Do Socioeconomic Determinants of Freedom Differ Across Income Groups? An Inferential and Nonparametric Modeling Approach Using Country-Level Data

Advanced Data Analysis Final Project

Jingyu Hu jh4862
Kai Xu kx2215
Liang-Po Yen ly2665
Yifei Duan yd2773
Sikai Gao sg4520

Introduction

Freedom is recognized as a multidimensional construct shaped by a country's socioeconomic environment. Understanding the determinants of human freedom across nations is crucial to those who care about individual liberties. However, the extent to which these determinants affect freedom may vary across different economic contexts. This project investigates the central question: Do the effects of socioeconomic determinants on human freedom vary across income groups?

To situate our empirical analysis within a broader theoretical and empirical framework, we draw on five recent studies that examine the complex interplay between human freedom, economic growth, institutional quality, and corruption. Collectively, these works underscore that freedom is not merely a normative ideal but a significant driver of development, whose effects vary by economic context and institutional structure. Makrevska Disoska and Shapkova Kocevska find that in Eastern Europe, higher levels of freedom are strongly associated with increased GDP per capita, offering no evidence of a trade-off between liberty and prosperity (Disoska). Yen's Granger causality analysis reveals a bidirectional relationship between freedom and economic growth in certain regions, moderated by generalized trust, echoing our findings of context-specific interactions (Kocevska). Abdelrahim's global study links personal and economic oppression to higher corruption levels, with political instability and governance weakness acting as key mediators—findings that parallel the significance of political stability in our model (Abdelrahim). Lastly, Dolan dissects the Human Freedom Index into its personal and economic components and finds a strong correlation between them, especially through legal institutions and trade openness, while cautioning that not all economic policies enhance freedom equally, highlighting the importance of structural factors in shaping freedom's effects (Dolan). Together, these studies reinforce our project's core insight: that the socioeconomic determinants of freedom are neither uniform nor linear, but depend critically on a nation's income level, governance capacity, and institutional foundations.

To practice these understandings, we employ both parametric and nonparametric modeling approaches using cross-national data that combine the Human Freedom Index with a range of macroeconomic indicators. Countries are stratified into low-, middle-, and high-income groups to account for economic context. Our analysis begins with linear modeling, incorporating interaction terms to detect whether the association between GDP per capita, infant mortality, and political stability, and human freedom changes by income category. Diagnostic tests reveal violations of linear model assumptions. Because of group-specific effects, we extend the analysis using flexible nonparametric methods, including LOESS smoothers, weighted natural cubic splines, weighted generalized additive models, and random forests. These techniques allow us to relax linearity assumptions and capture nuanced, nonlinear patterns in the data. Through model comparison and robust inference techniques, we demonstrate that income-level interactions significantly alter the strength and shape of relationships between key predictors and freedom outcomes.

Data Collection, Description, and Processing

Figure 1: Scatter plot of Human Freedom Score vs. other Predictive Variables by Income Group

Two major sources of data are the COTA institution and the World Bank Organization. The datasets include economic, social, and political indicators for 165 countries from 2000 to 2022. Key variables include human freedom, GDP per capita, income inequality, unemployment rate, intentional homicides, internet usage, political stability, and social spending on education and health. The dataset has 3,795 records, capturing both country-level and year-by-year trends. Variables with missing values over 40%, namely literacy_rate, Gini_index, were removed. Observations with missing Human Freedom Index score values and outliers were also removed. Missing values were imputed with medians.

The scatter plot above reveals distinct patterns in the relationship between human freedom and its socioeconomic determinants across income groups. Figure 1 in the Appendix illustrates that besides the Human

Freedom Index and Log GDP with roughly normal distribution centered around their means, other variables are skewed to different levels. Most other variables are skewed to the right: Infant mortality shows a strong right skew with most values below 50 but outliers reaching above 80; Intentional homicides exhibit a similar trend with a long tail, indicating that while most countries have low homicide rates, a few have extremely high values; Internet usage is also right-skewed with most countries having less than 40% usage yet some over 80%; Education expenditure, Medical expenditure, and Unemployment rate are more of a bell-shape with a mild right tail. The only left-skewing variable is CPI with a heavy tail, spanning a wide range from around 20 to near 100. Political stability is almost uniform, ranging from 0 to 100. This diversity in distribution types suggests the need for normalization or transformation for some variables in statistical modeling.

Figure 2 in the Appendix shows a clear stratification, with high-income countries exhibiting significantly higher median freedom scores than middle- and low-income groups, accompanied by less variation and fewer extreme low outliers. Figure 3 in the Appendix confirms this trend, suggesting that economic status correlates positively with both the level and stability of freedom. Figure 4 in the Appendix further quantifies these associations: HFI is most strongly correlated with internet use at correlation coefficient about 0.63, political stability at correlation coefficient about 0.63, and log GDP per capita at correlation coefficient about 0.63, while showing a strong negative relationship with infant mortality at correlation coefficient about -0.58, which is an inverse proxy of development. Notably, education expenditure shows a relatively weak correlation with freedom at correlation coefficient about 0.19, consistent with our later regression findings. Figure 3 also illustrates heterogeneous slope patterns: internet usage and medical expenditure display steeper positive associations with freedom in high-income countries, while the effect of intentional homicides on freedom is more pronounced in middle-income nations. Political stability shows a fairly consistent positive trend across all groups, suggesting its universally beneficial role. These visualizations support our hypothesis that the impact of socioeconomic factors on human freedom is not uniform but varies significantly by income level, warranting flexible modeling approaches to capture these interactions.

Statistical Modeling (Linear Models)

Our initial hypothesis posited that the effect of socioeconomic determinants on human freedom varies across income levels. The null hypothesis corresponded to the baseline linear model, which includes socioeconomic variables — log GDP per capita, CPI, infant mortality, unemployment rate, internet use, education expenditure, medical expenditure, intentional homicides, and political stability — along with the three-category income group variable, without interactions. The alternative hypothesis corresponded to the interaction model that incorporates interaction terms between each socioeconomic predictor and the three income group categories, in addition to the main effects included in the baseline model. Comparing these two models using ANOVA revealed that the inclusion of interaction terms significantly improves the model's explanatory power (p-value < 2.2e-16), providing strong evidence that the relationship between socioeconomic determinants and freedom differs across income groups.

Figure 5 in Appendix presents the residual diagnostic plots used to assess whether the assumptions of linear regression are satisfied for the interaction model. The Residuals vs Fitted plot displayed a mild fanning pattern, and the Scale–Location plot showed a modest downward trend, indicating notable heteroscedasticity. The Breusch-Pagan test further confirmed the presence of heteroscedasticity (p-value < 2.2e-16). The Normal Q–Q plot indicated slight deviations at the tails, suggesting some degree of non-linearity. Multicollinearity was evaluated using Generalized Variance Inflation Factors (GVIFs). The results indicated moderate multicollinearity among predictors, with GVIF^(1/(2*df)) values falling below 5, the commonly accepted threshold. Despite the inclusion of many potentially correlated socioeconomic indicators and their interactions with income groups, multicollinearity does not appear to pose a serious threat to the model's validity.

