Economics Department

Economics Working Papers

The University of Auckland *Year* 1999

On A New Measure of Human Capital and Its Impact on Gross Domestic Product

Debasis Bandyopadhyay* P Lahiri†

Feng Yu‡

*University of Auckland, d.bandyopadhyay@auckland.ac.nz

†

‡

This paper is posted at ResearchSpace@Auckland.

http://researchspace.auckland.ac.nz/ecwp/196

ON A NEW MEASURE OF HUMAN CAPITAL AND ITS IMPACT ON GROSS DOMESTIC PRODUCT

Debasis Bandyopadhyay P. Lahiri and Feng Yu[1]

University of Auckland University of Nebraska-Lincoln

Abstract

The general goal of this paper is first to develop an operationally simple measure of human capital using the relative frequency histogram of the highest educational attainment and then to analyze the cross-country variations of the proposed measure. Visual inspection and the matrix of rank correlation coefficients show that relative frequency distributions of the highest educational attainment are similar for countries with similar Gross Domestic Products (GDP) level, but they are very different for countries whose GDP levels are quite different. Guided by

---

[1] Debasis Bandyopadhyay is a Lecturer, in the Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand. P. Lahiri is a Professor and Feng Yu is a Graduate Student, both in the Department of Mathematics and Statistics, University of Nebraska-Lincoln, Lincoln, Nebraska 68588-0323, U.S.A. The research of the first two authors are supported in part by the Marsden Fund (Grant #96-UOA-SOC-0018) of the Royal Society of New Zealand. The research of the second and third authors are also supported in part by the United States National Science Foundation SBR-9705574 and a Graduate Student Fellowship of the Department of Mathematics and Statistics, University of Nebraska - Lincoln, respectively.

intuition, we define a simple descriptive statistic, EER measured by the relative proportion of labor force with education beyond the secondary level to those with no formal education. This simple statistic turns out to extract the most essential information contained in the relative frequency histogram of the highest educational attainment to forecast future economic growth of a country. Consequently, we propose this statistic EER as a new measure of human capital. Non-parametric tests show that both the means and variances of the distribution of log (EER) for the high GDP countries are significantly higher than the corresponding means and variances for the low GDP countries. A chi-square test reveals that for the two groups of low and high GDP countries, the distributions of EER can be characterized by a unified class of gamma distributions with the *same* shape parameter but with very different scale parameters. Based on the data created by Barro and Lee (1993), we note that our new measure of human capital (*i.e.*, EER) *alone* can explain cross-country variations in per capita GDP much better than the other growth models such as Solow (1956) and Mankiw, Romer, and Weil (1992). Those models include population growth rate and investment rate as covariates and the latter model use an additional covariate SCHOOL measured by the average secondary school enrollment rate or in addition to those two covariates. We explain the better performance of our model by noting that the statistic EER is significantly negatively correlated with population growth rate and positively correlated with investment rate and SCHOOL.

KEY WORDS: Chi-square test; Highest educational attainment; Investment rate; Kruskal-Wallis test; Population growth rate, Relative frequency distribution.

1. INTRODUCTION

Economists are still trying to understand why countries experience sharp divergence in Gross Domestic Products (GDP) and growth rates with a resulting experience of dramatically different standards of living. The recent literature on income distribution and growth based on human capital theory (e.g., Galor and Zeira, 1993) add a new dimension to macroeconomic dynamics by making the distribution of human capital a fundamental determinant of the macroeconomic aggregates. Concurrently, there have been renewed interests in the controversies surrounding alternative paradigms of growth (e.g., Solow, 1994) facing the fact of cross-country growth disparities (e.g., Quah, 1996) and the fact of a growth-inequality relationship (e.g., Chang, 1994).

The existing literature demands but does not provide systematic or stylized observations on an international comparison of distribution of human capital. Robert M. Solow, a Nobel Laureate in economics, encouraged future researchers to fill that vacuum in the literature in the 1992 George Seltzer Distinguished Lecture Series at the University of Minnesota entitled *Growth with Equity with Investment in Human Capital.* The main objective of this paper is to address this important issue.

Several attempts have been made to compile data on human capital distribution. Mincer (1991) and Krueger (1993) are examples of work related to gathering data on human capital distribution to be used as evidence on theoretical models. They are, however, mainly concerned with the US data.

Barro and Lee (1993) compiled data on highest educational attainment, one of the measures of human capital, among adult population (25 and older) for a broad cross section of countries. The data was given over five-year intervals from 1960 to 1985. The data provides the fractions of population belonging to seven categories: no formal education, incomplete primary, complete primary, first cycle of secondary, second cycle of secondary, incomplete higher, and complete higher. They created the data using census information on school attainment for adult population which were obtained from UNESCO publication and other sources. School enrollment ratios were used to fill in the missing observations. See Barro and Lee (1997) for an update of their data set for the population aged 15 and over.

Nehru, Swanson and Dubey (1995) created a series of estimates of stock of education in 85 countries over 28 years (1960-87) for the population between the ages 15 and 64. They used enrollment data from UNESCO sources and corrected their estimates for grade repetition among school-goers and country specific dropout rates for primary and secondary students.

Prior to Barro and Lee (1993, 1997) and Nehru, Swanson and Dubey (1995), there were several studies on the international comparisons of various measures of human capital. A few notable papers in this area of research include Psacharopoulos and Arriagada (1992), Lau, Jamison and Louat (1991) and Kaneko (1986) among others.

The average years of schooling, a conventional measure of human capital, has been increasing in almost all countries since the 60s (see, e.g., Barro and Lee, 1997), but only a small group of countries has been enjoying more than 3% annual average growth rate between 1965-90. The following quotation (see the web page maintained by the National Bureau of Economic Research, Inc., http://www.nber.org/programs/efg/efg.html) reiterate the above fact.

