Analyzing the Relationship Between Total Trade, GDP Per Capita, and Quality of Government

ECO 5720 - 101

Presentation 3

April 29, 2024

Sharon Ware, Will Carswell, Alexia Moehling,
Ashton Roess, Taylor Lautz

ABOUT THE DATA

- The QoG Basic dataset is a subset of the larger QoG Standard dataset, containing around 400 variables sourced from 80 different data providers focused on aspects related to the Quality of Government.
- If data for a specific country is unavailable for 2020, data from the nearest available year within a range of +/- 3 years is included, such as 2021 or 2019.
- Additionally, the QoG Basic Time Series (TS) dataset spans from 1946 to 2023, analyzing data on a country-year basis (e.g., Sweden-1946, Sweden-1947, and so forth), providing a comprehensive historical perspective.
- In the Codebook, you can find a description of all data sources and variables. We provide a list of the variables grouped into nineteen main topics.

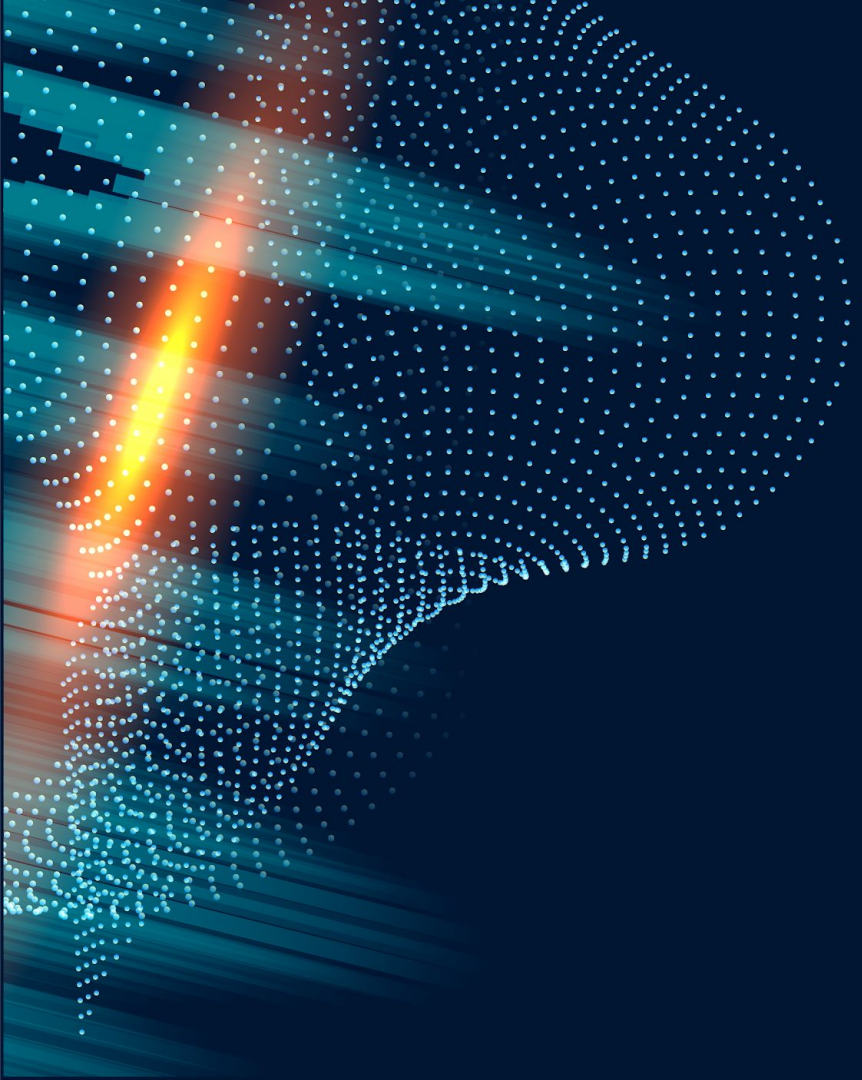
Our research has two primary objectives:

- We aim to understand how high-quality political institutions are established and sustained.
- We investigate the impact of government quality on various policy domains, including health, environment, social policy, and poverty.

Question

- **We are interested in exploring how a country's total trade (gle_trade) relates to the quality of government (icrg_qog).**

    - Understanding this relationship provides insights into economic and governance dynamics, shedding light on how international trade impacts the quality of governance within countries. Furthermore, this relationship can inform policymakers about potential strategies to enhance governance standards through economic measures.



Variables

- **Independent Variables**:
  - Total Trade (gle_trade)
    - Represents the total trade (imports and exports) of a country in millions of current-year US dollars
  - GDP per Capita (gle_cgdpc)
    - Measures the average economic output per person in a country.

- **Dependent Variable**:
  - Quality of Government (icrg_qog)
  - The mean value of the ICRG variables 'Corruption', 'Law and Order' and 'Bureaucracy Quality', scaled from 0 to 1. Higher values indicate higher quality of government.

