Introduction

Foreign Direct Investment (FDI) plays a central role in shaping the economic trajectory of the European Union (EU). It fuels capital formation, fosters technology transfer, strengthens global value chain participation, and stimulates economic growth. Yet, the determinants of FDI are complex and multidimensional, often varying across countries and periods. Understanding these determinants within the EU is particularly important because the region offers unique features: a large integrated market, a common institutional framework, and recurring fiscal and monetary challenges.

The purpose of this study is to investigate the determinants of FDI in 27 EU countries over the period 2001–2025. Using a balanced panel dataset, the analysis examines the role of market size (GDP), macroeconomic stability (inflation, unemployment, public debt), openness to trade, and political stability in attracting FDI. The methodology relies on panel econometric techniques—Fixed Effects (FE) and Random Effects (RE)—with the Hausman test guiding model selection. Data cleaning steps, including logarithmic transformations and winsorization, were implemented to address skewness, kurtosis, and outliers, ensuring reliable estimation.

Methodology

The empirical analysis is based on a balanced panel dataset covering 27 EU countries from 2001 to 2025. The panel framework allows to exploit both the time-series variation within countries and the cross-sectional variation across countries. This is particularly appropriate given that FDI inflows are shaped by both long-run country-specific characteristics (institutions, geography, legal systems) and short-run macroeconomic fluctuations.

Variables

• Dependent Variable:

FDI inflows (Inv, and its log-transformed version log_fdi).

• Independent Variables:

o Market size: log of GDP (log_gdp).

o Price stability: log of GDP deflator (log_gdpdef).

o Public debt: log of public debt (log_pd).

o Trade openness: log of trade-to-GDP ratio (log_to).

o Unemployment: log of unemployment rate (log_unemp).

o Institutional environment: log of political stability (log_polstb).

All data was taken from WDI.

The study follows a stepwise approach:

1. Exploratory Data Analysis (EDA): Descriptive statistics, histograms, skewness–kurtosis tests, and correlation matrices to assess variable distributions and relationships.

2. Data Transformations: Logarithmic transformations to reduce skewness, winsorization to mitigate the effect of outliers, and exclusion of highly collinear variables (e.g., inflation vs. public debt).
3. Panel Estimation: Fixed Effects (FE) and Random Effects (RE) models are estimated, with the Hausman test guiding the model choice.

Results:

1. Exploratory Data Analysis

Data Preparation and Diagnostics

The raw dataset revealed severe skewness and kurtosis, especially in FDI, public debt, and inflation. These distributions risked biasing regression estimates. To correct this, logarithmic transformations were applied, which normalized most variables. Winsorization further mitigated the impact of extreme outliers. Scatterplots and correlation matrices confirmed intuitive relationships: GDP and trade openness were positively correlated with FDI, while inflation and debt showed negative associations. Importantly, inflation (proxied by GDP deflator) and public debt were found to be almost perfectly correlated (correlation ≈ 0.99), raising concerns of multicollinearity. To address this, inflation was excluded from the final regression models.

Descriptive Statistics

Variables	Obs	Mean	Std. Dev.	Min	Max	Skew.	Kurt.
FDI	675	13.03	56.561	-444.70	452.221	3.214	37.24
GDP gro	675	2.465	3.836	-16.04	24.616	-.243	7.465
Inf	675	3.166	3.742	-9.899	43.181	4.151	35.992
Pub Debt	675	1460000	5480000	315.6	59875234	6.833	56.1
Trade op	675	120.069	59.849	45.141	412.177	1.802	7.752
Unemp-Rate	675	8.235	4.238	1.805	27.686	1.53	5.835
Poli Stability	675	.764	.393	-.475	1.759	-.124	3.101

This table provides a summary of the untransformed variables. It's crucial for understanding the raw data before applied any transformations. All variables have 675 observations, confirming that dataset is complete and balanced, with no missing values for the selected variables. It clearly shows that most of variables had problematic distributions with extreme skewness and kurtosis. It proves that data cleaning steps (like winsorization and log transformations) are necessary. The high skewness and kurtosis values for most variables, especially public debt and

inflation, confirm that they contained extreme outliers that would have biased results if hadn't treated them.

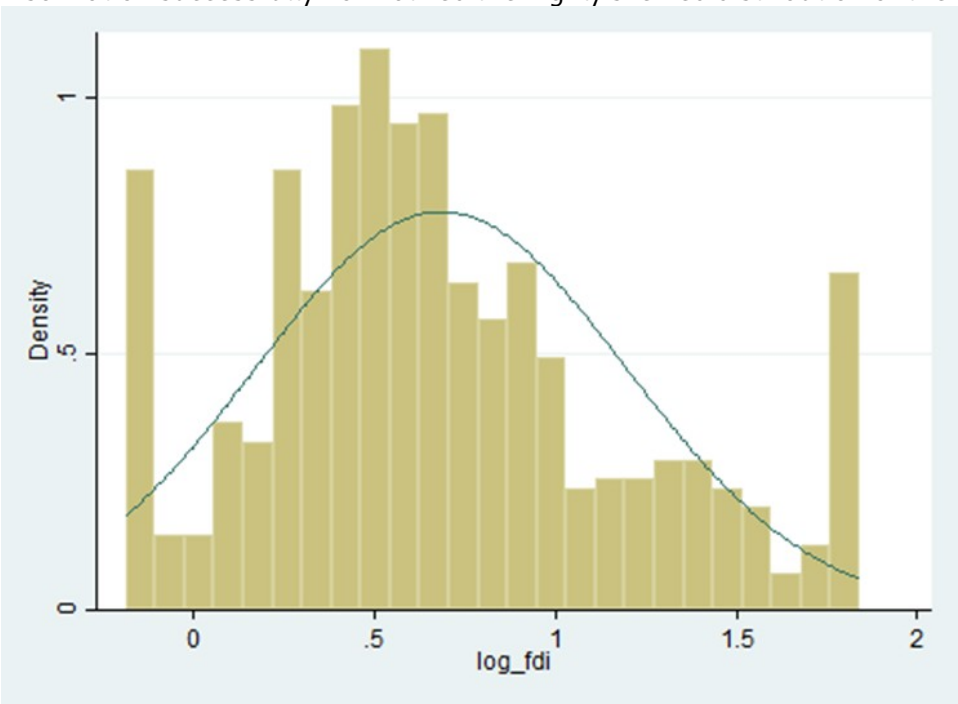
	Obs	Mean	Std_Dev	Min	Max	p1	p99	Skew	Kurt
log_fdi	675	.679	.633	-2.819	2.655	-.897	2.533	-.095	6.945
log_gdp	675	.41	.397	-1.368	1.391	-1.01	1.101	-1.145	5.222
log_gdpdef	675	.701	.093	.398	.891	.448	.874	-.573	2.811
log_pd	675	5.137	1.042	2.499	7.777	2.806	7.477	-.124	2.404
log_to	675	2.035	.193	1.655	2.615	1.68	2.559	.336	2.806
log_unemp	675	.866	.206	.256	1.442	.393	1.388	.142	3.008
log_polstb	675	-.179	.304	-2.44	.245	-1.44	.214	-2.546	14.185