"The first meeting of the newly formed "Growth Group" focused on the accumulation and development of human capital, finding some surprisingly paradoxical results and developing exciting avenues for future research. Lant Pritchett of the World Bank presented cross-sectional evidence that the growth of human capital, as measured by years of education, is completely uncorrelated with the growth of output.

This result is surprisingly robust to the use of different data sets, as confirmed by conference participant Jong-Wha Lee, NBER and Korea University who, together with Robert J. Barro, NBER and Harvard University, has developed a broad international database on education.

The conventional measure of human capital, the years that students devote to education, is extraordinarily crude, providing inadequate assessment of the value and growth of human capital...."

The above discussions encourage us to explore a new measure of human capital based on the relative frequency histogram of the highest educational attainment. While several researchers have been successful in creating meaningful databases concerning human capital based on highest educational attainment, systematic or stylized statistical analysis on these databases is not available. It is difficult to characterize the distributions of highest educational attainment for different countries by standard discrete probability distributions, (see, e.g., Bandyopadhyay, 1999). This is because of the fact that these probability distributions are too smooth to account for the fact that schooling is more likely to be terminated at the completion of a category of schooling (e.g., primary, secondary, and higher education) than during a category.

In this paper, the research is conducted in two phases. In the first phase, we attempt to identify a single key measure of the relative frequency distribution of the highest educational attainment in the labor force. In the second phase, we investigate the impact of this key measure on GDP.

Throughout the paper, we analyze the data compiled by Barro and Lee (1993). In section 2, we introduce a summary measure EER mentioned in the Abstract and carry out a detailed statistical analysis based on that statistic. Our investigation suggests the EER is an important feature of the relative frequency distribution of the highest educational attainment in the labor force. Our investigation reveals the surprising fact that it is possible to describe the EER distributions for countries with low GDP and high GDP by two different gamma distributions with the same shape parameter (0.5) but very different scale parameters. The scale parameter of the gamma distribution for the high GDP countries is much larger than the corresponding scale parameter of the gamma distribution for the low GDP countries.

In section 3, we explored the relationship between GDP and EER. We discover that there is a very strong relationship between log (GDP) and log (EER). In this connection, we mention two important papers. First paper was by Solow (1956) who used investment and population growth rate in order to explain GDP. Later on Mankiw, Romer, and Weil (1992) added a third covariate SCHOOL mentioned in the abstract to the Solow's model. The predictive power of a *single* covariate *EER* defined by the distribution of highest educational attainment of the labor force appears to be quite noteworthy when compared to these well-known models due to Solow (1956) and Mankiw, Romer, and Weil (1992). In particular, the adjusted R^2 value for the model with a single covariate EER is 70 % compared to 53% for the Solow's model with two covariates and 70% for the Mankiw-Romer-Weil model with three covariates. The adjusted R^2 value for our model can be further improved to 86% by considering different intercepts for the low GDP and high GDP countries. More importantly, our proposed model is very simple and the slope of our proposed regression is significantly positive with the implication that one way to improve on GDP is to increase the relative proportion of educated adults in the labor force. Our measure is also significantly correlated with the covariates used in the Solow and Mankiw-Romer-Weil models.

2. A NEW MEASURE OF HUMAN CAPITAL

We begin by plotting relative frequency histograms of highest educational attainment for countries with low and high GDP's. There is a striking similarity among these histograms within each of these two groups. However, this histogram for a low GDP (see Figure 1) country is very different from that of a high GDP country (see Figures 2). To confirm this conclusion derived from visual inspection, we calculated a matrix of rank correlation coefficients for 6 low GDP countries (Bangladesh, Benin, Congo, India, Indonesia and Sudan) and 6 high GDP countries (Belgium, Canada, Germany, Japan, New Zealand and the United States). The results are given in Table 1. All the rank correlation coefficients in the low GDP countries are positive and are usually very high (mostly 1, but the correlation coefficient among all 5 countries with Indonesia is 0.8, which is also very high). Similarly, all the corresponding rank correlation coefficients in the high GDP countries are positive and most of them are very high numbers (with the exception of Belgium whose correlation is moderate (0.4) with USA and New Zealand). The corresponding rank correlation coefficients for countries from the two different groups are mostly negative and often high negative numbers.

Figure 1 guides us toward an intuitive new measure of human capital, EER. It is the ratio of the number of all individuals in the labour force with at least secondary level education to those without any formal education. The ratio EER captures the essential information contained in the relative frequency distributions of the highest educational attainment in the labour force relevant for explaining future economic growth. Using a well-known statistical package (SAS), we carry out a test for normality of the distribution of log (EER). The normality assumption is rejected with very low *p-value* (0.0001) for both the low GDP (40 countries) and high GDP (40 countries) groups. Note that we put 7 countries between the low and high GDP countries in order to avoid ambiguity in defining low and high GDP countries. When all the 87 countries are combined, the hypothesis of normality is not rejected (*p-value* = 0.6). The Q-Q plots are displayed in Figure 3. This apparent paradox can be explained by the fact that both low and high values of log (EER) are observed when all the countries are pooled and thereby achieving symmetry in the distribution. The means for the log (EER) for low and high GDP groups are -1.76 and +1.72 and an application of Kruskal-Wallis test reveals that the mean for the high GDP group is significantly higher than that in the low GDP group at 5% level of significance. The variances of these two groups are 1.96 and 3.16 respectively and the variance of the high GDP group is significantly higher than that of the low GDP group.

