Open access • Journal Article • DOI:10.30541/V33I1PP.1-18

Mincerian Earnings Function for Pakistan — Source link

Tayyeb Shabbir

Published on: 01 Mar 1994 - The Pakistan Development Review (Pakistan Institute of Developemnt Economics (PIDE))

Topics: Earnings

The Pakistan Development Review
33:1 (Spring 1994) pp. 1–18

Mincerian Earnings Function for Pakistan

TAYYEB SHABBIR

Due to its central role in various debates about the determinants of individual earnings, the Mincerian earnings function (MEF) as given in Mincer (1974) has attracted the attention of many economists. The MEF has been estimated virtually for every country except Pakistan, where a necessary condition has been missing, i.e., national level data on the *exact* number of years of schooling completed has not been available; instead, in a majority of the relevant micro-level surveys, schooling has been measured only in terms of a 'categorical' variable with possible values being 'Primary and Incomplete Middle', 'Middle and Incomplete Matric', etc. At best, this data deficiency has restricted the existing estimated earnings functions to what we refer to as the 'Dummies earnings functions' (DEF) since they are constrained to specify schooling in terms of a set of dichotomous dummy variables.

Using a nationally representative data on male earners, this study tries to fill the above gap by estimating the MEF both in its 'strict' as well as the 'extended' forms. In terms of the 'strict' MEF, i.e., the one analogous to Mincer's (1974) specification which essentially treats earnings as a function of schooling and job-market experience, the main findings are that the marginal rate of return to schooling is 8 percent, the experience-earnings profile is consistent with the pattern suggested by the human capital theory and as much as 41 percent of the variance in log earnings is accounted for by the strictly defined MEF. By and large, these findings are consistent with those implied by estimated MEFs for comparable LDCs. Further, the present study also estimates 'extended' MEF, whose specification supplements that of the 'strict' MEF by adding variables to control for urban vs rural background, occupational categories, employment status, and provincial heterogeneity. The 'extended' MEFs are also estimated separately for urban and rural samples and for each province. Formal 'Chow-type *F* tests' conducted to test for homogeneity of the parameters of MEF across different sub-samples reveal 'pervasive' segmentation across the above strata.

1. INTRODUCTION

The Mincerian earnings function (MEF), as implied by the analytical model given in Mincer (1974), is one of the standard specifications that have been estimated for virtually every country of the World.[1] As is well-known, the MEF posits the

Tayyeb Shabbir is Senior Research Economist at the Pakistan Institute of Development Economics, Islamabad.

Author's Note: I am grateful to several colleagues at PIDE and an anonymous referee for their helpful comments and to Ayaz Ahmad for his excellent research assistance.

[1]Interestingly, the intellectual history of the 'Mincerian earnings function' is far more involved than is generally realised. For a brief sketch of this history, see [Fleisher and Kniesner (1984), p. 314 fn 12].

natural logarithm of individual earnings to be a linear function of the *exact* number of years of schooling completed by the individual, labour market experience, and its quadratic term (and, of course, a stochastic error term). On the one hand, such a function has often been estimated by the proponents of the traditional (i.e., Becker-Mincerian) human-capital school in the hope of finding empirical support for its various hypotheses regarding the productivity-enhancing role of investments in schooling and on-the-job training. On the other hand, ongoing debates regarding the determinants of individual earnings often employ the MEF as a point of departure.[2] In any event, estimated MEF serves as a very useful point of reference or means of validating some critical theoretical hypotheses; as such, it often serves as the basis of worthwhile economic policy prescriptions.

Considering its importance, it is noteworthy that there exist no estimates of the MEF for Pakistan.[3] Essentially, the apparent reason for this somewhat unusual situation is the fact that, due to a peculiarity of the questionnaire design, the majority of the relevant micro-level surveys report individual's schooling not as a 'continuous' variable, i.e., in terms of the total number of years of schooling completed, but as a 'categorical' or discrete response variable, with possible responses such as 'Primary and Incomplete Middle', 'Middle and Incomplete Matric', and so on.[4] This data deficiency has precluded the calculation of any national level MEF and restricted the existing estimated earnings functions to a specification that we refer to as the 'Dummies Earnings Function' (DEF)–since they are constrained to specify schooling in terms of a set of dichotomous dummy variables.[5] [For instance, see Haque

[2]In this regard, the studies reviewed in [Behrman (1990), pp. 48–54] and Behrman and Birdsall (1987) are relevant, particularly those dealing with the debate about the 'ability' bias in the regression estimates of the marginal rate of return to schooling as obtained from the traditional MEF specification.

[3]Although Haque (1984) or Sabot (1989) may be considered as exceptions, yet both these studies are based on rather restrictive samples. Haque's study is based on a sample pertaining to just a single city in Pakistan while Sabot's is based on 800 rural households. In any event, Sabot (1989) does not even report the actual estimates for the MEF and merely notes (p. 424; fn 3) that "the years of schooling variable is positive, large, and significant when the (wage) equation is estimated without the cognitive skills or ability variables". Incidentally, it may be noted as well that in Shabbir (1991), a study of the credentialist effects of schooling in Pakistan, I have reported the estimates for a MEF that are based essentially on the same national data set as has been used in the present paper. However, as these estimates were reported merely as an adjunct to the main objective of that research endeavour, many important issues related to MEF were either not addressed there in sufficient detail or were ignored altogether.