Simple Linear Regression

```
. regress gle_trade icrg_qog
```

Source	SS	df	MS	Number of obs	=	2,151
				F(1, 2149)	=	446.39
Model	8.7664e+12	1	8.7664e+12	Prob > F	=	0.0000
Residual	4.2203e+13	2,149	1.9638e+10	R-squared	=	0.1720
				Adj R-squared	=	0.1716
Total	5.0969e+13	2,150	2.3707e+10	Root MSE	=	1.4e+05

gle_trade	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
icrg_qog	267947.1	12682.06	21.13	0.000	243076.7	292817.5
_cons	-92460.94	7741.548	-11.94	0.000	-107642.6	-77279.23

Multiple Linear Regression

```
. regress gle_trade icrg_qog gle_cgdpc
```

Source	SS	df	MS	Number of obs	=	2,151
				F(2, 2148)	=	304.61
Model	1.1262e+13	2	5.6310e+12	Prob > F	=	0.0000
Residual	3.9707e+13	2,148	1.8486e+10	R-squared	=	0.2210
				Adj R-squared	=	0.2202
Total	5.0969e+13	2,150	2.3707e+10	Root MSE	=	1.4e+05

gle_trade	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
icrg_qog	138272.6	16611.83	8.32	0.000	105695.6	170849.5
gle_cgdpc	5.3337	.4590504	11.62	0.000	4.43347	6.233929
_cons	-59037.96	8042.923	-7.34	0.000	-74810.68	-43265.23

Improvements

Increasing Adjusted R-Squared

- The Adjusted R-Squared value of the simple linear regression is low, 0.1716. This means that the model we used was not a very good fit for the data. We are looking for a value closer to 1 which indicates that all variability in the dependent variable is explained by the independent variables.
- The Adjusted R-Squared value of the multiple linear regression is better than the SLR but not by very much, 0.2202. This also means that the model we used was not a very good fit for the data. For both of these estimation strategies, I would recommend looking at other models to analyze our data and through cross-validation.

Decreasing the Root MSE Value

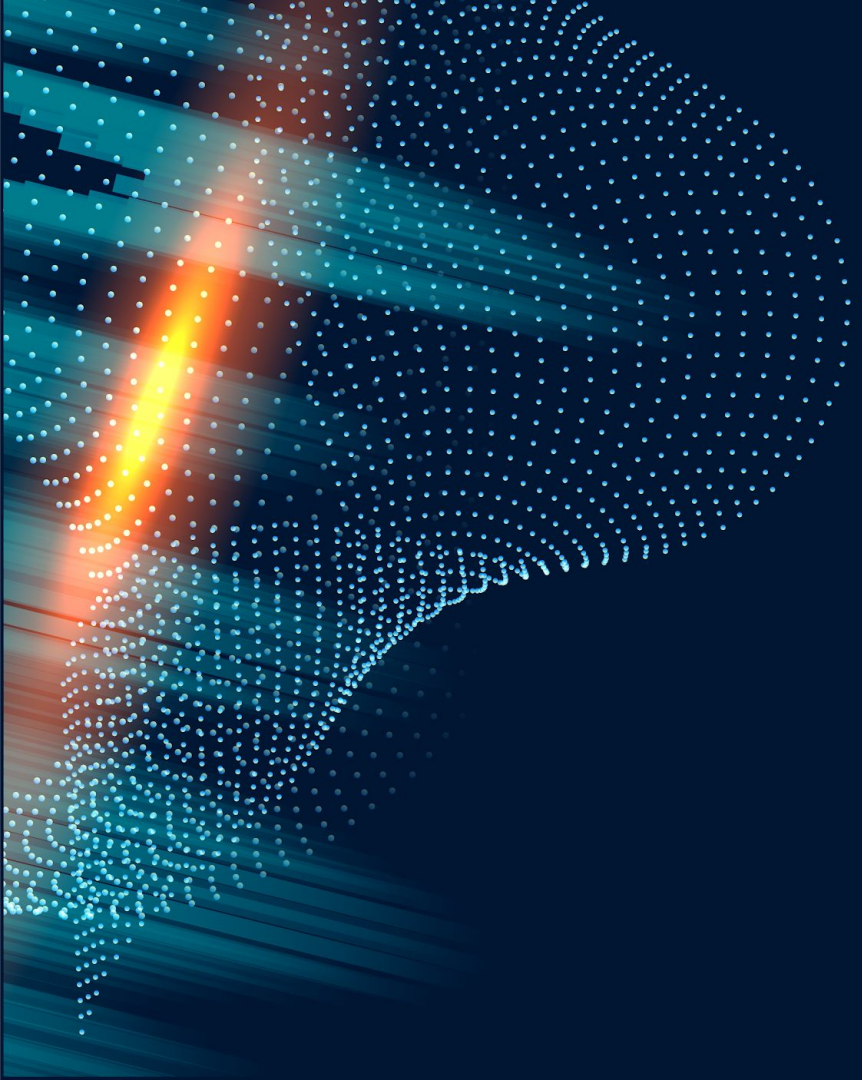
- In both the SLR and MLR results, there is a Root MSE value of 140,000. RMSE measures the amount of error between the predicted value and actual values of the data set. This large Root MSE value could indicate the possibility of overfitting. It could also indicate that the model is not accurately seeing the underlying patterns in the data. Ways to fix this include improving model selection and through cross-validation.

Conclusions

For the simple linear regression model: y = -92460 + 267947icr_qog

For the multiple linear regression: y = -59037 + 138272icr_qog + 5.3337gle_cgdpc

- Provided equations for SLR (y = -92460 + 267947icr_qog) and MLR (y = -59037 + 138272icr_qog + 5.3337gle_cgdpc) serve as starting points for model reassessment and improvement.
- Based on our improvements and the variables tested within the data we can conclude that:
  - Adjusted R-Squared values for both simple linear regression (SLR) and multiple linear regression (MLR) are low, indicating poor model fit.
  - Root Mean Squared Error (RMSE) values of 140,000 in both SLR and MLR suggest potential overfitting and failure to see data patterns
  - The low Adjusted R-Squared values and high RMSE values suggest that the current models may not effectively address the research question regarding the relationship between a country's total trade (gle_trade) and the quality of government (icrg_qog).
  - Overall objective is to develop models that better capture data complexities, enhancing predictive analytics effectiveness.



Thank you!