To account for the presence of heteroscedasticity, robust standard errors were used for coefficient inference in the final linear model. This ensures that the resulting p-values and confidence intervals remain valid and reliable. Note that middle-income countries account for approximately 49% of the dataset, compared to 37% high-income and only 11% low-income countries. Due to this imbalance in the number of observations across income groups, observation-level weights were introduced to prevent the overrepresented group from disproportionately influencing the estimation results. Table 1 shows the output of the final linear model estimated with robust standard errors and group weights.

Table 1: Results of Linear Regression Model with Interactions

Term	Estimate	Std. Error	t value	Pr(>|t|)	Signif.
(Intercept)	3.5941	0.3233	11.117	<2e-16	***
log GDP	-0.0273	0.0649	-0.421	0.6739	
CPI	0.0183	0.0025	7.471	1.65e-13	***
Infant Mortality	0.0016	0.0034	0.485	0.6276	
Unemployment Rate	0.0452	0.0106	4.261	2.22e-05	***
Internet Use	-0.0303	0.0117	-2.581	0.0100	**
Education Expenditure	-0.0349	0.0316	-1.105	0.2693	
Medical Expenditure	0.0382	0.0230	1.661	0.0971	.
Intentional Homicides	0.1173	0.0163	7.193	1.19e-12	***
Political Stability	0.0176	0.0031	5.621	2.43e-08	***
High Income Group	6.8074	0.9856	6.907	8.51e-12	***
Middle Income Group	3.6656	0.6234	5.880	5.50e-09	***
log GDP × High Income	-0.0491	0.1115	-0.441	0.6594	
log GDP × Middle Income	-0.1472	0.0961	-1.532	0.1259	
CPI × High Income	-0.0496	0.0059	-8.400	<2e-16	***
CPI × Middle Income	-0.0157	0.0035	-4.451	9.47e-06	***
Infant Mortality × High Income	-0.1788	0.0166	-10.749	<2e-16	***
Infant Mortality × Middle Income	-0.0109	0.0043	-2.507	0.0123	*
Unemployment Rate × High Income	-0.0274	0.0165	-1.663	0.0965	.
Unemployment Rate × Middle Income	-0.0664	0.0124	-5.365	9.96e-08	***
Internet Use × High Income	0.0404	0.0120	3.370	0.0008	***
Internet Use × Middle Income	0.0367	0.0125	2.936	0.0034	**
Education Expenditure × High Income	-0.0700	0.0498	-1.405	0.1602	
Education Expenditure × Middle Income	-0.0930	0.0418	-2.227	0.0262	*
Medical Expenditure × High Income	0.1472	0.0306	4.813	1.70e-06	***
Medical Expenditure × Middle Income	0.1051	0.0330	3.186	0.0015	**
Intentional Homicides × High Income	-0.0160	0.0196	-0.818	0.4135	
Intentional Homicides × Middle Income	-0.1091	0.0165	-6.609	6.12e-11	***
Pol. Stability × High Income	-0.0085	0.0037	-2.312	0.0210	*
Pol. Stability × Middle Income	0.0022	0.0037	0.603	0.5464	

Extension: Nonparametric Modeling

Due to the violation of homoscedasticity and linearity found in the diagnosis of the linear regression assumptions, and the income group is unbalanced, nonparametric models were developed with weights to investigate the relationship between the predictors and human freedom index. The LOESS smoother was first used to assess the shape of the association between each feature and the human freedom index, and the results showed that several predictors and distinct nonlinear patterns, including log gdp, infant mortality, unemployment rate, government expenditures on both education and health care, and intentional homicide. In contrast, the consumer price index (CPI), internet usage, and political stability present some near-linear trends. These LOESS plots indicate the modeling strategy of retaining appropriate linear specifications while introducing flexible structures for variables that exhibit nonlinear effects.

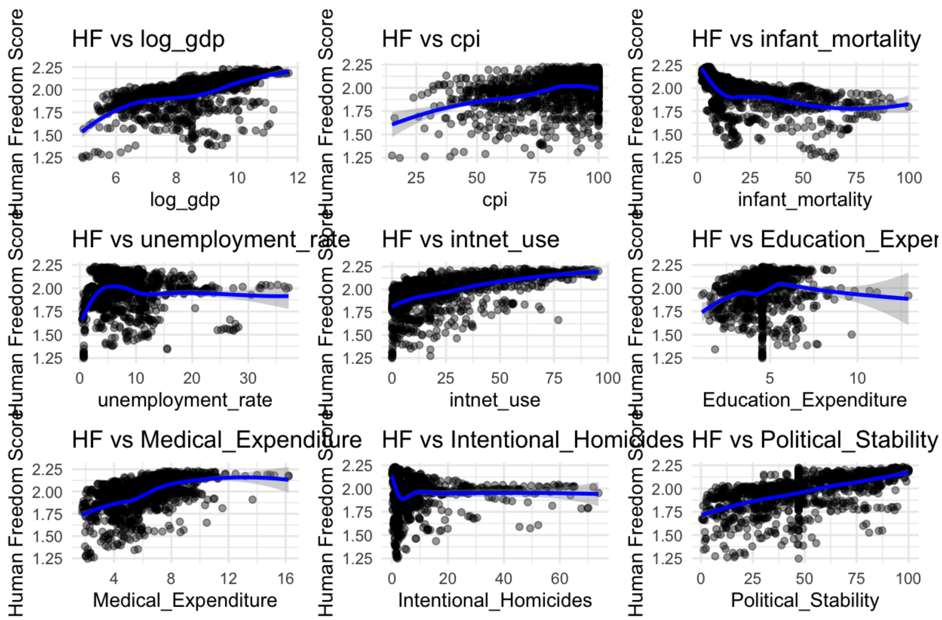

Figure 3: Nonparametric LOESS Regression Plots

In order to flexibly capture the nonlinear effects, a partially linear regression model with natural cubic spline was first employed. Nonlinear features revealed in LOESS smoothers use spline functions with four degrees of freedom, thus giving their relationship with the human freedom index a local curvature. While the CPI, internet usage, and political stability were retained as linear predictors, consistent with the approximately linear trend. In addition to the main effects, three levels of income categorization variables (low, medium, and high) were included to capture the structural differences across countries. To account for potential heterogeneity of effects across income levels, the model was extended to include interaction terms between the income group variables and each of the continuous predictors, which allows the shape and slope of predictor effects to vary flexibly across economic strata. Model comparisons using AIC show that interaction spline model substantially outperforms the non-interaction model (Table 2 in Appendix), suggesting that both nonlinear structure and group-level moderating effects help to account for variations in human freedom index. Table 3 in Appendix However, the spline relies on manually specified degrees of freedom and does not optimize smoothness from the data. To provide a more data-driven treatment of nonlinearity and accommodate group-specific variations in smooth shapes, a weighted generalized additive model was applied.