From the above discussions, it is clear that the distributions of log (EER) are different for the two groups of countries (i.e., high GDP and low GDP countries). Is it possible to smooth the distribution of EER by a theoretical probability distribution? If so, what kind of distribution can characterize the distribution of EER? Since EER ranges from zero to infinity, gamma distribution seems to be a natural choice. Also, gamma distribution can take a variety of shapes. We applied the maximum likelihood method to estimate the shape and scale parameters of the two gamma distributions for the two groups. For the low GDP group, estimates of shape and scale parameters are 0.52 and 1.14 respectively. For the high GDP group the corresponding estimates are 0.52 and 36.78 respectively. In order to avoid complexity associated with the Pearson's chi-square goodness-of-fit test with estimated parameters, we decide to test if a gamma distribution with known shape parameter 0.5 and scale parameter 1 fits the EER's for the low GDP countries. Similarly, we test if a gamma distribution with known shape parameter 0.5 and scale parameter 37 will fit EER's for the high GDP countries. The Pearson's chi-square statistics for these two cases turn out to be 8.0 and 6.8 respectively implying that there is no reason why we should reject our research hypothesis (this is true at 1% level of significance). The empirical cumulative distribution functions (cdfs) and the true cdfs for the two groups are plotted in Figure 4. The probability density functions (pdf) for the gamma distributions are also plotted in Figure 4.

3. RELATIONSHIP BETWEEN EER AND GDP

In this section, we will explore if EER *alone* can explain GDP. TABLE 2 compares regression analyses (in the log scale) of our models (i.e., Model 1 and Model 2) with the two important models due to Solow (1956) and Mankiw, Romer, and Weil (1992). Note that the Solow's model uses investment and population growth rate as covariates. Mankiw, Romer, and Weil (1992) added SCHOOL measured by the average secondary school enrollment rate to the Solow's model. First of all, it is clear from Table 2 that education data makes a difference. Based on the data for 87 countries given in the data set created by Barro and lee (1993), the Mankiw-Romer-Weil model yields an adjusted R^2 of 70% compared to 53% for the Solow's model. It is interesting to note that our measure EER *alone* can produce an adjusted R^2 of 70% (see Model 1). Model 1 performs equally well as the Mankiw-Romer-Weil model. But, the simplistic nature of our model is very appealing and has the implication of improving GDP by simply investing on basic education. The above results can be explained by the fact that in the log scale EER is significantly negatively correlated with the population growth rate (correlation=-0.57, *p-value*= 0.0001) and significantly positively correlated with the investment rate (correlation = 0.68, *p-value* = 0.0001) used in the Solow model as covariates. Our measure is also significantly positively correlated (correlation = 0.81, *p-value* = 0.0001) with SCHOOL, which was the additional covariate used by Mankiw, Romer, and Weil (1992).

Figure 5 provides the scatter plots along with the fitted regressions of our model. We have the combined plot as well as plots for low GDP and high GDP groups. The plots for the two groups are suggestive of similar slope but different intercepts. A statistical test shows that the intercepts are significantly different but the slopes are insignificant (at 1% level of significance). This encourages us to consider a regression model (Model 2) with different intercepts but the same slope for the two groups of countries. The results are given in Figure 6. The adjusted R^2 improves to 86%.

4. CONCLUSIONS

The general goal of this paper is first to develop an operational measure of human capital using the database created by Barro and Lee (1993) and cross-country variations of the proposed measure. Our goal has been to find the minimum set of human capital related variables that could provide a significant explanation of cross-country GDP. Our research suggests that a single measure (EER) of the relative frequency distribution of the highest educational attainment as a possible powerful measure of human capital.

It will be interesting to examine the robustness of the different claims made in this paper with respect to other data sets available in the literature. The relationship between our proposed measure (EER) of human capital with various social development indicators (e.g., life-expectancy rate, fertility rate, education of women, teacher-pupil ratio, etc.) will be an interesting future research topic.

In sum, the objectives of the our research include the general goal of filling up the vacuum in the literature that exists with regard to the availability of operationally meaningful data on human capital and quantitative analysis of possible links between the distribution of human capital economic growth. The results will, therefore, improve the understanding in those areas and complement recent advances in theoretical models that establish links between economic growth and the distribution of human capital. The output of this research may also help us to determine future trends in economic growth in different countries.

REFERENCES

Bandyopadhyay, D. (1999) "On the Characterization of Human Capital Distribution - An Empirical Study," *Journal of Statistical Planning and Inference*, (forthcoming in 2000).

Barro, R. J. and J.-W. Lee, (1993) "International comparisons of Educational Attainment," Journal of Monetary Economics 32, 363-394.

Barro, R. J. and J.-W. Lee, (1997) "International Measures of Schooling Years and Schooling Quality," American Economic Review 86(2), 218-223.

Chang, R., 1994, Income Inequality and Economic Growth: Evidence and Recent Theories, Federal Reserve Bank of Atlanta Economic Review 79, 1-10.

Galor, O. and J. Zeira, (1993), "Income Distribution and Macroeconomics," *Review of Economic Studies*, 60, 35-52.

Kaneko, M., (1986), *The Educational Composition of the World's Population: A database.* Washington, DC, The World Bank, Education and Training Department. Report No. EDT 29.

Krueger, A. B., (1993), "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989," *Quarterly Journal of Economics*, 108, 33-60.

Lau, L. J. et-al, (1993) "Education and Economic Growth: Some Cross-Sectional Evidence from Brazil," Journal of Development Economics 41(1), 45-70.

Mankiw, N. G., D. Romer, and D. N. Weil, (1992) "A Contribution to the Empirics of Economic Growth," Quarterly Journal of Economics 107, 407-437.

Mincer, J., (1991), "Human Capital, Technology, and the Wage Structure: What Do Time Series Show?" NBER Working Paper No. 3581.

