A study on the determinants of growth and political stability in the neoclassical model framework

January 2024

Roger Casas

1. Introduction

1.1 Theoretical Framework: Convergence in Neoclassical Growth Models

In alignment with contributions like Robert J. Barro's *Economic Growth in a cross section of countries* (Barro, 1991) the main scope of this project is to study whether or not convergence, as predicted in neoclassical growth models, is a reasonable result. Therefore, this empirical study aims to further contribute to the validation (or not) of the results and assumptions of neoclassical growth theory. Moreover, further analysis of the determinants of long-run growth and the factors that condition political stability is also performed.

Specifically, convergence may be tested in the context of Solow's (Solow, 1956) model, referred to in the appendix for further understanding of its mechanisms and implications. Yet, it is worth stating one of its main conclusions, convergence. In the Solow model (also called this way), growth in output per capita can be expressed by the following equations:

$$\frac{\dot{y}_t}{y_t} = a + \alpha \frac{\dot{\tilde{k}}_t}{\tilde{k}_t}$$

$$\frac{\dot{\tilde{k}}_t}{\tilde{k}_t} = A_t^{\frac{1-\alpha}{\alpha}} \left(\frac{s}{y_t^{\frac{1-\alpha}{\alpha}}} - \frac{s}{{y_t}^{*\frac{1-\alpha}{\alpha}}} \right)$$

This implies that, for countries with similar parameters, the more distant the output level of a country is from its steady state (or Balanced Growth Path), represented by ${y_t}^*$ in the equation, the larger its growth rate of output per capita ($\frac{\dot{y}_t}{y_t}$). Notice that other values are assumed non-negative. As a consequence, when considering similar steady-state levels of output per capita for countries, or, in practice, when controlling for variables other than y_t, the growth rate of output per capita should be inversely related to the level itself. Again, this premise is tested in the obtained cross-section of countries, which is explained in the following section.

1.2 Data Discussion: Dataset and Included Variables

Cross-sectional data containing observations from 93 countries was obtained by integrating diverse data sources. Mainly, the obtained variables come from files provided by The World Bank, although others were also integrated from Our World in Data, Penn World Tables or The Heritage Foundation. For specification of the source of each variable please direct to Table 1 in the appendix.

Variables considered for causal relation assessment and the study of convergence respond, also, to the arguments provided by scientific literature (Barro, 1991). To measure the output per capita level, real GDP per capita in terms of 2017 US dollars in 1990 was used. The growth rate of output per capita was computed using the same variable for the years 1990 and 2019. Notice that considering output per capita instead of the level itself is a way, per se of controlling for population growth. In

the same sense, the fact that GDP90 and GDP19 are real variables accounts for different price developments within countries. Notice also that the time span considered for computing the growth rate is non-trivial, as the Solow model mainly explains long-run growth. Thus, considering a rather large time period for growth rate computation also helps to neutralize short-run variations, the study of which is outside the scope of this paper.

Furthermore, ES90 acts as a proxy accounting for human capital stock. In various endogenous economic growth models, human capital is regarded as one of the main determinants of the quality of the research sector, which in turn conditions technological progress. Therefore, the addition of an educational enrollment score responds to the need to control for human capital stock and technological progress (variable A_t and its growth rate a in the previous equation). Similarly, PPI90DEV relates to price distortions in the market for capital stock and both WGI and GovernmentConsumption do inform on distinct areas of public sector performance. It is argued, moreover, that government consumption negatively affects growth through the distorting effect on savings of expenditure programs (Barro, 1990) (Barro, 1989). Finally, it might be intuitively clear that the economic regime of a country can be related to its growth, the significance and direction of such relation are also tested.

2. Descriptive Analysis

A prior graphical inspection of the testable relations was conducted through scatter plots (for numerical regressors) and bar plots (for EconomicRegime). Please find such plots in Figures 1-9 in the Appendix.

In principle, the relation between GDP90 and GR9019 would seem weak and polluted by outliers (Figure 1). To omit outliers, filtering according to box-plot criterion was performed for both, the dependent and the independent variable. According to such a rule, a value of the distribution of a variable is labeled as an outlier if it is either above the third quartile plus 1.5 times the interquartile range or below the first quartile minus 1.5 times the interquartile range. Outliers were cleared, firstly, for GR9019, and for GDP90 as a second step. This led to data depicting a more compressed relation between both variables, as reported in Figure 4, although still apparently weak. Nonetheless, looking at the trend line, the hypothesis of a negative relation between output level and growth rate seems sensible.

Afterward, plots for other continuous regressors confirmed the hypotheses formulated a priori. The relation between governance performance (including political stability) and the educational score seems positive, while government consumption and market distortions apparently affect growth negatively. Additionally, any strong conclusion apart from average comparison between groups can be made by inspecting Figure 9.

Lastly, a correlation matrix for the continuous regressors was created and inspected to check potential multicollinearity problems (Table 2). Logically, the real GDP per capita level and the governance index seem significantly correlated. As correlation is not perfect, such regressors can be included in the regression model, yet, their condition should be considered when assessing significance and estimating standard errors.

3. General Multivariate Regression Model

Given the fact that the objective of the project is finding and revisiting causal relations (which is different in terms of approach to predicting or forecasting), a general linear model was estimated. Such specification was regarded as the base one, which may be subject to further refinement in the sections below. Therefore, composite hypothesis testing and results analysis are done in subsequent sections.

3.1 Results

As shown in Figure 10, the empirical findings presented in this study lend support to the outcomes delineated by Barro (1991) and, more broadly, align with the principles of conditional convergence. Notice that coefficients on GovernmentConsumption, WGI, ES90, and GDP90 are found significant at a 0.01 or lower level. Estimates for which the confidence interval intersects with the 0 vertical line are, logically, not regarded as significant at a 5% confidence level.

As the results confirm, when considering control indicators, such as human capital approximated through educational enrollment rates, or the governance index, a discernible pattern of convergence emerges. The sign of the estimated coefficient in this context adheres to expectations: ceteris paribus, lower real GDP per capita levels imply higher average growth rates.

In principle, the significance of human capital proxies should coincide with observations of conditional convergence. As this is the case in the estimated model, the reported findings do not contradict the conclusions drawn in Barro (1991). Yet, it could still be argued that educational enrollment rates may be imperfect proxies for capturing information on human capital. A parallel line of reasoning may account for the non-significance of the coefficient on PPI90DEV, as deviation from the sample mean of the PPP index for capital stock might not be the most apt variable for accounting for market distortions.

Furthermore, the sign of the coefficients associated with the governance index and the share of government consumption in GDP remains consistent with existing literature (Barro, 1989) (Barro, 1990).

The non-significance of the binary variables representing the categories of the economic regime could potentially be attributed to the imperfect nature of the categorization derived from the economic freedom index. Once again, this acts as an imperfect proxy that may not fully encapsulate the nuances of the economic regime. Notice also that, in the general model and all other specifications below, the dummy variable related to the first category was omitted to avoid perfect multicollinearity.

3.2 Heteroscedasticity and Checks on Normality Assumptions:

Before delving into the final regression model, it is essential to address the potential presence of heteroscedasticity across countries. Heteroscedasticity, as suggested by Barro (1991), can have notable implications in the context of cross-country data. Therefore, a comprehensive examination

is deemed necessary before proceeding with the final regression estimates, as it lays the foundation for the subsequent utilization of White's (1980) heteroscedasticity-consistent covariance matrix, if deemed essential.

Initially, a graphical inspection was conducted to assess variance equality by examining residuals concerning every numeric predictor and fitted values (figures 11-16). Upon scrutinizing the graphs, no discernible pattern was observed for the residuals along different x-axes. To further investigate heteroscedasticity, the Goldfeld-Quandt test (Goldfeld, Quandt, 1965) was employed, akin to the conventional variance test. This test involves eliminating a fraction of central observations of a preidentified explanatory variable, with the fraction normally set to the closest integer to 20% of the sample size. The results indicate insufficient evidence to reject the null hypothesis regarding homoscedasticity. Thus, considering both graphical and test-related outcomes, it was deemed feasible to proceed with the regression model without employing heteroscedastic errors. However, a second verification is conducted in the regression diagnostic section following the selection of the final model.

Subsequently, attention was directed to assessing the normality of residuals through graphical methods (figures 17 and 18), the Shapiro-Wilk test and a test for the true mean of the residuals. The Q-Q plot certainly validates the residuals' normality assumption, with an approximate adjustment to the bisector line. The histogram also exhibits a general resemblance to that of a normal sample. Further statistical analysis, including the mean test, provides insufficient evidence to reject the null hypothesis that the true sample mean of the residuals differs from zero. Additionally, the null hypothesis implying the normality of the disturbance term is not rejected in the Shapiro-Wilk test.

