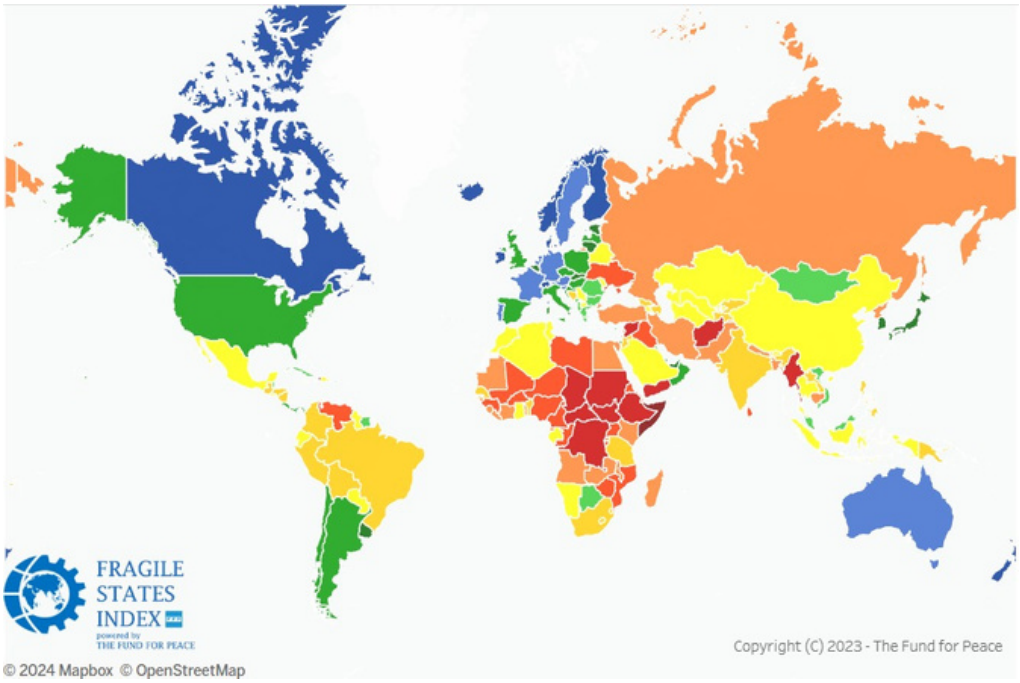
DEPARTMENT OF BUSINESS ECONOMICS
UNIVERSITY OF DELHI

TOPIC:

A Multiple Regression Analysis of Fragile States Index Indicators to Understand the Impact on a Nation's Economic Inequality



SUBMITTED TO:
Dr. Sonal Katyal

SUBMITTED BY:
Aditya Chandramauli - 23005
Akshat Goyal - 23010
Devanshu Hembrom - 23027
Harsh Bengani - 23036
Ranjeet Kumar - 23071
Mohd Zeeshan Ali - 23058
Saurya Karn - 23085

INDEX

1. Introduction: A Deep Dive into the Fragile States Index Methodology

The data we have chosen is the The Fragile States Index (FSI) 2023 Report, a crucial tool developed by The Fund for Peace (FFP), sheds light on a critical issue: a state's vulnerability to instability. This index, meticulously constructed, ranks 178 countries based on their risk of succumbing to internal pressures that could lead to state fragility. Understanding the methodology behind the FSI is fundamental, as it reveals the intricate process of identifying and measuring these vulnerabilities.

The cornerstone of the FSI is the Conflict Assessment System Tool (CAST) framework. CAST, developed in the 1990s, aimed to equip policymakers and field practitioners with a framework for comprehending and gauging the drivers and dynamics of conflict in complex environments. This framework's enduring relevance is evident in its continued use by professionals and even local civil society organizations within conflict zones. Interestingly, the FSI itself emerged from CAST, as researchers sought to determine if the framework could be leveraged to assess and rank state fragility on a national level.

The methodology employed by the FSI goes beyond a simple, standardized approach. It embraces a mixed-method strategy, meticulously integrating both qualitative and quantitative data sources. This multifaceted approach ensures a comprehensive picture of a country's vulnerabilities.

At the heart of the FSI's data collection process lies content analysis. Researchers delve into a vast ocean of information – millions of news articles, research reports, and other publicly available documents. For each of the twelve key indicators within the FSI framework, meticulously crafted Boolean search phrases are applied to this global media data. The objective? To determine the salience, or prominence, of issues related to each

indicator within each country. The raw data for this analysis comes from a commercial content aggregator, encompassing information from over 10,000 English-language sources worldwide. Every year, a staggering 45-50 million articles and reports are meticulously analyzed. Based on the assessed salience for each sub-indicator within a country, provisional scores are assigned.

While content analysis delves into the qualitative realm, the FSI also leverages the power of quantitative data. Pre-existing quantitative datasets, meticulously compiled by international and multilateral statistical agencies such as the United Nations, World Bank, and World Health Organization, are meticulously examined. These datasets are chosen for their ability to statistically represent key aspects of the FSI's indicators. The raw data undergoes a process of normalization and scaling to facilitate comparative analysis across countries. The trends identified in this quantitative analysis are then thoughtfully integrated with the results gleaned from the content analysis phase.

However, the FSI goes beyond solely relying on pre-existing data sources. Recognizing the dynamic nature of state fragility, the methodology incorporates qualitative expert analysis. A dedicated team of social science researchers independently reviews each of the 178 countries. This review process involves a meticulous assessment of key events throughout the year, drawing comparisons with the previous year. This approach is crucial, as it helps to capture dynamic year-on-year trends across different indicators that might not be readily apparent in quantitative datasets measuring long-term structural factors. Additionally, it serves as a safeguard against potential biases or inaccuracies that might arise from the content analysis data.

Once these three independent data streams – content analysis, quantitative data analysis, and qualitative expert analysis – are compiled, a critical process of triangulation commences. This triangulation involves applying a set of established rules to ensure that the data sets are integrated in a way that capitalizes on the strengths of

each approach. Furthermore, it helps to mitigate any inherent weaknesses, gaps, or biases that might be present in any single data source. It's important to note that the raw data underpinning the FSI is already freely available electronically; the true strength lies in the meticulous methodological rigor and the systematic integration of this diverse range of data sources.

Through this rigorous process, final indicator scores for each country are meticulously crafted. A panel review, conducted by the FSI research team, ensures that the final scores are proportionate across the entire spectrum of countries analyzed.

