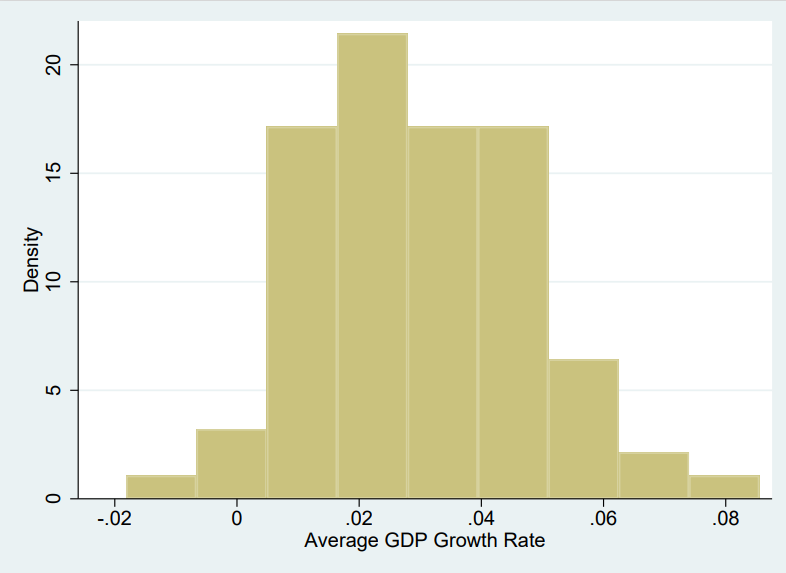
**Part 1**

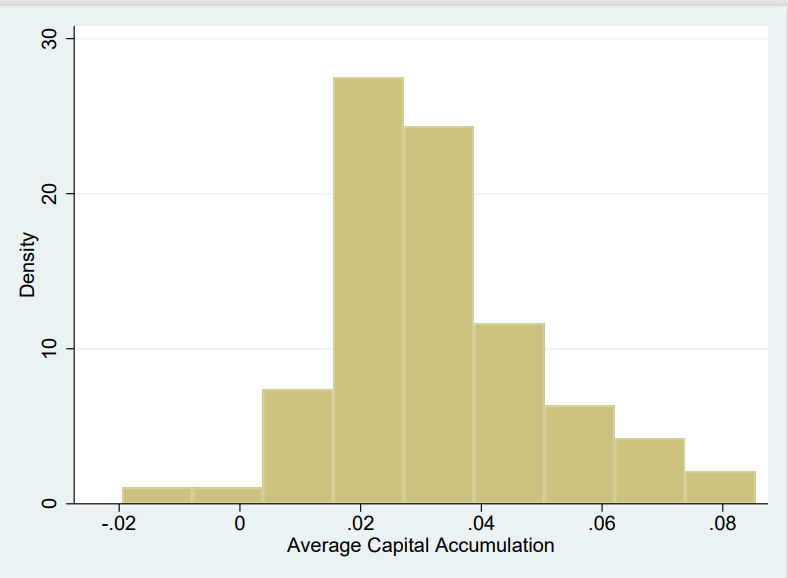
**Section A**

I am conducting a study that explores the cross-section of economic growth between 1990 and 2015. My dataset has 82 countries and 51 variables that are all useful measures of economic growth, including GDP growth, labor force changes, corruption index, education, population, access to electricity, internet, telephone, vaccines, drinking water, health, and life expectancy.

