A Survival Analysis of Global GDP Per Capita Growth

Joe Margolis, Erik Hager, Declan Elias

February 21, 2023

1 Introduction

We aim to examine the effects of education, rural population percentage, and government type on the growth rates of per capita GDP for 128 countries from 1960 to 2021. Both neoclassical economic theory and recent economic growth literature motivated the choice of these predictors.

The generally-accepted Augmented Solow Model of economic growth includes human capital as a key component in driving a country's total economic output (Mankiw et al., 1992), and this compelled us to look at measures of human capital in our analysis. The relationship between growth and human capital is typically characterized with some measure of education as a proxy for human capital, so we decided to include a measure of secondary school completion as a small part of our analysis. However, the extreme sparsity of this data, as well as the lack of available global education data in general, motivated our search for another variable to use as a proxy for human capital for our primary analysis. Barkley et al. (2004) examined the relationship between human capital and growth in an interesting context, focusing on growth correlations in rural areas of the United States. 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 economic growth rates. Their findings motivated our use of a measure of countries' rural population percentages as a predictor for growth in our modeling, since the supposed negative correlation of rurality with human capital suggests it could be used as an inverse proxy for it.

The Solow model is set up to exclude the effects of institutions such as government, though there is some debate about whether institutions may affect growth in addition to the inputs included in the augmented model. Glaeser et al. (2004) explored possible effects of political institutions on growth, in addition to the effects of human capital. Although they suggest human capital is a more fundamental source of growth than institutions, they found evidence that dictatorships can help initiate growth in poor countries, which tends to subsequently lead to a reform of political institutions in order for growth to continue. This motivated the extension of our analysis to include a measure of countries' government types as a potential predictor of growth.

We chose to measure economic growth rates with the doubling time in years of countries' real GDP per capita at various income thresholds, and used this data as the outcome variable in each of our separate models centered around human capital and political institutions.

In Section 2, we describe the data and variables we used to build our growth models. We provide the results of our analyses in Section 3 and a discussion of these results in Section 4.

2 Data and Methods

2.1 GDP per Capita Doubling Time and Income Thresholds

To construct our time-status outcome variable, we collected World Bank data from 1960 to 2021 on the GDP per capita of 128 countries, chained to 2015 U.S. dollars. Next, we classified each data point into an income threshold based on its value as follows: Low Income ($1,000-$2,000), Lower-Middle Income ($2,000-$5,000), Upper-Middle Income ($5,000-$10,000), Developed ($10,000-$20,000), Wealthy ($20,000+). We then recorded the first year each country entered a given threshold, as well as the GDP per capita value at that year, and tracked the country across the years to find the first year during which the country reached double that value. We assigned the difference between the starting year and ending year over these periods of doubling for each country at each threshold to be the values

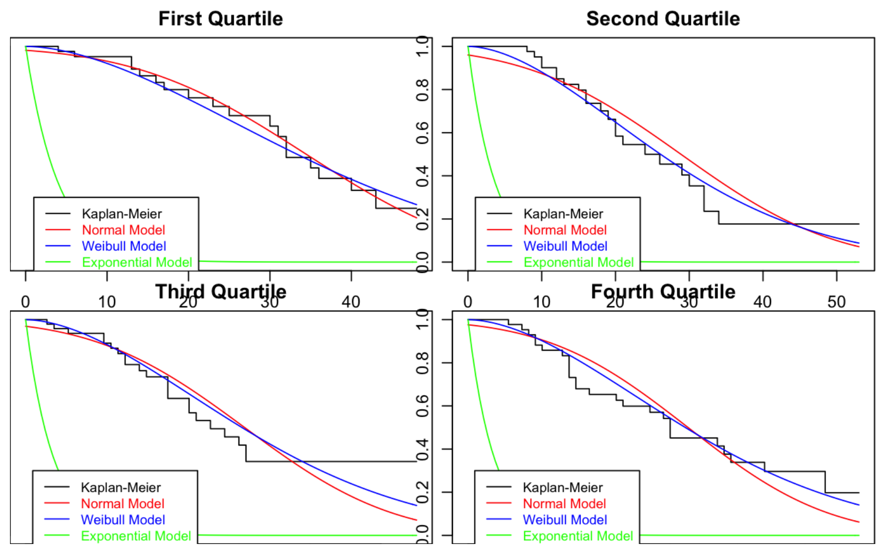

Figure 1: This figure displays the Kaplan-Meier, normal, weibull, and exponential curves for the 4 quartiles of the Rural Population data

for our "doubling time" outcome variable. Finally, we included a status variable to keep track of data that represented a failure to double by 2021 and denote it as right-censored.

