Survival Analysis of GDP Per Capita Doubling Times by Country

Joe Margolis, Erik Hager, Declan Elias

December 13, 2022

1 Introduction

We aim to examine the effects of government type and rural population percentage on the growth of per capita GDP for 128 countries from 1960 to 2021. Both neoclassical economic theory and recent economic growth literature motivated the choice of our two explanatory variables. Prevailing theory suggests a model for GDP growth built on the institutions, level of physical capital, and level of human capital in any given country, and this traditional model has been supported by empirical evidence from numerous researchers.

The relationship between growth and human capital in the form of education is extensively documented across economic literature, with numerous studies finding strong correlations. This motivated our choice of secondary school completion as a covariate in our modeling. However, the significant sparsity of this data, as well as the lack of available global education in general, motivated our search for another variable as a proxy for human capital. Barkley et al. (2004) examined the relationship between human capital and growth in various contexts, focusing on growth correlations in rural areas. They found that rural areas had lower levels of human capital than local metropolitan areas, and also that higher human capital levels within rural areas corresponded with higher growth rates. Their findings motivated our use of the readily available rural population percentage metric as a growth indicator in our model, due to the supposed negative correlation between rurality and human capital.

Empirical support for the use of institutions in predicting growth is less documented than for human capital, due to difficulty in settling on an adequate metric to use. Esfahani and Ramirez (2003) found a substantial positive correlation between institutions and growth, building a model of institutions based primarily on national infrastructure capital. Ideally, we would build a similar model, incorporating both the capital-based and institutional elements of the theoretical growth model into an explanatory variable. However, the World Bank lacks sufficient data on measurements of capital and infrastructure across countries from 1960-2021. Instead, we decided to explore the governmental element of institutions as a predictor of growth. Glaeser et al. (2004) explored the use of political institutions as a growth predictor, noting that reverse causality has often been ignored in previous studies. They found evidence that dictatorships tend to initiate growth in poor countries, which subsequently leads to a reform of political institutions. This motivated us to look for an approach that sequentially puts a measure of government before the corresponding growth metric, leading to our inclusion of a country's government type preceding a period of growth as a predictor in our model.

In Section 2, we describe the data and variables we used to build our growth models. In Section 3, we provide some preliminary results and figures from the initial analyses of our data.

2 Data and Methods

2.1 GDP Doubling Time and Threshold

For this project we implemented three separate data sources, the first of which came from the World Bank Group, a group which is made up of five worldwide banking institutions to collect data from as many different companies as possible (The World Bank). The main goal of the original starter data set was to find our main outcome variable, which is time until gross domestic product (GDP) per capita doubled, standardized by U.S. Dollars with inflation levels from 2015. Since the starting

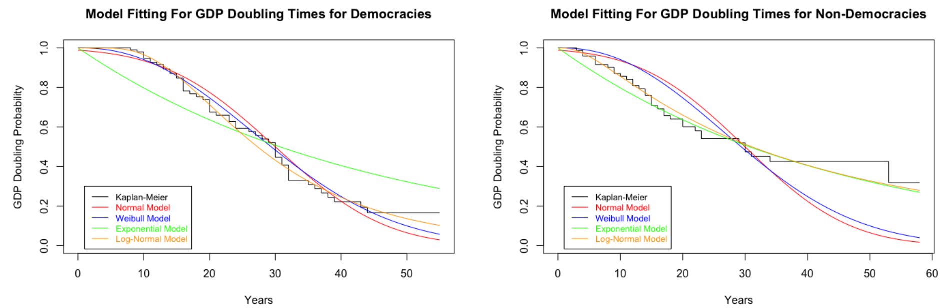

Figure 1: This figure displays the comparison between parametric models and Kaplan-Meier curves for Democratic countries and Non-Democratic Countries.

years of each country's data differed, we decided on starting each country's start time at the first year they entered a certain level of economic standing. To do this we had to create a 'Threshold' variable which grouped countries into a categorical economic standing based on their starting GDP per capita, placing each country in at least one of the groups (countries that entered a new threshold were given another observation for their doubling time after their upgrade): Low Income ($1,000-$2,000), Low-Mid Income ($2,000-$5,000), Upper-Mid Income ($5,000-$10,000), Developed ($10,000-$20,000), Wealthy ($20,000+). Entering each threshold gave us our initiating point, and the years until the GDP per capita at that point doubled became our main output variable, with all censoring relating to countries that had not doubled their GDP by 2021, the final year the data set contained.