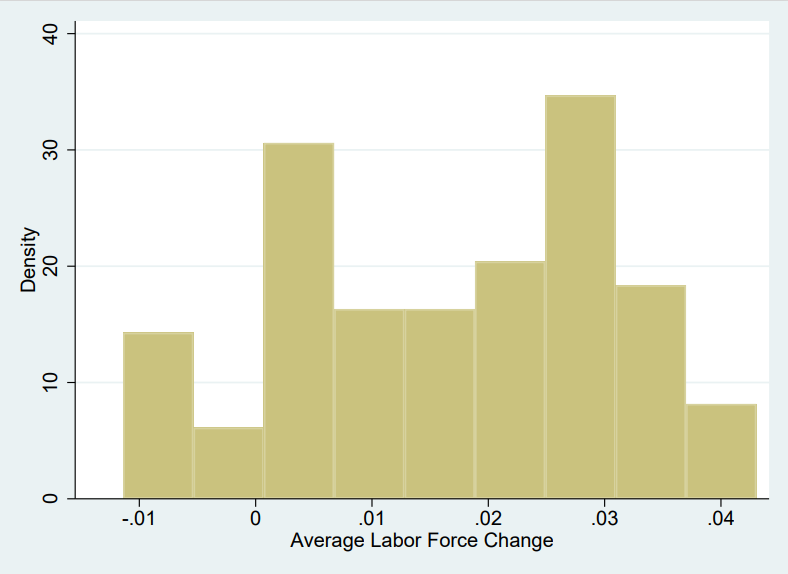
**Histogram of Average GDP Growth Rate**



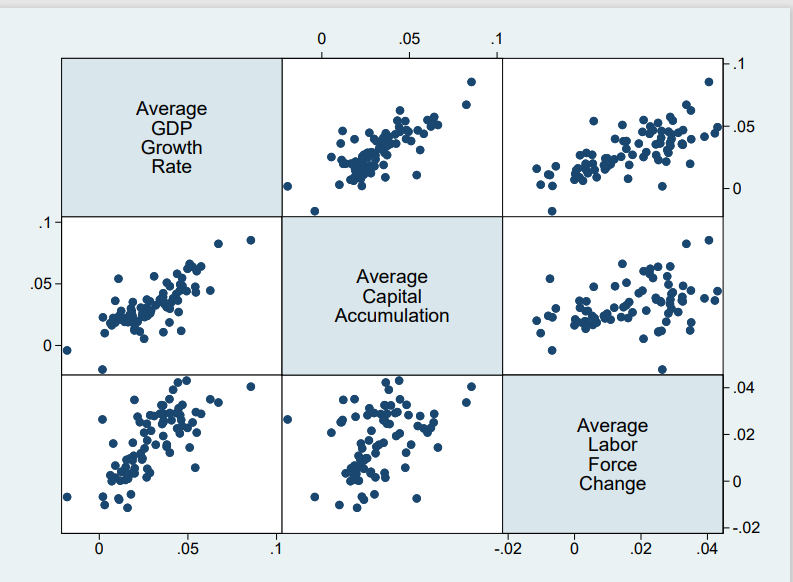
**Histogram of Average Capital Accumulation**



**Histogram of Average Labor Force Change**



After inspecting the histograms for average GDP growth rate, average capital change, and average labor force change, I do not think that there are any marked outliers that might overly influence my results.

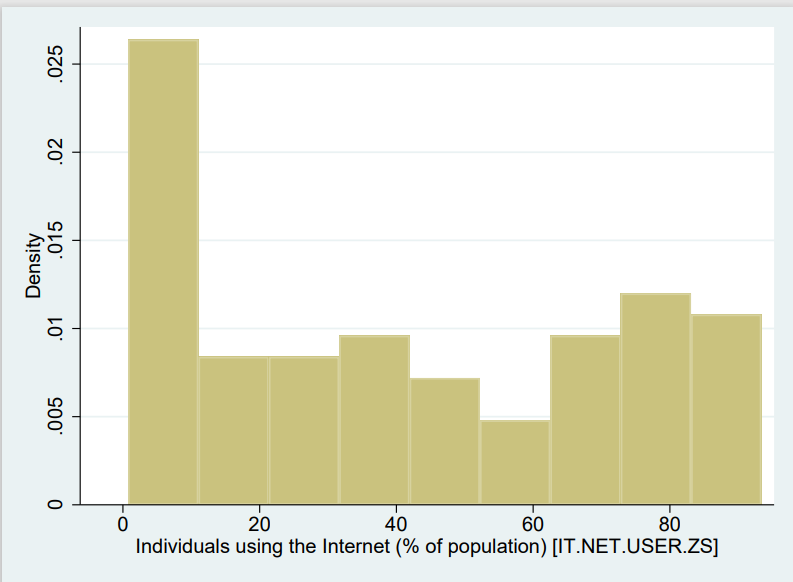


I then plotted several scatterplots to study the relationship between the average GDP growth rate, average capital accumulation, and average labor force change. There is a strong positive relationship between each of the variables.