[4]Incidentally, the various 'certificate' levels in the general educational system of Pakistan are given below, followed in parenthesis by the required number of years of schooling: Primary (5), Middle (8), Matric (10), Intermediate (12), Bachelor's (14) and Master's (16).

[5]Two points may be noted here. Firstly, the availability of data on schooling as a 'continuous' variable is a necessary but not a sufficient condition for the estimation of a MEF since a researcher may hold strong *a priori* beliefs as to how schooling and individual earnings are related. Secondly, it turns out that the lack of a 'continuous' schooling variable not only precludes the estimation of MEF but, as shown in Shabbir (1993), the 'noise' or measurement problem inherent in the 'categorical' nature of the schooling data for Pakistan biases the regression coefficient estimates for DEF.

(1977); Guisinger *et al.* (1984); Khan and Irfan (1985), and a related more recent study by Ashraf and Ashraf (1993)]. Though people often tend to treat DEF and MEF as belonging to the common genre of 'human-capital-type' earnings functions, as explained further in Section 2, in many important ways they are quite distinct entities. Thus, the absence of an estimated MEF represents a void in the earnings function literature about Pakistan.

Interestingly, a unique opportunity for surmounting the above-mentioned data constraint has existed since the 1979 Pakistan Labour Force and Migration Survey (PLMS), which happens to contain all the information needed to estimate the MEF, albeit the required data is split between two separate modules or sets of questionnaires of the PLMS–Household Income and Expenditure Survey (HIES) and the Migration Survey. Merging the appropriate data from the above two modules of the PLMS has provided us with a nationally representative sample of male earners which is used to obtain our estimates of the MEF.

The primary objective of this paper is to fill the above gap in the relevant literature by estimating the MEF for Pakistan. In this regard, the two specific questions of interest are: (a) what is the marginal rate of return to schooling and (b) what proportion of the variance in the dependent variable (natural logarithm of earnings) can be accounted for by the strictly defined MEF, i.e., one given only in terms of individual investments in schooling and on-the-job training. Though in this paper it is not our goal to try to arrive at the most complete specification of the earnings functions for Pakistan, or test alternate theories of the determinants of earnings for that matter, we do extend the strictly defined or parsimonious MEF by introducing controls for such characteristics of the earners as their place of birth (urban vs rural), employment status (self-employed vs employee), and occupational groups. With the help of the 'extended' MEF, we are able to test the assumption of the homogeneous labour market that is implicit in the strict Mincerian Earnings Function. In particular, in order to investigate this issue, the 'extended' MEF has been estimated separately for urban and rural Pakistan, as well as for each of its four provinces, to ascertain if the labour markets are segmented along these lines.[6]

The rest of this paper is organised as follows: Section 2 briefly presents the analytical background needed to clarify the precise interpretation we have in mind for the MEF. Section 3 describes the data used in this study, while Section 4 discusses the empirical estimates for the MEF for Pakistan. Finally, Section 5 offers a few concluding remarks regarding this study.

2. ANALYTICAL BACKGROUND AND EXISTING LITERATURE

In the last section we briefly described the MEF in an intuitive fashion. Since

[6]I am grateful to an anonymous referee for these suggestions to extend the parsimonious MEF.

the formal analysis that lies behind the Mincerian earnings function (MEF) is well-known and easily accessible in Mincer (1974), we would proceed directly to a comparison of the MEF and the DEF with a view to further motivate the need for this paper.

If, for a given individual, we represent the natural logarithm of earnings by *Ln Y*, the (exact) number of years of schooling completed by *S* and the years of job market experience by *EXP*, the following equation represents the MEF:

$$MEF: \quad Ln\ Y = a + r\,S + c\,EXP + d\,EXP^2 + U \ldots \quad \ldots \quad (1)$$

On the other hand, the DEF–the specification that typifies the existing earnings functions for Pakistan–can be represented by Equation (2) given below:

$$DEF: \quad Ln\ Y = \alpha + \sum_{i=1}^{k-1} (\beta_i D_i) + \gamma\,EXP + \delta\,EXP^2 + V \quad \ldots \quad \ldots \quad (2)$$

where the D_i consist of a set of dichotomous (0,1) dummy variables, one of which takes on the value unity corresponding to the 'category' in which the individual's educational level falls, e.g., $D_1 = 1$ if the individual's education falls in the group 'Primary and less than Middle' and $D_2 = 1$ if it falls in the group 'Middle and less than Matric', and so forth, for the k–1 categories. (The excluded category is 'Less than Primary and Illiterate'; where 'illiterate' is *defined* as someone with zero years of schooling.)

While the MEF and the DEF share some similarities, they also differ in important ways.[7] For one, note that the DEF differs from the MEF because the way the schooling variable is specified in each case implies a distinct hypothesis about the effect of additional schooling on the natural logarithm of earnings (*Ln Y*). While the MEF implies that *Ln Y* and *S* are linearly related, the DEF specification implies that schooling affects *Ln Y* in the manner of a (discontinuous) stepfunction where the stepsize increases only with the completion of the different certification levels. This suggests that the DEF is relatively more in the spirit of credentialism.[8]

[7]In terms of similarities, the MEF and the DEF are related to each other since, as pointed out by an anonymous referee, the MEF can be considered as a restricted form or a special case of the DEF and, in principle, one can test this restriction. However, here we want to keep the major goal of the paper, i.e., to estimate the MEFs *per se*, in sharper focus.