The FSI, however, transcends a simple ranking system. It aspires to be a springboard for deeper, interpretive analysis for users. While the index inherently ranks countries, highlighting some as more fragile than others, the ultimate goal is to illuminate trends in the internal pressures faced by each individual state. By pinpointing the most salient pressures within a country, the FSI empowers policymakers and practitioners to conduct in-depth analysis and craft targeted plans to bolster each state's resilience and safeguard against instability. In essence, the FSI serves as a crucial tool for conflict prevention, enabling proactive measures to be taken before internal pressures escalate into state fragility.

Here's a breakdown of these variables and what they tell us about a country's stability:

1. **Cohesion Indicators**

C1: Security Apparatus: This indicator evaluates the effectiveness of a state's security forces (police, military) in maintaining order and preventing violence. A weak security apparatus can create a power vacuum, allowing armed groups to flourish.

C2: Factionalized Elites: This indicator measures the level of competition and fragmentation among a country's elite groups. When elites prioritize personal gain over

national interests, it can lead to instability and undermine public trust.

C3: Group Grievance: This indicator assesses the presence and intensity of historical or ongoing grievances held by specific groups within a society. Unresolved grievances can fuel resentment and potentially lead to violence.

2. **Economic Indicators**

E1: Economy: This indicator measures the severity of a country's economic decline, including factors like GDP growth rate, unemployment, and inflation. Economic hardship can breed social unrest and weaken government legitimacy.

E2: Economic Inequality: This indicator examines the distribution of wealth and economic opportunities within a country. Significant disparities between different regions or groups can create tensions and fuel social unrest.

E3: Human Flight and Brain Drain: This indicator measures the rate at which skilled or educated people emigrate from a country. This loss of human capital weakens the economy and reduces the country's capacity to address its challenges.

3. **Political Indicators**

P1: State Legitimacy: This indicator assesses the degree to which the public views the government as a legitimate authority. A lack of legitimacy can lead to disobedience and undermine the government's ability to function effectively.

P2: Public Services: This indicator evaluates the quality and reach of essential government services like healthcare, education, and sanitation. Poor quality or limited access to public services can erode public trust and fuel discontent.

P3: Human Rights and Rule of Law: This indicator measures the government's respect for human rights and its adherence to the rule of law. Violations of human rights and a lack of rule of law can create an environment of fear and impunity, hindering development and stability.

4. **Social and Cross-Cutting Indicators:**

S1: Demographic Pressures: This indicator examines population growth rates, age structure, and resource scarcity. Rapid population growth or imbalances in the age structure can strain resources and lead to competition for scarce resources.

S2: Refugees and Internally Displaced Persons: This indicator measures the number of people displaced from their homes due to conflict or natural disasters. Large refugee or IDP populations can put a strain on resources and social services, potentially leading to instability.

X1: External Intervention: This indicator assesses the impact of foreign intervention (military, political, economic) on a state's stability. Depending on the nature of the intervention, it can exacerbate or mitigate existing tensions.

By analyzing these twelve variables, **the Fragile States Index provides a comprehensive picture of a country's strengths and weaknesses.**

2. Data Summary and Key Objectives:

```
. summ S1DemographicPressures, detail
```

S1: Demographic Pressures

	Percentiles	Smallest		
1%	1.2	1.1		
5%	2.4	1.2		
10%	3	1.4	Obs	179
25%	4.1	1.5	Sum of Wgt.	179
50%	5.9		Mean	5.955866
		Largest	Std. Dev.	2.278726
75%	8.1	9.7		
90%	8.9	9.7	Variance	5.192592
95%	9.6	9.8	Skewness	-.0500372
99%	9.8	10	Kurtosis	1.904241

```
. summ S2RefugeesandIDPs,detail
```

S2: Refugees and IDPs

	Percentiles	Smallest		
1%	1.1	.5		
5%	1.5	1.1		
10%	1.8	1.1	Obs	179
25%	2.8	1.2	Sum of Wgt.	179
50%	4.5		Mean	4.764246
		Largest	Std. Dev.	2.373935
75%	6.5	9.6		
90%	8.3	9.8	Variance	5.635568
95%	9.1	10	Skewness	.3777024
99%	10	10	Kurtosis	2.153523

. summ C3GroupGrievance,detail

C3: Group Grievance

	Percentiles	Smallest		
1%	.5	.3		
5%	2	.5		
10%	2.3	.5	Obs	179
25%	3.6	.7	Sum of Wgt.	179
50%	5.5		Mean	5.57486
		Largest	Std. Dev.	2.367757
75%	7.6	9.5		
90%	8.8	9.6	Variance	5.606275
95%	9.3	9.7	Skewness	-.0895089
99%	9.7	9.7	Kurtosis	2.014634

E1: Economy

	Percentiles	Smallest		
1%	1	1		
5%	1.5	1		
10%	2.6	1	Obs	179
25%	4.1	1.1	Sum of Wgt.	179
50%	6		Mean	5.687151
		Largest	Std. Dev.	2.200741
75%	7.2	9.5		
90%	8.5	9.6	Variance	4.843261
95%	9.2	9.6	Skewness	-.2947943
99%	9.6	9.9	Kurtosis	2.414107

. summ P1StateLegitimacy,detail

P1: State Legitimacy

	Percentiles	Smallest		
1%	.3	.3		
5%	.4	.3		
10%	.8	.3	Obs	179
25%	3.6	.3	Sum of Wgt.	179
50%	6.4		Mean	5.741341
		Largest	Std. Dev.	2.901853
75%	8.1	9.8		
90%	9.3	9.9	Variance	8.420753
95%	9.7	9.9	Skewness	-.4189213
99%	9.9	10	Kurtosis	2.048629

. summ C1SecurityApparatus,detail

C1: Security Apparatus

	Percentiles	Smallest		
1%	.3	.3		
5%	1.2	.3		
10%	1.8	.4	Obs	179
25%	3.3	.4	Sum of Wgt.	179
50%	5.1		Mean	5.014525
		Largest	Std. Dev.	2.37981
75%	6.7	9.6		
90%	8.3	9.7	Variance	5.663496
95%	9.1	9.9	Skewness	.0364979
99%	9.9	10	Kurtosis	2.268114

. summ C2FactionalizedElites,detail

C2: Factionalized Elites

	Percentiles	Smallest		
1%	1.1	1		
5%	1.8	1.1		
10%	2.7	1.4	Obs	179
25%	4.9	1.4	Sum of Wgt.	179
50%	7.2		Mean	6.618436
		Largest	Std. Dev.	2.427869
75%	8.6	9.9		
90%	9.5	10	Variance	5.894546
95%	9.7	10	Skewness	-.637628
99%	10	10	Kurtosis	2.366222

```
. summ X1ExternalIntervention,detail
```

