ECON350: Applied Econometrics
A look at Well-Being and Infant Mortality Rates
Empirical Project Paper

Jared Balakrishnan

Drexel University
jared0@drexel.edu

December 10, 2019

Abstract

This paper investigates the effect of different socioeconomic indicators, primarily the GDP per capita on Infant Mortality Rate. Infant Mortality Rates have historically been an important topic of discussion when it comes to the areas of international development and public health. There has always been placed a huge emphasis on the Gross Domestic Product (GDP) of a country while evaluating its well-being. But whether this is the sole important explanatory variable is a question that needs to be examined? Using data provided by the World Bank on 170 countries, we regress Infant Mortality per 1,000 live births on different measures of GDP per capita, Population, Literacy and Agriculture to study the effects of variables not contained in the GDP such as literacy and the percentage of the population engaged in agriculture. Upon investigation, evidence points to the presence of a signficant relationship between the Infant Mortality rates with respect to literacy and agricultural occupation in differing directions.

1 Introduction

One of the major aims of developmental economists has been to study the overall health of a society, given health is the most important variables when it comes to the survival of populations as well as human beings in general. When it comes to the question of measuring a society's overall health, some of the major factors under consideration include the kinds of illnesses affecting the people, the amount of healthcare infrastructure that is available and the number of infants that pass away before the age of one.

One of the prime factors that comes under consideration while adjudging a country's well-being is the level of income in the country, which could be best explained in terms of the GDP and GDP per capita.

However, it is important to note that GDP, however useful as it might be, is not the best measure of the overall standard of living of a country. This is because of the fact that although the changes in the output of goods and services pr person (that is, the change in GDP divided by the population, called the GDP per capita) are often used a measure of whether the average citizen of a country is better or worse off, it fails to capture many extraneous variables that could be considered as being instrumental in determining the general well-being of a population.

This brings attention to the other explanatory variables. How big of a role does a country's level of education play? How important is the country's occupational structure when it comes to determining IMR?

This study therefore will be focused on studying the impacts of the various variables that affect a country's Infant Mortality Rate. We will be looking at data from 2017 for 170 countries in terms of their population, GDP per capita, percentage of population primarily involved in agriculture and the literacy rates statistically in order to determine their collective influence on the IMR.

2 Review of Literature and Economic Theory

2.1 Gross Domestic Product

The Gross Domestic Product (often abbreviated shortly as *GDP*) is a measure for the monetary value of the final goods and services produced within the borders of a country in a given period of time. This interval of time used for measurement usually tends to be in terms of years or three-month periods called *Quarters*. The reason as to why the word final is being specifically mentioned is in order to allude to the fact that these are goods and services purchased by the final end user (International Monetary Fund, n.d) .

The GDP usually comprises of the goods and services produced for sale in the market and also includes non-market production, such as defense (relating to national security) or education services that are provided by the government.

The GDP of a country is usually calculated and reported by a country's national statistical agency. This agency is tasked with compiling information from a multitude of sources and following a series of established international standards while calculating the same. The current international standard for the measurement of GDP is contained in the *System of National Accounts, 1993*, compiled by the International Monetary Fund (IMF), the European Commission, the Organization for Economic Cooperation and Development (OECD) and the World Bank.

According to the most recent statistics and figures available at the moment of writing this report as sourced from the *FRED* or *Federal Reserve Economic Data* (FRED, n.d), a database maintained by the Research division of the Federal Reserve Bank of St.Louis, the GDP of the United States at the end of Q3 in 2019 was estimated to be approximately USD 21,542.104 billion.

2.2 Infant Mortality Rate

According to the Centers for Disease Control and Prevention, Infant Mortality is defined as the "death of an infant before his or her first birthday". The Infant Mortality Rate (IMR) is the number of infant deaths recorded every 1,000 live births. The United States recorded an IMR of 5.8 infant deaths per 1,000 live births as of 2017 (Centers for Disease Control and Prevention, n.d) .

Infant Mortality Rate is often considered to be a very important marker of the health levels of a society. This is so not because of the fact that the survival of newborns is a largely studied public health area, but because of the fact that high infant mortality rate numbers could be indicative of unmet human needs in the areas of sanitation, medical care, nutrition and very importantly, education.

