**The Impact of Gross Domestic Savings on COVID-19 Cases**

**A Cross Country Analysis**

ECON 385 Econometrics

Professor Abigail Hornstein

Ivan Yang

**Abstract**

The COVID-19 pandemic has taken a pause on everyone’s life, forcing countries across the globe to tighten travel restrictions and quarantine procedures, citizens to commit to robust testing and vaccine mandates. While Gross Domestic Savings (which consists of savings of household, private corporate, and public sector) appears to be an attractive indicator of a country’s financial health and its readiness to combat the coronavirus, few independent studies have uncovered the impact of Gross Domestic Savings on the pandemic. By analyzing data from 118 countries, this paper attempts to study the relationship between Gross Domestic Savings as a percentage of GDP and a country’s response to COVID-19, which is measured in accumulative cases. This study employs single and multiple linear regression models to assess the relationship between the two variables. Secondary independent variables include, Human Development Index (HDI), median income, life expectancy, Gini coefficient, and polity2 score. The null hypothesis is that Gross Domestic Savings has no effect on total cases of COVID-19, whereas the alternative hypothesis would be Gross Domestic Savings has effect on total cases of COVID-19.

**Introduction**

The world is slowly recovering and opening again, but the long-lasting effects of COVID-19 pandemic are still widespread. As of today, the unprecedented global crisis so far has caused 5.3 million deaths toll and 271 million confirmed cases, leaving businesses shut down and millions of people become unemployed around the world (WHO). In the context of globalization, countries are becoming more and more interconnected to one another, the trend of which will likely to continue in the future. One drawback of globalization entails those countries that are exposed to greater pandemic crisis with less preparedness would likely spill over their impacts, extending impact that affects other countries throughout the world. Thus, it is essential for us to understand countries’ level of preparedness prior the pandemic and its relationship with countries’ current response to COVID-19, so that countries can better align themselves for the future if there were to be another global pandemic taking place.

According to Investopedia, Gross Domestic Savings is an economic indicator tracked by U.S. Commerce Department’s Bureau of Economic Analysis, and it measures the amount of income that households, private businesses, and government save, calculated by GDP subtracting final consumption expenditure. Essentially, Gross Domestic Savings offers a good indication of a country’s financial health as investments in infrastructure for instance, and purchase of goods and services are generally generated through savings. This study expects to see that an increase in Gross Domestic Savings will also increase confirmed COVID-19 cases in a country. The rationale behind it is that countries that have higher saving rate are likely to have more resources available to invest in healthcare infrastructure, and their citizens are more likely to be able to afford healthcare goods and services in the time of crisis. Thus, ease of testing that comes with high accessibility in these countries will also lead to higher confirmed cases, simply because these countries are able to conduct more testing than other countries.

This study attempts to uncover the relationship between levels of savings on a national level and their response to COVID-19 pandemic by utilizing pooled cross-sectional data from various sources to construct simple and multiple linear regression models.

**Literature Review**

Ross Levine, Chen Lin, Mingzhu Tai, and Wensi Xie (2021) studied the sharp increase in bank deposits seen its 2020 and attempted to uncover its relationship with coronavirus. They used data from Center for Systems Science and Engineering at Johns Hopkins University for county level infection rates, Google trends for individual anxiety level over future incomes and savings, Census Household Pulse Survey database for individual reduction in spending and public responses, RateWatch for deposit rate data, and FDIC and Call Reports for deposit flow at branch level, etc. Ross et al. then created regression models that assess the association between deposit rates (12-month CDs offered by branch) and COVID-19 cumulative cases per capita at county level. The regression analysis suggests that deposit rates decrease more sharply in counties that have higher infection rates. In other words, deposit rates have a strong and negative relation with COVID-19 exposure. The research also examined whether this negative association is across all banks with different characteristics such as size, geographic locations, and safety level. The data suggests that effect of COVID-19 exposure on deposit rates are insensitive to banks characteristics across all counties. The research also suggests that surging expectations of future income loss (thus less spending), concerns for unemployment, and seeing bank deposits as a safer investment choice are major forces that drove up deposit rates during the pandemic. Therefore, the paper reached the conclusion that “There is a negative relationship between the deposits interest rates at the branch-week level and COVID-19 infection rates at county level” (Ross et al.).

