An Economic Analysis of Personal Earnings in Rawalpindi City

NADEEM UL HAQUE*

The intent of this paper is to delineate the determinants of the distribution of income in Rawalpindi city. The basic hypothesis to be tested is that for each individual, personal earnings are a function of his socio-economic characteristics such as age, sex, education, and the like. If the hypothesis bears out, the disparity of incomes within the city would then prove to be a consequence of the variation in these individual characteristics. It would then be of interest to quantify the effect of each of these determinants on the eventual distribution of income.

In this paper first some questions of methodology are examined and the theoretical framework for the anaylsis is set out. Subsequently some previous research which is of relevance to the topic is summarised. A description of the data set and any reservations about it are discussed next. After the discussion of the data set, the results are presented. In the conculsion, some econometric problems and the main implications of the results of analysis are discussed.

THEORETICAL FRAMEWORK

The basic frame work for the following analysis is provided by the theory of human capital. The corner-stone of this theory is that human beings invest in themselves in a variety of ways, i.e. incur present costs for future benefits. Education, on-the-job-training, work experience are all important manifestations of this phenomenon of self-investment. The acquisition of human capital raises an individual's productivity and, because employers pay in proportion to productivity, the individual's earnings. Each person invests in human capital to the point where the marginal return equals marginal cost [2].

*The author is a Research Economist at the Pakistan Institute of Development Economics. Much gratitude is owed to K. Hamdani Research Economist and Dennis Detray Research Adviser for the considerable help and advice that they were always willing to give.

In this study education is used as a direct human capital variable while age is taken as a proxy for the human capital variable, work experience.[1] The hypothesis therefore is that personal earnings are a positive function of age and education. This relationship is expected to explain much of the observed variation in earnings.

However, the observed relation between earnings on the one hand and education and age on the other may just be a statistical illusion. For example, if only the children of the rich can afford to go to school, then the well-paid jobs which they might get after finishing school may not be a result of the amount of human capital acquired but of their family connections. Similarly, if only the very bright get educated then higher earnings are a consequence of innate intelligence and not of education alone. Again structural defects in the market explain the disparity in incomes. Women may be discriminated against, irrespective of age and education. The prevalence of the dual markets might hinder the human capital process at work. In this case these human capital variables will be of greater relevance in the formal sector than in the informal.

There are therefore likely to be a large number of variables which will determine the eventual distribution of income. The use of multiple regression analysis will allow us to separate the effects of a number of independent variables on a particular dependent variable. This technique will enable us to explain the variation in personal earnings in terms of variation in individual characteristics such as education, age, sex and occupation. Our model therefore is the classical least squares equation with both dummy and continuous regressors:

$$y_i = \alpha_o + \sum_{J=1}^{n} \alpha_j X_{ji} + u_i$$

where j stands for a variable, i stands for an individual, y_i is the natural logarithm of the level of earnings or the wage rate, $X_{1i} \ldots X_{ni}$ are n observable characteristics (continuous or dummy) used to explain y_i, and u_i is the random unobserved disturbance with zero mean and constant variance.

The regression model will identify and rank for us the determinants of the distribution of income. The regression package will at the same time provide us with the means and variances of the variables used in the equation. Considerable useful information about the sample or any breakdown of the sample that we may wish to consider can be obtained from these summary statistics. Of particular interest would be the log variance of income, a much used measure of income inequality.[2]

Being a relative measure of inequality, the log variance is useful in making sub-sectoral comparisons, when various sub-samples are considered.

PREVIOUS RESEARCH

Income distribution studies in economic literature tend to be one of three distinct types. The first is concerned with the statistics of the observed distribu-

[1] In the human capital specification education should be a continuous variable measured as the number of years of formal schooling. The Rawalpindi Survey however picked up only the educational level attained.

[2] See 'Sen' [14] about the merits and de-merits of the measure. Basically the log variance is a relative measure of income inequality. The measure is free of any change of units. It is sensitive to income transfers at the lower end of the scale and therefore useful to us as we are probably dealing with the middle portion of the income distributions.

tion of income. Statistical distributions which best fit the observed distribution of income or the distribution above a certain level are examined. Underlying stochastic processes which could generate such distribution are also studied. [6, 9]. For the second type the inequality measures like the Gini coefficient are calculated. Comparisons of these measures amongst nations, amongst regions within nations or amongst various groupings within a population are then made in an attempt to understand inequality [11, 13].

The third type of study is relatively new and unlike the other two is derived from economic theory. This is the theory of human capital that has already been discussed. As noted earlier, here an individual's earnings are a positive function of the amount of human capital he possesses [3, 10|]. The distribution of income is therefore a consequence of individual supply and demand for human capital.

In Pakistan, research on the distribution of income has been mainly of the second type, i.e. the calculation of inequality measures. In her study on the "Measurement of Inequality in Urban Incomes in Pakistan", Khadija Haq calculated the Gini coefficients for the years 1948-49 to 1960-61 and estimated the trend over time of inequality.[3] For her data, she used the estimates of personal income based on the "All Pakistan Income Tax Returns" published by the Central Statistical Office (C.S.O.). Her analysis is therefore restricted to the very high income groups (over Rs. 3,500 per month) which constitutes some 0.01% of the pupulation. For this income group, her results show that (a) income is more unequally distributed in Pakistan than in most developed and developing countries, and (b) income is skewed in favour of the rich but the trend is towards the reduction of disparities within the high income bracket.[4] Her analysis however does not include any estimate of leakages due to misreporting of taxes. Also, over time the proportion of fringe benefits in personal income may have increased.[5] An allowance for either of these factors could increase inequality.[6]

For our purpose, a more interesting study and one to which frequent reference will be made is Blaug's study of earnings in Thailand [5]. Blaug had a very large sample of about 9000 observations[7] A large amount of detailed information was collected for each individual. For example, questions were asked on: all the jobs held by an individual; fringe benefits in cash or kind; social background of respondents; the type of school attended; occupations;

[3] She calculates both the Gini coefficient and the Puareto coefficient for all the years between 1948-49 and 1960-61. Both these coefficients reveal a negative trend. The Gini coefficient falls from 0.61 in 1948-49 to 0.45 in 1960-61.

[4] It would be fair to point out that she qualifies these results rather heavily. She sets out clearly the limitations of her data. On her results the relative position of new entrants and lower income groups in the tax-paying population is improving. She also notes that the tax-paying population has been receiving an increasing share of the national income while paying a decreasing percentage of it in taxes.

[5] The author is conscious of these.

[6] The only other study on personal income distribution for Pakistan was Asbjorn Bergen [4]. He measures the Gini coefficients for East Pakistan, West Pakistan, rural areas urban areas etc. Using the CSO's Quarterly survey data. He also calculates saving ratios. His main conclusions however is that the data base is too weak for the analysis he was trying to do—Most other studies in Pakistan deal with national income and its distribution.

[7] He used the household economic survey of 4,600 observations. As college graduates were under-represented, another survey was conducted. 2000 random observations and 3000 men and women were "interviewed purposively" to reach predetermined quotas defined in terms of age, sex and education.

hours of work; sector of employment; and even income from property and self-employment.

From those individual characteristics Blaug derived 69 new variables. Using a specification similar to the one used in this paper and stepwise regression procedures, he sifted out the effects of these variables on income, the dependent variable. The results of his analysis were:

(*i*) An almost linear age-earnings profile. There was a very shallow concavity in the age-earning profile which could not be picked up by the normal human capital specification using age and the square of age.

(*ii*) Education was an important determinant of the distribution of income. The hypothesis of a positive relationship between the two was accepted.

(*iii*) The two human capital variables, education and age, together explained most of the variation in income.[8]

(*iv*) Amongst the education variables, the higher levels of education were the ones that contributed the most towards explaining the variation in income.

(*v*) Family background, employment status and occupations were not insignificant when explaining income disribution.