Unlike the weighted natural cubic spline, the weighted generalized additive model automatically estimates the degree of smoothing for each covariate through generalized cross validation with greater flexibility to pick up local patterns without overfitting. The weighted generalized additive model preserves the three income classes and introduces interaction terms between the income class and each of the continuous predictors, using the "by" parameter in the smoothing term to allow different functional forms for each class. For the linear covariates (CPI, internet usage, and political stability), interaction terms are included in the parameter component in order to maintain interpretability while adjusting for the heterogeneity in the slopes of the groups.

The weighted generalized additive model with interaction lowered the AIC from 2247 to 1939 (Table 1 in Appendix), compared to a basic weighted generalized additive model with only main effects and no interaction terms. The significance of the smoothing terms' analysis shows that the predictors of the smoothing terms are nonlinear, which concurs with the LOESS smoother. The nonlinear income-specific effects of education expenditure were observed with middle-income countries presenting the strongest curvilinear relationship. Finally, The weighted generalized additive model with interaction accounted for 85.7% variance and also 87.2% deviance explained, demonstrating the excellent capture of socioeconomic structural processes.

Table 2: Parametric Coefficients in Weighted Generalized Additive Model with Interactions

Term	Estimate	Std. Error	t value	Pr(>|t|)	Signif.
(Intercept)	-12.52	47.91	-0.261	0.7939	
CPI	0.005396	0.002982	1.810	0.0707	.
Internet Use	-0.02721	0.01219	-2.231	0.0259	*
Political Stability	0.009129	0.002618	3.487	0.0005	***
High Income Group	25.57	48.36	0.529	0.5972	
Middle Income Group	18.38	47.91	0.384	0.7013	
CPI × High Income	-0.02361	0.005359	-4.405	1.17e-05	***
CPI × Middle Income	-0.006249	0.003569	-1.751	0.0803	.
Internet Use × High Income	0.03379	0.01236	2.734	0.0064	**
Internet Use × Middle Income	0.04163	0.01298	3.207	0.0014	**
Pol. Stability × High Income	-0.0007931	0.003058	-0.259	0.7954	
Pol. Stability × Middle Income	0.009332	0.003129	2.982	0.0029	**

Table 3: Smooth Terms in Weighted Generalized Additive Model with Interactions

Term	edf	Ref.df	F	Pr(>|t|)	Signif.
s(Log GDP)	6.98	8.10	4.467	2.09e-05	***
s(Infant Mortality)	0.76	0.76	0.156	0.731	
s(Infant Mortality × Low Income)	5.28	6.00	4.608	0.000168	***
s(Infant Mortality × High Income)	2.95	3.15	1.690	0.0822	.
s(Infant Mortality × Middle Income)	5.88	6.94	4.031	0.000170	***
s(Unemployment Rate)	3.75	4.37	2.505	0.0491	*
s(Unemployment Rate × Low Income)	8.40	8.64	16.482	<2e-16	***
s(Unemployment Rate × High Income)	5.83	6.47	3.826	0.00239	**
s(Unemployment Rate × Middle Income)	3.19	3.73	2.390	0.0524	.
s(Education Expenditure)	2.05	2.62	0.951	0.523	
s(Education Expenditure × Low Income)	5.37	6.22	4.357	0.000173	***
s(Education Expenditure × High Income)	0.86	0.91	0.216	0.657	
s(Education Expenditure × Middle Income)	6.44	7.52	3.203	0.00139	**
s(Medical Expenditure)	1.53	1.80	0.711	0.454	
s(Medical Expenditure × Low Income)	5.21	6.13	3.732	0.00124	**
s(Medical Expenditure × High Income)	2.45	3.07	2.713	0.0456	*
s(Medical Expenditure × Middle Income)	6.27	7.16	4.564	3.57e-05	***
s(Intentional Homicides)	0.76	0.76	0.804	0.433	
s(Intentional Homicides × Low Income)	0.75	0.76	1.070	0.368	
s(Intentional Homicides × High Income)	4.57	5.32	4.187	0.000803	***
s(Intentional Homicides × Middle Income)	7.99	8.56	5.119	1.76e-06	***

To assess the strength and predictive significance of the predictors in a purely nonparametric setting, the random forest regression model was fitted with 500 trees and three variables selected randomly for each split. The model includes all socioeconomic variables and three levels of income group variables. The final model's mean residual square is 0.151, explaining about 90.61% of the variance in the human freedom index, reflecting excellent sample prediction accuracy. Variable importance, as measured by the mean rate of decrease in accuracy and the mean rate of decrease in node impurity, recognizes intentional homicide, political stability, and government expenditure on healthcare as the most influential predictors (Figure 5 in Appendix). Interestingly, the variable importance ranking is partially consistent with the patterns revealed in the weighted generalized additive mode, lending credibility to the nonlinear and income dependent effects revealed previously. The random forest results not only strengthen the validity of the covariates revealed, but also provide a model selection robustness check.

Results and Research Questions

Figure 1 presents the corresponding interaction plots, illustrating that most of the relationships between socioeconomic determinants and human freedom vary across income groups. From Table 1 and Figure 1, main effects of CPI, internet use, unemployment rate, intentional homicides, and political stability are statistically significant at 5% significance level. This implies that, controlling for other variables, these socioeconomic indicators contribute meaningfully to variations in human freedom. Predictors such as CPI, infant mortality, internet use, medical expenditure, and unemployment rate show strong statistical and visual evidence of varying effects on human freedom across income levels, as indicated by statistically significant interaction terms and clear differences in slope among income groups. However, education expenditure, political stability, and log GDP indicated some degree of inconsistency between the statistical output and the interaction plots. Education expenditure showed significant interaction for middle-income countries, and political stability showed significant interaction for high-income countries, but this significance was not clearly reflected in the visualizations. Note that log GDP per capita showed visibly distinct slopes across income groups, particularly a steeper effect in high-income countries, but its interactions with high-income and middle-income groups were not statistically significant. These discrepancies suggest that nonparametric models may be more appropriate for capturing the complex, potentially nonlinear relationships between socioeconomic variables and human freedom across income groups.

The weighted generalized additive model reveals that internet use and political stability have considerable linear impacts. In some interaction terms show statistically significant heterogeneity for specific groups, (most importantly, CPI * high income and political stability * middle income). This model also displays the ubiquitous nonlinearity of the connection between socioeconomic indicators and human freedom index. Log GDP exhibits a very strong nonlinear relationship (edf = 6.98). Income specific smooths have differential importance and complexity of shape for infant mortality, unemployment rate, education expenditure, medical expenditure, and intentional homicides. Particularly, infant mortality shows nonlinear patterns in low income (edf = 5.28) and middle income (edf = 5.88) countries but not in high income settings. Similarity holds for intentional homicides, which is extremely nonlinear in middle income (edf = 7.99) and high income (edf = 4.57) countries. Moreover, the trend of unemployment, education and healthcare expenditure also exhibits income dependent nonlinearities, especially prevalent in middle income countries. These trends validate the structural economic conditions' impact on the human freedom index.