Moreover, a country’s healthcare expenditure reflects a country’s healthcare capabilities and is a good indicator of accessibility, availability of health-related goods and services in the country. The unprecedented pandemic has exposed different countries’ levels of healthcare capacities and responsiveness. Jahidur Khan, Nabil Awan, Md Islam and Olav Muurlink (2020) conducted research on the association between healthcare capacities (a multi-factored index that measures the number of nurses, midwives, physicians, hospital beds per 1,000 people) and COVID-19 case fatalities. They used the John Hopkins University database for confirmed cases and deaths, Worldometer for number of COVID tests carried out by countries in 2020. Researchers found that, “with each additional unit increase in the healthcare capacity index associated with a 42% decrease in case fatalities.” (Khan, et al.) In other worlds, results of the study show that countries with more robust healthcare capacities lead to fewer fatality cases in COVID-19; the result is unsurprising, but it is consistent with existing research that has proved healthcare as a key determinant of mortality rate overall. Overall, the study shows that government’s role is critical in managing the effects of COVID-19, and that countries would be better off if they were to invest and have abundant healthcare resources available ahead of the pandemic.

While Covid-19 serves as a negative, unanticipated income shock, there are research done on consumer behaviors following a positive income shock due to government policy as well. Sumit Agarwal and Wenlan Qian (2014) examined how consumers responded to an unanticipated stimulus fiscal program announced in Singapore back in 2011. By using a difference-in-difference method and analyzing consumer’s monthly response following the announcement, Agarwal and Wenlan found that consumer spending jumped sharply following the announcement, especially in the first two-month period via Credit Card purchases. Specifically, consumers spent 80 cents on every dollar received from the program. The research also found that individuals with low credit card limit and asset liquidity reflect the strongest spike in consumption shortly after the income shock (Agarwal et al.).

**Data**

To understand the relationship between Gross Domestic Savings and a country’s response to COVID-19, pooled cross-sectional data was collected from the World Bank, WHO, United Nations, and World Inequality. The main dependent variable is the natural log of cumulative cases of a country starting from the outbreak of COVID-19. The natural log is used to prevent outliers which will skew the data otherwise. The data is sourced from WHO, which gets updated daily. This study uses two reported dates: “2020-12-30” for total cases in year 2020, and “2021-11-11” for total cases in year 2021 up until the specified date. The primary independent variable is Gross Domestic Savings (GDS) as percentage of GDP, sourced from World Bank in 2017 and 2018. More recent years of data is purposely not chosen as the many countries faced unprecedented crisis due to the pandemic; countries also tend to report their macro-data relatively slow, thus data in recent years is more limited than the ones before. The study chose GDS as a percentage of GDP, as opposed to the actual amount of savings in current U.S. dollars, since wealthier countries would likely to have a higher GDS in actual amount even if it is low as percent of GDP, skewing the data against poorer countries. Hence, expressing GDS as percent of GDP allows all countries to have the chance to have high Gross Domestic Savings rate.

A scatterplot of the natural logarithm of COVID cases and GDS (as a percent of GDP) is shown below as Figure 1.

**Chart, scatter chart

Description automatically generatedFigure 1. Scatterplot of l\_cases vs GDS**

The results show a general, mild correlation between the two variables. As GDS increases, the natural logarithm of COVID cases also increases, indicating the level of Gross Domestic Savings has a positive effect on confirmed cumulative COVID cases in a country.

In addition to the primary explanatory variable of GDS as a percent of GDP, this study includes five other secondary explanatory variables to strengthen and support the multiple linear regression models for the dependent variable of natural logarithm of cumulative cases. This is done to discover the true ceteris paribus effect (holding all else constant) of GDS (% of GDP) on a country’s cumulative confirmed COVID-19 cases (in natural logarithm). These variables are: Human Development Index (HDI), life expectancy (LifeExp), Polity2 Score (polity2), median income (Med\_Inc), Gini coefficient (gini). The data for HDI is sourced from United Nations and created for 2017 and 2018. According to the U.N., HDI is a comprehensive index that measures key human development of a country in three aspects: a long and healthy life, being knowledgeable, and have a decent standard of living. HDI scores between 0 and 1, with score closer to 1 entails a higher level of human development. Data on life expectancy is sourced from World Bank and created for both year 2017 and 2018; it measures a person’s average longevity in years within a country. Polity2 Score is sourced from Ferdi (SCO) for year 2017 and 2018. This variable, according to SCO, is a revised combined POLITY score that captures political regime spectrum, ranging from -10 to +10 (-10 to -6 are “autocracies; -5 to +5 are “anocracies”; +6 to +10 are “democracies”). Median income is sourced from World Bank and obtained for the year 2017 and 2018 and uses purchase power parity in current U.S. dollars. Gini coefficient captures income inequality of a country, in which scores range from 0 to 100. The scores that tend toward 0 entail perfect equality, as opposed to scores that tend toward 100 which entail perfect inequality.

