CHAPTER 4

EMPIRICAL ANALYSIS OF THE EDUCATIONAL FUNDING AND INCOME INEQUALITY IN MALAYSIA

4.1 Introduction

In this chapter, we will present on both from an international and a Malaysian perspective, analysis of the relationship between educational funding and the policy objective of poverty reduction and income redistribution. We will test the hypothesis that education improves the incomes of the poor not only indirectly through growth, but also in a direct manner by providing a level playing field. Utilizing a sample of 111 countries and a measurement of human capital that takes into consideration international differences in the quality of education, we will attempt to find empirical support for our hypothesis that would suggest that education policies are a first-best poverty reduction strategy.

The rest of this chapter is structured as follows: Section 4.2 present the traditional models of human capital theory from which we derive our hypotheses. Section 4.3 presents the international evidence that education improves the incomes of the poor not only indirectly through growth, but also in a direct manner by providing a level playing field. Part one of the Section 4.3 presents the data and the basic specification for the preliminary empirical analysis from the international perspective; part two will then presents the empirical results. In Section 4.4, we will move our focus to Malaysia where the empirical analysis of the relationship between equality in educational funding and attainment and equality in the distribution of income in Malaysia will be carried out. Finally, Section 4.5 summarizes the argument and points out some limitations of the study.

4.2 The "Human Capital Earnings Function"

As discussed in Chapter 2, there are a number of empirical studies that have investigated the relationship between educational factors and income equality (see a survey by Psacharopoulos and Woodhall (1985, pp.264-270) and Ram (1989)). Earlier work shows that there is a close relationship between education and income distribution in developed countries. Becker and Chiswick (1966) show that, across regions in United States, income inequality is positively correlated with inequality in schooling and negatively correlated with the average level of schooling. Chiswick (1971), based on cross-section data from nine countries, suggests that earning inequality increase with educational inequality. Subsequent studies find that a higher level of schooling reduces income inequality, while inequality of educational attainment increases it (Adelman and Morris (1973), Chenery and Syrquin (1975), Ahluwalia (1975), Marin and Psacharopoulos (1974), Psacharopoulos (1977) and Winegarden (1987)). On the other hand, Ram (1984 and 1989) finds, with slightly different specifications and data that mean schooling and schooling inequality do not have any statistically significant effects on income inequality.

Although results from previous empirical studies were ambiguous, most of these studies have emphasized on The "human capital earnings function" (HCEF), which has become a fundamental tool in research on earnings, wages and incomes in developed and developing economies. It is an accepted procedure in litigation involving earnings, such as cases involving the value of lost earnings due to injury, death or discrimination. It is also frequently used to make educational policy decisions based on estimates of the rate of return from schooling (see, for example, Psacharopoulos and Mattson, 1996).

The basic feature of the HCEF is that it relates the natural logarithm of earnings to investments in human capital measured in time, such as years of schooling and years of post-school work experience. It has several desirable features:

(1) It is not an ad hoc specification. It is derived from an identity. As a result, the coefficients of the equation have economic interpretations

(2) Because of the positive skewness of earnings and the rise in earnings inequality as schooling level increases, by using the natural logarithm of earnings rather than earnings as the dependent variable the residual variance in the HCEF is less heteroskedastic and the distribution of the residuals is closer to normal.

(3) It is an efficient user of data. Although data on earnings, years of schooling and years since leaving school are readily available, data on individual schooling costs are not readily available. The HCEF procedure involves converting a relationship between earnings and dollar investments in human capital to one between the natural logarithm of earnings and years of investment in schooling and training.

(4) The HCEF is flexible, allowing for easy incorporation of additional variables appropriate for the particular purpose of the study.

(5) Finally, the coefficients of the HCEF are devoid of units, facilitating comparisons across space (e.g., countries) or across time periods (e.g., decades).

Following Chiswick and Mincer (1972) and Mincer (1974, 1976), the human capital estimating form is:

$$\log y = \beta_0 + \beta_1 S + \beta_2 T - \beta_3 T^2 \qquad \beta_I \geq 0,\ I = 0, 1, 2, 3$$

where y = annual income; S = number of years of formal schooling; and T = number of years of labor force experience. Years of labor market experience, T, are assumed to be measured by age A, minus schooling, S, minus 5; that is $T = A - S - 5$, where six is assumed to be the age at the commencement of schooling.

The model of optimal investment in human capital, which underlies the earnings function for an individual, was initiated by Be-Porath (1967). It predicts a declining rate of investment in human capital with age. The intuitive reasoning behind this result is that most of the investment is made at younger ages to give individuals a longer period in their finite lifetimes over which they can receive returns. But the entire investment is not made instantaneously (before beginning the working life) because the marginal cost of human capital rises within each period, so that it pays to spread the investment over time. The investment declines over time both because marginal benefits decline and because the marginal cost curve itself shifts upward with advancing age. There is also the depreciation of human capital with age (owing to obsolescence and physiological factors), that accentuates this decline in investment.

Qualitatively, the Ben-Porath analysis implies three distinct phases of investment in human capital over the life cycle. In initial phase, all available time is spent acquiring human capital. This period of complete specialization is one of full-time schooling and no earnings, and it can end before the completion of schooling. In the second phase, there is positive but declining investment in human capital. This is a period of on-the-job training and includes part-time schooling, when a declining fraction of available labor time spent on the further acquisition of human capital. In the final phase, all available time is spent earning, and none is spent acquiring the additional human capital—indeed, there is a net loss arising from depreciation.

These considerations lead to a declining rate of investment in human capital over the life span, which becomes negative in the final phase. The decline itself implies that earnings rise to peak (at zero net investment) and then begin to fall off. But the exact shape of the earnings function depends on the particular rate of decline assumed, that is, on the shape of the life cycle investment schedule. A linear decline in the post school investment schedule generates the following quadratic earnings function:

$$\log y = \beta_0 + \beta_1 S + \beta_2 T - \beta_3 T^2 \qquad \beta_i \geq 0,\ I = 0, 1, 2, 3$$

Apart from those mentioned, the main assumptions subsumed in the derivation of this earnings function are: (1) a constant labor market return (β_1) for every year of schooling, and (2) independence between the return to formal schooling and to post school investment (that is, no interaction effect between education and on-the-job experience).

The simplicity and econometric tractability of this earnings function make it agreeable to work with. As stated above, quite apart from its interpretation in terms of human capital theory, it furnishes some useful by-products. Since the dependent variable is the logarithm of income, the estimated regression equation explains the variance of $\log y$ – a familiar index of inequality. The computed R^2 can then be interpreted as the percentage of inequality that is explained by the model.

Following the above human capital theory, we will try to prove that public expenditure on education and its dispersion will reduce the income inequality through our cross-countries and cross time analysis. This will be carried out in Section 4.3 and 4.4, where we look at the relationship between educational funding, income distribution, the level and dispersion of education, and the level of income across countries (the international cross countries analysis) and across time periods (Malaysia analysis). Our empirical analysis will prove that in Malaysia, higher educational funding, higher education levels and dispersion among the population have contributed to a more equal income distribution.

4.3 Cross Countries Analysis

The World Bank in its World Development Report 2000 highlighted that effective anti-poverty strategies should not only focus on economic growth alone, but also on three additional issues: strengthening the participation of poor people in local decision-making and fighting discrimination; reducing vulnerability of the poor to economic and natural shocks, sickness and violence; and lastly, expanding economic opportunity and access to assets, such as education, capital and land. This study argues that human capital appears to be the main aspect of most poor people. Hence, investment in the human capital of the poor should be a powerful way to augment their assets, redress asset inequality and reduce poverty.