Another study of interest is Sudbir Anand's "Size Distribution of Income in Malaya" [1]. Anand, too, had a large sample: 6000 observations. He was not concerned however with estimating the effects of variables other than the human capital variables on income. He mainly tested the applicability of the human capital model in Malaya. The intention was to see how well the model explained the disparity in incomes. He used what may be termed a "pure" human capital specification:

$$\text{Log } Y = B + B_1S + B_2T + B_3T^2$$

where

Y = Income

S = Years of schooling

T = Years of labour force experience and was measured as $T = \text{Age} - S - 5$

The equation was run for the whole sample and for various subsamples selected by occupation, sex, various age groups, social groups and educational levels. The results indicated that the human capital model explained a large part of the variance in incomes in Malaya. The basic hypothesis of increasing returns to education and age and the concavity of the age-earnings profile were all supported by the analysis.

THE DATA

The data used in the analysis were obtained in the Rawalpindi Socio-Economic Survey conducted by the Pakistan Institute of Development Econo-

[8] Blaug includes sex in his "basic" variables and it is the three of these prove to be most significant.

mics in 1975. As a detailed description of the survey, the sampling design and the possible sampling and non-sampling error is found in Hamdani [8], this discussion will concentrate only on the points of interest to the present Study. Briefly, basic socio-economic, information was collected from a thousand households in Rawalpindi. As Hamdani noted, "a tight budget necessitated the small sample households in size and the simple sampling design". He concluded that the sampled households were representative of Rawalpindi [8, p. 148].

The summary showed that the labour force consisted of 1641 individuals of whom 1541 were males. A little less than half of the labour force, 49.24 percent, were regular employees while 33.4 percent were self-employed. Surprisingly only 1.4 percent reported as being casual employees and 5.3 percent as unemployed. Under-employment however was substantial: 23.8 percent would like to work more hours.[9] A minority of 0.2 percent labour force were apprentices, and unpaid family helpers constituted 7.5 percent.

The use of these data for an income distribution analysis has certain reservations. A brief description of the city will bring out the first of these reservations. Rawalpindi is the fifth largest city in Pakistan with an estimated population of 6,73,000 individuals in 1975.[10] An important regional metropolis with primarily administrative functions, the city was the country's interim capital in the early sixties. The development of the country's new capital, Islamabad, on the outskirts of Rawalpindi has aided the development and expansion of its wholesale trade and construction activities. Manufacturing activity is virtually non-existent in the city. Essentially the economic activities of the people in the city are trade, construction and administration (i.e. government employees).[11] The absence of manufacturing activity and the city's close proximity to the national capital make it somewhat unrepresentative of the other urban areas of Pakistan [8, p. 148]. To that extent results of this study would be applicable particular to Rawalpindi.

A bias stems from the sampling design. The sampling frame consisted only of structured and semi-structured dwellings. This could yield an income distribution truncated at the lower tail, i.e. not enough representation of the very poor in the sample. If housing were to be regarded as a normal good, with a positive income effect, then it may be assumed that the very poor live in unstructured or *Kutcha* dwellings.[12] Our sample in that case underestimates the inequality or the variation in incomes.[13] The distinction between structured and unstructured dwellings however fades if we accept what is a fact in underdeveloped countries that the structured dwellings of the poor tend to be overpopulated slums. The mean income in the sample is Rs. 376, which is really very low, especially when we take into account that there are on average 2.68 people dependent on an earner. The per capita income in the sample is therefore

[9] This is however, a very unsatisfactory measure of under-employment.

[10] The figure was arrived at by projecting the 1972 census estimate at a 3.2% annual growth rate [8, p. 148].

[11] We note a sampling bias. Rawalpindi has had since the pre-independence days one of the largest army cantonments. The military personnel were however deliberately excluded from the sample because it is illegal to gather any information on them.

[12] It could also be that living in the unstructured dwellings may be recent migrants earning about the same as the others but in transit to move into the structured houses in which case there is no bias.

[13] As mentioned earlier the mean income for the sample is lower than the per capita income for the whole coming. The sample probably did not pick up enough of the rich. In that case there is reason to assume that inequality has been underestimated.

Rs. 102.1 per month whereas the per capita income per month for the country as reported by the *Economic Survey* is Rs. 153.5.[14] For both these reasons therefore the lower tail of the distribution given by the sample would be representative of those in unstructured dwellings. Inequality however may have been underestimated.

As mentioned earlier, the data yield a low figure for unemployment: only 5.3% of the labour force. It seems unlikely that this reflects the true probability of being unemployed in Pakistan. Underdeveloped economies, it is often thought, are characterised by high figures for unemployment. But in these economies, there being no form of unemployment benefits, and property and savings being luxuries specifically of the rich, not many can afford to remain unemployed.[15] Thus, rather than reported unemployment, disguised unemployment should be expected to be very high. Our somewhat unsatisfactory measure of underemployment seems to substantiate this hypothesis. For a proper assessment of all forms of unemployment, however, individual employment records need to be examined. In particular one would expect a lot of the "unemployed" to seek refuge in the informal sector, i.e. among the self-employed and the casuals.[16]

For this paper, however, we shall use only the subsample of the gainfully employed. The unemployed, unpaid family helpers and apprentices are dropped. This is possible because the focus of this paper is to study the determinants of the distribution of income or how individual characteristics affect individual incomes. The unemployed are not normally out of work as a result of some unidentifiable individual quirk. In fact it seems reasonable to assume that individuals with the same characteristics have the same probability of being unemployed.

[14] On the other hand this is what can be expected as the sample figure includes personal earnings only, while the Economic survey figure is a G.N.P. population ratio. Per capita incomes yielded by micro-data are normally lower than such macro calculations.

[15] Some recent labour force surveys have substantiated this point. Lower unemployment rates than 5.3% have been reported by these surveys See [15].

[16] Our sample may be reporting a low unemployment figure because it is drawn from structured and semi structured dwellings only. The unemployed being very poor lives in non-structured dwellings. But for reasons given in the text this statement would not be true. First our sample is most likely not unrepresentative of those living in non-structured dwellings. Secondly unemployment is just as much a luxury for these people as for those living in structured dwellings.

Table 1

*The Variables**

Original	*Derived (Regressors)***				
1. Age	X_1	=	Age		
	X_1^2	=	Age^2		
	X_2	=	Age1	1 if 16-19 years	0 otherwise
	X_3	=	Age2	1 if 20-24 years	0 otherwise
	X_4	=	Age3	1 if 25-29 years	0 otherwise
	X_5	=	Age4	1 if 30-34 years	0 otherwise
	X_6	=	Age5	1 if 35-39 years	0 otherwise
	X_7	=	Age6	1 if 40-44 years	0 otherwise
	X_8	=	Age7	1 if 45-49 years	0 otherwise
	X_9	=	Age8	1 if 50-54 years	0 otherwise
	X_{10}	=	Age9	1 if 55-59 years	0 otherwise
	X_{11}	=	Age10	1 if over 60 years	0 otherwise
2. Sex	X_{12}	=	Sex	1 if female	0 otherwise
3. Marital Status	X_{13}	=	Marital status	1 if single	0 otherwise
4. Education	X_{14}	=	Education 1	1 if Primary educated	0 otherwise
	X_{15}	=	Education 2	1 if Secondary educated	0 otherwise
	X_{16}	=	Education 3	1 if higher educated i.e. degree	0 otherwise
	X_{17}	=	Technical education 1	1 if any on-the-job training or apprenticeship.	0 otherwise
	X_{18}	=	Technical education 2	1 if greater than 6 month on-the-job training or apprenticeship	0 otherwise

Continued—

*In the analysis, age, education, sex and marital status are referred to as the set of basic variables. The human capital variables are age and education.

**The subcategories which have been excluded are: for age, the under 16; for sex, males; for marital status, the married; for education, the uneducated; for migration, the non-migrants; for employment status, the regular employees and for occupation, the production workers.