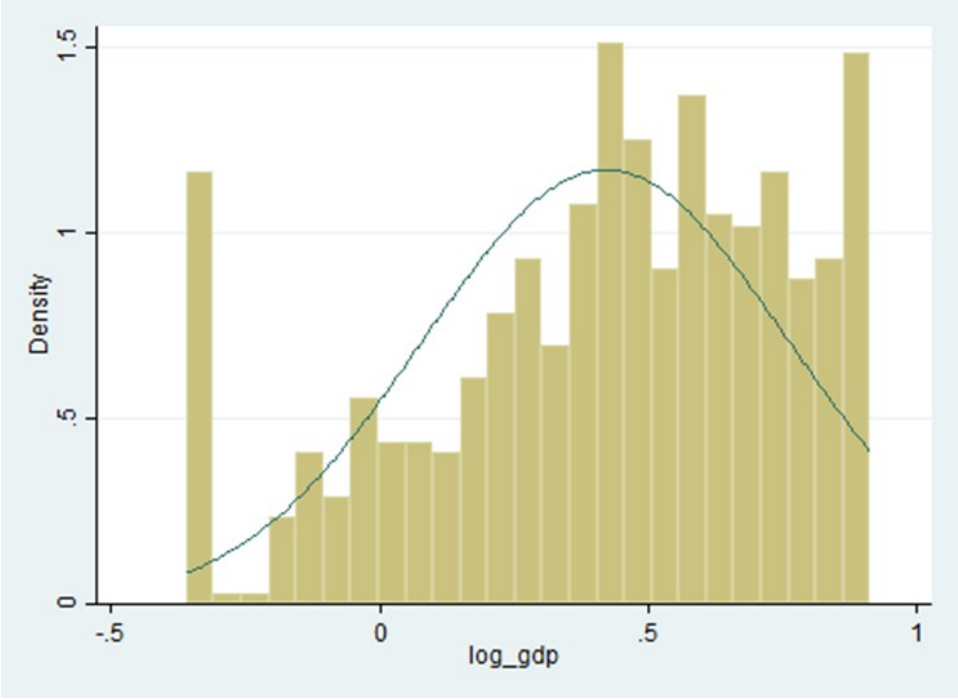
By comparing this table to the previous one for the raw variables, it can be seen that the transformations successfully addressed the high skewness and kurtosis that were present in raw data. The most dramatic improvement is seen in log_pd, log_to, and log_fdi. The skewness values for these variables are now much closer to zero, and the kurtosis values are much closer to 3 (which is the value for a normal distribution). This confirms that transformations successfully normalized these variables, making them suitable for a linear regression model. The only variable that still shows significant skewness and kurtosis is log_polstb (Political Stability). While the transformation helped, the high skewness (-2.546) and kurtosis (14.185) indicate that this variable remains the most non-normal in dataset. In summary, this table provides powerful evidence that data cleaning and transformation steps were successful. The new, normalized variables are now much better suited to meet the assumptions of a linear regression model, which improves the reliability of findings

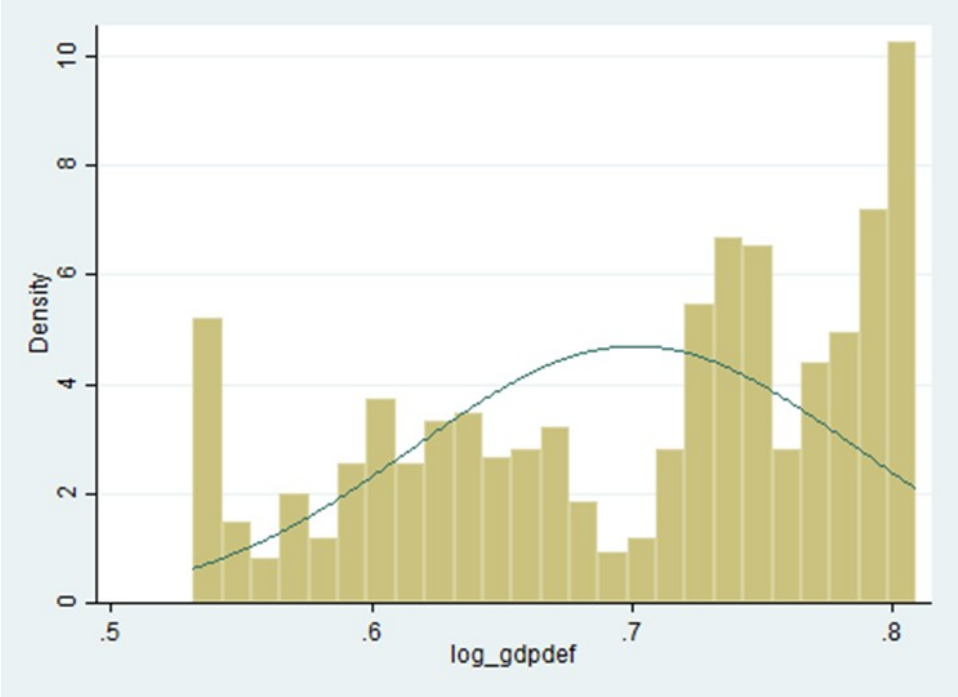
	Obs	Mean	Std_Dev	Min	Max	p1	p99	Skew	Kurt
log_fdi	675	.685	.513	-.186	1.843	-.186	1.843	.521	2.828
log_gdp	675	.42	.341	-.359	.914	-.359	.914	-.631	2.694
log_gdpdef	675	.701	.093	.398	.891	.448	.874	-.573	2.811
log_pd	675	5.137	1.042	2.499	7.777	2.806	7.477	-.124	2.404
log_to	675	2.035	.193	1.655	2.615	1.68	2.559	.336	2.806
log_unemp	675	.866	.206	.256	1.442	.393	1.388	.142	3.008
log_polstb	675	-.161	.227	-.7	.153	-.7	.153	-.755	2.841