The study expects to see additional positive and/or negative impact on COVID-19 cases due to the addition of secondary explanatory variables. Median income, HDI, and life expectancy would likely to have a negative impact on COVID-19 cases as countries that are more developed are expected to have more testing ability and report more cases. Meanwhile, Gini coefficient would likely to have a more positive impact. A summary and description of all variables used in this study can be found below shown as Table 1.

**Table 1. Variable Descriptions**

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Year | Source |
| GDS | Gross Domestic Savings (as % of GDP) | 2017, 2018 | World Bank |
| HDI | Human Development Index, score from 0 to 1, measured in three dimensions | 2017, 2018 | United Nations |
| Med\_Inc | Median Income, current purchase power parity US $ | 2017, 2018 | World Bank |
| LifeExp | Life Expectancy, measured in average lifespan of an individual in years | 2017, 2018 | World Bank |
| Polity2 | Captures political regime authority spectrum, score from -10 to 10 (categorized in: autocracies, anocracies, democracy) | 2017, 2018 | Ferdi (SCO) |
| L\_cases | Total COVID-19 cases, natural logged | 2020, 2021 | WHO |
| Gini | Gini Coefficient, measures income inequality, range from 0 -- 100 | 2017, 2018 | World Inequality Database |

The descriptive statistics for each variable can be found in Table 2 as shown below.

**Table 2. Variable Descriptive Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Obs | Mean | Std. Dev. | Min | Max |
| l\_cases | 423 | 10.541 | 3.248 | 0 | 17.653 |
| gini | 342 | .563 | .08 | .376 | .746 |
| polity2 | 332 | 4.117 | 6.17 | -10 | 10 |
| GDS | 446 | 20.644 | 16.988 | -68.407 | 66.658 |
| LifeExp | 494 | 72.318 | 7.39 | 52.24 | 84.934 |
| Med Inc | 446 | 10621.794 | 14116.654 | 203.641 | 65957.409 |
| HDI | 360 | .718 | .151 | .386 | .956 |
|  | | | | | |

Data on natural logarithm of COVID-19 accumulative cases, Gross Savings, life expectancy, and median income are relatively consistent in number of observations, while Gini coefficient, Polity2 score, and HDI have similar number of observations within the group, but they are significantly fewer than the first four variables mentioned.

This study utilizes Ordinary Least Squares, and before running regression models, Gauss-Markov assumptions (Classical Linear Model) are checked to ensure we have best linear unbiased estimates for the study. The CLM assumptions are checked as follows (Wooldridge, 2012):

1. **Linearity**: linear parameters and correct model specification.

Models employed in later sections of the study follow: y = β₀ + β₁X₁ + β₂X₂ + ... + βkXk + ε, where β are coefficients parameters and ε as the error term. The model is correctly specified. Thus, the model passes this assumption.

1. **Random Sampling**: all data are obtained from random sampling.

All data are extracted from random populations and samples of countries across the world with no deliberation. Thus, the model satisfies random sampling assumption.

1. **No Perfect Collinearity**: in the sample and population.

Perfect collinearity should not exist between explanatory variables. To test for this assumption, the study used pairwise correlation command in STATA, as can be seen in Appendix. The statistics results showed no perfect collinearity between the explanatory variables.

1. **E (u | x) = 0**: Error term has 0 conditional mean.

As GDS does not fully explain the variation in COVID-19 accumulative cases, because other unobserved variables could have had impacts on COVID cases, which are captured by the error term, but are not included in the study, it is challenging to state with certainty that models used in the study have met this assumption. Thus, the study will take precautions and close consideration for the results will be reported accordingly.

