AN ECONOMETRIC INVESTIGATION INTO THE DETERMINANTS OF INFANT MORTALITY RATES

NIALL SHERRY

Junior Sophister

Given the unprecedented worldwide economic progress of the past century, the fact that premature death amongst infants is still widespread constitutes one of the greatest humanitarian travesties of our time. In this paper Niall Sherry interestingly points out, that the wealth of a country is not necessarily a safeguard against this issue. Examining the relationship between various developmental indicators and their influence on infant mortality rates across a broad range of countries, his results suggest that while income growth is a determining factor, short-term medical relief may in fact be more effective than long-term policy measures.

Introduction & Motivation

'I believe that children are our future…'
Whitney Houston, 'The Greatest Love of All'

It is quite evident today that high infant mortality rates are a major problem in the developing world. However, there is also scope for investigation into the determinants of disparate infant mortality rates across the developed world, where economic advancement does not automatically bring about a corresponding drop in mortality rates. Even within some of the wealthiest nations in the world, there exists a relatively large disparity in infant mortality levels. These differences can be clearly seen when one compares Qatar (21 per 1000) with Japan (4 per 1000); or Saudi Arabia (26 per 1000) with Singapore (3 per 1000). Even within highly developed countries like the USA, infant mortality rates can at times remain high relative to rates in poorer nations. This project aims to measure infant mortality rates across the world by considering a number of variables linked to income, healthcare, equality and female education in order to determine whether causal relationships can be established. A thorough understanding of the causes of infant mortality can be of use to governments in both the developed and developing world as they strive to reduce mortality rates. Such knowledge can potentially help ensure that governments of developing nations achieve a more equitable distribution of resources. This paper will also analyse how relevant these causes are at particular stages of economic development by running the overall model on specific subsets of the data.

Theory and Variables

The dependent variable (LINFMOR) for this project was taken to be the infant mortality rate, specifically the rate of neonatal death in children under one year of age per 1,000 live births, as measured by the World Bank.

The first independent variable (LINCOMEQ) to be included in the model was average income per person in each country as measured by GNI per capita. This approach follows the Atlas method. The variable acts as a general indicator of the economic status of the country. Here, we expect GNI per capita to have a negative impact on the mortality rate. Interestingly, past research may suggest that this relationship is not intuitive. From his study of infant mortality rates in England over the 16th, 17th and 18th centuries, Schellekens points out that: 'Since a rise in living standards may have improved parents' ability to acquire breast-milk substitutes, post-neonatal mortality could have risen with living standards', as breast milk provides a vital boost to the immune system of newborn infants (Schellekens, 2001: 4). However, the results of his research show a negative correlation between infant mortality and economic development. Schellekens suggests a number of reasons for this. For instance, he asserts that the eighteenth century decline in the neonatal mortality rate was: 'due to an improvement in the health and nutrition of mothers' (Schellekens, 2001: 9).

The second independent variable (LHEALTHQ) was the level of governmental expenditure per capita on healthcare. Again, a negative relationship was expected here, since more government spending in the area of healthcare should logically lead to better post-natal treatment for infants and the better provision of medicine. However, Andes stresses the importance of providing not just more healthcare, but good healthcare; consequently a third independent variable (DIAR) was included to measure the effectiveness of treatment within the given countries (Andes, 1989). Specifically, this variable refers to the percentage of children who receive fluids and treatment when suffering from diarrhoea (which was selected over immunisation rates for DPT and measles based on their poorer performance in the model).We would anticipate that higher percentages of treatment would lead to lower mortality rates. In practice, however, the latter of these two ended up being the sole measure of healthcare effectiveness in the sample countries. The reasons for this are detailed fully below.

The fourth independent variable (FEMLIT) was female literacy rates. One of the most important factors in preventing infant mortality is well educated mothers. Waldmann posits that:

> 'greater female literacy could reduce infant mortality because literate women are more likely to know how to protect their infants' lives…Alternatively, high female literacy rates may indicate an egalitarian government devoted to primary education for all, which also places a high priority on public health' (Waldmann, 1992: 1294).