X1: External Intervention

Percentiles		Smallest		
1%	.4	.3		
5%	.5	.4		
10%	1.1	.4	Obs	179
25%	3.1	.4	Sum of Wgt.	179
50%	5.3		Mean	5.072067
		Largest	Std. Dev.	2.577801
75%	7	10		
90%	8.3	10	Variance	6.645058
95%	9.4	10	Skewness	-.0911419
99%	10	10	Kurtosis	2.187248

```
. pwcorr
(Country ignored because string variable)
(Rank ignored because string variable)
```

	Year	Total	S1Demo~s	S2Refu~s	C3Grou~e	E3Huma~n	E2Econ~y
Year	.						
Total	.	1.0000					
S1Demograp~s	.	0.8657	1.0000				
S2Refugees~s	.	0.8077	0.6797	1.0000			
C3GroupGri~e	.	0.7049	0.4999	0.6179	1.0000		
E3HumanFli~n	.	0.7291	0.6154	0.5736	0.3825	1.0000	
E2Economic~y	.	0.8348	0.8537	0.5934	0.4618	0.5558	1.0000
E1Economy	.	0.8411	0.7391	0.6680	0.4537	0.7220	0.7304
P1StateLeg~y	.	0.8606	0.6571	0.5871	0.6446	0.4845	0.6662
P2PublicSe~s	.	0.8970	0.9161	0.7205	0.4945	0.6903	0.8916
P3HumanRig~s	.	0.8394	0.6730	0.5751	0.6337	0.4350	0.6523
C1Security~s	.	0.8822	0.7409	0.6944	0.6301	0.6094	0.6993
C2Factiona~s	.	0.8759	0.6793	0.6581	0.7096	0.5473	0.6477
X1External~n	.	0.8107	0.6064	0.6864	0.4472	0.7321	0.6044

	E1Econ~y	P1Stat~y	P2Publ~s	P3Huma~s	C1Secu~s	C2Fact~s	X1Exte~n
E1Economy	1.0000						
P1StateLeg~y	0.6229	1.0000					
P2PublicSe~s	0.8055	0.6683	1.0000				
P3HumanRig~s	0.5854	0.9104	0.6614	1.0000			
C1Security~s	0.7047	0.7378	0.7749	0.7584	1.0000		
C2Factiona~s	0.6429	0.8632	0.6764	0.7921	0.7506	1.0000	
X1External~n	0.7573	0.6220	0.6780	0.5803	0.6722	0.6911	1.0000

Our objectives involving the Fragile States Index (FSI) dataset aims to delve deeper into the complex dynamics of state fragility. By utilizing multiple regression analysis within the Stata environment, we have several key objectives:

a. **Identifying Drivers of Fragility:** Our core objective is to isolate the most influential variables among the FSI's twelve indicators that contribute to a state's overall fragility. Through multiple regression, we seek to quantify the relationship between each independent variable and the dependent variable **E2: Economic Inequality**, while controlling for other factors. This will reveal which aspects of a nation's social, economic, and political landscape are most strongly associated with elevated levels of fragility.

b. **Assessing Predictive Power:** Beyond identifying key drivers, we'll examine the overall predictive power of our regression model. Using the **R-squared value**, we can determine what proportion of the variation in FSI scores is explained by our chosen independent variables. A high R-squared suggests a strong model, allowing us to draw clearer inferences about the complex factors behind fragility.

c. **Testing Model Validity:** Rigor is essential in any econometric analysis. We will perform a suite of diagnostic tests on our model to ensure its robustness. A key focus is the **test of homoscedasticity**, which examines the assumption of constant error variance. Unequal variance could undermine our model's accuracy. Using tools like the **Breusch-Pagan test** within Stata, we'll check for **heteroskedasticity** and, if detected, take corrective measures such as using robust standard errors.

d. **Exploring the Impact of Specific Factors:** Our analysis may involve focusing on specific variables of particular interest. For example, we could investigate whether external intervention has a **disproportionate influence** on state fragility, or we might concentrate on the effects of economic inequality and brain drain. By creating models focused on these specific scenarios, we can develop detailed insights into their unique roles within the broader context of fragility.
e. **Understanding Vulnerability and Resilience:** Ultimately, our project aims to illuminate not only the pathways towards fragility but also the characteristics associated with greater resilience. Highlighting variables that have a mitigating effect on the FSI score will inform policies and interventions designed to strengthen fragile states and foster their ability to withstand shocks and challenges.

3. Test Model

E2: Economic Inequality = β0 + β1(S1: Demographic Pressures) + β2(S2: Refugees and IDPs) + β3(C3: Group Grievance) + β4(E3: Human Flight and Brain Drain) + β5(E1: Economy) + β6(P1: State Legitimacy) + β7(P2: Public Services) + β8(P3: Human Rights) + β9(C1: Security Apparatus) + β10(C2: Factionalized Elites) + β11(X1: External Intervention) + ε

Where β0 is the intercept, β1 to β11 are the coefficients representing the impact of each independent variable on E2:Economic Inequality.

The regression equation aims to model the complex factors contributing to a state's Economic Inequality (E2). It posits that economic inequality is influenced by a wide range of variables encompassing social, political, economic, and cross-cutting indicators as defined by the Fragile States Index.

We also incorporate Demographic Pressures (S1), Refugees and IDPs (S2), and Group Grievance (C3), suggesting that societal factors like population dynamics, displacement, and social tensions can exacerbate economic inequality.

The presence of Human Flight and Brain Drain (E3), Economy (E1), State Legitimacy (P1), Public Services (P2), and Human Rights (P3) highlights that both economic conditions and governance-related factors potentially shape income disparities. A poorly performing economy, lack of government legitimacy, insufficient public services, or human rights abuses could fuel economic inequality.