1. **Homoskedasticity**: Var (u | x1,…,xk) = σ².

The expected value of the variance for the error term must hold constant given any values of the explanatory variables. To test if there exists heteroskedasticity in the models, the study uses Breusch-Pagan and Cook-Weisberg test, and we found no presence of heteroskedasticity. Thus, the assumption is satisfied.

1. **Normality**: normal distribution of the error term in population.

This assumption requires the error term to be normally distributed with zero mean and variance. It cannot be said with certainty that this assumption is met. However, to proceed with simple and multiple linear regressions, we assume this assumption is satisfied. The results will be taken with close consideration and interpreted with caution.

All six CLM assumptions are satisfied, the study then proceeds with regression models computed with STATA.

**Results**

Model 1

The simple linear regression is performed to examine the relationship between GDS and natural logarithm of COVID-19 cases, without the presence of other explanatory variables. This regression aims to identify the direct impact that an increase/decrease of GDS has on COVID cases.

**Model 1: l\_cases = β₀ + β₁(GDS) + ε**

**Estimated Equation: l\_cases = 10.361 + 0.044(GDS)**

**(0.229) (0.009)**

**N = 340 R² = 0.072**

This simple linear regression model captures a total of 340 observations. Standard errors of the coefficients are displayed in the parentheses. The sum of squared residuals R² has a value of 0.072, which entails GDS only explains 7.2% of the variation in natural logarithm of COVID cases, indicating a weak correlation between the two variables. Taking log-level implemented in the model into account, the coefficients results can be interpreted as: a 1% increase in GDS will increase COVID-19 cases by 4.4%. This preliminarily confirms our expectation, that there is a positive relationship between GDS and confirmed COVID-19 cases. GDS is statistically significant at 1% level. The simple regression serves as a good baseline for further study. To reduce the potential omitted variable bias and make our model more consistent and accurate with results, several other variables are included through multiple linear regressions.

Model 2

**Model 2: l\_cases = β₀ + β₁(GDS) + β**₂**(LifeExp) + ε**

**Estimated Equation: l\_cases = 2.480 + 0.009(GDS) + 0.12(LifeExp)**

**(1.34) (0.009) (0.019)**

**N = 333 R² = 0.142**

This multiple linear regression model captures a total of 333 observations. R² has a value of 0.142, which increased slightly from Model 1 since we included an additional explanatory variable, but is still relatively low, as GDS and LifeExp together only explains 14.2% of the variation in the dependent variable. The coefficient for GDS shows that a 1% increase in GDS will increase COVID-19 cases by 0.9% (calculated by (exp(0.009)-1)\*100%). The coefficient for LifeExp suggests an additional year of life expectancy increases COVID cases by 12.75%, which comes surprisingly as we expected life expectancy would have a negative effect on COVID-19 cases. The p-values indicate that LifeExp is statistically significant at 1% level, while GDS is not statistically significant.

Model 3

**Model 3: l\_cases = β₀ + β₁(GDS) + β**₂**(LifeExp) + β3(HDI) + ε**

**Estimated Equation: l\_cases = 6.429+ (-0.001)(GDS) + (-0.019)(LifeExp) + 8.956(HDI)**

**(1.851) (0.01) (0.044) (2.301)**

**N = 306 R² = 0.214**

This multiple linear regression model captures a total of 306 observations. R² has a value of 0.214, which is still relatively low in explanatory power. Here we see HDI is statistically significant at 1% level, whereas LifeExp and GDS are not statistically significant at 10% level. This might be due to the fact that LifeExp is a part of the three measurements of HDI, which is why the two variables have a high correlation (0.91). The coefficient of HDI suggests HDI has a positive relationship with natural logarithm of COVID-19 cases, and that 1% increase of HDI leads to 8.95% increase in COVID cases. Surprisingly, GDS now has a negative relationship with natural logarithm of COVID-19 cases, as 1% increase in GDS now leads to 0.1% decrease in COVID cases.