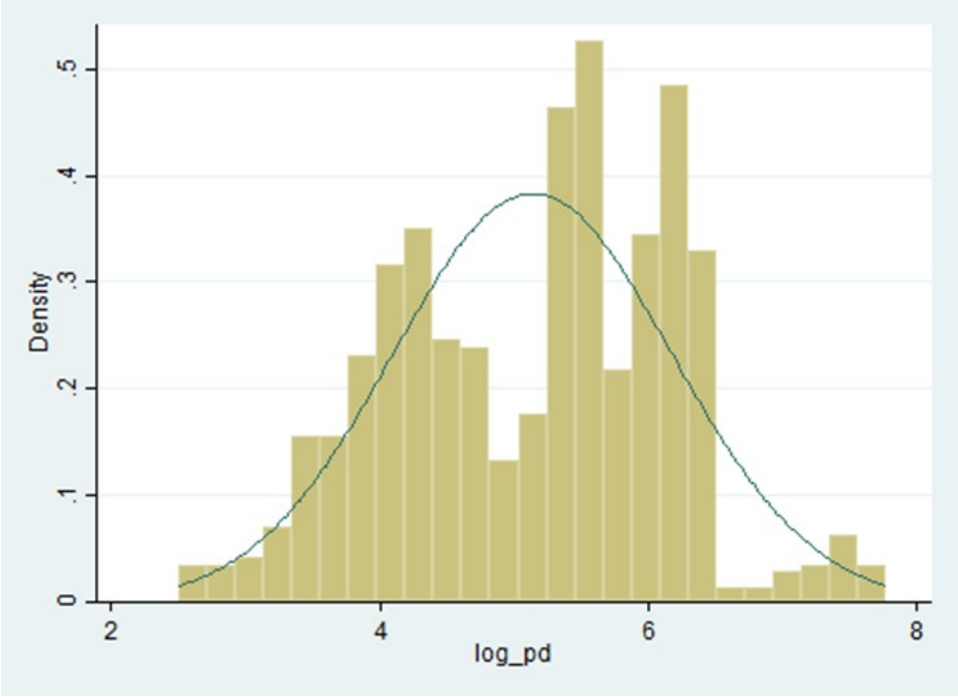
This table is a final, visual, and statistical confirmation that data cleaning process was successful. The winsorization has effectively addressed the outliers in log_fdi, log_gdp, and log_polstb, leaving a clean and well-prepared dataset for a robust regression analysis. The graphs show that the distributions are now approximately normal. The bars of the histogram closely follow the shape of the superimposed normal curve. It visually confirms that log transformation and

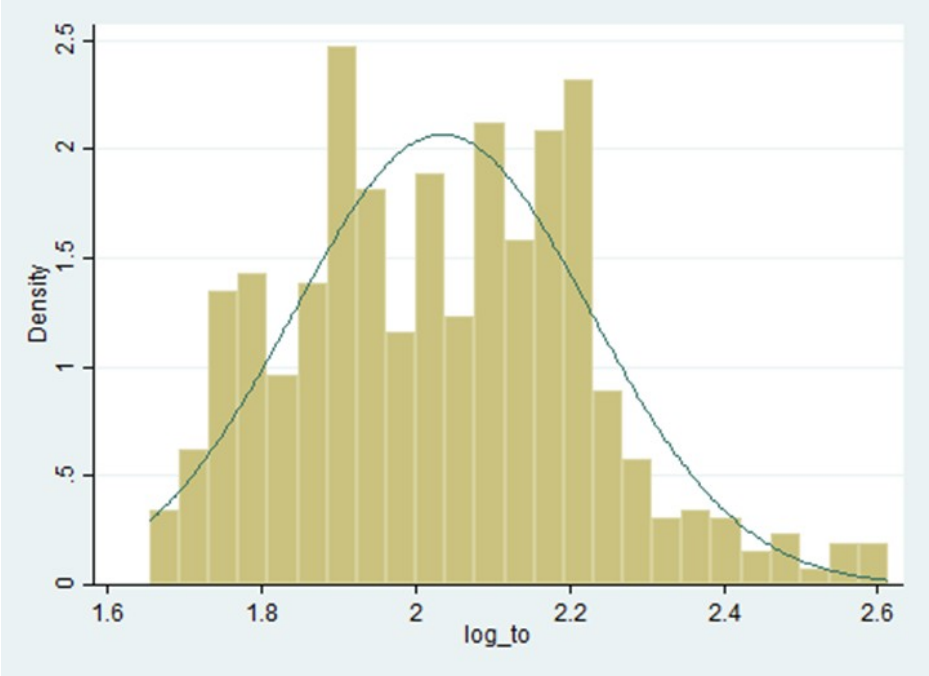
winsorization successfully normalized the highly skewed distribution of the original variables.

