The Impact of Research and Development Expenditure on Economic Growth

Group 1: Andrew Mandel and John Darcy

November 21, 2019

ECON 3161 Dr. Shatakshee Dhongde

Abstract

Expenditure on research and development often signifies that a country is committed to making advancements in the fields of science and technology. This study attempts to reveal the relationship between economic growth (as measured by GDP per capita) and gross domestic expenditure on research and development as a percentage of gross domestic product (GERD). Other explanatory variables including the unemployment rate, GINI index, education expenditure as a percentage of GDP, labor productivity (as measured by GDP per hour worked), gross savings, and foreign direct investment inflow are also analyzed and utilized to determine this relationship. A time lag of 3 years between each country's GERD (measured in 2012) and corresponding GDP per capita (measured in 2015) is used to allow time for the expenditure to be transformed into new developments that can impact GDP per capita. A positive correlation between GERD and GDP per capita is hypothesized and supported by the linear regression models developed in this study.

I. Introduction

Research and development (R&D) embodies the actions necessary to introduce or improve new products, processes, and services. Expenditure on research and development is a good indicator of how dedicated a country is to creating technological growth, as countries with greater faith in their research and development efforts will be willing to invest in it more. Typically, within a few years of investment, the effects of research and development will begin to be realized as they make an impact on the economy.

Technological progress has long been noted as an integral factor in the development and advancement of humankind. Even from early times, it can be seen that the civilizations which prospered the most and had the highest qualities of life were ones that embraced technological growth and actively took steps to better understand it and harness its many beneficial impacts. The Stone Age was revolutionized by the invention of tools that could be used to gather food and build shelter; two things absolutely essential for daily life. During the Renaissance, the invention of the printing press served as a huge leap for increasing literacy and spreading new ideas around the world by allowing for the mass production of books. This innovation serves as a prime example of how technology can impact many different facets of life and the economy.

Today, technological growth via research and development still has an important impact on the wellbeing of humankind and economic growth. In addition to the direct products of research and development such as new inventions or manufacturing methods, several beneficial byproducts are also generated. These byproducts can include increased labor productivity, increased quality of living, and new job creation, all of which are expected to have a positive impact on a country's economic growth. Therefore, being that research and development is so valuable to society, it is equally important to understand how its impacts can be modeled from an economic standpoint.

This paper will draw the relationship between economic growth and expenditure on research and development by using cross-sectional data to create both simple and multiple linear regression models. Our hypothesis is that there will be a positive correlation between expenditure on research and development and economic growth. The economic rationale used to fuel this hypothesis is that more expenditure on research and development is likely to lead to more efficient products and manufacturing processes, thus resulting in greater productivity and higher outputs, as well as emerging technological fields that will create new jobs, motivate individuals to obtain higher levels of education, and provide new areas for investment. All of these outcomes will in turn bolster the economy even more.

II. Literature Review

In order to study the effect of Research and Development on Economic Growth, Yazgan and Yalçinkaya (2018) researched the effect of Real Fixed Capital Investments per capita on GDP per capita for OECD (Organisation for Economic Co-operation and Development) countries. These countries were split into two groups, one containing countries whose GDP per capita is below $30,000 USD and another for countries with a GDP per capita above that threshold. Multiple regression models were created for both groups to test the effect of a variety of variables. Each of these models estimated the effect of Real Fixed Capital Investments per capita and Employed Manpower on GDP per capita. These two variables were consistent through each regression while a few others were tested alongside them to see if they had any effect on the data. The third variables tested included: Real R&D Investments of The Public Sector, Real R&D Investments of The Private Sector, Real R&D Investments of The Universities, Total Real R&D Investments and Total R&D Staff. For each regression, optimal lag data was determined to allow for the investment in R&D to be utilized and for any discoveries made in the research to influence the economy. While optimal lag data was determined for each regression, a 3-year lag was by far the most common. For the most relevant data, and that of which Yazgan and Yalçinkaya were most interested in, where Capital Investments are compared to change in GDP per capita, all regressions had positive coefficients and supported the correlation at a 1% significance level. While all variables in the regression had positive coefficients, not all coefficients were determined to be statistically significant at the 1% significance level. Those that were significant in all regressions, which include those variables most relevant to our research, include R&D investments in the private sector, universities, and total R&D investments.

Nekrep, Strašek, and Boršič (2018) completed similar research by determining how R&D expenditure as a percent of GDP affects economic growth. However, in this instance, they measured economic growth with labor productivity per hour of work. They believe that R&D leads to technological advancements which will make workers more productive. This research was completed using data from all European Union member states except for Croatia, which there was no available data for. Before a regression was determined, correlation coefficients and p-values were calculated for each country to determine the general strength and direction of the relationship between R&D expenditure and a country's labor productivity. Of the 27 countries tested, 9 had a positive and strong correlation between R&D expenditure and economic growth, 7 had a positive and moderate correlation, 1 had a positive and weak correlation, 1 had a negative correlation and 9 had statistically insignificant data at a 5% significance level. A regression was also made taking the form $y_i = \beta_0 + \beta_1 x_1 + \beta_2 (x_1)^2 + u_i$. The regression determined that there was a positive correlation between the two variables, where a 1% increase in R&D expenditure should lead to a

22.411 EUR per hour increase in labor productivity. With an R-squared value of 0.614, there is a moderately strong correlation between R&D expenditure and labor productivity, where 61.4% of the data can be explained by the regression model. It is interesting to note that their β_3 was determined to be negative, thus there is a point where increasing R&D expenditure begins to have a negative effect on the economy. This value was determined to be 2.85%, thus 2.85% is what they believe to be the optimal percentage of GDP to be spent on R&D.