[8]However, Chiswick (1973) makes the case that even a specification akin to the DEF is consistent with the human capital view that schooling affects earnings via increasing productivity, albeit such increase presumably occurs only after a threshold of years of schooling is reached and conceivably such thresholds could correspond to the certification levels. Thus, Chiswick implicitly assumes that intra-diploma years contribute little or nothing to productivity till the threshold level is reached. Here, once again, the shortcoming inherent in having schooling data available only in the 'categorical' form should be apparent since we cannot test the above assumption unless we can identify individuals whose years of completed schooling fall 'in between the diploma years'.

The above distinction between the MEF and the DEF may also be drawn in terms of the respective 'marginal' rates of return to schooling as implied by these specifications. While the marginal rate of return in the case of MEF is constant, continuous over the domain of S, and well-defined for intra-diploma years, the comparable rate in the case of DEF is variable and discontinuous; in particular, its dynamic profile for intra-diploma years is theoretically not well-defined.[9]

At this point, when we are contrasting the MEF and the DEF, it may be useful to recall that the goal of this paper is not to test as to which of the two is the more appropriate specification or, for that matter, to search globally for the best earnings function specification for Pakistan, but instead it is to remedy the lack of an estimated MEF caused by the data deficiency alluded to earlier. Obtaining estimates of the MEF is an important issue not only in its own right but also because, despite some similarities, the MEF and the DEF are distinct enough to warrant such an exercise.

3. DATA DESCRIPTION

As mentioned earlier, the lack of a sample with information on the *exact* years of schooling completed by an individual may have been one of the obstacles in the estimation of the MEF for Pakistan. We overcome this long-standing constraint by merging the relevant information from two modules of the 1979 Population, Labour Force and Migration Survey (PLMS),[10] i.e., the Migration Modules (for data on 'continuous' schooling variable) and the Household Income and Expenditure Survey (for information on all the remaining variables needed, such as monthly earnings, age, employment status, and the occupational group of the individuals). This enables us to estimate presumably the first MEF for Pakistan that is based on a nationally representative sample.

For the present study, most of the empirical estimates are for a sample of 3017 male earners (wage earners or salaried employees) for whom the natural logarithm of reported monthly earnings was positive, i.e., $\text{Ln } Y > 0$, and years of schooling completed or $S \geq 0$. (However, any variations in the sample described above are noted in the relevant tables given in the next section, which presents the empirical

[9]Of course, one such profile is suggested by Chiswick (1973) and has been noted in detail in the previous footnote of this paper.

[10]Conducted as a joint project of the Pakistan Institute of Development Economics (PIDE) and ILO-UNFPA, the PLM, a nationally representative survey, was based on a two-stage stratified random sample of 11,288 households. Each household was asked to respond to four sets of the questionnaires, i.e., income-expenditure, labour force participation, migration, and fertility–two of which, the Household Income and Expenditure (HIES) and the Migration, are relevant here. Whereas the HIES is a survey that is conducted with some regularity, the Migration Survey was a one-shot thing done only in 1979. These surveys were conducted during the last two quarters of 1979. However, the Migration Survey spilled over into the first couple of months of 1980 as well.

estimates.) Meanwhile, Table 1 provides the definitions of some of the important variables, their sample means, and standard deviations for the above sample.

Table 1

*Means, Standard Deviations, and Bivariate Correlations for Some Important Variables**

(Male Earners, S ≥ 0, Sample Size = 3017)

Variable	Bivariate Correlations			Means	Standard Deviations
	Ln Y	S	EXP		
Ln Y	1.00			6.23	0.67
S	0.40	1.00		4.47	4.97
EXP	0.26	–0.37	1.00	23.12	14.52

*Definitions of the Variables:

Ln Y = Natural logarithm of individual's monthly earnings (may consist of wages or salary).

S = Years of completed schooling.

EXP = Years of labour market experience calculated as (AGE-S-6).

4. EMPIRICAL ESTIMATION AND INTERPRETATION OF RESULTS

(a) Results for the 'Strict' or 'Traditional' MEF

Table 2 presents the estimated regression coefficients for the 'traditional' Mincerian earnings function, i.e., one which has been strictly defined *à la* Mincer (1974) and has been presented earlier as Equation (1). These estimates are divided in two panels, i.e., Panel A and Panel B, which consist of three columns each. The first three columns (i.e., columns 1 through 3) pertain to the case where S ≥ 0 (i.e., the individuals with zero years of schooling are also included) while the last three columns (i.e., columns 4 through 6) pertain to the case where S > 0 (i.e., those with zero years of schooling are excluded).

