Term Project

For
T. Nosanchuk
53.512

A Model for the Prediction/Interpretation of Infant Mortality Rates

Warren Munroe
239013
December 22, 1999

Table of Contents

Introduction

Why do so many more children die within the first year after birth in some countries and not in others? In order to become familiar with some of the factors which may influence the survival of infants it may be helpful to employ a framework. In this exploration, infant mortality rates (IMR) will be examined from two perspectives. Structural elements will be consider in tandem with societal priorities. Of course frameworks may include and exclude variables inappropriately, but they can help to quickly focus the inquiry towards important elements. As well, they can help weigh the level of specificity (or generality) of the indicators.

Structural elements considered here will be 'the percentage of people living in large cities' in a country as well as 'the percentage of the labor force involved in agriculture'. Increased urbanization is a general indicator of resource refinement and therefore value added to raw goods. The amount of labor involved in agriculture indicates the availability of people (or lack of) for service in other economies. The thinking here is that the more people involved in agriculture, the fewer there will be involved in wealth production and other activities such as health care. Within these variables are included distribution networks, systems for providing health care and other services. As well, other less physical aspects are important to consider such as techniques for the development of medicines and the commitment to care.

A commitment to care can be crudely examined by looking at societal priorities. Two variables which are readily accessible are levels of military expenditures and levels of education expenditures as a percentage of the GNP. Here we can begin to examine different orientations for different countries in relation to the general welfare of the inhabitants. Military spending may take money and resources away from health care services and increase IMR more directly through the disruption of health services, as well as indicate an increase in conflict. Education expenditure on the other hand may provide an opportunity to improve care by improving knowledge for the development of tools and techniques for both prevention and cure. Using these general indicators, regression analysis can further our understanding of the possible causes of infant mortality.

Methodology

Using these indicators, a series of steps have been followed in order to analysis the data more fully. They are:

PART I - IMR, DEFENSE, URBANITY

Step 1. Regression - IMR against Defense spending
Scattergram - examine for linearity - transformation
Calculate standardized residuals

Step 2. Regression - IMR against Urbanized population
Scattergram - examine for linearity - transformation
Calculate standardized residuals - Y'

Step 3. Regression - Urbanized population against Defense spending
Scattergram - examine for linearity
Calculate r and r2
Calculate standardized residuals - X2'
Propose model

Step 4. Regression - Y' against X2'
Calculate r, r2, F, t, & p.

Step 5. Regression - IMR against Defense spending for low Urban population
Scattergram - calculate beta (standardized slope)

Step 6. Regression - IMR against Defense spending for high Urban population
Scattergram - examine for linearity - transformation

Step 7. Create multiplicitive term.
R, R2

PART II - ADD AGRICULTURE

Step 8. Partial correlation coefficient matrix

Step 9. Agriculture against Urbanity
test for colinearity (r, r2)
interaction - IMR, Defense controlling low and high agriculture.

PART III - ADD EDUCATION

Step 10. Scattergram of Defense and Education expenditures showing Agricultural labor

Step 11. Multiple regression using Defense, Urbanity, X1X2, Education
Calculate R, R2, F,t,p,b,beta.

PART IV - PROPOSED MODEL

PART V - Categorize countries by region with dummy variables.

Since an analysis using these steps would be cumbersome for a full data set of all the countries in the world, a random selection of a portion of the middle income countries is utilized. The data used is from the World Bank - World Development indicators, 1999. The indicators used are explained as follows: Infant Mortality Rate is the number of infants who die before reaching one year of age, 1,000 live births in a given year; Urbanization: the percent of the total population of a country living in cities; Military Expenditures: percent of central government expenditure; Labor involved in Agriculture: the percent of the total population of a country; and Education expenditure: percent of central government expenditure. A random selection of eighteen middle income countries (with a Gross Domestic Product from 3,001 to10,000 US dollars - 1996) was made resulting in the following table 1. For a more detailed explanation, refer to the World Develpoment Indicators and it's sources.

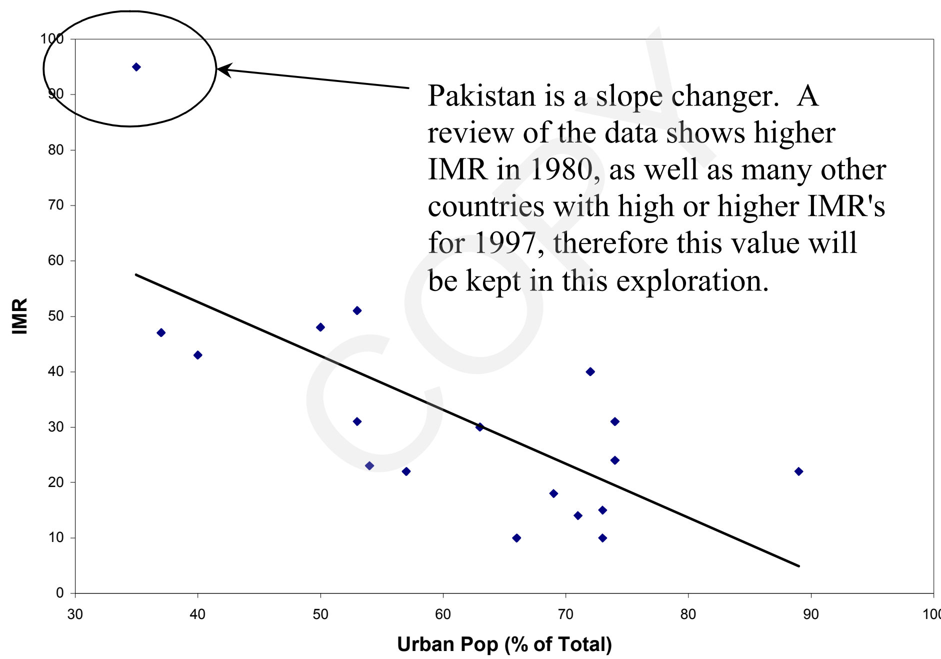
Table 1. Eighteen randomly selected countries with Defense spending, Urban population and Infant Mortality.