4. Building a More Succinct Model

4.1 Step-down Greedy Search Method

A step-down greedy search algorithm was employed to systematically eliminate highly non-significant variables. The resultant model, hereafter referred to as the "base" compressed model, served as the foundation from which various tests and specifications were derived.

Notice that the greedy search algorithm can be applied using alternative criteria, such as the AIC and BIC, or methods (step-up or combined step-up and step-down). The selection of the step-down and p-value approach stems from the objective of studying causal inference.

The algorithm used expectably discarded the binary variables relative to the economic regime and PPI90DEV.

4.2 Hypothesis Testing in the Step-down Model Context

The results for the step-down base model can be found in Figure 19. As explained in the general model section, the observed directions of the relationships are logical. It is worth stating that all coefficients exhibit high levels of statistical significance (0.01 or lower), and that the F-test for the joint significance of the covariates was decisively rejected.

Nonetheless, double-check of the non-significance of the EconomicRegme and PPI90DEV variables was done through F-tests for nested models, summarized in Table 3, which depicts an anova (instruction in R) output summarizing all information. Any of the three nested models which were tested were found significantly better in terms of performance than the others. Therefore, PPI90DEV was considered in the final specification for control purposes, as the relation between market distortions and output growth has already been proven by the literature (Barro, 1991). Contrarily, as the inclusion of the economic regime variables was a mere hypothesis, such covariates have been dismissed. Notice also that in any of the specifications shown in figures from 19 to 21 neither the coefficients on EconomicRegime and PPI90DEV are found relevant, while the others are significant. Estimates on model 3 are basically the ones of the general regression.

Further checks on polynomial trends were computed for GDP90 and WGI. It is worth stating that such hypotheses are theory-based, as returns to scale in terms of higher growth to improved governance may be diminishing, or the negative relations between the level of output and the growth rate may be positive when a certain level is reached. Nonetheless, Figures 22-24 show that any quadratic term, as well as PPI90DEV was found significant for different specifications, while estimates on GovernmentConsumption, WGI, GDP90 and ES90 were found relevant.

The finally selected specification would be, thus, model 5, which keeps the quadratic term on WGI and PPI90DEV for control reasons, and also because consideration of such variables is theory-based. Heteroscedasticity and normality checks were conducted again for the final equation, leading to essentially equal results than in section 3.2 (see figures 25 and 26 for histogram and Q-Q plot of the residuals).

5. Diving into Political Stability

5.1 ANOVA for Political Stability

According to previous investigations (Miljkovic, Rimal, 2008), an ANOVA model was formulated, including the Continent, Independence, and PoliticalRegime variables as categorical factors, and PoliticalStability as the dependent variable. The continent in which a given country is located may affect its political stability as, when being a neighbor of potentially dangerous states, the credibility of its institutions can be jeopardized. Additionally, the period in which a country gained independence can affect political stability as, hypothetically, rather ancient countries may have had more time to develop social, political, and economic institutions. The contrary could happen, for instance, to post-colonialist countries. As Table 4 shows, the main effects on Continent and

PoliticalRegime seem significant, as well as the three-way interaction effect. Contrarily, pair interaction effects and the main effect related to Independence are not found significant.

5.2 Logistic Regression

For further analysis of the relations between the included variables and the political stability index, a logistic regression model was developed. As a discrete variable, a binary variable (dTopStability), which equals 1 if a given observation's political stability index lies above the sample 70% quantile of such variable and 0 otherwise, was considered. In general, results only found significant thre coefficient on the dummy variable representing the third category related to PoliticalRegime. Such conclusions are equal for the two specifications considered (one including categories related to independence and the other not doing so), as shown in table 5.

6. Conclusions and Limitations of the Project

Further analysis should meticulously assess the magnitude of the coefficients, not only the direction of the relations, which has not been deeply commented on due to multiple bias signals. However, the findings made are not trivial at all, as hypotheses on many covariates have been confirmed.

To improve the quality of the coefficients, proxy enhancement would be an important step. Literacy rates instead of education scores or other measures approaching the effect of technological stock should be added. Bootstrap aggregation techniques or instrumental variable regression could also be helpful in this sense.

Regarding the logistic regression, another discrete variable accounting for political stability could be used, and the inclusion of additional regressors could be considered. It is also noteworthy that there can exist reciprocal causality between regressands like GR9019 and political stability or governance indices, leading to biases in both multivariate and logistic frameworks.

Finally, the analysis of this paper is static, i.e., cross-sectional. The time dimension, also found in many of the variables included (although not everyone) could be leveraged to see how factors are interrelated over time. Research in this line has been conducted by Barro and Sala-i-Martín (1992).

Appendix

Table 1: Variable definition and source specification
GDP90: Real GDP per capita in terms of current US$. Source: The World Bank. **GDP19:** Real GDP per capita in terms of current US$. Source: The World Bank. **GR9019:** Growth rate of real GDP per capita from 1990 to 2019 in terms of current US$. Generated from GDP90 and GDP19. **PE90:** Primary education enrollment rate in 1990. Source: The World Bank. **PE19:** Primary education enrollment rate in 2019. Source: The World Bank. **TE90:** Tertiary education enrollment rate in 1990. Source: The World Bank. **TE19:** Tertiary education enrollment rate in 2019. Source: The World Bank. **ES90:** Educational score computed as a simple average of PE90 and TE90. **ES19:** Educational score computed as a simple average of PE19 and TE19. **ControlofCorruption:** Control of corruption index provided by The World Bank. **RuleofLaw:** Rule of Law index provided by The World Bank. **RegulatoryQuality:** Regulatory Quality index provided by The World Bank. **PoliticalStability:** Political Stability index provided by The World Bank. **GovernmentEffectiveness:** Government effectiveness index provided by The World Bank. **WGI:** World Governance Index computed as an average of the five previous variables. **GovernemntConsumption:** Government consumption as a share of GDP. Source: The World Bank. **Independence:** Period in which a given country gained independence. Categorized: - B.C. Era to 10th Century = 1 - 1000 to 15th Century = 2 - 1500 to 1700 = 3 - 1800 = 4 - 1901-1949 = 5 - 1951-1960 = 6

- 1961-1968 = 7
- 1970-1979 = 8
- 1980-Present = 9

Source: author creation.

Continent: Geographical zone to which a given country belongs (rigorously may not exactly refer to conventional continent division). Categorized:

- North America = 1
- Asia = 2
- Africa = 3
- Europe = 4
- South America = 5
- Oceania = 6

Source: author creation.

PoliticalRegime: Categorical variable accounting for the electoral regime of a given country. Coded:

- Closed autocracy = 0
- Electoral autocracy = 1
- Electoral democracy = 2
- Liberal democracy = 3

Source: Our World in Data.

PPI90DEV: Absolute deviation from the sample average of Purchasing Power Parity index for capital goods in 1990. Source: Penn World Tables.

EconomicRegime: Categorical variable based on the Index of Economic Freedom by The Heritage Foundation. Coded:

- Observation's index below the 20% quantile = 1
- Observation's index between the 20% and 40% quantile = 2
- Observation's index between the 40% and 60% quantile = 3
- Observation's index between the 60% and 80% quantile = 4
- Observation's index above the 80% quantile = 5