Let us first discuss the S ≥ 0 case which is perhaps more relevant for a developing country, given the fact that a relatively larger proportion of their population is generally without any schooling. The results in Table 2 indicate the following findings. The signs of S and the EXP terms are consistent with the Mincerian view, and the estimated coefficients for each of these variables are significant at the 99 percent level. The private marginal rate of return to schooling is 5 percent in column 1 and goes up to 8 percent in column 2 when we control for the labour market experience, which removes the downward bias for the coefficient estimate of S (younger cohorts

Table 2

Regression Estimate of the Mincerian Earnings Function

(OLS; Dependent = Ln Y; Male Earners)

	Panel A (S ≥ 0)			Panel B (S > 0)		
	1	2	3	4	5	6
Constant	5.99*	4.98*	5.02*	5.78*	4.87*	5.30*
	(395.43)	(155.06)	(156.46)	(135.94)	(94.69)	(67.12)
S	0.05*	0.08*	0.03*	0.07*	0.10*	–0.02
	(23.60)	(38.74)	(4.27)	(16.33)	(24.42)	(–1.16)
$(S)^2$			0.004*			0.007*
			(8.66)			(7.03)
EXP		0.06*	0.06*		0.06*	0.05*
		(27.18)	(27.53)		(16.27)	(16.69)
$(EXP)^2$		–0.001*	–0.001*		–0.001*	–0.001*
		(–18.72)	(–19.15)		(–8.31)	(–9.42)
Adjusted R^2	0.15	0.41	0.42	0.14	0.41	0.42
N =	3017	3017	3017	1568	1568	1568

*Significant at 99 percent level, 2-tailed *t*-test; (*t*-statistics in parentheses).

have relatively more schooling and necessarily fewer years of experience).[11] Again, inclusion of EXP and EXP^2 (i.e., experience terms) in the specification more than doubles the explanatory power of the regression (Adjusted R^2 goes up from 0.15 in column 1 to 0.41 in column 2). Lastly, note in column 3 that when we add an S^2 term to the specification, its sign turns out to be positive and significantly different

[11]This rate of return of 8 percent is comparable to that of 7.7 percent reported by [Haque (1984), p. 12], who employs a similar earnings function specification but his estimate is based on a sample of just one city of Pakistan. In any event, for a broader cross-national comparison, let us note that [Psacharopoulos (1980), pp. 89, 90] reports 12.9 percent, 9.7 percent, and 7.7 percent as the respective (average) rates of return to schooling for 'Developing (Asian)', 'Intermediate', and 'Advanced' countries. Though, based on these averages alone, the estimate for Pakistan would appear to lean relatively more towards the result for the 'Advanced' (and 'Intermediate') group of countries rather than the 'Developing (Asian)' group, the intra-group variation in the last group is instructive too. The country-wise breakdown of the rate of return for the 'Developing (Asian)' group, in the ascending order, is Taiwan (6.0 percent), Singapore (8.0 percent), Thailand (10.4 percent), S. Vietnam (16.8 percent), and Malaysia (22.8 percent). Then, perhaps, the first three countries can be considered to belong to a sub-group 'Low-rate countries' (average rate of return: 8.13 percent), while the remaining two countries belong to the sub-group 'High-rate countries' (average rate of return: 19.8 percent). The above suggests that there may be no hard and fast empirical regularity linking the rate of return on schooling for different countries to their stage of development.

from zero. This implies that, in the case of Pakistan, the marginal returns to schooling may be an increasing function of the level of schooling. This is in line with the earlier estimates of the rates of return obtained for the various levels of education in Pakistan.[12]

Let us now refer to Panel B of Table 2 in order to say a few words about the results based on the sample where only those observations are included where $S > 0$. In particular, note that S_B, regression coefficient of S in column 5, now increases to 0.10 as compared to the value of 0.08 for its counterpart, S_A, in column 2. The null hypothesis is that $S_A = S_B$ vs the alternative that $S_A \neq S_B$. *A priori*, one may expect $S_A = S_B$ in light of the traditional argument that like gender, age, etc., education is exogenous to the income determination process and, thus, selecting on S should not make any difference to the estimates.[13] On the other hand, however, it can be argued that in terms of the income determination process, those with no schooling may satisfy a different law from those with 'positive' schooling.[14] Thus, theory may not be the ultimate arbiter of the argument regarding whether S_A is equal to S_B. Further, the empirical test may also be warranted if one considers that there may be a selectivity problem considering the fact that, unlike for the developed countries such as U.S., whose data were used to originally test the MEF, most developing countries have a significant proportion of population with zero schooling. Thus, one could argue that this last feature may imply structural differences across these two types of economies with respect to income determination. In any event, considering samples where $S > 0$ vs those where $S \geq 0$ is not without precedent in the literature [for instance, see Psacharopoulos (1977)].

In our particular case, a Chow-type *F*-test does not reject the null hypothesis of $S_A = S_B$ vs the alternative $S_A \neq S_B$ at the 1 percent level of significance.[15] With this perspective in mind, it may be noted that the empirical results presented in the

[12]See [Khan and Irfan (1985), Table 2 on p. 625]. In general, however, there exists no consistent stylised fact regarding the sign of S^2 in the earnings function. For the U.S., Mincer found the sign to be negative and significant, which however becomes insignificant once the number of weeks worked in a year are controlled for. [Mincer (1974), Table 3.3 on p. 53 and comments on p. 54.] For a number of developing countries, the evidence is one of perhaps a V- or U-shaped relationship between the marginal rate of return and the level of S; see [Psacharopoulos (1980), Table 1]. It seems that in order to resolve the above issue, it may be necessary to control for the amount of time worked per year in the earnings function specification.

[13]This was pointed out by an anonymous referee. A similar discussion can be noted in [King (1980), pp. 249–55] as well.

[14]Those with zero schooling may be less likely to find paid employment. On the other hand, they may also have a stronger preference for self-employment as evidenced by the general observation that a relatively larger proportion of the self-employed as against employees have $S = 0$.