Country	IMR	LaborAg	Ed expend	Defense	%urban
Argentina	22	12	3.5	1.7	89
Bulgaria	18	13	3.3	2.8	69
Colombia	24	27	4.4	2.6	74
Guatemala	43	52	1.7	1.3	40
Hungary	10	15	4.7	1.5	66
Indonesia	47	55	1.4	1.8	37
Latvia	15	16	6.5	0.9	73
Lithuania	10	18	5.6	0.5	73
Mexico	31	28	4.9	1	74
Morocco	51	45	5.3	4.3	53
Pakistan	95	52	3	6.1	35
Paraguay	23	39	3.9	1.4	54
Romania	22	24	3.6	2.5	57
South Africa	48	14	7.9	2.2	50
Syria	31	33	4.2	7.2	53
Tunisia	30	28	6.7	2	63
Turkey	40	53	2.2	4	72
Ukraine	14	20	7.2	2.9	71

Part I. IMR(Y) against Urbanity (X1) & Defense Spending (X2)

In order to check for linearity quickly, scatterplots of each of the independent variables (Defense spending and Percentage of the Total Population living in cities of 1,000,000 people or more) were run against IMR. In each case, there was a slight slope in the graph. The values were then squared and then logged (log 10) and run in new scatterplots. But the transformation did very little to change the graph. In each case the residuals were standardized and placed in scatterplots and compared again. The transformations did nothing to improve the interpretability of the data suggesting that the slope is very slight and therefore the raw data is used.

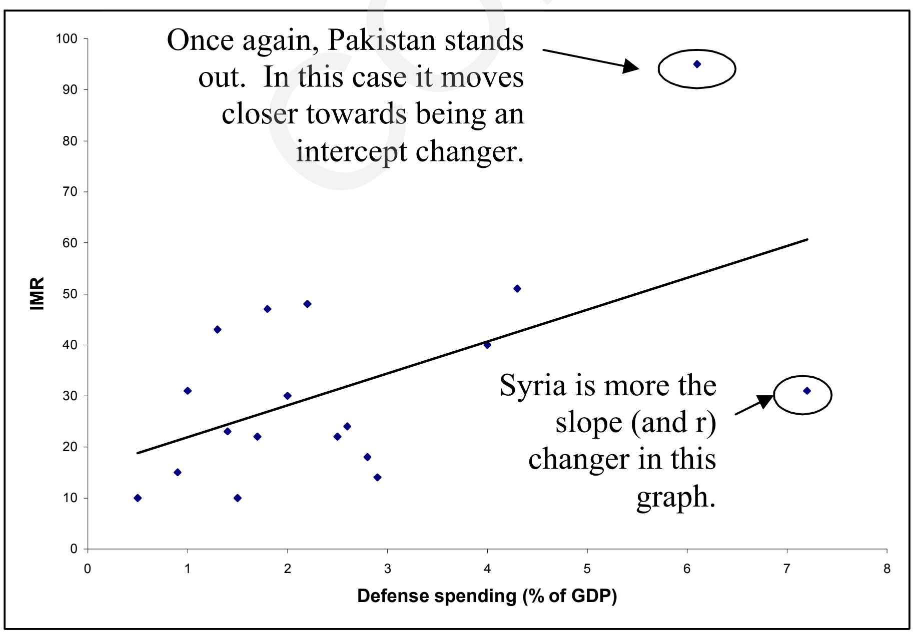
Figure 1. Scatterplot of Urban population (X) against IMR (Y)



Looking more closely at the distribution of the data, it appears as though there is a wide range of IMR around the seventy five to eighty percent marks for Urbanized Population. Some countries have a very low IMR with a relatively high urban population while others have a much higher IMR. But there is an overall oval shape, with the exception of Pakistan as noted in the graph. The regression equation here is Y = 91.5 - .97 X with an r squared of .50. For this sample then, every percent increase in urban population see a drop in infant mortality by almost one unit. This approach begins to break open the garbage can variable, wealth. Wealth product increases with cities since cities often are engaged in refinement of raw resources. Cities also tend to have distribution systems for the provision of services which meet basic necessities and additional medical care. Distribution systems into rural settings are by definition less extensive. But Urbanization only explains 50% of the variation in IMR and therefore it is only a small part of the provision of health care.

In any country, whether highly urbanized or not, choices can be made regarding how revenues (or borrowed capital) may be spent. People may choose to allocate money and human resources in very different directions. A countries peoples choice may affect the citizens immediate welfare. For example, priorities may be for education on one hand or military spending on the other. High military expenditures would divert resources from the health care of the general population while also possibly indicating armed conflict, neither of which are good for the reduction of Infant mortality. Yet not all countries are so engaged and therefore we should see an adequate amount of variation to make it comparable to IMR.