2.2 Modeling by Education

Despite the extreme sparsity of global education data, we decided to explore the supposedly robust relationship between education and growth suggested by economic literature. We look at the percentage of those in a country 25 and older who have completed lower secondary school, which we obtained from the World Bank.

2.3 Modeling by Government Type

Next we began to study new predictors to truly investigate why GDP doubling times vary by different countries. The next we studied was the type of government. To do this we joined a data set uploaded to GitHub by Bastian Herre, an employee for Our World in Data, which included a variable called 'regime_bmr_owid', which gave a flag for whether or not a country was a democracy (Bastian Herre Github). In addition, Wikipedia research was used to manually find democracy statuses for the 48 observations which were missing from Herre's data set (Wikipedia).

Once we had the data joined, Kaplan-Meier curves were compared to exponential, normal, log-normal, and weibull models to see if certain parametric models describe the data well as seen in Figure 1. What was discovered form this data is, that both the weibull and log-normal models were the closest representations of each Kaplan-Meier model.

To continue our investigation we then created cox-snell residuals for weibull and log-normal models. Figure 3 shows these Cox-Snell residuals for non-democracies, and what we can see is that while both display a pretty linear relationship nearing a slope, of one, but the log-normal shows a slightly closer relationship. This idea is backed up by the comparison of log-likelihoods in which the log-normal has a log-likelihood of -162.1, slightly outperforming the weibull model's log-likelihood of -165.1.

The same story was seen when checking models for countries deemed as a democracy. Once again, the Cox-Snell residuals for both, hold close to the linear relationship with slope of one, but log-normal shows to slightly outperform the weibull, as seen in Figure 3, and once again the log-likelihoods support this, with log-normal outperforming the weibull -222.8 to -226.4. For this reason, we decided upon log-normal distribution for both democracies and non-democracies, displaying right skewed distributions for each.

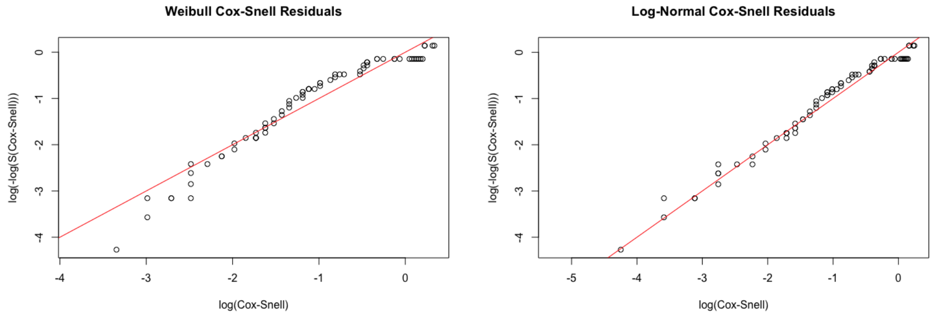

Figure 2: This figure displays Cox-Snell residuals for the weibull and log-normal models of non-democracies.

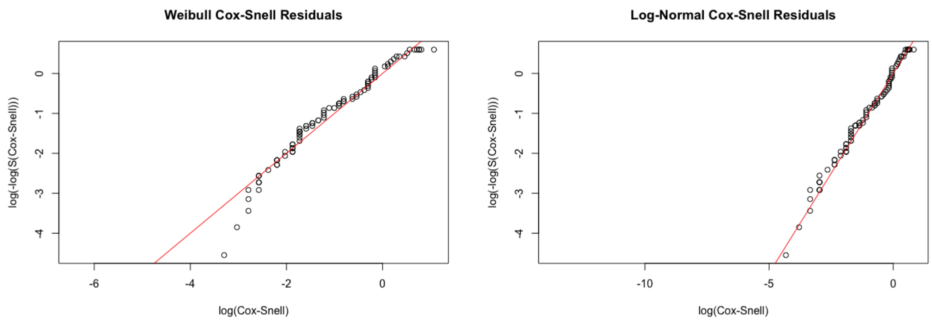

Figure 3: This figure displays Cox-Snell residuals for the weibull and log-normal models of democracies.