[15]Specifically, we find Calculated $F_{1449,1569} = 0.938 <$ Tabulated $F_{1449,1569} = 1.0$ at 1 percent level.

remainder of this paper are only for the sample where $S \geq 0$.

(b) Variations on a Theme: 'Extended' MEF As Evidence of Labour Market 'Segmentation'

We have already discussed empirical estimates of the 'traditional' MEF for Pakistan. Mincer presumed a 'smooth' or homogenous labour market in his theoretical as well as empirical analysis for the United States. However, it is important to ascertain if the income determination process is in fact 'smooth', or is instead 'segmented' along such dimensions as the spatial characteristics of the labour market, urban vs rural origins of a person, an individual's employment status as an employee vs a self-employed person, and the various occupational groups a worker may belong to.

How segmented is the labour market in Pakistan? In order to answer this question, let us refer to the empirical results in Tables 3 through 5, wherein we have relaxed the assumption of a homogenous labour market.

Let us start with Table 3, which gives the results for the 'extended' MEF–obtained by extending the 'traditional' MEF by including dichotomous (0,1) dummy variables to account for possible segmentation along four important dimensions–urban vs rural origin (represented by dummy variable URBAN), provincial labour market heterogeneity (represented by the set of dichotomous (0,1) dummy variables PUNJAB, NWFP, and BALOCHISTAN, with SINDH being the excluded dummy variable), employment status, i.e., being an employee vs being self-employed (represented by dummy variable SE) and occupational group-based-heterogeneity (represented by the set of dichotomous (0,1) dummy variables PROF, CLER, AGR, with PROD being the excluded category). [Note: For additional detail on how the above variables have been defined, see the glossary given at the bottom of Table 3.] Note that the introduction of the above dummy variables represents the familiar 'shift' variable or 'intercept adjustment' approach to testing labour market segmentation. A significant estimated coefficient of a given dummy variable would imply that workers with the same levels of schooling, experience, and other characteristics included in the regression receive different incomes in the labour market depending upon the characteristics represented by the dummy variable. Let us now look at the empirical results pertaining to each of the four possible dimensions of market segmentation that have been suggested above.

Inclusion of the dummy variable URBAN can be justified on several grounds. Generally, it is argued that the workers with an urban origin or background enjoy a relative advantage on account of their exposure to a more varied set of influences, and access to opportunities for easier acquisition of information, as well as to better-quality schools. The results in the first column of Table 3 confirm the above *a priori*

Table 3

'Extended' MEF for Pakistan
(OLS; Dependent = Ln Y; Males; S ≥ 0)

	1	2	3	4
Constant	4.88*	4.96*	5.02*	4.93*
	(147.02)	(136.05)	(131.49)	(168.19)
S	0.07*	0.07*	0.07*	0.08*
	(35.61)	(36.03)	(28.59)	(40.68)
EXP	0.06*	0.06*	0.06*	0.06*
	(27.74)	(27.59)	(27.22)	(38.40)
$(EXP)^2$	–0.001*	–0.001*	–0.001*	–0.001*
	(–19.25)	(–19.21)	(–18.98)	(–28.46)
URBAN	0.20*	0.17*	0.17*	0.18*
	(10.03)	(8.24)	(7.58)	(11.04)
PUNJAB		–0.14*	–0.14*	–0.14*
		(–6.71)	(–6.77)	(–9.57)
NWFP		0.07	0.07	0.03
		(1.69)	(1.80)	(0.89)
BALOCHISTAN		0.04	0.04	0.08*
		(1.34)	(1.17)	(3.81)
SE				0.27*
				(18.64)
PROF			0.17*	0.21*
			(4.70)	(6.35)
CLER			–0.08*	0.01
			(–3.40)	(0.84)
AGR			–0.08*	0.12*
			(–2.58)	(6.80)
Adjusted R^2	0.43	0.44	0.45	0.34
N =	3017	3017	3017	8655

*Significant at 99 percent level; 2-tailed *t*-test (*t*-statistics are given in the parentheses).

GLOSSARY:

URBAN = Dichotomous; equals 1 if the person was born in urban Pakistan, 0 otherwise.
SE = Dichotomous; equals 1 if the person is self-employed, 0 if an employee.
PUNJAB = A set of (0,1) dummy variables, each member of which assumes value 1 only if the individual resides in that province, 0 otherwise. [Note: SINDH is the excluded dummy variable.]
NWFP = A set of (0,1) dummy variables, each member of which assumes value 1 only if the individual resides in that province, 0 otherwise. [Note: SINDH is the excluded dummy variable.]
BALOCHISTAN = A set of (0,1) dummy variables, each member of which assumes value 1 only if the individual resides in that province, 0 otherwise. [Note: SINDH is the excluded dummy variable.]
PROF = Dichotomous; equals 1 if the individual's occupation group is 'Professional, Technical, and Related Workers', 0 otherwise.
CLER = Dichotomous; equals 1 if the individual's occupation group is 'Clerical and Related Workers', 'Sales Workers' or 'Service Workers', 0 otherwise.
AGR = Dichotomous; equals 1 if the individual's occupation group is 'Agricultural, Animal Husbandry, Forestry Workers, Fishermen, and Hunters'.
PROD = (Excluded) dichotomous; equals 1 if the individual's occupation group is 'Production and Related Workers, Transport and Equipment Operators, Labourers', 0 otherwise.