Gocer, Alatas, Peker (2016) looked to determine how R&D expenditure and innovation affects income. In their research, they began by selecting 11 European Union countries. Data for R&D expenditure and patents filed (representing innovation) was collected for each of the 11 countries and a regression was created to determine how these affected a country's Gross National Product (GNP). A logarithmic model was created to determine the elasticities for each of the explanatory variables. In this case, it was again decided to determine an optimal lag time for the R&D and innovation to have an effect on the country's income. The coefficients for R&D and innovation were determined to be 0.19 and 4.05 respectively. This leads to the conclusion that a 1% increase in R&D expenditure should lead to an increase in that country's GNP of approximately 0.19%. Furthermore, a 1% increase in innovation, or patents filed, should lead to an increase in GNP of approximately 4.05%. The significance level of the R&D expenditure coefficient and that of innovation were determined to be statistically significant at the 10% and 1% significance levels respectively. This again shows the positive relationship between R&D and economic growth. It also shows the significance of using R&D expenditure efficiently. The correlation between innovation and income has a much larger magnitude than that of R&D expenditure. If a 1% increase in R&D expenditure was able to be turned into a 1% increase in innovation (which is likely impossible), then the increase in income each year would likely be much larger.

In research investigating the relationship between Gross Domestic Expenditure on Research and Development (GERD) and GDP, Szarowská (2018) compares R&D spending and a multitude of other variables to economic growth. The explanatory variables tested are investment ratio to GDP, share of the population either holding a tertiary degree or employed in science or technology, and the openness of borders in terms of the sum of imports and exports divided by the GDP. The R&D expenditure is also broken up into three subcategories in a separate regression in order to determine which is the most effective. These subcategories are business, government, and higher education research spending. This data was collected for 8 different Central and Eastern European countries and evaluated in multiple regression models. Again, a time lag was introduced to allow for R&D effects to reach the economy, being one year in this case. All of the above variables were determined to have a positive effect on the country's GDP. It was determined that a 1% increase in GERD led to a .032% increase in GDP per capita. It was also determined that government research spending was the most effective subcategory, with a 1% increase leading to a 0.035% increase in GDP per capita. Business research spending was determined to be the next most effective, followed lastly by higher education research spending. This data again supports our hypothesis that an increase in R&D spending should lead to an increase in economic growth.

There is a lot of research investigating the correlation between R&D and economic growth, often in terms of GERD vs GDP per capita, however this research will be slightly different. This research will be analyzing some explanatory variables that have not been touched upon by previous studies which examine R&D. The secondary explanatory variables of GINI index, unemployment rate, education expenditure, productivity of labor, gross savings, and inflow of foreign direct investment (FDI) are much less visited. Examining these additional variables will provide a more comprehensive view than previous research and also allow for the controlling of more variables. Much of the previous research has also focused on European countries or a certain subset of countries. This is often valuable when researching a specific topic or trying to answer a specific question, however, this research attempts to look more generally. In order to do so, data was acquired from as many different countries as possible to get a more comprehensive view of R&D's affect throughout the whole world. For these reasons, this study should help provide a more thorough view of how R&D affects economic growth.

III. Data

To characterize the relationship between expenditure on R&D and economic growth, cross-sectional data was gathered. The dependent variable used was the natural logarithm of GDP per capita, as this statistic is most commonly used to reflect the economic growth of a country. The data for GDP per capita were sourced from World Bank. The main explanatory variable used is gross domestic expenditure on R&D as a percentage of gross domestic product (GERD), with data taken from UNESCO. This variable was picked instead of a different variable such as total expenditure on R&D in order to keep the analysis fair between different countries, as countries with larger economies would be able to spend a greater amount of money on R&D compared to countries with smaller economies and lower budgets. An initial scatterplot of the natural logarithm of GDP per capita versus GERD shows a positive and mild correlation between the two variables and can be seen below in Figure 1.