In the linear framework, allowing every slope to vary by income group improved fit dramatically: according to the anova test, adding the 18 interaction terms (2 non-base income levels * 9 predictors) cut the residual sum of squares from 670.9 to 521.2 (p-value < 2.2e-16). When we tested each predictor's pair of interaction coefficients via joint F-tests, seven exhibited highly significant heterogeneity—CPI (p-value < 2.2e-16), infant mortality (p-value < 2.2e-16), unemployment rate (p-value = 2.086e-07), internet use (p-value = 0.003251), medical expenditure (p-value = 7.87e-06), intentional homicides (p-value < 2.2e-16), and political stability (p-value = 0.0003796). In contrast, the interaction terms for log GDP (p-value = 0.305) and education spending (p-value = 0.07835) were not statistically distinguishable from zero.

Moving to nonparametric methods, the spline-based models again also favored interactions: fitting natural splines with four degrees of freedom yielded a reduction in residual sum of squares (ΔRSS=176.58, p-value < 2.2e-16) when comparing models with and without income-group interactions. In the GAM framework, incorporating interaction terms led to a 165.69 unit reduction in deviance (p-value < 2.2e-16). The by-group smooth terms including infant mortality for both middle-income (p-value = 1.704e-04) and low-income countries (p-value = 1.6775e-04), unemployment for low-income (p-value < 2.2e-16) and high-income countries (p-value = 0.0024), intentional homicides for middle-income (p-value = 1.76e-06) and high-income countries (p-value = 8.03e-04), and medical spending for low-income (p-value = 0.0012), middle-income (p-value = 3.57e-05), and high-income countries (p-value = 4.56e-02) were statistically significant. In contrast, the global smooth terms for infant mortality, education expenditure, medical expenditure, and intentional homicides were not significant. Their influence on human freedom cannot be adequately captured by a homogeneous trend, but instead varied meaningfully across income levels. This reinforces the importance of modeling heterogeneous relationships.

Finally, the random forest's variable importance analysis based on permutation confirmed that all variables contributed significantly (all p≈0.01), but with different importance levels. Political stability, intentional homicides, and medical expenditure are the most influential predictors of human freedom. These results underscore the robustness and stability of nonparametric approaches in capturing the nonlinear predictor effects.

Variable	%IncMSE	%IncMSE.pval	IncNodePurity	IncNodePurity.pval
Political_Stability	49.10	0.0099	209.64	0.0099
Intentional_Homicides	48.65	0.0099	144.91	1.0000
Medical_Expenditure	42.89	0.0099	212.10	0.0099
Unemployment_Rate	42.25	0.0099	139.88	1.0000
Infant_Mortality	28.38	0.0099	412.75	0.0099
Education_Expenditure	26.16	0.0099	51.98	1.0000
Log_GDP	25.06	0.0099	216.32	0.0099
Income_Group_3cat	20.37	0.0099	152.06	0.0099
Internet_Use	20.09	0.0099	159.95	1.0000
CPI	18.42	0.0099	38.34	1.0000

Table 4: Permutation-Based Random Forest Variable Importance Metrics

In addition to the above analysis, we also tested for overall differences in HF across income groups without relying on distributional assumptions. The Kruskal–Wallis test of HF by income group yielded a p-value<2.2×10^{-16}, unequivocally rejecting the null of equal distributions. To pinpoint which pairs differ, we applied Dunn's post-hoc test with Bonferroni correction: high-versus-low-income countries produced Z=18.61 (p<4.16e-77), high-versus-middle-income Z = 18.87 (p<3.02e-79), and middle-versus-low-income Z = –6.84 (p<1.18e-11). Thus, every adjacent pair of income strata differs strongly in Human Freedom, corroborating our interaction findings from previous models.

Model Validation & Generalization

To assess the robustness and generalizability of the fitted models, we performed 10-fold cross-validation on four candidate specifications: baseline linear regression, weighted partial linear spline model, weighted generalized additive model , and random forest model. Performance was evaluated using the root mean square error and coefficient of determination calculated using the retained data.

Table 5: 10 Fold Cross Validation across 4 Models

	RMSE	Rsq
Linear	0.719	0.670
Spline	0.713	0.655
GAM	0.641	0.736
RF	0.398	0.900

Of these models, random forest has the best prediction power with smallest root mean square error with a value of 0.398 and largest R-squared with a value of 0.900, which indicates that the random forest can smoothen out nonlinear and interaction effects without specifying a priori structural specifications. While random forest is highly informative about variable importance and has improved fitting capacity, its black-box nature limits interpretability and makes it less appropriate for inference. Weighted generalized additive model is less precise but also possesses strong generalization with a root mean squared error of 0.641 and an R-squared of 0.736. Conversely, the spline model and linear regression model have larger mean square error, i.e., 0.713 and 0.719, respectively, and smaller R squared, i.e., 0.670 and 0.655, respectively, suggesting the potential constraints in capturing the true functional form, especially under nonlinear and heterogeneous situations.

These findings support the applicability of weighted generalized additive model as the preferred inferential model-it provides interpretable smoothing functions and interaction effects while maintaining good predictive power. The RF results further emphasize the predictive importance of features such as intentional homicide, political stability, and healthcare expenditures, aligning with those identified as significant in the weighted generalized additive model and spline model. In summary, the consistency of the effect patterns of flexible parametric and nonparametric learners reinforces the stability of the observed relationships.

Although the weighted generalized additive model and random forest regression differ in terms of interpretability and structure, their comparability suggests that the nonlinear associations found are not model-specific artifacts. Thus, we expect that the main findings-particularly the income-dependent effects of education, health, and demographic factors on freedom can be generalized to similar populations in similar sociopolitical contexts.

Conclusion

This study concentrates on the statistical investigation of socioeconomic determinants of the human freedom index through parametric and nonparametric model frameworks. The initial assumptions diagnostic analysis indicated the violation of linearity and homoscedasticity, which makes the application of nonparametric modeling efficient and reliable. Specifically, the weighted generalized additive model with interaction structures suggesting the substantial nonlinearities and income heterogeneity in patterns of socioeconomic effects. Compared to the weighted linear regression and weighted spline models, the weighted generalized additive model achieved explanatory capability, (with adjusted R squared = 0.857), accuracy, and interaction effects.

Above all, the impacts of critical predictors, such as infant mortality, unemployment rate, government expenditure on education and healthcare, and intentional homicides differ across the income levels. Low income countries exhibited strong nonlinear relationships for infant mortality (edf = 5.28), education expenditure (edf = 5.37), and health expenditure (edf = 5.21), indicating the early investments in public welfare are associated with accelerated improvements in human freedom index at key development thresholds. Middle income nations exhibit widespread nonlinear effects across all predictors. Particularly, intentional homicides (edf = 7.99) and education expenditure (edf = 6.44) show the complex and non-monotonic relationships with human freedom index, suggesting that in such transition countries, marginal changes in violence or investment do not have constant impacts. This pattern reflects the tension between institutional capacity and economic development, where the human freedom index is shaped by thresholds, saturation points, and conflicting influences. High income countries have more stable and attenuated relationships, with moderate nonlinearity in features, such as intentional homicides (edf = 4.57), unemployment (edf = 5.83), and medical expenditure (edf = 2.54). These trends suggest that even in sophisticated countries, freedom index remains vulnerable to institutional disruptions and healthcare investments, though crude development indicators such as infant mortality and education expenditure appear less influential, consistent with the hypothesis that investments in these areas may have saturated. These findings underscore the necessity of modeling income-stratified nonlinear effects, as policy levers do not operate uniformly across the developmental spectrum.

