in the variable $log(distance)$ corresponds to a decrease of 1.454 in the dependent variable $log(bilateral\_trade)$, holding all other variables constant.
- All coefficients are statistically significant
- The standard error of 2.162 indicates that the coefficient estimates can vary by an average of 2.162 units around their estimated values
- The results suggest that the multiple regression model has a **good ability to explain the variation in the dependent variable considered.**

original gravity equation

	(1)
(Intercept)	−17.001***
	(0.279)
log(distance)	−1.454***
	(0.020)
log(gdp_origin)	1.220***
	(0.009)
log(gdp_dest)	0.859***
	(0.008)
Num.Obs.	17410
R2	0.605
R2 Adj.	0.605
AIC	370131.2
BIC	370170.0
Log.Lik.	−38127.817
F	8892.420

TRADE FACILITATION VARIABLES

We analyzed four different models, one for each dummy variable representing the countries' membership to:

- **EU**
- **GATT**
- **WTO**
- **Crossed model for all of them**

The most meaningful of the first three models was the WTO based one, with highest R2 and F-stat.

The **last one's R2** was slightly **better,** but the **F-stat dropped** to half of the models (1), (2), (3), to a value of 2296.

For this reason we chose to consider only the WTO dummies in our final model.

	(1)	(2)	(3)	(4)
(Intercept)	−17.346***	−16.576***	−16.758***	−17.145***
	(0.280)	(0.281)	(0.325)	(0.330)
log(distance)	−1.396***	−1.467***	−1.477***	−1.418***
	(0.022)	(0.021)	(0.020)	(0.022)
log(gdp_origin)	1.202***	1.218***	1.206***	1.158***
	(0.009)	(0.009)	(0.009)	(0.009)
log(gdp_dest)	0.865***	0.845***	0.840***	0.845***
	(0.006)	(0.006)	(0.006)	(0.006)
eu_member_origin	0.306***			0.275***
	(0.040)			(0.040)
eu_member_dest	−0.152**			−0.189***
	(0.052)			(0.051)
eu_member_origin × eu_member_dest	0.662***			0.605***
	(0.114)			(0.113)
gatt_o		−0.263***		−0.407***
		(0.069)		(0.072)
gatt_d		−0.175*		−0.306***
		(0.073)		(0.076)
gatt_o × gatt_d		0.553***		0.506***
		(0.092)		(0.095)
wto_member_origin			0.149	0.373+
			(0.197)	(0.205)
wto_member_dest			−0.101	0.148
			(0.204)	(0.212)
wto_member_origin × wto_member_dest			0.587**	0.257
			(0.209)	(0.215)
Num.Obs.	17410	17410	17410	17410
R2	0.606	0.607	0.610	0.613
R2 Adj.	0.606	0.607	0.609	0.613

LINK

DISTANCE-RELATED AND CULTURAL VARIABLES

For the **distance-related** variables we could notice that:

- Jetlag is not relevant
- The distance from capitals is not relevant
- Contiguity affects positively the trades
- Distance from capital cities is a not relevant variable, even in a model without distance.

For the **cultural** variables we could notice that:

- Religion similarity is not so relevant
- The interaction between the dummies of thetwo main languages of each country is very relevant
- Diplomatic disagreement is vary relevant

LINK	(1)
(Intercept)	−14.063***
	(0.447)
log(distance)	−1.874***
	(0.052)
log(gdp_origin)	1.217***
	(0.009)
log(gdp_dest)	0.860***
	(0.008)
contiguity	1.218***
	(0.132)
same_gmt_offset_dummy	−0.141*
	(0.057)
distcap	0.000***
	(0.000)
Num.Obs.	17410
R2	0.612
R2 Adj.	0.611
AIC	369853.2
BIC	369915.3
Log.Lik.	−37985.856
F	4566.223

LINK	(1)
(Intercept)	−18.537***
	(0.303)
log(distance)	−1.434***
	(0.022)
log(gdp_origin)	1.253***
	(0.009)
log(gdp_dest)	0.903***
	(0.008)
religion_sim	0.169*
	(0.071)
common_language_primary	0.031
	(0.111)
common_language_second	0.022
	(0.111)
diplo_disagreement	−0.153***
	(0.023)
common_language_primary × common_language_second	0.753***
	(0.168)
Num.Obs.	15312
R2	0.625
R2 Adj.	0.624
AIC	329062.0
BIC	329130.3
Log.Lik.	−33072.052
F	3193.780

CONCLUSIONS

CONCLUSIONS

After comparing the main five models, which were:

1) The basic Gravity Model
2) The GM + Membership dummies
3) The GM + Distance variables
4) The GM + Cultural variables
5) The GM + overall best variables

We could conclude that:

LINK

	(1)	(2)	(3)	(4)	(5)
Num.Obs.	17410	17410	17410	15312	15312
R2	0.605	0.613	0.612	0.625	0.630
R2 Adj.	0.605	0.613	0.611	0.624	0.630
AIC	370131.2	369801.5	369853.2	329062.0	328831.1
BIC	370170.0	369910.2	369915.3	329138.3	328938.0
Log.Lik.	−38127.817	−37954.009	−37985.856	−33072.052	−32952.622
F	8892.420	2296.003	4566.223	3183.780	2175.371
RMSE	2.16	2.14	2.14	2.10	2.08

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

- The R2 gets always higher, showing better fitting models as we add meaningful variables, but f-stat decreases until becomes a quarter of the first model's. RMSE goes just slightly lower.
- Adding many variables allowed us to increase the R2 of only **0,025**
- **We conclude that the first model, with its simplicity, alredy has good values.**
- The improvement that we would get in applying the last model does not justify the variables cleaning and gathering required.
- **Simple models** like the first one could be applied easily in **multiperiodal analysis** that could follow our work.

REFERENCES

- *Anderson, J. E., & van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. American Economic Review, 93(1), 170-192.*
- *Santos Silva, J. M., & Tenreyro, S. (2006). The log of gravity. Review of Economics and Statistics, 88(4), 641-658.*
- *Helpman, E., Melitz, M. J., & Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. Quarterly Journal of Economics, 123(2), 441-487.*
- *Egger, P., & Larch, M. (2008). Interdependent preferential trade agreement memberships: An empirical analysis. Journal of International Economics, 76(2), 384-399.*
- *Head, K., Mayer, T., & Ries, J. (2010). The erosion of colonial trade linkages after independence. Journal of International Economics, 81(1), 1-14.*
- *"Gravity Equations: Workhorse, Toolkit, and Cookbook" - Keith Head and Thierry Mayer*