

Assessing the Impacts of Economic Factors on Funding for Tech Startups

Nana Adjoa Ansah, Parunjodhi Munisamy, Michel Ruiz-Fuentes

May 03, 2022

Abstract

Our analysis aims to investigate the correlation between the funding stage of a tech startup (Series C or not), the GDP per capita of the country the startup is based in, and the funding amount they receive from a given investor. Our data comes from the WorldBank, a Continents CSV and a Kaggle dataset compiling startup information from TechCrunch, Crunchbase, etc. We combined the datasets to conduct a multiple regression analysis investigating the correlation between the variables listed above for tech startups in North America. This subject is fascinating because startup companies are crucial for the growth of a country's economy. Our research could provide insightful findings that are relevant and beneficial to all of us because more startups create a healthier economy. We find that GDP per capita has a positive linear relationship with funding amounts and companies in Series C have more funding amount than the other funding stages.

Introduction

Our team is investigating the research question:

How the funding amount a tech company receives is affected by the GDP per capita of the country it is based in and moderated by whether it is in the Series C funding stage or not.

We are assessing the correlation between the funding amount (startups receive from investors) and economic variables that can reveal trends that help tech startups raise capital and boost the country's economy. Furthermore, our research could show what economic factors contribute to the likelihood of tech companies receiving funding from investors. We are modeling our data with the intent to: classify, summarize patterns, and test a theory. First, we would like to classify what variables positively correlate with the funding amount that tech startups receive. Then, we will summarize our trends, and lastly, we will test the theory that Series C stage companies are more likely to receive funding from investors.

Background and Significance

As ventures or young and growing companies, these startups must gather capital to develop with their ideas. This funding will come from investors like venture capitalists, banks, hedge funds, or high net-worth individuals. The funding amount offered by an investor is based on several factors: their level of interest in the market, the company's valuation, and what the current funding stage reveals about the startup's growth potential. Each funding stage possesses distinct characteristics that demonstrate the company's track record and risk, hence its Valuation score. In our data analysis, we will investigate how the funding stage of a company affects the funding amount they receive from an investor.

Our first key variable will investigate if being Series C stage company has a moderating effect on the funding amount a startup receives. Series C companies are thriving businesses in their markets. They are lower risk and have consistent revenue from customers purchasing their products and services. When a Series C company is looking for an investor to provide funding, its objective is to expand to new markets and acquire other companies ("How the Stock Market Affects GDP" n.d.). Due to their reputation of success, they are more attractive and viable investments. When investors invest in a Series C company, they benefit from a high investment appreciation due to the successful nature of being a Series C startup ("LEF 2021 – Liberec Economic Forum" n.d.). Therefore we believe focusing on whether or not a company is in the Series C stage will be the most insightful.

Our second key variable explores whether the GDP per capita of the country the startup is based in affects the funding amount they will receive from an investor. Based on prior research, we found that a stronger GDP per capita allows for more venture opportunities. When a country's economy is well-developed, investors have more discretionary funding, which leads to higher levels of investment and the emergence of more fintech startups, with venture capital being readily available (Haddad and Hornuf 2019). With an increasing developing economy of a country, there is a higher likelihood of individuals needing financial services such as asset management. As a result, the stock markets of such countries become active, inadvertently having a positive influence on tech startup formations and supporting the prosperity of venture capital and entrepreneurship (Black and Gilson, 1999). With stock prices rising, investors and consumers have more wealth and optimism about future prospects. Eventually, there are increased sales and earnings for corporations, further boosting the GDP per capita of the country ("How

the Stock Market Affects GDP” n.d.).

As mentioned above, we will investigate the impacts of being a Series C stage company and the GDP per capita of the country the company is based in. We are interested in the relationship between these two factors because established financial markets are often key determinants to providing Series C funding to startup tech companies. Better financial markets would lead to economic growth and, therefore, GDP per capita in the country. Through the tremendous improvements in technology that blossom from the tech startups, there are robust boosts in the economic development of a country (“Financial Development” n.d.). Consequently, since a more stable and stronger financial structure is positively associated with economic growth (Prochniak and Wasiak 2017), holding everything else constant, on average, a company based in a country where the GDP per capita is high is more likely to be associated with being in the Series C funding stage because there would be more funding available for them to invest in.

Hypothesis

Therefore, our **two alternate hypothesis** are:

- 1) We expect that when the GDP per capita of the country the startup is based in increases, so does the funding amount.
- 2) We expect that the GDP per capita of the country the startup is based in and funding amount will increase significantly if the startup is in the Series C funding stage.

Methods

Data

Our main dataset is [Tech Company Fundings](#). The dataset owner is Shivam Bansal, an experienced data scientist who has won multiple Kaggle Analytics Competitions. They are active on Kaggle with numerous datasets in their toolkit. It consists of 3575 observations with 8 variables looking at different features of the tech company. Out of these 8 variables, we choose to focus on funding amount (key dependent variable), funding stage (key independent variable), vertical (independent variable) and region the company is based in. Its unit of analysis is a unique tech company.

To perform a more in-depth regression analysis, we decided to join this dataset with the [World Statistics Dataset](#). This dataset uses data from the World Bank and its unit of analysis is a specific country. To match this dataset, having 16104 observations, with our main one mentioned above, we are going to use the 2019 data from the World Bank dataset to collect information for each country about GDP per capita (key independent variable) and total population (independent variable). We chose these 2 variables since it provides more macro-level data on where the company is based. Then, we did a full join in R to create our final dataset. It is important to mention that we applied country-level data to firm level data. However, since the tech company dataset is our primary data, our resulting unit of analysis is a unique tech company. Furthermore, the time coverage of our data is from January 2020 to September 2021.