Despite the advantages of modeling strategy, some limitations should be taken into consideration. The observational data restrict the causal interpretation, and complex model forms might face the risk of overfitting. Future studies may apply some causal inference approaches, such as propensity score model, to better manage the treatment effects and recognize the confounders.

Appendix

Summary Statistics for All Numeric Variables

Variable	NA Count	Mean	Std	Min	25%	Median	75%	Max	Skew	Kurt
Log GDP	303	8.570	1.505	4.879	7.474	8.507	9.887	11.699	-0.124	-0.825
CPI	247	80.329	17.237	14.991	69.872	84.336	94.217	100.000	-0.937	0.311
Infant Mort	115	22.372	20.871	2.100	5.700	15.300	31.600	99.600	1.279	0.825
Unemploy	116	7.961	5.716	0.398	4.120	6.596	10.469	37.161	1.782	4.361
Internet	185	26.326	25.829	0.000	4.600	17.750	41.700	95.600	0.947	-0.244
Edu Exp	953	4.588	1.323	1.100	3.919	4.546	5.193	12.902	0.670	2.915
Med Exp	347	6.085	2.369	1.756	4.212	5.926	7.863	16.201	0.458	0.246
Homicides	1424	8.277	12.522	0.000	1.447	2.896	9.210	73.791	2.707	7.632
Pol Stab	258	49.987	26.696	0.474	28.643	47.090	71.498	100.000	0.155	-0.972
HF	75	7.249	1.269	3.480	6.420	7.370	8.290	9.320	-0.574	-0.188