2.4 Modeling by Rural Population

We then decided to explore if the rural population had any relationship with the time it took for a country's GDP to double. We continued using years until GDP doubled as our explanatory variable, but added a new predictor, 'RuralPopulation' taken from the World Bank. This gave us a measure of the percentage of the population living in rural communities.

The first step in analyzing the data was to create a categorical variable based on the percentage of rural population. We split the population into 4 groups separated by the 4 quartiles: [0, 30.61%], (30.61%, 44.23%], (44.23%, 57.74%], (57.74%, 100%]. From there we fit a Kaplan-Meier curve and compared it to exponential, normal, and weibull models. As shown in Figure 4, the normal, log-normal and weibull models do a good job of approximating the Kaplan-Meier curve.

To explore the parametric models further, we create Cox-Snell residual plots for each model. We then select the distribution that fits the best, and perform a log-likelihood test to check if the inclusion of Rural Population is significant.

3 Results

3.1 Visual Analysis of Education

Due to the lack of observations present for our measure of educational attainment, we decided to stick to a brief visual analysis of the relationship between education and growth. Figure 5 shows a scatterplot of values for educational attainment versus GDP doubling times, and indicates a negative relationship. We recognize that this plot doesn't account for the censoring of doubling times, and cannot suggest results of statistical significance. We merely wish to include this visual as grounds to help imagine the building of a useful educational model of growth given less sparse data.

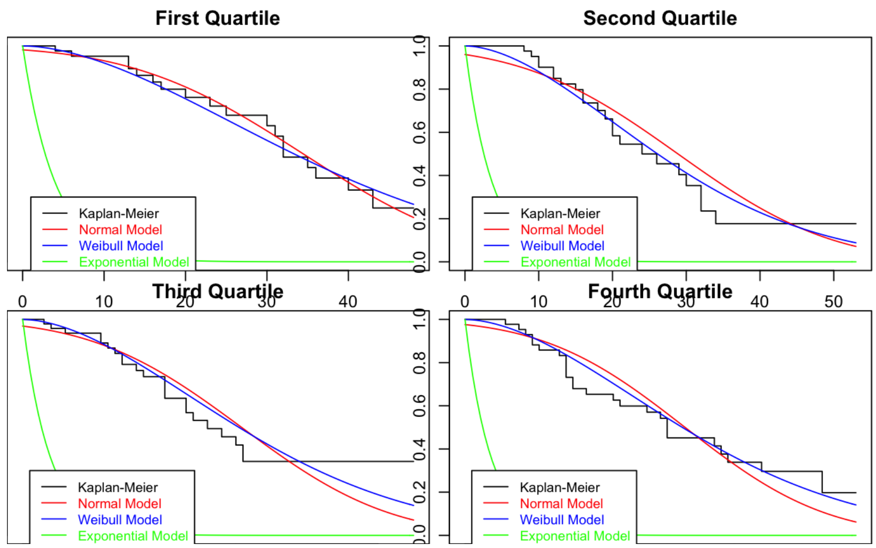

Figure 4: This figure displays the Kaplan-Meier, normal, weibull, and exponential curves for the 4 quartiles of the Rural Population data