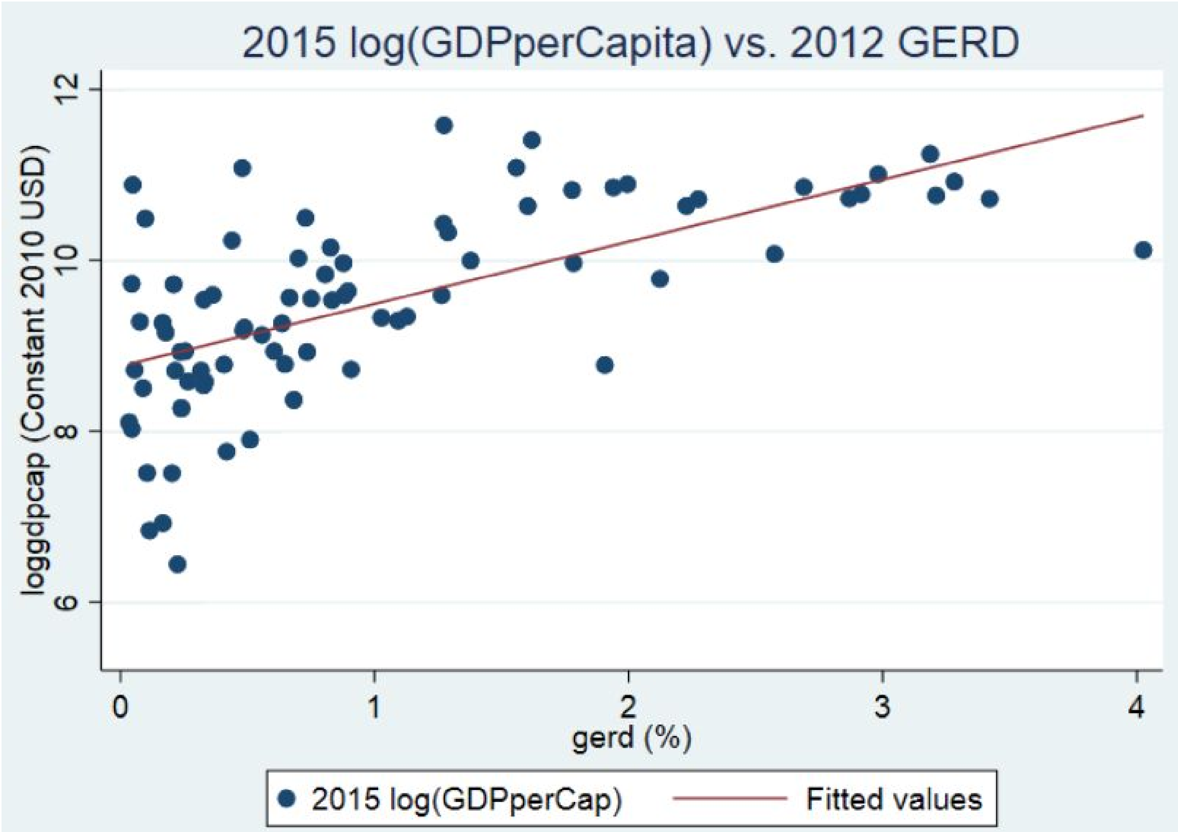

Figure 1 - Scatterplot of Natural Logarithm of GDP per Capita vs. GERD

In addition to these main variables, several other explanatory variables were used to strengthen the multiple linear regression models in order to uncover the ceteris paribus effect of GERD on GDP per capita. These variables include the GINI index, unemployment rate, education expenditure, labor productivity, gross savings, and inflow of foreign direct investment. Data for GINI index, unemployment rate, education expenditure, gross savings, and inflow of foreign direct investment were taken from the World Bank. Data for labor productivity was gathered from the Organisation for Economic Cooperation and Development (OECD). The GINI index measures the distribution of income across a population, with a value of 0% showing perfect income equality and a value of 100% showing perfect income inequality. Therefore, a high GINI index conveys that people who make a higher income also receive a larger percentage of the total income of the population. In the regression models, this variable is predicted to have a coefficient with a negative sign, as greater economic growth is predicted to result in more income equality (a lower GINI index), as more individuals will have access to higher paying jobs that are created by R&D.

The unemployment rate reflects the percentage of the total labor force that is not employed but available for and seeking employment. Similar to the GINI index, this variable is predicted to have a coefficient with a negative sign, as greater economic growth is likely to create more jobs and reduce the unemployment rate. Total government expenditure on education as a percentage of GDP is used to examine the quality and availability of education, as it is hypothesized that countries experiencing higher economic growth will also spend more on education, giving this variable a coefficient with a positive sign. This is because more economic growth due to R&D will not just generate new jobs, but specifically new jobs that require a higher degree of education in order to perform higher-level research. Also, with more jobs created that are related to R&D, the country would be more likely to invest in education in order to motivate individuals to attain higher levels of education and fill these jobs. Labor productivity is measured by GDP per hour worked, and as labor becomes increasingly productive, less hours will need to be worked to obtain the same output. Countries with higher economic growth can expect their labor to be more productive, so the labor productivity variable is predicted to have a coefficient with a positive sign in the multiple linear regression model. Gross savings as a percentage of GDP are calculated as gross national income less total consumption, plus net transfers. This variable is expected to have a coefficient with a positive sign, as increased savings help to finance larger and more long-term investments, which are necessary for expenditure on R&D. A country with low savings levels would be less likely to commit to long-term investments such as R&D, as R&D typically takes several years before it can produce beneficial economic byproducts. The final explanatory variable used is inflow of foreign direct investments as a percentage of GDP, as a country experiencing successful R&D work will attract more foreign investments into the country, causing this variable to have a coefficient with a positive sign. A summary of each variable can be found below in Table 1.

Table 1 - Variable Descriptions

Variable Name	Description	Year	Units	Source
loggdpcap	Natural logarithm of GDP per capita.	2015	Constant 2010 USD	World Bank
gerd	Gross domestic expenditure on research and development as a percentage of GDP.	2012	Percentage	UNESCO
gini	Measure of income equality across a population.	2015	Percentage	World Bank

unemploy	Percentage of the total labor force that is not employed.	2015	Percentage	World Bank
educ	Total government expenditure on education as a percentage of GDP.	2015	Percentage	World Bank
logproduc	Natural logarithm of GDP per hour worked.	2015	Constant 2010 USD	OECD
savings	Gross savings, as a percentage of GDP.	2015	Percentage	World Bank
fdi	Inflow of foreign direct investment (FDI), as a percentage of GDP.	2015	Percentage	World Bank

Descriptive statistics for each variable can be found below in Table 2.