Lacking assets is both a cause and an outcome of poverty. Poor people have few assets in part because they live in poor countries or in poor areas within countries. They also lack assets because of stark inequalities in the distribution of wealth and the benefits of public action. In West and Central Africa the rich-poor gap in school enrollment ranges from 19 percentage points in Ghana to almost 52 percentage points in Senegal (World Bank, 2000a). Poor women and members of disadvantaged ethnic or racial groups may lack assets because of discrimination in the law or customary practices. Low assets and low income is mutually reinforcing: low education translates into low income, which translates into poor health and reduced educational opportunities for the next generation.

Between the mid-1980s and mid 1990s public spending on education and health increased in a large number of low-income countries, though slowly. For 118 developing and transition economies, real per capita spending increased on average by 0.7 percent a year for education and 1.3 percent a year for health. Such spending also rose as a share of total spending and national income (ibid). Table 4.1 shows that the World average of public expenditure on education has increased from 3.9 percent of GNP to 4.8 percent in 1997. Most of the regions have shown increased in public expenditure of about 25% between 1980 and 1997. However, evidence suggest, for example, that public spending on education is not progressive but is frequently regressive (table 4.2)

Table 4.1
International data of public expenditure on education

	Public expenditure on education % of GNP		Net enrollment ratio % of relevant age group* Primary		Secondary	
	1980	1997	1980	1997	1980	1997
World	3.9	4.8	81	90	60	68
Low income	3.4	3.3	66	76	38	51
Middle income	3.8	4.8	86	97	63	71
Lower middle income	3.5	4.8	85	98	64	70
Upper middle income	4.0	5.0	88	96	59	75
Low and middle income	3.5	4.1	78	89	53	63
East Asia & Pacific	2.5	2.9	86	99	59	67
Europe and & Central Asia	-	5.1	92	100	84	81
Latin America & Caribbean	3.8	3.6	85	94	55	66
Middle East & North Africa	5.0	5.2	74	87	46	66
South Asia	2.0	3.1	64	77	38	55
Sub-Saharan Africa	3.8	4.1	-	-	-	-
High Income	5.6	5.4	97	100	87	96

Note:

* UNESCO enrollment estimates and projections as assessed in 1999.

Source: World Bank (2000a)

Table 4.2
Public spending on education by income quintile
In selected developing countries (percent)

Country	Year	Quintile 1 (poorest)	2	3	4	5 (richest)
Armenia	1996	7	17	22	25	29
Cote d'Ivoire	1995	14	17	17	17	35
Ecuador	1998	11	16	21	27	26
Ghana	1992	16	21	21	21	21
Guinea	1994	9	13	21	30	27
Jamaica	1992	18	19	20	21	22
Kazakhstan	1996	8	19	23	27	26
Kenya	1992/93	17	20	21	22	21
Kyrgyz Republic	1993	14	17	18	24	27
Madagascar	1993/94	8	15	14	21	41
Malawi	1994/95	16	19	20	20	25
Morocco	1998/99	12	17	23	24	24
Nepal	1996	11	12	14	18	46
Nicaragua	1993	9	12	16	24	40
Pakistan	1991	14	17	19	21	29
Panama	1997	20	19	20	24	18
Peru	1994	15	19	22	23	22
Romania	1994	24	22	21	19	15
South Africa	1993	21	19	17	20	23
Tanzania	1993	13	16	16	16	38
Vietnam	1993	12	16	17	19	35

Note: * Data taken are around 1993/4 and are considered `high quality' by the World Bank. In cases of non-availability, the data are drawn from a most recent published household survey, i.e.: Ecuador (1998) and Morocco (1998/99)

Source: World Bank (2000a)

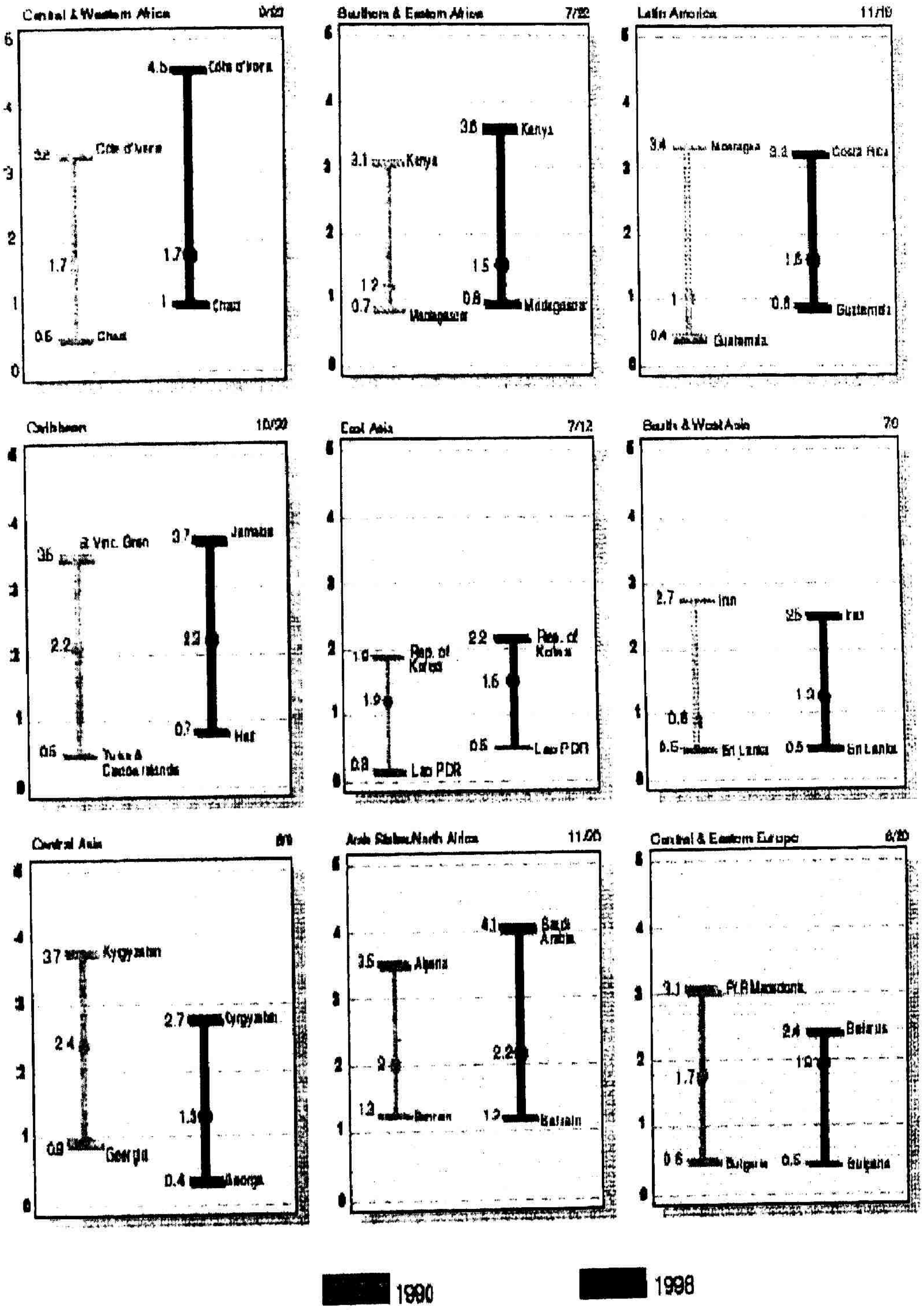
The following indicators will give us some insight on the current world development and trends on expenditure on education, particularly primary education: -

The public current expenditure on primary education as a percentage of GNP[14] in the less developed countries was less than 1.7 per cent GNP in 1998. One tenth reported spending less than 0.7 per cent and one tenth over 3.6 per cent. The variations reported between regions are quite large although they have narrowed over the assessment period: the median values for each of the nine regions for which data are available were between 0.8 per cent and 2.4 per cent in 1990 and between 1.3 per cent and 2.3 per cent in 1998 (Fig. 4.1).