2.3 Economic Theory and Literature Review

The topic of Infant Mortality has always been a recurrent visitor in the arena of problems pertaining to public health and economic development in general. It is estimated that the victims of apparently 30 percent of all deaths can be attributed to children in developing nations, while it is at less than 1 percent in developing countries (Asiedu et al.). Reducing child mortality always has been a high priority of economists and other social scientists considering how it is very openly indicative of unmet standards in the most basic factors involving the well-being of Mankind. In fact, one of the Millennium Development Goals (MDGs) set up by the United Nations reads "To reduce child

mortality" (United Nations, n.d). The observations made by many social scientists tend to point to the fact that a country's level of well-being greatly impacts the said country's Infant Mortality Rate. This in turn leads to the GDP per capita being considered the paramount explanatory variable that might have a significant effect on the Infant Mortality Rate. Erdogan, Ener and Arica found evidence relating a significant negative relationship between GDP and IMR. In fact, Pritchett and Summers (1996:863) in their work have explicitly noted that "wealthier nations are healthier nations" and that "gains from rapid economic growth flow into health gains". This is a completely plausible and simple statement, as in general a higher GDP per capita translates into higher household incomes which in turn implies that people can afford to allocate more money on their personal health, enabling the entire household to be healthier. Cutler, Deaton and Lleras-Muney (2006:106-107) noted that wealthier nations fared better in terms of the major factors that improved the public health excluding income: be it the education system, health infrastructure or better institutions.

The United States had a per-capita GDP of 62,641 USD in 2018 (World Bank, n.d). Yet, Mississippi has an IMR of 8.6 (Centers for Disease Prevention and Control, n.d). Canada has a per-capita GDP of 46,210.5 USD yet the Northwestern Territories (NWT) reported an IMR of 9.6 in 2018 (Statistics Canada, n.d). This necessitates that we pay attention to the other explanatory variables, ultimately bringing us to the question: What say do the extraneous variables other than GDP surrounding a society's well-being have on its IMR?

The link between the parent's education level and the chances of ensuring a child's survival is very well established: there is an negative and significant effect of the mother's level of education on the infant mortality rate (Corman and Grossman, 1985). This is perfectly plausible, given the fact that a more well-educated and well-informed society tends to be more aware of the perils involved in raising a child and ensuring its health. Jamison, Jamison and Hanushek (2006) were able to prove that an increase in the quality of education available translated into an effective decline in the IMR in open economies as compared to closed economies. In addition, Cutler, Deaton and Muney (2006) endorse the fact that education has a direct and positive effect on the health levels of a society.

A good measure of uniformly measuring the levels of education in a nation would be through its literacy rate, which is defined as the percentage of a population over the age of 15 that is literate. In an econometric model, it would be expected that higher levels of literacy among a country's societies have a negative effect on their infant mortality rates.

A very common occurrence among developing countries in general is a larger percentage of the population being involved in agricultural activity. Given the fact that being resident in rural locales, poverty and low levels of a parent's education are prime factors in terms of contribution to the IMR, it also is safe to assume that this segment of the population involved in blue-collar agricultural work is the most vulnerable to child deaths, and that these populations warrant the most counseling in order to reduce the IMR. In a 2012 study, Headey, Chiu and Kadiyala (2012) found that in India, agricultural workers possessed a Body Mass Index (BMI) that was lower as compared to that of people not engaged in agriculture when controlling for health, education and location. Another study by Mustafa and Odimegwu from 2008 points to the conclusion that child mortality is observed to be higher among the children of agricultural workers owing to the nature of agricultural employment and the absence of these parents in caring for their children for prolonged intervals of time (Mustafa and Odimegwu, 2008). Similar observations were obtained from studies conducted in Pakistan and Bangladesh comparing agricultural workers and those not employed in agriculture (Uddin, Hossain and Ullah, 2009 and Kiani, 1992).

This helps us reach a conclusion that parents involved in agricultural labor have a harder time in

ensuring the survival of their children owing to the socio-economic demographics, lower levels of education and limited access to good quality health infrastructure (Singh, Tripathi).

This in turn implies that we could expect the Infant Mortality Rates for a country to increase, with an increase in the percentage of the population that is engaged in agricultural activity as the primary occupation. This intuitively makes sense given the fact that a larger population engaged in agricultural labor indirectly translates into lesser levels of education among the parents, thereby making them prone to child deaths.

3 Description of Econometric Model

3.1 Data Used

This study utilizes data on different socioeconomic indicators from the countries of the world in order to estimate their Infant Mortality Rates. The dataset was forked from the World Bank's open-source repository of datasets. It is from the year of 2017, and contains data on 170 countries. The dataset from 2017 was so chosen as to ensure that the dataset is clean, complete and relatively recent in terms of time.