Figure 2. Defense spending (X) Against IMR (Y)

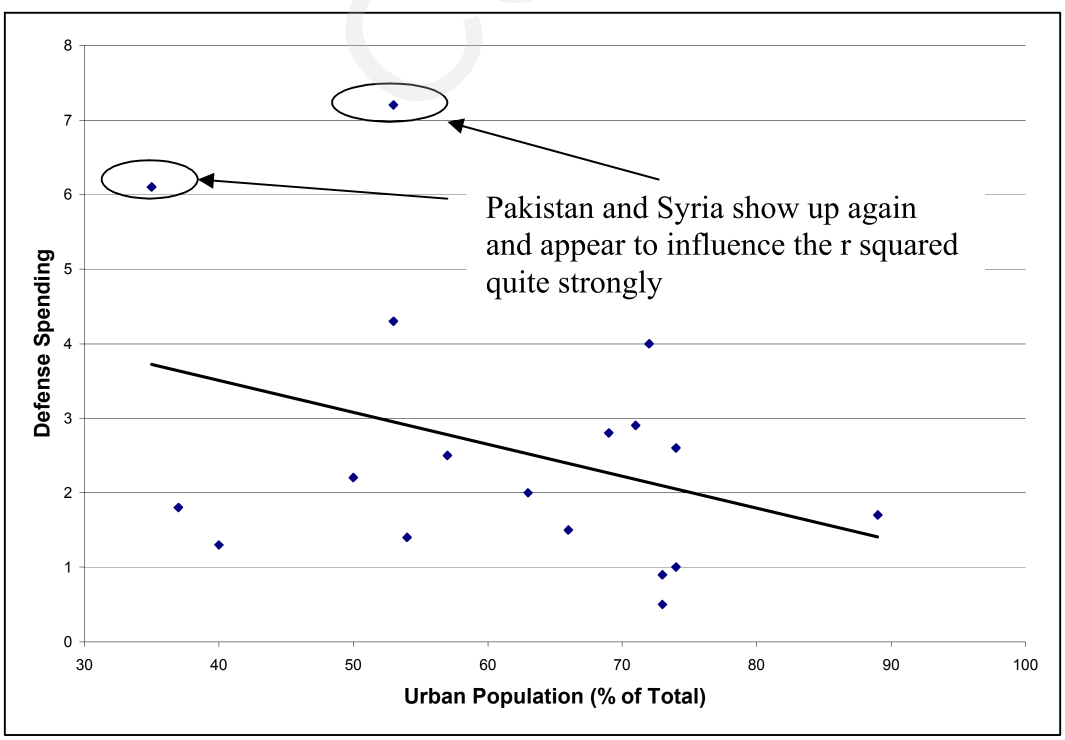


Here again, there are high leverage points which may change the intercept as well as the r squared. But they are roughly equidistant on either side of the line, and the line seems to emanate from the tilted watermelon of the other points outward between the points in question. The regression equation here is Y = 15.7 + 6.3 X with an r squared of .30. In this case, there is a positive relationship between IMR and defense spending. Interestingly, the regression equation suggests that the IMR will be no higher than 15.7 per one thousand when defense spending is reduced to zero. As well, IMR goes up by 6.3 for every extra percent of GNP increase in military spending.

Now it would be interesting to see how these variables work together in explaining the variation in infant mortality. In order to explore possible relationships and consider causal models, partial correlations and partial regressions will be used, controling the examination of the variables directly and explicitly.

As such, IMR is first compared with Urban population and Defense spending. We have the opportunity to look at IMR without the influences of urbanization. In this way, all countries can be regarded as having the same amount of people in cities in order to see other potential influences. It is important therefore to remove any possible influence urbanization may have on defense spending. As well, a visual examination of these variables in a scattergram shows that there is very little curviness.

Figure 3. Urban Population Against Defense Spending



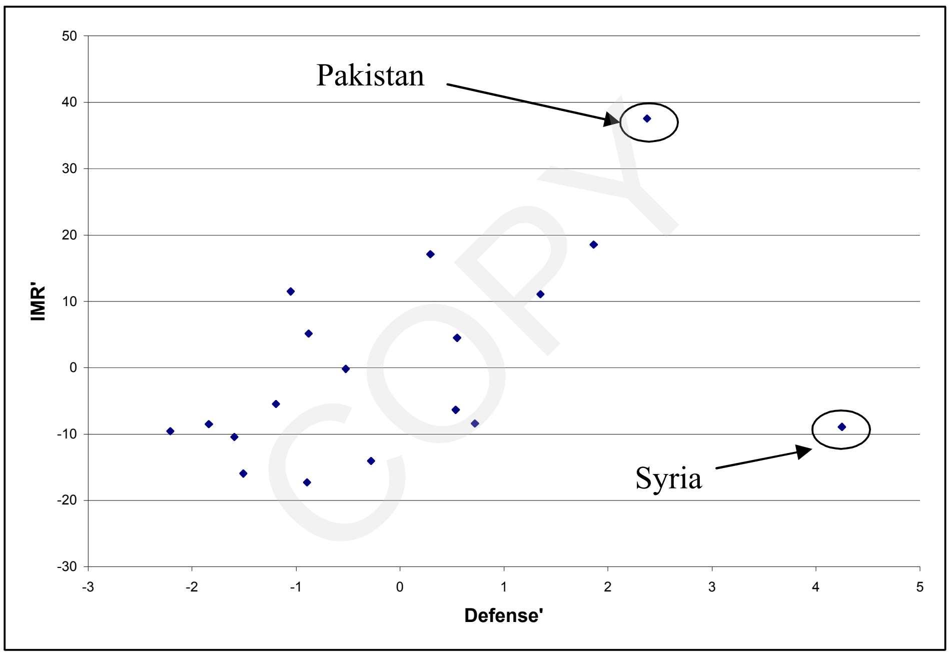
The slope indicates that there is a negative relationship between defense spending and urban population. Again Syria and Pakistan are high leverage points. A review of the population data set shows that there are many other examples of high defense spending and low urban population so in this exploration they will still be kept. Nonetheless, they are beginning to bring the whole equation into doubt. The regression equation for these variables is $Y = 5.2 - .04X$. According to this equation, defense spending would be 5.2 percent of GNP if there were no urban population. Also defense spending drops marginally, less then half a percent, for every unit increase in urban population. The r squared is very low though, where urban population explains only 13 percent of the variance in defense spending.