There are, however, large variations within regions ranging between about 1.5 and 3.5 percentage points between the highest and lowest reported values. Within each region, these variations are as great or greater than the median value itself. In 1998, the largest variations reported within regions were in Central and Western Africa (3.5 percentage points) and the Caribbean (3.0 percentage points). In four of the nine regions the gap between the highest and lowest spenders has narrowed over the assessment period. In all of these cases the maximal reported decreased -- especially in the transition countries of Central Asia and Central and Eastern Europe -- although in Latin America the larger factors in the reduction of variation within the region were the increases reported at the lower end of the scale.

[14] This indicator measures public current expenditure on primary education (central, provincial and local) expressed as a percentage of the GNP.

Figure 4.1
Public current expenditure on primary education as a percentage of GNP, 1990 and 1998 (median values and variation within regions)



Source: UNESCO (2000)

Care needs to be taken in interpreting these results, for example, when attempting to draw conclusions about high levels of expenditure relative to GNP. It is true that they may be associated with high ratios of enrolment or low pupil-teacher ratios, smaller classes or relatively more contact hours between pupils and teachers - all of which are generally regarded as desirable features and likely to encourage the provision of good quality education. Yet high levels of expenditure can occur even where enrolment rates are relatively low or pupil teacher ratios are relatively high and require further analysis in order to interpret and understand the underlying causes and reasons.

Next is the *public current expenditure on primary education per pupil as a percentage of GNP per capita.*[15] In 1998, the regional average (median) expenditure per pupil varied between 8 per cent and 20 per cent of GNP per capita in the eight regions for which data are available compared with between 6 per cent and 19 per cent in 1990. All but one region - Central Asia - showed increases in the median values reported over the assessment period indicating that expenditure per student had increased relative to GNP per capita. This may have been the result of real increases in expenditure per pupil or decreases in GNP per capita or a combination of both. Nevertheless, in relative terms, the results suggest that countries have given a higher priority to funding allocated to primary education over the assessment period.

[15] This indicator measures the average cost of a pupil in primary education in relation to the theoretical average income of individuals in each country. It is a proxy measure of a country's ability to afford to pay for education and avoids problems of international comparability that result if expenditures need to be converted to a common currency.

Table 4.3
Public expenditure on primary education per pupil as a percentage of GNP per capita, 1990 and 1998 (median values and variation within regions)

	Median		Intra-regional variation	
	1990	1998	1990	1998
Southern & Eastern Africa	6	13	13	13
Latin America	8	10	21	26
Central Asia	16	11	25	9
East Asia	8	11	12	19
South & West Asia	7	8	24	10
Caribbean	12	16	11	18
Arab States/North Africa	19	20	21	32
Central & Eastern Europe	14	17	12	13

Source: UNESCO (2000)

As with the first indicator, greater variations were reported within regions than between regions. For six of the eight regions for which data are available for this indicator, the variations within regions widened over the assessment period. In the other two regions - Central Asia and South and Western Asia - the gap narrowed considerably but largely because of very large decreases in the highest reported values down from 37 per cent to 16 per cent in Central Asia and from 28 per cent to 15 per cent in South and West Asia (Table 4.3).

The third indicator is *public expenditure on primary education as a percentage of total public expenditure on education.*[16] This indicator measures the relative priority given to primary education within overall public expenditure on education. In 1998, the regional variations reported in the proportions of public education expenditure devoted to primary education were not as great as for the previous indicator. In the ten regions for which data are available, the average (median) proportions of public

[16] The share of public expenditure on primary level compared to the whole of the public expenditure of education.

expenditure devoted to primary education varied between about 36 per cent and 46 per cent. The variation between regions has narrowed over the assessment period with reductions in the highest averages reported in 1990 and increases in the lowest reducing the range between regions from 23 percentage points in 1990 to 10 percentage points in 1998. However, as with the other finance indicators, the largest variations were reported within regions (Table 4.4).

Table 4.4
Public expenditure on primary education per pupil as a percentage of total public expenditure on education, 1990 and 1998 (median values and variation within regions)

[illegible]	[illegible]		Intra-regional variation	
	1990	1998	1990	1998
Central & Western Africa	49	46	61	36
Southern & Eastern Africa	34	42	23	25
Latin America	40	47	53	51
Central Asia	26	36	36	34
East Asia	41	43	64	35
South & West Asia	43	44	49	45
Caribbean	47	46	37	36
Arab States/North Africa	47	43	62	46
Central & Eastern Europe	39	44	66	60
Pacific	40	36	50	57

Source: UNESCO (2000)

There remains a need to supplement the indicators used here with statistics which would make it possible to better capture and understand the variations covered in this analysis. The development of associated methodologies for analysis will make it possible in future to enrich understanding of the financing of education throughout the world. All three finance indicators are affected by external factors which can make it difficult to compare countries - especially those in very different circumstances. In particular, they are dependent on: the duration of the primary phase of education in each country (which varies between 4 and 8 or 9 years between countries); the number of children of school-age in the population - the 'demand' for education - or, at least, the proportion of the demand which is actually met (that is, the actual number of school places provided); the levels of teachers' salaries and other remuneration; the country's ability to 'afford' to pay for education as indicated by the level of GNP per capita and; the level of financial support from the private sector which can be large in some countries.

We will now test the hypothesis that education improves the incomes of the poor not only indirectly through growth, but also in a direct manner by providing a level playing field. First of all, we will examine the data source and specification of the variables.

4.3.1 Income Distribution

Is the source for internationally comparable data on the distribution of income, the data set is from Deininger and Squire (1996). This data set contains Gini coefficients[17] and cumulative quintile shares for 111 countries over a period of 40 years. The average per capita income of the poor is defined as the average per capita income of the poorest 20 percent of the population.

---

[17] The Gini coefficient represents a ratio that indicates the extent of inequality in the distribution of income and ranges from 0 to 1, with zero indicating perfect equality, and one indicating perfect inequality. Named after Corado Gini, the Gini coefficient is the ratio of the area between the 45 degrees line and the Lorenz curve and the area of the entire triangle. The Gini coefficient captures the disparities in the percentages of income that each percentile of the population receives. If each percentile receives one percent of the income, there is no disparity and the Gini coefficient is zero. If one percentile receives all the income, there is maximum disparity, and the Gini coefficient is 1.

At this point, it is worth stressing that this definition does not provide a very homogenous measure of poverty, neither across countries nor across time. For example, in Indonesia it was only in 1997, just before the Asian crises, that absolute poverty (as defined by the World Bank) was reduced to 20 percent of its population. In this case, our measure would be an appropriate indicator of absolute poverty. However, countries such as Bangladesh have 60 percent of their population living on less than one dollar a day. In that case, our measure will only reflect how poorest of the poor are faring—without capturing the extent of absolute poverty. Another drawback of this measure is that its capacity to register changes in the mass of the desperately poor across time is not very accurate. Again, if extraordinary growth in Bangladesh were to halve absolute poverty, our measure may not reflect any change at all. These ambiguities should be kept in mind when we use the term "incomes of the poor". This approach focuses on relative poverty rather than on absolute poverty.

The poverty data are taken from the Deininger and Squire (1996) data set. First, we derive a sample of 102 countries for which 'high quality' Gini coefficients are available.[18] We use data around 1990 and restrict the sample to one observation per country. For 89 of the 102 countries with high quality Gini coefficients (see Appendix 4.1), there is also information about the share of the income accruing to the poorest 20 percent of the population (quintile 1). For these countries, we measure average per capita income of the poor as average per capita income times the share of income accruing to the poorest quintile divided by 0.2, where data for average per capita income are taken from the Penn World Tables (PWT 1994).