Table 1—*Contd.*

5. Migration	X_{19}	=	Migration 1	if 1-3 years of migration	0 otherwise
	X_{20}	=	Migration 2	1 if 4-6 years of migration	0 otherwise
	X_{21}	=	Migration 3	1 if 7-15 years of migration	0 otherwise
	X_{22}	=	Migration 4	1 if 16-30 years of migration	0 otherwise
	X_{23}	=	Migration 5	1 if migrant	0 otherwise
6. Employment	X_{24}	=	Emp. Status 1	1 if self-employed with no Employees	0 otherwise
	X_{25}	=	Emp. Status 2	1 if self-employed with employees	0 otherwise
	X_{26}	=	Emp. Status 3	1 if casual employees	0 otherwise
7. Occupation	X_{27}	=	Occupation 1	1 if professional and technical related workers	0 otherwise
	X_{28}	=	Occupation 2	1 if administrative and managerial workers	0 otherwise
	X_{29}	=	Occupation 3	1 if clerical and related workers	0 otherwise
	X_{30}	=	Occupation 4	1 if sales workers	0 otherwise
	X_{31}	=	Occupation 5	1 if services workers	0 otherwise
	X_{32}	=	Occupation 6	1 if agricultural workers	0 otherwise
8. Hours worked	X_{33}	=	Hours worked		
9. Years on the job.	X_{34}	=	Years on the job		
10. Number of unpaid helpers.	X_{35}	=	Number of unpaid family helpers		

SOME FINDINGS

The regression results are presented in Table 2.[17] Of the eighth, seven, have the logarithm of total monthly earnings as the dependent variable, while only one uses the log of the hourly wage rate. Of the number of different functional forms and specification that were tried the best are presented here. Different specifications with varying numbers of variables were used to identify the effects of particular individual characteristics or groups of individual characteristics on individual earnings. The results indicate that the majority of the variables used are significant in their effect on personal incomes. In regression 1, of

[17] Table 1 presents the definitions of the variables used in the analysis. As is evident since individual characteristics are mainly qualitative much use has been made of dummy variables. Age for example has been divided into 10 categories, education in 5, employment status in 5 and occupation in 6, i.e. from the 9 original characteristics 35 new variables have been derived. Following the theory of dummy variables the nth sub-category from each set of qualitative variables has been excluded.

Table 2

Regression Results

		Mean (Standard deviation)	Regression 1 log monthly earnings, all variables	Regression 2 log monthly earnings, age dummies, sex, education	Regression 3 log monthly earnings, human capital variables	Regression 4 log monthly earnings, basic variables hours worked	Regression 5 log monthly earnings, basic variables	Regression 6 log monthly earnings, non-basic variables	Regression 7 log hourly wage rate, all variables	Regression 8 log monthly earnings, non-human capital variables
X_1	Age	37.432 (13.675)	0.059 (0.006)		0.076 (0.006)	0.065 (0.0063)	0.064 (0.006)		0.059 (0.006)	
X^2_1	Age^2		—0.0007 (0.00007)		—0.0008 (0.00007)	0.0007 (0.00007)	—0.0007 (0.00007)		—0.0006 (0.00007)	
X^2	Age 16-19	0.046 (0.209)		—0.013 (0.101)						
X_3	Age 20-24	0.115 (0.319)		0.428 (0.082)						
X_4	Age 25-29	0.142 (0.349)		0.488 (0.078)						
X_5	Age 30-34	0.109 (0.332)		0.601 (0.082)						
X_6	Age 35-39	0.126 (0.332)		0.616 (0.082)						
X_7	Age 40-44	0.116 (0.320)		0.728 (0.082)						
X_8	Age 45-49	0.107 (0.309)		0.725 (0.083)						
X_9	Age 50-54	0.085 (0.279)		0.623 (0.086)						
X_{10}	Age 55-59	0.054 (0.227)		0.592 (0.096)						

Continued—

Table 2—*contd.*

		Mean (Standard deviation)	Regression 1 log monthly earnings, all variables	Regression 2 log monthly earnings, age dummies sex, education	Regression 3 log monthly earnings, human capital variables	Regression 4 log monthly earnings, basic variables hours worked	Regression 5 log monthly earnings, basic variables	Regression 6 log monthly earnings, non-basic variables	Regression 7 log hourly wage rate, all variables	Regression 8 log monthly earnings, non-human capital variables
X_{11}	Age Above 60	0.039 (0.194)		0.512 (0.104)						
X_{12}	Sex (Female)	0.060 (0.237)	—0.614 (0.064)	—0.714 (0.070)		—0.639 (0.066)	—0.723 (0.065)		—0.415 (0.063)	—0.666 10.0691
X_{13}	Marital Status (Single)	0.296 (0.457)	—0.126 (0.040)	—0.165 (0.044)		—0.163 (0.042)	—0.177 (0.043)		—0.091 (0.040)	—0.318 (0.035)
X_{14}	Edu. 1 Primary	0.039 (0.462)	0.130 (0.036)	0.132 (0.039)	0.154 (0.040)	0.121 (0.038)	0.120 (0.038)		0.115 (0.036)	
X_{15}	Edu. 2 Secondary	0.284 (0.451)	0.390 (0.044)	0.339 (0.041)	0.464 (0.041)	0.420 (0.039)	0.381 (0.039)		0.418 (0.044)	
X_{16}	Edu. 3 Higher	0.059 (0.235)	0.784 (0.073)	0.912 (0.067)	0.919 (0.070)	0.896 (0.068)	0.853 (0.068)		0.833 (0.073)	
X_{17}	Tech. Edu. 1	0.287 (0.453)								
X_{18}	Tech Edu. 2	0.241 (0.428)	0.057 (0.036)	—0.006 (0.038)		0.024 (0.036)	0.024 (0.036)		0.020 (0.037)	
X_{19}	Migration 1 1-3 Years	0.042 (0.201)	0.201 (0.073)						0.114 (0.073)	0.247 (0.079)
X_{20}	Migration 2 1 if 4-6 years	0.044 (0.206)	0.231 (0.060)						0.234 (0.069)	0.235 (0.075)
X_{21}	Migration 3 1 if 7-15 years	0.166 (0.372)	0.126 (0.039)						0.078 (0.039)	0.136 (0.042)

Continued—

Table 2—*Contd.*

		Mean (Standard deviation)	Regression 1 log monthly earnings, all variables	Regression 2 log monthly earnings, age dummies sex, education	Regression 3 log monthly earnings, human capital variables	Regression 4 log monthly earnings, basic variables hours worked	Regression 5 log monthly earnings, basic variables	Regression 6 log monthly earnings, non-basic variables	Regression 7 log hourly wage rate, all variables	Regression •8 log monthly, earnings non-human capital variables
X_{22}	Migration 4	0.654						0.112		
		(0.476)						(0.035)		
X_{23}	Migration 5 1 if Migrant	—	—	—	—	—	—	—	—	—
X_{24}	Employment	0.354	0.024					—0.062	0.003	—0.038
	Status : 1	(0.478)	(0.038)					(0.042)	(0.038)	(0.041)
X_{25}	Employment	0.044	0.585					0.760	0.463	0.628
	Status : 2	(0.204)	(0.074)					(0.082)	(0.074)	(0.081)
X_{26}	Employment	0.016	0.208					—0.437	—0.209	0.309
	Status : 3	(0.125)	(0.114)					(0.132)	(0.114)	(0.124)
X_{27}	Occupation : 1	0.109	0.174					0.339	0.272	0.479
	Professional and Tech.	(0.305)	(0.060)					(0.058)	(0.059)	(0.057)
X_{28}	Occupation : 2	0.119	0.439					0.857	0.404	0.785
	Administrative and Managerial	(0.136)	0.110)					(0.123)	(0.110)	(0.114)
X_{29}	Occupation : 3	0.164	0.072					0.317	0.137	0.291
	Clarical and Related	(0.370)	(0.052)					(0.050)	(0.052)	(0.049)
X_{30}	Occupation : 4	0.228	0.105					0.189	0.017	0.109
	Sales workers	(0.420)	(0.044)					(0.045)	(0.044)	(0.044)
X_{31}	Occupation : 5	0.112	—0.088					—0.180	—0.081	0.146
	Service workers	(0.315)	(0.051)					(0.057)	(0.052)	(0.053)
X_{32}	Occupation : 6	0.020	0.104					—0.052	0.017	0.021
	Agricultural	(0.139)	(0.106)					(0.122)	(0.106)	(0.113)