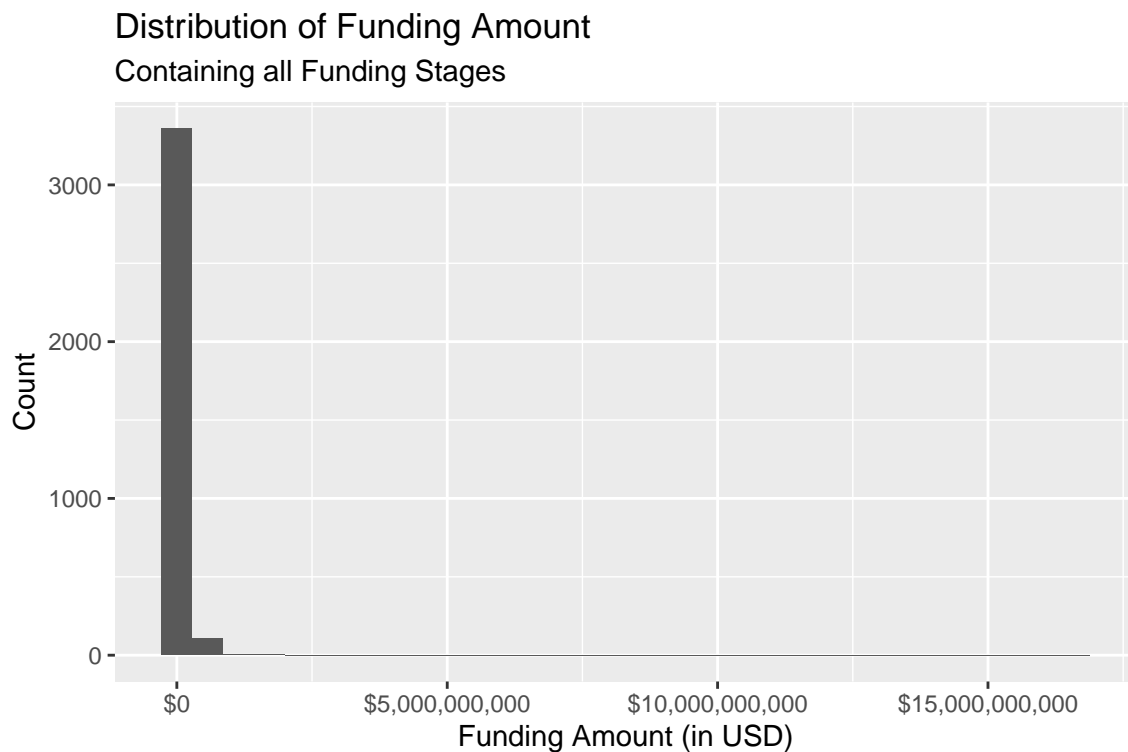
Additionally, one last data wrangle we performed was assigning each country a company is based in to a continent using a related [dataset](#) from Daina Bouquin from github. Looking at the continent the company

is based in rather than in a specific region allows for more uniformity when it comes to interpretation of results.

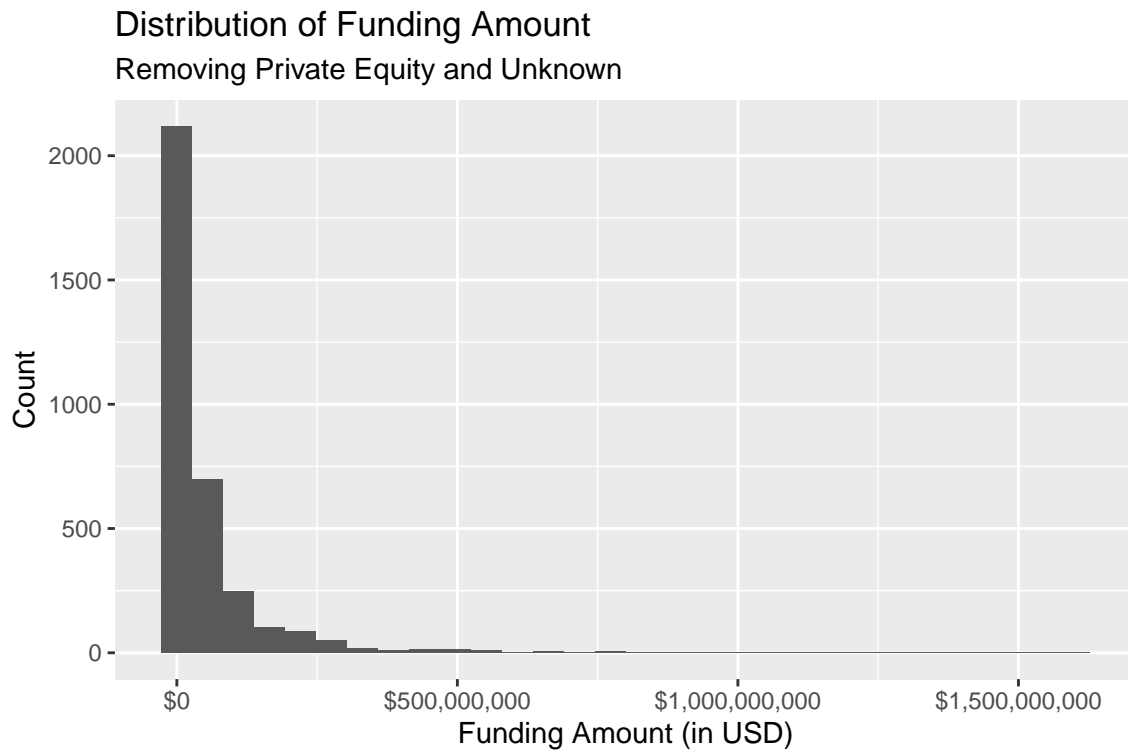
Variables

Backed by the above background and literature analysis, the variables that we are emphasizing on in our models are:

- 1) Funding amount which is our dependent variable with unit USD. It is a continuous variable. When plotting a histogram of this parameter, the distribution is found to be very right skewed.