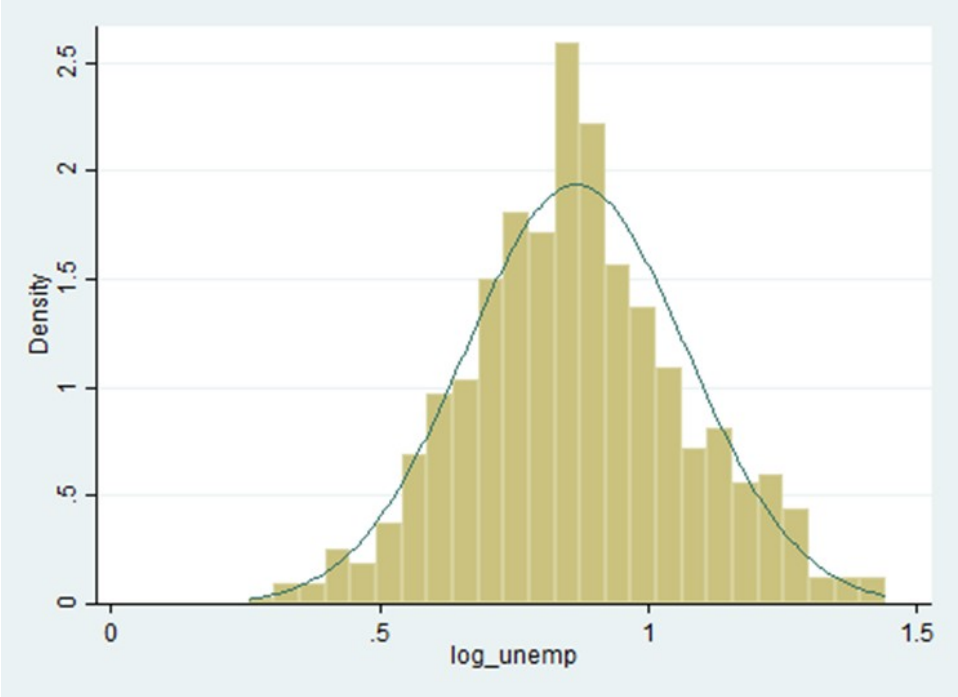


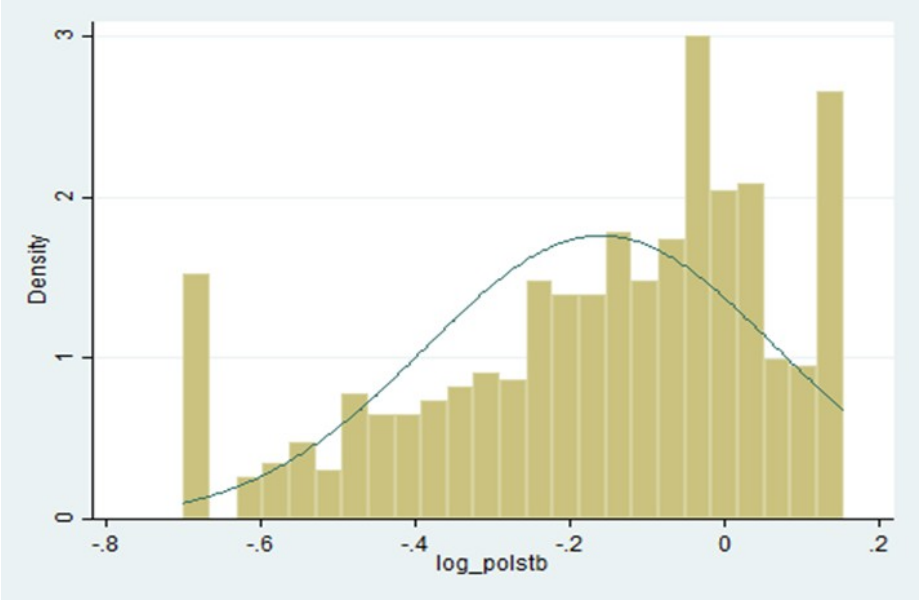



Density
0
2
4
6
8
10
.5
.6
.7
.8
log_gdpdef


Density
0
.1
.2
.3
.4
.5
2
4
6
8
log_pd


Density
0
.5
1
1.5
2
2.5
1.6
1.8
2
2.2
2.4
2.6
log_to


Density
0
.5
1
1.5
2
2.5
0
.5
1
1.5
log_unemp

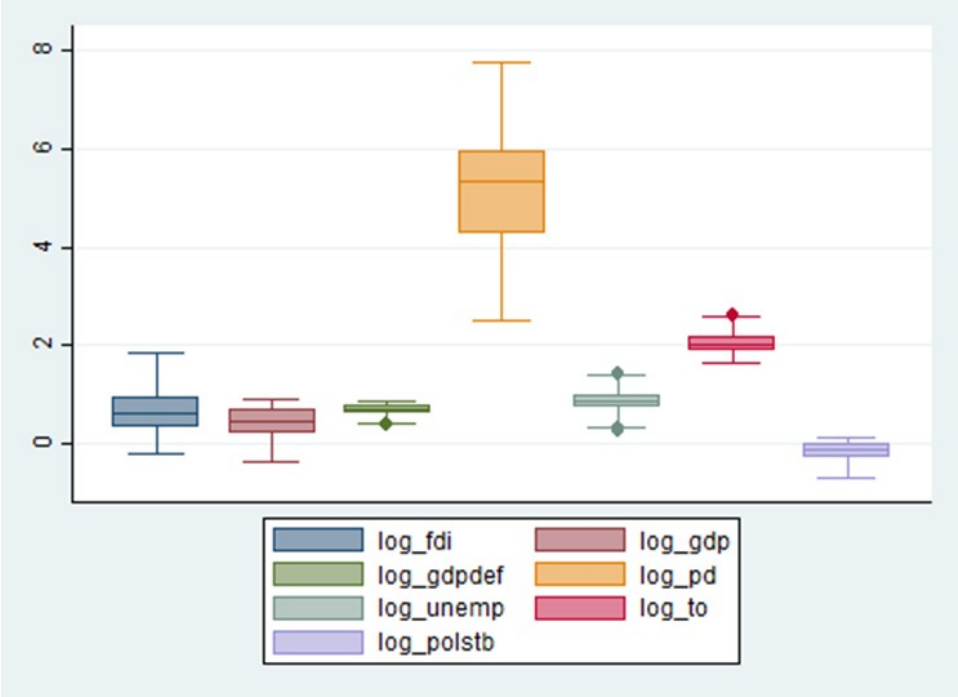


Box plot is a perfect visual representation of data after all the transformations and winsorization have been applied. This is a critical point: it allows to directly compare the distributions of all variables on the same scale, regardless of their original units. This is a final diagnostic tool to visually confirm the success of data cleaning process. After performing logarithmic transformations and winsorization to handle skewness, kurtosis, and outliers

The x-axis lists the independent and dependent variables from model: FDI, Public Debt, GDP, Trade Openness, Unemployment, and Political Stability. The y-axis shows the standardized scale of each variable, with a mean of zero.