Interestingly, our model considers Security Apparatus (C1), Factionalized Elites (C2), and External Intervention (X1). The inclusion of these variables suggests that you're exploring whether security dynamics, internal power struggles, and foreign involvement play a role in amplifying or mitigating economic inequality.

```
. regress E2EconomicInequality S1DemographicPressures S2RefugeesandIDPs C3GroupGrievance E3HumanFl
> ightandBrainDrain E1Economy P1StateLegitimacy P2PublicServices P3HumanRights C1SecurityApparatus
>  C2FactionalizedElites X1ExternalIntervention
```

Source	SS	df	MS		
Model	631.051331	11	57.3683028	Number of obs	= 179
Residual	130.590122	167	.781976776	F(11, 167)	= 73.36
				Prob > F	= 0.0000
				R-squared	= 0.8285
				Adj R-squared	= 0.8172
Total	761.641453	178	4.27888456	Root MSE	= .88429

E2EconomicInequality	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S1DemographicPressures	.195421	.076424	2.56	0.011	.0445393	.3463027
S2RefugeesandIDPs	-.1416769	.0484826	-2.92	0.004	-.2373948	-.045959
C3GroupGrievance	.0267347	.0442433	0.60	0.546	-.0606135	.1140829
E3HumanFlightandBrainDrain	-.1509686	.0531707	-2.84	0.005	-.2559419	-.0459954
E1Economy	.0688004	.0597128	1.15	0.251	-.0490888	.1866896
P1StateLegitimacy	.0997527	.0681077	1.46	0.145	-.0347104	.2342158
P2PublicServices	.5955397	.0780082	7.63	0.000	.4415304	.749549
P3HumanRights	-.0257718	.0675617	-0.38	0.703	-.1591568	.1076132
C1SecurityApparatus	-.0225374	.0558162	-0.40	0.687	-.1327336	.0876588
C2FactionalizedElites	-.0096093	.0657769	-0.15	0.884	-.1394708	.1202522
X1ExternalIntervention	.0647326	.0500499	1.29	0.198	-.0340795	.1635446
_cons	1.241409	.2543374	4.88	0.000	.7392782	1.74354

```
. vif
```

Variable	VIF	1/VIF
P2PublicSe~s	9.23	0.108347
P1StateLeg~y	8.89	0.112468
P3HumanRig~s	7.04	0.142090
S1Demograp~s	6.90	0.144854
C2Factiona~s	5.81	0.172257
C1Security~s	4.02	0.248983
E1Economy	3.93	0.254391
X1External~n	3.79	0.263918
S2Refugees~s	3.02	0.331637
E3HumanFli~n	2.78	0.359314
C3GroupGri~e	2.50	0.400318
Mean VIF	5.26	

The statistical analysis reveals whether these independent variables indeed have a significant influence on economic inequality, as well as the direction and magnitude of their potential effects. We remove Public Services, State Legitimacy and Human Rights as they have higher VIF.

4. Rectified Model:

E2: Economic Inequality = β0 + β1(S1: Demographic Pressures) + β2(S2: Refugees and IDPs) + β3(C3: Group Grievance) + β4(E3: Human Flight and Brain Drain) + β5(E1: Economy) + β6(C1: Security Apparatus) + β7(C2: Factionalized Elites) + β8(X1: External Intervention) + ε

```
. regress E2EconomicInequality S1DemographicPressures S2RefugeesandIDPs C3GroupGrievance E3HumanFl
> ightandBrainDrain E1Economy C1SecurityApparatus C2FactionalizedElites X1ExternalIntervention
```

Source	SS	df	MS		
				Number of obs	= 179
				F(8, 170)	= 67.92
Model	580.138733	8	72.5173417	Prob > F	= 0.0000
Residual	181.502719	170	1.06766305	R-squared	= 0.7617
				Adj R-squared	= 0.7505
Total	761.641453	178	4.27888456	Root MSE	= 1.0333

E2EconomicInequality	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S1DemographicPressures	.6174415	.0606564	10.18	0.000	.4977048	.7371782
S2RefugeesandIDPs	-.0898355	.0553652	-1.62	0.107	-.1991274	.0194564
C3GroupGrievance	.0019164	.0510622	0.04	0.970	-.0988813	.1027141
E3HumanFlightandBrainDrain	-.0933396	.0599902	-1.56	0.122	-.2117613	.025082
E1Economy	.2009873	.0669873	3.00	0.003	.0687533	.3332213
C1SecurityApparatus	.0704493	.0605739	1.16	0.246	-.0491246	.1900233
C2FactionalizedElites	.0426819	.061502	0.69	0.489	-.0787241	.164088
X1ExternalIntervention	.0637361	.0582877	1.09	0.276	-.0513247	.1787969
_cons	.4452094	.2688514	1.66	0.100	-.0855077	.9759264

Name	Variable	Variable type	Variable represents
y	E2EconomicInequality	Numerical	Dependent Variable
β1	S1DemographicPressures	Numerical	Independent Variable
β2	S2RefugeesandIDPs	Numerical	Independent Variable
β3	C3GroupGrievance	Numerical	Independent Variable
β4	E1Economy	Numerical	Independent Variable
β5	E3HumanFlightandBrain Drain	Numerical	Independent Variable
β6	C1SecurityApparatus	Numerical	Independent Variable
β7	C2FactionalizedElites	Numerical	Independent Variable
β8	X1ExternalIntervention	Numerical	Independent Variable

5. Hypothesis formulation

With our collected data we are trying to answer the following basic question:

“Is that Economic Inequality affected by Demographic Pressures, Refugees and IDPs, Group Grievance, Human Flight and Brain Drain, Economy, Security Apparatus, Factionalized Elites and External Intervention ”

Following hypothesis shall be tested using multiple regression analysis:

(1) Demographic Pressures

Ho: Economic Inequality is not affected by Demographic Pressures.

Ha: Economic Inequality is affected by Demographic Pressures.

(2) Refugees and IDPs

Ho: Economic Inequality is not affected by Refugees and IDPs

Ha: Economic Inequality is affected by Refugees and IDPs.

(3) Group Grievance

Ho: Economic Inequality is not affected by Group Grievance.

Ha: Economic Inequality is affected by Group Grievance.

(4) Human Flight and Brain Drain

Ho: Economic Inequality is not affected by Human Flight and Brain Drain.

Ha: Economic Inequality is affected by Human Flight and Brain Drain.

(5) Economy

Ho: Economic Inequality is not affected by the Economy.

Ha: Economic Inequality is affected by Economy.

(6) Security Apparatus

Ho: Economic Inequality is not affected by Security Apparatus.

Ha: Economic Inequality is affected by Security Apparatus.

(7) Factionalized Elites

Ho: Economic Inequality is not affected by Factionalized Elites.

Ha: Economic Inequality is affected by Factionalized Elites.

(8) External Intervention

Ho: Economic Inequality is not affected by External Intervention.

Ha: Economic Inequality is affected by External Intervention.