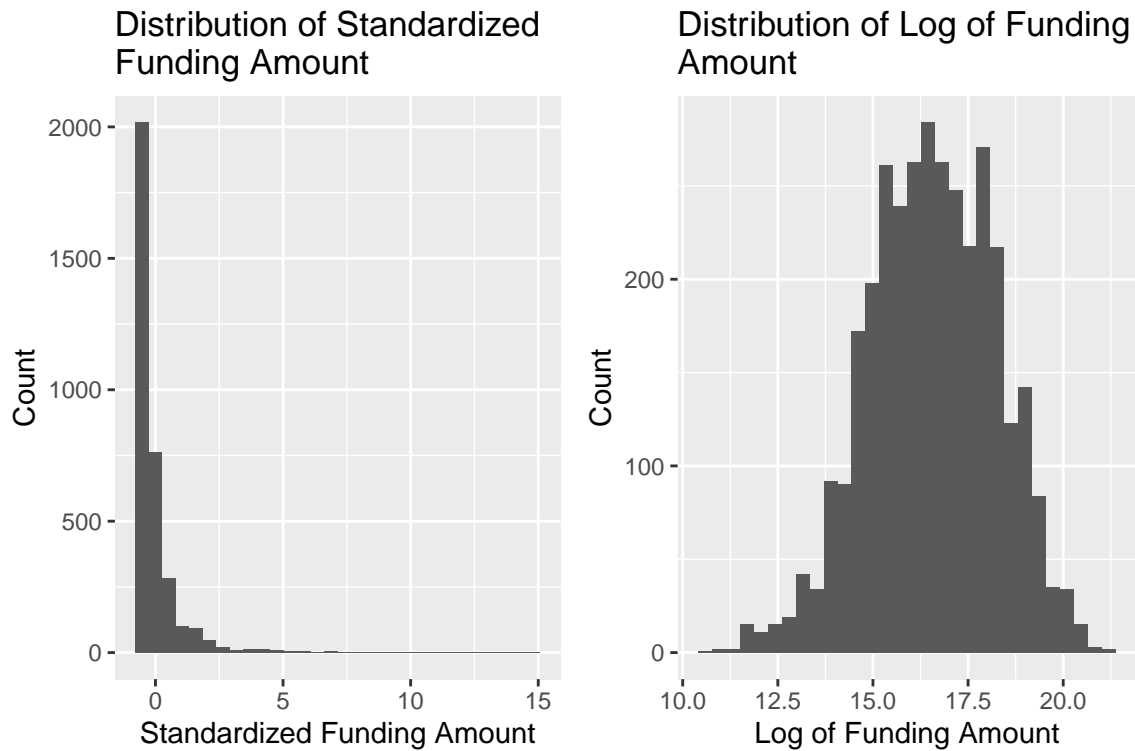


When we plotted the histogram for the funding amount, the highest funding amounts were private equity and unknown and they skewed our distribution significantly. Since private equity is not a funding stage and the unknowns do not bring insights, we can remove these two funding stages from our data without creating bias. This reduced the skewness of the distribution.



However, after we removed private equity and unknown data, the data is still skewed. So we will transform funding amount values in an attempt to have a uniform distribution. We used two approaches: standardization and log transformation.

First, we standardized the funding amount because it helps us interpret the intercept in a more sensible way and allows us to look at the funding amount in terms of standard deviation from the mean. Next, we conducted a log transformation of the funding amount to scale the magnitude of our data and transform the data to be more linear.



After viewing the transformation plots above, we found that there was a more uniform distribution for funding amount in the log transformation thereby having less of a skew and more linear. We did not perform the power transformation because we were satisfied with the uniform distribution for the funding amount produced by the log transformation.

- 2) The second variable that we will be putting emphasis on in this model is GDP per capita. This is an independent continuous variable in terms of USD. GDP per capita is a measure of economic growth relative to the population of a country. It is calculated by dividing GDP by population. The values range from USD 261.25 to USD 114,704.60 for the year 2019. To make this variable more interpretable, we are going to divide it by 1,000 and change the units to USD (in thousands)

When plotting a histogram of this variable, the distribution is continuous to a certain extent with significant modes.

To explore our options, we will still transform this parameter and compare the standardized transformation of GDP per capita with the log transformation.

For the standardized transformation curve, there is not much of a difference from using the original data while for the log transformation curve, the data becomes left-skewed. And so to meet the condition of constant variance the most, we will use GDP per capita, untransformed, in our model.

- 3) Last but not least, we will focus on is a binary indicator of whether the tech company is in the funding stage Series C or not. Plotting a histogram of this indicator, we find that the number of companies not in Series C is around six times more than those in Series C.

The other variables that we will still be including in our model but not put much emphasis on include:

- whether the company is a B2B company or not → binary indicator – is B2B company or is not B2B company
- which continent is the company located in → categorical variable with 6 values
- Total population → continuous variable having a range of 11,646 - 7,673,533,972 people

Modeling approach

Moving on to our modeling approach, in agreement with our above alternative hypothesis, we are going to look at two main multiple regression models.

To test our first hypothesis we conducted a parallel slopes model. In this model, we are measuring the constant rate of change in GDP per capita of the country the startup is based in, and the funding amount regardless of whether they are in Series C stage or not. This model gives us the benefit of having independent variables that do not interact with each other.

To test our second hypothesis we conducted an interaction model between GDP per capita of the country the startup is based in and the funding amount moderated by whether the company is in Series C or not. When we use an interaction model we are explicitly making the two key independent variables depend on each other.

Since we are conducting a multiple regression with 5 independent variables, we use a backward selection approach that is estimating models excluding some of the independent variables. This process will be based upon the results of the VIF (Variance Inflation Factor) test to identify multicollinearity which will be included in our appendix. When we assess for multicollinearity we are looking at whether our independent variables are correlated with one another. We will then choose the best model by looking at the maximum adjusted R^2 and minimum AIC and BIC values.

We first performed a VIF test on our parallel slopes model since that would give a more accurate picture of multicollinearity amongst our variables. From the VIF test results, we find that the independent variable looking at what continent the company is located in still has a GVIF above 5, indicating multicollinearity. Therefore, we decided to construct another model similar to the parallel slopes model one but excluding the continent variable.

After not including which continent the company is based in, the GVIF values in our subset parallel slopes model are all below 5, indicating none to little multicollinearity.

Results

Table 1: Predicting Funding Amount received by a Tech Company

	<i>Dependent variable:</i>		
	Log of Funding Amount (USD)		
	(1)	(2)	(3)
GDP per Capita (USD, in thousands)	0.005* (0.003)	0.012*** (0.001)	0.005* (0.003)
Company in Series C	1.871*** (0.248)	1.534*** (0.080)	1.533*** (0.079)
Company not a B2B software	0.290*** (0.070)	0.293*** (0.071)	0.290*** (0.070)
Total population (in ten thousands)	0.00000** (0.00000)	0.00001*** (0.00000)	0.00000** (0.00000)
Company in Asia	1.451*** (0.299)		1.464*** (0.298)
Company in Europe	1.386*** (0.296)		1.396*** (0.296)
Company in North Africa	1.774*** (0.324)		1.775*** (0.324)
Company in Oceania	1.453*** (0.365)		1.464*** (0.365)
Company in South America	2.521*** (0.355)		2.518*** (0.355)
Interaction term - Series C X GDP per capita	−0.006 (0.004)		
Constant	14.126*** (0.280)	15.322*** (0.114)	14.143*** (0.280)
Observations	3,395	3,395	3,395
Adjusted R ²	0.142	0.126	0.142
Akaike Inf. Crit.	12,633.870	12,688.900	12,633.940
Bayesian Inf. Crit.	12,707.430	12,725.680	12,701.370

Note:

*p<0.1; **p<0.05; ***p<0.01

To find the model that best fits our data, we look at the lowest BIC and AIC values as well as the highest

adjusted R^2 values. Therefore, we have narrowed it down to model 1 (interaction model) and model 3 (full parallel slopes model) since they have the highest adjusted R^2 values and the lowest AIC and BIC values compared to model 2 (subset parallel slopes model).

To choose between model 1 and model 3, since we have conflicting AIC and BIC values, we decided to conduct an anova test. As a result, we found that adding the interaction term to model 1 does not significantly reduce the variation in our dependent variable. And so holding everything else equal as well as to be parsimonious, we conclude that model 3 is best fitted for our data.