In order to see how defense spending may influence IMR, both IMR and defense without the influence of urbanity will be regressed together, $Y' = a + bX^{2'}$. The equation turns out to be $Y' = 0 + 3.89X^2$, and the r squared is .2 . So with the influence of urbanization removed, defense spending explains approximately twenty percent of the variation in IMR. The slope value could be rounded up of course, but isn't because it will prove to be a very valuable check when a multiple correlation is run later. It would be interesting to see the results of a regression of this kind using a larger sample size including countries in other income categories. It is also helpful to consider the definitions used and the particular circumstances of the countries considered. For example Urbanity may focus on countries with a few large cities rather then many small ones and Defense spending may be Offense Spending.

Now that some exploration has indicated that there may be a relationship between the variables, a test of significance can be utilized. First though, it is important that the assumptions of a regression be checked. This can be done (though roughly) by looking at the scattergram of the first order correlation of IMR and defense spending with urbanity held constant (Figure 4).

Though the shape of the points of the matched pairs is rather oval, there are two high leverage points. Indeed, Syria is a high leverage point which pulls down the r squared as it has high defense spending with a much lower than expected infant mortality. As for the bulk of the other points (excluding Pakistan), they do very much make up a fairly even oval shape with no other discernable features between high and low x and y values then to suggest that defense spending does explain some of the variance in IMR with urban held constant. But when above average defense spending includes the Syria and Pakistan, the relationship with IMR appears to be reduced, since the shape of these points together is largely round with a strong dip in towards Pakistan and back down to Syria; more of an inverse U shape. Nonetheless, the strong showing from defense spending suggests that how a country spends it's money is as important as how much money it has to spend. Defense spending may indicate reduced access to health care, as well as, increased deaths due to conflict.

Figure 4. Defense against IMR holding Urbanity constant



Since the points are fairly well behaved, a confirmatory test can be run and a calculated t score can be generated. The r (partial correlation) can be easily calculated by hand and is written out below. The r values can be found in the regression tables generated along with the scattergrams. But they can also be easily generated simply as partial correlation in most computer statistics packages.

Table 2. Partial Correlation

	IMR	Defense	%urban
IMR	1		
Defense	0.547545	1	
%urban	-0.70394	-0.3546	1

The formula is as follows:

Partial correlation = rx2y .x1 = $\underline{\text{r } x_2\text{y - (rx2x1)(ryx1)}}$ OR

Srt of $1\text{-}r^2x_2x_1$ sqrt of $1\text{-}r^2y\ x_1$

$\underline{.55 - (-.35)\ (.70)}$ = $\underline{.305}$ = .45

sqrt of $1\text{-}.35^2$ sqrt of $1\text{-}.70^2$ .669

And to get an F statistic (to see if there is a relationship between the independent variable and the dependant) the following formula can be used: (three degrees of freedom have been used since one of the variables was held constant - urbanity is involved but not seen here)

F1, n-3 = r2x2yx1 or

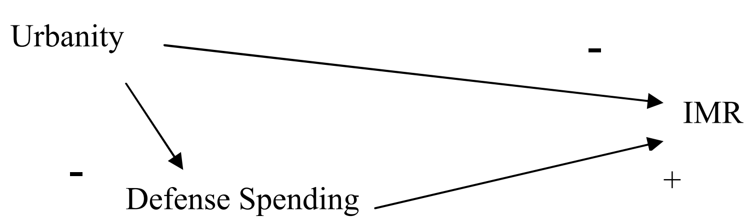
F1,15 = $\underline{.20}$ (15) = 3.77

1-.20

Since the critical value for F with 1 and 15 degrees of freedom is 4.5 allowing for five percent error, and the calculated F value is closer to 0 then the null hypothesis must be accepted that there is no linear effect of defense on infant mortality.

With the urbanity variable held constant, the influence of defense spending does drop somewhat, from 30% to 20%. Though the influence of defense spending is reduced by controlling for urbanity, it has not disappeared. The partial correlation suggests that an increase in the percentage of a countries population living in large cities contributes to a reduction in IMR as well as defense spending with all other things remaining equal. A diagram can be drawn which indicates the relationship derived from the partial correlation (Figure 5).

Figure 5.



Now that a causal model has been proposed suggesting various relationships between the variables, a test can be run which holds each independent variable constant while the other is tested against the dependant. In this case, a regression will be run between urbanity and IMR (without the influence of defense spending) and this relationship will be added to the previous linear equation. The equation reads:

$$IMR = 71 - .805868\ X1 + 3.893615\ X2$$

where b1 = the change in Y per unit change in X1 with X2 held constant
and b2 = the change in Y per unit change in X2 with X1 held constant

Therefore, when defense is held constant, urbanity has a negative relationship with IMR, but when Urbanity is held constant defense spending has a positive relationship with IMR. These values have little meaning otherwise since they are generated relative to scale and strength of each variable towards Y. But there is at least one real value which provides an important check with the first equation where a variable was directly and explicitly controlled. The value for the slope of X2 in the equation above should be the same as that for defense when urbanity was held constant...and it is.[1]

Instead of using values which have little meaning, it is helpful to generate a linear equation using values which at least relate to one another, which can be done easily by calculating standardized residuals. These values will provide a more meaningful estimate of the relationship between the independent variables (with the influence of each removed) and Y. Though the values are generated using a formula similar to a partial correlation, it is important to remember that they can range beyond -1 to +1. As well, since standardized residuals are used the value of Y where X is zero will have to be zero also. The multiple regression provides the following results:

$$Y = -.583097 + .340779$$

With t scores of -3.326 and 1.944

Here we see once again that urbanity has a negative relationship with IMR, while defense spending is held constant. On the other hand defense spending has a positive relationship with urbanity held constant. As well, it becomes more apparent that urbanity has a little less than twice the effect that defense spending has, but in the opposite direction. The amount of the variance explained by both of the variables is nearly sixty