Model 4

**Model 4: l\_cases = β₀ + β₁(GDS) + β**₂**(LifeExp) + β3(HDI) + β4 (Med\_Inc) + ε**

**Estimated Equation: l\_cases = 4.31 + 0.005(GDS) + (-0.008)(LifeExp) + 11.38378 (HDI) + (-.00004)(Med\_Inc)**

**(1.878) (.0099) (.043) (2.437) (.00001)**

**N = 300 R² = 0.254**

This multiple linear regression model captures a total of 300 observations, relatively consistent to what we had before. R² has a value of 0.254, which is still relatively new. Median income variable is included as it can be a good indicator of citizens’ purchasing power and ability to save prior to the pandemic. HDI and Med\_Inc are both statistically significant at 1% level, while GDS and LifeExp are not statistically significant at 10% level. The new coefficient for GDS suggests a small positive relationship with the dependent variable, as 1% increase in GDS will lead to a 0.5% increase in COVID cases.

Model 5

**Model 5: l\_cases = β₀ + β₁(GDS) + β**₂**(LifeExp) + β3(HDI) + β4 (Med\_Inc) + β5(polity2)+ ε**

**Estimated Equation: l\_cases = -.389 + .017(GDS) + .101(LifeExp) + 7.398(HDI) + -.00005(Med\_Inc) + .016(polity2)**

**(3.264) (.014) (.073) (3.802) (.00002)**

**(.033)**

**N = 120 R² = 0.324**

After including variable Polity2 score, the number of observations significantly decreased to 120. The reason is that data on Polity2 for countries across the world is quite limited. We included the variable to see whether the type of regime a country has effects on the cases of COVID, as an autocratic government like UAE who took a top-down approach to manage the outbreak of pandemic via mandatory lockdowns and quarantines will likely to have a different outcome in contrast to a democratic country that took a more bottom-up, laid back approach like USA. The p-values suggest Med\_Inc is statistically significant at 1% level, and HDI is statistically significant at 10% level, while the rest of variables are not statistically significant at 10% level. The coefficients suggest GDS has a positive relationship with the dependent variable as 1% increase in GDS will lead to a 1.7% increase in COVID cases.

Model 6 Full Model

**Model 6: l\_cases = β₀ + β₁(GDS) + β**₂**(LifeExp) + β3(HDI) + β4 (Med\_Inc) + β5(polity2) + β6(gini) + ε**

**Estimated Equation: l\_cases = -1.6 + .009(GDS) + .011(LifeExp) + 13.332(HDI) + -.00005(Med\_Inc) + .0178 (polity2) + 6.415(gini)**

**(3.493) (.013) ( .069) (3.643) (** **.00001) (.030) ( 2.368)**

**N = 118 R² = 0.403**

By including the last variable Gini coefficient, we now have the full model that has all explanatory variables for the model. Gini is included to uncover the effects GDS have on countries with different income inequality levels and how this impacts COVID-19 cases. Almost all coefficients are positive, indicating positive impacts on COVID cases. Only variable that has a negative relationship with the dependent variable is Med\_Inc, and it is statistically significant at 1% level. However, the coefficient is small, and it suggests for 1 unit increase of Med\_Inc will decrease COVID cases by 0.004%. Gini and HDI are also statistically significant at 1% level. GDS has a relatively low t statistic but a high p-value, hence we cannot reject the null hypothesis that GDS has no effect on the natural logarithm of COVID-19 cases.

Table below uses STATA command outreg2 to create a summary of regression statistics for all models. Codes can be found in do.file.

**Table 3. Summary Regression Statistics (Model 1- 6)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | M(1) | M(2) | M(3) | M(4) | M(5) | M(6) |
| VARIABLES | l\_cases | l\_cases | l\_cases | l\_cases | l\_cases | l\_cases |
|  |  |  |  |  |  |  |
| GDS | 0.045\*\*\* | 0.009 | -0.001 | 0.005 | 0.017 | 0.009 |
|  | (0.009) | (0.009) | (0.010) | (0.010) | (0.014) | (0.013) |
| LifeExp |  | 0.120\*\*\* | -0.019 | -0.009 | 0.101 | 0.011 |
|  |  | (0.019) | (0.044) | (0.043) | (0.073) | (0.069) |
| HDI |  |  | 8.956\*\*\* | 11.384\*\*\* | 7.398\* | 13.332\*\*\* |
|  |  |  | (2.301) | (2.438) | (3.802) | (3.644) |
| Med\_Inc |  |  |  | -0.000\*\*\* | -0.000\*\*\* | -0.000\*\*\* |
|  |  |  |  | (0.000) | (0.000) | (0.000) |
| polity2 |  |  |  |  | 0.016 | 0.018 |
|  |  |  |  |  | (0.033) | (0.030) |
| gini |  |  |  |  |  | 6.416\*\*\* |
|  |  |  |  |  |  | (2.368) |
| Constant | 10.361\*\*\* | 2.480\* | 6.429\*\*\* | 4.310\*\* | -0.389 | -1.599 |
|  | (0.230) | (1.340) | (1.851) | (1.878) | (3.264) | (3.494) |
|  |  |  |  |  |  |  |
| Observations | 340 | 333 | 306 | 300 | 120 | 118 |
| R-squared | 0.072 | 0.142 | 0.214 | 0.254 | 0.324 | 0.403 |