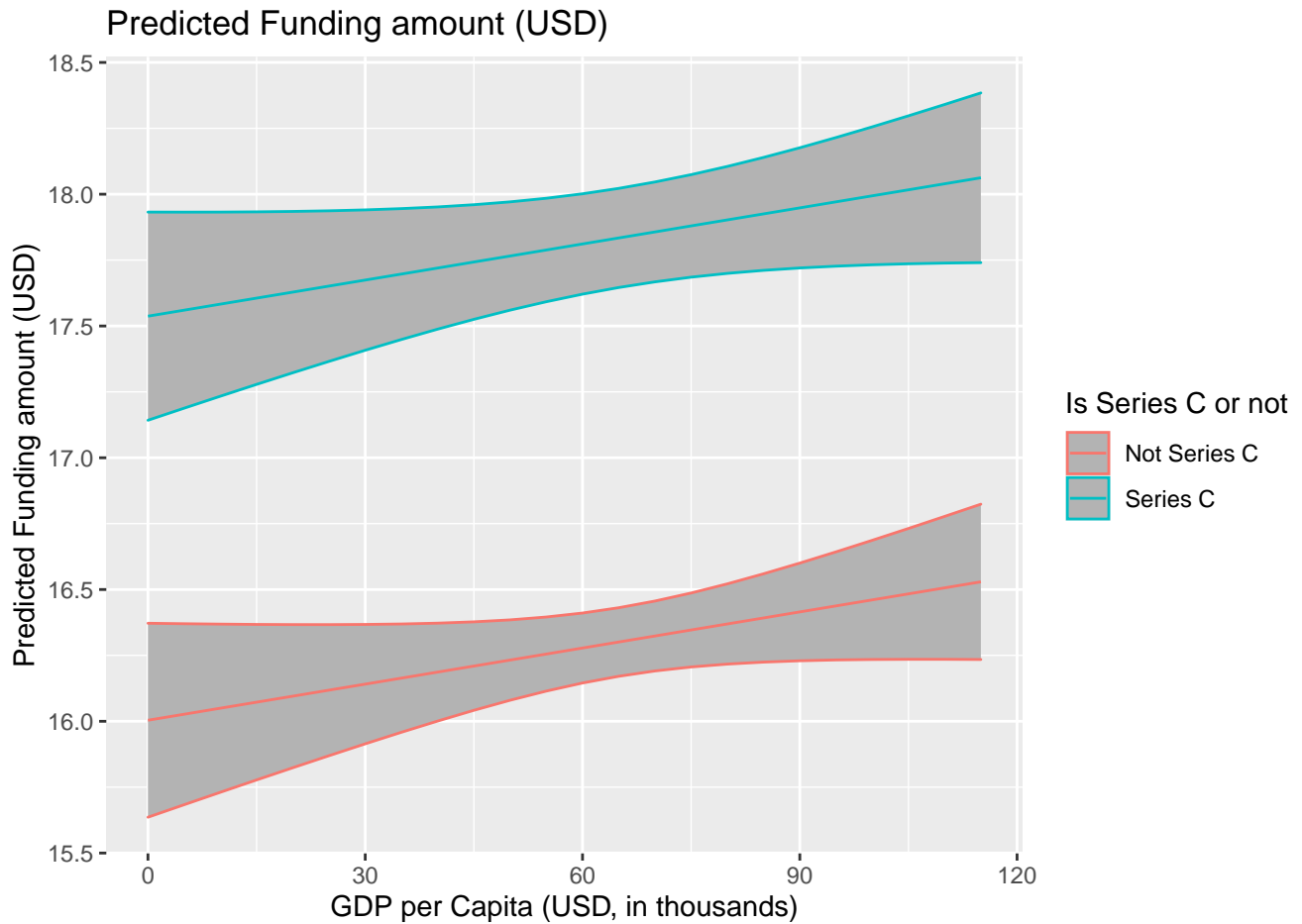
Therefore, we found evidence to support our first hypothesis but not enough evidence to support our second hypothesis.

From our chosen model, holding everything else constant, on average, a unit increase in GDP per capita in thousands of USD is associated with an increase of 0.501% in funding amount received by a tech company. Since there is one star associated with the coefficient, it is statistically significant.

Moreover, holding everything else constant, on average when compared to a company in Series C, a company in the Series C funding is associated with 363.205% greater in funding amount received by a tech company. Since there are three stars associated with the coefficient, it is statistically significant.

Regardless, looking at practical significance, the adjusted R^2 of all of the 3 models are very small, ranging from 0.126 - 0.142. This overall tells us that the models generated don't explain much of the unexplained variation in the funding amount tech company receives.

To summarize our results, we can hold other variables other than our key variables constant and model predictions for the funding amount of a tech company based on our data.



Looking at how GDP per capita can predict funding amount of a company based on whether it is Series C company or not, we can look at the graph above showcasing the predicted funding amount with GDP per Capita in thousands USD. Holding everything else constant, the general trend observed is that as GDP per capita increases, the tech company is predicted to receive more funding, more so if the company is in the series C funding stage than not. This is in agreement with our first hypothesis that states that there is a significant association between the GDP per capita of the country the company is based in and the funding amount for a company regardless of whether the company is Series C or not.

Conclusion

The question we are addressing in this paper is how much funding does a tech company receive based on the GDP per capita of the country it is based in and whether it is in the Series C funding stage or not. We conducted a multiple regression analysis looking at both a parallel slope and interaction models. We found that the parallel slopes model looking at GDP per capita and whether the company is in series C or not the best fit while accounting for population of the country the tech company is in, the region the company is in and whether the company is a B2B software or not has the best goodness of fit for our data. Therefore, there is more evidence to support the alternate hypothesis that there is a significant association between the GDP per capita of the country the company is based in and the funding amount for a company regardless of whether the company is Series C or not. Holding everything else constant, there is a linear positive relationship between GDP per capita and funding amount, both

in USD. Moreover, on average, companies in the Series C funding stage receive more funding than those which are not.

However, the models generated don't provide much practical significance since they don't explain much of the unexplained variation in the funding amount received by tech companies. This research has important implications in terms of identifying what are the possible variables that might help determine the funding amount of a tech company. We have already emphasized on GDP per capita of the country the tech company is based in as well as whether the company is in Series C or not. Further research can put emphasis on which vertical (category) the tech company is in or even the region the company is based in.

The conclusion that Series C companies are the preferable funding stage for investors is true for the culture of startups in the United States, but not applicable for all startups worldwide. If we compared the environment for startups between the United States and European countries, we would find that the funding stage that investors prefer will differ ("The 3 Key Differences Between European Vs US Startups | Startup Grind" n.d.). In European countries, investors are strong stakeholders in seed and early stage funding, but not as much in the growth stage funding (Series A-C). For every US dollar available for a European startup, there is USD6 for an American startup and this speaks to the difference between Series A funding and Series C (Law 2017). Therefore, European countries are more likely to invest in earlier stage companies due to financial restrictions and capacities. Thus, this potentially requires further investigation in order to get a more significant understanding of how funding amount is allocated to tech companies. Results from such research questions are crucial to help start-ups have an idea of what to expect given the variables being analyzed.