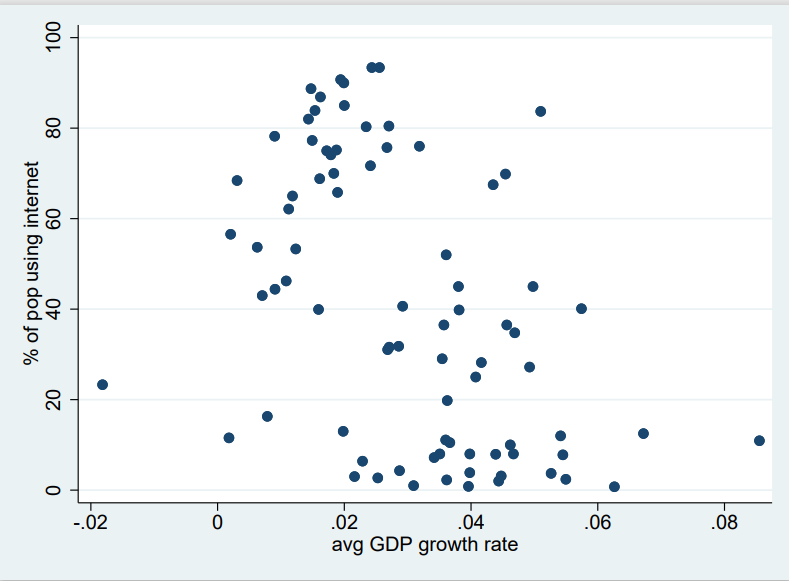
**Section B**

I would like to investigate internet\_usage - individuals who have used the internet in the last 3 months via a computer, mobile phone, personal digital assistant, game machine, digital TV, etc. Individuals benefit from the internet by taking more online courses, which helps them further their education and earn a higher income. Businesses benefit from the Internet because they shift from selling locally to anywhere in the world. Consequently, the country's exports will go up which will improve economic growth.