Figure 1: GR9019 vs. GDP90

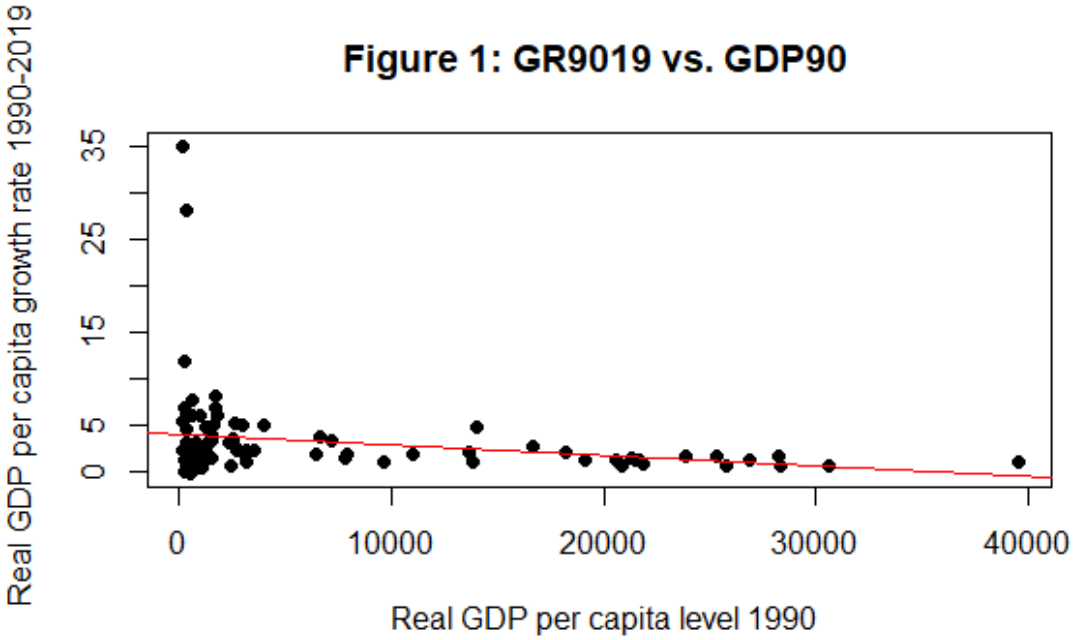

Figure 2: GR9019

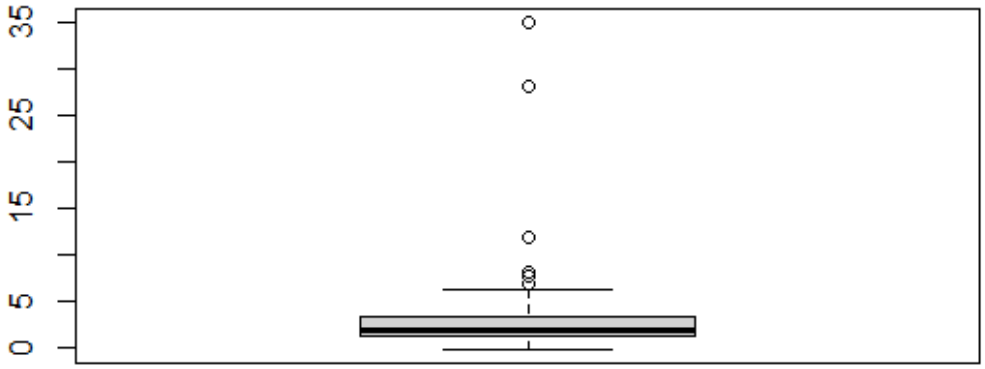

Figure 3: GDP90

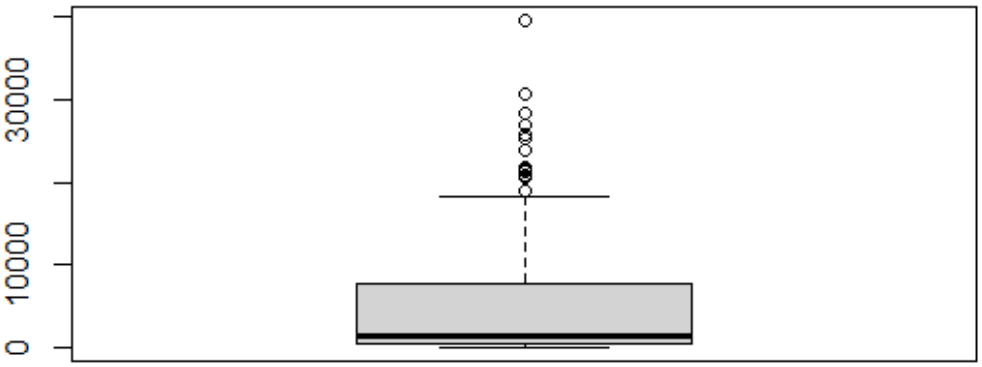

Figure 4: GR9019 vs. GDP90

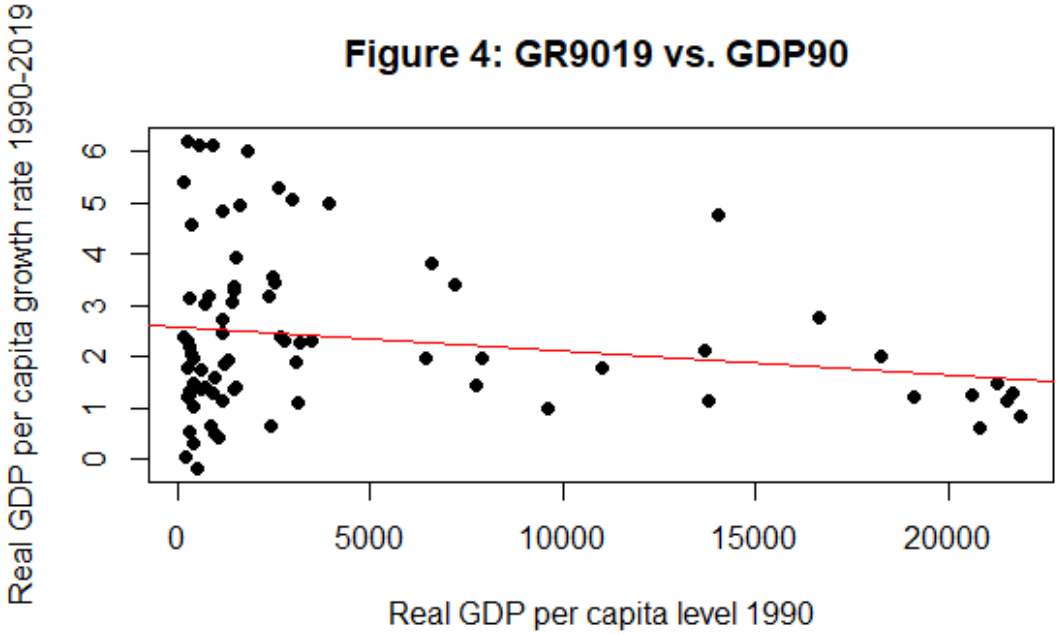

Figure 5: GR9019 vs. ES90

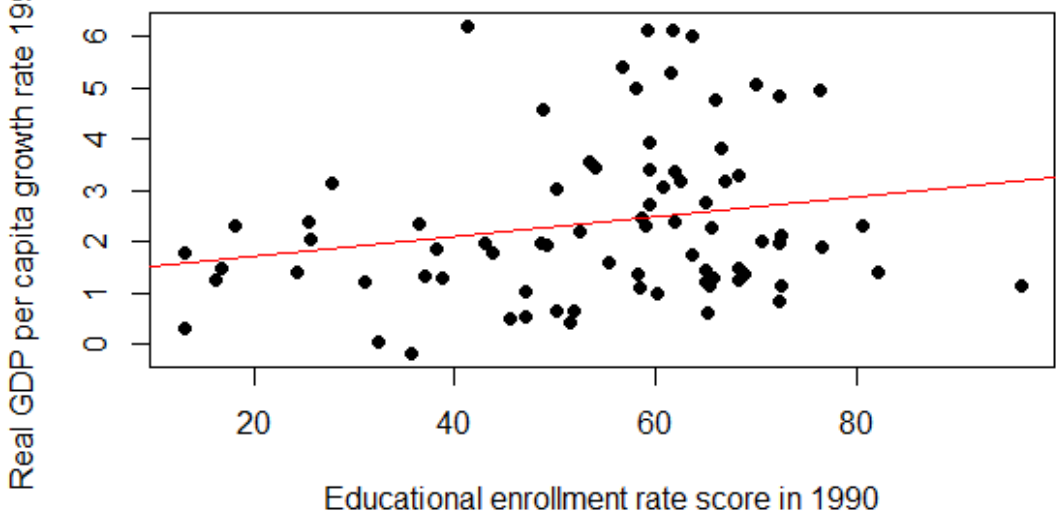

Figure 6: GR9019 vs. WGI

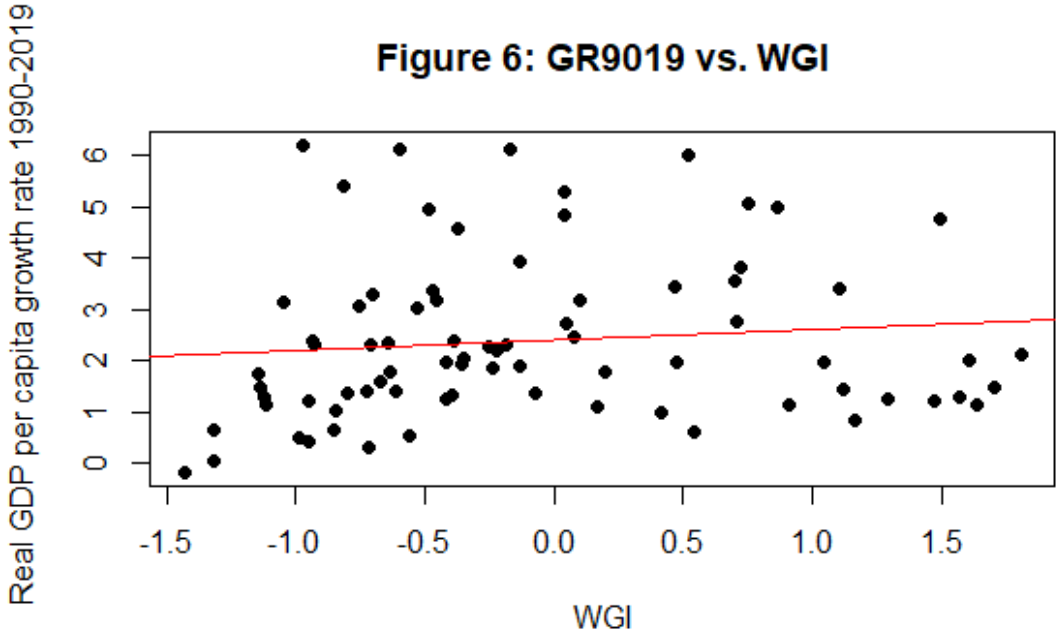

Figure 7: GR9019 vs. GovernmentConsumption

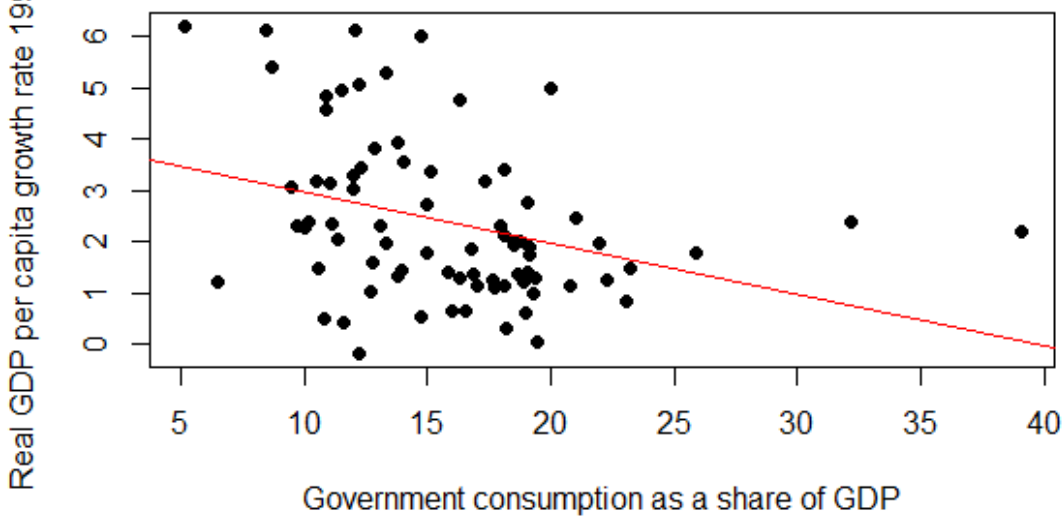

Figure 8: GR9019 vs. PPI90DEV

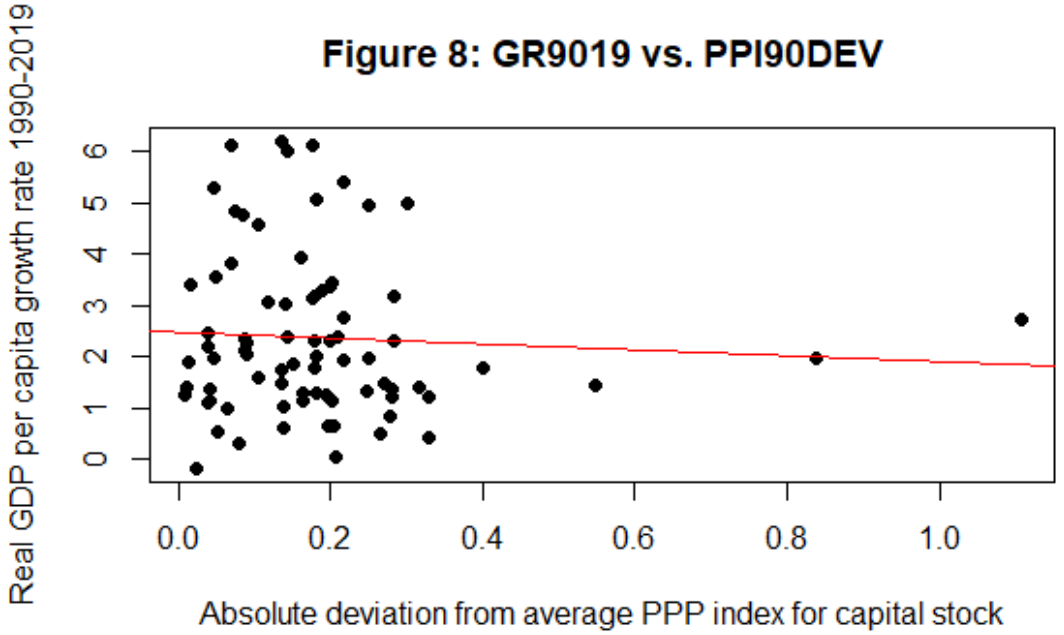