We estimate the average per capita income of the poor for the remaining 13 countries in our sample under the assumption that the distribution of income is lognormal. If so, we can approximate the missing quintiles for these countries on the basis of Gini coefficients by using

$$\ln y_p = -\gamma.G + \ln y ....(4.1)$$

Where $\ln y_p$ denotes the natural logarithm of average per capita income in the poorest quintile of the population; G denotes the Gini coefficients; ln y denotes the natural logarithm of average per capita income in the entire population, and $-\gamma = 0.036$ is a constant. The resulting numbers for the average per capita income of the poor are listed in the Appendix 4.2, together with all other variables used in the analysis.

With the data set, regressing the incomes of the poor on average per capita income yields an R-squared[19] of 0.86, and a slope coefficient[20] of 1.06, which is not statistically different from 1 (see Figure 4.2). The result obtained is summarized as follows (for full regression output, see Appendix 4.3):

[18] In order to be included in 'high quality' set, an observation must be drawn from a published household survey, provide comprehensive coverage of the population and be based on a comprehensive measure of income or expenditure. Deininger and Squire (1996)
[19] A R-squared lies between 0 and 1. If it is 1, the fitted regression line explains 100 percent of the variation in the explained variables (in this case *Yp*). Typically, however, R-squared lies between these extreme values. The fit of the model is said to be "better" the closer is R-squared to 1. (Gujarati D. N. 1995)
[20] Slope coefficient, also know as the Beta coefficient in portfolio theory. The interpretation is that a 1 percent increase in the average per capita income leads to an increase of about 1.06 percent increase in the income of the poor. (*Ibid)*

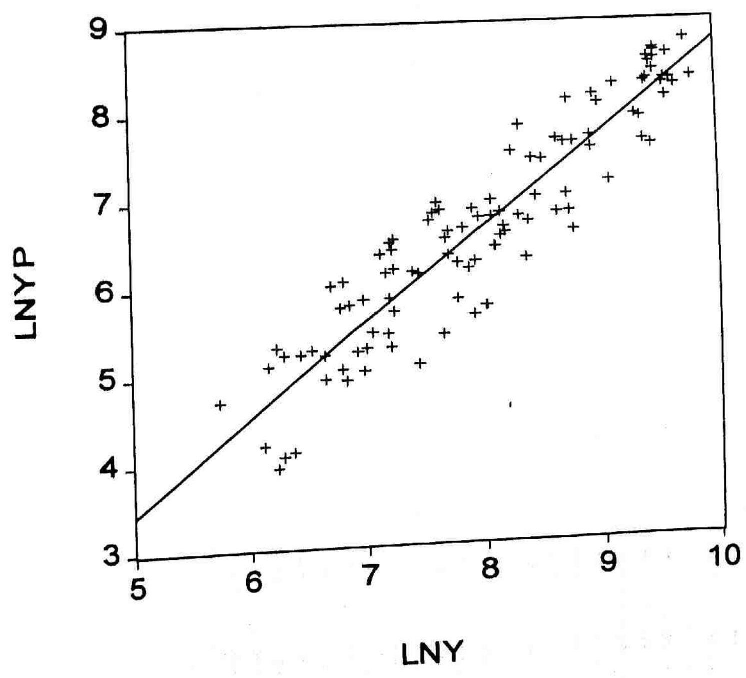
$$\ln yp = -1.88 + 1.06 \ln y$$

$$(-5.45) \quad (24.95)$$

R-squared = 0.86 D.W.[21] = 2.2695

Hence we can conclude at this stage that growth is good for the poor: higher average income would translate one-for-one into higher income of the poor. The question is whether other variables could have additional positive impact on the income of the poor. Our focus is on education.

Figure 4.2
Economic Growth is good for the Poor



[21] Durbin-Watson *d*-statistic: a tool commonly used to detect serial correlation, which is simply the ratio of the sum of squared differences in successive residuals to the RSS. In this case, we do not reject the null hypothesis, meaning no autocorrelation either positive or negative is detected. (*ibid)*

4.3.2 Education

In the empirical growth literature, it has been common practice to use enrollment rates or average years of education as proxies for the change and the level in the stock of human capital. As discussed in WoBmann (2000), the standard specification of human capital in macroeconomic production function is problematic for methodological and empirical reasons. For example, a large body of micro-econometric evidence based on Mincerian wage equation would suggest a semi-logarithm and not a log-linear relation between output per worker and average years of education. In addition, rates of return to education tend to decline with rising levels of schooling (Psacharopoulos 1994), and the quality of a year of education may substantially differ across countries. All these aspects should be taken into account when constructing an empirical measure of the stock of human capital. Hall and Jones (1999) address these problems by specifying the stock of human capital (H) in a way that is consistent with a microeconomic Mincerian wage equation. Their measure of human capital is given by

$$H_i = e^{\sum r_j . S_{ij}} L_i ....(4.2)$$

where r_j is the world average of the Mincerian rate of return to investment in the j-th level (primary, secondary, or higher) of education, S_{ij} is average years of schooling taken from Barro and Lee (1996) at the j-th level of education in country i, and L_i is the number of working-age persons in country i.

Gundlach et al. (1999) improve this empirical measure of human capital by using social rates of return to education derived on the basis of the so-called elaborate method as reported in Psacharopoulos (1994) and by accounting for country-specific duration of each level of education as reported in UNESCO's Statistical Yearbook. In addtion, Gundlach et al. (1998) address the second part of the second problem. They use an index of schooling quality calculated by Hanushek and Kimko (2000) on the basis of international cognitive achievement tests of students in mathematics and natural sciences to account for international differences in the quality of education. The resulting measure of human capital per working-age person in country *i*, is given by

$$\ln(H_i / L_i) = \begin{cases} r^{\Pr i}.S_i.Q_i & if S_i \leq \Pr i_i \\ (r^{\Pr i} \Pr i_i + r^{Sec}.(S_i - \Pr i_i)).Q_i & if \Pr i_i < S_i \leq \Pr i_i + Sec_i . \\ r^{\Pr i}.\Pr i_i + r^{Sec}.Sec_i + r^{High}.(S_i - \Pr i_i - Sec_i)).Q_i & if S_i > \Pr i_i + Sec_i \end{cases}$$

.... (4.3)

Where $r^{\Pr i}$, r^{Sec} and r^{High} are world-average social rates of return to primary, secondary, and higher education (20 percent, 13.7 percent, and 10.7 percent, respectively).[22] $\Pr i_i$ and Sec_i are country-specific measures of the duration of the primary and the secondary level of schooling; S_i is average years of educational attainment in country *i* taken from Barro and Lee (1996), and Q_i is an index of schooling quality, measured on a 0 to 1 scale.[23] Multiplying quantity of schooling by

[22] Barro and Lee (1996)
[23] For details of the calculation. Including the inputation of missing values for selected countries, see Gundlach et al. (1998). Given that the data from Gundlach et al. (1998) refers to 1990, but distribution data from Deininger and Squire does not do so in all cases, necessary adjustments are made by

quality of schooling to arrive at a measure of quality-adjusted schooling appears to be justified because estimated regression coefficients on quantity and quality did not differ when the log values of these variables were entered separately on the right-hand side of a conventional production function by Hanushek and Kim (2000).

4.3.2 Empirical Results

To estimate the potential impact of quality-adjusted human capital on the incomes of the poor, we estimate an OLS-regression, which controls for the impact of average per capital income. Accordingly, our regression equation reads

$$\ln yp = c + a_1 \ln y + a_2 \ln(H/L) + a_i X_i \quad \text{.....................(4.4)}$$

Where

$\ln yp$ = natural logarithm of average per capita income in the poorest quintile of the population,

$\ln y$ = natural logarithm of average per capita income in the entire population,

$\ln(H/L)$ = natural logarithm of human capital per working-age person,

X_i = a set of further possible control variables.