Appendix

Summary statistics for key variables

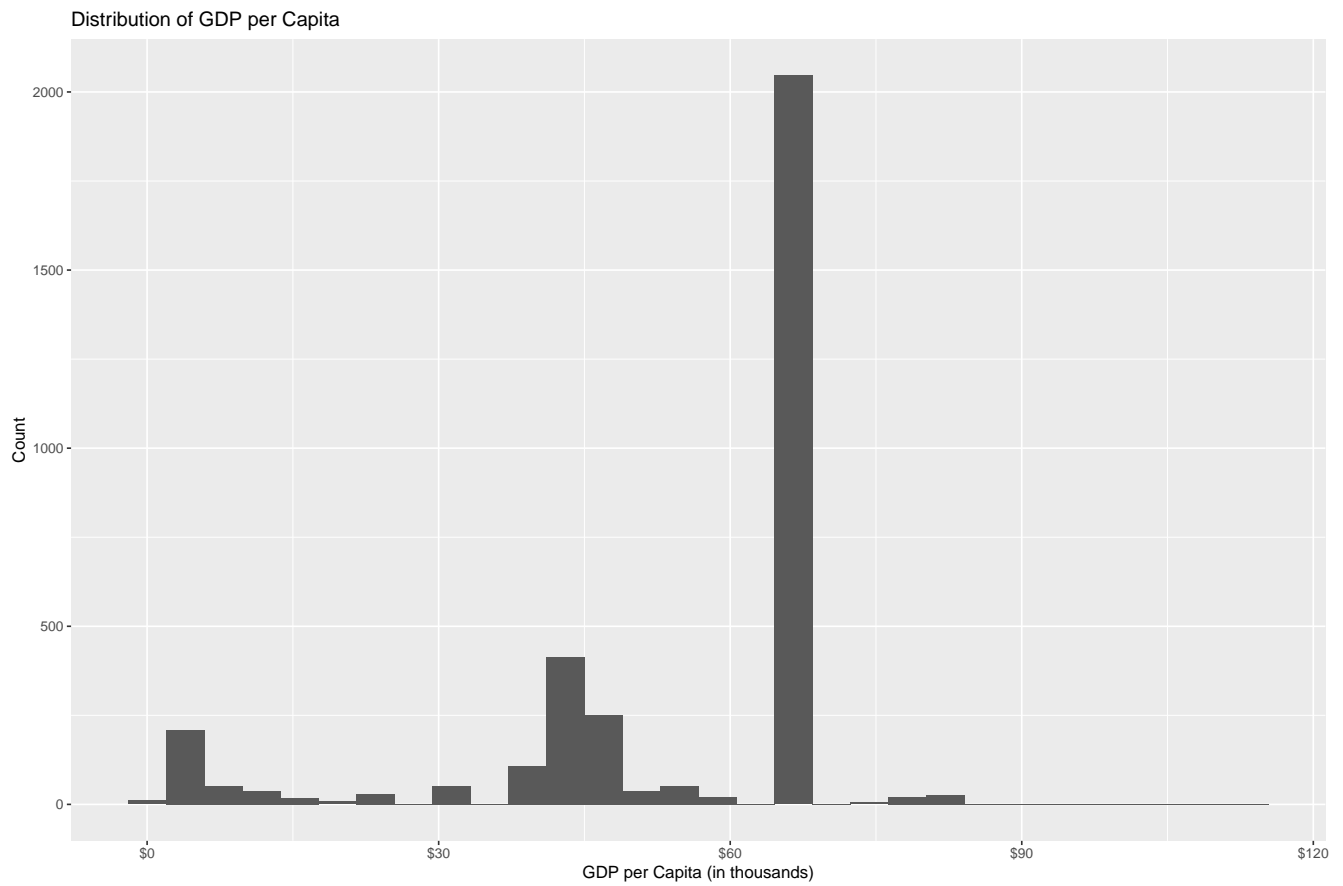
##	Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75
## 1	funding_amount	3395	51876080.662	104309284.282	40000	5e+06	5e+07
## 2	gdp_per_capita	3401	53.313	19.122	1.285	43.641	65.118
## 3	is_series_c	3401					
## 4	... Not Series C	2954	86.9%				
## 5	... Series C	447	13.1%				
##	Max						
## 1	1.6e+09						
## 2	114.705						
## 3							
## 4							
## 5							

The table above gives us an overview of the range, the median, the interquartile range amongst other statistics of our key variables in our models: funding amount, GDP per capita and whether the company is in the Series C funding stage or not.