---

[1] Indeed, these were not the same and an error was found by seeing a difference in these values.

percent. The high t-scores suggest that each of them has a linear effect on IMR. How well does the observation fit the predicted Y? Urbanity explained 50 percent while Defense spending explained 30 percent of the variation in IMR. Together they explain 60 percent of IMR. Since the independent variables are also related to each other, defense spending only adds ten percent to the explanation. To test to see if the two independent variable have a linear effect on IMR, a test for significance can be run. The null hypothesis would read 'that X1 and X2 have no linear effect on IMR'. The F statistic is calculated as follows:

$$F2,15 = \frac{.59705}{.40295} \left(\frac{15}{2}\right) = 11.1127$$

If the R^2 for this sample size gives a F statistic above the critical value of 3.7 for a confidence level of .05 (one tailed test) then the null hypothesis, that there is no linear effect on Y can be rejected. It is important when referring to the F statistic that the basic assumptions of normality were meet before generating a correlation coefficient. This can be very roughly checked by looking at the plot of a regression while one of the variables is controlled. In this way, the points on the graph can be checked for their overall shape as well as for any difference in slope. The shape should be oval and the slope should not vary if there is no difference between values of the variable which is held constant. In this case, since low versus high amounts of defense spending may have a very different effect on IMR, this variable will be controlled. It should be noted though, that either variable can be controlled to see if there are non-additive relationships. Plotting the points in this manner can prove beneficial when looking for interaction effects (where the values in one variable have a varied effect on values in another variable). In other words, does the linear effect of Urbanity on IMR change between low defense spending and high defense spending countries? First though a look at the shape shows that the values are within a roughly oval shape. In order to check for interaction effects a line is fit to each graph and compared (see Figure 6).

Figure 6. Urbanity (Y) against IMR (X) for low defense spending countries

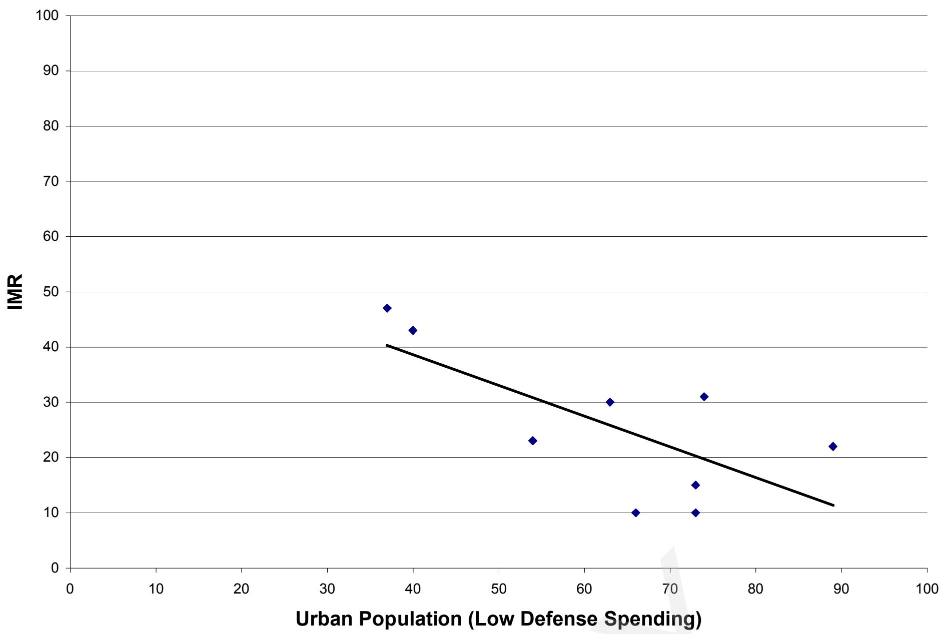
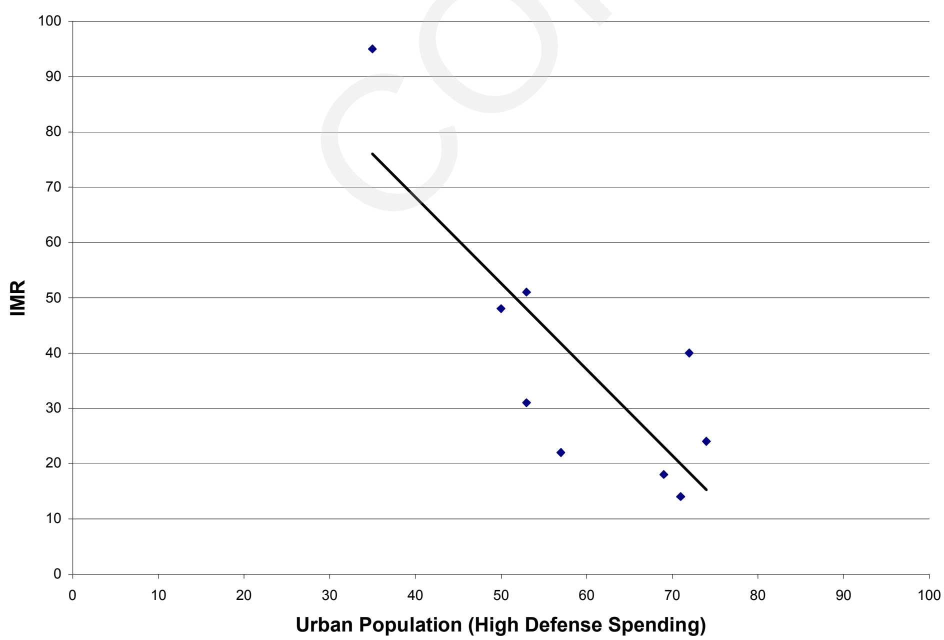