The features contained in the dataset include:

- **Country**: This is the name of the country, and the primary key in the dataset that uniquely identifies each row that is present in the dataset.
- **GDPPC**: This is the country's GDP per capita in US dollars as of 2017, which is basically the GDP of the country for that year divided by its population.
- **Literacy**: The literacy rate of the country. This is defined as the percentage of the population over the age of 15 that is capable of reading and writing.
- **InfantMortality**: The Infant Mortality Rate of the country. This is the number of infants that lost their lives before the age of 1, for every 1,000 live births.
- **Population**: The population of the country in 2017.
- **Agriculture**: This field indicates the percentage of the country's population that is engaged in agriculture as a primary source of employment.
- **NetMigration**: The net migration of people from the country for every 1,000 people.

It is a known fact that good results cannot be achieved via data analysis without cleaning the data. This is so because cleaning the data ensures that all of the data present is of a consistent format, and that missing values are appropriately taken care of. While inspecting the dataset for any potential changes to be made in terms of feature introduction or feature rescaling, it was found that there were some things that needed to be taken care of prior to conducting any form of regression analysis on the data.

Firstly, it was noticed that the values of GDPPC and Population were spread over a huge interval, sparking concerns about a potential presence of outliers and the skewing of data to one extreme or the other. To check this, boxplots of the GDPPC and Population were made with the help of the Python programming language. A boxplot is a standardized way of displaying the data in a distribution based on a five number summary consisting of the *minimum*, *first quartile Q1*, *median*,

third quartile Q3 and the *maximum*. It is an effective method to visualize data since it gives one a very clear idea about the presence of outliers, the symmetry of the dataset, how tightly the data is grouped and whether the data is skewed or not.

The boxplot of the GDPPC feature came out to be as shown below:

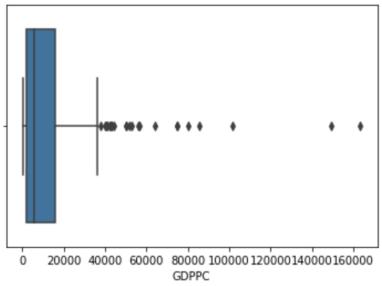

Figure 1: Boxplot of the GDPPC variable

As could be seen from the picture, it could be seen that using the GDPPC variable as is could throw the results of the regression tests in a haywire direction, towards one extreme or the other. From classroom knowledge of econometrics, it is also known that whenever the value of the dependent variable is larger than 0, using the natural logarithm of the dependent variable would satisfy the CLM assumptions in a better fashion as compared to using the level of the dependent variable. As a consequence, it was decided that the natural logarithms of the GDPPC and Population would be used, leading to the birth of two features by the name of **logGDP** and **logPopulation** respectively.

In order to see if the natural logarithms perform better, box plots of the new distribution for **logGDP** were made, as could be seen below:

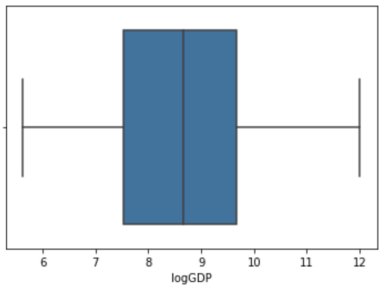

Figure 2: Boxplot of the GDPPC variable

From the above figure, it could be seen that the problem of outliers has been handled in a much more efficient manner by using the natural logarithms of GDPPC and Population.

Other modifications made to features include the creation of the variables **AgroPer** from the variable *Agriculture* and **LitPer** from the variable *Literacy*. These new variables are nothing but merely the

values of the parent variables multiplied by 100. This was done so because these percentage variables were divided by 100, which when used in a regression model would yield abnormally high values for their coefficients as compared to the other explanatory variables which would have coefficients that are far lower in magnitude.

3.2 Model Description

With the dataset in possession, the econometric model constructed comes out as being shown below:

$$IMR = \beta_0 + \beta_1 log(GDP) + \beta_1 log(Population) + \beta_2 \%(literacy) + \beta_3 \%(agriculture) + u$$

With the introduction of the new simplified variables **logGDP**, **logPopulation**, **LitPer** and **AgroPer**, the model can be re-written as:

$$IMR = \beta_0 + \beta_1 (logGDP) + \beta_1 (logPopulation) + \beta_2 (LitPer) + \beta_3 (AgroPer) + u$$

where the variables are as described in the previous section.