Distribution for key Variables

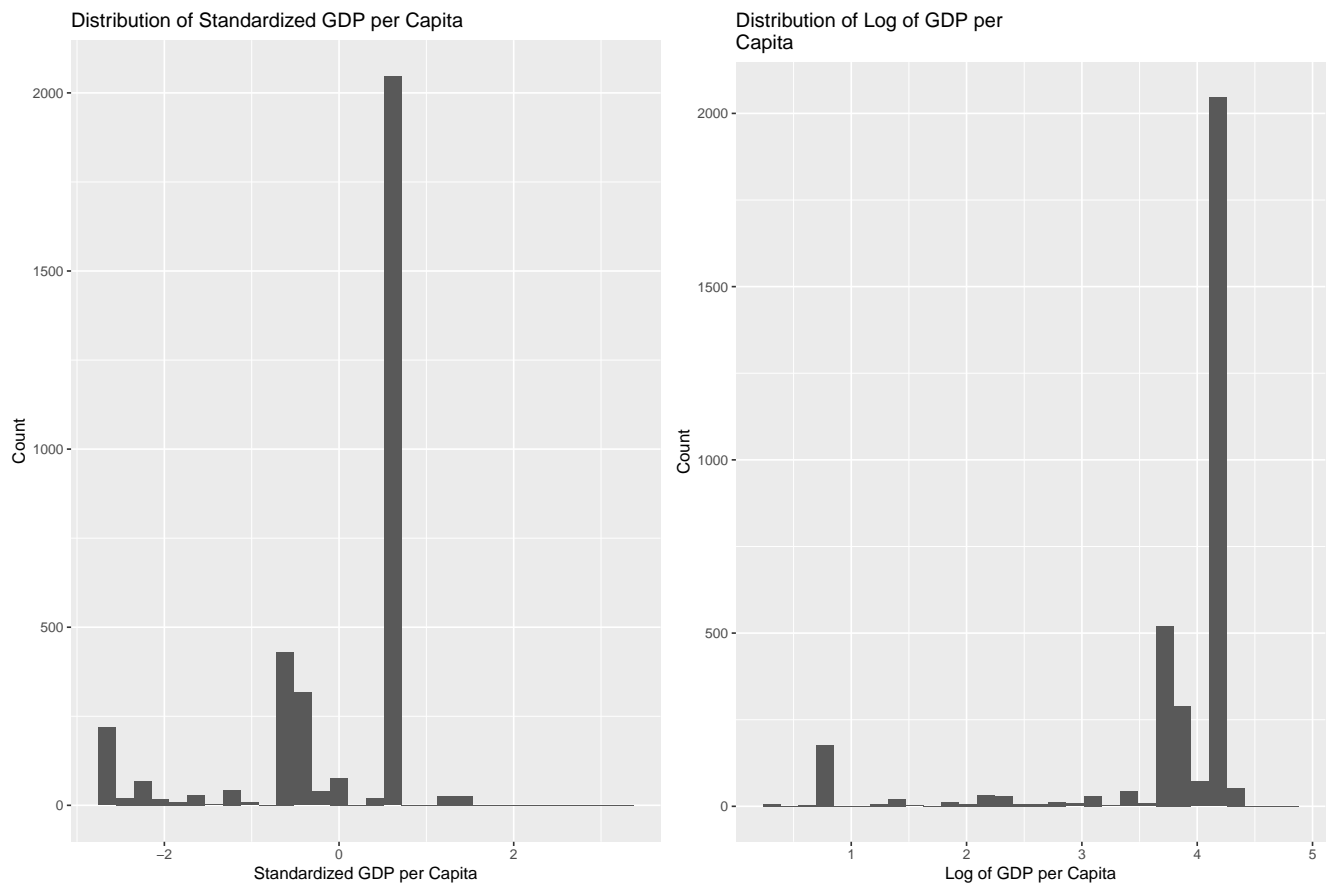
GDP per Capita Distribution

No transformation



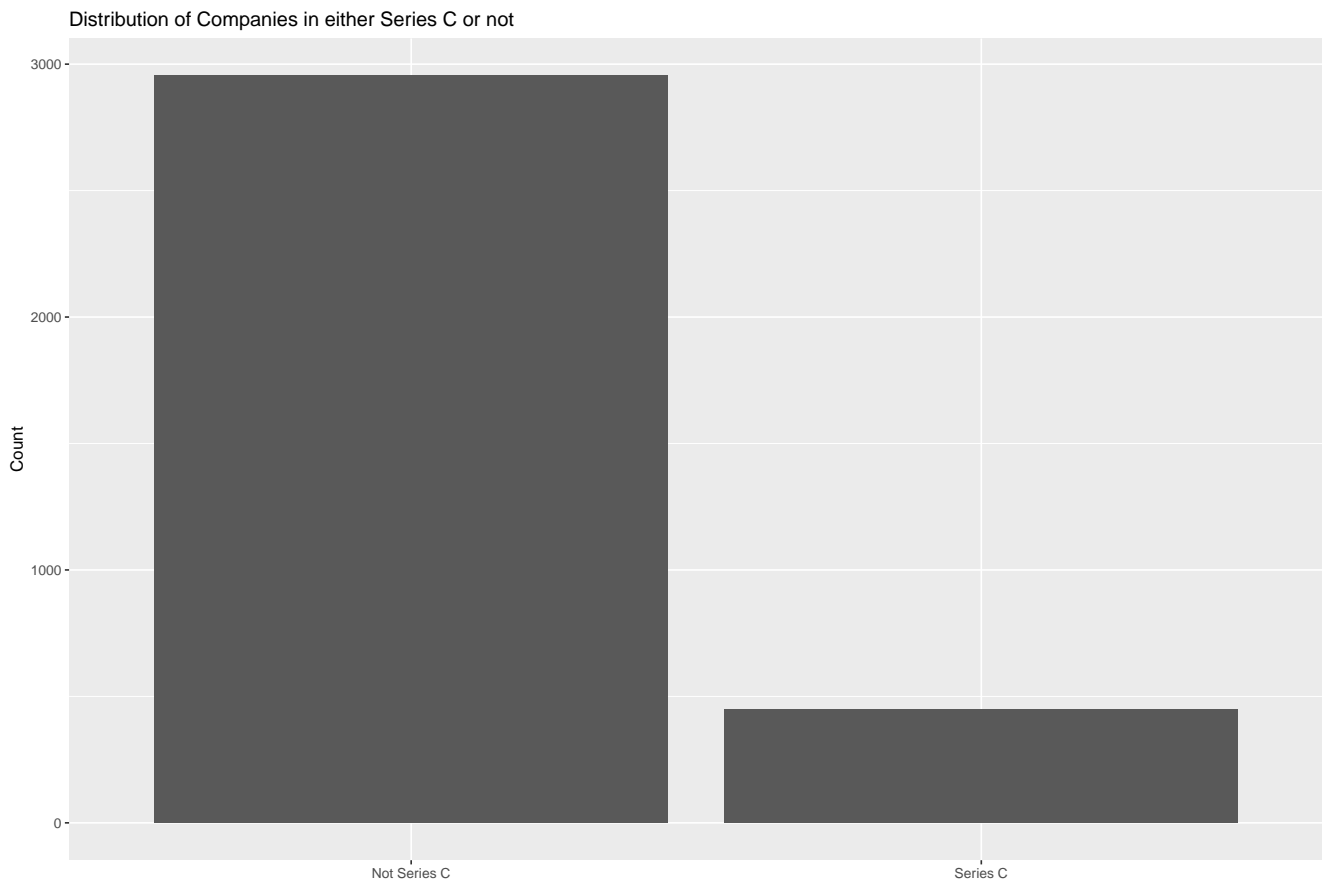
The above histogram looks at the distribution for GDP per capita before any transformation.

Standardization vs Log transformation



Comparing the two histograms above (left one – standardization; right one - log transformation), we note not much of a difference between the distributions before and after transformations.

Series C or not Distribution



The above histogram looks at the distribution of the binary indicator, whether the tech company is in the Series C funding stage or not.

Assessing for explanatory

VIF test – Parallel slopes model

```
##                GVIF Df  GVIF^(1/(2*Df))
## gdp_per_capita 3.757940  1      1.938541
## is_b2b         1.013201  1      1.006579
## population     1.942398  1      1.393699
## is_series_c    1.009027  1      1.004503
## continent_f    5.204059  5      1.179327
```

The VIF test above is for our full parallel slopes model. We see that the independent variable, continent the tech company is based in, is the only one with a VIF value of greater than 5, indicating multicollinearity.

VIF test – Parallel slopes model (subset)

```
## gdp_per_capita    is_b2b    population    is_series_c
##      1.096710      1.011157      1.092266      1.008496
```

The VIF test above is for our subset parallel slopes model (excluding the independent variable, the continent the tech company is based in). All of the variables have VIF values less than 5, indicating little to no multicollinearity.