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. In a brief visual analysis, we use the World Bank to look at the percentage of those in a country 25 and older who have completed lower secondary school and how this correlates with our outcome variable.

2.3 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 1, 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.

2.4 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).

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

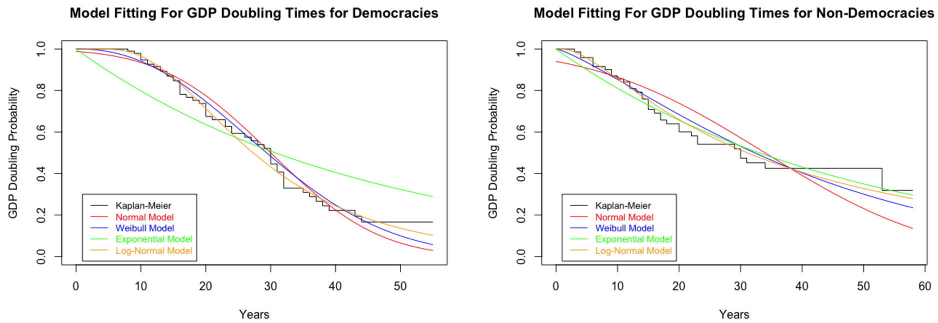

Figure 2: This figure displays the comparison between parametric models and Kaplan-Meier curves for Democratic countries and Non-Democratic Countries.

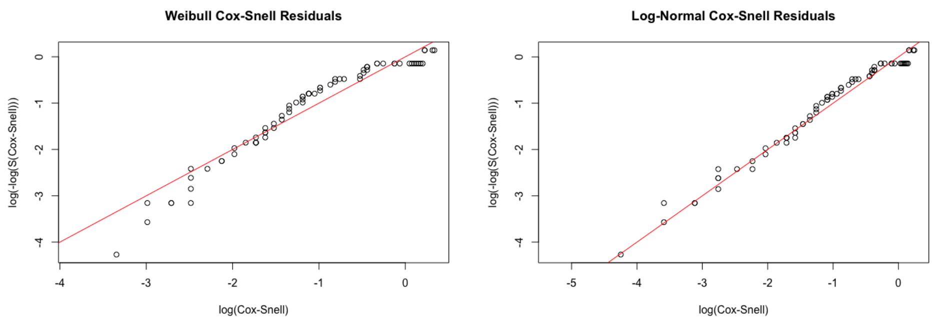

Figure 3: This figure displays Cox-Snell residuals for the weibull and log-normal models of non-democracies.

2. 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 4 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 4, 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.

3 Results

3.1 Results by 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.

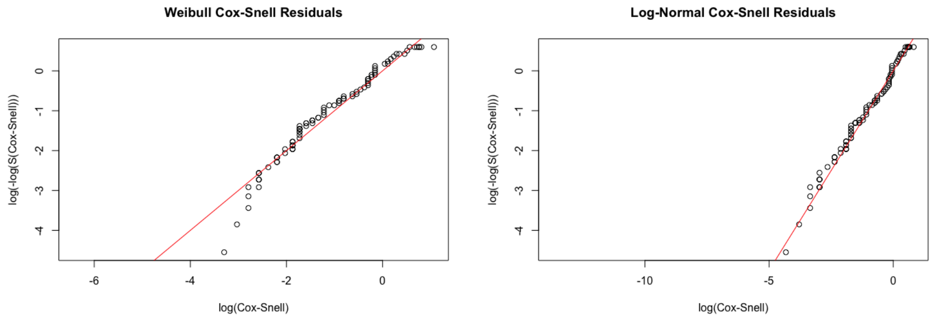

Figure 4: This figure displays Cox-Snell residuals for the weibull and log-normal models of democracies.