Figure 9: Average GR9019 per economic regime category.

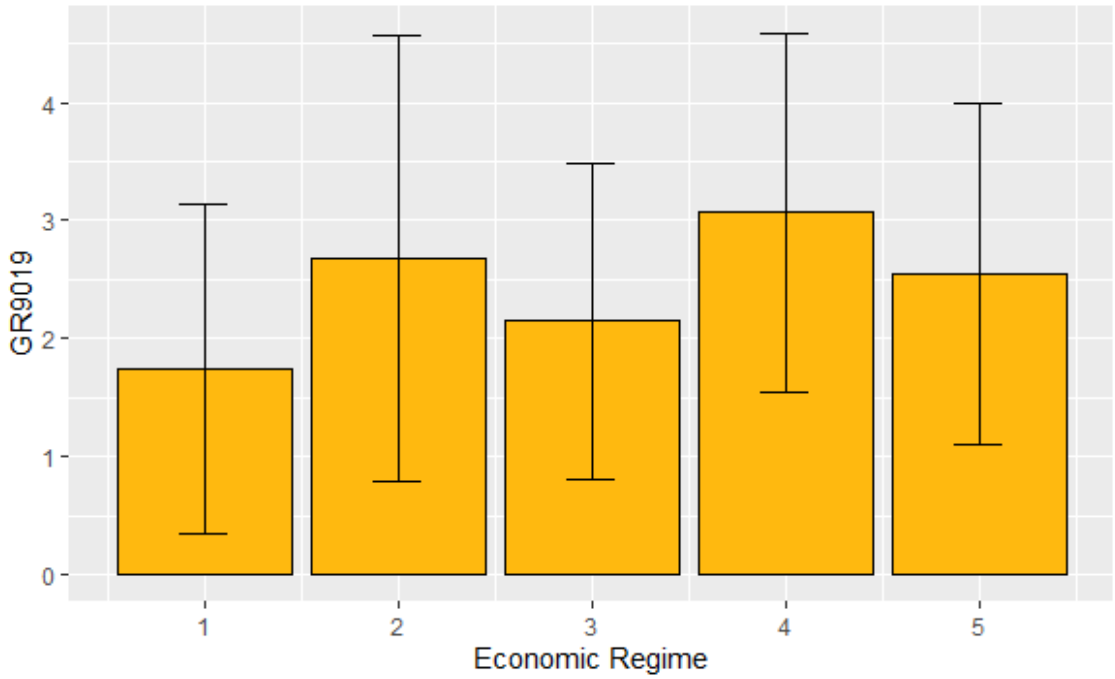

Table 2: Correlation matrix of all regressors (with p-values for the correlation test)

	GDP90	ES90	WGI	GovernmentConsumption	PPI90DEV
GDP90	1	0.48	0.82	0.42	-0.01
ES90	0.48	1	0.53	0.25	0.03
WGI	0.82	0.53	1	0.38	-0.01
GovernmentConsumption	0.42	0.25	0.38	1	-0.07
PPI90DEV	-0.01	0.03	-0.01	-0.07	1
N = 78					
P-values					
	GDP90	**ES90**	**WGI**	**GovernmentConsumption**	**PPI90DEV**
GDP90		0	0	0	0.94
ES90	0		0	0.028	0.8
WGI	0	0		0	0.91
GovernmentConsumption	0	0.028	0		0.5483
PPI90DEV	0.94	0.8	0.91	0.5483	