Notes:

(i) Ho – Null hypothesis; Ha – Alternative hypothesis

(ii) Significance level for all the tests conducted shall be 0.05.

```
. regress E2EconomicInequality S1DemographicPressures S2RefugeesandIDPs C3GroupGrievance E3HumanFl
> ightandBrainDrain E1Economy C1SecurityApparatus C2FactionalizedElites X1ExternalIntervention
```

Source	SS	df	MS	Number of obs	=	179
				F(8, 170)	=	67.92
Model	580.138733	8	72.5173417	Prob > F	=	0.0000
Residual	181.502719	170	1.06766305	R-squared	=	0.7617
				Adj R-squared	=	0.7505
Total	761.641453	178	4.27888456	Root MSE	=	1.0333

E2EconomicInequality	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S1DemographicPressures	.6174415	.0606564	10.18	0.000	.4977048	.7371782
S2RefugeesandIDPs	-.0898355	.0553652	-1.62	0.107	-.1991274	.0194564
C3GroupGrievance	.0019164	.0510622	0.04	0.970	-.0988813	.1027141
E3HumanFlightandBrainDrain	-.0933396	.0599902	-1.56	0.122	-.2117613	.025082
E1Economy	.2009873	.0669873	3.00	0.003	.0687533	.3332213
C1SecurityApparatus	.0704493	.0605739	1.16	0.246	-.0491246	.1900233
C2FactionalizedElites	.0426819	.061502	0.69	0.489	-.0787241	.164088
X1ExternalIntervention	.0637361	.0582877	1.09	0.276	-.0513247	.1787969
_cons	.4452094	.2688514	1.66	0.100	-.0855077	.9759264

6. Identification of significant predictors

With the results of the regression analysis, we can check our formulated hypotheses, before doing so we shall take a look at the values of coefficients determined by the model. The rejection rule using p-values for rejecting or not rejecting a null hypothesis is as follows

Reject a null hypothesis if: p-value < Significance level

Do not reject a null hypothesis if: p-value > Significance level

Interpreting the results

The Stata outputs depict the results of a multiple linear regression analysis and various diagnostic tests conducted on the provided dataset. The primary variable of interest is **"E2EconomicInequality,"** which is being modeled as a function of several predictor variables, including demographic pressures, refugee and IDP situations, group grievances, human flight and brain drain, economic factors, security apparatus, factionalized elites, and external intervention.

To conduct our hypotheses test we shall use the concept of “p-values” which are provided in the regression output itself.

Regression Model:

The regression model summary (Image 1) indicates that the model includes **8 predictor variables** and has an **R-squared value of 0.7617**, suggesting that approximately **76% of the variation in economic inequality can be explained by the included predictors**. The adjusted R-squared value of **0.7505** further confirms the **model's reasonable fit, accounting for the number of predictors.**

The coefficients table reveals that several predictor variables have statistically significant relationships with economic inequality.

Specifically, **"S1DemographicPressures" (coefficient = 0.6174415, p < 0.001)** and "**E1Economy" (coefficient = 0.2009873, p = 0.003)** exhibit **positive and significant** associations with economic inequality. This suggests that higher demographic pressures and stronger economic conditions are associated with increased economic inequality.

On the other hand, **"S2RefugeesandIDPs" (coefficient = -0.0898355, p = 0.107)** and **"E3HumanFlightandBrainDrain" (coefficient = -0.0933396, p = 0.122)** have **negative coefficients,** although their relationships are not statistically significant at the conventional **5% level**. The remaining predictors, including group grievances, security apparatus, factionalized elites, and external intervention, do not exhibit significant associations with economic inequality in this model.

Variance Inflation Factors (VIFs):

VIFs are used to assess the presence of multicollinearity, which occurs when predictors are highly correlated with each other, potentially leading to unstable and unreliable coefficient estimates.

The mean **VIF of 3.21** suggests a moderate level of multicollinearity among the predictors. However, individual VIF values above 5 or 10 are generally considered problematic. In this case, none of the predictors exceed a VIF of 4, indicating that multicollinearity is not a major concern in the model.

```
. vif
```

Variable	VIF	1/VIF
X1External~n	3.76	0.265682
C2Factiona~s	3.72	0.269020
E1Economy	3.62	0.275989
C1Security~s	3.46	0.288641
S1Demograp~s	3.19	0.313962
S2Refugees~s	2.88	0.347218
E3HumanFli~n	2.59	0.385388
C3GroupGri~e	2.44	0.410337
Mean VIF	3.21	

Heteroskedasticity Tests:

The Breusch-Pagan / Cook-Weisberg test: Tests the null hypothesis of constant variance against the alternative of heteroskedasticity. The chi-square statistic of **0.14** and the corresponding p-value of **0.7127 suggest** that there is no evidence to reject the null hypothesis of homoscedasticity (constant variance) at the 5% significance level.

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of E2EconomicInequality

         chi2(1)      =     0.14
         Prob > chi2  =   0.7127
```

```
. imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(44)     =     40.86
         Prob > chi2  =    0.6069

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	40.86	44	0.6069
Skewness	9.27	8	0.3202
Kurtosis	0.26	1	0.6116
Total	50.39	53	0.5765

However, the **White's test** provides a more comprehensive assessment of heteroskedasticity by testing against unrestricted heteroskedasticity. The chi-square statistic of 40.86 and the associated p-value of 0.6069 indicate that the null hypothesis of homoscedasticity cannot be rejected at the 5% significance level.

It is worth noting that the White's test decomposes the overall test statistic into components attributable to heteroskedasticity, skewness, and kurtosis. In this case, the heteroskedasticity component contributes the largest portion **(40.86)** of the overall test statistic **(50.39)**, while the skewness and kurtosis components are relatively small **(9.27 and 0.26, respectively).**

Correlation Matrix:

```
. corr E2EconomicInequality S1DemographicPressures S2RefugeesandIDPs C3GroupGrievance E3HumanFligh
> tandBrainDrain E1Economy C1SecurityApparatus C2FactionalizedElites X1ExternalIntervention
(obs=179)
```

	E2Econ~y	S1Demo~s	S2Refu~s	C3Grou~e	E3Huma~n	E1Econ~y	C1Secu~s	C2Fact~s	X1Exte~n
E2Economic~y	1.0000								
S1Demograp~s	0.8537	1.0000							
S2Refugees~s	0.5934	0.6797	1.0000						
C3GroupGri~e	0.4618	0.4999	0.6179	1.0000					
E3HumanFli~n	0.5558	0.6154	0.5736	0.3825	1.0000				
E1Economy	0.7304	0.7391	0.6680	0.4537	0.7220	1.0000			
C1Security~s	0.6993	0.7409	0.6944	0.6301	0.6094	0.7047	1.0000		
C2Factiona~s	0.6477	0.6793	0.6581	0.7096	0.5473	0.6429	0.7506	1.0000	
X1External~n	0.6044	0.6064	0.6864	0.4472	0.7321	0.7573	0.6722	0.6911	1.0000

Image presents the correlation matrix, which displays the pairwise correlations among the predictor variables and the dependent variable (E2EconomicInequality). Correlation coefficients range from -1 to 1, with values closer to 1 or -1 indicating stronger positive or negative linear relationships, respectively.