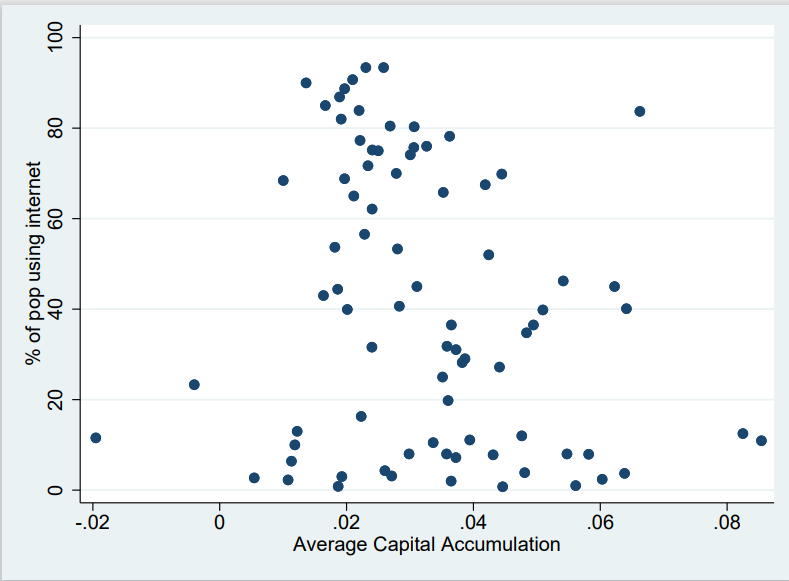
**Histogram of Internet Usage**



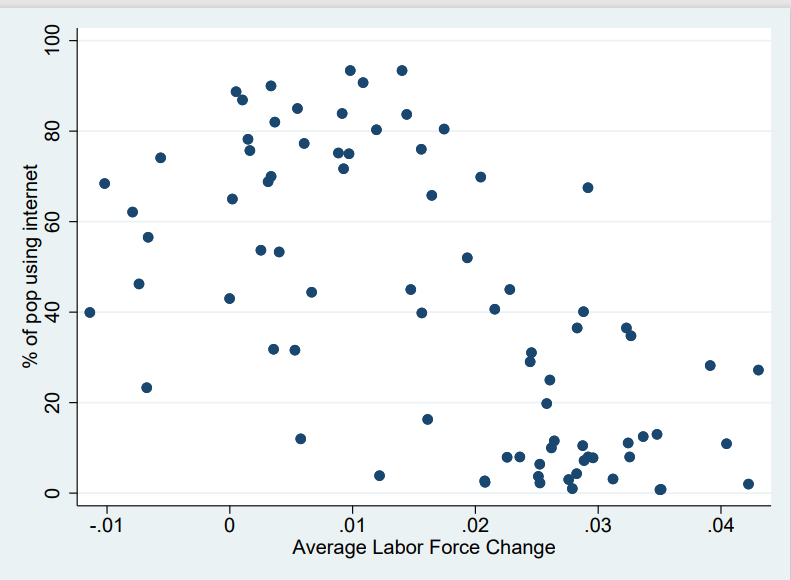
**Scatterplot of Internet Usage Vs Average GDP Growth Rate**



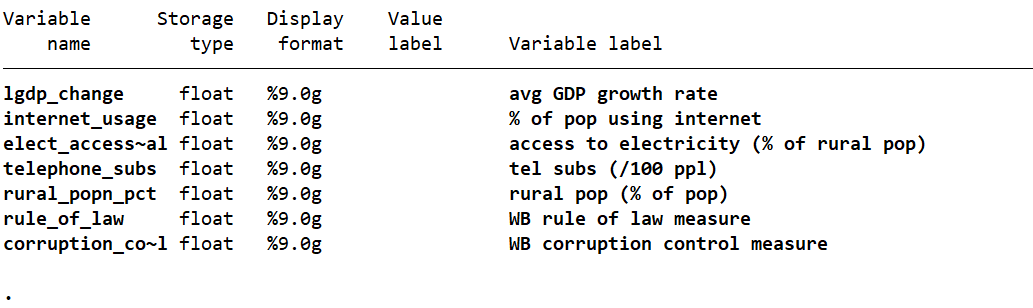
**Scatterplot of Internet Usage Vs Average Capital Accumulation**



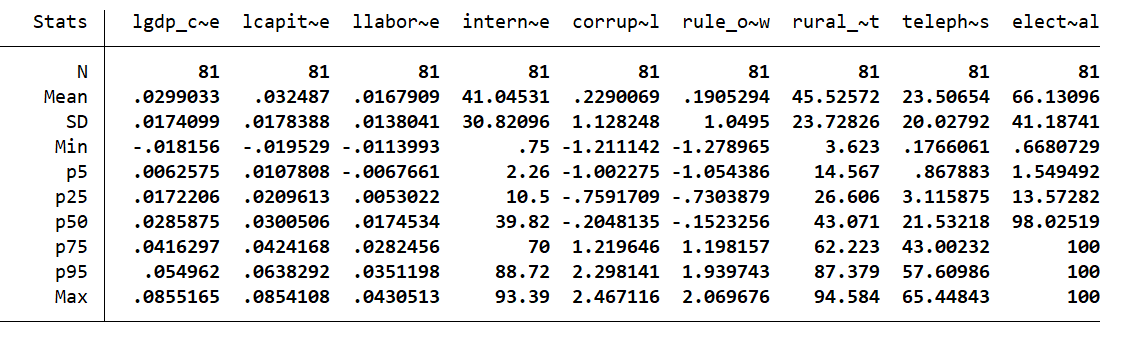
**Scatterplot of Internet Usage Vs Average Labor Force Change**