Table 1: Summary Statistics of all variables

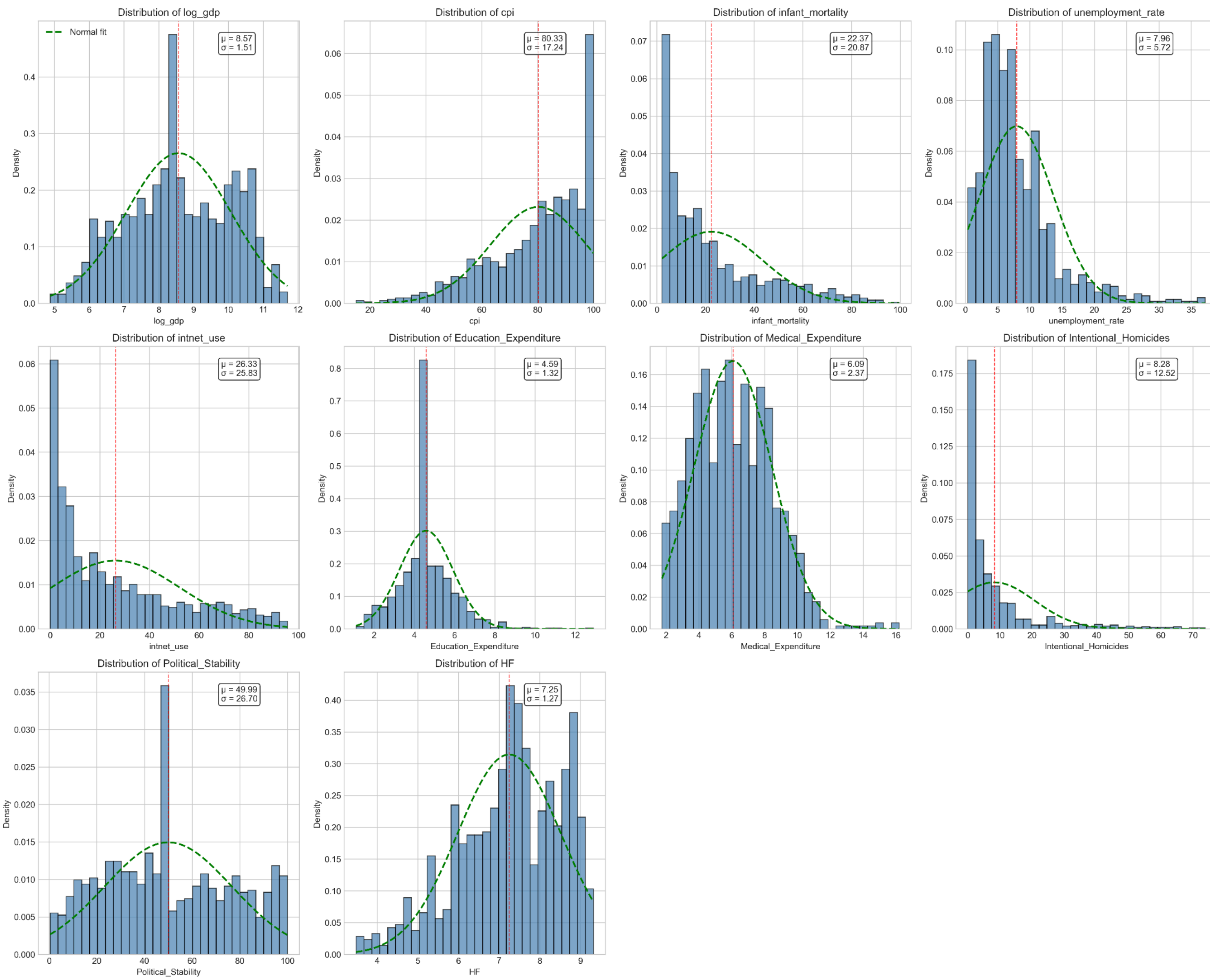

Figure 1: Histogram of Variables

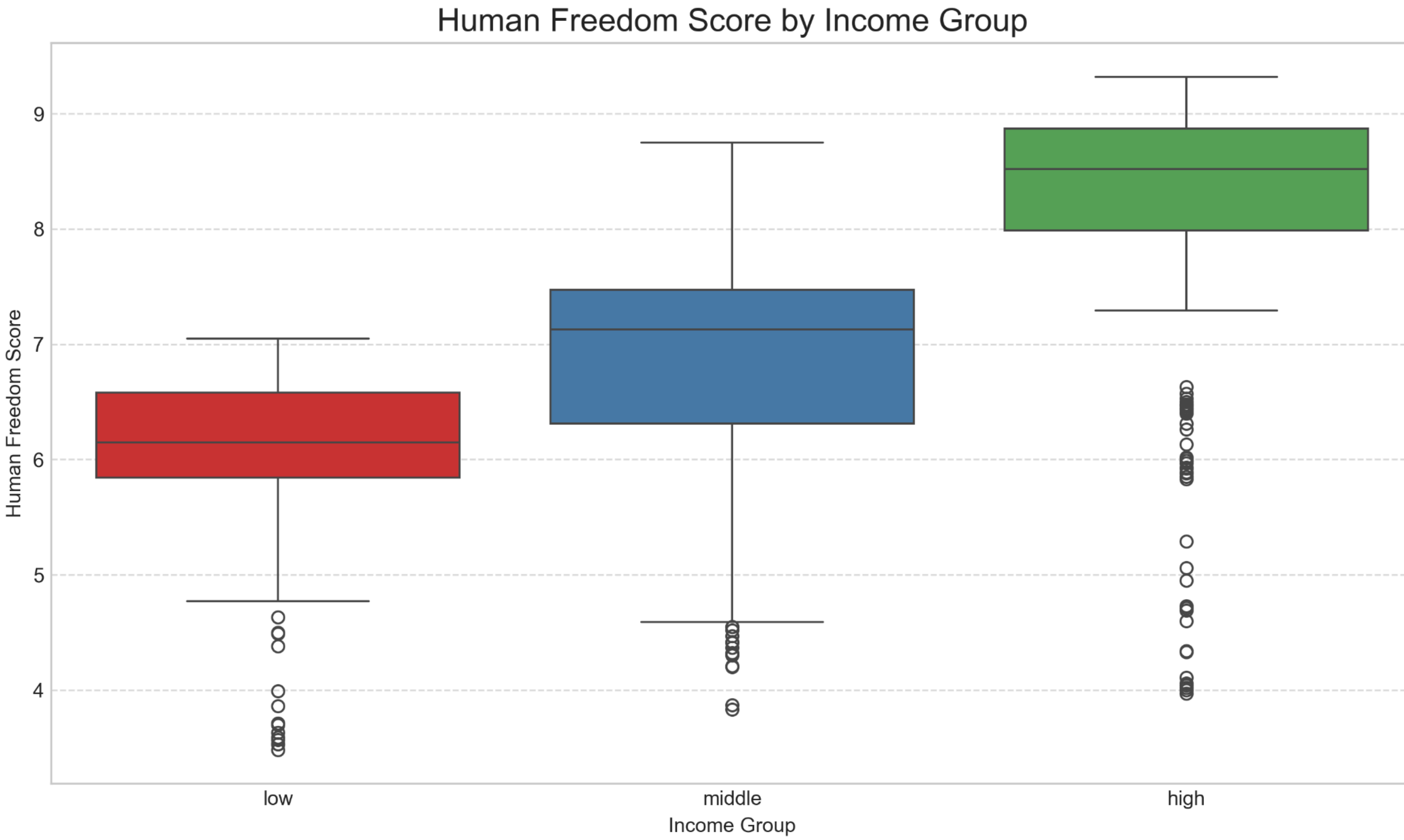

Figure 2: Boxplot of Human Freedom Score

Figure 3: Bivariate Plot of Human Freedom Score and the most Relevant Predictors

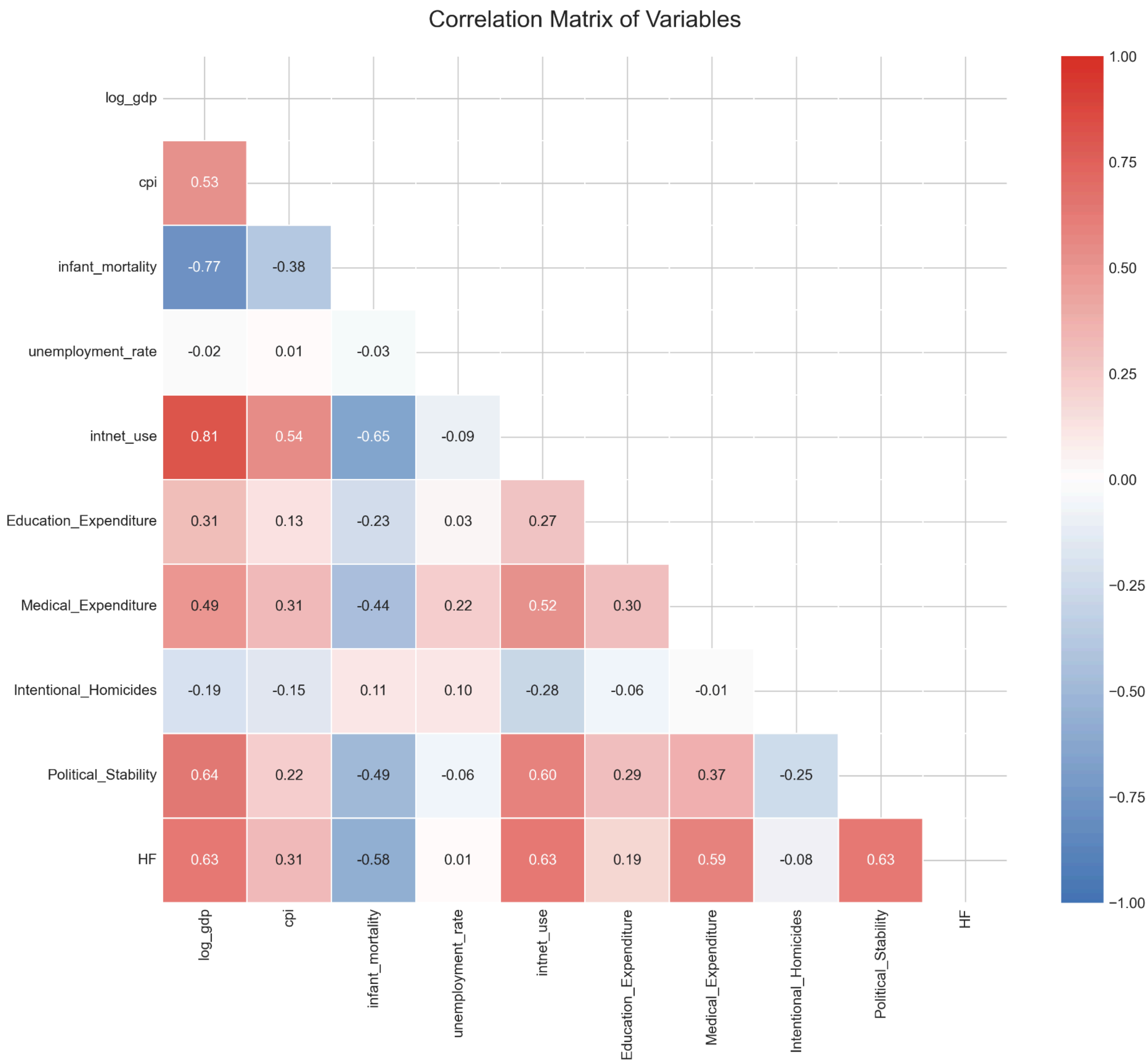

Figure 4: Correlation Matrix of All Variables

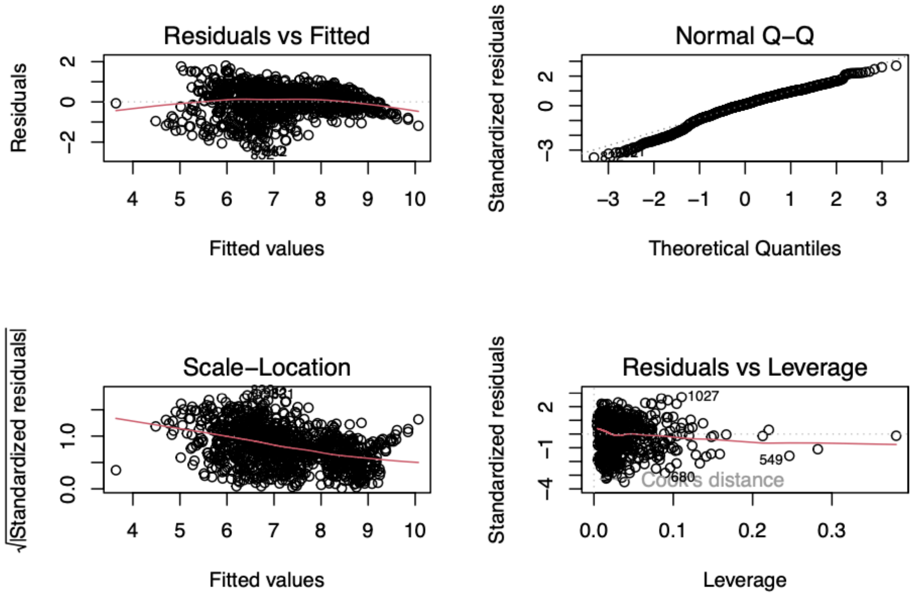

Figure 5: Residual Diagnostics for the Linear Regression Model with Income Group Interactions

	df	AIC
model_spline	31.00	2477.68
model_spline_interact	76.00	2100.04
model_gam	52.77	2321.45
model_gam_interact	100.27	1872.38

Table 2: AIC Comparison Between Spline and Generalized Additive Model (with/without Interactions)

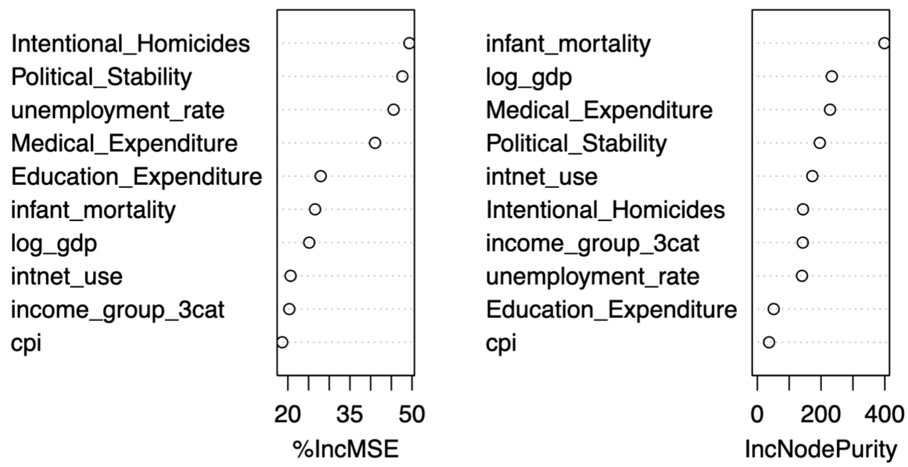

Figure 6: Variance Important Plot in Random Forest Regression Model

Table 3: Results for Spline Model with Interactions

Term	Estimate	Std. Error	t-value	p-value	Signif.
(Intercept)	206.511	73.965	2.792	0.005	**
ns(log_gdp, df = 4)1	-0.526	0.262	-2.007	0.045	*
ns(log_gdp, df = 4)2	-0.018	0.281	-0.064	0.949	
ns(log_gdp, df = 4)3	0.382	0.611	0.625	0.532	
ns(log_gdp, df = 4)4	-0.382	0.316	-1.210	0.226	
cpi	0.011	0.003	3.530	0.000	***
ns(infant_mortality, df = 4)1	- 199.762	73.853	-2.705	0.007	**
ns(infant_mortality, df = 4)2	-41.570	14.392	-2.888	0.004	**
ns(infant_mortality, df = 4)3	- 562.727	207.497	-2.712	0.007	**
ns(infant_mortality, df = 4)4	-1.505	0.418	-3.598	0.000	***
ns(unemployment_rate, df = 4)1	0.340	0.267	1.270	0.204	
ns(unemployment_rate, df = 4)2	0.314	0.804	0.391	0.696	
ns(unemployment_rate, df = 4)3	9.633	2.217	4.346	0.000	***