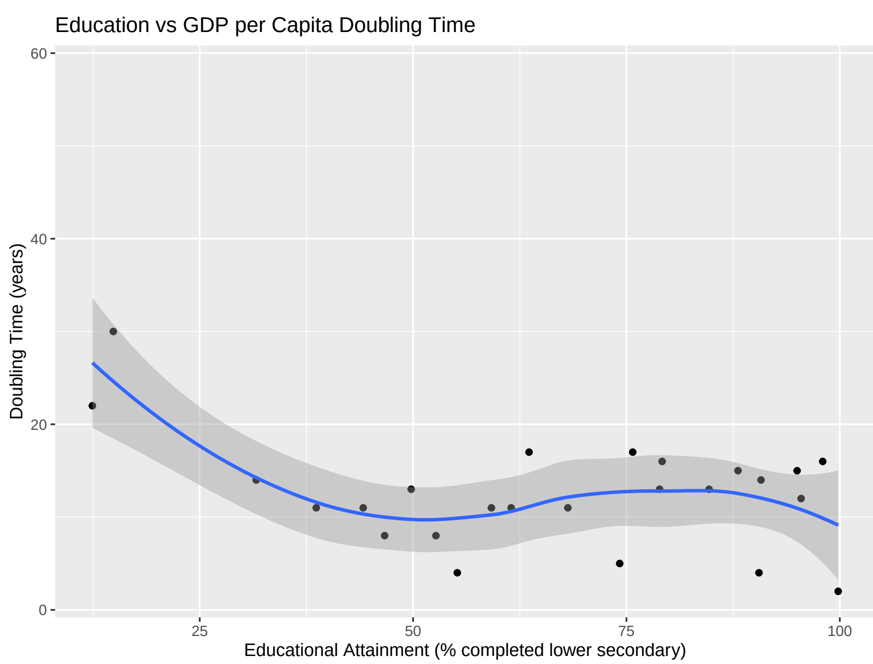

Figure 5: This figure shows a scatter plot and smoothed line of doubling times vs educational attainment.

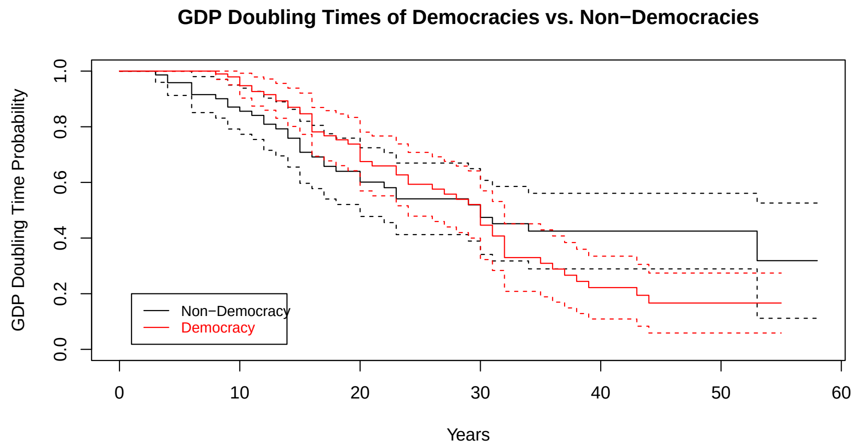

Figure 6: This figure displays the comparison between Kaplan-Meier curves for Democratic countries and Non-Democratic Countries.

Log Rank Table Comparing GDP Doubling Times for Democracies and Non-Democracies

Democracy Status	Number of Countries	Observed GDP Doubling Time	Expected GDP Doubling Time	(O-E)^2/E	(O-E)^2/V
Non-Democracy	75	34	36.7	0.194	0.355
Democracy	111	52	49.3	0.144	0.355

Statistical Significance Output Statistics

Test Statistic	P-Value	Hazard Ratio Point Estimate	Hazard Ratio 95% Confidence Interval
0.338	0.5609859	0.8783274	[0.5728957-1.346596]

Figure 7

3.2 Results by Government Type

To begin assessing the difference in survival experiences for democracies and non-democracies, we produce Figure 6 which displays the Kaplan-Meier curves for each, where each drop in a curve indicates countries doubling their GDP per capita. What we notice in this figure is that non-democracies which display early success, actually expand economically quicker than democracies. Eventually though, if non-democracies are not successful early, democracies in later years become more likely to actually reach the ending point of their GDP doubling time as we notice the intersection where the trends flip around the 30 year without doubling mark. It is important to note though that the overlap in confidence intervals denoted by the dashed lines displays that there is not actually enough information to notice a statistically significant difference in their survival experiences.

To continue to investigate if there is a difference in survival experiences of democracies and non-democracies, we produced a log-rank test on the data, continuing to use GDP doubling time as the output and grouping by democracy or non-democracy. The output of this test is shown in the top table of Figure **??**. From this test we used the Mantel-Cox method to produce a test statistic, p-value, and hazard ratio for the data, the results of which are displayed in the table beneath the log rank table. Figure **??** also includes the results of this table, both the p-value of 0.5609859 (observed at threshold $\alpha = 0.05$) and the hazard ratio confidence interval of [0.5728957 - 1.346596] support what was noticed in the Kaplan-Meier curve confidence intervals. Since the p-value is much greater than 0.05 and the hazard ratio contains 1, when there is no stratification applied, we accept the null hypothesis that there is not a statistically significant difference between the time it takes for democracies and non-democracies to double their GDP per capita.