---

replacing data on average years of education in 1990 with data relating to the years in which the income distribution was actually measured.

Table 4.6
OLS Estimates

	Dependent Variable: ln yp					
	(1)	(2)	(3)	(4)	(5)	(6)
C	-0.84	-0.66	-0.66	-0.6	-0.87	-0.69
	(-1.83)	(-1.25)	(-1.25)	(-1.04)	(-1.85)	(-1.41)
ln y	0.89	0.88	0.88	0.87	0.89	0.88
	(-13.3)	(12.63)	(12.63)	(11.60)	(13.21)	(12.47)
ln (H/L)	0.32	0.31	0.31	0.3	0.31	0.32
	(3.13)	(3.00)	(3.00)	(2.85)	(2.98)	(3.07)
ln INV	-	0.04	0.04	0.01	-	-
		(0.7)	(0.7)	(0.37)		
MINING	-	-	-	0.02	0.02	-
				(0.83)	(0.90)	
MALARIA	-	-	-	-0.01	-	-0.01
				(-0.99)		(-0.99)
Sample	101	101	98	88	98	91
R-squared	0.87	0.87	0.87	0.87	0.87	0.87

Note: t-statistics in parenthesis[24]
For detailed results, see Appendix 4.4

Without including any further control variables, we find that the regression coefficients are statistically significant and have the expected sign (Table 4.6 column (2)). The coefficient a_1 is statistically not different from 1, which preserves the finding that growth in average income is translated one-for-one in growth of income of the poorest quintile of the population. This distributional effect comes on top of the

[24] *t*-statistic: In the language of significant tests, a statistic is said to be statistically significant if the value of the test statistic lies in the critical region. In this case the null hypothesis is rejected. By the same token, a test is said to be statistically insignificant if the value of the test statistic lies in the acceptance region. In this situation, the null hypothesis is not rejected. In the above result, the *t* test is significant and hence we reject the null hypothesis.

growth effect or rising quality-adjusted human capital, which works through higher average income. Our point estimates suggest that a 10 percent increase in stock of quality adjusted human capital per worker would increase the average income of the poor by an additional 3.2 percent.

We include further variables in our regression equation (4) to further test the robustness of the basic result. In Levine and Renelt (1992) and most empirical growth studies, a measure of physical capital accumulation is found to be a robust variable. We include and measure physical capital accumulation (*INV*) as the average share of real investment in GDP in 1960-90.[25] In our specification, this variable yields a statistically insignificant regression coefficient. This result most likely reflects that the inclusion of average income as a conditioning variable already accounts for the potential distributional effect of physical capital accumulation on the income of the poor. But conditioning for average income obviously does not fully account for the distributional effects of human capital accumulation, since the estimated regression coefficient remains statistically significant and more or less unchanged in size.

In further specifications, we include poverty-related variables such as the share of mining in GDP (*MINING*) and the incidence of malaria in a country (*MALARIA*) as further checks of the robustness of our results.[26] A high share of mining in GDP may lead to a relatively unequal distribution of income due to rent seeking activities, and hence to slower growth (Rodriguez and Sachs 1999). The incidence of malaria may limit economic development through poor health, high mortality, and absenteeism of

[25] The share of real investment in GDP is taken from the Penn World Tables (PWT 1994)

the workforce. Accordingly, Bloom and Sachs (1998) have argued for the importance of malaria in explaining African poverty. However, we find statistically insignificant regression coefficients both for *MINING* and for *MALARIA*.

Nevertheless, our basic results remain intact. Even after introducing the additional variables we find that quality-adjusted human capital has a statistically significant positive impact on the incomes of the poor. The size of its effect is somewhat reduced compared to our initial regression equation (4), but our point estimate still suggest that 10 percent increase in the stock of quality adjusted human capital would be associated with a direct increase of the average incomes of the poor by 3.1 percent in addition to its indirect effect through higher incomes.

[26] The share of mining in GDP is taken from Hall and Jones (1999); the proportion of a country's population at risk of falciparum malaria transmission is taken from McArthur and Sachs (2001).

4.4 Empirical Analysis for Malaysia

In the previous section we have show that growth in average income is translated one-for-one in growth of income of the poorest quintile of the population. This distributional effect comes on top of the growth effect or rising quality-adjusted human capital, which works through higher average income.

In this section we go one step further to by investigating the relationship between educational funding, its dispersion and income inequality in Malaysia. Considering the ambiguous theoretical predictions on the relation between education and income distribution (as discussed in Chapter 2), we look for empirical evidence based on Malaysian data. In the next section we discuss the data and present the results of estimating the effects of educational factors and its dispersion on income distribution.

4.4.1 The Model

As discussed in Chapter 2, income distribution is related to the population's average schooling and its dispersion. Income inequality increases with education inequality. In contrast, for a given distribution of education, an increase average schooling has an ambiguous effect on income distribution. To illustrate this, traditional models of human capital theory would suggest the following expression for the level of earnings (Y) of an individual with S years of schooling:

$$\log Y_s = \log Y_0 + \sum_{j=1}^{s} \log(1+r_j) + u \quad \ldots.(4.5)$$

where r_j is the rate of return to the jth year of schooling. The function can be approximated by:

$$\log Y_s = \log Y_0 + rS + u \; ..... (4.6)$$

The distribution of earnings can be written as:

$$Var(\log Y_s) = Var(rS) = \bar{r}^2 Var(S) + \bar{S}^2 Var(r) + 2\bar{r}\bar{S}Cov(r,S) \; ..... (4.7)$$

Hence, an increase in educational inequality (*Var(S)*) leads unambiguously to higher income inequality, with other variables held constant. If the rate of return (*r*) and schooling level (*S*) are independent, an increase in the level of schooling will also lead unambiguously to a more unequal income distribution. If, however, the covariance between the return to education and the level of education is negative, an increase in schooling can reduce income inequality. For example, we can think of an economy where improved access to education may allow people with high abilities to earn more income than people with low abilities, even when all of them have the same level of education (see for example, De Gregorio and Kim, (1999)). In this case, as education expands, income distribution may become more unequal.

Therefore, following the above, we relates the income inequality (which is measures by the Gini coefficients) to education inequality (also known as the distribution of education or dispersion of education, which is measured by CV); average years of education attainment (TYR); level of per capita GDP and public expenditure on primary and secondary education. All explanatory variables are expected to carry negative signs and to be significant except *CV* and *LNGDP*, which would then suggest that the income inequality can be reduced through higher education attainment, higher public expenditure on education and more equal distribution of education.

Using a lin-log model,[27] we estimate the following regression

$$GINI = a + \beta_1 CV + \beta_2 TYR + \beta_3 LNGDP + \beta_4 LNGEPRI + \beta_5 LNGESEC + \beta_6 Xs + e \qquad \ldots\ldots\ (4.8)$$

Where

GINI	= Gini coefficient for the households income distribution;
CV	= Coefficient of variation in the number of years of school completed for the total population over 15 years (defined as the standard deviation divided by the mean, measured in percentage);
TYR	= Educational Attainment measured by Average schooling years in the total population over 15 years (in percentage);
LNGDP	= Log of per capita GDP (at 1990 constant price);
LNGEPRI	= Log of real government current educational expenditure per pupil at primary school (PPP-adjusted 1985 international dollars);
LNGESEC	= Log of real government current educational expenditure per pupil at secondary school (PPP-adjusted 1985 international dollars);

[27] Linear-logarithmic model where the dependent variable (Y) is in linear form but the explanatory variables (Xs) is in log form. This functional form is chosen as we are interested to finding the absolute change in Y for a percent change in X. A model that can accomplish this purpose can be written as

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \quad \text{(Gujarati, 1999)}$$

Xs = Control variables i.e. Unemployment rate for all workers (URT); and, change in the Consumer Price index from the previous year (INFL); and

e = a random error term.