The correlation matrix reveals several notable observations:

1. The dependent variable **"E2EconomicInequality"** has the strongest positive correlation with **"S1DemographicPressures" (0.8537)** and **"E1Economy" (0.7304)**, which aligns with the significant positive coefficients observed in the regression model.

2. There are moderately strong positive correlations among some predictor variables, such as between **"C1SecurityApparatus"** and **"C2FractionalizedElites" (0.7506)**, and between **"E1Economy"** and **"X1ExternalIntervention" (0.7573)**. These correlations could contribute to the moderate levels of multicollinearity observed in the VIF analysis.

3. Negative correlations are observed between **"E2EconomicInequality"** and **"S2RefugeesandIDPs" (-0.5934)** and **"E3HumanFlightandBrainDrain" (-0.5558)**, which is consistent with the negative coefficient signs in the regression model, although the relationships were not statistically significant.

7. Model evaluation and validation

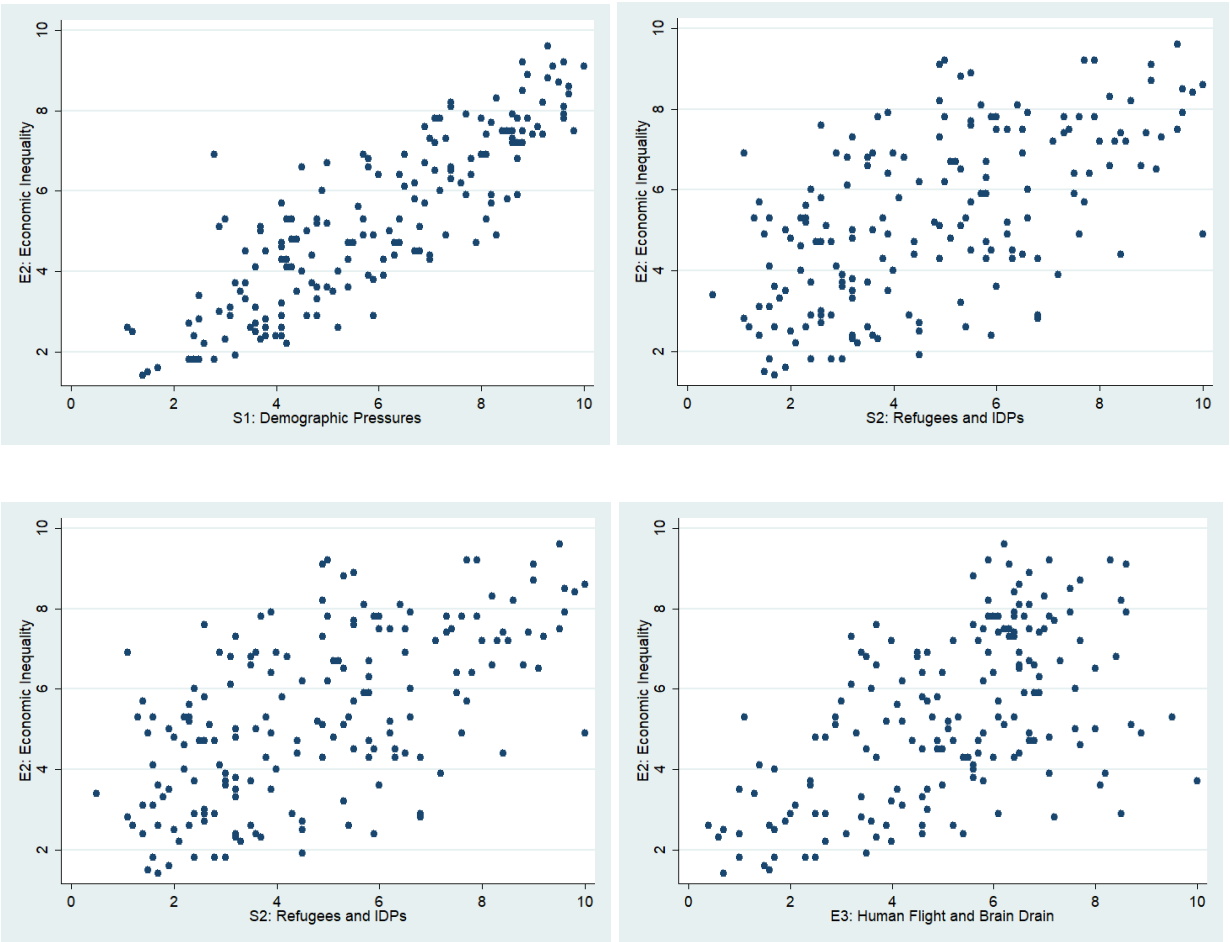
In conducting a regression analysis, it is also crucial to check the assumptions underlying the Ordinary Least Squares (OLS) regression model to ensure that the

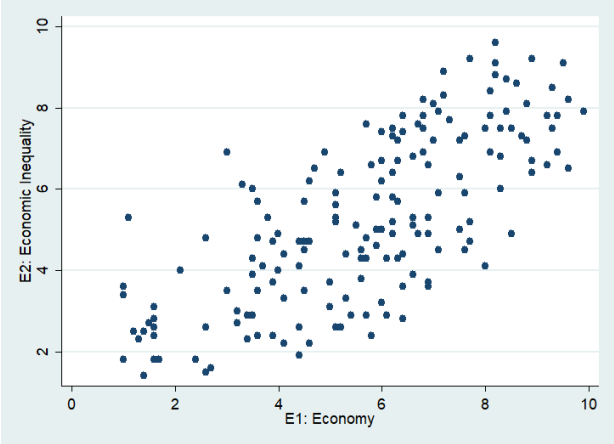
estimated coefficients are unbiased, efficient, and have the smallest variance among all unbiased estimators, known as the Best Linear Unbiased Estimator (BLUE) property. Violations of these assumptions can lead to biased estimates and inefficient estimators, destroying the BLUE property of the OLS Estimators.