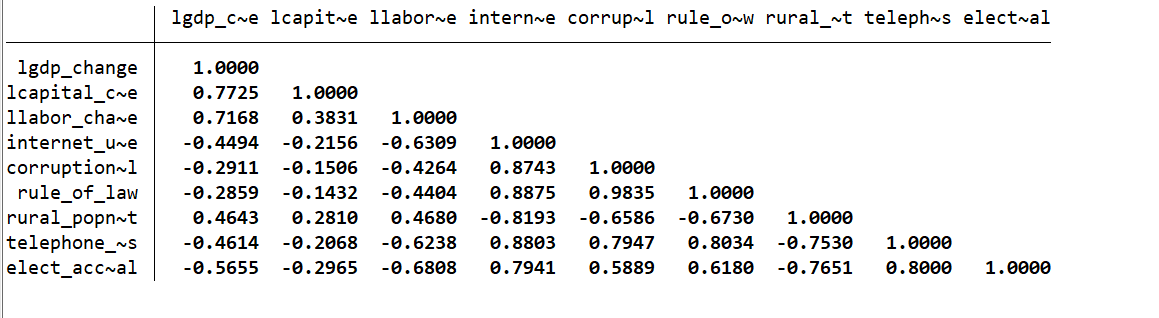
I will choose control variables on the basis of whether they have a high correlation with average GDP growth (measure of economic growth) or internet usage. Variables with high correlation with internet\_usage: corruption\_control (0.87), rule\_of\_law (0.8875), rural\_popn\_pct (-0.8193), telephone\_subs (0.8803). Variables with a high correlation with the average GDP growth rate: elect\_access\_rural (-0.5655).

My analysis will then focus on the following variables:  


**Section C**

Here are also the summary statistics for the above variables:  


I will also present the correlations between the variables in my analysis

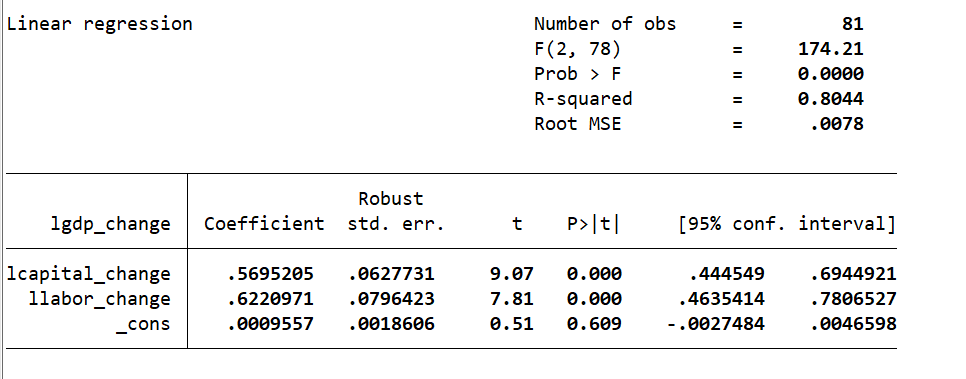


**Part 2**

**Section A**

Regression model: lgdp\_change = b0 + b1\*lcapital\_change + b2\*llabor\_change + u

In the model above, I am regressing the average GDP growth rate against average capital accumulation and average labor force change. The residual is attributable to changes in technology.



**Section B**

Regression model: lgdp\_change = b0 + b1\*lcapital\_change + b2\*llabor\_change + b3\*colonial\_origin + b4\*colonial\_origin\*lcapital\_change + b5\*colonial\_origin\*llabor\_change

H0: b3 = 0, b4 = 0, b5 = 0

To estimate a regression model for colonial and non-colonial countries, I added colonial\_origin and variables that interact between colonial origin and average capital accumulation and colonial origin and average labor force change. If colonial origin is 0, I will have a model that estimates the average GDP growth rate for non-colonial countries. If colonial origin is 1, I will have a model that estimates the average GDP growth rate for colonial countries. Colonial\_origin and interaction of colonial origin and llabor\_change are significant at the 5% level.

To determine whether colonial\_origin and the interaction variables of colonial origin and llabor\_change and colonial origin and lcapital\_change are jointly significant, I ran an F-test. The p-value is 0.1280, so the above-mentioned variables are not jointly significant at any reasonable significance level. There are no statistically significant differences in growth between colonial and non-colonial countries.