Continued—

Table 2—*Contd.*

	Mean (Standard deviation)	Regression 1 log monthly earnings, all variables	Regression 2 log monthly earnings, age dummies six, education	Regression 3 log monthly earnings, human capital variables hours worked	Regression 4 log monthly earnings, human capital variables	Regression 5 log monthly earnings, basic variables	Regression 6 log monthly earnings, non-basic variables	Regression 7 log hourly wage rate, all variables	Regression 8 log monthly earnings, non-human capital variables
X_{33} log Hours worked	5.279	0.190	0.165	0.386	0.267				0.164
	(0.338)	(0.045)	(0.049)	(0.041)	(0.047)				(0.049)
X_{34} Years on Job	11.055	0.010					0.011	0.008	0.009
	(10.293)	(0.002)					(0.002)	(0.002)	(0.C02)
X_{35} No. of unpaid family helpers	0.099	0.231					0.205	0.258	0.223
	(0.382)	(0.039)					(0.044)	(0.039)	(0.043)
Intercept	.	3.352	4.437	2.009	2.981	4.436	5.58	0.944	4.856
R^2		0.432	0.315	0.280	0.337	0.321	0.224	0.411	0.321
K		23	17	6	9	8	12	22	17
		1378	1378	1378	1378	1378	1378	1378	1378
$\overset{n}{\bar{R}^2}$		0.432	0.307	0.277	0.333	0.317	0.218	0.402	0.313
mean dependent variable	5.883							0.599	
St. dev. dependent variable	0.681							0.671	
mean income	446.789							2.284	

Note : Standard Errors in Parentheses.

the 23 variables used, 16 have coefficients significant at the 1% level of confidence, 1 at the 5% level and 2 at the 10% level, while only 4 are insignificant. A comparison of R^2 to R^{-2}, the adjusted R^2 shows that most of the variables used in the regression contribute to the explanation of the variation in the regression.

Two interesting features of the results are probably worth noting at the outset. Firstly observe the stability of the coefficients across regressions. The addition of new variables to a regression or the deletion of existing ones from it does not radically change the values of the coefficients. Secondly the regressions do quite well in explaining the variation in the dependent variable. Up to 43% of this variation is accounted for in the analysis (i.e. R^2 of 0.432 for regression 1). For cross section data such results are very reasonable.[18]

The regressions, because of their log form, must be read as giving for a unit increase in an independent variable a percentage change in personal earnings. Also because the constant term has not been constrained to equal zero, the regression coefficients associated with the dummy variables must be interpreted as giving the difference in personal earnings of an individual belonging to a particular category rather than to an excluded category, after holding all the other variables constant. In regression 1, for example, the coefficient of 0.271 for professional workers means that these individuals earn 27.1 percent more than the production workers, the excluded category. Similarly a 0.13 coefficient for the primary educated in regression 1 implies that these individuals earn 13 percent more than the uneducated.[19]

The functional form, regression of log of personal earnings on age and age^2 has been specifically used to capture any nonlinearities in the age-earnings profile.[20] The expectation is for the coefficients for age and age^2 to be positive and negative respectively; in short, a concave age-earnings profile. The results bear out this hypothesis. In each of the regressions the coefficients for age and age^2 are amongst the most significant and of the expected signs. Holding all other variables canstant, the age-earnings profile peaks at 42 years.[21] The income earned at this peak age however depends on the other characteristics of the individual; for example, the higher the education the higher the peak will be.

The above result however may have been constrained by the functional form that was used. If we did not have an everywhere dense normal age distribution, i.e. if there were gaps in the age distribution and the concentration was at the extremes, then this phenomenon could occur. In order to

[18] Compare this with Blaug's R^2 of 0.578 and this having a much larger information set then ours. Our results are also very reasonable when compared with other studies. The R^2 achieved by some of these are noted in Blaug [5] 'Ashenfelter and Mooney examined an extremely homogeneous group of recent Woodrow Wilson Fellows, obtained an R^2 of only 0.29 and observed "our equation does a very good job of explaining income differentials", [5].

[19] Note that we consider Regression 1 to be the best estimation as it yields the largest R^2 and has the greatest number of variables on the independent side. Most of the discussion in this section will centre around this equation.

[20] Blaug in his study on Thailand got an almost linear age-earnings profile.

[21] This is for regression 1. For other regressions the peak is at 45 years. The estimate for the first regression has been presented because it controls for the effects of all other characteristics whereas the other regressions do not.

confirm the concavity of the age-earnings profile, we ran a stepwise linear regression the results of which are presented in table 2. Here age was defined as a set of ten binary variables, one for each five-year age group. All age groups turn out to be significant and indicate a clear, concave age-earnings profile. Up to 45 years, the coefficient of each age group is greater than the previous one. After 45 the reverse holds. As the maximum age group 40-45 years is approached from either side, a limiting behaviour is observed—the difference between the coefficients start diminishing.[22]

Turning to the other human capital variable, education, we find that the coefficients for all the categories are significant. Income, as expected, is an increasing function of education, thus confirming the human capital hypothesis of education being a self-investment process. Successively, higher levels of education command increasingly higher levels of income, i.e. difference between the coefficients of higher and secondary levels of education is greater than that between the coefficients for secondary and primary education (0.394 and 0.260 respectively). Amongst the education categories, primary education has the smallest coefficient in all the regressions. The coefficients had however become small, insignificant and negative, when regression was tried for all members of the labour force including the unemployed apprentices and the unpaid family helpers.[23] This could be due to there being proportionately more of the primary educated amongst the unemployed and the other excluded classes than those from the other education categories.

Of the education coefficients, technical education is the only category that presents a surprising result. Being a human capital variable, the attainment of technical education was also expected to raise earnings. In all our regressions, however, the coefficient for this category is not significantly different from zero. This gives us the rather startling result that those with technical education earn no more than those without it. The only meaningful definition of this variable that the data would allow was more than six months of on-the-job training or apprenticeship. Formal technical education is available in very limited supply in Pakistan. Most skills are gained through a very long period of apprenticeship. It would therefore not be very surprising to find the drop-out rate to be very high. The regression coefficient may be indicating this phenomenon.

From the employment status variables, the excluded category is the regular employees. Of the included categories the self-employed with employees have the largest coefficient and the most significant. Individuals in this category (some sixty of them) earn about 58.5 percent more than regular employees. However, for these "richer" self-employed it is probably reasonable to expect that their earnings differential is a return on capital.

A somewhat surprising result is that for the self-employed without employees. These individuals, as it turns out, earn just about the same as the regular employees. The casual workers, however, according to expectation earn

[22] Ideally one should do a difference of the means test on the coefficients. However limitations of proper computer programming facilities prevented this. But rough calculations do show that the coefficients are all different.

[23] For the full sample of earners regression 5 was tried because this was the only specification permitted by the data. Remember we have no occupation and employment status record for these excluded classes.

Continued—

about 20 percent less than the regular employees. Roughly speaking the two categories, the self-employed and the casuals constitute what is known as the informal sector.[24] Theory suggests that individuals within these categories desire to get into the formal sector but market imperfections prevent them from doing so. For the self-employed higher earnings, it seems, is not one of the incentives for wanting regular employment. For the casuals however it must be noted that only those who admitted to having worked for at least the past week were admitted into the category and their work characteristics of the past week were recorded. No employment history for the past year or any long period was recorded. It would however not be unreasonable to assume that these workers would face periods of unemployment during a year. Average earning over the year for the casual should therefore be even lesser than those indicated by the regressions. Both the casuals and the self-employed however put in much longer hours than regular employees. The opportunity of any secondary employment is therefore severly limited for the former. Also these individuals are affected by the vagaries of the market whereas the regularly employed are not.