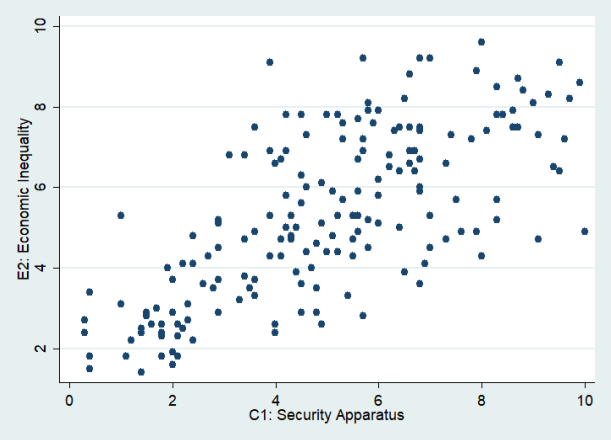
We will now proceed to validate these assumptions.

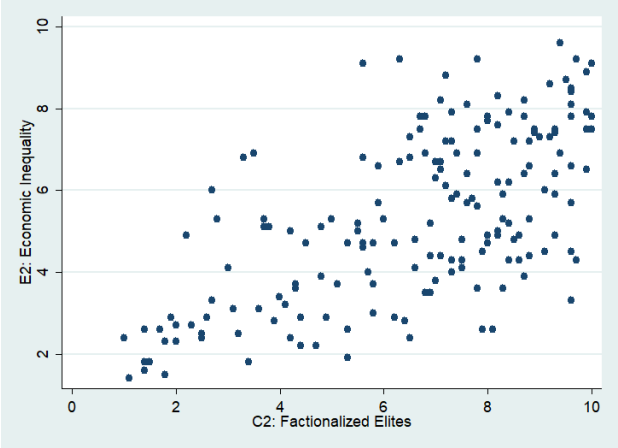
The regression model is linear in the parameters.

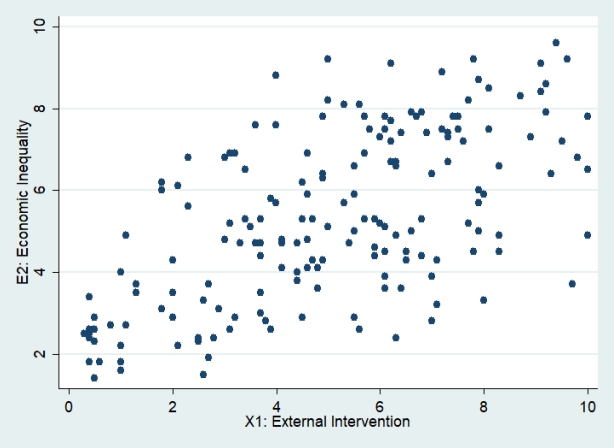
By creating scatterplots between the dependent variable and each independent variable, we can determine whether the model is linear or not.











From the above graphs, it is clear that the assumption of linearity holds true for our model.

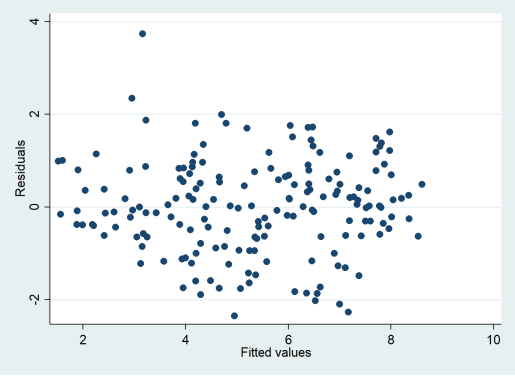
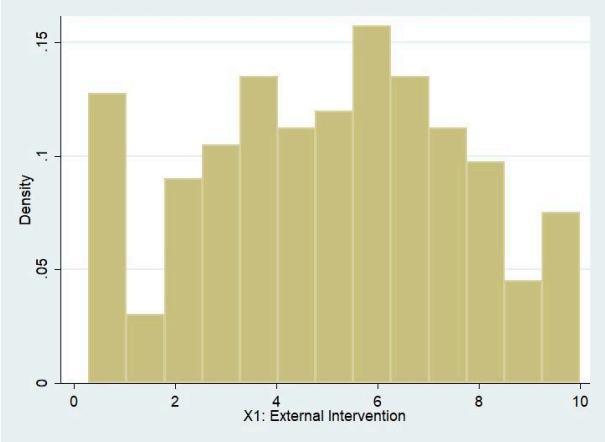
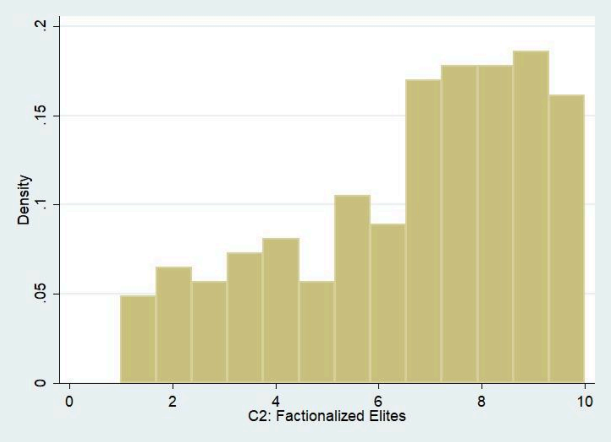


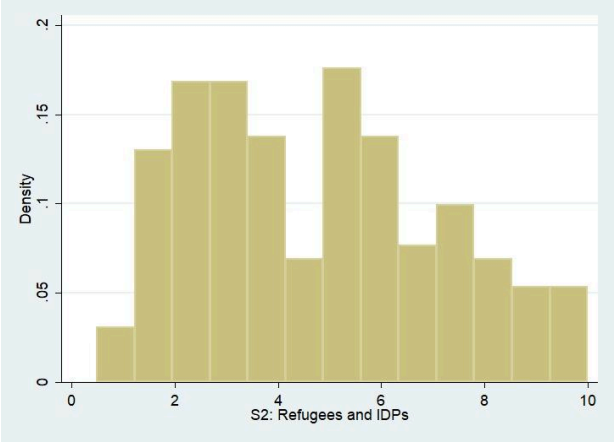
Figure: Examining the residuals plot(s) to assess whether the relationship between the residuals and the predicted values or independent variables appears to be random and evenly spread out around zero.

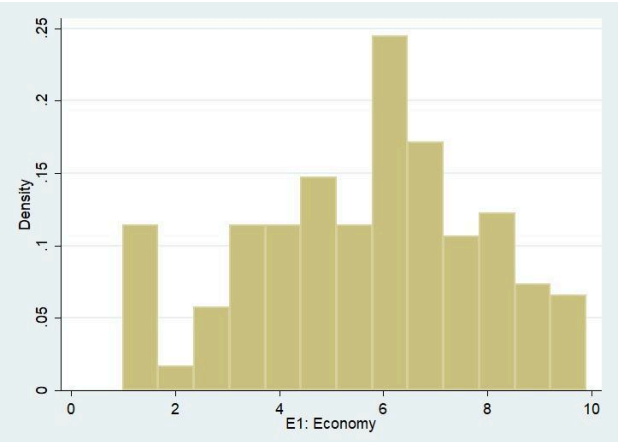
So, a random **scatter of points around zero** suggests that the assumptions of linearity and constant variance are met.