ns(unemployment_rate, df = 4)4	13.878	4.750	2.922	0.004	**
intnet_use	-0.056	0.012	-4.740	0.000	***
ns(Education_Expenditure, df = 4)1	-0.337	0.182	-1.854	0.064	.
ns(Education_Expenditure, df = 4)2	0.013	0.399	0.032	0.974	
ns(Education_Expenditure, df = 4)3	-1.397	0.709	-1.969	0.049	*
ns(Education_Expenditure, df = 4)4	-0.815	1.075	-0.758	0.448	
ns(Medical_Expenditure, df = 4)1	0.161	0.160	1.007	0.314	
ns(Medical_Expenditure, df = 4)2	2.182	0.465	4.695	0.000	***
ns(Medical_Expenditure, df = 4)3	-0.463	0.716	-0.647	0.518	
ns(Medical_Expenditure, df = 4)4	-4.752	1.603	-2.965	0.003	**
ns(Intentional_Homicides, df = 4)1	0.151	0.274	0.550	0.583	
ns(Intentional_Homicides, df = 4)2	-7.017	1.851	-3.791	0.000	***
ns(Intentional_Homicides, df = 4)3	60.976	22.226	2.744	0.006	**
ns(Intentional_Homicides, df = 4)4	126.922	44.057	2.881	0.004	**
Political_Stability	0.007	0.003	2.527	0.012	*
income_group_3cathigh	- 199.874	73.934	-2.703	0.007	**
income_group_3catmiddle	- 203.162	73.924	-2.748	0.006	**
ns(infant_mortality, df = 4)1:income_group_3cathigh	196.845	73.855	2.665	0.008	**
ns(infant_mortality, df = 4)2:income_group_3cathigh	23.902	16.537	1.445	0.149	
ns(infant_mortality, df = 4)3:income_group_3cathigh	1270.679	320.606	3.963	0.000	***
ns(infant_mortality, df = 4)4:income_group_3cathigh	1295.659	447.299	2.897	0.004	**
ns(infant_mortality, df = 4)1:income_group_3catmiddle	200.643	73.851	2.717	0.007	**
ns(infant_mortality, df = 4)2:income_group_3catmiddle	42.550	14.395	2.956	0.003	**
ns(infant_mortality, df = 4)3:income_group_3catmiddle	564.055	207.492	2.718	0.007	**