For the migrant, the results are both the most surprising and the most interesting. The expectation in this case was that migrants in their early days of migration would probably earn less than the residents but gradually upon acquiring location-specific human capital they would earn as much as if not more than the local residents. In the long run, migrants are expected to earn more than the residents, for migration, it must be remembered, is a self-selection process whereby only the more dynamic and the more capable migrate. Our results however show the recent migrants to be earning more than those who have settled in the city: the new arrivals about 20.1 percent more than the

Footnote 23 *Contd.*

The results are:

$$\begin{array}{llllll}
\text{Log } y = 0.418 + & 0.221 \text{ Age}, & -0.002 \text{ Age}^2, & -1.344 \text{ Sex} & -0.619 \text{ marital status} & \\
 & (0.017) & (0.0002) & (0.176) & (0.125) & \\
 & -\,0.0006 \text{ Edn1} & +\,0.666 \text{ Edn2} & +\,1.001 \text{ Edn3} & + & 0.195 \text{ Tech. Edn.} \\
 & (0.109) & (0.152) & (0.206) & & (0.104)
\end{array}$$

$n = 1641$; $R^2 = 0.337$, $F = 103.57$
mean y = 376.436 std. dev. log y = 2.185
where y = individual monthly earuings

These results are reliable to the extent that the probability of beng unemployed geiven by the sample reflects the actual probability. The coefficients of X_{15}, X_{16} and X_{18} now all rise in value. This is probably because inclusion of the low income groups has now decreased the value of intercept term. The mean income is also now lower than before. The coefficient of the primary educated X_{14} has now not only decreased drastically in value but is also negative and insignificant. This indicates a larger proportion of the primary educated amongst the unemployed and the excluded categories than of the other education groups. Alternatively the primary educated have a higher probability of being unemployed than the others and/or a higher propensity to apprenticeship work and unpaid family work.

[24] In the analysis, frequent reference will be made to the dual economy theory. The informal sector we consider to be roughly equivalent to the subsample of the self-employed and casual workers, while the formal sector constitutes the regular employees. The self-employed with employees are not stressed in either of these sectors as some of these have capital enough to be classified in the formal sector. We use these rough classifications as these correspond most closely to most definitions of the dual markets. By most definitions the bulk of the self-employed and the casuals will be classified in the informal sector while the bulk of the regular employees in the formal sector. The classification rule for the dual markets is a controversial affair [12] and we do not wish to get involved in it. Our results for the dual markets will therefore be just as acceptable as our classification rule.

inhabitants of the city, the relatively settled 23.1 percent more while the settled ones only 12 percent more. An explanation of these coefficients can be the increasing economic importance of Rawalpindi in recent years because of the development of the new capital on its outskirts. The city in the recent past has been offering better opportunities and thus providing incentives for better qualified people to migrate to it. Therefore existing earnings could be a reflection of differing market conditions at the time of migration.

Another surprising fact about migrants is revealed by the following cross-tabulation, in Table 3.

Table 3

Distribution of Migrants Amongst Employment Categories

	Regular Employed	Self-Employed without employees
Migration 1-3 years	84.5%	15.5%
Migration 4-6 years	77.6%	22.4%
Migration 7-15 years	56.2%	38.8%

The dual economy suggests that most of the migrants, especially the recent ones, would be found in the informal sector which by definition corresponds to our self-employed with employees. The above results however contradict this hypothesis. A large majority of the recent migrants, it seems, are able to find regular employment. This could be indicative of any number of things. For example, are these immigrants more productive than the others to be able to get regular jobs so quickly? Is it a result of the government's regional policy that these people get better jobs? Is the earning pattern indicative of this distribution amongst the sectors?

Returning now to the results in table 2, note that we are using the unconstrained wage rate specification in our regressions. Economic theory suggests that the wage rate be used as the dependent variable for it is this variable on which the individual labour supply decisions are based[25]. Although we do not use the wage rate explicitly as a dependent variable, controlling for hours worked in essence means that we are calculating a wage rate equation. An additional benefit is that we get a coefficient for hours worked.[26] A comparison of regressions 1 and 6 illustrates the above point. Both regressions use

[25] Each individual works upto the point where his marginal rate of substitution of leisure for money is equal to the wage rate.

[26] Another reason for using this specification is that we are not sure of our hours worked variable. In fact there is reason to believe that there may have been some misreporting.

the same number of variables and the same functional form; the only difference being in the dependent variable. It is evident that differences; where they exist, are not alarming.

The coefficient of hours worked, the results indicate, is 0.190. An elasticity coefficient, this should be interpreted as a 10 percent increase in the number of hours worked producing a 1.90 percent increase in income.[27] Decreasing returns to extra hours worked are indicated. Had the coefficient been equal to one we would have had an exact wage rate equation, and proportional returns to working additional hours. The low value of the coefficient is probably an indication of the costs of leisure being very high: people are probably willing to work at differing and lower wage rates in order to maximise total earnings.[28]

Amongst occupations, the highest paid are administrative and managerial workers. These are followed by professional and technical workers, sales workers, agricultural workers (mainly dairy workers), clerical and related workers and the service workers. The last mentioned are the only ones with a negative co-efficient. The excluded sub-category was that of production workers. A surprising result is that clerical, service and production workers should earn as much as dairy workers. A part of the dairy workers' earnings is however return on dairy animals. In regression 6 however all other occupations become significant except for agricultural workers, service workers it turns out earn definitely less than production workers.[29]

As expected the coefficient for both females and singles are negative. For the former the result is probably indicative of discrimination while for the latter of self-selection process. For females in fact a further testable hypothesis would be that not only does the market discriminate against them in terms of salaries and wages but also in terms of the number of employment opportunities open to them.

[27] Income is a concave function of hours worked.

[28] Another reason for this low value for the log of hours worked coefficient is that the regression includes only those who are gainfully employed. The observations are all centred around the mean hours worked. We may therefore be capturing only a small segment of the hours worked-income curve. Introduction of the part time workers and the unemployed may raise the value nearer to one. The following regression rather imperfectly illustrates this point.

$$\begin{array}{llllll}
\log y = & 2.688 & + \underset{(0.015)}{0.174\ \text{Age}} & - \underset{(0.0002)}{0.002\ \text{Age}^2} & - \underset{(0.122)}{1.084\ \text{Sex}} & \\
 & - \underset{(0.108)}{0.450} & \text{marital status} & \underset{(0.095)}{0.263\ \text{Edn1}} & + \underset{(0.101)}{0.850\ \text{Edn2}} & \underset{(0.178)}{1.33\ \text{Edn3}} \\
 & + \underset{(0.90)}{0.158} & \text{Tech Edn} + & \underset{(0.032)}{0.769\ \text{log hours worked}} & & \\
n = & 1641 & R^2 = 0.508 & F = 186.79 & &
\end{array}$$

where y=individul monthly earnings

This is the same regression as presented in footnote 26. The sample of observations is the same. The only difference is the inclusion of the $\log_{33}$ (hours worked) on the independent side of the equation. The coefficient for this new variable is now 0.769. i.e. nearer one thus proving the point made above. This has happened probably because we are now trying to constrain the hours worked income to pass through the origin by including some observations with zero income and no hours worked. If we had some intermediate observation on part-time workers, the coefficient may move even closer to one.

[29] In Thailand Blaug had an interesting result for occupations. In this analysis amongst the highest paying occupations was the military. The political and military climate in most underdeveloped countries seem to support this thesis. However for lack of any data we cannot firmly ascertain this.

An additional unpaid family helper adds about 23 percent to an individuals income. This figure however may be taken as an indication of the helpers productivity and therefore of their contribution to family income.

A comparison of regression 4 to regression 1 shows that most of the explained variance of earnings is accounted for by the basic variables, age, sex, education and marital status.[30] The addition of employment status migration and occupation to these variables raises the R^2 from 0.337 to 0.432, i.e. an addition of only about 0.095 to the explanation. Regression 3 using only the human capital variables age and education yields an R^2 of 0.280. About 64.8 percent of the explained variance is therefore due to these variables. It seems therefore that an important conclusion of the analysis would be that a large portion of the inequality in earning is attributable to differing levels of human capital that individuals acquire.

The above method however overestimates the importance of the human capital variables. If the subset of the human capital variables was orthogonal to the rest of the variables, the above result would then be correct. In that case the correlation matrix would be block diagonal. Such an assumption however is unseasonable and also not-supported by the data. In other words we have a full correlation matrix. Our estimate is in fact the upper limit of the true parameter. The lower limit can be obtained by reversing the procedure used above. This would involve regression all variables other than the human capital, on monthly income and comparing the explained sum of squares of this regression with that of regression 1 in table 2. This latter regression contains the complete set of independent variables. The former regression is regression 8 in table 2.

The human capital variables therefore contribute at least 25.7 percent to the explanation of the variable in income. Similarly for the basic variable, i.e. age, sex, education and marital status, the upper and lower limits are 78 percent and 45.3 percent respectively. The true value lies between these limits. On the average, the human capital variables explain about 45.25 percent of the income distribution and the basic variables 61.65 percent. Clearly these two subsets take up a large proportion of the explanation of the distribtuion of income.