Table 4

Urban vs Rural Mincerian Earnings Functions for the Whole of Pakistan

(OLS; Dependent Variable = Ln Y; Males; S ≥ 0)

	Urban			Rural		
	1U	2U	3U	1R	2R	3R
Constant	5.06*	5.20*	5.12*	4.89*	4.93*	4.87*
	(129.65)	(129.29)	(140.97)	(89.79)	(76.48)	(117.90)
S	0.08*	0.07*	0.07*	0.06*	0.07*	0.07*
	(31.31)	(24.53)	(31.02)	(16.69)	(14.07)	(25.02)
EXP	0.06*	0.06*	0.06*	0.06*	0.06*	0.06*
	(22.83)	(22.52)	(27.45)	(15.82)	(15.76)	(27.23)
$(EXP)^2$	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*
	(–15.92)	(–15.84)	(–20.16)	(–10.96)	(–10.99)	(–20.54)
PUNJAB		–0.18*	–0.20*		–0.002	–0.06*
		(–7.29)	(–9.69)		(–0.06)	(–3.08)
NWFP		0.09	0.01		0.13**	0.07
		(1.64)	(0.17)		(1.97)	(1.54)
BALOCHISTAN		–0.12*	–0.07		0.27*	0.20*
		(–3.03)	(–1.90)		(5.39)	(6.83)
SE			0.28*			0.27*
			(13.75)			(12.80)
PROF		0.29*	0.29*		–0.02	0.08
		(6.54)	(7.10)		(–0.31)	(1.52)
CLER		–0.04	0.05**		–0.17*	–0.03
		(–1.46)	(2.17)		(–3.95)	(–0.96)
AGR		–0.08	0.02		–0.13*	0.11*
		(–1.14)	(0.54)		(–3.57)	(5.73)
Adjusted R^2	0.42	0.46	0.41	0.33	0.37	0.28
N =	1894	1894	3535	1123	1123	5120

* Significant at 99 percent level; 2-tailed *t*-test (*t*-statistics are given in the parentheses).

** Significant at 95 percent level; 2-tailed *t*-test.

Table 5

Urban vs Rural Earnings Functions, by Province

(OLS; Dependent Variable = Ln Y; Males; S ≥ 0)

	Punjab				Sindh				NWFP				Balochistan			
	1P	2P	3P	4P	1S	2S	3S	4S	1N	2N	3N	4N	1B	2B	3B	4B
			Urban	Rural			Urban	Rural			Urban	Rural			Urban	Rural
Constant	4.80*	4.73*	4.81*	4.78*	5.17*	4.93*	5.19*	4.95*	5.04*	5.04*	5.07*	5.04*	5.45*	5.34*	5.64*	5.29*
	(136.41)	(130.04)	(91.35)	(102.18)	(116.88)	(96.68)	(99.79)	(59.13)	(59.20)	(57.81)	(37.05)	(45.52)	(50.07)	(48.74)	(37.18)	(34.77)
S	0.08*	0.08*	0.08*	0.08*	0.07*	0.07*	0.07*	0.07*	0.07*	0.07*	0.08*	0.07*	0.06*	0.05*	0.05*	0.04*
	(31.71)	(30.21)	(21.36)	(20.81)	(23.77)	(23.05)	(19.28)	(11.25)	(12.33)	(12.03)	(9.00)	(7.20)	(8.90)	(7.55)	(6.23)	(2.96)
EXP	0.06*	0.06*	0.06*	0.06*	0.05*	0.05*	0.06*	0.04*	0.06*	0.06*	0.06*	0.06*	0.04*	0.04*	0.04*	0.04*
	(29.60)	(29.70)	(19.82)	(22.03)	(19.51)	(19.76)	(17.62)	(9.63)	(12.67)	(12.65)	(6.74)	(10.67)	(6.72)	(6.97)	(3.90)	(5.40)
$(EXP)^2$	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001*	–0.001**	–0.001*
	(–21.65)	(–21.82)	(–14.38)	(–16.31)	(–14.50)	(–14.77)	(–13.59)	(–6.93)	(–10.01)	(–10.00)	(–4.93)	(–8.68)	(–4.77)	(–4.95)	(–2.04)	(–4.45)
SE	0.22*	0.24*	0.23*	0.25*	0.32*	0.34*	0.32*	0.40*	0.32*	0.32*	0.44*	0.24*	0.22*	0.20*	0.12	0.34*
	(10.73)	(11.64)	(7.43)	(8.94)	(12.55)	(13.69)	(10.64)	(8.57)	(7.03)	(7.00)	(5.70)	(4.18)	(4.39)	(4.18)	(1.66)	(5.06)
PROF	0.11**	0.12**	0.20*	0.03	0.36*	0.39*	0.44*	0.22	0.19	0.19	0.17	0.24	–0.06	–0.05	–0.11	0.15
	(2.32)	(2.54)	(3.09)	(0.39)	(6.58)	(7.26)	(7.29)	(1.88)	(1.87)	(1.87)	(1.25)	(1.55)	(–0.50)	(–0.44)	(–0.75)	(0.71)
CLER	0.02	0.01	0.05	–0.03	0.03	0.03	0.04	0.01	–0.01	–0.01	–0.04	–0.02	–0.02	–0.04	–0.03	–0.05
	(0.85)	(0.44)	(1.65)	(–0.95)	(1.17)	(1.15)	(1.43)	(0.11)	(–0.23)	(–0.25)	(–0.48)	(–0.23)	(–0.33)	(–0.74)	(–0.38)	(–0.57)
AGR	0.09*	0.15*	0.13	0.13*	–0.08*	0.11*	–0.08	0.14*	0.04	0.04	–0.16	–0.06	–0.05*	–0.05	–0.01	–0.12
	(4.29)	(6.70)	(1.87)	(5.23)	(–2.87)	(3.01)	(–1.17)	(2.66)	(0.71)	(0.73)	(–0.79)	(1.08)	(–2.30)	(–0.79)	(–0.07)	(–1.58)
URBAN		0.16*				0.28*				0.01				0.23*		
		(7.26)				(8.76)				(0.16)				(4.44)		
Adjusted R^2	0.32	0.33	0.42	0.28	0.36	0.38	0.42	0.29	0.29	0.29	0.36	0.24	0.27	0.31	0.25	0.28
N =	4816	4816	1534	3282	2523	2523	1502	1021	921	921	308	613	395	395	191	204