Glewwe, who specifically investigated the relationship between the two in Morocco, concurs with the reasoning that:

> 'Health knowledge appears to be the most important skill that mothers (indirectly) obtain from their schooling that prepares them to provide for their children's health' (Glewwe, 1999: 154).

He also provides data to demonstrate the relationship. The literacy rate of women aged 15-24, which is the primary age group for new mothers in most countries around the world, was used here as a measure of their human capital. This indicator was selected over the female primary education rate due to its better performance within the model. The literacy rate of women was expected to have a negative partial effect on the infant mortality rate.

Finally, inequality within countries needed to be examined for two reasons. First, to again quote Andes: 'a comprehensive account of the social processes important in reducing infant mortality includes an understanding of the socioeconomic context of the community' (Andes, 1989: 395). Inequality is one way of measuring the general social context of a country. Secondly, inequality has proven to be a strong influence on infant mortality in the past. Although Waldmann cannot conclusively offer any theory as to why this is, he fails to reject:

> 'the possibility that the results are caused by a positive correlation of the rich share and the relative price of e.g. health care…[or] the possibility that a larger fraction of babies are born to poor families in countries with a high rich share' (Waldmann, 1992: 1299).

Irrespective of the reasoning, inequality within countries has been demonstrated in the past to be a good determinant of the infant mortality rate. The findings below add some more weight to the theory, even if discussing or demonstrating any further reasoning or causality is beyond the scope of this project. To measure the effect of inequality, the Gini coefficient of income inequality (GINI) is used. This is the only independent variable to be included for which a positive parameter sign was expected, since as inequality rises, one would expect the infant mortality rate to rise as well.

Data

By and large, the data used in this project comes from the World Bank's database for the year 2000. This year was chosen as it contained the most complete and sufficient datasets for the largest number of countries. This allowed for complete data to be collected on 100 countries in total. However, several points should be noted.

For certain observations, in lieu of any data for a country in a given year, data from several years to either side of 2000 was used. It was anticipated that variables such as the female literacy rate wouldn't change too much within a short period of time. Thus, the results drawn from this project are still, by and large, statistically significant.

For two of the independent variables, diarrhoea treatment and literacy rates, data on first world countries wasn't available for some observations (due to the fact that a lack of diarrhoea treatment is not a problem in, say, the United States). In order to attain a feasible dataset, these were assumed for a number of first world countries to be virtually total, and set at 99%. While this somewhat diminishes the accuracy of the model, any discrepancy between the real data and the substitute data is likely to be minuscule and therefore insignificant.

For one independent variable, the Gini coefficient, the data was taken not from the World Bank but from the United Nations development programme as more data was available from the latter. Also of note is the large discrepancy in collection years for the data used in calculating these coefficients, ranging from the early 1990s up to 2007. Once again, while the limiting effect this divergence has on our ability to draw inferences from the data below is acknowledged, it is again contended that large-scale income inequality within a country should remain relatively constant over such a short time period.

Empirical Results

The population regression model that was utilised in this investigation took the following functional form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$$

where:

Y = Infant mortality rates per 1,000 live births
X_1 = GNI per capita, measured by the Atlas method
X_2 = Percentage of diarrhoea treatment
X_3 = Female literacy rates between 15-24
X_4 = The Gini coefficient of income inequality
u = The error term of statistical residuals

Note: the exogenous variable, government health expenditure, has been omitted from the model.