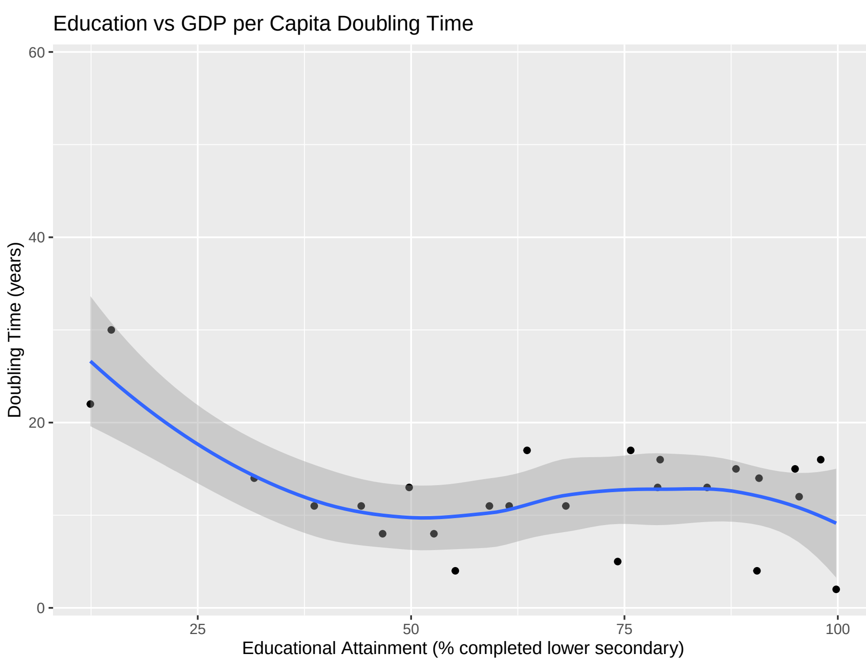

Figure 5: This figure shows a scatter plot and smoothed line of doubling times vs educational attainment.

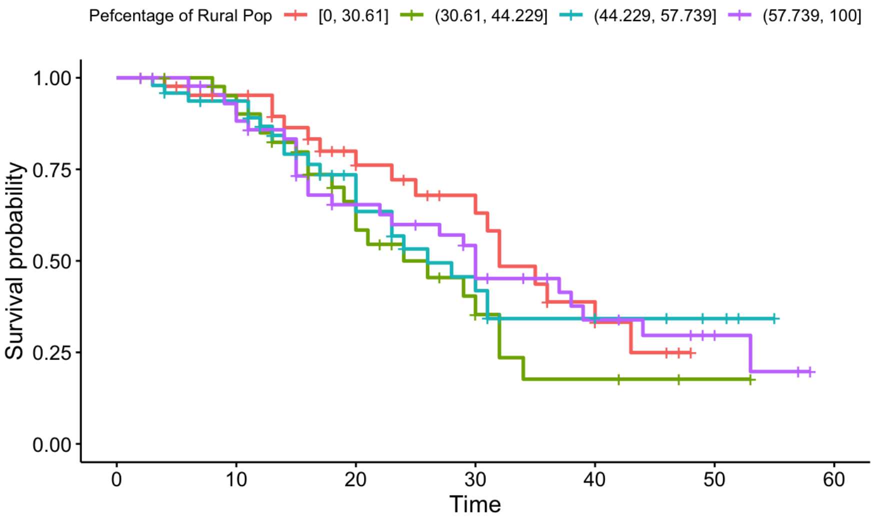

Figure 6: This figure displays the comparison between Kaplan-Meier curves for the 4 quartiles of the Rural Population data

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 7

3.2 Results by Rural Population Percentage

After fitting the Kaplan-Meier to the 4 different quartiles we were able to produce the visualization in Figure 6. 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 7. 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 8 and Figure 9 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 834.6, compared to the AIC of 842.6 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.