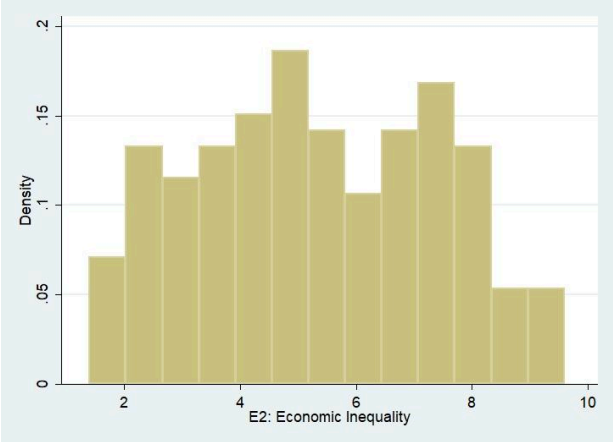
There should be enough variation in Xi to be qualified as an explanatory variable.

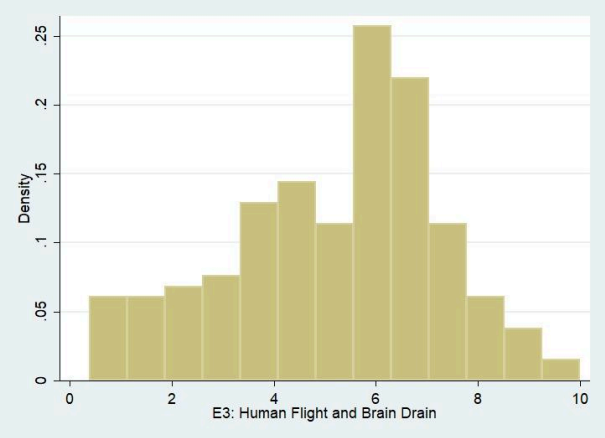


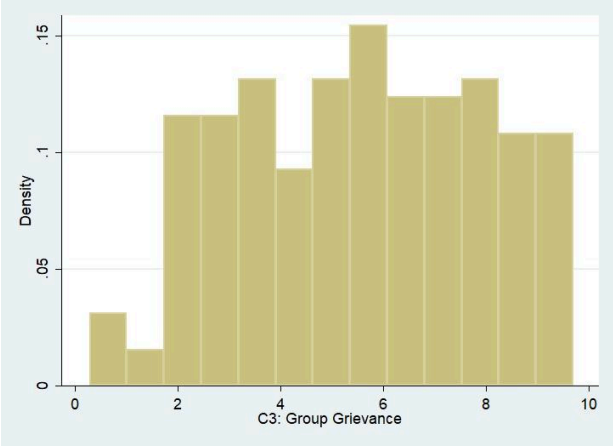


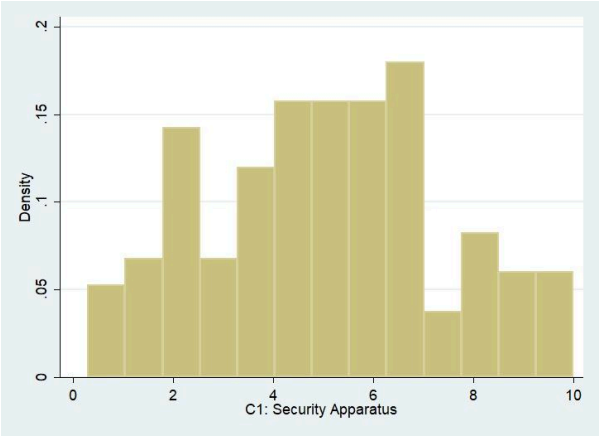












It is evident from the above discussion that there is enough variation in the values of minimum, maximum, mean, and standard deviation, indicating sufficient variability in the distribution. This suggests that the independent variable is strong.

The covariance between Xi and Ui is zero.

Based on the correlation matrix, it appears that there are no significant correlations between the independent variables and the residuals. This suggests that the assumption of zero covariance between the independent variables and the residuals holds in our regression model.

```
. corr E2EconomicInequality S1DemographicPressures S2RefugeesandIDPs C3GroupGrievance E3HumanFligh
> tandBrainDrain E1Economy C1SecurityApparatus C2FactionalizedElites X1ExternalIntervention
(obs=179)
```

	E2Econ~y	S1Demo~s	S2Refu~s	C3Grou~e	E3Huma~n	E1Econ~y	C1Secu~s	C2Fact~s	X1Exte~n
E2Economic~y	1.0000								
S1Demograp~s	0.8537	1.0000							
S2Refugees~s	0.5934	0.6797	1.0000						
C3GroupGri~e	0.4618	0.4999	0.6179	1.0000					
E3HumanFli~n	0.5558	0.6154	0.5736	0.3825	1.0000				
E1Economy	0.7304	0.7391	0.6680	0.4537	0.7220	1.0000			
C1Security~s	0.6993	0.7409	0.6944	0.6301	0.6094	0.7047	1.0000		
C2Factiona~s	0.6477	0.6793	0.6581	0.7096	0.5473	0.6429	0.7506	1.0000	
X1External~n	0.6044	0.6064	0.6864	0.4472	0.7321	0.7573	0.6722	0.6911	1.0000

The correlations between the independent variables are significant. Generally, correlations below 0.3 or 0.4 are considered weak.

8. Conclusion

The Stata outputs provide insights into the relationships between economic inequality and various socio-economic, political, and demographic factors. The multiple linear regression model identifies demographic pressures and economic conditions as significant positive predictors of economic inequality, while the effects of refugee/IDP situations and human flight/brain drain are negative but not statistically significant.

The diagnostic tests indicate that the model assumptions of homoscedasticity and absence of severe multicollinearity are reasonably met, lending credibility to the regression results. However, it is essential to interpret these findings within the context of the specific research question and to consider potential limitations or confounding factors not accounted for in the model.

Overall, this analysis offers a quantitative exploration of the determinants of economic inequality, highlighting the complex interplay between demographic, economic, and socio-political factors. Further research and validation may be warranted to gain deeper insights and inform policy decisions aimed at addressing economic disparities.