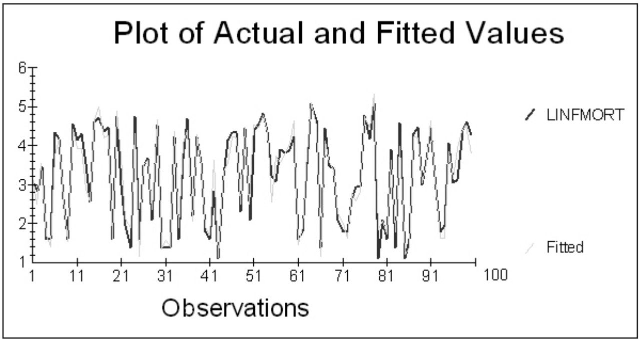
In this study, it was deemed appropriate to convert the endogenous variable as well as the X_1 and X_4 exogenous variables to log form in order to reduce heteroskedasticity and obtain a better functional form. Furthermore, the X_1 variable was also transformed to quadratic form. This allowed for a better reflection of the fact that the effect of increased income on infant mortality diminishes as income rises (as observed by running the model through *Microfit* in various forms). Once the data was passed through *Microfit* the subsequent set of results were attained.

Table 1: Regression Results

Regressor	Coefficient	Standard Error	T-Ratio	P-Value
INTERCEPT	3.8229	0.58226	6.5656	0.000
LINCOMEQ (X1)	0.20175	0.014517	-13.8975	0.000
DIAR (X2)	-0.010081	0.0015443	-6.527	0.000
FEMLIT (X3)	-0.006973	0.0018005	-3.8728	0.000
LINEQ (X4)	1.0040	0.14505	6.9213	0.000
R-Squared	0.94114			
R Bar-Squared	0.93866			
F-Statistic F(4,95)			379.7312	0.000
Functional Form CHSQ (1)			0.41333	0.520
Normality CHSQ(2)			2.9904	0.224
Heteroskedasticity CHSQ(3)			0.066697	0.796

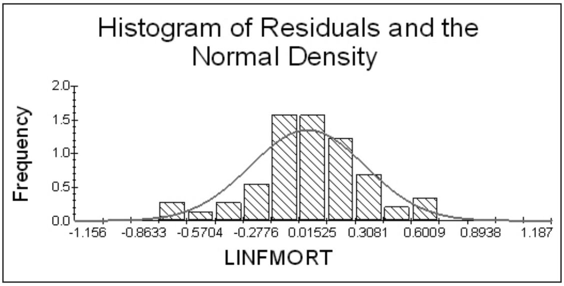
The R-squared result of 94.1% describes how much of the variation in the Y variable is explained by the model. Generally, it is prudent to also consider the R Bar-Squared result when a model consists of many exogenous variables. This result essentially imposes a penalty for adding extra independent variables to the regression. However, only four regressors are included in the model above and thus we would expect the two figures to be similar. A quick glance at the regression output table confirms this. An analysis of the t-ratios indicates that all the population parameters are statistically significant at the 1% level. This is also true of the intercept term. Alternatively, the same results can be inferred by looking at the p-values. Small p-values usually reflect evidence against the null hypothesis (H_0: $\beta_j = 0$). Furthermore, all of the signs for the coefficients fall in line with expectations. Given the large calculated F-statistic value of 379.7, we reject the null hypothesis that this model has no explanatory power. In fact the p-value suggests that there is no significance level at which it should not be rejected. As well as this, a plot of the fitted vs. actual values of Y implies that the sample regression function fits the data well.

Figure 1. Plot of Actual and Fitted Values



The heteroskedasticity result is also promising. Using the Koenker-Bassett test for heteroskedasticity, we fail to reject the null of homoskedasticity even.at the 79.6% significance level. Since the data used in this model is cross sectional, autocorrelation between residuals is not expected. Therefore, the results of the Durbin-Watson and Breusch-Godfrey tests are ignored. Ramsey's RESET test for correct functional form produces a small chi-square value and quite a high p-value. Thus we fail to reject the null hypothesis of correct functional form at any of the regular significance levels (1%, 5%, 10%). The Jacque-Berra test of normality of residuals reports a chi-square value of almost 3 and a p-value of 22.4%. Such figures result in a failure to reject the null of normally distributed residuals at the 10% significance level. In addition, the histogram of residuals seems to support the conclusion of the JB test. This solidifies the belief that the disturbances in the model follow a normal distribution.