The size of the least squers coefficient in the classical linear model, it is well known, is not a reliable measure of the relative importance of an independent variable in determining the variation in the regressioned. The size of these coefficients can easily be varied by changing the units of measurement. Of the three objective measures of the size of a coefficient we have chosen the beta-coefficients [7]. Table 4 presents the beta-coefficients for 3 of our regressions. The coefficients have been ranked according to size. The ranking reveals age and age^2 as being the foremost in affecting income distribution followed closely by the two education categories, secondary and higher. Thereafter the rankings show sex, the richer self-employed, unpaid family helpers, years on the job, hours worked, primary education, administrative occupation, marital status, migration 2, and the remaining occupation and employment status variables.

[30] In fact the most important variables are age, education and sex in that order. The effect of marital status is not very significant as indicated by the beta coefficients.

Table 4

Beta—Coefficients for Selected Regressions

	Regression 1 log monthly income	Rank	Regression 4 log monthly income	Rank	Regression 7 log wage rate	Rank
X_1 Age	1.176	1	1.304	1	1.198	1
X^2_1 Age^2	—1.123	2	−1.129	2	−1.109	2
X_{12} Sex (Female)	—0.213	5	−0.222	5	−0.215	5
X_{13} Marital Status (Single)	—0.084	12	−0.109	7	−0.062	14
X_{14} Education 1 Primary	0.088	10	0.082	8	0.080	11
X_{15} Education 2 Secondary	0.258	4	0.278	4	0.281	4
X_{16} Education 3 Higher	0.271	3	0.310	3	0.293	3
X_{18} Tech. Education 2	0.036	21	0.015	9	0.026	20
X_{19} Migration 1 1-3 Years	0.059	17			0.034	18
X_{20} Migration 2 1 if 4-6 Years	0.070	14			0.072	13
X_{21} Migration 3 1 if 7-15 „	0.069	15			0.044	15
X_{24} Employment Status 1	0.017	23			0.002	27
X_{25} Employment Status 2	0.173	6			0.141	7
X_{26} Employment Status 3	—0.038	20			−0.039	16
X_{27} Occupation 1 professional	0.078	13			0.124	9
X_{28} Occupation 2 Adm and Management	0.078	11			0.082	10
X_{29} Occupation 3 Clerical and related	0.039	19			0.076	12
X_{30} Occn 4 Sales workers	0.064	16			0.011	21
X_{31} Occn 5 Service workers	0.041	18			0.038	17
X_{32} Occn 6 Agricultural	0.021	22			0.024	19
X_{33} Hours worked	0.094	9	0.132	6		
X_{34} Years on Job	0.157	7			0.126	8
X_{35} No of unpaid family helpers.	0.129	8			0.144	6

Thus, again we find that age and education, the two human capital variables, are the most important in affecting the level of earnings of an individual and hence the distribution of income.

SOME ADDITIONAL FINDINGS

The regression results of the last section revealed that the bulk of the self-employed, i.e., those without employees (SE1) earn no more than the regular employees (RE). The self-employed with employees (SE2) make about 58 percent more and the casual employees (CE) make 20 percent less than the RE. In the rankings of the beta-coefficients, Coefficients for SE1 and CE were among the lowest. There is no indication however, of how well the model used explains the variation in earnings within these categories. The importance of the human capital variables in the previous results may have been due to the larger proportion of the RE in our sample (60%). For the SE and the CE, possibly the structural variables or some other unidentified forces may explain the income differential. The acquisition of human capital, especially education, may be more important to individuals in regular employment than in self-employment.[31]

There is therefore a need to examine in detail, the results of the previous section especially with regard to employment status. For this reason four subsamples were selected, one for each employment status.[32] For each sub-sample then, two regressions were run. The results are presented in Table 5 and the salient features of these are discussed below:

(*i*) Note first that in the CE category there are very few observations —too few in fact for a meaningful analysis—The high R^2's are not as much an indication of a good fit as of the lack of a reasonable number of observations. At best the results for this category are highly unreliable and will not be stressed.

(*ii*) Not surprisingly, the model fits the RE the best and the SE1 the least. In the earlier regressions for all the earners, therefore the greater part of the explanation was due to the presence of the RE. The R^2 of 0.361 for the SE, however, is better than expected. The model used thus explains well the variation in earnings within this category.

(*iii*) The age and education coefficients turn out as expected for each of the subsamples. The age—earnings profiles are all concave. The peaks are at 45 for the RE and the CE at 46 the SE1 and 50 for the SE2. The coefficient of age^2 is a relative measure of the peakedness or the flatness of the profile. Interestingly enough and true to hypothesis, the age earnings profile is the flattest for the CE and the SE2 followed by the SE1 and the most peaked for the RE.

[31] Remember investment in human capital raises an individuals earnings by raising his productivity. Employers will be willing to pay for higher productivity but customers at a shop may not. The possibility of the person using his human capital to improve his business however always remains. In this case the human capital hypothesis would hold amongst the SE too. Human capital is therefore expected to raise productivity in both sectors. Amongst the self-employed however education is not a prerequisite for entry whereas in some forms or regular employment it is. The bias is therefore evident.

[32] Alternatively, we could have used interaction terms but this procedure would have multipled to an uncontrolable limit the number of variables

Table 5

Result of Subsample Regressions

		Regular Employees (RE)			Self employed without employees (SEI)		
		Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variable	Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variables
X_1	Age	34.986 (12.46)	0.063 (0.008)	0.070 (0.008)	40.74 (14.836)	0.046 (0.010)	0.049 (0.010)
X_2	Age²	— —	−0.0007 (0.00009)	−0.0008 (0.00009)	— —	−0.0005 (0.0001)	−0.0005 (0.0001)
X_{12}	Sex (Female)	0.062 (0.241)	−0.459 (0.069)	−0.420 (0.067)	0.057 (0.233)	−0.755 (0.130)	−0.829 (0.131)
X_{13}	Marital Status (Single)	0.341 (0.414)	−0.086 (0.043)	−0.111 (0.044)	0.242 (0.429)	−0.226 (0.079)	−0.257 (0.081)
X_{14}	Education (Primary)	0.293 (0.455)	0.171 (0.044)	0.219 (0.044)	0.326 (0.469)	0.109 (0.064)	0.084 (0.065)
X_{15}	Education (Secondary)	0.388 (0.488)	0.458 (0.049)	0.563 (0.043)	0.129 (0.336)	0.262 (0.092)	0.186 (0.090)
X_{16}	Education (Higher)	0.083 (0.276)	0.848 (0.074)	1.041 (0.065)	0.018 (0.161)	0.746 (0.215)	0.530 (0.214)
X_{18}	Technical Education	0.260 (0.439)	0.024 (0.038)	0.018 (0.036)	0.205 (0.404)	0.130 (0.080)	0.093 (0.071)
X_{19}	Migration 1 1-3 Years	0.061 (0.239)	0.177 (0.067)	— —	0.018 (0.135)	0.110 (0.209)	— —
X_{20}	Migration 2 4-6 Years	0.056 (0.229)	0.283 (0.068)	— —	0.027 (0.161)	0.079 (0.174)	— —
X_{21}	Migration 3 7-15 Years	0.159 (0.366)	0.099 (0.043)	— —	0.182 (0.387)	0.084 (0.074)	— —
X_{22}	Migration 4 16-30 Years						