Figure 10: General regression estimates

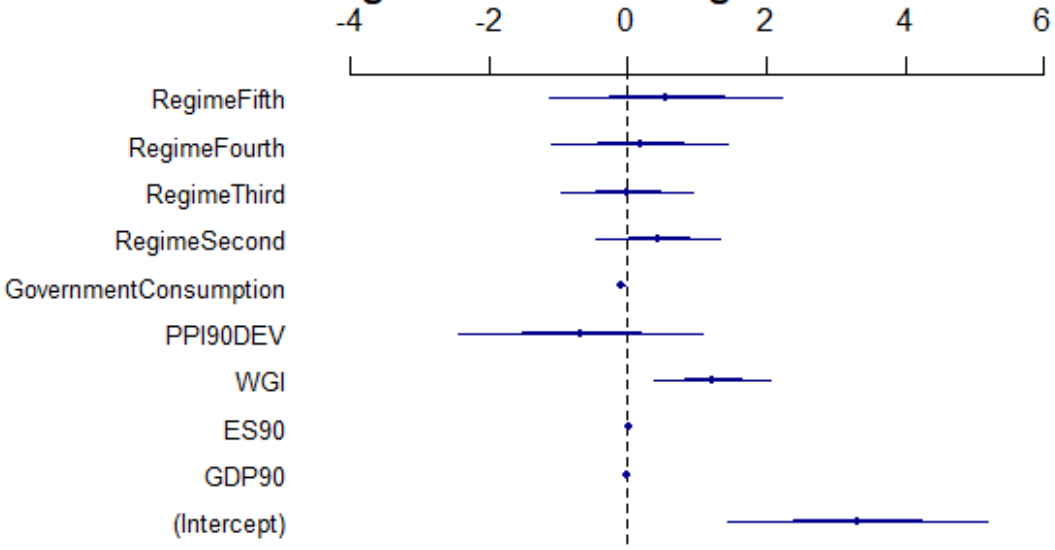

Figure 11: Residuals vs. ES90

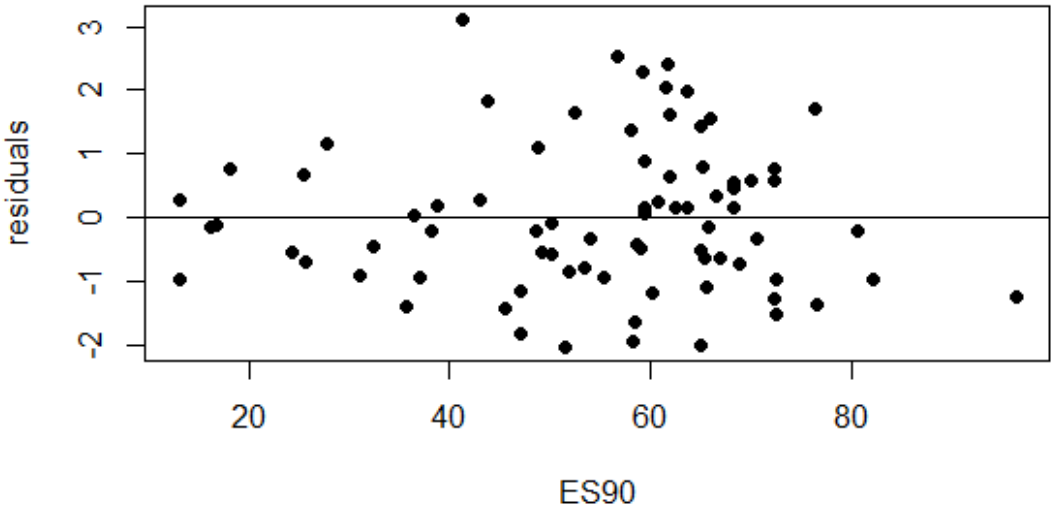

Figure 12: Residuals vs. WGI

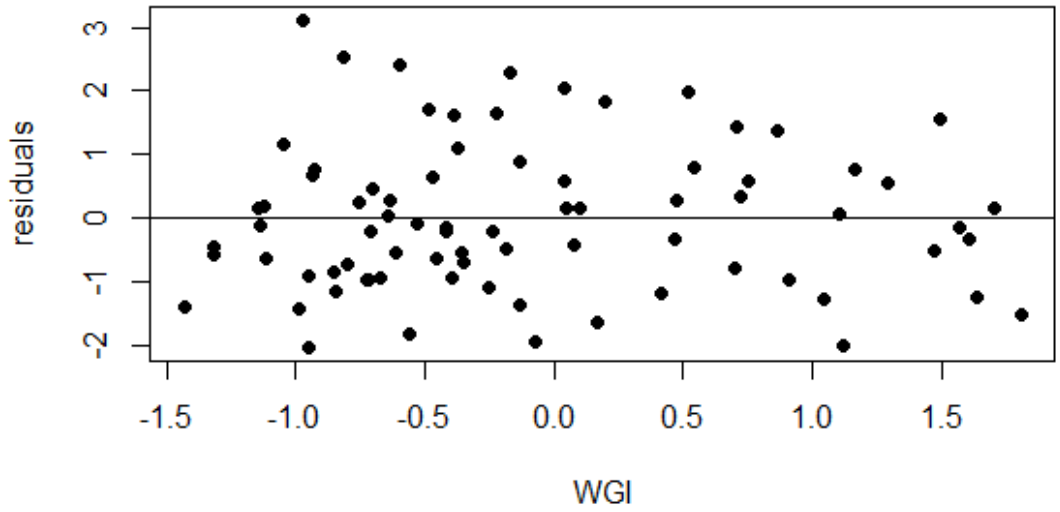

Figure 13: Residuals vs. GovernmentConsumption

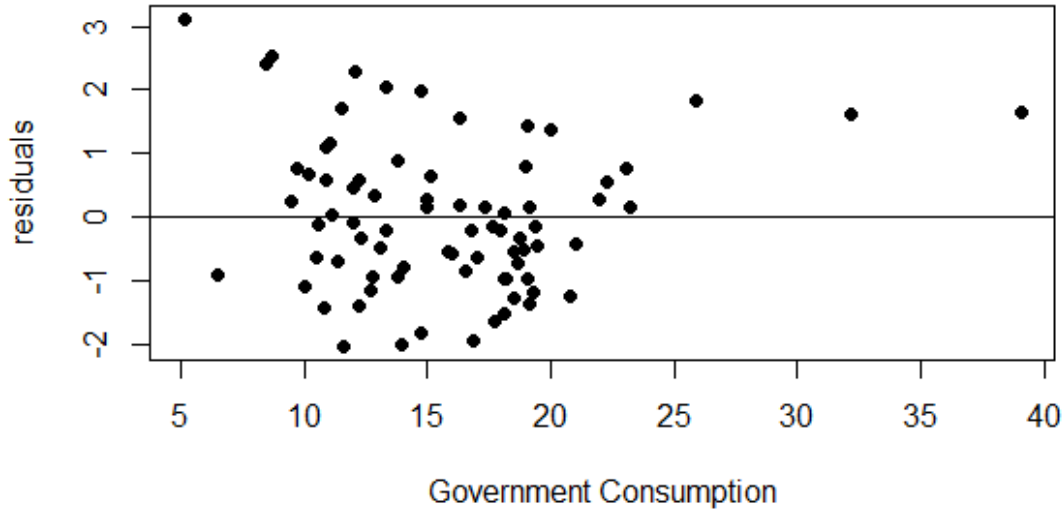

Figure 14: Residuals vs. GDP90

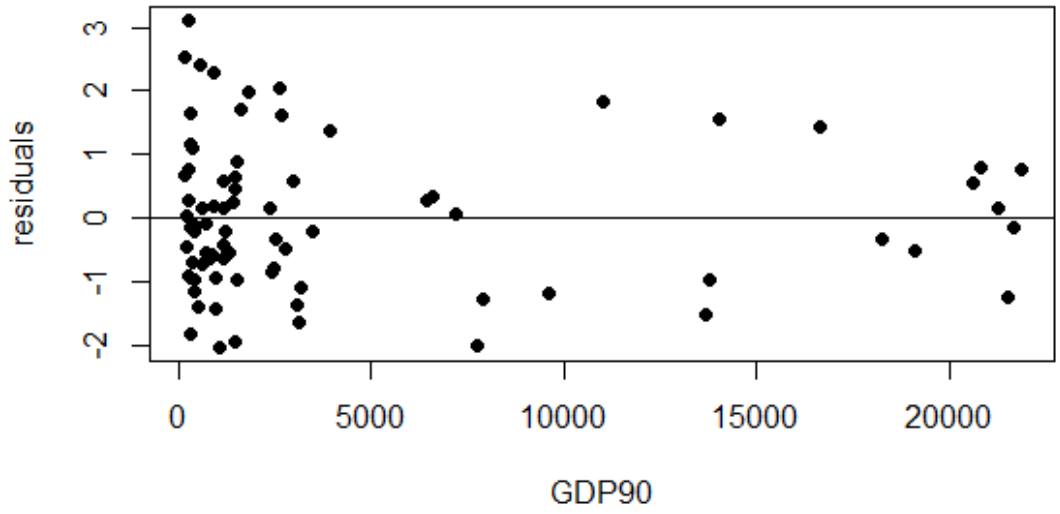

Figure 15: Residuals vs. PPI90DEV

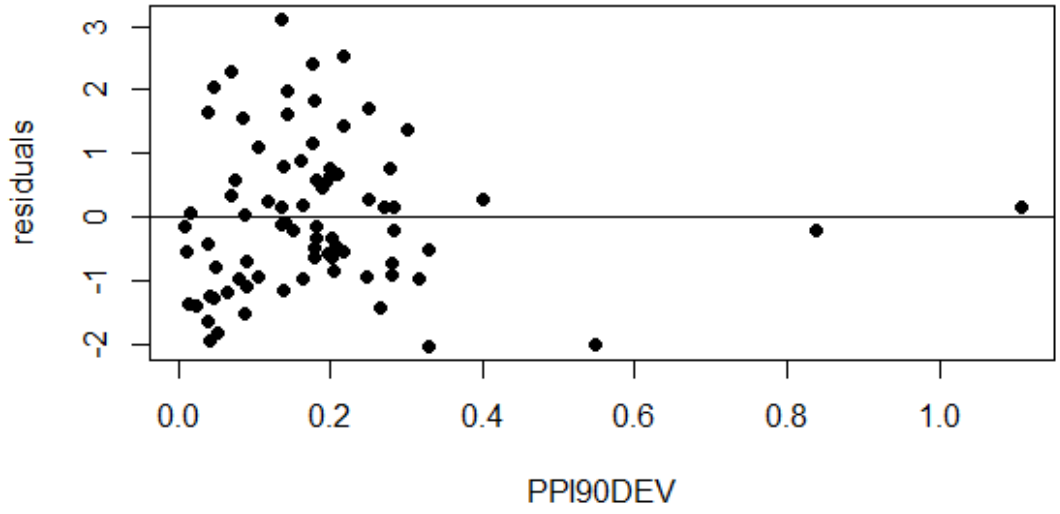