A change in the log of a number is a relative change. Therefore, the slope coefficients in Equation (4.8) measures

$$\beta_i = \frac{\Delta Y}{\Delta X / X} = \text{Absolute change in } Y \text{ over relative change in } X \text{ ......(4.9)}$$

where, ΔY and ΔX represent (small) changes in the dependent variable and explanatory variables. It can be written, equivalently, as

$$\Delta Y = \beta_i \left[\frac{\Delta X}{X} \right] \text{........(4.10)}$$

This equation states that the absolute change in *Gini*(= Δ*Gini*) is equal to β_i times the relative changes in the explanatory variables. If the latter is multiplied by 100, then Equation (4 .10) gives the absolute change in *Gini* for a percentage change in X. Thus, if $\Delta X / X$ changes by 0.01 unit (or 1 percent), the absolute change in Y is 0.01(β_i), holding all other X variables constant. If in our analysis, we find that β_3 = 2, the absolute change in *Gini* is (0.01)(2), or 0.02. Therefore, in our estimating regressions, if *LNGDP, LNGEPRI, LNGESEC* change by 10 percent, the absolute change in *Gini* is 0.1(β_i) where i = 3, 4 and 5. And since *CV* and *TYR* are both measures in percentage, a one-unit change in *CV* or *TYR* would result in one absolute change in *Gini*.

Therefore, following the specification, β_1 and β_3 are expected to carry positive signs, β_2, β_4, β_5 and β_6 are expected to carry negative signs. All explanatory variables must be statistically significant.

4.4.2 Preliminary Data Analysis

The question under study (i.e. the Gini coefficient) will be addressed using Malaysia time series data from 1970 through 2000. The Gini data for the period 1970 to 1995 are taken from various sources including Deininger and Squire (1996), Department of Statistic's various survey, the Third Outline Perspective Plan (OPP III), Malaysia Plan documents, EPU's Malaysian Quality of Life 1999 and other sources. Figures for 1996 through 1999 are taken from the Seventh Malaysia Plan and OPP III. Figure for year 2000 is estimated based on 1999 figure. In addition to data availability criteria, this period was chosen as it begins with the start of NEP. And given the explanatory nature of this paper, it is a time series of a sufficient length to illustrate some very interesting findings.

If we look at this data set we see a remarkable degree of stability in the degree of inequality in the distribution of income, which is measured by a Gini coefficient ranging from zero (perfect equality) to one (perfect inequality). The Gini coefficient, for all households was 0.50 in 1970 and 0.44 in 1999/2000; during this period the low was 0.44 in 1999/2000, and the high was 0.53 in 1976. This is a degree of stability, which, on the surface, is little short of remarkable. The Gini coefficient does, however, exhibit some signs of instability over the period. In particular, there is a negative time trend; an ordinary least squares regression of the Gini coefficient on a time trend yields a slope coefficient, which is negative and significant at the one percent, level. This is a first warning that visual examination of the income distribution data can be misleading.

To measure the degree of inequality in educational attainment, the distribution of the number of years of school completed for the total population over age 15 as given in Barro and Lee (2000). The standard deviation[28] and the mean were approximated from these relative frequency distributions. Although this is just an approximation, it is the best that could be done with the available data. The measure of educational inequality used is the coefficient of variation (the standard deviation divided by the mean). This is superior to the standard deviation as a measure of educational inequality, as it adjusts the measure of educational inequality for increases in mean attainment over time. For example, a distribution wherein 25 percent of individual had one year of primary school, 50 percent had two years of primary school, and 25 percent had three years of primary school would be judged to be less equal (a larger coefficient of variation) than one in which the same percentage had one, two, and three years of secondary school. This is clearly appropriate.

[28] The standard deviation is computed by assuming that each person has an educational attainment of log (1+years of schooling). Thus a person with no formal or zero schooling is assumed to have one (effective) year of educational human capital. The standard deviation (SD or s) can be calculated as follows:

$$s = \sqrt{\sum (x_i - \bar{x})^2 / (n-1)}$$

where $\bar{x}$ is the mean of the sample and n is the number of scores.

The coefficient of variation of education has been steadily decreasing in M from approximately 0.218 in 1970 to 0.129 in 2000, as shown in Appendix 4. is due partly to a decrease in the standard deviation and in part to an increas mean. In any case, the general pattern discussed in the introduction is appar distribution of education is becoming more equal alongside the more distribution of income. We present the data in Appendix 4.5.

The next two variables are real government current educational expenditure p at primary and that at secondary schools (PPP-adjusted 1985 international do Policymakers usually justify higher educational spending as a very effective reducing income inequality (as in the case of Malaysian NEP). Theor increased government expenditure on education could cause the income dist to become either more equal or less equal, but the evidence seems to support t that they contribute to added equality. Again, these are very important vari consider in our analysis. The data are taken from Barro and Lee (2000) shortcoming of these data is that the spending figures comprise only public (including subsidies to private education), whereas the number of pupils incl typically for all schools.

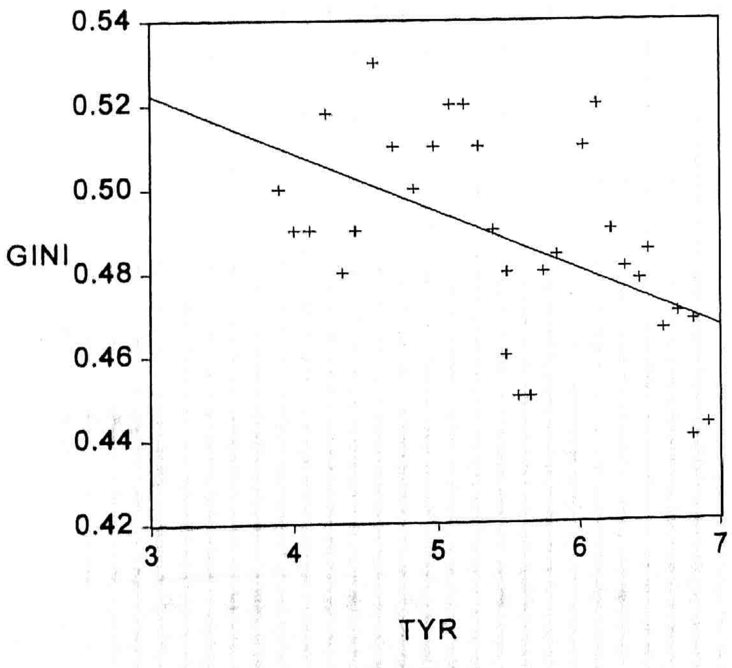
[29] The data are chosen following the suggestion in Ram (1988) that the internationally compar be used, see also Ram (1987).
[30] ***Ibid.***

Two other possible factors affecting the distribution of income studied by many researchers are unemployment and inflation. For example, Blinder and Esaki (1978) found strong disequalizing effects from higher unemployment, with relatively weaker equalizing effects from higher inflation in the United States. We will include these variables in our regression equation to test further the robustness of the result in the Malaysian context.

4.4.3 The Results

Before going into the details of the results it is useful to look at the simple cross-correlation between income distribution and educational variables. Figure 4.3 plots average years of schooling against the Gini coefficient for the period under review. The relationship is negative, indicating that increases in education reduce inequality.

Figure 4.3
Educational Attainment and Income Distribution 1970-2000



On the other hand, Figure 4.4 shows that there is a positive relationship between income and educational inequality.