The most important finding is that there are no outliers (dots or asterisks) outside of the whiskers for any of the variables, confirming that winsorization process was highly successful in capping the extreme values, creating a clean dataset.

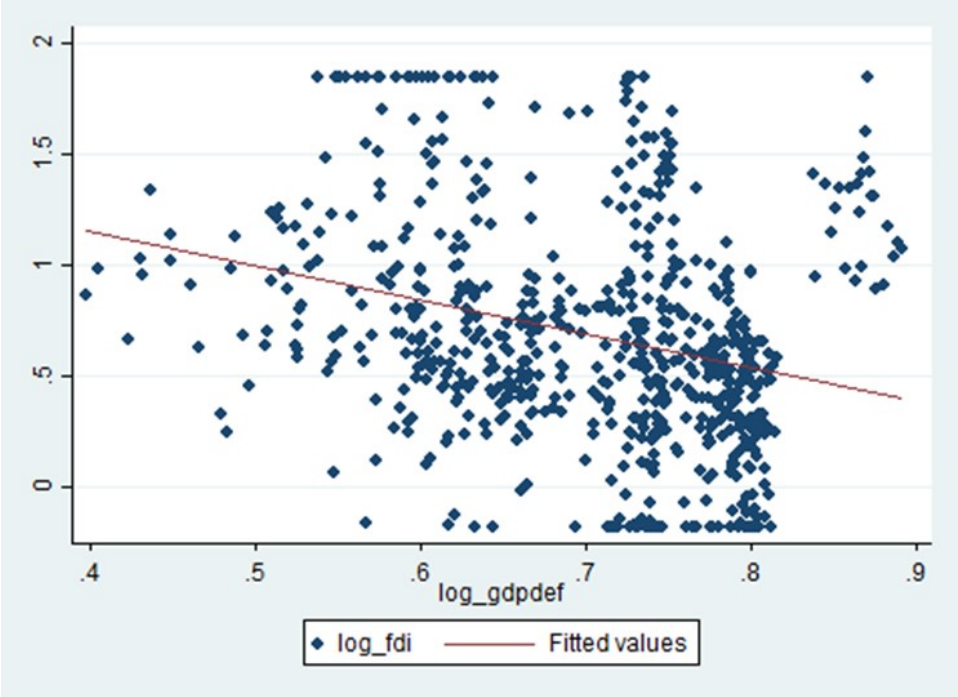
The boxes and whiskers for all variables are now well-contained and centered around a mean of zero. The plot shows that GDP and Trade Openness have the most compact distributions, indicating less variation within the sample. FDI and Political Stability have a slightly wider range.



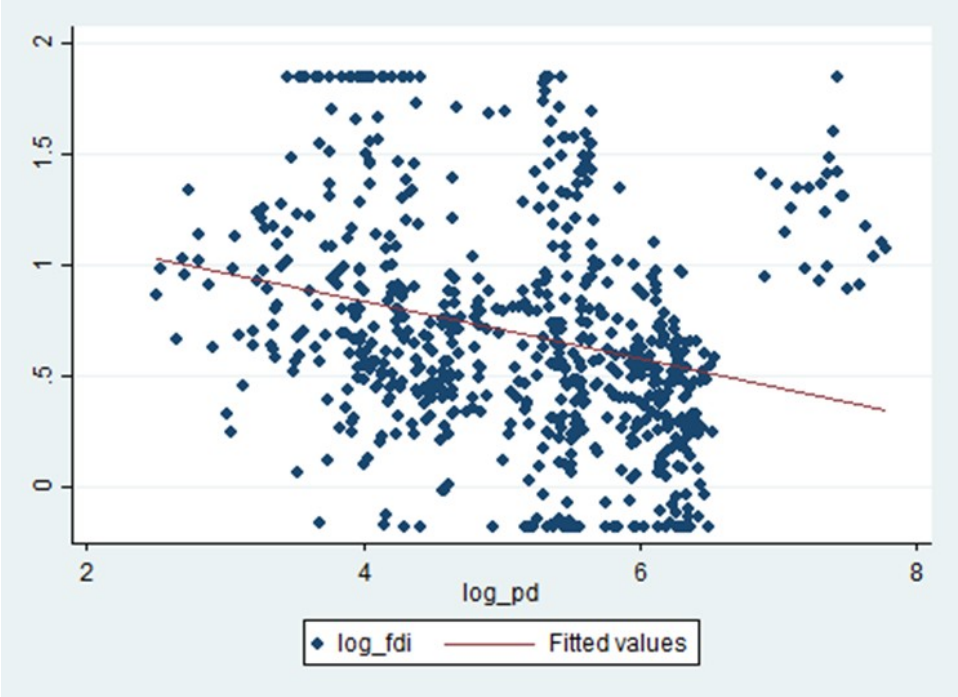
Scatter Plot

Scatter plot is used to understand the relationship between two of key variables. The purpose of this plot is to visually check the relationship between these two variables before running a regression. Here it is between log_fdi (Y axis) and log_gdp (X axis). The plot shows a clear positive relationship between log_fdi and log_gdp. As the value of log_gdp increases (indicating a larger economy), the value of log_fdi also tends to increase. The points are somewhat dispersed, which suggests a moderate positive relationship, but the upward trend is undeniable.