Figure 7. Urbanity (Y) against IMR (X) for high defense spending countries



The slope between Urban population and IMR is very different, indicating an interaction effect. Where defense spending is high, the slope is -1.558, almost three times greater than that for low spending at -0.557. To treat this non-additivity, a new variable can be created and added to the regression equation. The interaction can be made to fit to the comparison values. Therefore, to make this new variable, X1 and X2 can be multiplied together (see Appendix A). As such, this interaction term can act as a new predictor variable. The addition of this term results in the following equation when the standardized residuals are compared:

$$Y = -.1572\ X1 + 1.5048\ X2 - 1.1239\ X1X2 \quad \text{with a } R^2 \text{ of } .67544 \text{ and}$$

t scores of .555 2.302 -1.839

In this multiple regression equation, the slopes for X1 and X2 have changed considerably with urbanity losing most of its significance to the interaction term. The Urbanity t-score is also low. The X1X2 term has a negative relationship with IMR. The addition of the new term has increased the R^2 by almost eight percent to .67544 from .59705 suggesting that it has added significantly to the explanation of variance in IMR. Urbanity has dropped to one tenth the effect of that exerted by defense spending when it had been nearly twice as important previously. The beta values can be outside the -1 to +1 range and in this case both the interaction term and defense hold about as much weight, but in opposite directions. A closer look at the drop in urbanity shows that the p value (significant t) has risen to 60 % chance of making a type one error in saying that this variable effects the dependant when is does not. To see if the overall equation results have something to say about the variation in IMR, a test of significance can be run. This time the F cal must be for 3 and 18 - 3 - 1 degrees of freedom and therefore, above a F critical of 3.3. With an R^2 of .67544 (don't round this to .68 yet) the results are 9.7, well above the critical level and therefore the null hypothesis can be rejected; these variables do have an effect on IMR with very little chance of making a type one error (.001).

PART II. ADDING AGRICULTURE

Urbanity does help explain some of the variance in IMR and the claim here is that it addresses some of the structural components in health care. To broaden this inquiry another wealth creation and distribution indicator can be added, namely, the percentage of the labor force involved in agriculture. The problem with including this variable though is that it is likely the same measure as urban population only the exact opposite;

one may assume that as urban population rises, agrarian population drops. Therefore, a partial correlation matrix can be generated to examine these variables more closely. In particular, the potential for co-linearity between agriculture and urbanity can be checked.

Table 3. Partial Correlation co-efficient matrix

	IMR	Labor Ag	Ed expend	Defense	%urban
IMR	1.0000				
LaborAg	0.6823	1.0000			
Ed expend	-0.3015	-0.6309	1.0000		
Defense	0.5475	0.3752	-0.1656	1.0000	
%urban	-0.7039	-0.6674	0.3203	-0.3546	1.0000

The r for Labor involved in Agriculture and % urban is high at -.6674. Therefore, before Labor Ag is added to the multiple regression it will be regressed against IMR' with the linear effects of urbanity removed. The formula is as follows:

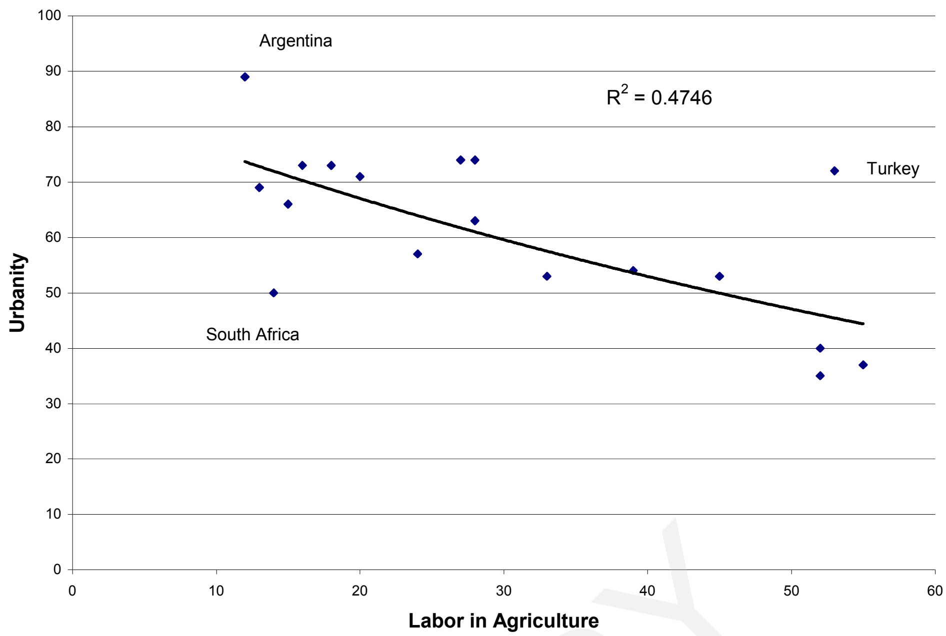
$$r\ \ ag\&imr\ .\ urb = \frac{.6823 - (-.6674)(-.7039)}{\text{sqrt of } 1 - -.6674^2 * \text{sqrt of } 1 - -.7039^2}$$

$$= \frac{.21251714}{.528938419} = .4017$$

The relationship drops nearly thirty percent from 0.6823 to 0.4017 when the influence of urbanity is removed from both IMR and agriculture. The F stat is easily calculated to be very near an r of 0 at .1924, therefore the null hypothesis that there is no linear effect can not be safely rejected. This first order relation between agriculture and IMR, though reduced is not removed. It is tempting to put agriculture in the multiple regression, but the standardized residuals of agriculture against IMR, and the standardized residuals of urbanity should be regressed. This approach will examine the similarity in that which is not explained by either variable. If there is a correlation then labor should not be included, though it can still be referred to. This results in a correlation r of just under .7 which is indeed very high.