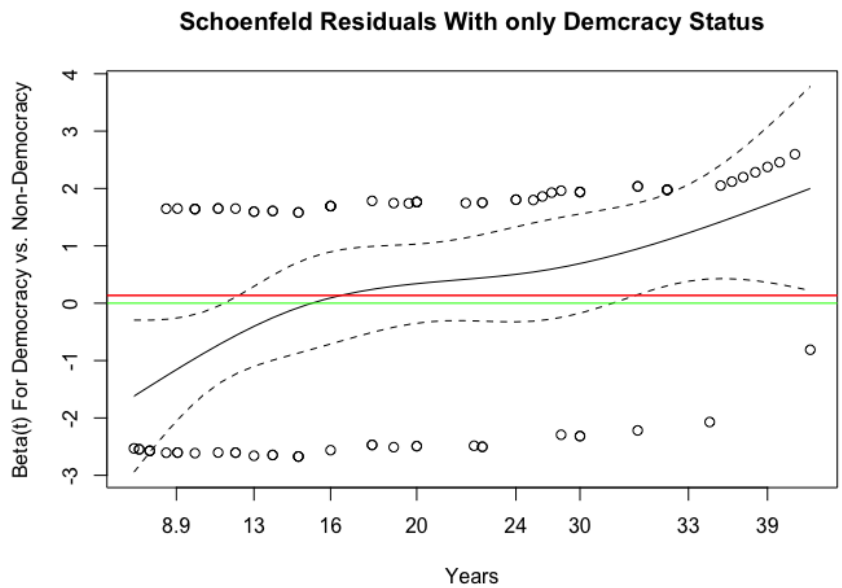

Figure 8: This figure displays the Schoenfeld Residuals for a Cox-PH model which only uses democracy status as a covariate

That being said, the fact that the Kaplan-Meier curves intersect displays cancellation between the models that leads to the two models showing statistical insignificance in the presence of practical significance. To investigate this we developed Shoenfeld residuals for a Cox-PH model, first only using democracy status as a covariate in the model. While we can already cancel out proportionality of hazards from the survival curves, we can still use the Schoenfeld residuals to estimate the change in hazard ratios over time. Figure 8 backs up the story described by our Kaplan-Meier curves from Figure 1. The green horizontal line represents where the log of the hazard ratio between democracies and non-democracies is at zero, and therefore where the chance of a country reaching their GDP doubling point is not effected by democracy status. The graph shows that early on, non-democracies are outperforming democracies and have a higher chance of reaching their GDP doubling point. This lasts until slightly over 13 years since their starting GDP time, when the intersection occurs and democracies begin to outperform non-democracies in growth rate.

In order to improve the accuracy of the model, we also accounted for the confounding variable we found to have significance by stratifying for threshold. When including threshold, we still noticed a similar relationship, with non-democracies outperforming democracies in their growth rate earlier into their tenure, with that comparison flip-slopping slightly past 13 years as is evident in Figure 9. The main difference lies in the time where democracy has overtaken non-democracies, as opposed to continuing to expand the gap in hazard ratios that was seen in Figure 8, these residuals show that democracies tend to even out and show a proportional relationship in hazard ratios to non-democracies as time continues.

Since it is apparent that the proportional hazards assumption does not apply to the difference between democracy and non-democracy GDP doubling times, the next step was to model the change in the hazard ratio over time. While 'regime_bmr_owid' is not a time varying coviariate in this data set, there is still an obvious time varying effect of the variable on the hazard ratio. To model this we ran Cox-PH models on 'regime_bmr_owid', treating it this time as a time varying covariate. Figure 10 displays the output coefficients from these model, with the top table representing the model that only includes our democracy declaration variable and the bottom table representing the model that stratifies for threshold. In the top table, the first exponentiated coefficient represents the multiplicative affect on the hazard function at time 0 when changing from a non-demoracy to a democracy. This tells us that at the beginning of a country's GDP Doubling Time, if they are a democracy, the likelihood that they reach their GDP doubling time at that moment is only 0.21506 times that of a non-democracy, a very significant difference as is represented by the P-Value of 0.002929. The second exponentiated