Table 3 (continued): Results for Spline Model with Interactions

Term	Estimate	Std. Error	t-value	p-value	Signif.
ns(unemployment_rate, df = 4)1:income_group_3cathigh	-0.117	0.538	-0.218	0.827	
ns(unemployment_rate, df = 4)2:income_group_3cathigh	0.054	1.500	0.036	0.971	
ns(unemployment_rate, df = 4)3:income_group_3cathigh	-10.398	4.006	-2.596	0.010	**
ns(unemployment_rate, df = 4)4:income_group_3cathigh	-17.222	8.634	-1.995	0.046	*
ns(unemployment_rate, df = 4)1:income_group_3catmiddle	-0.378	0.516	-0.733	0.464	
ns(unemployment_rate, df = 4)2:income_group_3catmiddle	-0.695	1.451	-0.479	0.632	
ns(unemployment_rate, df = 4)3:income_group_3catmiddle	-11.520	3.941	-2.923	0.004	**
ns(unemployment_rate, df = 4)4:income_group_3catmiddle	-17.298	8.493	-2.037	0.042	*
ns(Education_Expenditure, df = 4)1:income_group_3cathigh	0.591	0.744	0.794	0.427	
ns(Education_Expenditure, df = 4)2:income_group_3cathigh	-0.950	0.887	-1.071	0.285	
ns(Education_Expenditure, df = 4)3:income_group_3cathigh	1.293	1.987	0.650	0.516	
ns(Education_Expenditure, df = 4)4:income_group_3cathigh	1.609	2.478	0.649	0.516	
ns(Education_Expenditure, df = 4)1:income_group_3catmiddle	0.562	0.381	1.477	0.140	
ns(Education_Expenditure, df = 4)2:income_group_3catmiddle	-1.761	0.749	-2.350	0.019	*
ns(Education_Expenditure, df = 4)3:income_group_3catmiddle	1.661	1.370	1.213	0.226	
ns(Education_Expenditure, df = 4)4:income_group_3catmiddle	0.530	1.960	0.271	0.787	
ns(Medical_Expenditure, df = 4)1:income_group_3cathigh	1.326	0.380	3.486	0.001	***
ns(Medical_Expenditure, df = 4)2:income_group_3cathigh	-0.435	0.862	-0.505	0.614	
ns(Medical_Expenditure, df = 4)3:income_group_3cathigh	3.761	1.365	2.755	0.006	**
ns(Medical_Expenditure, df = 4)4:income_group_3cathigh	5.802	2.832	2.049	0.041	*
ns(Medical_Expenditure, df = 4)1:income_group_3catmiddle	-0.320	0.329	-0.973	0.331	
ns(Medical_Expenditure, df = 4)2:income_group_3catmiddle	-1.132	0.867	-1.306	0.192	
ns(Medical_Expenditure, df = 4)3:income_group_3catmiddle	-0.671	1.382	-0.485	0.628	
ns(Medical_Expenditure, df = 4)4:income_group_3catmiddle	3.517	2.974	1.183	0.237	
ns(Intentional_Homicides, df = 4)1:income_group_3cathigh	0.260	0.552	0.472	0.637	
ns(Intentional_Homicides, df = 4)2:income_group_3cathigh	12.553	3.462	3.626	0.000	***

Table 3 (continued): Results for Spline Model with Interactions

Term	Estimate	Std. Error	t-value	p-value	Signif.
ns(Intentional_Homicides, df = 4)3:income_group_3cathigh	-57.757	22.262	-2.594	0.010	**
ns(Intentional_Homicides, df = 4)4:income_group_3cathigh	-126.230	44.167	-2.858	0.004	**
ns(Intentional_Homicides, df = 4)1:income_group_3catmiddle	1.531	0.477	3.212	0.001	**
ns(Intentional_Homicides, df = 4)2:income_group_3catmiddle	8.227	1.871	4.398	0.000	***
ns(Intentional_Homicides, df = 4)3:income_group_3catmiddle	-58.252	22.245	-2.619	0.009	**
ns(Intentional_Homicides, df = 4)4:income_group_3catmiddle	-125.812	44.059	-2.856	0.004	**
cpi:income_group_3cathigh	-0.029	0.006	-4.880	0.000	***
cpi:income_group_3catmiddle	-0.009	0.004	-2.438	0.015	*
intnet_use:income_group_3cathigh	0.062	0.012	5.149	0.000	***
intnet_use:income_group_3catmiddle	0.071	0.013	5.585	0.000	***
Political_Stability:income_group_3cathigh	0.002	0.003	0.598	0.550	
Political_Stability:income_group_3catmiddle	0.012	0.003	3.547	0.000	***

Reference

Datasets:

1. *Cato Institute*. “Human Freedom Index 2023.” *Cato Institute*, 2023, https://www.cato.org/human-freedom-index/2023.

2. *World Bank*. “Unemployment Rate (% of Total Labor Force) (Modeled ILO Estimate).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/WB_SSGD_UNEMPLOYMENT_RATE.

3. *World Bank*. “GDP per Capita (Current US$).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/WB_WDI_NY_GDP_PCAP_PP_CD.

4. *FAO Statistics Division*. “Consumer Price Index (CPI).” *World Bank Data360*, https://data360.worldbank.org/en/dataset/FAO_CP.

5. *World Bank*. “Individuals Using the Internet (% of Population).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/WB_SSGD_PCT_HHS_INTERNET.

6. *International Monetary Fund (IMF)*. “General Government Education Expenditure (% of GDP).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/IMF_COFOG_GEE_GF09?view=trend.

7. *World Bank*. “Current Health Expenditure (% of GDP).” *World Bank Open Data*, https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?most_recent_value_desc=false.

8. *World Bank*. “Intentional Homicides (Per 100,000 People).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/WB_WDI_VC_IHR_PSRC_P5.

9. *World Bank*. “Political Stability and Absence of Violence/Terrorism (Percentile Rank).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/WB_WDI_PV_PER_RNK.

10. *World Bank*. “Infant Mortality Rate (Per 1,000 Live Births).” *World Bank Data360*, https://data360.worldbank.org/en/indicator/WB_WDI_SP_DYN_IMRT_IN.

Papers:

1. Abdelrahim, Yousif. “The Influence of Personal and Economic Oppression on a Country’s Corruption Levels Worldwide.” *50 Cell Press*, 2024, https://www.cell.com/action/showPdf?pii=S2405-8440%2824%2908722-X.

2. Disoska, Makrevska, and Shapkova Kocevska. “Human Freedom and Economic Prosperity: Evidence from Eastern Europe.” *Visio Journal*, vol. 4, 2019, pp. 37–48. https://visio-institut.org/wp-content/uploads/2020/01/Elena-Makrevska-Disoska-and-Katerina-Shapkova-Kocevska-Visio-Journal-4-37-48.pdf

3. Dolan, Ed. “Economic Freedom and Personal Freedom: What Can We Learn from the Cato and Fraser Indexes?” *Niskanen Center*, 2023,

https://www.niskanencenter.org/economic-freedom-and-personal-freedom-what-can-we-learn-from-the-cato-and-fraser-indexes/.

4. Kocevska, Katerina, and Disoska, Elena. "Human Freedom and Economic Development: A Granger Causality Analysis of Panel Data." *ResearchGate*, 2021, https://www.researchgate.net/publication/357328505.