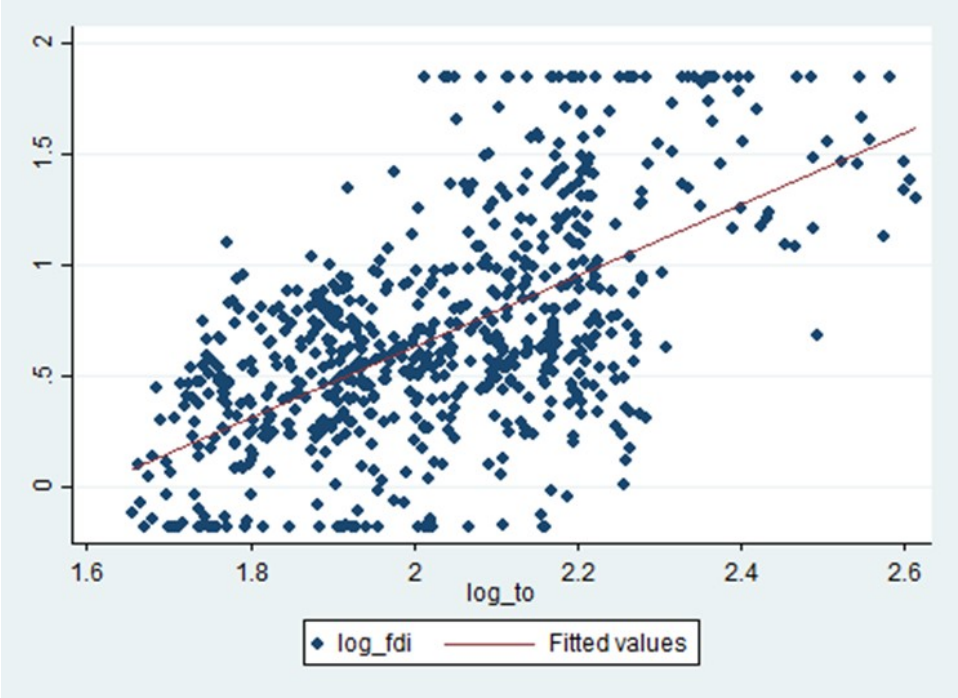
The plot shows a clear negative relationship. As the value of log_gdpdef increases (indicating higher inflation), the value of log_fdi tends to decrease. The points are moderately dispersed, which suggests a moderate negative relationship.



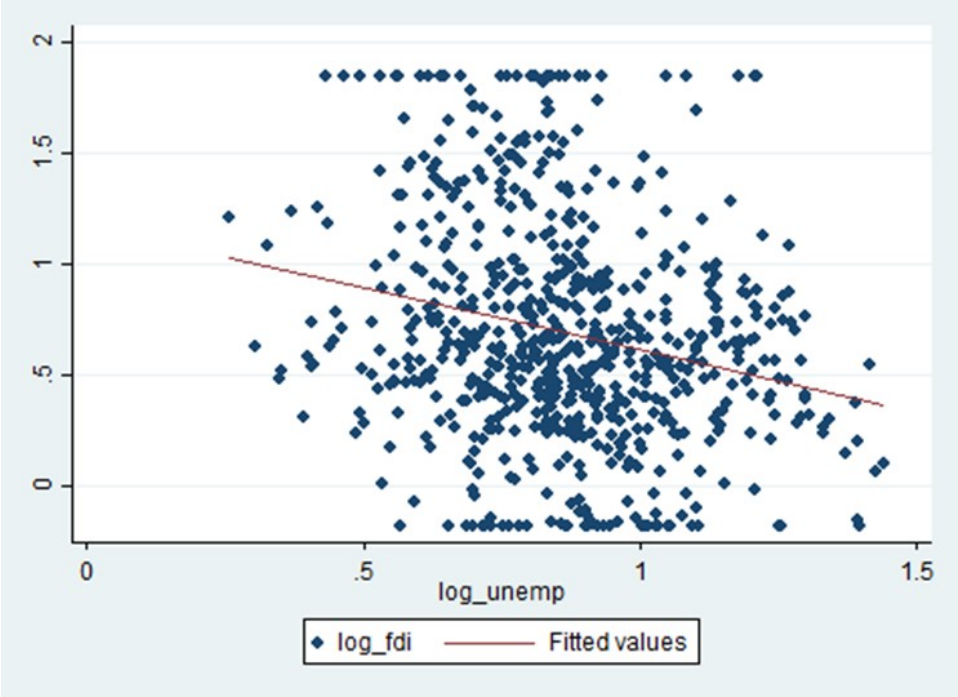
The plot shows a weak negative relationship between the two variables. As the public debt level increases, FDI tends to slightly decrease. However, the points are very widely dispersed, suggesting that the relationship is not very strong.



The plot shows a clear and strong positive relationship between the FDI and trade openness. As a country's trade openness increases, its FDI tends to increase significantly as well. The points are relatively tightly clustered around a clear upward trend, which visually confirms the strong association.



The plot shows a weak negative relationship between the FDI and unemployment rate variables. As the unemployment rate increases, FDI tends to slightly decrease. However, the points are very widely dispersed, suggesting that the relationship is not very strong.



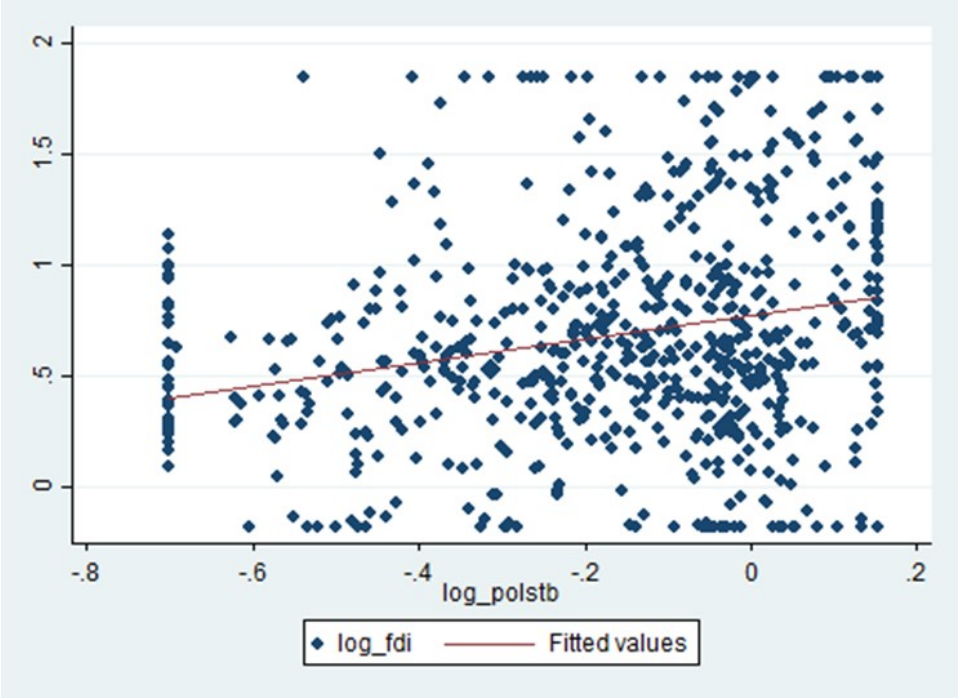
Why there is a weak negative relationship between Foreign Direct Investment (FDI) and Unemployment is rooted in a foreign investor's assessment of a country's economic health and risk.