Figure 16: Residuals vs. fitted values

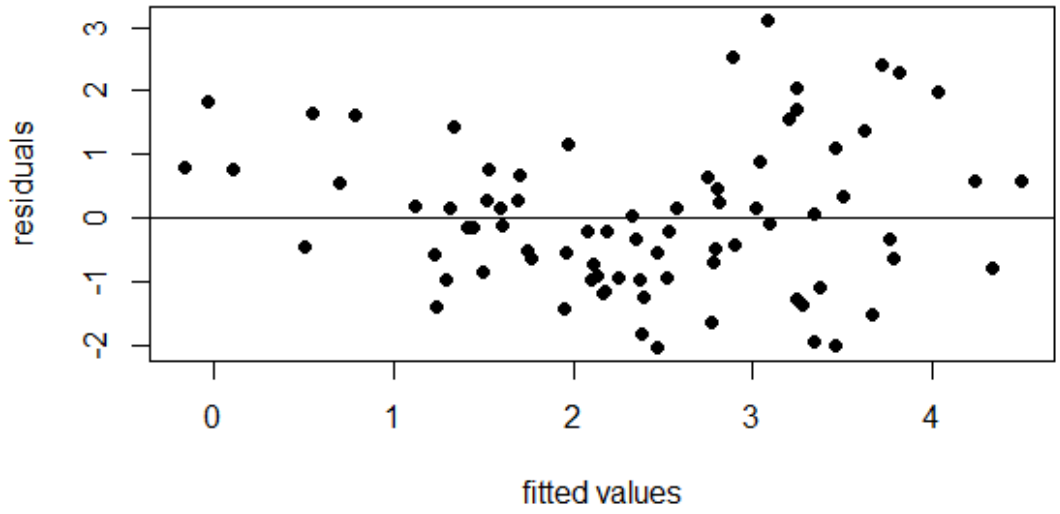

Figure 17: Normal Q-Q plot for model residuals

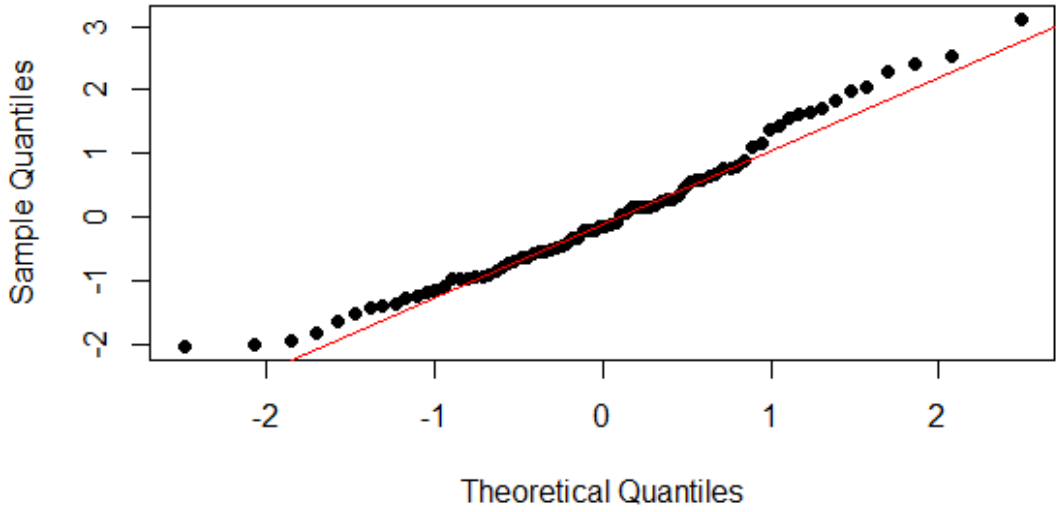

Figure 18: Histogram of general model residuals

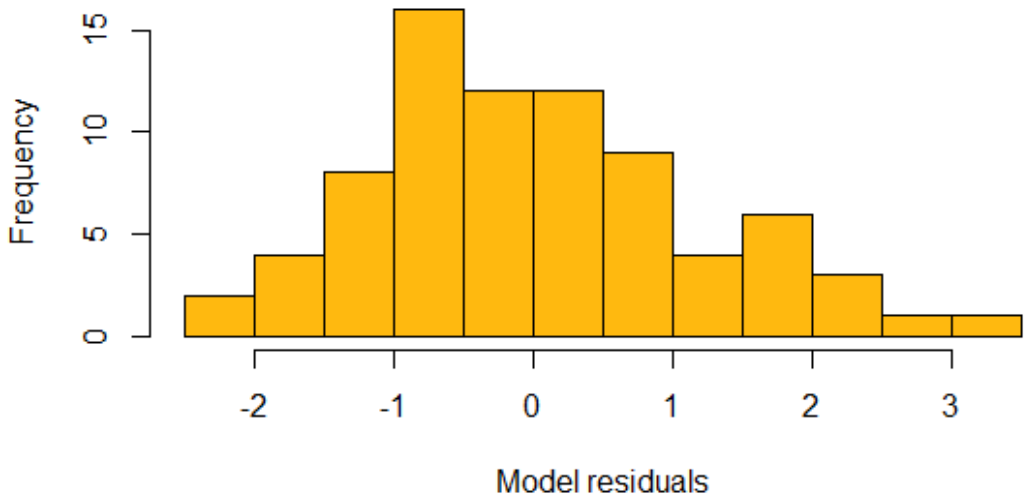

Figure 19: Step-down base regression estimates

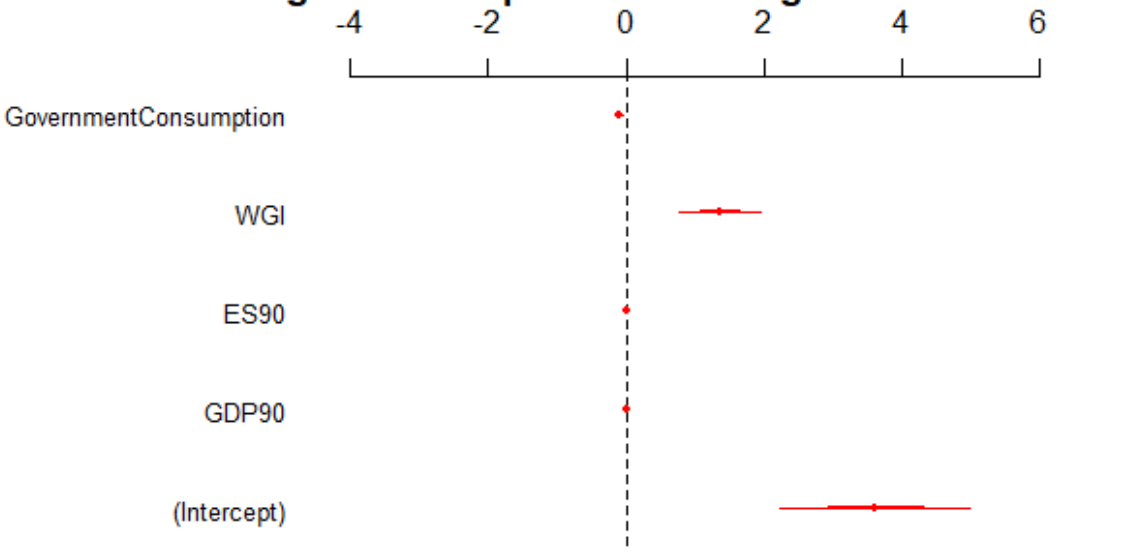

Table 3: F tests for the joint significance of EconomicRegime and for significance of PPI90DEV