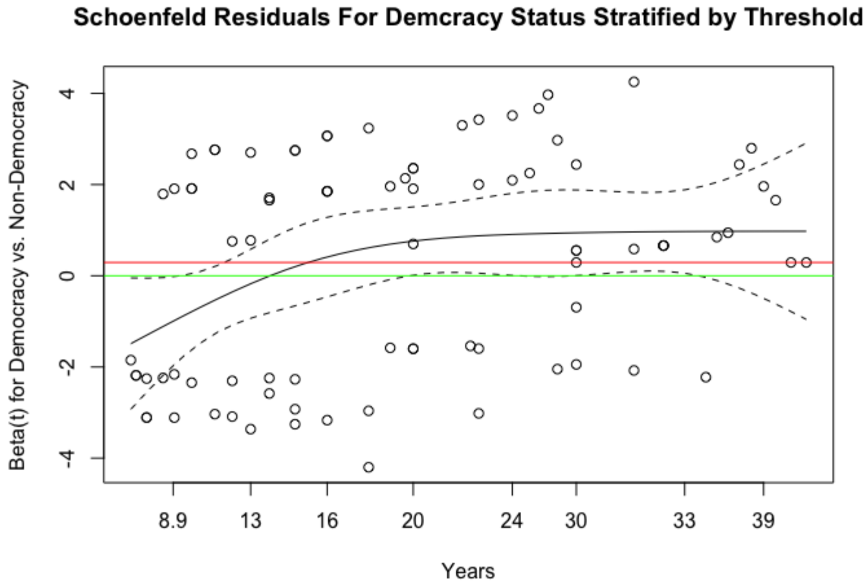

Figure 9: This figure displays the Schoenfeld Residuals for a Cox-PH model which uses democracy status as a covariate aloong with stratifying for threshold

Time Varying Coefficients of Cox-PH Model

Time Varying Representation	Coefficient	Exponentiated Coefficient	P-Value
regime_bmr_owid = 1	-1.53685	0.21506	0.002929
regime_bmr_owid = 1:Stop	0.08400	1.08763	0.000562

Time Varying Coefficients of Cox-PH Model Stratified by Threshold

Time Varying Representation	Coefficient	Exponentiated Coefficient	P-Value
regime_bmr_owid = 1	-1.39743	0.24723	0.02247
regime_bmr_owid = 1:Stop	0.09397	1.09853	0.00336

Figure 10

coefficient represents how that hazard ratio changes over time. With a value of about 1.08763, we determine that with each year that passes, the hazard ratio gets multiplied by 1.08763. Knowing this we are then able to calculate that non-democracies are more likely to reach their GDP Doubling Time than democracies up until 18.29583 years have passed, when the two hazard functions intersect and democracies become more likely to double their GDP than non-democracies.

The second table in Figure 10, that represents the same type of model only stratified by threshold, shows a similar story. According to this model, at the initiating point of a country's GDP doubling time a country that is a democracy has only 0.24723 times the probability of doubling their GDP, but every year that continues that ratio is multiplied by 1.09853. The stratified model does show a pretty substantial difference in the time it takes for the hazard functions to intersect, as it calculates that at 14.87102 years into the GDP doubling time a democracies and non-democracies will have an equal chance of doubling their GDP, and from that point on democracies become more likely. It is also important to note the p-values for all of the coefficients in this table, particularly the 'Stop' variables show statistical significance, reinforcing the claim that the proportional hazards assumption is not applicable to democracy status and 'regime_bmr_owid' does have a time varying effect on GDP Doubling Time.