Nehru, V.; Swanson, E. and A. Dubey, (1995), "A New Database on Human Capital Stock in Developing and Industrial Countries: Sources, Methodology, and Results," *Journal of Development Economics*. Vol 46 (2), 379-401.

Psacharopoulos, G. and M. Arriagada, (1986) "The Educational Composition of the Labour Force: An international comparison," International Labour review 125(5).

Quah, D., 1996, Convergence Empirics Across Economies with (Some) Capital Mobility," Journal of Economic Growth (March), 95-123.

Solow, R. M., (1956) "A Contribution to the theory of Economic Growth," Quarterly Journal Economics LXX, 65-94.

Solow, Robert M., 1994, Perspectives On Growth Theory, Journal of Economic Perspectives 8, 45-54.

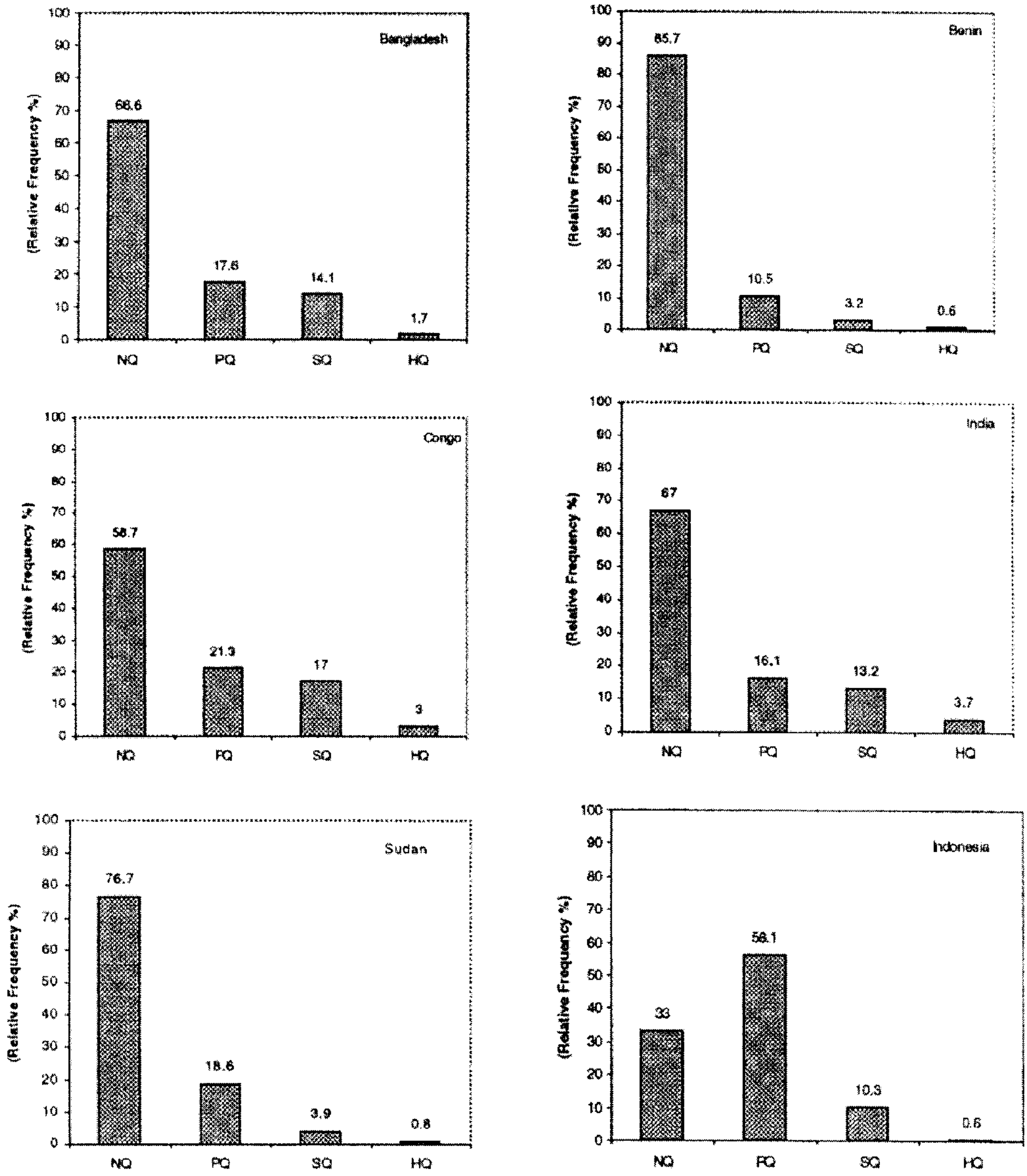


Figure 1. Histogram of highest educational attainment for low GDP countries

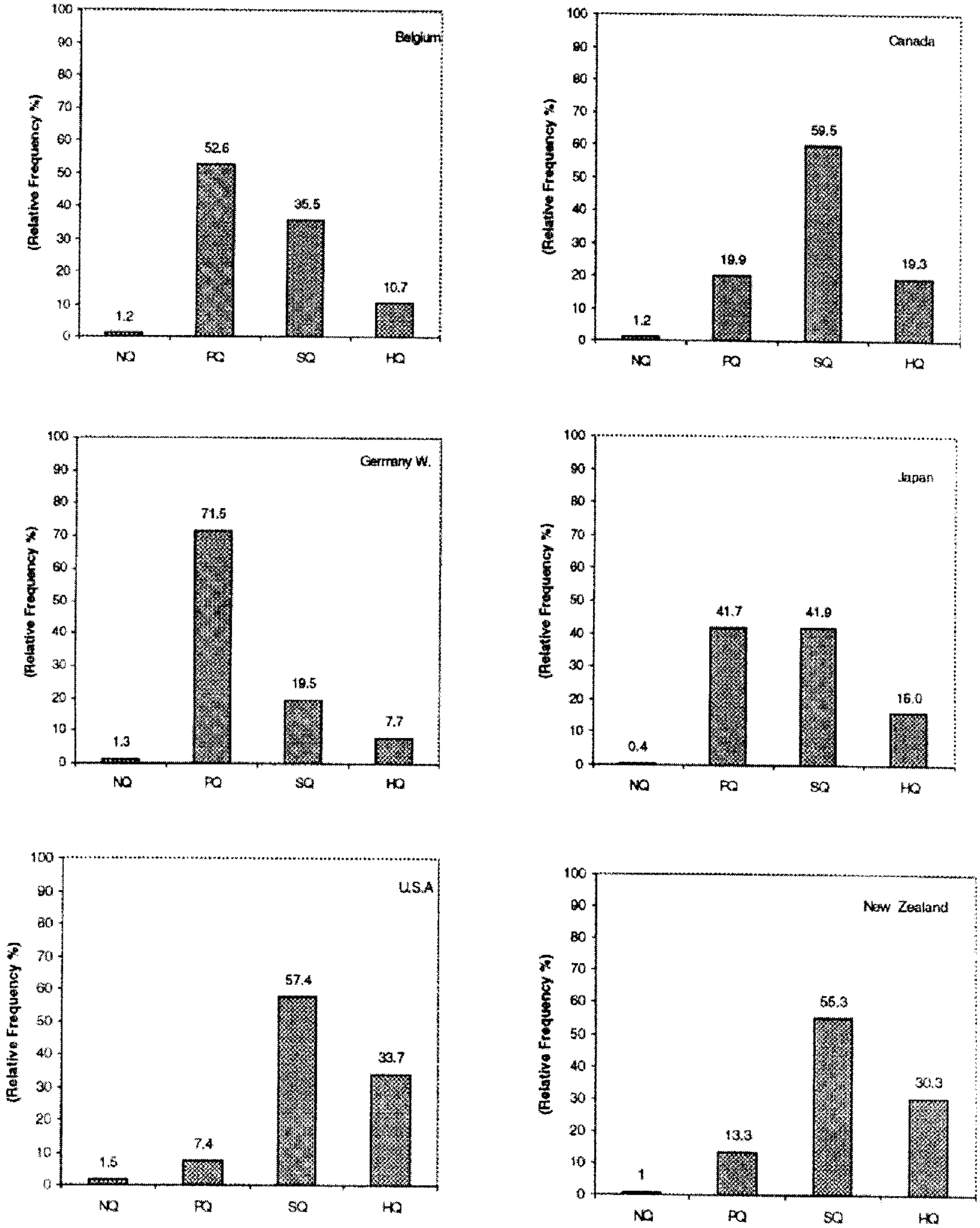