Another check can be of the scattergram of Urbanity against Agriculture which shows the relationship graphically. Indeed there appears to be a fairly good correlation.

Figure 8. Scattergram of Agricultural labor force against Urban population



But the percentage of people involved in agriculture only explains 44% of the variation the percentage of people living in cities. The notable exceptions are Turkey with ~50% involved in Agriculture and over 70% living in cities while in South Africa 50% live in cities and only 14% are involved in agriculture. This may point to a highly resource based economy, including mining, and fewer people living in cities of over one million. Yet, if we are looking at the difference between resource based and resource manipulation economies, we should consider a variable which more accurately separates these two types of economies. Nonetheless, the labor force involved in agriculture includes those people working in agriculture, hunting, forestry, and fishing and therefore is a very good indicator of resource based economies with of course the notable exception of mining. Therefore for our purposes, it is a good indicator of resource based economies.

One other approach which may prove helpful when comparing urban population and agrarianism would be to refer to the interaction effect found between urbanity and IMR for low and high levels of defense spending. In this case, agricultural labor force can be regressed against IMR for both low and high levels of defense spending. The slopes found are much closer together: .725 for low, and 1.05 for high military expenditures. Therefore, the percentage of the labor force involved in agriculture has similar effects on IMR for both low and high defense spending, whereas, urbanity had very different effects.

Though there does appear to be some co-linearity between urban population and labor force involved in agriculture, these two variables can be put together in a multiple regression to see how they might influence infant mortality rates. As a structurally oriented indicator, this variable looks less at infrastructure and distribution and more at availability of goods and services which can be put towards general welfare and therefore possibly a reduction in infant mortality.

When agriculture is added to the multiple regression (which already includes urbanity, defense spending, and the interaction term) the R2 rises again, this time by nine percent to .76588 from .67544 (see Appendix D). Agriculture receives a t score of 2.241 well above urbanity which has changed from a low negative score of -.555 to a positive .711 (this switch may mean that Agriculture should not be included). The high t score suggests that Agriculture is adding to the explanation of variance in IMR. Urbanity has the lowest beta still since the interaction term seems to have taken most of its effect. The other signs remain the same. An examination of the residuals shows that most have moved closer to the expected though a few have moved very slightly away. Only South Africa has jumped significantly outward, from under one standard deviation to two. South Africa does have a low urban population and low agrarian sector. As well, more effort has been going into education recently. Perhaps this will be picked up with the addition of education expenditures.

PART III. ADDING EDUCATION

As the structural component is broadened, so it should be that the choices for the allocation of wealth should also be considered more closely. As we have seen already, an increase in military expenditures does say something about infant mortality. What other choices might a country have for the allocation of wealth? An obvious one is the percent of the GNP put towards education. Here again though, are not defense spending and education spending the same only opposite? Either money is spent on education and books or on the military and guns. It is perhaps useful to remember that at the roots of Canadian history existed this very tension. The English and Dutch traders sold guns to the Iroquois (400 shot guns at well below market value) while the French, particularly the Jesuits, would not allow gun or alcohol sales but emphasized education.

In order to check whether resource based economies and education expenditure affect IMR, another exploration of the data is conducted. Again, the new variable can be transformed if necessary and matched pairs can be visually inspected in scattergrams. Once this is done, a regression can be performed and the r^2 calculated. First, the education expenditure data is set against IMR and does not need transformation (this was squared and logged but these did not add to the interpretability) but the correlation of determination r2 is very small at .1139.

Though this is a small relationship another approach can be taken in order to examine the possibility of a dichotomous relationship with defense spending. When used in a scattergram with defense expenditures, an interesting picture emerges (Figure 9). The regression results show that there is very little relationship between the two variables. In fact, only 3 and a quarter percent was accounted for. The slope is very low and it would appear that there is no relationship (the F stat is less then .5). To help draw out some of the information in this graph median lines (dashed) have been added. Only one country, Morocco, has above median military and education expenditure, though it is very close to the middle of the graph. The majority of the points lay below the median for military expenditures; only four are above. There are equal numbers of points above and below the education median for the countries with low military expenditures; seven and seven. Interestingly, Indonesia and Guatemala have a very high percentage of their labor force involved in agriculture. South Africa and the Ukraine on the other hand have low percentages at 14 an 20 respectively. Therefore, there may be fewer resources available (assuming agrarianism is an indicator of wealth) for any type of expenditure for the countries in the low military and education spending quadrant. A quick look at agriculture in the high military and low education spending quadrant reveals that have a large agricultural sector does not necessarily reduce military expenditures. Turkey has 53%, Pakistan 52%, and Syria 33% of labor involved in agriculture. It would appear that countries with high percentages of labor involved in agricultural do have the opportunity for some expenditures.

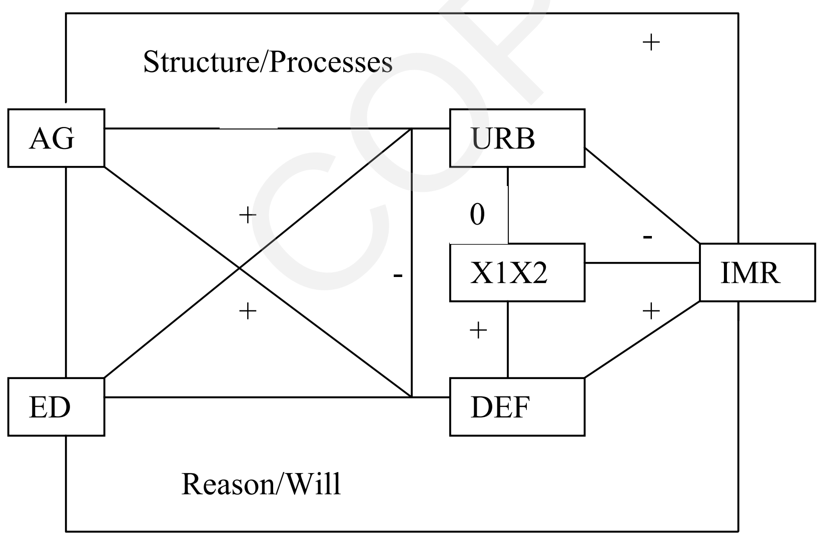
When education expenditures is added to the multiple regression, the residual statistics show that the variance in standard deviation has been reduced to between