* Significant at 99 percent level; 2-tailed *t*-test (*t*-statistics are given in the parentheses).

**Significant at 95 percent level; 2-tailed *t*-test.

expectation since they show that the regression coefficient of URBAN is positive and significant; it remains robust even when we include additional control variables (i.e., moving across columns 1 through 4 in Table 3). Incidentally, this result is in keeping with the available evidence for Pakistan and other developing countries in general.[16]

The third column of Table 3 presents results for the specification where province as well as occupational group dummy variables have been introduced to capture inter-provincial and inter-occupational differences in the determinants of individual earnings.

The cross-province-variation in the parameters of the MEF may be expected for a number of reasons. Firstly, spatial heterogeneity of labour markets and variations in the ethnic mix and cultural norms can often act as barrier to completely free mobility. Secondly, inter-provincial differences in budgetary allocations can lead to important differences in the labour market environment, such as information flows and access to schooling. The latter factor can indirectly affect labour market opportunities and thus limit job mobility of workers. In order to control for the above type of inter-provincial differences, three binary dummy variables, i.e., PUNJAB, NWFP, and BALOCHISTAN, are included in the specification while SINDH is the excluded dummy variable which represents the 'reference' province.

Similar to segmentation along provincial lines, the labour market may also be segmented along occupational categories or groups. We have tried to capture these latter effects by introducing three binary variables, i.e., PROF, CLER, and AGR, where the excluded category is PROD. (These variables have been described at the foot of Table 3.)

Turning now to the empirical findings with regard to inter-provincial and inter-occupational group differences, first note that, as given in column 3 of Table 3, the estimated coefficient for PUNJAB is 'negative and significant' at the 99 percent level, while for NWFP it is positive but significant only at the 90 percent level, and for BALOCHISTAN it is positive but not significant. (The excluded provincial dummy variable is SINDH.) Though the above empirical findings are consistent with similar evidence reported in a related context by Khan and Irfan (1985) and Ashraf and Ashraf (1993), these results may still be somewhat surprising to some since the Punjab province is generally considered to be relatively the most 'prosperous' one. However, once we consider the differential influence across provinces of outmigration from Pakistan to the Gulf countries, the results can be rationalised. Since a relatively greater proportion of the populations of the NWFP and

[16]For Pakistan, a similar finding of a positive and significant estimate for URBAN dummy is reported in the DEF-type earnings functions estimated by Ashraf and Ashraf (1993) and Khan and Irfan (1985), while Fields (1978) confirms it for Colombia, a typical developing country.

Balochistan provinces outmigrated during the 1970s, this lead to a relatively greater tightening of their labour markets, resulting in a greater average increase in the general wage rate in these provinces.[17]

Further, in terms of the inter-occupational differences, the evidence given in column 3 shows that, on average, the workers represented by the category PROF earn significantly more (to the extent of 17 percent) while those in the CLER group or AGR group each earn 8 percent less compared to those belonging to the excluded category of PROD. One of the interesting implications of these results is that the production and blue-collar workers (represented by the excluded group PROD) have higher mean earnings compared to the white-collar clerical or sales workers. However, this may once again be considered as a consequence of the nature of outmigration to the Middle East since the bulk of the foreign demand for the Pakistani labour during the 1970s was for skilled and semi-skilled production workers.

Another noteworthy feature of the results given in column 3 of Table 3 surfaces when we compare them to those in column 2 of Table 2. While it is clear that due to the introduction of variables representing different provinces and the occupational groups, the explanatory power of the earnings function increases (Adjusted R^2 increases from 0.41 to 0.45), the coefficient of S as a measure of the marginal rate of return to education is lowered from .08 previously to .07 in Table 3 –approximately a 12 percent reduction.