Figure 4.4
Education Inequality and Income Distribution

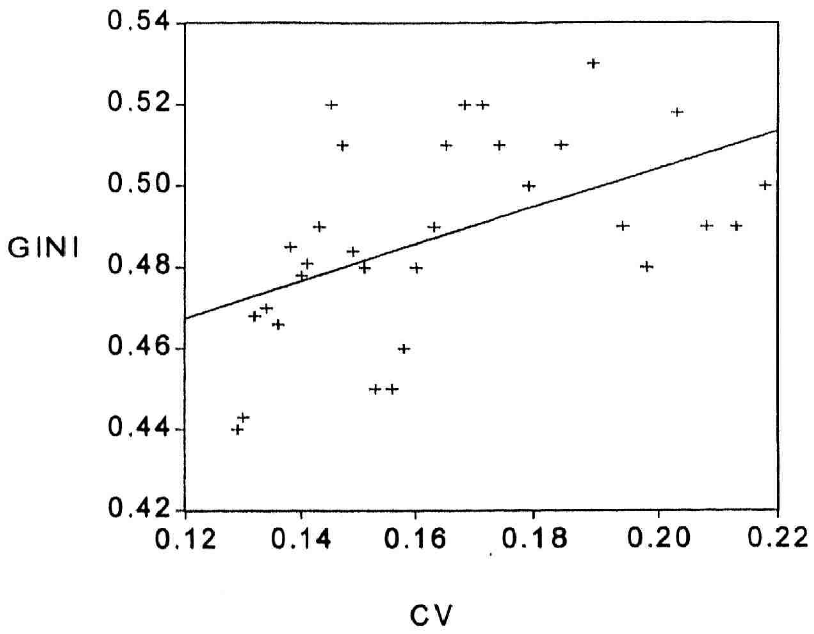
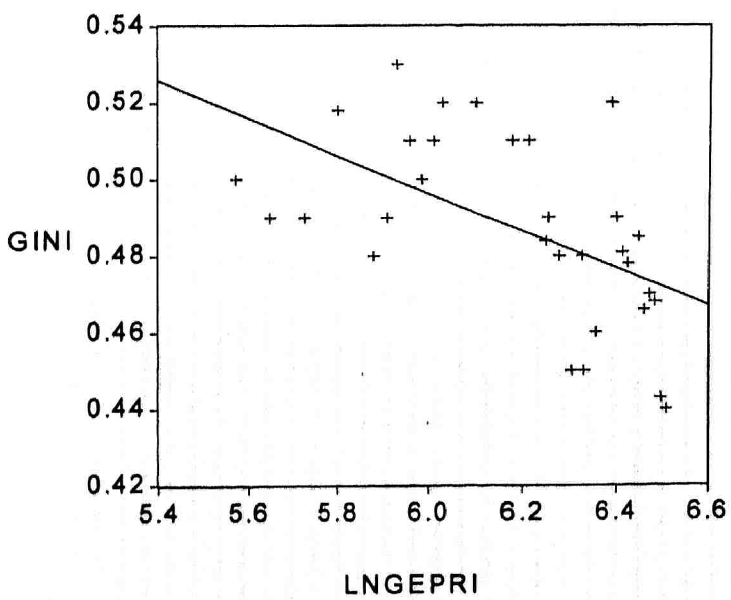


Figure 4.5
Real government current educational expenditure per pupil at primary school and income distribution



Both Figure 4.5 and 4.6 show that increased real government education expenditure has a negative effect on income distribution. Although these figures are suggestive, further statistical analysis is required to examine their robustness and obtain orders of magnitude for the importance of educational factors in explaining changes in income distribution in Malaysia.

Figure 4.6
Real government current educational expenditure per pupil at secondary school and income distribution

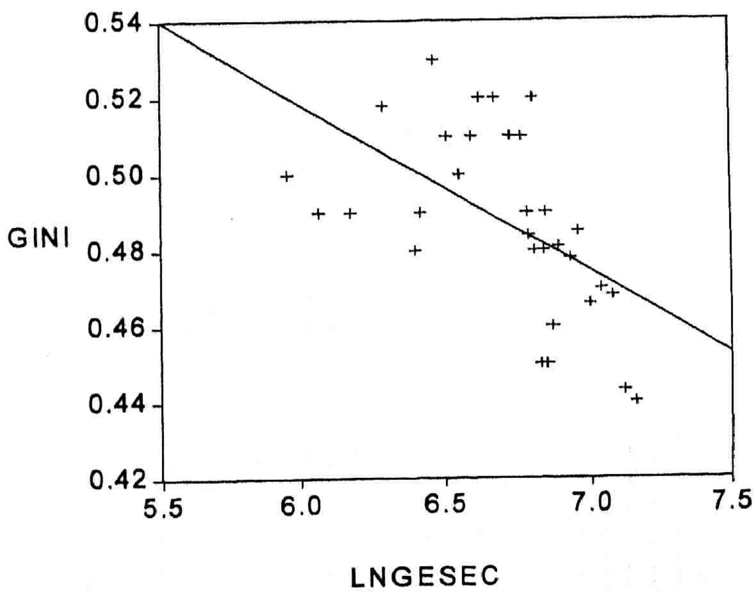


Table 4.7 OLS Estimates for Malaysia

Dependent Variable: GINI												
Regression:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
C	0.740 (1.045)	0.401 (0.586)	0.345 (0.497)	-0.029 (-0.050)	-0.392 (-0.752)	-0.164 (-0.293)	0.863 (5.836)	0.966 (5.495)	-0.255 (-0.455)	-0.094 (-0.157)	0.761 (5.219)	0.816 (4.353)
CV	0.533 (0.385)	0.957 (0.692)	1.263 (0.926)	2.331# (2.496)	2.800# (3.130)	2.498# (2.678)	-	-	2.514# (2.545)	2.328# (2.294)	-	-
TYR	-0.033* (-1.712)	-0.034* (-1.706)	-0.024 (-1.230)	-	-	-	-0.045# (-3.631)	-0.037# (-2.830)	-	-	-0.061# (-4.039)	-0.054# (-3.194)
LNGDP	0.132# (4.376)	0.133# (4.314)	0.133# (4.254)	0.146# (4.878)	0.148# (4.825)	0.144# (4.762)	0.118# (5.504)	0.112# (5.262)	0.136# (3.900)	0.136# (3.918)	0.118# (5.401)	0.110# (4.860)
LNGEPRI	-0.098 (-1.474)	-0.148 (-3.246)#	-	-0.101 (-1.458)	-	-0.155# (-2.870)	-	-0.195# (-4.226)	-	-0.150# (-2.599)	-	-0.149# (-2.924)
LNGESEC	-0.097* (-1.714)	-	-0.169# (-3.087)	-0.069 (-1.232)	-0.121# (-2.735)	-	-0.166# (-4.448)	-	-0.119# (-2.548)	-	-0.133# (-3.542)	-
Control variables												
URT	-	-	-	-	-	-	-	-	-0.001 (-0.385)	-0.001 (-0.239)	-0.006 (2.416)	-0.005 (-1.935)
INFL	-	-	-	-	-	-	-	-	-0.001 (-0.812)	-0.001 (-0.557)	-0.001 (-0.760)	-0.001 (-0.601)
No. of observations	31	31	31	31	31	31	31	31	31	31	31	31
D.W.	1.3186	1.1687	1.3141	1.2344	1.1058	1.2517	1.1034	1.2644	1.0496	1.1902	1.2858	1.28
F-statistic	11.476	13.207	12.667	12.671	15.538	16.081	17.794	16.692	8.9923	9.1083	13.364	11.410
R^2	0.70	0.67	0.66	0.66	0.63	0.64	0.66	0.65	0.64	0.65	0.73	0.70

Source: Appendix 4.6

Note: T-statistics in parenthesis

* Significant at 10% level

\# Significant at 5% level

The OLS results are presented in Table 4.7. Regression (1), (2) and (3) are basic regressions. The regression explains about 70 percent of the variance of income distribution, except for the latter period, where its explanatory power once drops to 63 percent.