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Robustness**

Two F-tests are performed in order to explain the natural logarithm of COVID-19 cases more accurately with the primary explanatory variable GDS and test the potential joint significance GDS has when combined with other secondary explanatory variables, which include life expectancy, HDI, Gini coefficient, polity2 score, and median income. Individual statistical significance of GDS is found in Model 1. To compare the predictive power of the unrestricted model and the predictive power of the restricted model, we create the first F-test, where Model 6 will serve as the unrestricted model, and Model 1 will serve as the restricted model.

F-test 1

*H0:* ***β****2 = 0,* ***β****3 = 0,* ***β****4 = 0,* ***β****5 = 0,* ***β****6 = 0*

*H1:* ***β****2 ≠ 0,* ***β****3 ≠ 0,* ***β****4 ≠ 0,* ***β****5 ≠ 0,* ***β****6 ≠ 0*

The null hypothesis states that life expectancy, HDI, median income, polity2 score, gini coefficient do not have effects on the dependent variable, whereas the alternative hypothesis states the null hypothesis is false. The formula for F-test is: F = [(SSRr - SSRur)/q] / [SSRur /(n-k-1)] (Wooldridge, 2012). We take the restricted residual sum of squares (SSR) from the Model 1 (2605.59865), and subtracts SSR from Model 6 (355.839712), then divided by the number of restrictions (5), to get the numerator of the F statistic (449.952). For the denominator of the F statistic, we use SSRur (355.839712) divided by the degrees of freedom (117), and we get 3.041. Finally, we get the F-statistic which equals 11.4. We can directly calculate the F-test in STATA by running the unrestricted model first (Model 6), then do "test (restricted variables)”. We get:

(1) HDI = 0

(2) Med\_Inc = 0

(3) LifeExp = 0

(4) polity2 = 0

(5) gini = 0

F (5, 111) = 11.40

Prob > F = 0.0000

The p-value suggests that all variables are statistically significant at 1% level. Thus, we reject the null hypothesis in which they do not have effects on natural logarithm of COVID cases.

F-test 2

For the second F-test, unrestricted model is still Model 6, which has all the independent variables. The restricted model will be Model 3, which excludes Med\_Inc, polity2, gini.

*H0:* ***β****4 = 0,* ***β****5 = 0,* ***β****6 = 0*

*H1:* ***β****4 ≠ 0,* ***β****5 ≠ 0,* ***β****6 ≠ 0*

The null hypothesis states that life Med\_Inc, HDI, polity2 score, and gini coefficient do not have effects on the dependent variable, whereas the alternative hypothesis states the null hypothesis is false. We repeat the same process as the first F-test, where we get the result from STATA as follows:

(1) Med\_Inc = 0

(2) polity2 = 0

(3) gini = 0

F (3, 111) = 7.34

Prob > F = 0.0002

The p-value suggests that all variables are statistically significant at 1%, 5%, and 10% level. Therefore, all variables are jointly significant, and we reject the null hypothesis. With the results from two F-tests, we can conclude that while the coefficient on GDS is not statistically significant individually (from Model 6), it is jointly significant when combined with other explanatory variables.

**Extensions**

After analyzing GDS’ relationship with natural logarithm of COVID-19 cases via OLS, it is appropriate to examine the relationship in a different function form using Fixed Effects, as we are interested in individual effects that are unique to each country, which is not captured by OLS. By accounting for the unobserved heterogeneity, we can improve the accuracy of our estimates and reduce potential bias. Since for all independent variables, only data in 2017 and 2018 are collected, it is appropriate to hold only country fixed, instead of both holding both country and time fixed. Each country is assigned with a unique number countryid, and each country shows up twice in the dataset (for year 2017 and for year 2018). The regression statistics are shown as below.