Figure 2. Histogram of highest educational attainment for high GDP countries

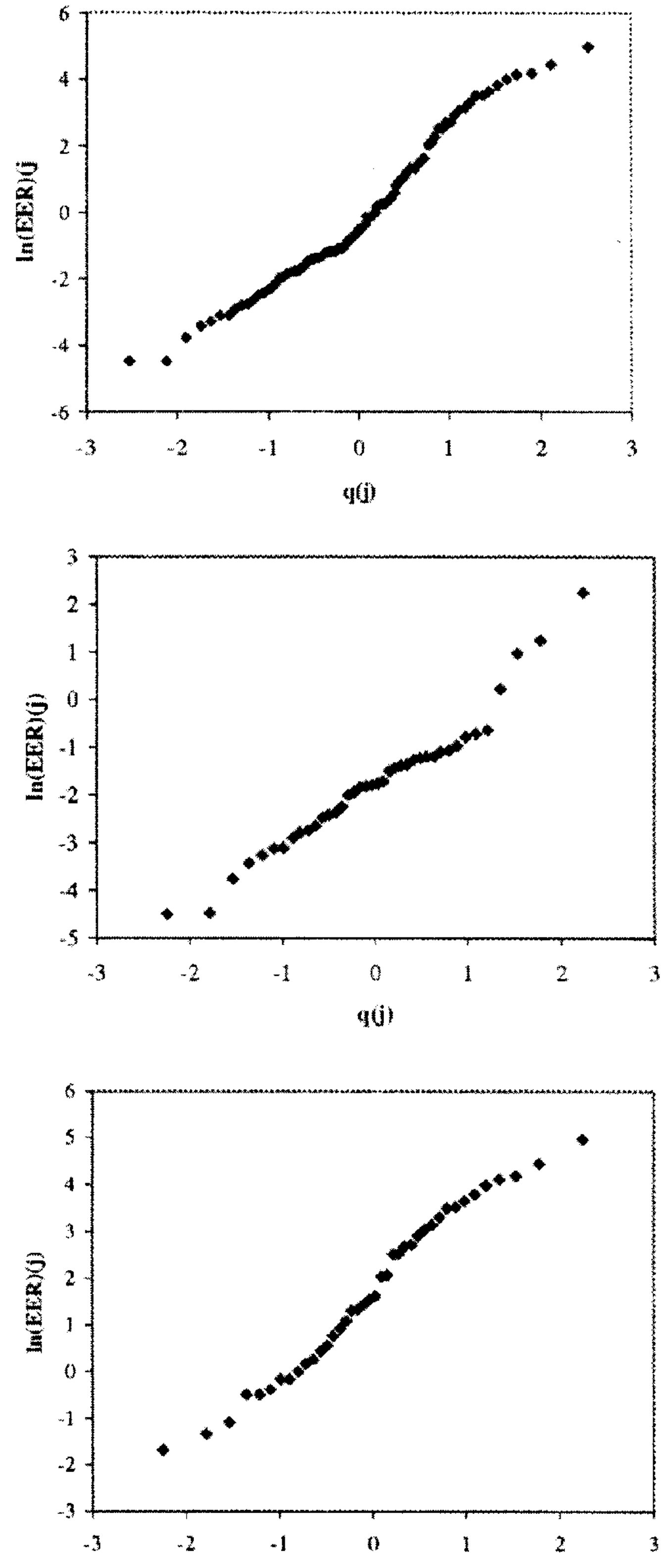


Figure 3. Q-Q plots of ln(EER) (Top: 87 countries; Middle: Low GDP countries; Bottom: High GDP countries)

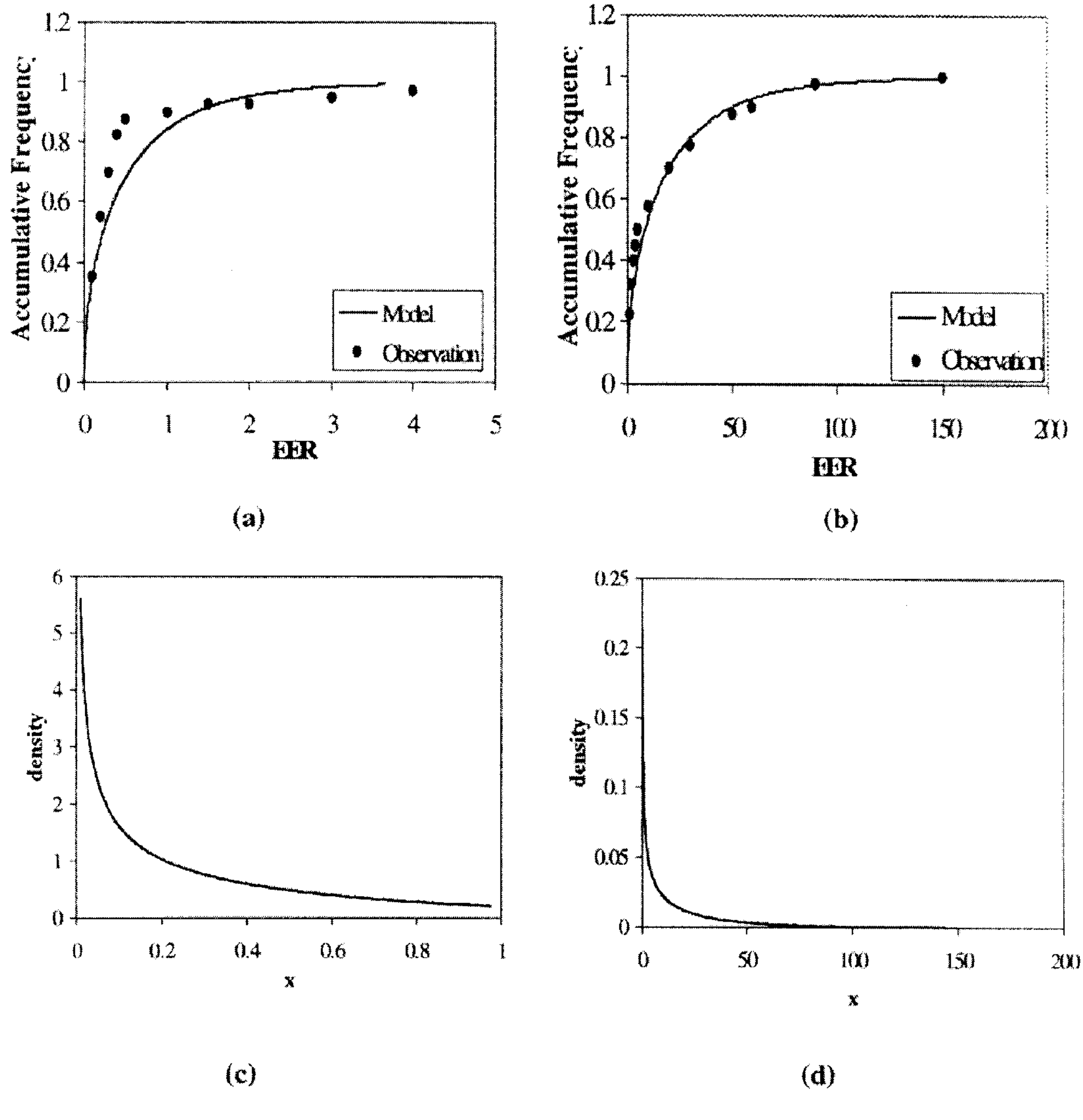


Figure 4. Empirical and hypothetical cdf's of highest educational attainment for low GDP (a) and high GDP (b) countries; pdf's of Gamma distributions with α=0.5, β=1.0 (c) and α=0.5, β=37 (d)