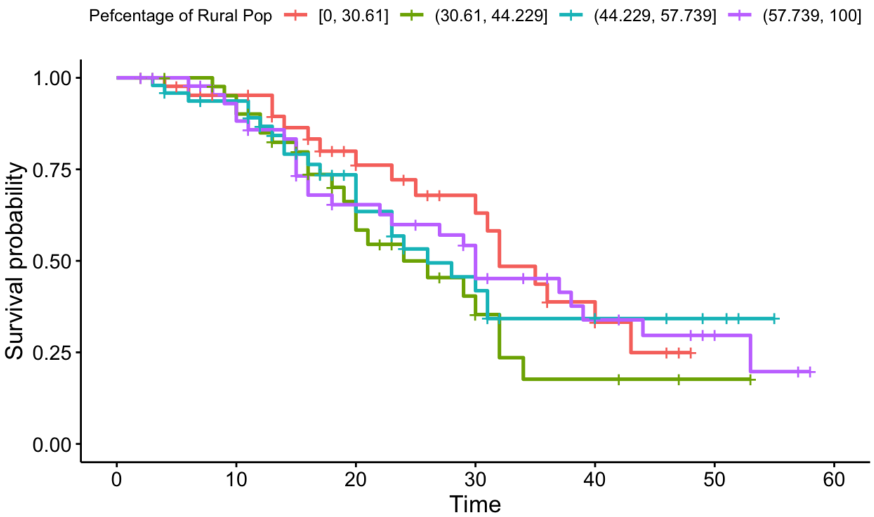

Figure 11: This figure displays the comparison between Kaplan-Meier curves for the 4 quartiles of the Rural Population data

Log Rank Table Comparing GDP Doubling Times for RuralPopulationCat Quartiles

Rural Population Category	Number of Countries	Observed GDP Doubling Time	Expected GDP Doubling Time	(O-E)^2/E	(O-E)^2/V
[0, 30.61]	44	18	22.0	0.71261	0.98553
(30.61, 44.229]	48	22	17.6	1.08878	1.42192
(44.229, 57.739]	48	22	21.6	0.00636	0.00871
(57.739, 100]	48	26	26.8	0.02361	0.03562

Figure 12

3.3 Results by Rural Population

After fitting the Kaplan-Meier to the 4 different quartiles we were able to produce the visualization in Figure 11. The figure does not show enough to suggest a difference in survival curve for the different quartiles as well as significant overlap in the confidence intervals.

To confirm what we saw in the visualization we carried out a log-rank test to compare the different survival curves. The output is shown in Figure 12. We used the Mantel-Cox method to produce a test statistic: 1.83136 and p-value: 0.6081345. The results in this table, both the p-value of 0.6081345 (observed at threshold $\alpha = 0.05$) support what was noticed in the Kaplan-Meier curve confidence intervals. There is no statistically significant difference between the time it takes for countries to double their GDP per capita based on how rural the population is using a non-parametric approach.

We wanted to check if the conclusion from the non-parametric model held up with AFT models. We fit a gaussian, exponential, weibull, and lognormal model. To check the fit, we graphed the Cox-Snell residuals for each distribution. The Cox-Snell residuals for the weibull and log-normal models, shown in Figure 13 and Figure 14 respectively, both were approximately linear with slope 1. So to decide which distribution to use, we compared AIC's. We selected the log-normal model with an AIC of 842.6, compared to the AIC of 834.6 for the weibull model.

When then performed a Likelihood Ratio Test to decide if the inclusion of RuralPopulation as a covariate was necessary. We failed to reject the null hypothesis of the simpler model with just the intercept over the alternate hypothesis of the model with the addition of RuralPopulation with a Test Statistic of .2 and a p-value of 0.6547208. Thus we come to the same conclusion with parametric models as we did with a non-parametric model: rural population has no statistical significance.