Table 2 - Variable Descriptive Statistics

Variable	Observations	Mean	Std. Dev.	Min.	Max.
loggdpcap	183	8.63	1.46	5.43	11.58
gerd	82	1.01	0.99	0.03	4.03
gini	75	36.88	8.10	25.40	59.10
unemploy	179	7.72	5.78	0.16	27.65
educ	104	4.70	1.60	1.36	12.46
logproduc	42	4.67	0.054	4.54	4.86
savings	155	21.05	11.0	-30.96	57.12
fdi	178	5.15	10.07	-7.82	80.79

While economic growth data was gathered on 183 countries (country names shown in Appendix A), the sample sizes used in each regression model are notably smaller due to countries with unreported or unknown values for many of the other variables used in the model.

Before constructing each regression model, all of the Gauss-Markov assumptions were checked. The five Gauss-Markov assumptions can be seen below:

1. **Model is linear in parameters such that:** $y = B_0 + B_1X_1 + B_2X_2 + ... + B_kX_k + u$

   Where B_0, B_1, B_2, ..., B_k are the unknown parameters of interest and u is the error term.

   All models in section IV of this paper satisfy this assumption, as they are linear in parameters.

2. **Data obtained from random sampling**

   Data was sourced from every country around the world where available, so there was no consideration in the countries used in the sample. This proves that the sampling was random.

3. **No perfect collinearity between explanatory variables**

   STATA software was used to check for perfect collinearity between each explanatory variable. This analysis showed that Assumption 3 was not violated, as there were no exact linear relationships between the explanatory variables and none of the explanatory variables were constant. The result of this analysis can be found in Appendix B.

4. **Zero Conditional Mean**

   The expected value of the error term, u, is zero given any value of the explanatory variables. This assumption is difficult to assume, as there are likely to be other unobserved factors that can influence economic growth. Therefore, all results will be interpreted with caution.

5. **Homoskedasticity**

   The variance of the error term, u, has a constant variance given any value of the explanatory variables. This assumption is also difficult to assume, as the values of the explanatory variables may contain information about the variability of the unobserved factors. Therefore, the results will be interpreted with caution.

IV. Results

With all Gauss-Markov assumptions addressed, several different regression models are formulated to test the hypothesis. All STATA regression outputs can be found in Appendix C.

<u>Model 1:</u>

First, a simple linear regression model is constructed to test the relationship between the natural logarithm of GDP per capita and GERD. This model is written as:

Model 1: $log(gdpcap) = B_0 + B_1(gerd) + u$

This model has a sample size of 80 countries. From the STATA output, the estimated equation can be written as:

$$\textbf{Estimated Equation 1: } log(gdpcap) = 8.76 + 0.728(gerd)$$

The model has an R-squared value of 0.41, denoting a weak/mild correlation between GERD and GDP per capita. The coefficient on *gerd*, B_1, has a positive sign as predicted, showing that GERD and GDP per capita have a positive linear relationship. Furthermore, since a log-level model is used, this coefficient can be interpreted to show that a 1% increase in GERD results in a 72.8% increase in GDP per capita. Since a 1% increase in GERD is a relatively large increase, it may be more appropriate to describe a 0.1% increase in GERD as a 7.28% increase in GDP per capita.

While this simple linear regression model provides a good baseline for testing the relationship between GERD and economic growth, a multiple linear regression model can provide higher accuracy by adding more explanatory variables to explain more of the variation in the dependent variable. Holding these added explanatory variables fixed will help to better uncover the ceteris paribus relationship in question.

Model 2:

Model 2 is a multiple linear regression model constructed by adding in all secondary explanatory variables. This model is written as:

Model 2:

$$log(gdpcap) = B_0 + B_1(gerd) + B_2(gini) + B_3(unemploy) + B_4(educ) + B_5(logproduc) + B_6(savings) + B_7(fdi) + u$$

This model has a sample size of 27 countries, which is noted as being relatively small, since there is limited data available for labor productivity. The issue of micronumerosity can arise from an insufficient sample size, however we continue our regression analysis with caution, as the economic benefits realized by increased labor productivity are a very important byproduct of R&D. From the STATA output, the estimated equation can be written as:

Estimated Equation 2:

$$log(gdpcap) = -15.81 + .36(gerd) - .01(gini) + .02(unemploy) + .09(educ) + 5.31(logproduc) + .02(savings) + .02(fdi)$$

This model has an R-squared value of 0.68, denoting a relatively strong correlation between GERD and GDP per capita. The coefficient on *gerd*, B_1, has a positive sign as predicted, showing that GERD and GDP per capita are positively correlated. Compared to the simple linear regression model, this B_1 coefficient has a smaller value, which is expected since the simple linear regression model overestimates the impact of GERD on GDP per capita due to omitted variable bias. B_1 can be interpreted to show that a 0.1% increase in GERD results in a 3.6% increase in GDP per capita. The sign of the coefficient for *produc* is positive as expected, and this coefficient can be interpreted to show that a 1% increase in GDP per hour worked results in a 5.31% increase in GDP per capita.