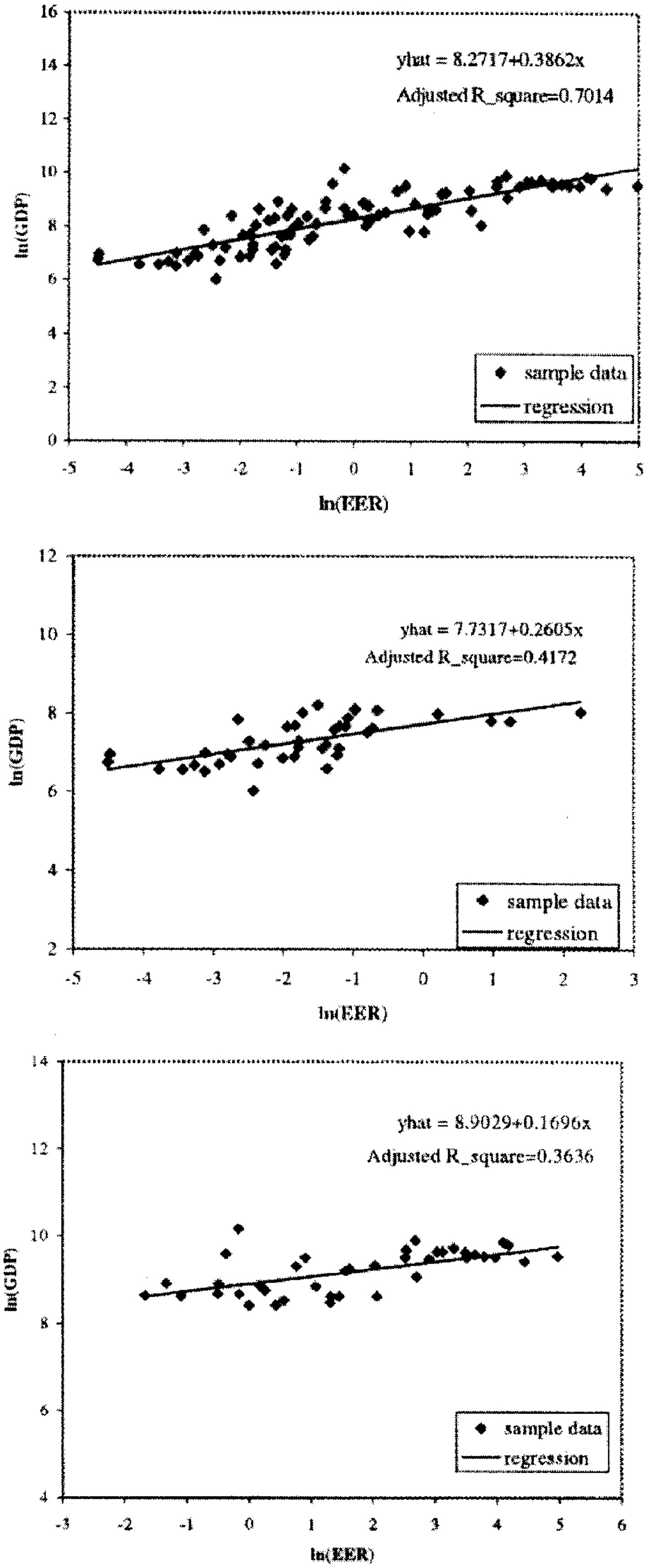
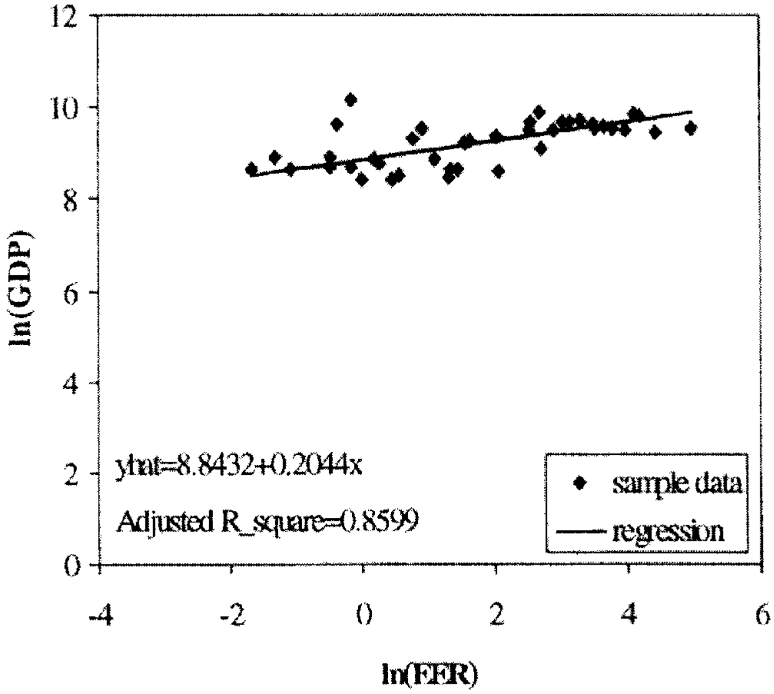


Figure 5. Linear regression of ln(GDP) vs. ln(EER) (Top: Total 87 countries; Middle: Low GDP countries; Bottom: High GDP countries)



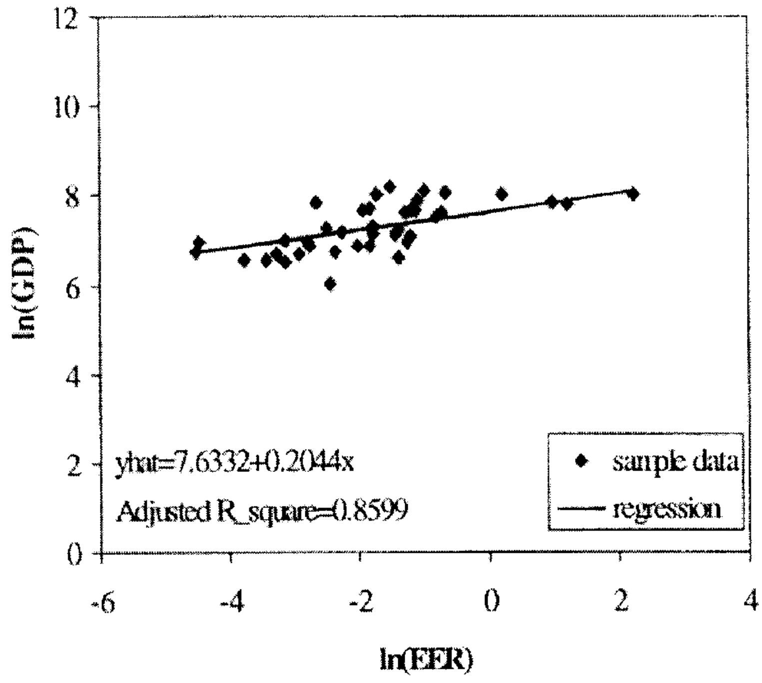


Figure 6. Linear regression of ln(GDP) vs. ln(EER) (Top: High GDP countries; Bottom: Low GDP countries)