-1.8403 to 1.6308 from the spread in the previous equation of -1.6101 to 2.0936 (see Appendix E). But the Standard Deviation itself grows slightly and the R2 only grows by a little over one half percent from .76588 to .77129. In fact, the adjusted R which accounts for the growing number of variables (but does not account for how small education expenditures effects IMR) drops to .676 from .69384. Perhaps the influence of education is spread out over time and space and therefore less of a direct measure then is defense spending. Indeed, South Africa does not move in the casewise plot of the standardized residuals. Nonetheless, other variables which measure increased knowledge and understanding are important to consider. The final equation in this exercise reads:

Y = .222763 X1 (urb) + 1.71924 X2 (def) - 1.40146 X1X2 (interaction term)
+ .497033 X4 (ag) + .096725 X5 (education) with an R^2 of .77

PART VI. PROPOSED CAUSAL MODEL

In order to help interpret the results refinements can be made to the orginal model.



Several changes have been made but the original variables have maintained their relationships. The interaction term has been added as well as agriculture and education. The relationships between structure/processes and reason/will exemplify some of the countries examined. A country with high agricultural labor may emphasis education or military activity. For the most part, agriculture has a slight negative relationship with education. Other highly agricultural countries gravitate to military expenditures. There is a very slight negative relationship between education and defense spending, though it is almost non existant. Urbanity and agriculture are strongly negative to the point of legitimate concerns regarding colinearity. Urbanity and defense have an interaction effect creating the multiplicitive term X1X2 which has a large negative relationship with IMR in the multiple regression, but when placed alone against IMR, it has an r of nearly 0. This is true for it's relation with Ag and ED as well. The interaction term also has no relation with urbanity but has a good fit with Defense spending ($R2=.81$). Defense has a positive relationship with IMR as does agriculture, though much milder. Education and urbanity have the opposite effect, again much weaker. In the multiple regression equation, the interaction term has a strong negative relationship with IMR, while defense spending has a strong positive effect. The interaction term picks up most of the urbanity variance in relation to IMR. The addition of agriculture reduces the influence of defense and X1X2 slightly. The addition of education drops the adjusted R.

PART V. REGIONAL CATEGORIES - DUMMY VARIABLES

As the variables were added to the equation, various regions seemed to have similarities to those countries within and differences with those countries without. For example, Middle Eastern countries seemed to have high military expenditures while Eastern European countries had lower infant mortality rates.

Table 4. Dummy Variables for Latin America (LA) Eastern Europe (EE) and the Middle East and Northern Africa and a reference of Other

Country	IMR	Defense	LA	EE	ME/Naf
Argentina	22	1.7	1	0	0
Bulgaria	18	2.8	0	1	0
Colombia	24	2.6	1	0	0
Guatemala	43	1.3	1	0	0
Hungary	10	1.5	0	1	0
Indonesia	47	1.8	0	0	0
Latvia	15	0.9	0	1	0
Lithuania	10	0.5	0	1	0
Mexico	31	1	1	0	0
Morocco	51	4.3	0	0	1
Pakistan	95	6.1	0	0	1
Paraguay	23	1.4	1	0	0
Romania	22	2.5	0	1	0
South Africa	48	2.2	0	0	0
Syria	31	7.2	0	0	1

Tunisia	30	2	0	0	1
Turkey	40	4	0	0	1
Ukraine	14	2.9	0	1	0

The regions can be categorized using dummy variables and then regressed against IMR first, and then defense. The reference category uses South Africa and Indonesia since neither are close to other countries and both are nearer each other though separated by the Indian ocean (making this category questionable); nonetheless as test is performed to examine it's possible usefulness. The first multiple regression results in the following equation when referring to the beta scores (see Appendix F):

IMR = -.774069 Eastern Europe -.425527 Latin America +.042778 Middle EastN. Africa

t scores of -2.604, -1.471, .148

and a R of .73228 and a R2 of .53624

The Latin American score lays between 1 and 2 and has a high p value (.1635) and therefore is a border line for significance in effect on IMR. The Middle East/N.Africa has a wide range of results, from Pakistan with high IMR and Syria with relatively low, reducing the t for this region. Nonetheless, the categories seem to offer some information.

Defense spending seems to be higher in countries in the Middle East/ N. Africa region. The beta value for this region is .7, while Latin America receives -.103 and Eastern Europe has -.04 (see Appendix F). A view of regional information may help further the inquiry into variables influencing IMR.

Conclusion

This report has examined infant mortality rates recorded at the national level using world development indicators from the World Bank (1999). An attempt was made to focus on structural and decison making elements regarding the general welfare of differing states. Of particular interest is the direct influence of urbanity and defense spending and the use of the interaction term on IMR in relation to the less tangible influence of knowledge and techniques. Further studies may explore other indicators which may account for these elements. Other considerations may look more closely at why the interaction term has proven so strong. As well, urban population and types of cities should be refined and examined along with rates of urbanization. Perhaps administrative systems can be included with a view to unions and co-operatives. Also, implicit in studies of infant mortality is the emphasis on development. Perhaps societies which are not easily measured at the national level should also be considered. These would include those engaged in subsistance economies.

The proposed model also must be examined more thoroughly. Yet, as a stepping stone towards better understanding the elements at play in considering infant mortality the overall framework has provided interesting results.