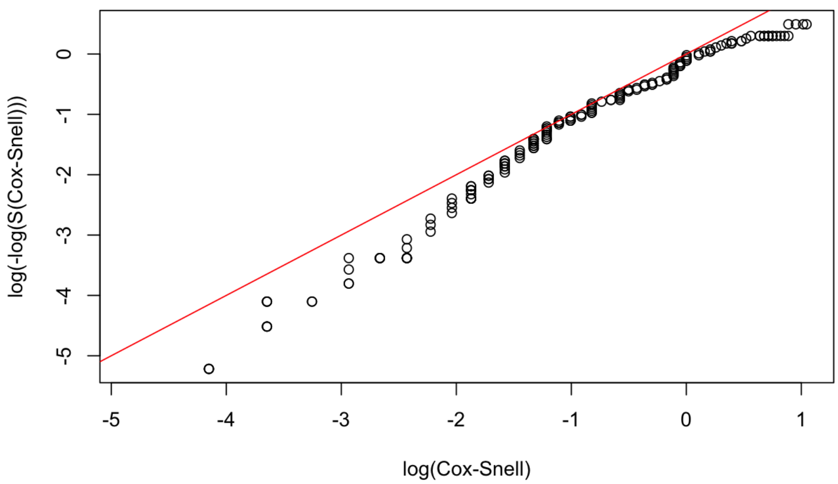

Figure 8: Cox-Snell Residual of the weibull RuralPopulation model

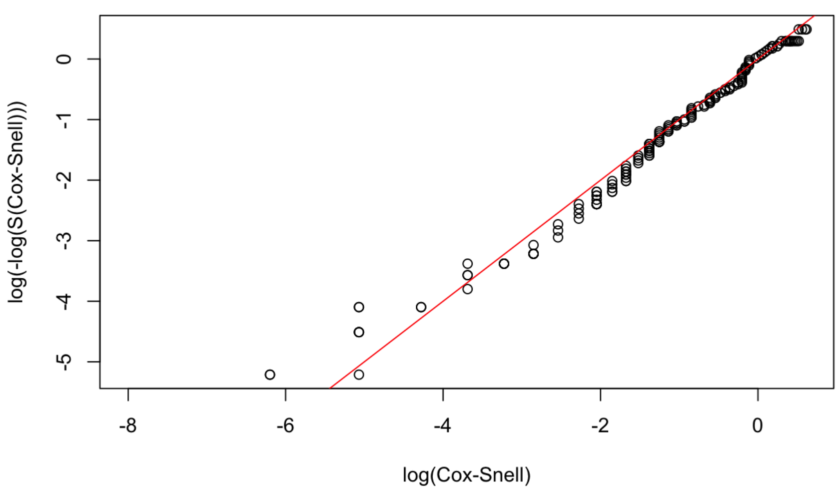

Figure 9: Cox-Snell Residual of the log-normal RuralPopulation model

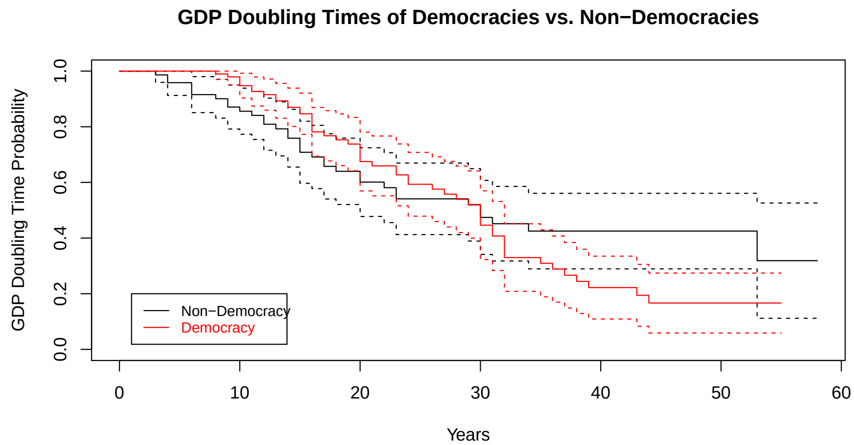

Figure 10: This figure displays the comparison between Kaplan-Meier curves for Democratic countries and Non-Democratic Countries.

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 11

3.3 Results by Government Type

To begin assessing the difference in survival experiences for democracies and non-democracies, we produce Figure 10 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 11. 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 11 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.

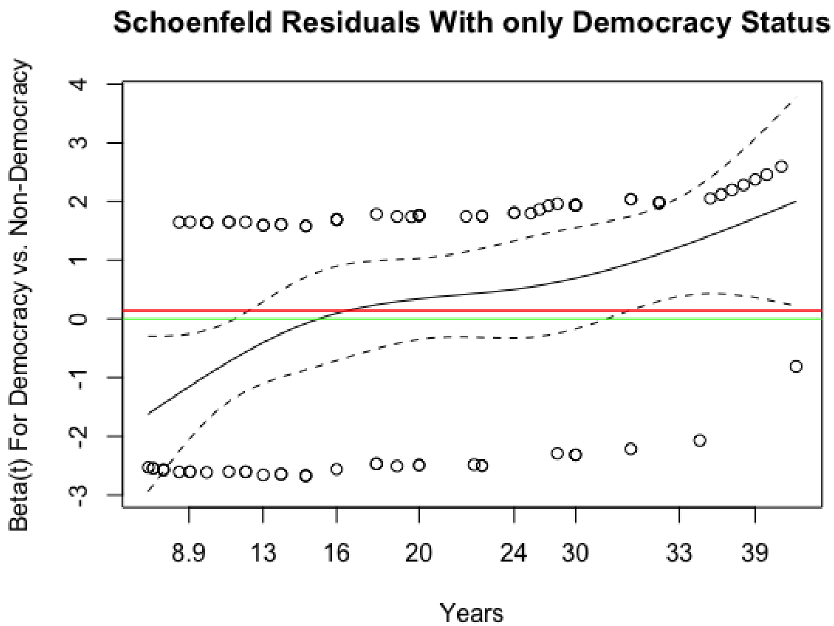

Figure 12: This figure displays the Schoenfeld Residuals for a Cox-PH model which only uses democracy status as a covariate