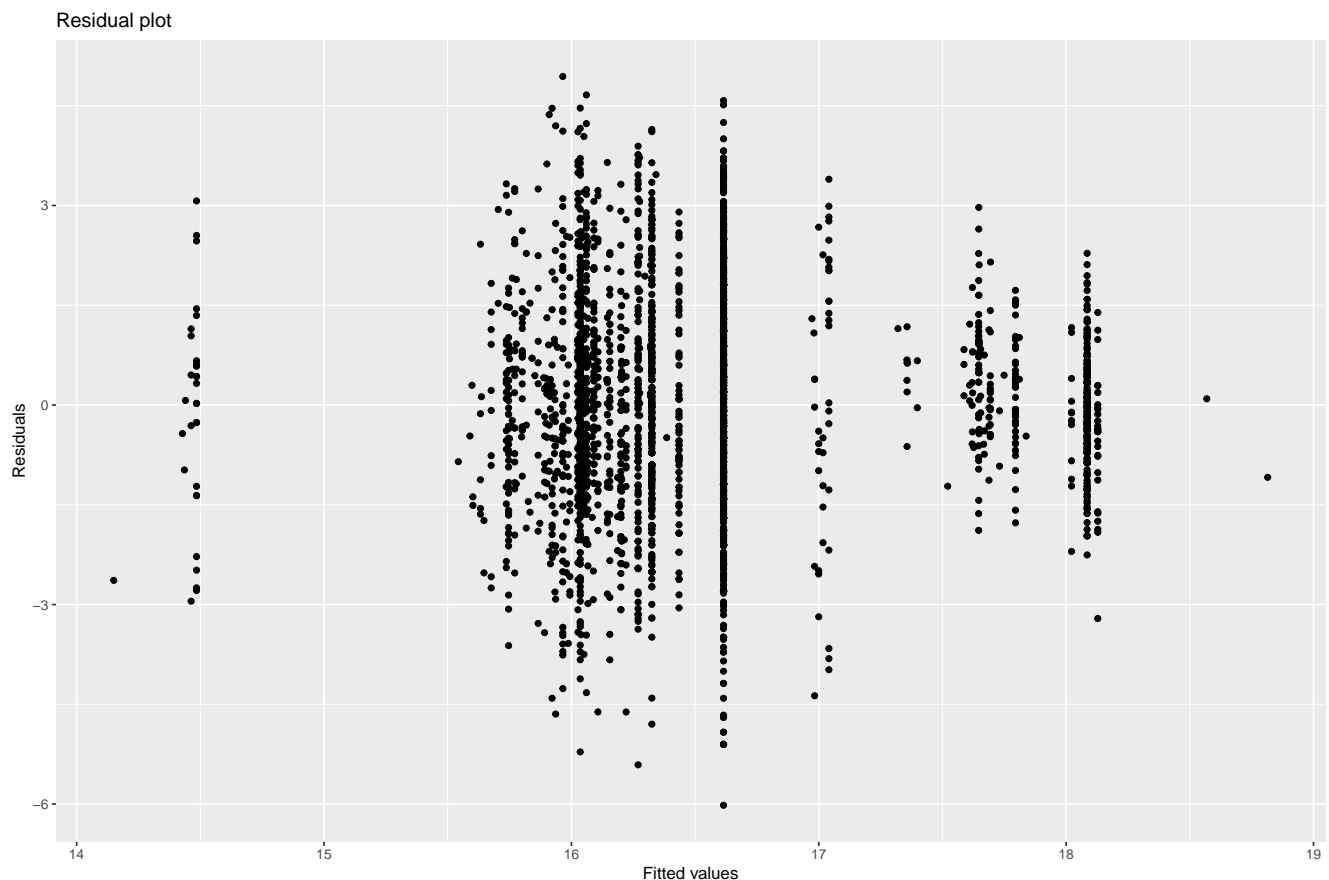
Anova test

```
## Analysis of Variance Table
##
## Model 1: funding_amt_log ~ gdp_per_capita + gdp_per_capita * is_series_c +
##   is_b2b + population + is_series_c + continent_f
## Model 2: funding_amt_log ~ gdp_per_capita + is_b2b + population + is_series_c +
##   continent_f
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     3384 8155.4
## 2     3385 8160.3 -1      -4.979 2.066 0.1507
```

From the anova test, we note that the p-value is greater than our default significance level which is 0.05. Therefore it is not statistically significant. We fail to reject our null hypothesis stating that adding an additional independent variable will significantly reduce the unexplained variation in our dependent variable.