Base step-down model: $GR9019_i = GDP90_i + GovernmentConsumption_i + ES90_i + WGI_i$						
Model 3: $GR9019_i = GDP90_i + GovernmentConsumption_i + ES90_i + WGI_i + RegimeSecond_i + Regimethird_i + Regimefourth_i + Regimefifth_i$						
Model 2: $GR9019_i = GDP90_i + GovernmentConsumption_i + ES90_i + WGI_i + PPI90DEV_i$						
	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
Base	73	109.74				
Model 2	72	108.23	1	1.5058	0.9735	0.3273
Model 3	68	105.19	4	3.05	0.4929	0.7409

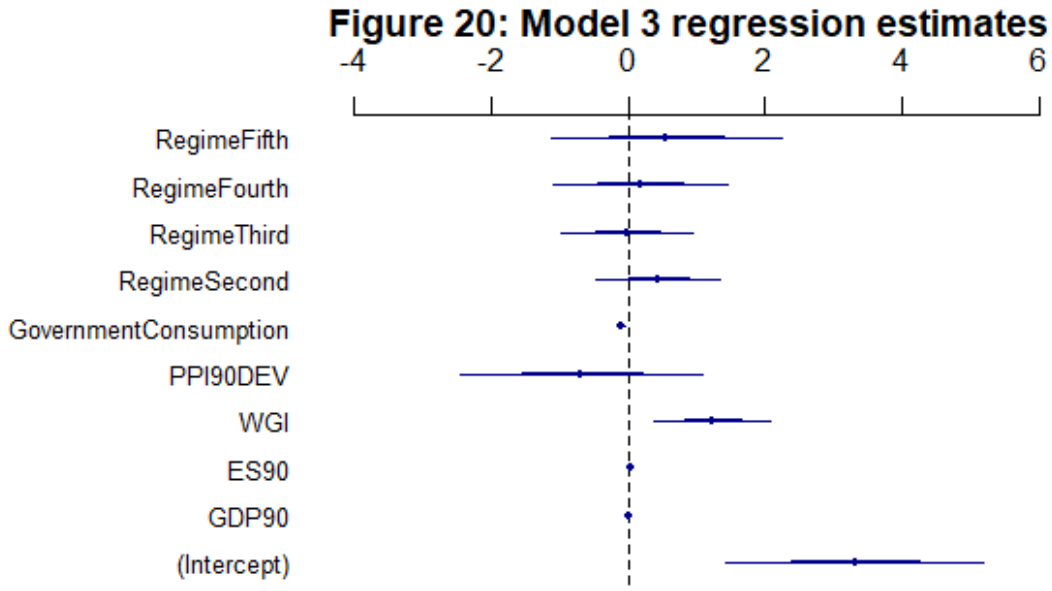

Figure 21: Model 2 regression estimates

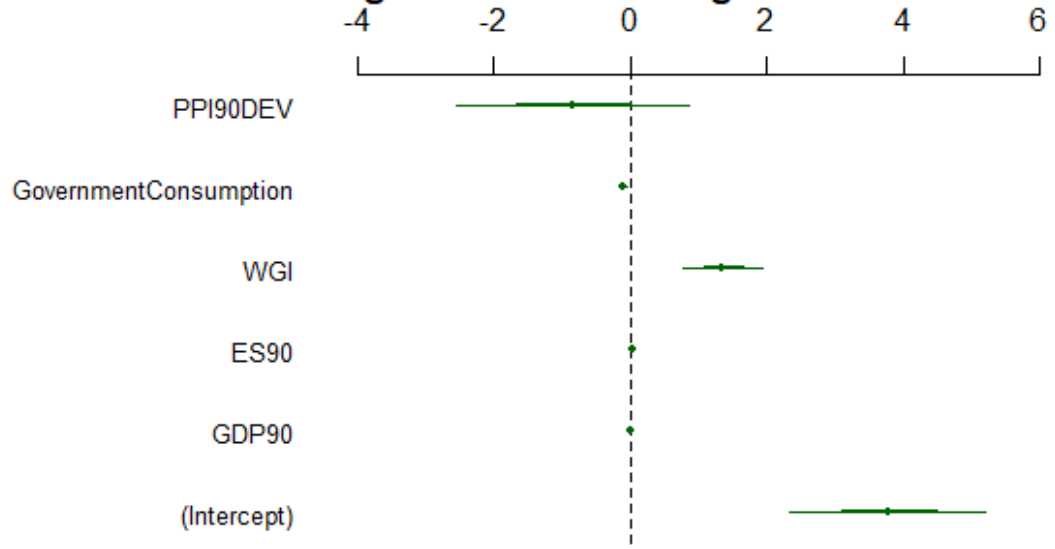

Figure 22: Model 4 (with a quadratic term on GDP)

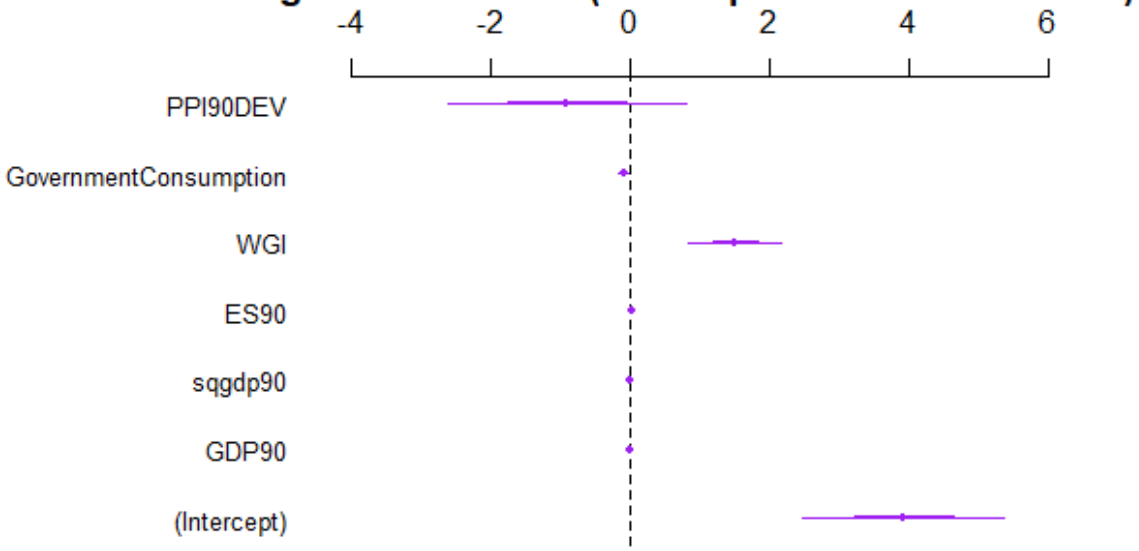

Figure 23: Model 5 (with a quadratic term on WGI)

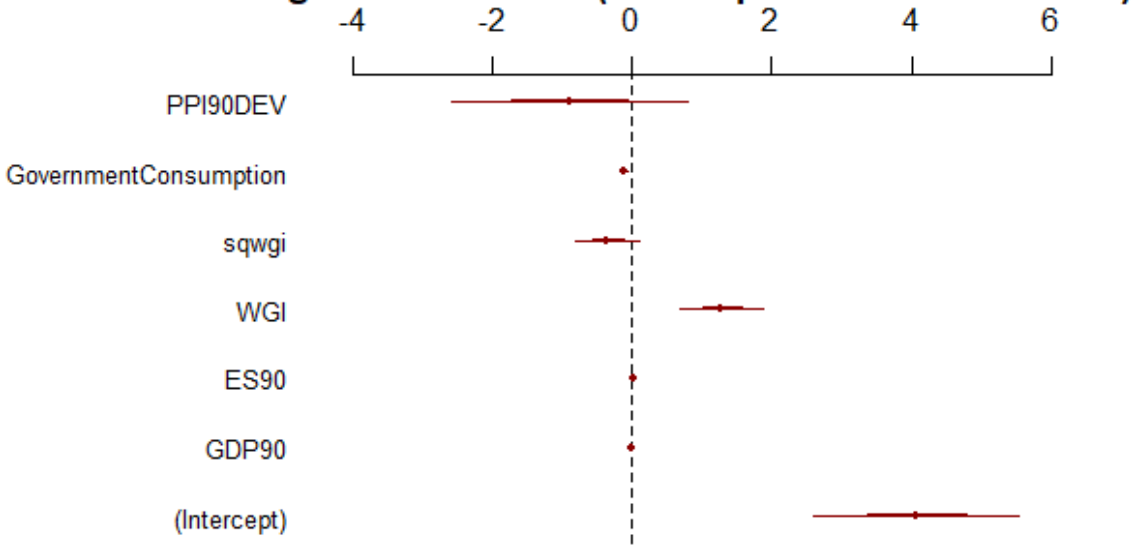

Figure 24: Model 6 (with both quadratic terms)