**Part 3**

**Section A**

I ran a simple regression model of the average GDP growth rate, average capital accumulation, and average labor force change on internet\_usage. I reported the coefficients and robust standard errors for internet\_usage (treatment variable) below. I have also included the R^2 for the simple regression models.

|  | lgdp\_change | lcapital\_change | llabor\_change |
| --- | --- | --- | --- |
| Coefficient | -0.0002539 | -0.0001248 | -0.0002826 |
| Robust Std Error | 0.0000492 | 0.0000619 | 0.0000298 |
| R^2 | 0.202 | 0.0465 | 0.398 |

Internet usage has a statistically significant effect on average GDP growth and average labor force change at the 1% level and internet usage has a statistically significant effect on average capital accumulation at the 5% level. Despite being statistically significant, internet usage does not have a large practical significance for the average GDP growth rate, average capital accumulation, or average labor force change because the estimated coefficients for internet usage are small in the three simple regression models. The standard errors for internet usage in the above three simple regression models are small. Additionally, the simple regression model of internet usage on average labor force change has the largest R^2 (39.8%) and the simple regression model of internet usage on average capital change has the smallest R^2 (4.65%). The R^2s indicate that variation in internet usage explains variation in average labor force change the most and variation in average capital change the least.

**Section B**

|  | (1) | (2) | (3) | (4) |
| --- | --- | --- | --- | --- |
| VARIABLES | Simple Model | Solow Model | Controlled Model | Solow with Controls |
|  |  |  |  |  |
| % of pop using internet | -0.000254\*\*\* | -1.16e-05 | -0.000126 | 6.99e-05 |
|  | (4.92e-05) | (4.14e-05) | (0.000196) | (0.000102) |
| WB corruption control measure |  |  | -0.00867 | -0.00465 |
|  |  |  | (0.00896) | (0.00519) |
| WB rule of law measure |  |  | 0.0146 | 0.00622 |
|  |  |  | (0.00909) | (0.00617) |
| rural pop (% of pop) |  |  | 5.39e-05 | 0.000101 |
|  |  |  | (0.000145) | (7.14e-05) |
| tel subs (/100 ppl) |  |  | -8.32e-05 | -4.63e-05 |
|  |  |  | (0.000207) | (7.67e-05) |
| access to electricity (% of rural pop) |  |  | -0.000198\*\* | -4.00e-05 |
|  |  |  | (9.25e-05) | (4.99e-05) |
| Average Capital Accumulation |  | 0.570\*\*\* |  | 0.541\*\*\* |
|  |  | (0.0636) |  | (0.0639) |
| Average Labor Force Change |  | 0.605\*\*\* |  | 0.577\*\*\* |
|  |  | (0.109) |  | (0.111) |
| Constant | 0.0403\*\*\* | 0.00169 | 0.0469\*\*\* | -0.00122 |
|  | (0.00309) | (0.00368) | (0.0122) | (0.00819) |
|  |  |  |  |  |
| Observations | 81 | 81 | 81 | 81 |
| R-squared | 0.202 | 0.805 | 0.356 | 0.823 |

In the simple regression model of the average GDP growth rate on internet usage, internet usage is significant. In the Solow model, control model, and Solow model with controls, internet usage is not significant at any reasonable significance level. As other variables in the controlled and solow models are highly correlated with internet usage or the average GDP growth rate, more of the variation in internet usage is absorbed, and internet usage becomes insignificant. Additionally, the estimated coefficients on internet usage are small in all 4 models so the effect of internet usage on the average GDP growth rate is not practically significant.

Through my study, I have found that internet usage’s effect on the average GDP growth rate is not significant at any reasonable significance level because other factors such as average capital accumulation, average labor force change, corruption, telephone subscriptions, and access to electricity are highly correlated with internet usage. Despite that, I only have 82 observations in my dataset so I lose a lot of degrees of freedom when I run my regression with the controlled model and solow with controls. To address this problem, I would like to have more data points in a future study. I would also incorporate more control variables in the *controlled model* and *solow with control models* that are also positively or negatively correlated with the average GDP growth rate and internet usage. I also want to compare regression models where the y-variable is log vs the y-variable is not log to see whether internet usage becomes significant.