The summary statistics for the explanatory variables are as shown in the below table:

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
IMR	170	24.34294	23.26419	1.8	100
GDPPC	170	15227.18	24233.63	276	163352
logGDP	170	8.642355	1.49715	5.620401	12.00366
Population	170	4.15e+07	1.50e+08	15000	1.37e+09
logPopulation	170	15.795	2.040346	9.615806	21.03881
AgroPer	170	12.31941	13.10839	0	70.1
LitPer	170	85.32118	18.82417	6.4	100

Table 1: Summary Statistics of the variables involved in the Econometric Model

Initially, before the construction of the final model, a bunch of other models were considered and regressed upon the IMR. Most of the other models considered failed to explain more of the variance as the above model could, because of which they ultimately ended up being disregarded. Some of the models used included:

- Regressing just the GDP per capita (GDPPC) on the IMR.
- Regressing the level form of all the given variables in the dataset on the IMR.
- Employing a dummy variable titled poverty that takes on the value of 1 when the GDPPC ¡ 1550 and 0 otherwise, and then regressing it on the IMR along with the explanatory variables from the above model.

The reason as to why this model was chosen over the others will be discussed in the results section below.

4 Results

As coule be inferred from the tables that follow, the regression of the IMR against the logarithm of the GDP per capita, the logarithm of the population, the percentage of the population involved in agriculture as a primary occupation and the literacy rate is a good fit for the data. In fact, together, these 4 variables explain about 76.9% of the variation in the Infant Mortality Rates, based on the R-Squared value, which is actually quite high considering the number of explanatory variables included in the regression. The intercept term of the regression equation as seen in the model explained above cannot be considered meanginful to the study given the fact that it is impossible for countries to have an infant mortality rate if their population is equal to zero.

The coefficient of the natural log of the GDP per capita, the loggdp shows that when all the other explanatory variables are controlled for, the Infant Mortality Rate is projected to decline by 5.306 % for every one percentage increase in the GDP per capita. This negative effect is as expected, given a higher GDP per capita entails a higher level of income among the country's residents, who in turn would also be more capable of accessing a good healthcare system. But at the same time, the p-value indicates that the population is not as significant, that is, barely insignificant. This is a very good way to see that the problem of Infant Mortality is pervasive in all kinds of societies regardless of the population, although wealthy countries with extremely small populations might not have to deal with the issue as much.

Controlling for all the othe explanatory variables, the coefficient on the natural log of the population indicates that there is a -0.35% decline in the Infant Mortality Rate for every one percent increase in the population. This is also in agreement with what is naturally expected by the study, considering the fact that it is easier to allocate financial as well as infrastructural resources to a smaller population.

Similarly, the negative coefficients on the litper and agroper variables indicate that there is a significant, and negative relationship of the two variables in relation with the Infant Mortality Rate. It is observed that there is a -0.586% decline in the IMR for every one percent increase in the literacy rate of the country. This makes sense and is in agreement with our ideas given the fact that a more educated population would be naturally more aware as well as be knowledgable of what needs to be taken care of in the event of child birth, thereby pointing to a lower infant mortality rate. In the case of the percentage of the population engaged in agriculture, there is an increase in the IMR by 0.28% for a percentage increase in the proportion of people engaged in agriculture. This is also as expected, given the fact that a higher percentage of the population engaged in agriculture is usually associated with more rural areas and lesser wealth, which in turn implies lesser education levels and a lack of access to good health infrastructure.

The other models used in comparison were not as effective in explaining the relationship between the explanatory variables and the explained variable. Using a level model of the IMR and the GDP per capita resulted in just 20% of the variation in the Infant Mortality Rates, as could be seen from the R-Squared value of the regression model. The coefficients of the explanatory variables were also very skewed, with the coefficient on the GDP per capita being -0.00044 and the intercept term being 31.049. A similar observation was made in the case of the level model regression of infant mortality against the GDP per capita, literacy rate, percentage of population in the agriculture and the population, with population becoming a totally insignificant variable and the presence of very small coefficients on the explanatory variables. The idea to create the dummy variable titled poverty was scrapped owing to potential issues of multicollinearity since it is basically a condition placed on the natural logarithm of the GDP per capita.

4.1 Tables

Independent Variable	Model 1: Level-Log Model (Best Choice)
loggdp	-5.3069 (1.0648)
logpopulation	-0.3548202(0.445313)
litper	-0.5865578 (0.0677683)
agroper	0.2791889 (0.1154402)
intercept	122.4179 (13.61972)
Observations	170
R-Squared	0.7696
Adjusted R-Squared	0.7640

Table 2: Regression Results of Infant Mortality on Log(GDP per capita), Log(Population), Literacy Rate, Percent of Population Engaged in Agriculture as a primary occupation.