**Table 4. Fixed Effects Model on Country**

Fixed-effects (within) regression Number of obs = 118

Group variable: countryid Number of groups = 59

R-squared: Obs per group:

Within = 0.7281 min = 2

Between = 0.2777 avg = 2.0

Overall = 0.2598 max = 2

F(6,53) = 23.65

corr(u\_i, Xb) = -0.9983 Prob > F = 0.0000

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| l\_cases | Coef. | | St.Err. | t-value | | p-value | [95% Conf | | Interval] | | Sig |
| HDI | 101.036 | | 39.639 | 2.55 | | .014 | 21.532 | | 180.541 | | \*\* |
| Med\_Inc | 0 | | 0 | 2.08 | | .043 | 0 | | 0 | | \*\* |
| LifeExp | 2.69 | | .519 | 5.18 | | 0 | 1.648 | | 3.732 | | \*\*\* |
| GDS | .021 | | .046 | 0.46 | | .644 | -.071 | | .113 | |  |
| polity2 | -.078 | | .057 | -1.38 | | .173 | -.192 | | .036 | |  |
| gini | 7.142 | | 13.098 | 0.55 | | .588 | -19.129 | | 33.412 | |  |
| Constant | -267.167 | | 29.879 | -8.94 | | 0 | -327.097 | | -207.237 | | \*\*\* |
|  | | | | | | | | | | | |
| Mean dependent var | | 12.360 | | | SD dependent var | | | 2.257 | |
| R-squared | | 0.728 | | | Number of obs | | | 118 | |
| F-test | | 23.648 | | | Prob > F | | | 0.000 | |
| Akaike crit. (AIC) | | 85.399 | | | Bayesian crit. (BIC) | | | 104.794 | |
| *\*\*\* p<.01, \*\* p<.05, \* p<.1* | | | | | | | | | | | |
|  | | | | | | | | | | | |

From the FE model, we can see coefficient of GDS increased from 0.009 (Model 6) to 0.021, which entails the unobserved heterogeneity has some but a relatively small effect. HDI and Med\_Inc are both statistically significant at 5% level, life expectancy is statistically significant at 1% level, whereas GDS, polity2, and gini are not statistically significant at all key levels.

Since the data is a pooled cross-sectional and only have two years, we can also perform the first-differencing model as it regresses not on the level of the data, but on the change of data between 2017 and 2018. Results can be seen in Table 5.

**Table 5. First Differencing**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D.l\_cases | Coef. | | St.Err. | t-value | | p-value | [95% Conf | | Interval] | | Sig |
| D.1 HDI | 25.148 | | 27.768 | 0.91 | | .369 | -30.573 | | 80.869 | |  |
| D.1 Med\_Inc | 0 | | 0 | -0.06 | | .952 | 0 | | 0 | |  |
| D.1 LifeExp | .253 | | .452 | 0.56 | | .578 | -.653 | | 1.16 | |  |
| D.1 GDS | -.025 | | .031 | -0.81 | | .419 | -.087 | | .037 | |  |
| D.1 polity2 | -.057 | | .038 | -1.50 | | .139 | -.132 | | .019 | |  |
| D.1 gini | 6.47 | | 8.666 | 0.75 | | .459 | -10.919 | | 23.859 | |  |
| Constant | 1.069 | | .129 | 8.31 | | 0 | .811 | | 1.327 | | \*\*\* |
|  | | | | | | | | | | | |
| Mean dependent var | | 1.173 | | | SD dependent var | | | 0.453 | |
| R-squared | | 0.088 | | | Number of obs | | | 59 | |
| F-test | | 0.833 | | | Prob > F | | | 0.550 | |
| Akaike crit. (AIC) | | 81.620 | | | Bayesian crit. (BIC) | | | 96.163 | |
| *\*\*\* p<.01, \*\* p<.05, \* p<.1* | | | | | | | | | | | |
|  | | | | | | | | | | | |

We can see that no independent variables are statistically significant at all key levels.

Finally, to test if heteroskedasticity exists in the study, the paper used Breusch–Pagan/Cook–Weisberg test in STATA. The test regresses squared residuals from the original OLS model on all explanatory variables, and test for the overall significance of the second regression. See below as Table 6.