Continued—

Table 5—*contd.*

		Regular Employees (RE)			Self employed without employees (SEI)		
		Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variable	Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variables
X_{27}	Occupation 1 Profession and technical	0.145	0.196	—	0.041	−0.345	—
		(0.352)	(0.059)	—	(0.198)	(0.156)	—
X_{28}	Occupation 2 administrative and Managerial	0.020	0.475	—	0.012	0.498	
		(0.139)	(0.118)	—	(0.110)	(0.256)	—
X_{29}	Occupation 3 Clerical	0.277	0.026	—	0.002	−0.046	—
		(0.448)	(0.048)	—	(0.045)	(0.612)	—
X_{30}	Occupation 4 Sales work	0.074	−0.036	—	0.480	0.131	—
		(0.262)	(0.064)	—	(0.500)	(0.075)	—
X_{31}	Occupation 5 Service work	0.142	−0.090	—	0.041	−0.125	—
		(0.349)	(0.052)	—	(0.198)	(0.126)	—
X_{32}	Occupation 6 Agricultural work etc.	0.007	−0.111	—	0.105	0.122	—
		(0.086)	(0.180)	—	(0.306)	(0.150)	—
X_{33}	Hours worked	5.219	0.070	0.041	5.361	0.477	0.564
		(0.349)	(0.046)	(0.047)	(0.307)	(0.100)	(0.101)
X_{34}	Years on the job	9.342	0.011	—	13.492	0.006	—
		(8.472)	(0.002)	—	(12.320)	(0.003)	—
X_{35}	No of unpaid family helpers	—	—	—	0.232	0.261	—
		—	—	—	(0.557)	(0.051)	
	Intercept	—	19	9	—	20	9
	K	—	3.852	3.94	—	2.172	1.839
	R^2	—	0.501	0.453	—	0.361	0.288
	n	—	809	809	—	488	488
	F	—	41.76	73.54	—	13.20	21.53
	R^2	—	0.490	0.448	—	0.335	0.276
	Mean Income	—	417.216	—	—	440.77	—
	Std. dev. Log Income.		0.601	—	—	0.739	—

Continued—

Table 5—Contd.

		Self employeed with employees (SE2)			Casual Employees (CE)		
		Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variables	Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variables
X_1	Age	46.1 (9.605)	0.130 (0.078)	0.076 (0.073)	30.130 (12.440)	0.145 (0.066)	0.184 (0.081)
X_2	Age^2	— —	−0.0013 (0.0008)	−0.0006 (0.0008)	— —	−0.0016 (0.009)	—0.002 (0.001)
X_{12}	Sex (Female)	— —	— —	— —	0.174 (0.388)	−1.767 (0.451)	—0.978 (0.574)
X_{13}	Marital Status (Single)	0.033 (0.180)	−0.069 (0.434)	0.070 (0.427)	0.522 (0.511)	0.076 (0.412)	0.424 (0.536)
X_{14}	Education (Primary)	0.311 (0.467)	0.219 (0.185)	−0.042 (0.172)	0.478 (0.511)	−0.631 (0.224)	—0.183 (0.302)
X_{15}	Education (Secondary)	0.197 (0.401)	0.700 (0.259)	0.214 (0.210)	0.087 (0.288)	−0.404 (0.560)	0.417 (0.572)
X_{16}	Education (Higher)	0.082 (0.473)	0.582 (0.434)	0.539 (0.286)	— —	— —	— —
X_{18}	Technical Education	0.246 (0.434)	0.234 (0.207)	−0.020 (0.173)	−0.304 (0.470)	0.187 (0.447)	0.236 (0.424)
X_{19}	Migration 1 1-3 Years	— —	— —	— —	— —	— —	— —
X_{20}	Migration 2 4-6 Years	0.049 (0.218)	0.651 (0.343)	— —	0.130 (0.344)	0.542 (0.408)	— —
X_{21}	Migration 3 7-15 Years	0.098 (0.300)	—0.033 (0.290)	— —	0.217 (0.422)	0.475 (0.459)	— —
X_{22}	Migration 4 16-30 Years						

Continued—

Table 5—*Contd.*

		Self employed with employees (SE2)			Casual Employees (CE)		
		Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variables	Mean (Std. dev).	Log monthly income All variables	Log monthly income Basic variables
X_{27}	Occupation 1 Profession and technical	0.066	0.077	—	0.087	2.370	—
		(0.250)	(0.459)	—	(0.288)	(0.578)	—
X_{28}	Occupation 2 administrative and managerial	0.066	0.465	—	0.0	—	—
		(0.250)	(0.452)	—	—	—	—
X_{29}	Occupation 3 Clerical	0.016	0.191	—	0.0	—	—
		(0.128)	(0.580)	—	—	—	—
X_{30}	Occupation 4 Sales work	0.344	—0.069	—	—	0.087	0.138
		(0.479)	(0.198)	—	—	(0.288)	(0.397)
X_{31}	Occupation 5 Service work	0.155	—0.332	—	—	0.087	−0.277
		(0.321)	(0.257)	—	—	(0.288)	(0.608)
X_{32}	Occupation 6 Agricultural work etc.	0.016	1.206	—	—	0.0	—
		(0.539)	(0.539)	—	—	—	—
X_{33}	Hours worked	5.461	0.350	0.031	5.265	0.218	0.495
		(0.223)	(0.414)	(0.039)	(0.297)	(0.576)	(0.613)
X_{34}	Years on the job	15.607	0.021	—	8.00	−0.023	—
		(10.190)	(0.009)	—	(6.941)	(0.021)	—
X_{35}	No of unpaid family helpers	0.377	0.051	—	0.0	—	—
		(0.489)	(0.114)	—	—	—	—
	Intercept	—	1.242	4.329	—	2.771	−0.731
	K	—	18	8	—	13	8
	R^2	—	0.442	0.193	—	0.857	0.541
	n	—	61	61	—	23	23
	F	—	1.85	1.56	—	4.13	2.06
	R^2	—	0.221	0.086	—	0.685	0.327
	Mean Income	—	982.95	—	—	270.22	—
	Std. dev. Log Income.	—	0.576	—	—	0.777	—

Note : Standard Errors in Parentheses.

(*iv*) Education has the largest and most significant effect on earnings for the RE such sample. Surprisingly however, the education coefficients are all highly significant in the SE1 regressions. Increasing returns to higher levels of education are also indicated and Schooling has a similar effect across the employment catetgories, i.e. the human capital hypothesis on schooling holds true for all employment statuses in our sample. At each schooling level however the RE earn approximately 10 percent more than the SE.

(*v*) Most of the SE1 and SE2 are sales workers, while the RE are mainly clerical workers. In view of the description of Rawalpindi given above, this result is hardly surprising. The most paying occupation however is still the administrative and managerial and the vast majority of these are classified as regular employees.

(*vi*) The mean income is the highest for the SE2 but for these individuals we know a part of their income is a return on capital. Between the SE1 and the RE, the former have a slightly higher mean income, (a difference of Rs. 25.6). It was because of this difference that in the full sample regression the coefficient for the SE1 was small, positive and insignificant.

(*vii*) Inequality as given by the standard deviation of the logarithm of incomes is the least amongst the SE 2 and the most amongst the CE. Income amongst the SE 1 however is more unequally distributed than among the RE. For the SE 1 and the CE the inequality index is greater than the equivalent index for the full sample.

(*viii*) As mentioned earlier the SE 1 and the CE constitute the informal sector while the RE the formal. Given this our results serve to throw some light on the dual economy hypothesis.[33] According to this theory there are barriers to entry into the formal sector those in the informal sector. The latter would like to move into the former sector for incentives like higher earnings, shorter working hours, job security etc. According to our results however they may not earn more in the formal sector. But as expected incomes are more equally distributed in the formal sector than in the informal.

(*ix*) One of the derived variables in the data was the formal/informal. The formal sector was defined as, the government and municipality employees, the professionals, employees of large scale manufacturing and other firms (more than 20 employees).[34] A cross tabulation of this variable with the employment status is shown in Table 6.

[33] This is not intended to be a verification of the dual economy hypothesis. We only provide evidence for some facts of this hypothesis which are incidental to our analysis. We have no evidence for the existence or non-existence of barriers to entry into the formal sector.

[34] This definition can easily be criticised on grounds that it classifieds all those in the higher income brackets in the formal sector. See footnote (24).

Table 6

	ES1	ES2	RE	CE
Formal	27%	13.3%	70.7%	9.1%
Informal	97.3%	86.7%	29.3%	90.9%

There are 29.3% of RE who should be classified in the informal sector. To be extent that these people are migrants and low-income-earners the above results will be biased.