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 12 backs up the story described by our Kaplan-Meier curves from Figure 2. 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 13. 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 12, 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 covariate 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 14 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-democracy 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

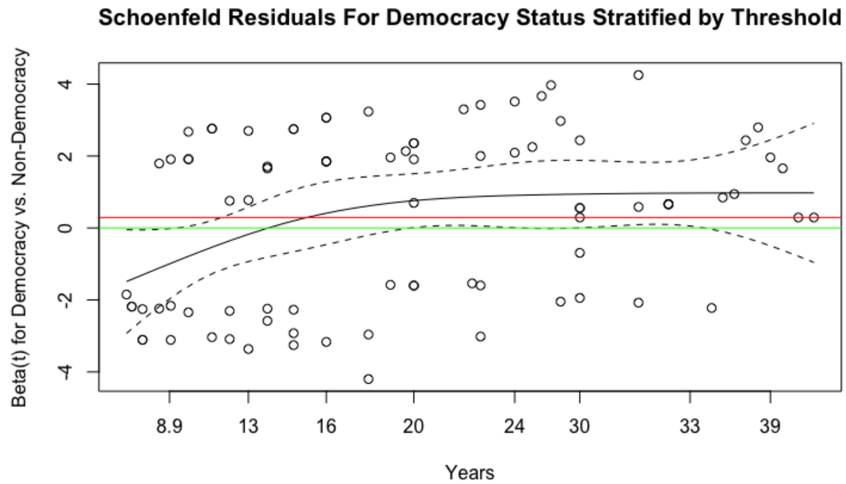

Figure 13: This figure displays the Schoenfeld Residuals for a Cox-PH model which uses democracy status as a covariate along with stratifying for threshold

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 14

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 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 14, 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.

4 Discussion

4.1 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.

4.2 Government Type

Although we found an insignificant effect of government type on doubling time using both non-parametric and log-normal models, further investigation found that this was due to overlapping in the survival curves and hazard functions of democracies and non-democracies. This led us to study the time varying effect of government type on GDP doubling time. Literature suggests that both the shortest and longest doubling times happen in non-democracies, with democratic countries in the middle. Our study directly supported this theory as we noticed in the overlapping Kaplan-Meier curves, meaning that this effectively cancels out any significance in our time-constant model. We did find significance though when treating government type as time varying, finding that earlier in GDP doubling time non-democracies are more likely to double their GDP, but a favorable trend in the hazard ratio over time for democracies shows that later in GDP doubling time democracies overtake non-democracies higher growth rate. This allowed us to reject the null hypothesis, realizing that government type does have a significant effect on GDP doubling time for a country.

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. Literature suggests that human capital and institutions are potential drivers of economic growth, so we chose to test variables we thought might represent these concepts well and analyze their effects on GDP growth.

We decided to conduct only a visual analysis of education's effects on growth, and created a scatterplot of secondary school completion versus doubling times. Though there appears to be a negative relationship, we cannot draw any statistical conclusions from this.

We ended up modeling human capital's growth effects using rural population percentage. However, we found no evidence of any significant effect of a country's rural population percentage on its economic growth. Though this is contrary to the findings in the study done by Barkley et al., this might be explained by the differences in scope between the two analyses. Their study was conducted locally, analyzing differences in rurality within a particular region of a single country, while our study was conducted using global panel data. It is entirely possible that their findings do not generalize to our context, and that rurality is a poor proxy for human capital across countries.

For institutions, we chose a variable classifying a country's government type as either democratic or non-democratic, and conducted regressions to analyze its correlation with doubling times. We ended up finding a significant time-varying effect, in which non-democracies outperform democracies amongst countries with quicker doubling times, but democracies outperform non-democracies amongst those with slower doubling times. This could represent the effect discussed in the paper by Glaeser et al., in which dictatorships tend to initiate the quickest growth, but democracies have higher rates of sustained growth over longer periods of time.

Going forward, we would be interested in finding a better way to analyze the growth effects of human capital in the form of education, perhaps on a smaller scale where ample data can be found. We would also like to dig deeper into the question of political institutions, and explore why exactly it is that democracies and non-democracies exhibit growth experiences characterized by the time-varying effect we found.

6 References

1. N. Gregory Mankiw, David Romer, David N. Weil, A Contribution to the Empirics of Economic Growth, The Quarterly Journal of Economics 107.2 May 1992, Pages 407–437

2. 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.

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

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

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