The significant statistics for *LNGDP*, *LNGEPRI* and *LNGESEC* indicates that increases in the Real GDP per capita contributes to inequality (a higher Gini coefficient) in the distribution of income. Although the coefficients for real government expenditure per pupil (primary and secondary schools) are not significant at the ten percent level, its sign and its reasonably high t-statistics suggest the possibilities that higher educational spending make the distribution of income more equal. Statistical results improve in regression (2) and (3) when either *LNGEPRI* or *LNGESEC* were excluded.

The coefficients of *CV* and *TYR* however, are insignificant at the five percent level. The Durbin-Watson statistics, although in the indeterminate range, is quite low and therefore suggest that autocorrelation[31] in the error term may be a problem. We performed additional tests to detect autocorrelation and found no clear evidence of autocorrelation.[32] Other test such as zero-order correlation coefficients among regressors was done and it was found that multicollinerility[33] was serious among the

[31] Autocorrelation may be defined as "correlation between member of series of observations order in time [as in time series data] or space [as in cross-sectional data]. (Gujarati, 1995)

[32] The Breusch-Godfrey (BG) test of higher-order autocorrelation or serial correlation LM test were performed for AR(1), AR(2) and AR(3). The results shown that at 5 percent level, (n-p).R^2 were below the critical chi-square value at the chosen level of significance. The full results of the test are show in Appendix 4.7. (*Ibid*)

[33] Multicollinearity originally meant the existence of a "perfect," or exact, linear relationship among some or all explantory variables of a regression model. (*Ibid*)

explanatory variables, i.e. *CV* with *TYR* and *LNGEPRI* with *LNGESEC*. The correlation matrix is shows in Appendix 4.8.

In the data used in this paper, *CV* and *TYR* are highly correlated, with a correlation coefficient of –0.98 (Appendix 4.9). This, from a statistical perspective, means that we are faced with a problem of severe multicollineality between these two explanatory variables, implying that it is very difficult (if not impossible) to statistically separate their individual effects. This suggests that the relatively high standard errors of the coefficient estimates associated with these variables may be due to their strong correlation.[34]

In order to solve the above statistical problem, we dropped some variables from the model. In regression (5), *TYR* and *LNGEPRI* were omitted from the model and in regression (6), *TYR* was dropped together with *LNGESEC*. On the other hand, *CV* was omitted in regression (7) together with *LNGEPRI*, and in regression (8), *CV* was dropped with *LNGESEC*. The results from this estimation are presented in Table 4.7.

[34] As is often the case in time series data sets, the explanatory variables are general mutually correlated. The correlation matrixes are shown in Appendix 4.4. The primary problem, which results from multicollinearity, is that it becomes difficult to separate the effects of different variables: this appears econometrically as inflated standard errors of coefficient estimates. To the extent, however, that a coefficient estimate is significantly different from zero, the multicollinearity has not prevented us from determining that the associated variables has a significant effect. The LNGDP, LNGEPRI and LNGESEC variables generally have significant effects; thus, the focus of our discussion on multicollinearity problem is on CV and TYR.

These results are consistent with the earlier results, with the exception that the coefficients of education attainment and education inequality are now significant at the five-percent level. The results on the coefficients of the real GDP per capita and real government educational expenditure are consistent with the discussion of their possible effects given above. If *TYR* is omitted, both regressions (5) and (6) indicate that *LNGDP*, *LNGEPRI* and *LNGESEC* have significant effects (as before), the results Further indicate the role of educational expenditure and its dispersion on income distribution. Higher educational expenditures will reduced the income inequality by 12 to 15 percent. The *CV* now has a significant positive effect on *GINI*; the decreases in *CV* over time are associated with decreases in the Gini coefficient: less inequality in the distribution of income. That is, when distribution of education becomes more equal so too does the distribution of income. If *CV* is omitted, both regressions (7) and (8) indicate that *TYR* has a significant negative effect. That is, increases in educational attainment, holding everything else constant, cause the degree of inequality in the income distribution to decrease.

We add the square of log of per capita GDP in order to capture the inverted-U curve proposed by Kuznets for the relationship between income distribution and the level of income[35]. The results (Appendix 4.10) confirm that there is a Kuznets curve. We use the log of GDP to estimate this relationship, because the relationship was not found when measured with the level of per capita GDP. The Kuznets curve resulting from the regression in Appendix 4.9 indicates that income distribution becomes more

[35] More than forty years ago, in his Presidential Address to the American Economic Association, Kuznets (1955) suggested that income inequality was generally rising in the early stages of economic development. In the latter phases of the development process, inequality declines, he argued, and this

unequal with higher levels of income up to a range of income between RM1, 900 and RM2, 900 (constant at 1990 prices), and then income distribution starts equalizing. Another specification for the Kuznets curve has been proposed by Anand and Kanbur (1993) and also estimated by Deininger and Squire (1996b). It includes income in the regression as y and $1/y$. Our results show that for different specifications the nonlinearity in the relationship between income and its distribution in significant.

To check further the robustness of our results, we include the variables unemployment rate and inflation. The results are shown in Table 4.7, (regression (9) through (12)). Unemployment is said to have strong diequalizing effects while inflation with relatively weaker equalizing effects (Blinder and Esaki, 1978). However, we find statistically insignificant regression coefficients both for *UNRT* and *INFL*.

4.5 Concluding Remarks

Income distribution is related to the population's average schooling and its dispersion. Income inequality increases with education inequality. In contrast, for a given distribution of education, an increase in average schooling has an ambiguous effect on income distribution. From a political economy perspective as well as according to some endogenous growth models, a more equal distribution of income, which would reduce social conflict and guarantee a greater protection of private property rights, is considered to be conducive to growth. If, for instance, imperfect capital markets are

---

hypothesis of an inverted relationship between inequality and development has since been known as the Kuznets Curve. Kuznets (1955) responsible for observed inequality, then a certain amount of redistribution is believed to enhance growth and welfare because it would transfer resources to agents with potentially higher returns to investment. Redistribution through state-funded access to primary and secondary education for all children might be an efficient way to generate such a transfer of resources.

Overall, our empirical results confirm that education expenditure and education expansion is not distribution-neutral. Education seems to improve the income distribution directly and thus may allow the poor to benefits from growth to a greater extent. Accordingly, a focus of economic policies on education in order to reduce poverty and to speed up development appears to be justified. However, rather than merely expanding access to education, our empirical findings indicate that improving the quality of education should play a crucial role in development strategies.

From our analysis several issues for future research are immediately apparent. First, the direction of causality between inequality and human capital accumulation is an open question. For instance, as suggested by Ram (1989), it might well be that the causality runs from income distribution to human capital accumulation, rather than vice versa as in our interpretation. Second, while our findings provide an encouraging impetus for the use of education policies as part of anti-poverty programs, a rigorous theoretical framework supporting such claim is still missing. Third, highlighting the importance of education policy should be accompanied by a more precise identification of education policies that actually generate the expected effects. WoBmann (2000) stress that the positive effects of additional schooling expenditures can only be expected if the schooling system is governed by incentives to improve performance and to reduce costs, which is apparently not the case in most developed countries.

Many analysts have argued that redistributive effects of educational factors on income are weak if not inconsistent. In this chapter, although that view is not shown to be invalid, it is shown that more sophisticated analyses of the aggregate data do not necessarily lead to this conclusion.

However, one need to be careful in drawing a conclusion that the expansion of educational factors alone can make income inequality decline substantially in a short period. Especially in the case of Malaysia, where the connection between income distribution, education, macroeconomic factors, and government policy and many other factors have to be explored. We will discuss this further in the concluding chapter.