(x) In the previous section, it was shown that most of the explained variation in incomes is due to the human capital variables. Surprisingly enough this is true for all the employment status categories too. As before, upper and lower limits for the true proportion of the explained variation taken up by the human capital variables have been estimated. The Regression results are presented in tables 5 and 7. The estimated limits are presented in Table 8.[35]

Table 7

Summary Result for Some Subsample Regressions

		R^2	$\bar{R}^2$	F
Regular Employees	*Regression* 1 Log monthly earnings on the human capital variables.	0.422	0.418	97.437
	Regression 2 Log monthly earnings on variables other than human capital.	0.339	0.328	26.06
	Regression 3 Log monthly earnings on the non-basic variables.	0.261	0.250	34.450

Continued—

[35] Results for the CE are not presented for reasons noted in (i)

Table 7—Contd.

	Regression 4			
	Log monthly earnings on the human capital variables.	0.208	0.198	21.023
	Regression 5			
Self-employed without Employees	Log monthly earnings on variables other than human capital.	0.300	0.279	14.489
	Regression 6			
	Log monthly earnings on the non-basic variables.	0.196	0.175	9.624
	Regression 7			
	Log monthly earnings on the human capital variables.	0.204	0.114	2.267
	Regression 8			
Self-employed with Employees	Log monthly earnings on variables other than human captial.	0.316	0.159	2.020
	Regression 9			
	Log monthly earnings on the non-basic variables.	0.316	0.141	1.806

Table 8

	Proportion of explained variation attributable to human capital variables.		Proportion of explained variation attributable to the basic variables.	
	Upper limit	Lower limit	Upper limit	Lower limit
RE	84.2%	33.3%	90.4%	47.9%
SE1	57.6%	16.9%	79.7%	45.7%
SE2	46.2%	28.5%	43.7%	28.5%
CE	—	—	—	—

The human capital variables are of overwhelming importance for the RE subsample. But the interesting fact is that in the other subsamples too these variables have a large effect on individual earnings. On an average approximately 40% of the explained variation in these categories is due to these variables. The addition of sex and marital status increases the explanation significantly only among the SE. For the self-employment the sex and unpaid family helpers have larger beta-coefficient than some of the education dummies. For regular employees however the human capital variable are much stronger. As before marital status, is the least important of the basic variables.

CONCLUSIONS

It is quite possible when using ordinary least squares in the classical linear model to get spurious results. In particular biased results are likely to result if (a) the regression errors are not randomly distributed with zero mean and constant variance and (b) if these errors are not uncorrelated with each other or with the explanatory variables.[36] These doubts however were easily dispelled when the residuals in standardised form, for our main regression (regression 1) were plotted against the estimated value of income, also standardised. A symetrical, spherical pattern emerged indicating that there was no departure from the assumptions of homoskedasticity and independence.

When using such a large number of variables especially as most of them are in the binary form, there is a likelihood of multicolinearity affecting the results. The presence of this problem would bias both the size of the regression coefficients and the value of their standard errors. No special tests for colinearity were conducted but several checks were made to detect its presence. First an inspection of the correlation matrix for the independent variables showed that there was no cause for alarm. Second each of the independent variables was in turn regressed on the other independent variables to test for linear dependence. The R^2 was mostly below 0.30 and never more than 0.55. Third, the size of the regression coefficients was not changed much by the addition or deletion of variables from regressions. The presence of multicolinearity on the other hand, would have caused violent changes in the size of the coefficients with changes in the number of variables. We may conclude, therefore, that multicolinearity does not affect our results.

From our analysis we may therefore conclude that the human capital variables explain a large part of the income differential.[37] Age and education

[36] On the other hand our results may be suffering from a simultaneous equation bias. The occupation, employment status, education etc., may not all be predetermined variables but some may be choice variables. We would therefore have a simultaneous equation model on our hands of which we have estimated one under-identified equation. In this case the wrong estimation technique ordinary least squares was used. It is not however clear what the correct simultaneous equation model is. A lot would depend on what assumptions one makes, which variables are the endogenous ones etc. For our purposes however the single equation model is all right for we take the characteristics of the individual as given and ask the question how these effects his income.

[37] Generalisations beyond the city however must be made bearing in mind the socio-economic description of the city given in section 3. The lack of large scale manufacturing industry and the unusually large presence of the government sector in the form of Islamabad the capital probably biases the results in favour of the human capital variables.

both affect income as expected. Given greater equality in education therefore it would be reasonable to expect more closely grouped together Individual age-earnings profiles and therefore reduced inequality. Returns to education followed the same pattern across sectors; increasing returns to increasing levels of education. Education however commanded the greatest premium among the regular employees. At each level of education these individuals earned at least 10 percent more than the self-employed and the casuals. Interestingly enough there were positive and significant returns to education everywhere and successive levels of education commanded a higher premium every where.

For the dual market hypothesis our results show that individuals in the informal sector earn about as much or may be slightly more than those in the formal. Income however is more unequally distributed in the former sector than in the latter. The barriers to entry if any however we cannot identify in this analysis. But it seems as if there is no financial incentive for moving from the informal sector to the formal. Two incentives for such a move can however be identified from our analysis; (a) shorter working hours and hence opportunity for secondary employment, (b) a more equally distributed guaranteed monthly earnings. There are however a number of other incentives which one can hypothesis like freedom from the vicissitudes of the market, old age pension, paid holidays and other such employment benefits.

The results for migration were interesting enough to deserve a mention here. The pattern of earnings for migrants was concave; the relatively settled earned the most followed by the recent arrivals and "the settled". However migrants every where earned more than the local residents. Contrary to the dual economy theory the majority of the migrants were found in regular employment. In fact 85 percent of the recent migrants are in regular employment.[38] These individuals are not therefore as theory predicts using the informal sector as a point of entry.

To conclude therefore, our main result is the applicability and importance of the human capital variables in explaining the income differential. Both work experience and higher levels of education seem to command a premium in the market. But do the educated earn, more because they are more productive? or more able? or because they happen to belong to influential families? Unfortunately none of these questions can be answered with any degree of accuracy by our analysis. Data on family background was not available, and so the effects of parental education, income etc. on earnings and education could not be observed. Ability as is obvious would be an exceedingly difficult variable to measure. No assessments of the filtering process of education, i.e., the sorting out of the more able from the lesser can therefore be made.

[38] As noted earlier, a large proportion of recent migrants may be found unstructured dwellings. An inclusion of these may lower the figure of 85%.

REFERENCES

1. Anand, S. *The Size Distribution of Income in Malaya.* Oxford University (Unpublished).
2. Becker, Gary S. *Human Capital.* New York: Columbia University Press, 1974.
3. ——————. "Human Capital and the Personal Distribution of Income". *Journal of Economic Literature.* March 01, 1970.
4. Bergen, Asbjorn. "Personal Income Distribution and Personal Savings in Pakistan 1963-64". *The Pakistan Development Review.* Vol. VII, No. 2. Summer 1967.
5. Blaug, M. "An Economic Analysis of Earnings in Thailand". *Economic Development and Cultural Change.* Vol. 23. No. 1, October 1974.
6. Champernowne, D. G. "A Model of Income Distribution". *Economic Journal.* June 1953.
7. Goldberger, A. S. *Econometric Theory.* New York: John Wiley and Sons, Inc.
8. Hamdani, K. "Education and the Income Differential: An Estimation for Rawalpindi City". *The Pakistan Development Review.* Vol. XVI, No. 2. Summer 1977.
9. Mandelbrot, P. "The Pareto Levy Law and the Distribution of Income". *International Economic Review.* May 1960.
10. Mincer, J. "The Distribution of Labour Incomes Survey". *Journal of Economic Literature.* March 01, 1970.
11. Morgan, James. "The Anatomy of Income Distribution". *Review of Economics and Statistics.* Vol. 44, No. 3, August 1962.
12. Mazumdar, D. *The Urban Informal Sector.* World Bank Staff Working Paper No. 211,
13. Ojha, D.P. and V.V. Bhatt. "Pattern of Income Distribution in an Underdeveloped Economy: A Case Study of India". *American Economic Review.* September 1964.
14. Sen, A.K. *On Income Inequality.* Oxford Clarendon Press, 1973.
15. Shah, Nasra M., Makhdoom A. Shah and Tauseef Ahmed. *Labour Force and Unemployment Statistics in Pakistan. in* Pakistan Manpower Institute. *Manpower and Employment Statistics in Pakistan.* Islamabad, 1977.