Model 3:

To address the issue of small sample size in the previous model, Model 3 is created with all variables except for labor productivity. This model can be written as:

Model 3:

$$log(gdpcap) = B_0 + B_1(gerd) + B_2(gini) + B_3(unemploy) + B_4(educ) + B_5(savings) + B_6(fdi) + u$$

This model has a sample size of 43 countries, which is an improvement over Model 2. From the STATA output, the estimated equation can be written as:

Estimated Equation 3:

$$log(gdpcap) = 6.68 + 0.80(gerd) + 0.01(gini) + 0.04(unemploy) - 0.02(educ) + 0.05(savings) + 0.02(fdi)$$

This model has an R-squared value of 0.63, denoting a mild correlation between GERD and GDP per capita. The coefficient on *gerd*, B_1, can be interpreted to show that a 0.1% increase in GERD results in an 8% increase in GDP per capita. The coefficient on *educ* has a negative sign, which is the opposite of what was originally predicted. This may be because if a country spends more on R&D, it has less money remaining to invest in education. Nonetheless, this coefficient can be interpreted to show that a 0.1% increase in education expenditure corresponds to a 2% decrease in GDP per capita. The signs of the coefficients for both *gini* and *unemploy* are positive despite the fact that they were predicted to be negative, as economic growth would likely cause a decrease in the GINI index (more income equality) as well as a decrease in the unemployment rate. This discrepancy can be explained if economic growth due to an increase in R&D resulted in a loss of jobs, as more jobs may become automated due to advancements in machinery and production methods. This loss of jobs would result in a higher unemployment rate, and the remaining jobs focused on supervising the automated processes would likely be higher-paying jobs due to their complex nature. This would therefore trigger greater income inequality, as individuals making a higher income also receive a larger percentage of the total income of the population.

Out of all the variables tested in this model, GERD has the lowest p-value of 0.00 and a 95% confidence interval spanning [0.44, 1.16], showing that it is a highly significant variable. Using a two-sided T-test, GERD is significant at the 1% level. Gross savings and inflow of FDI are significant at the 5% level.

Model 4:

Using the results from the previous regression, the variables *gini*, *unemploy*, and *educ* were dropped from the next two models due to their statistical insignificance. In Model 4, labor productivity is added back into the regression to assess its impact (with a small sample size in mind). Model 4 can be written as:

$$\textbf{Model 4:}\ log(gdpcap) = B_0 + B_1(gerd) + B_2(produc) + B_3(savings) + B_4(fdi) + u$$

This model has a sample size of 37, which is still small enough to warrant some caution during analysis. From the STATA output, the estimated equation can be written as:

$$\textbf{Model 4:}\ log(gdpcap) = -2.27 + 0.47(gerd) + 2.50(logproduc) - 0.00(savings) + 0.02(fdi)$$

This model has an R-squared value of 0.51, conveying a mild correlation between GERD and GDP per capita. Compared to Model 2, the coefficients for labor productivity and gross savings are both smaller, denoting a smaller impact on economic growth. This change is unexpected, as this model has fewer secondary explanatory variables in it and would therefore suggest increased effects stemming from these two variables. GERD and inflow of FDI are the only significant variables, and they are both significant at the 1% level.

Model 5:

Lastly, Model 5 is constructed by dropping labor productivity from Model 4 in order to increase sample size and assess any changes in variable significances. Model 5 can be written as:

$$\textbf{Model 5:}\ log(gdpcap) = B_0 + B_1(gerd) + B_2(savings) + B_3(fdi) + u$$

This model has a larger sample size of 76. From the STATA output, the estimated equation can be written as:

$$\textbf{Model 5:}\ log(gdpcap) = 0.27 + 0.10(gerd) + 0.01(savings) + 0.01(fdi)$$

This model has an R-squared value of 0.48, showing a mild correlation between GERD and GDP per capita. This R-squared value is the smallest of the four multiple regression models constructed and it also incorporates the least number of explanatory variables. This makes sense, as the R-squared value of a model increases when more explanatory variables added. GERD and gross savings are both significant at the 1% level and inflow of FDI is significant at the 5% level.

A summary of all five regression models presented in this section can be seen below in Table 3.

Table 3 - Regression Models Summary

Dependent Variable: log(gdpcap)					
Independent Variables	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
gerd	0.73*** (0.10)	0.36** (0.14)	0.80*** (0.18)	0.47*** (0.10)	0.63*** (0.01)
gini		-0.01 (0.02)	0.01 (0.02)		
unemploy		0.02 (0.03)	0.04 (0.03)		
educ		0.09 (0.09)	-0.02 (.12)		
log(produc)		5.31** (2.09)		2.50 (1.53)	
savings		0.02 (0.02)	0.05** (0.03)	0.00 (0.02)	0.03*** (0.01)
fdi		0.02** (0.01)	0.02** (0.01)	0.02*** (0.01)	0.02** (0.01)
Intercept	8.76 (0.14)	-15.81 (9.99)	6.68 (1.44)	-2.27 (7.18)	8.03 (0.27)
No. of Observations	80	27	43	37	76
R-squared	0.41	0.68	0.56	0.51	0.48

V. Extensions

After Model 2 and Model 3 were created, F-Tests were conducted to ensure significance of all of the variables. In order to determine which variables might be worth removing, variables with low t-statistics were identified. The GINI index, unemployment rate, and education expenditure all appeared to have low significance. All of them were determined to be individually insignificant in each of the models they appeared in. Thus, F-tests were completed for the variables jointly, for both Model 2 and 3. Model 4 was created as restricted version of Model 2 in order to test the following hypotheses for Model 2:

$$H_0 : B_2 = 0, B_3 = 0, B_4 = 0$$

$$H_1 : H_0 \text{ is false}$$

Using the residual sum of squares for both the restricted (Model 4) and unrestricted model (Model 2), a F-value of 0.702 was determined. At the 10% level, the critical value $F_{3,27}$ is 2.31. Since the critical value of the F-distribution is larger than the calculated F-value, we fail to reject the null hypothesis at the 10% significance level, and can conclude that the GINI index, unemployment rate and education expenditure are jointly insignificant. This led to the removal of the *gini*, *unemploy*, and *educ* from Model 2.