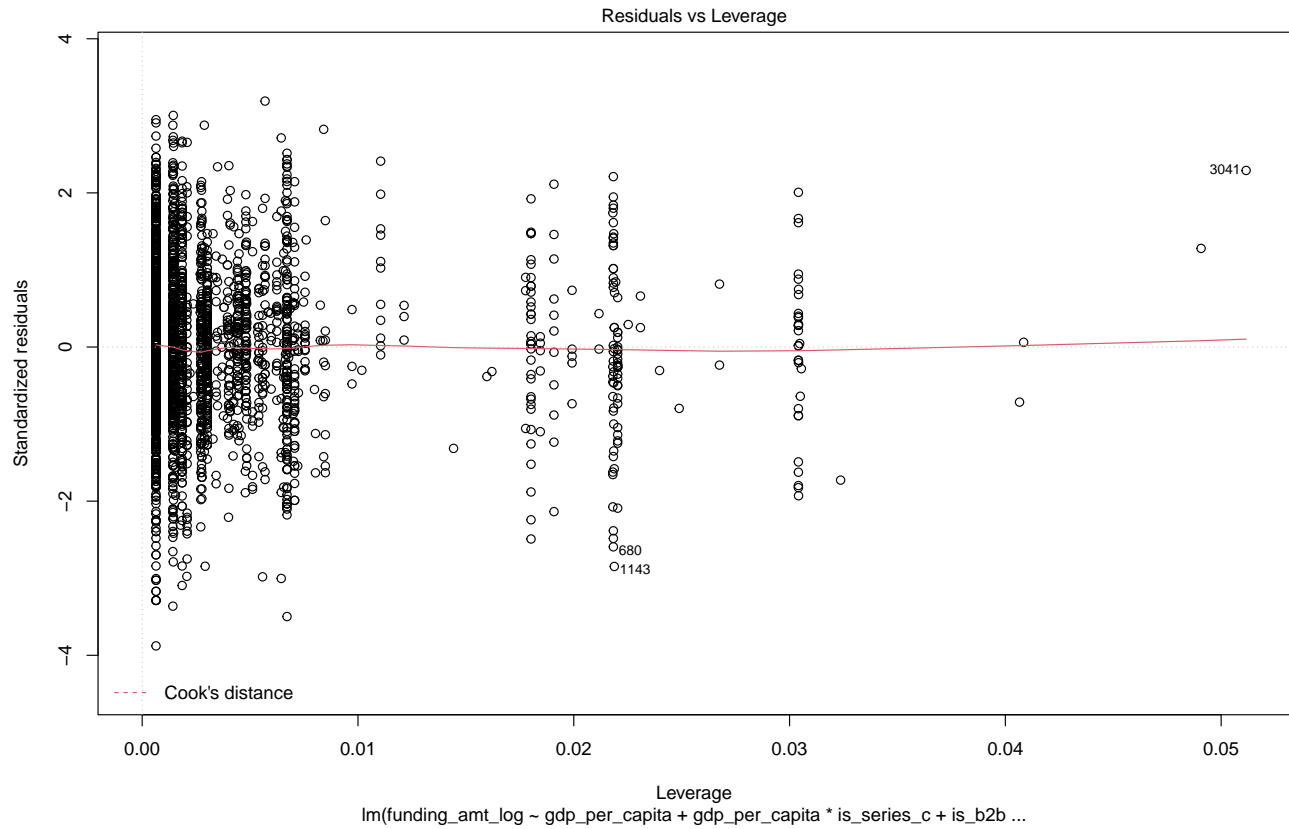
Conditions for multiple regression

Linearity



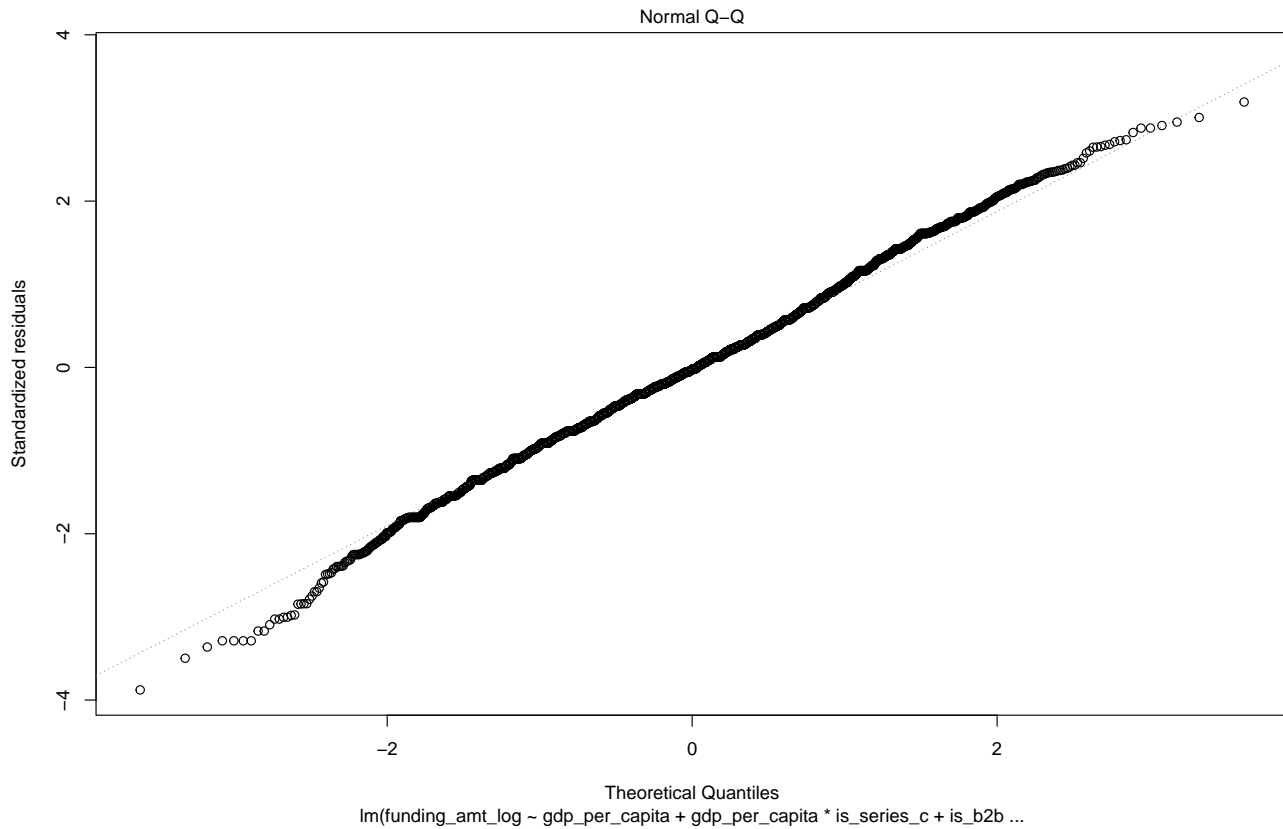
From the residual plot given, the data points are scattered symmetrically above the 0 horizontal line. Therefore we can assume that the condition for linearity is satisfied.

Homoskedasticity



From the plot above, the data points are concentrated in a larger cloud on the left side. Therefore, we can assume that the condition for homoskedasticity is not satisfied.

Normal distribution of residuals



From the Q-Q plot above, the data points trail off at the beginning and at the end of the graph. Therefore, we can assume that the condition for normal distribution of residuals is not satisfied.

Randomness and Independence

We cannot specifically plot graphs to test for randomness and independence. But we do believe that our data sets were collected with these conditions in mind.

Bibliography

- “Financial Development.” n.d. Text/{HTML}. *World Bank*. Accessed April 2, 2022. <https://www.worldbank.org/en/publication/gfdr/gfdr-2016/background/financial-development>.
- Haddad, Christian, and Lars Hornuf. 2019. “The Emergence of the Global Fintech Market: Economic and Technological Determinants.” *Small Business Economics* 53 (1): 81–105. <https://doi.org/10.1007/s11187-018-9991-x>.
- “How the Stock Market Affects GDP.” n.d. *Investopedia*. Accessed April 2, 2022. <https://www.investopedia.com/ask/answers/033015/how-does-stock-market-affect-gross-domestic-product-gdp.asp>.
- Law, Ryan. 2017. “The Valley Vs the World: How Startup Funding Varies by Country.” *The SaaS Growth Blog*. <https://medium.com/the-saas-growth-blog/the-valley-vs-the-world-how-startup-funding-varies-by-country-66677b88f7a3>.
- “LEF 2021 – Liberec Economic Forum.” n.d. Accessed April 2, 2022. <http://lef.tul.cz/>.
- Prochniak, Mariusz, and Katarzyna Wasiak. 2017. “The Impact of the Financial System on Economic Growth in the Context of the Global Crisis: Empirical Evidence for the EU and OECD Countries.” *Empirica* 44 (2): 295–337. <https://doi.org/10.1007/s10663-016-9323-9>.
- “The 3 Key Differences Between European Vs US Startups | Startup Grind.” n.d. Accessed April 2, 2022. <https://www.startupgrind.com/blog/the-3-key-differences-between-european-vs-us-startups/>.