Figure 2. Histogram of Residuals



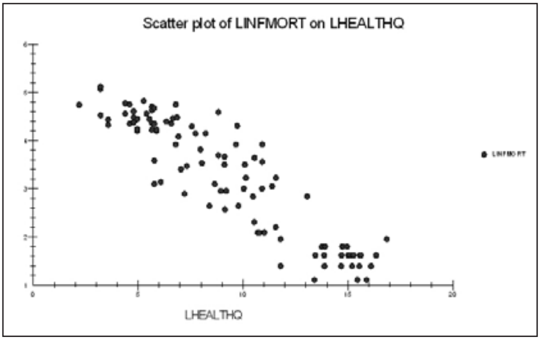
It is interesting to note at this point why one of the original variables, government health expenditure, has been excluded. The following are the regression results when it is included:

Table 2. Regression Results Including the Exogenous Variable LHEALTHQ

Regressor	Coefficient	Standard Error	T-Ratio	P-Value
INTERCEPT	3.9710	0.80383	4.9401	0.000
LINCOMEQ	-0.21942	0.06735	-3.2580	0.002
LHEALTH	0.01616	0.060120	0.26879	0.789
DIAR	-0.010167	0.0015847	-6.4156	0.000
FEMLIT	-0.0070433	0.0018282	-3.8527	0.000
LINEQ	0.99812	6.7724	6.7724	0.000
R-Squared	0.94118			
R Bar-Squared	0.93805			

As can be seen, the result here for health expenditure is insignificant at all of the usual levels. It adds almost nothing to the explanatory power of the model. However, this is not the truly worrying part. This model suggests that health expenditure has a positive partial effect on infant mortality. That is, the more money the government spends on healthcare, the more likely newborn babies are to die. Aside from being intuitively ridiculous, this is clearly proven wrong by examining the relationship purely between health expenditure and mortality (see Figure 3).

Figure 3. Scatter Plot of LINFORM on LHEALTHQ



We can clearly see a negative relationship between the two variables. Indeed, when the model is constructed with healthcare but without income included among the independent variables, the result is not dissimilar from the original model. In this case, healthcare becomes the most important variable with a strong negative relationship between it and the dependent variable. The reason for this effect is the incredibly strong correlation of .9683 between average income and government healthcare expenditure, close to perfect positive correlation.

This suggests an interesting effect whereby, the world over, government spending on welfare is completely dependent on the general welfare of the country irrespective of the type of government or party currently in power. However, a further investigation of this issue would go beyond the scope of this project.

Before we move on to analyse the results themselves, it is also worth including here the results of two sub-tests that were performed on the richest 25 and poorest 25 countries included in the original data set.

Table 3. Rich Countries Regression Results

Regressor	Coefficient	Standard Error	T-Ratio	P-Value
INTERCEPT	3.9296	21.4031	0.01836	0.856
LINCOMEQ	-0.04287	0.046487	-0.92221	0.367
DIAR	0.013004	0.21451	0.60620	0.551
FEMLIT	-0.2012	0.057522	-3.4977	0.002
LINEQ	1.637	0.24662	6.6379	0.000
R-Squared	0.83089			
R Bar-Squared	0.79707			

Table 4. Poor Countries Regression Results

Regressor	Coefficient	Standard Error	T-Ratio	P-Value
INTERCEPT	4.2911	1.6990	2.5256	0.020
LINCOMEQ	-0.17572	0.092475	-1.9002	0.072
DIAR	-0.007983	0.005481	-1.4566	0.161
FEMLIT	-0.009295	0.002563	-3.6266	0.002
LINEQ	0.80172	0.32445	2.710	0.023
R-Squared	0.70070			
R Bar-Squared	0.64084			