A similar F-test was completed for Model 3, testing the significance of GINI index, unemployment rate, and education expenditure in the model. A restricted regression, identified as Model 4, was created to test the following hypotheses for Model 3:

$$H_0 : B_2 = 0, B_3 = 0, B_4 = 0$$

$$H_1 : H_0 \text{ is false}$$

A F-value of 2.21 was calculated for the test of the above hypotheses. The critical value of the F distribution at the 10% level for $F_{3,36}$ is 2.25. Again, the critical value is larger than the calculated F-value for the regression. Thus, we fail to reject the null hypothesis, and it can be concluded that GINI index, unemployment rate, and education expenditure are jointly insignificant in Model 3. Given the results of the F-tests, *gini*, *unemploy*, and *educ* were removed from Model 3. The combined results of the two F-tests, both concluding that GINI index, unemployment rate, and education expenditure are jointly insignificant, allowed us to be confident that these variables do not have major significance in our model.

To examine an alternate functional form of the models generated in Section IV, the initial scatter plot of the logarithm of GDP per Capita vs. GERD was further analyzed. Based off its shape, the line of best fit may be better approximated as a natural logarithm rather than a linear function. Therefore, a new explanatory variable, *loggerd* (natural logarithm of GERD), was created and added into the original Model 1 to create Model 6:

Model 6: $log(gdpcap) = B_0 + B_1(gerd) + B_2(loggerd) + u$

This model has a sample size of 80 countries. Although the Gauss-Markov assumptions are still met for Model 6, there is a correlation coefficient of 0.86 between *gerd* and *loggerd*, which is a very high value, meaning that the results should be interpreted with caution due to near multicollinearity. From the STATA output, the estimated equation can be written as:

Estimated Equation 6: $log(gdpcap) = 9.15 + 0.48(gerd) + 0.24(loggerd)$

This alternate model has an R-squared value of 0.42, which is very similar to the R-squared value of 0.41 for Model 1. In this alternate model, *gerd* is significant at the 5% level, compared to *gerd* being significant at the 1% level in Model 1. Also, in this alternate model, *loggerd* is found to be insignificant at the 10% level. Based off these results, *loggerd* does not seem to be an alternative functional form of *gerd* worth evaluating further.

VI. Conclusions

Ultimately, our initial hypothesis of a positive correlation between GERD and GDP per capita was supported by each of our linear regression models. With each linear regression model having an R-squared value conveying a relatively mild/strong correlation and a positive linear relationship proven between GERD and GDP per capita, increasing GERD can be viewed as a productive solution to boost GDP per capita with other variables held fixed. Therefore, increasing expenditure on R&D can result in economic growth.

When examining secondary explanatory variables, it was demonstrated that GINI, unemployment rate, and education expenditure did not have significant impacts on economic growth. However, gross savings and inflow of FDI were both consistently significant variables across the different models constructed. This supports the economic rationale that an increase in savings allows for the financing of greater investments and specifically ones of long-term nature, such as R&D. Also, an increase in FDI inflow for a country due to successful R&D efforts will result in positive economic growth.

With the results of this study noted, future research can be conducted to extend the models constructed in order to find more significant secondary explanatory variables. Also, a study could be conducted using multiple different time lags between the main dependent and explanatory variables, such as 3 years, 5 years, and 10 years. This would allow for greater analysis of the optimal time it takes for R&D efforts to materialize and make economic impacts. Acquiring data for a larger cross-section of countries would also help provide a better understanding of the effect of R&D on economic growth. Given the importance of technological growth on society as well as the many emerging technological fields that will create new jobs, shift educational priorities, and establish new areas for investment, R&D is poised to further stimulate many aspects of the global economy for years to come.

Appendix A. List of countries used in study:

Aruba	Congo, Dem. Rep.	Haiti	Myanmar	Sao Tome and
Afghanistan	Congo, Rep.	Hungary	Montenegro	Principe
Angola	Colombia	Indonesia	Mongolia	Suriname
Albania	Comoros	India	Mozambique	Slovak Rep.
Andorra	Cabo Verde	Ireland	Mauritania	Slovenia
United Arab	Costa Rica	Iran, Islamic Rep.	Mauritius	Sweden
Emirates	Cuba	Iraq	Malawi	Eswatini
Argentina	Cyprus	Iceland	Malaysia	Turks and Caicos
Armenia	Czech Republic	Israel	Namibia	Chad
Australia	Germany	Italy	Niger	Togo
Austria	Djibouti	Jamaica	Nigeria	Thailand
Azerbaijan	Dominica	Jordan	Nicaragua	Tajikistan
Burundi	Denmark	Japan	Netherlands	Turkmenistan
Belgium	Domincan Republic	Kazakhstan	Norway	Timor-Leste
Benin	Algeria	Kenya	Nepal	Tonga
Burkina Faso	Ecuador	Kyrgyz Republic	New Zealand	Trinidad and
Bangladesh	Egypt, Arab Rep.	Cambodia	Oman	Tobago
Bulgaria	Spain	St. Kitts and Nevis	Pakistan	Tunisia
Bahrain	Estonia	Korea, Rep.	Panama	Turkey
The Bahamas	Ethiopia	Kuwait	Peru	Tanzania
Bosnia and	Finland	Lao PDR	Philippines	Uganda
Herzegovina	Fiji	Lebanon	Papua New Guinea	Ukraine
Belarus	France	Liberia	Poland	Uruguay
Belize	Micronesia, Fed. Sts	Libya	Puerto Rico	United States
Bermuda	Gabon	St. Lucia	Portugal	Uzbekistan
Bolivia	United Kingdom	Sri Lanka	Paraguay	St. Vincent and the
Brazil	Georgia	Lesotho	Qatar	Grenadines
Barbados	Ghana	Lithuania	Romania	Venezuela, R.B.
Brunei Darussalam	Guinea	Luxembourg	Russian Federation	Virgin Islands
Bhutan	Gambia	Latvia	Rwanda	Vietnam
Botswana	Guinea-Bissau	Macao SAR	Saudi Arabia	Vanuatu
Central African	Equatorial Guinea	Morocco	Sudan	Samoa
Republic	Greece	Moldova	Senegal	Kosovo
Canada	Guatemala	Madagascar	Singapore	Yemen, Rep.
Switzerland	Guam	Maldives	Solomon Islands	South Africa
Chile	Guyana	Mexico	Sierra Leone	Zambia
China	Hong Kong SAR	North Macedonia	El Salvador	Zimbabwe
Côte d'Ivoire	Honduras	Mali	Serbia	
Cameroon	Croatia	Malta	South Sudan	

Appendix B. Correlation coefficients between each variable to fulfill Gauss-Markov Assumption 3:

	loggdpcap	*gerd*	*gini*	*unemploy*	*educ*	*logproduc*	*savings*	*fdi*
loggdpcap	1.00							
gerd	0.63	1.00						
gini	-0.47	-0.58	1.00					
unemploy	-0.12	-0.21	0.14	1.00				
educ	0.28	0.45	-0.21	-0.20	1.00			
logproduc	0.32	0.02	-0.03	-0.11	-0.02	1.00		
savings	0.43	0.42	-0.44	-0.39	0.09	-0.07	1.00	
fdi	0.38	0.01	-0.03	0.01	-0.23	-0.01	0.30	1.00

Appendix C. STATA Regression Model Outputs

<u>Model 1:</u>

```
      Source |       SS           df       MS      Number of obs   =        80
-------------+----------------------------------   F(1, 78)        =     53.69
       Model |  41.1503357         1  41.1503357   Prob > F        =    0.0000
    Residual |  59.7806593        78  .766418709   R-squared       =    0.4077
-------------+----------------------------------   Adj R-squared   =    0.4001
       Total |  100.930995        79  1.27760753   Root MSE        =    .87545

------------------------------------------------------------------------------
   loggdpcap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gerd |   .7282087   .0993807     7.33   0.000     .5303568    .9260605
       _cons |   8.763889   .1417776    61.81   0.000     8.481632    9.046147
------------------------------------------------------------------------------
```

Model 2:

```
. regress loggdpcap gerd gini unemploy educ logproduc savings fdi
```

Source	SS	df	MS			
Model	8.72697711	7	1.24671102	Number of obs	=	27
Residual	4.11089874	19	.216363092	F(7, 19)	=	5.76
				Prob > F	=	0.0011
Total	12.8378759	26	.493764456	R-squared	=	0.6798
				Adj R-squared	=	0.5618
				Root MSE	=	.46515

loggdpcap	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gerd	.3552806	.1416094	2.51	0.021	.0588886	.6516726
gini	-.0132233	.0208186	-0.64	0.533	-.056797	.0303505
unemploy	.021791	.0286864	0.76	0.457	-.0382503	.0818322
educ	.093339	.0930318	1.00	0.328	-.1013787	.2880568
logproduc	5.306513	2.088217	2.54	0.020	.9358247	9.6772
savings	.0168574	.0212916	0.79	0.438	-.0277064	.0614213
fdi	.0151792	.0057524	2.64	0.016	.0031393	.0272191
_cons	-15.80787	9.985786	-1.58	0.130	-36.70836	5.092623

Model 3:

```
. regress loggdpcap gerd gini unemploy educ savings fdi

      Source |       SS           df       MS      Number of obs   =        43
-------------+----------------------------------   F(6, 36)        =     10.07
       Model |  37.5258735         6  6.25431226   Prob > F        =    0.0000
    Residual |  22.3687565        36  .621354347   R-squared       =    0.6265
-------------+----------------------------------   Adj R-squared   =    0.5643
       Total |    59.89463        42  1.42606262   Root MSE        =    .78826

------------------------------------------------------------------------------
   loggdpcap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gerd |   .7995376    .1775536     4.50   0.000     .4394422    1.159633
        gini |   .0126355    .0208965     0.60   0.549    -.0297446    .0550157
    unemploy |   .0413633     .032705     1.26   0.214    -.0249655    .1076921
        educ |   -.016478    .1238325    -0.13   0.895    -.2676219    .2346658
     savings |   .0532101    .0276732     1.92   0.062    -.0029138     .109334
         fdi |   .0151042    .0081181     1.86   0.071      -.00136    .0315685
       _cons |    6.67774    1.441788     4.63   0.000     3.753659    9.601821
------------------------------------------------------------------------------
```