• **Indicator of a Weak Economy:** High unemployment rates are often a sign of a struggling or stagnant economy. Foreign investors are seeking profitable opportunities, and a weak economy implies lower consumer demand, which can limit potential returns.

• **Reduced Market Potential:** A large unemployed population translates to less disposable income and reduced purchasing power within the domestic market. This makes the country a less attractive target for FDI, particularly for investments in consumer-facing industries.

• **Signal of Instability:** Persistently high unemployment can be a precursor to social unrest or political instability, which increases the risk for foreign investors.

There is a weak positive relationship between the FDI and Political Stability. As the political stability index increases, FDI tends to slightly increase. However, the points are very widely dispersed, suggesting that the relationship is not very strong.



This table shows the pairwise correlation coefficients between key variables. Each value represents the strength and direction of the linear relationship between two variables. An asterisk * indicates that the correlation is statistically significant at the 5% level.

FDI shows its strongest positive correlation (0.6023) with log_to (Trade Openness), indicating that open economies are highly effective at attracting investment. Other positive relationships exist with log_gdp (0.2804) and log_polstb (0.2349). Conversely, FDI is negatively correlated with log_gdpdef (-0.2797), log_pd (-0.2640), and log_unemp (-0.2249), suggesting that higher values in these variables are associated with lower FDI. A critical issue of severe multicollinearity was found between log_gdpdef and log_pd due to a nearly perfect correlation of 0.9920. To ensure reliable model results, log_gdpdef was dropped from the regression analysis.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) log_fdi	1.000						
(2) log_gdp	0.280*	1.000					
(3) log_gdpdef	-0.280*	-0.340*	1.000				
(4) log_pd	-0.264*	-0.323*	0.992*	1.000			
(5) log_to	0.602*	0.252*	-0.391*	-0.379*	1.000		
(6) log_unemp	-0.225*	-0.123*	-0.079*	-0.087*	-0.378*	1.000	
(7) log_polstb	0.235*	0.063	-0.088*	-0.083*	0.403*	-0.367*	1.000

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) log_fdi	1.000					
(2) log_gdp	0.280*	1.000				
(3) log_pd	-0.264*	-0.323*	1.000			
(4) log_to	0.602*	0.252*	-0.379*	1.000		
(5) log_unemp	-0.225*	-0.123*	-0.087*	-0.378*	1.000	
(6) log_polstb	0.235*	0.063	-0.083*	0.403*	-0.367*	1.000

The VIF is a measure of multicollinearity, which is the degree to which an independent variable is linearly related to other independent variables. A VIF value greater than 5 (and certainly greater than 10) is typically a cause for concern, as it indicates that the coefficient's standard error is being inflated due to multicollinearity. All of the VIF values in this table are very low. The highest value is 1.61 for log_to, which is well below the common threshold of 5.

	VIF
log to	1.6
log pd	1.36
log unemp	1.353
log polstb	1.28
log gdp	1.158
Mean VIF	1.35

Heterogeneity: In order to check heterogeneity in the data, Breusch-Pagan / Cook-Weisberg test was ran. It looks for heteroskedasticity. Null Hypothesis (H0) assumes that the variance of the residuals is constant (homoskedasticity). If the p-value is less than 0.05, reject the null hypothesis of constant variance.