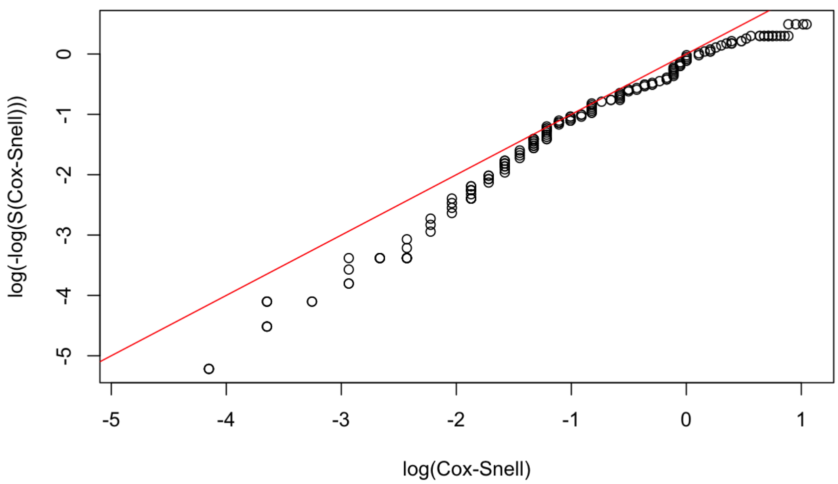

Figure 13: Cox-Snell Residual of the weibull RuralPopulation model

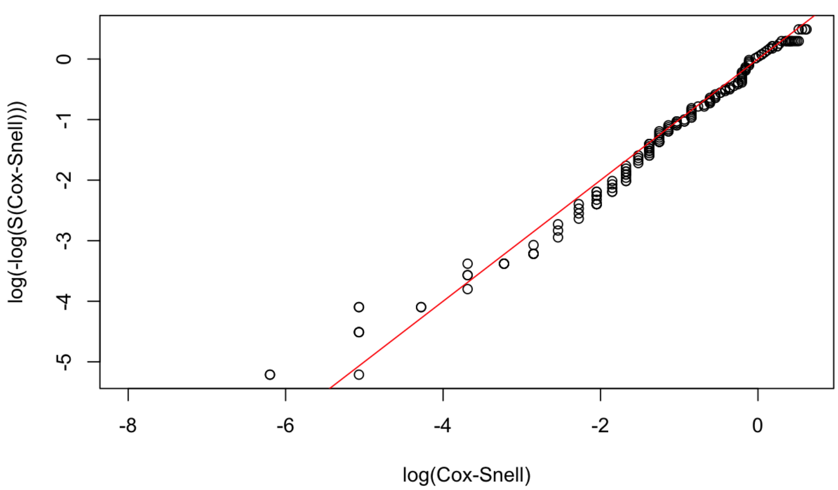

Figure 14: Cox-Snell Residual of the log-normal RuralPopulation model

4 Discussion

4.1 Government Type

Although we found an insignificant effect of government type on doubling time using both non-parametric and log-normal models, we have yet to account for time-varying effects of government type. Literature suggests that both the shortest and longest doubling times happen in non-democracies, with democratic countries in the middle. This could be the case in our data, meaning that this effectively cancels out any significance in our time-constant model. If this does turn out to be the case, this would provide an interesting piece of the puzzle of institutions' effect on growth.

4.2 Rural Population Percentage

We found no evidence of any significant effect of rural population percentage on growth, robust across numerous tests. We chose to look at rural population percentage because we were curious whether might serve as a good proxy for the level of human capital in a given country. Theory suggests human capital as a core element of output growth, and use of education metrics as a representation of human capital is well documented in economic literature. Our findings suggest, contrary to our initial hypothesis, that rural population percentage is likely a much poorer proxy for human capital than education metrics.

5 Conclusion

We aimed to model the time taken for various countries' per capita GDP to double from 1960 to 2021, using predictors motivated by economic theory. Economics suggests human capital and institutions as core components of growth, so we chose test variables we thought might represent these concepts well and analyze their effects on GDP growth. We ended up deciding to model human capital's growth effects using rural population percentage. However, we found no evidence of any significant effect of rural population percentage on growth.

For institutions, we chose a variable classifying a country's government type as either democratic or non-democratic. Our preliminary findings show no effect of this variable on output growth, although we plan to extend our analysis to include time-varying effects in hopes of exploring evidence we found suggesting this might unveil a significant effect.

6 References

1. Barkley, David L., Mark S. Henry, and Haizhen Li. "Does human capital affect rural economic growth?" Evidence from the South. No. 630-2016-41390. 2004.

2. Glaeser, Edward L., et al. "Do institutions cause growth?." Journal of economic Growth 9.3 (2004): 271-303.

3. Esfahani, Hadi Salehi, and Maria Teresa Ramirez. "Institutions, infrastructure, and economic growth." Journal of development Economics 70.2 (2003): 443-477.

4. Herre, Bastian. "Democracy Datasets." https://github.com/owid/owid-static

5. "Lists of Countries and Territories." https://en.wikipedia.org/wiki/Listsofcountriesandterritories

6. The World Bank. https://databank.worldbank.org/source/world-development-indicators