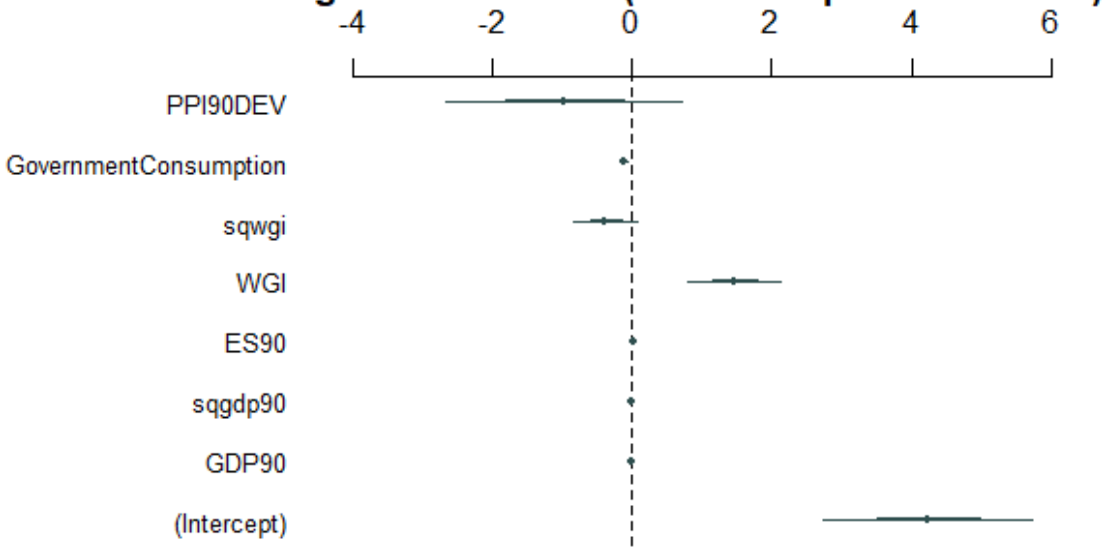

Figure 25: Normal Q-Q plot for final model residuals

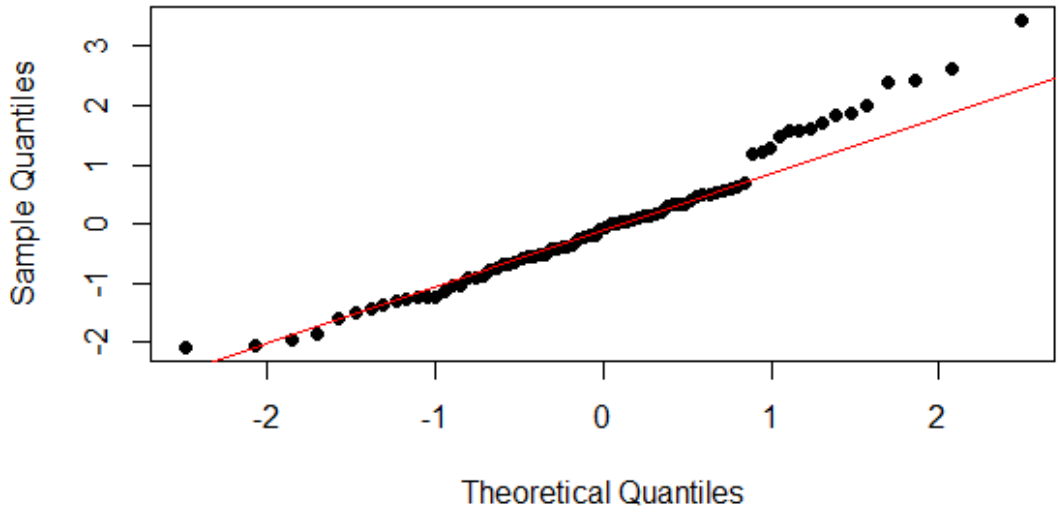

Figure 26: Final model residuals

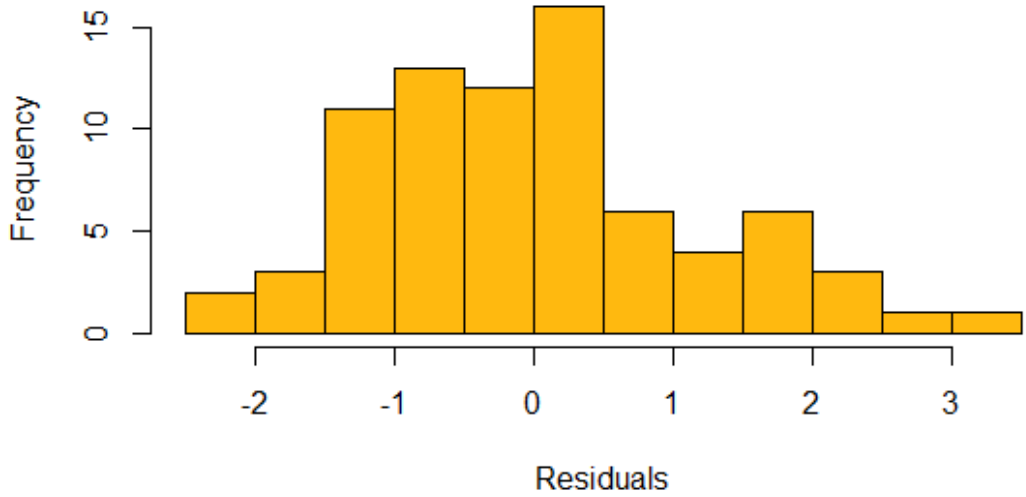

Table 5: ANOVA model estimates

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Continent	1	7.065	7.065	16.047	0.0002
PoliticalRegime	1	10.274	10.274	23.336	7.74e-06
Independence	1	0.732	0.732	1.663	0.2015

Continent:PoliticalRegime	1	1.601	1.601	3.637	0.061
Continent:Independence	1	0.004	0.004	0.008	0.9286
Politicalregime:Independence	1	0.826	0.826	1.877	0.175
Continent:Politicalregime:Independence	1	2.081	2.081	4.7272	0.0332
Residuals	70	30.818	0.440		

Table 6: Logistic regression model estimates

	Model 1	Model 2
Intercept	-1.04 (2.64)	14.62 (2797.44)
Continent 2	0.06 (2.62)	-0.27 (3.48)
Continent_3	-1.58 (2.68)	-4.48 (3.85)
Continent 4	-0.13 (2.65)	-1.16 (3.52)
Continent 5	-1.10 (2.68)	-1.98 (3.49)
Continent 6	14.26 (1696.74)	15.06 (2797.44)
Politicalregime 1	-1.27 (1.20)	-1.72 (1.61)
Politicalregime 2	0.94(0.98)	1.12 (1.46)
Politicalregime 3	3.34 (1.21)	4.10 (1.68)
Independence 2		-15.10 (2797.44)
Independence 3		-15.96 (2797.44)
Independence 4		-15.81 (2797.44)
Independence 5		-16.22 (2797.44)
Independence 6		-12.87 (2797.44)
Independence 7		-12.36 (2797.44)
Independence 8		-14.39 (2797.44)
Independence 9		-15.46
N	78	78
AIC	71.52	79.33
BIC	92.73	119.39
Pseudo – R2	0.6	0.68

References

- Barro, R. J. (1989). A cross-country study of growth, saving and government. NBER, working paper number 2855.
- Barro, R. J. (1990). Government spending in a simple model of endogeneous growth. Journal of Political Economy, 98(5, Part 2), S103-S125.
- Barro, R. J. (1991). Economic growth in a cross section of countries. The Quarterly Journal of Economics, 106(2), 407.
- Barro, R. J., & Xavier Sala-i-Martin. (1992). Convergence. Journal of Political Economy, 100(2), 223–251.
- Goldfeld, S. M., & Quandt, R. E. (1965). Some Tests for Homoscedasticity. *Journal of the American Statistical Association, 60*(310), 539–547.
- Miljkovic, D. & Rimal, A, (2008). The impact of socio-economic factors on political instability:A cross-country analysis. The Journal of Socio-Economics, 37(6), 2454-2463.
- Solow, R. M. (1956). A contribution to the theory of economic growth. The Quarterly Journal of Economics, 70(1), 65.
- White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica, 48*(4), 817–838.