**Table 6. Heteroskedasticity Test**

Table

Description automatically generated

We see the p-value at 0.1447 which is not statistically significant at all key levels. Thus, we cannot reject the null hypothesis that there is no joint significance with the error term, which suggests there is no heteroskedasticity present in the model. It is also consistent with the CLM assumption of homoskedasticity as the assumption is not violated.

**Conclusion**

The null hypothesis of the study is that Gross Domestic Savings has no effect on total cases of COVID-19. The hypothesis is rejected at 1% level in the simple linear regression model, indicating that 1% increase in GDS will increase COVID-19 cases by 4.4%. While GDS is not individually statistically significant in any of the multiple regression models, the F-tests show that GDS is jointly significant with all other explanatory variables. The Fixed Effects model suggests HDI, median income, and life expectancy are all individually statistically significant by holding country fixed. However, GDS is not found to be statistically significant in the Fixed Effects model. The first-differencing model suggests there all variables are not statistically significant at key levels. The study has found no presence of heteroskedasticity. Some of the limitations of the study include but not limited to the following: high correlation between HDI and life expectancy, as well as HDI and median income; can include more secondary independent variables that can more accurately explain the effects on COVID-19 cases, like the public healthcare expenditure; could potentially include Fixed Effects for regions/continents by setting dummy variables; could take into account of spillover effects amongst adjacent countries; could also look into the measurement error across independent variables, but more importantly measurement error in count of COVID-19 cases. This study contributes as a baseline for examining factors that impact COVID-19 cases. Particularly, achieving a higher Gross Domestic Savings could potentially help countries better prepare for COVID-19, and future pandemic crisis in general, as higher saving rates allow countries to utilize more resources to test, identify, and strategize their responses to these crises.

Works Cited

Agarwal, Sumit, and Wenlan Qian. 2014. "Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore." American Economic Review, 104 (12): 4205-30.

Donaldson, Jacquelyn. “Relationship between Income Inequality and Economic Productivity: A Cross-Country Analysis.” Handle Proxy, Georgia Institute of Technology, 1 Apr. 2021, http://hdl.handle.net/1853/64484.

“Home - WID - World Inequality Database.” WID, https://wid.world/.

“Human Development Reports.” | Human Development Reports, http://hdr.undp.org/en/indicators/137506.

Kagan, Julia. “National Savings Rate.” Investopedia, Investopedia, 7 Dec. 2021, https://www.investopedia.com/terms/n/nationalsavingsrate.asp.

Khan JR, Awan N, Islam MM and Muurlink O (2020) Healthcare Capacity, Health Expenditure, and Civil Society as Predictors of COVID-19 Case Fatalities: A Global Analysis. Front. Public Health 8:347. doi: 10.3389/fpubh.2020.00347

Kadam, Sachin. “The Impact of a Country’s Health Expenditure on COVID-19 Cases.” Georgia Tech Library , Apr. 2021, http://hdl.handle.net/1853/64481.

Levine, Ross Eric and Lin, Chen and Tai, Mingzhu and Xie, Wensi, How Did Depositors Respond to Covid-19? (October 2020). Available at SSRN: <https://ssrn.com/abstract=3714456>

Wooldridge, Jeffrey M., 1960-. Introductory Econometrics: a Modern Approach. Mason, Ohio: South-Western Cengage Learning, 2012.

“What Is Gross Domestic Saving? Definition of Gross Domestic Saving, Gross Domestic Saving Meaning.” The Economic Times, https://economictimes.indiatimes.com/definition/gross-domestic-saving.

“WHO Coronavirus (COVID-19) Dashboard.” World Health Organization, World Health Organization, https://covid19.who.int/table.

“Polity2 (Polity IV).” SCO, https://competitivite.ferdi.fr/en/indicators/polity2-polity-iv.

“The World Bank” Data, <https://data.worldbank.org/>.

Appendix

Model 1

Graphical user interface, text, table

Description automatically generated

Model 2

Table

Description automatically generated

Model 3

Table

Description automatically generated

Model 4

Table

Description automatically generated

Model 5

Table

Description automatically generated

Model 6

Table

Description automatically generated

Pairwise-Correlation

Text

Description automatically generated with medium confidence