Model 4:

```
. regress loggdpcap gerd logproduc savings fdi

      Source |       SS           df       MS      Number of obs   =        37
-------------+----------------------------------   F(4, 32)        =     10.49
       Model |  10.7057621         4  2.67644053   Prob > F        =    0.0000
    Residual |  8.16502973        32  .255157179   R-squared       =    0.5673
-------------+----------------------------------   Adj R-squared   =    0.5132
       Total |  18.8707919        36  .524188663   Root MSE        =    .50513

------------------------------------------------------------------------------
   loggdpcap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gerd |   .4716535     .100693     4.68   0.000     .2665486    .6767584
   logproduc |   2.501188    1.527207     1.64   0.111    -.6096313    5.612007
     savings |   -.004553    .0175289    -0.26   0.797    -.0402582    .0311523
         fdi |    .017089    .0058494     2.92   0.006     .0051742    .0290037
       _cons |  -2.267043    7.183789    -0.32   0.754    -16.89994    12.36586
------------------------------------------------------------------------------
```

Model 5:

```
. regress loggdpcap gerd savings fdi

      Source |       SS           df       MS      Number of obs   =        76
-------------+----------------------------------   F(3, 72)        =     24.44
       Model |  48.2113568         3  16.0704523   Prob > F        =    0.0000
    Residual |  47.3354782        72  .657437197   R-squared       =    0.5046
-------------+----------------------------------   Adj R-squared   =    0.4839
       Total |  95.5468349        75   1.2739578   Root MSE        =    .81083

------------------------------------------------------------------------------
   loggdpcap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gerd |   .6349452   .0982166     6.46   0.000     .4391541    .8307364
     savings |   .0323036   .0115862     2.79   0.007      .009207    .0554002
         fdi |   .0158545   .0069331     2.29   0.025     .0020337    .0296754
       _cons |   8.028713   .2727427    29.44   0.000      7.48501    8.572415
------------------------------------------------------------------------------
```

Model 6:

```
. regress loggdpcap gerd loggerd

      Source |       SS           df       MS      Number of obs   =        80
-------------+----------------------------------   F(2, 77)        =     28.36
       Model |  42.8152471         2  21.4076236   Prob > F        =    0.0000
    Residual |  58.1157479        77  .754749973   R-squared       =    0.4242
-------------+----------------------------------   Adj R-squared   =    0.4092
       Total |  100.930995        79  1.27760753   Root MSE        =    .86876

------------------------------------------------------------------------------
   loggdpcap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        gerd |   .4769284   .1958317     2.44   0.017     .0869777    .8668792
     loggerd |   .2410682   .1623102     1.49   0.142    -.0821328    .5642691
       _cons |   9.150372   .2958176    30.93   0.000     8.561324    9.739421
------------------------------------------------------------------------------
```

References

"Foreign Direct Investment, Net Inflows (% of GDP)." *The World Bank*, data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS

"GDP Per Capita (Constant 2010 US$)." *The World Bank*, data.worldbank.org/indicator/NY.GDP.PCAP.KD.

"GINI Index (World Bank Estimate)." *The World Bank*, data.worldbank.org/indicator/SI.POV.GINI.

Gocer, Ismet, Sedat Alatas, and Osman Peker. 2016. "Effects of R&D and Innovation on Income in EU Countries: New Generation Panel Cointegration and Causality Analysis." Theoretical and Applied Economics 23 (4): 153–64. doi:http://www.ectap.ro/theoretical-and-applied-economics-archive/.

"Government Expenditure on Education, Total (% of GDP)." *The World Bank*, data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS.

"Gross Savings (% of GDP)." *The World Bank*, data.worldbank.org/indicator/NY.GNS.ICTR.ZS.

Nekrep, Andreja, Sebastjan Strasek, and Darja Borsic. 2018. "Productivity and Economic Growth in the European Union: Impact of Investment in Research and Development." Nase Gospodarstvo/Our Economy 64 (1): 18–27. doi:http://www.degruyter.com/view/j/ngoe.

OECD (2019), "GDP per hour worked" (indicator), https://doi.org/10.1787/1439e590-en (accessed on 17 October 2019).

"Science, Technology and Innovation." *United Nations Educational, Scientific and Cultural Organization*, data.uis.unesco.org/Index.aspx?DataSetCode=SCN_DS&lang=en.

Szarowská, Irena. 2018. "Importance of R&D Expenditure for Economic Growth in Selected Cee Countries." E+M Economics & Management 21 (4): 108–24. http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=eoh&AN=EP133694535&site=ehost-live.

"Unemployment, Total (% of Total Labor Force) (Modeled ILO Estimate)." *The World Bank*, data.worldbank.org/indicator/SL.UEM.TOTL.ZS

Yazgan, Sekip, and Omer Yalcinkaya. 2018. "The Effects of Research and Development (R&D) Investments on Sustainable Economic Growth: Evidence from OECD Countries (1996-2015)." Review of Economic Perspectives 18 (1): 3–23. doi:http://www.degruyter.com/view/j/revecp.