Independent Variable	Model 1: Level-Level Model
gdppc	-0.0001014 (0.0000427)
population	-8.23e-10(6.14e-09)
litper	-0.6892711 (0.0666406)
agroper	0.5715273 (0.0992079)
intercept	77.68914 (6.59102)
Observations	170
R-Squared	0.7437
Adjusted R-Squared	0.7374

Table 3: Regression Results of Infant Mortality Rate on GDP per capita, Population, Literacy Rate, Percent of Population Engaged in Agriculture as a primary occupation.

Independent Variable	Model 1: Level-Level Model
gdppc	-0.0004404 (0.0000658)
intercept	31.04901 (1.879592)
Observations	170
R-Squared	0.2105
Adjusted R-Squared	0.2058

Table 4: Regression Results of Infant Mortality Rate on GDP per Capita

5 Conclusion

From the analysis, it is very clear that the issue of child deaths is something important that needs to be solved across the world in order to ensure a proper level of development of human society as a whole. Statistical evidence has pointed to a significant relationship between the levels of income, the level of education and the percentage of the population that is involved in agriculture as a primary occupation. In fact, judging by the t-statistics of the variables used in the model for this study, it could be seen that all of our variables are significant at the 95% level, although population is barely insignificant. The increase in levels of education as well as the amount of counselling provided to expectant mothers have proven to be effective in bringing the Infant Mortality Rates (Uddin, Hossain and Ullah, 2009). On a similar note, the influx of populations into more organized and salaried forms of employment, unlike agriculture, have proven to be effective as well.

International developmental goals and policies have also proved to be instrumental in bringing the rates of child death down, worldwide. However, from the analysis as well as a review of literature, it is seen that countries and organizations should take a concerted initiative in improving the parameters associated with a population's well being other than just the income levels. Although policy change and global policy initiatives are harder to implement, alternative options should be taken advantage of, in the areas of policymaking by shifting the area of focus from just wealth.

6 References

- Asiedu, Elizabeth, Neepa B Gaekwad, Malokele Nanivazo, Mwanza Nkusu, and Yi Jin. "On the Impact of Income per Capita on Health Outcomes: Is Africa Different?," n.d. Accessed October 15, 2019.
- Erdoğan, Engin, Meliha Ener, and Feyza Arıca. "The Strategic Role of Infant Mortality in the Process of Economic Growth: An Application for High Income OECD Countries." Procedia - Social and Behavioral Sciences 99 (2013): 19–25. https://doi.org/10.1016/j.sbspro.2013.10.467.
- Jamison, Eliot, Dean Jamison, and Eric Hanushek. "The Effects of Education Quality on Income Growth and Mortality Decline," 2006. https://doi.org/10.3386/w12652.
- "Infant Mortality." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 27 Mar. 2019, https://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm.
- "Finance & Development." Finance & Development — F&D, https://www.imf.org/external/pubs/ft/fandd/basics/gdp.htm.
- "Gross Domestic Product." FRED, 27 Nov. 2019, https://fred.stlouisfed.org/series/GDP.
- "United Nations Millennium Development Goals." United Nations, United Nations, https://www.un.org/millenniumgoals/childhealth.shtml.
- Pritchett, Lant, and Lawrence H. Summers. "Wealthier Is Healthier." The Journal of Human Resources, vol. 31, no. 4, 1996, p. 841., doi:10.2307/146149.
- Cutler, David, et al. "The Determinants of Mortality." 2006, doi:10.3386/w11963.
- Statistics Canada: Canada's National Statistical Agency / Statistique Canada : Organisme Statistique National Du Canada,

https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310071301&pickMembers[0]=1.13.

- Corman, et al. "Determinants of Neonatal Mortality Rates in the U.S.: A Reduced Form Model." NBER, 1 June 1984, http://www.nber.org/papers/w1387.
- Uddin J, Hossain Z, Ullah MO (2009). Child Mortality in a Developing Country: A Statistical Analysis, Journal of Applied Quantitative Methods; 4: 270-283.
- Singh, Rajvir, and Vrijesh Tripathi. "Under-Five Mortality among Mothers Employed in Agriculture: Findings from a Nationally Representative Sample." PeerJ, vol. 3, Aug. 2015, doi:10.7717/peerj.710.
- Mustafa HE, Odimegwu C. 2008. Socioeconomic determinants of infant mortality in Kenya: Analysis of Kenya DHS 2003. Journal of Humanities and Social Sciences 2(2):1–16.
- Kiani MF. 1992. Differentials in child mortality and health care in Pakistan. Pakistan Development Review 31(4.2):1051–1058.
- Headey D, Chiu A, Kadiyala S. 2012. Agriculture's role in the Indian enigma: help or hindrance to the malnutrition crisis? Food Security 4(1):87–102 DOI 10.1007/s12571-011-0161-0.