Analysis & Conclusions

Using the original data set, the obvious conclusion is that income is the most important determinant of infant mortality. This makes perfect sense. More money in a country and in the average family means more money available to keep children alive. Taking the high correlation between income and healthcare expenditure discussed earlier into account allows us to also conclude that the most powerful way to prevent infant mortality is to promote economic development. Although this is a simplistic and perhaps obvious result, it is an interesting one from an economic point of view. Also worth noting is the strong emphasis placed on the distribution of wealth as equally as possible, and the necessity of providing basic healthcare services such as diarrhoea treatments to children. This last one is perhaps the most interesting in terms of immediate action as it suggests that swift provision of basic medical care is more important than long-term education and improvement programmes.

However, once we start to analyse the separate sub-data sets for rich and poor countries, the results change slightly. For the ‘rich countries’ sub-set, income has become an almost irrelevant factor. As mentioned in the introduction, this can be seen empirically in the contrasting mortality rates of the various rich countries. Instead, by far the most important factor in developed countries is the distribution of wealth. As noted previously, it may be impossible to determine the exact causal link in this relationship but there does appear to be a very clear and strong correlation between the two. Interestingly, female literacy is still a relatively important measure, suggesting that even countries in the developed world have some way to go in terms of educating their populations equally. Less surprising, however, is the fact that the population parameter for the exogenous variable, diarrhoea treatments, has now become statistically insignificant at the 10% level. These are relatively standard across all first world countries, and the question of basic healthcare provisions is a much less pressing one than in poorer countries.

Indeed, looking at the 'poor countries' sub-set, we can see that the population parameter for diarrhoea treatments is surprisingly insignificant at the 10% level, while education rates for females still have a statistically significant effect on the infant mortality rate. Income and inequality play a vital role in these areas as the internal distribution of wealth is often imbalanced. The differences between the various countries are often more pronounced than in richer countries due to the greater marginal value of a dollar in poorer nations. This is why the income variable entered the model in quadratic form. Also worth noting is the drop in the R-squared and R bar-squared figures to 70% and 64% respectively. This suggests, unsurprisingly, that there are more variables involved in dealing with less developed countries such as the availability of clean water.

Thus, our conclusions are somewhat perplexing. The original model seems to lack at least some explanatory power due to the widely varying socioeconomic nature of the countries being observed. The sub-sets, however, are too small to be of any real statistical value. Rather they offer mere suggestions of what could be uncovered by using additional observations in a more focused way. Therefore, the aim of this project was not to offer definitive results, but instead to propose suggestions. The first of these is that, at the most basic level, economic growth is the best way to prevent infant mortality. The second is that the distribution of economic wealth within a country defines the general socioeconomic outlook of that nation and thus is immensely important. Finally, immediate medical provision appears to be a significant factor in preventing infant mortality and should perhaps be the focus of those nations aiming to deal with the problem in the short-term. The problem of infant mortality will not be resolved overnight, but well targeted short-term solutions can help to prevent immediate loss and harm.

Bibliography

Andes, N. 1989. 'Socioeconomic, Medical Care, and Public Health Contexts Affecting Infant Mortality: A Study of Community-Level Differentials in Peru'. *Journal of Health and Social Behaviour* 30:4:386-397.

Glewwe, P. 1999. 'Why Does Mother's Schooling Raise Child Health in Developing Countries? Evidence from Morocco'. *The Journal of Human Resources* 34:1:124-159.

Schellekens, J. 2001. 'Economic Change and Infant Mortality in England, 1580-1837'. *Journal of Interdisciplinary History* 32:1:1-13.

Waldmann, R. 1992. 'Income Distribution and Infant Mortality'. *The Quarterly Journal of Economics* 107:4:1283-1302.

Data drawn from the World Bank website: (http://devdata.worldbank.org/dataonline/) and the UN Human Development Report: (http://hdrstats.undp.org/indicators/147.html)