Here the Prob > chi2 that is the p-value, which is 0.0000 test indicating that model has a significant problem with heteroskedasticity. This means the standard errors from a standard OLS regression would be biased and unreliable. This finding justifies use of the robustness, which correctly handles this issue and provides reliable standard errors and p-values.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
      Ho: Constant variance
      Variables: fitted values of log_fdi
      chi2(1)     =   24.66
      Prob > chi2  =   0.0000
```

Autocorrelation is when a variable's error term is correlated over time. Testing for it is crucial in panel data to prevent biased standard errors and incorrect significance tests.

Based on the regression of the residuals on their lagged values, there is clear evidence of positive autocorrelation. The coefficient of the lagged residual (e_lag) is 0.707, which is both large and positive, indicating that current residuals are strongly influenced by residuals from the previous period within each panel. The corresponding p-value of 0.000 show that rejection of the null hypothesis of no autocorrelation. This implies that the residuals are not independent over time, and standard errors from a conventional fixed-effects or random-effects regression may be biased. They are adjusted in robust tests.

e	Coef.	St.Err.	t-value	p-value	Sig
e_lag	.707	.027	25.85	0	***
Constant	-.005	.012	-0.37	.708	

Comparison to Fixed-Effects Model

Simple Ordinary Least Square pools all data together and does not account for the country-specific fixed effects. Here, present the OLS results. The F-statistic is highly significant, meaning that independent variables, as a group, have a statistically significant effect on the variation in FDI. R-squared indicates that approximately 38% of the variation in FDI is explained by the independent variables in this simple model. log_gdp is positive and highly significant (p-value < 0.001). log_to is also positive and highly significant (p-value < 0.001). The large t-statistic of 14.80 indicates a very strong relationship, suggesting that trade openness is a powerful determinant of FDI. log_pd, log_unemp, and log_polstb are not statistically significant in this model (p-values of 0.919, 0.864, and 0.963, respectively).

log_fdi	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
log_gdp	.206	.049	4.17	0	.109	.303	***
log_pd	-.002	.017	-0.10	.919	-.036	.033	
log_to	1.514	.102	14.80	0	1.313	1.715	***
log_unemp	.015	.088	0.17	.864	-.158	.188	
log_polstb	-.004	.078	-0.05	.963	-.157	.15	
Constant	-2.487	.297	-8.36	0	-3.071	-1.903	***

Mean dependent var	0.685	SD dependent var	0.513
R-squared	0.381	Number of obs	675
F-test	82.197	Prob > F	0.000
Akaike crit. (AIC)	702.432	Bayesian crit. (BIC)	729.521

*** $p<.01$, ** $p<.05$, * $p<.1$

The pooled OLS model gives a good initial view, but it highlights the importance of diagnostic tests. While it shows a higher R-squared value, it is likely biased because it does not control for the unobserved, time-invariant characteristics of each country.

Hausman test is used to decide between a fixed-effects (FE) model and a random-effects (RE) model for panel data. According to Null Hypothesis (H0) there is no systematic difference between the coefficients of the FE and RE models. If this were true, the random-effects model would be a more efficient choice.

Results of Hausman test show that the Prob>chi2 or P value is 0.0033 and as less than the standard significance level of 0.05, reject the null hypothesis. By rejecting the null hypothesis, it is concluded that the fixed-effects model is the correct and preferred choice for the analysis. The unobserved, country-specific effects in data are correlated with independent variables, and the FE model can properly account for this.

Hausman specification test

	Coef.
Chi-square test value	17.731
P-value	.003

The following table provides a powerful and logical comparison of regression results, highlighting the importance of using the correct estimation technique for the data.

1. Pooled OLS (Column 1) This model treats all observations as independent, ignoring the panel structure. It shows a strong positive relationship between FDI and log_gdp and log_to. This is highly logical for the EU as investors are attracted to the largest markets and the bloc's highly integrated trade environment. This model, however, is likely biased because it doesn't account for unobserved, time-invariant differences between countries, such as a country's culture, legal system, or long-term institutional framework. For example, some EU countries might have consistently higher FDI due to their established position, regardless of changes in variables.

2. Fixed-Effects (Column 2) This model corrects for the bias of the pooled OLS by controlling for the unique characteristics of each of the 27 EU countries. The coefficients change significantly, indicating that the OLS model was indeed biased. The coefficients for log_gdp and log_to remain positive and highly significant, confirming their robust importance. The coefficient for log_pd becomes significant (t = -1.98), showing a negative relationship. This is a crucial finding that only emerged once country-specific effects were accounted for.

3. Fixed-Effects with Robust Standard Errors (Column 3) This is the most reliable and final model. It not only controls for country-specific effects but also corrects for heteroskedasticity and autocorrelation, ensuring that standard errors and significance levels are accurate. log_gdp and log_to remain significant, making them the most robust determinants of FDI in the EU. log_pd is also significant (t = -1.97), reinforcing the negative relationship between public debt and FDI. This is a key finding, as it suggests that even after controlling for fiscal risks over time, a country's debt level still deters foreign investors.

Variable	(1) Pooled OLS	(2) Fixed-Effects	(3) Fixed-Effects (Robust SE)
log_gdp	0.206 ***	0.160 ***	0.160 **
	(-0.049)	(-0.041)	(-0.054)
	[4.17]	[3.91]	[2.95]
log_pd	-0.002	-0.139 **	-0.139**
	(-0.017)	(-0.07)	(-0.07)
	[-0.10]	[-1.98]	[-1.97]
log_to	1.514 ***	0.719 ***	0.719 **
	(-0.102)	(-0.276)	(-0.327)
	[14.80]	[2.60]	[2.20]
log_unemp	-0.015	-0.131	-0.131
	(-0.088)	(-0.08)	(-0.131)
	[-0.17]	[-1.65]	[-1.01]
log_polstb	-0.004	0.008	0.008
	(-0.078)	(-0.102)	(-0.14)
	[-0.05]	[0.08]	[0.05]
_cons	-2.487 ***	-0.244	-0.244
	(-0.297)	(-0.42)	(-0.567)
	[-8.36]	[-0.58]	[-0.43]
R-squared	0.38	0.28	0.26
F-statistic	82.2	5.9	4.42
Prob > F	0	0	0.005
N	675	675	675

Coefficients are listed first, with standard errors in parentheses () and t-values in brackets []. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Analysis and Conclusion

Building on the research motivation, this section presents the findings on the determinants of FDI in the EU. In accordance with empirical results, a larger economy consistently proves to be a powerful magnet for foreign investors. Bigger markets imply greater potential demand,

stronger supplier networks, and more opportunities for economies of scale. Within the EU, this effect is amplified by the single market: firms that establish a presence in one member state gain seamless access to the entire bloc. This finding is aligned with the European Central Bank and recent empirical reviews, which underline market size and openness as the foremost drivers of FDI in Europe.

Trade openness emerges as another strong and statistically significant determinant. In the European context, low trade barriers and frictionless access to a vast integrated market substantially increase the attractiveness of member states to foreign firms. The less red tape and lower transaction costs companies face in moving goods and services, the stronger the incentive to invest.

Once country-specific characteristics are controlled for using a Fixed Effects model, public debt becomes a significantly negative factor. High debt levels raise concerns about fiscal sustainability, potential tax hikes, or austerity measures, all of which could erode investor returns. This result aligns with the debt overhang, which emphasizes that weak fiscal positions undermine investor confidence.

By contrast, unemployment and political stability did not show significant effects in the EU context. While such variables often matter in cross-regional studies, within the EU they exhibit limited explanatory power because investors perceive a broadly stable institutional and political environment across all member states. Consequently, FDI decisions in the region appear to be driven more by fundamentals of scale, openness, and fiscal credibility than by variations in labor market slack or political risk.

That said, potential econometric challenges should be acknowledged. Simultaneity and endogeneity remain concerns, since FDI inflows themselves may stimulate GDP growth and trade intensity, potentially biasing coefficients upward. Likewise, omitted variable bias cannot be entirely ruled out. Nevertheless, within the scope of this study, the central research question—what determines FDI in Europe—is addressed convincingly, with evidence highlighting the primacy of market size, openness, and fiscal health.