Table 1. Rank correlation coefficients for low GDP and high GDP countries

Corr. Coeff.	BDG	BEN	COG	IND	IDN	SDN	BEL	CAN	DEU	JPN	NZL	USA
BDG	1	1	1	1	0.8	1	-0.2	-0.4	-0.2	-0.4	-0.8	-0.8
BEN	1	1	1	1	0.8	1	-0.2	-0.4	-0.2	-0.4	-0.8	-0.8
COG	1	1	1	1	0.8	1	-0.2	-0.4	-0.2	-0.4	-0.8	-0.8
IND	1	1	1	1	0.8	1	-0.2	-0.4	-0.2	-0.4	-0.8	-0.8
IDN	0.8	0.8	0.8	0.8	1	0.8	0.4	0	0.4	0	-0.6	-0.6
SDN	1	1	1	1	0.8	1	-0.2	-0.4	-0.2	-0.4	-0.8	-0.8
BEL	-0.2	-0.2	-0.2	-0.2	0.4	-0.2	1	0.8	1	0.8	0.4	0.4
CAN	-0.4	-0.4	-0.4	-0.4	0	-0.4	0.8	1	0.8	1	0.8	0.8
DEU	-0.2	-0.2	-0.2	-0.2	0.4	-0.2	1	0.8	1	0.8	0.4	0.4
JPN	-0.4	-0.4	-0.4	-0.4	0	-0.4	0.8	1	0.8	1	0.8	0.8
NZL	-0.8	-0.8	-0.8	-0.8	-0.6	-0.8	0.4	0.8	0.4	0.8	1	1
USA	-0.8	-0.8	-0.8	-0.8	-0.6	-0.8	0.4	0.8	0.4	0.8	1	1

BDG: Bangladesh
BEN: Benin
COG: Congo
IND: India
IDN: Indonesia
SDN: Sudan

BEL: Belgium
CAN: Canada
DEU: Germany, W.
JPN: Japan
NZL: New Zealand
USA: United States

Table 2. Comparison of Model 1 and Model 2 with the Solow and Mankiw-Romer-Weil Growth Models

	Estimated Coefficient (Standard Error) Dependent Variable: LN real GDP Per Adult in 1990			
Models / Independent Variable	Solow (1956)	Mankiw-Romer -Weil (1992)	Model 1	Model 2
LN (*I/Y*)	1.41 (0.18)	0.65 (0.18)	****	****
LN (*n*+0.05)	-0.36 (0.14)	-0.31 (0.11)	****	****
LN (SCHOOL)	****	0.73 (0.10)	****	****
LN EER	****	****	0.39 (0.03)	0.20 (0.03)
Intercept	4.51 (0.55)	5.50 (0.46)	8.27 (0.06)	7.63 (0.08) [a] 8.84 (0.08) [b]
Number of Countries	87	87	87	87
R-Square (Adjusted)	0.53	0.70	0.70	0.86

Note: *I/Y* = investment (saving) rate, *n* = population growth rate, and SCHOOL = percentage of working age population in secondary school averaged between 1960-85) Data Source, Mankiw, et al (1992). Data for constructing the statistic EER are obtained from Barro and Lee (1993). [a]. Intercept for low GDP countries. [b]. Intercept for high GDP countries.

Table 2A

VARIABLES:	Dependent Variables: Natural Log of Average Investment Rate (*I/Y*) and Natural Log of Population Growth Rate (*n*) '60-'90 (Mankiw, et al, 1992) Estimated Coefficients (Standard Error) Number of Observations:87*			
Dependent / Independent	***LN (I/Y)*** (1)	***LN (I/Y)*** (2)	***LN (n+0.05)*** (3)	***LN (n+0.05)*** (4)
LN (*HQ*)	0.35 (0.04)	0.27 (0.06)	****	0.01 (0.01)
LN (NQ)	****	-0.13 (0.06)	0.08 (0.01)	0.09 (0.01)
Intercept	-0.86 (0.15)	-1.34 (0.26)	-2.52 (0.02)	
Standard Error	0.47	0.46	0.09	0.09
F-Stat				
R-Squared (Adjusted)	0.48	0.51	0.57	0.57

Table 2B

Correlation Matrix	Ln (SCHOOL)	Ln(*n+0.05)*	Ln(*I/Y*)	Ln(EER)	Ln(*NQ*)	Ln(*HQ*)
Ln (SCHOOL)	1.00					
Ln(*n+0.05)*	-0.48	1.00				
Ln(*I/Y*)	0.70	-0.41	1.00			
Ln(EER)	0.83	-0.72	0.70	1.00		
Ln(*NQ*)	-0.67	0.76	-0.61	-0.96	1.00	
Ln(*HQ*)	0.86	-0.44	0.70	0.81	-0.68	1.00
Ln(GDP/Adult)	0.77	-0.73	0.71	0.88	-0.82	0.87