Another issue of interest relates to the nature of the effect of self-employment–an important yet inadequately studied question for the LDCs in general, and Pakistan in particular. To capture this effect we include SE, dichotomous (0,1) dummy variable, which takes on unity value for the self-employed. The coefficient estimate for SE is positive and significant at the 99 percent level (column 4, Table 3). This evidence of higher mean earnings for the self-employed is consistent with the finding reported by Haque (1977) for the Rawalpindi City. However, the above results ought to be treated as of exploratory nature at best, since estimating earnings functions when the data about employees and the self-employed (whose 'earnings' also contain capital income) are pooled is not a straightforward matter.[18] The self-employed may differ systematically from employees in terms of the labour supply and risk-taking behaviour. Thus, the limitations of the present exercise of trying to control for the self-employment status by merely introducing a dummy variable in

[17]As pointed out by Khan and Irfan (1985), an additional explanation may lie in the fact that due to a sampling error, relatively more of the poorer (far-flung) areas in the NWFP and Balochistan were left out of the sample which may have artificially raised the 'prosperity' measures of these provinces.

[18]The lack of data on capital stock and the absence of control for selectivity are two probable reasons of the significant drop of adjusted R^2 when going from column 3, i.e., only employees, to column 4, i.e., when employees and the self-employed both are included in the sample.

the specification should be kept in mind while interpreting the empirical results.

Regarding the light shed on labour market segmentation by the above results, it reveals evidence of significant segmentation with respect to the urban, provincial, occupational group, and employment status of the earner. However, as mentioned earlier, the models in Table 3 presume that all segmentation effects can be captured by the 'intercept' adjustment alone. This may not be so, and instead the complete set of parameters of the earnings function may change across these market heterogeneities.

In order to entertain the above possibilities, we move to Table 4 and Table 5. For the purposes of this additional exercise, however, we would limit ourselves to the urban vs rural and provincial segmentation, though occupational categories and employment status would continue to be included as control variables in the specification.

Table 4 provides the sample with separate estimates for the MEF for urban and rural Pakistan. Thus, unlike Table 3, it assumes that the complete regression specification (and not just the intercept term) may vary across these sub-samples. Even an informal examination of these results for urban vs rural samples yields some interesting observations: First, while the coefficient estimate of S for the 'strict' MEF is substantially lower for the rural sample (i.e., 1U compared with 1R), this divergence disappears when we employ the 'extended' MEF, which is a more complete specification (i.e., 3U and 3R). However, in other instances, such as in the case of coefficient estimates for provincial dummies (PUNJAB, NWFP, and BALOCHISTAN) as well as occupational dummies (PROF, CLER, and AGR), there are significant differences across urban and rural samples. In particular, Balochistan's 'disadvantage' in the urban case turns into an 'advantage' since the relevant coefficient estimate reverses its sign (from –0.12 to +0.27) and is still significant.[19] Another interesting finding is in terms of the occupational group dummies where, as expected, AGR for the rural sample is relatively more significant while PROF loses its relative significance. From the segmentation point of view, even the above cursory analysis points out that the income determination process is different for urban vs rural markets. However, we also conducted a formal analysis to investigate if the parameters of the 'extended' MEF (i.e., column 3U vs 3R) are different across urban and rural samples. In this regard, the null hypothesis of homogeneity of the parameters of the MEF across two subsamples is rejected at 1 percent level since it turns out that $F_{Calculated}$ (11,8633; 0.01) = 19.24 > $F_{Tabulated}$ (11,8633; 0.01) = 2.24. Further, this heterogeneity is not limited to just the

[19]It is possible that this result is only reflecting a sampling anomaly since, as mentioned previously, due to difficulties of access to the far-flung regions particularly in rural Balochistan, the poorer households may have been left out of the sample.

intercepts being different, but is 'pervasive', since the null hypothesis of only the intercepts being different across the urban vs rural sub-samples is also rejected at 1 percent level. [$F_{Calculated}$ (10,8633; 0.01) = 21.76 > $F_{Tabulated}$ (10,8633; 0.01) = 2.327.] In fact, these formal *F*-tests were also conducted for urban vs rural sub-samples of each province. Except for the NWFP, the above national result was upheld across each province's urban vs rural samples. The above formal analysis has confirmed the need for undertaking separate within-sample estimation for urban and rural samples and, importantly enough, it has reaffirmed the presence of market segmentation along these dimensions.

Finally, let us now turn to an analysis of inter-provincial variation in the MEF. Table 5 gives four sets of columns for a given sub-sample corresponding to each province. In terms of some general observations, note that the evidence in Table 5 suggests that segmentation along such dimensions as the individual's employment status (SE being positive and significant), and that along the urban vs rural dimensions, exists across provinces too. In terms of the occupational group dummy variables, in many instances, PROF's coefficient is positive and significant but is generally relatively lower for rural samples, whereas AGR's coefficient is positive and relatively more significant for the same areas. Further, the coefficient of S, i.e., the marginal rate of return to schooling is positive and significant but does not vary much across the provinces except for Balochistan, where not only the sample size is relatively small but also the province is the least developed one, in particular its rural areas.

Again, formal *F*-tests comparing the results for the 'extended' MEF estimated for within urban samples across the four provinces (i.e., 3P, 3S, 3N and 3B) enabled us to reject at 1 percent the null hypothesis of homogeneity of the MEF across provinces. Similar tests also show that the above heterogeneity extends beyond merely differential intercepts across the provinces. On the other hand, *F*-tests comparing rural samples across the four provinces reject at 1 percent the null hypothesis of homogeneity of the parameters of the MEF, but do not reject the null hypothesis that only the intercepts differ; it shows that the heterogeneity is less 'pervasive' in the case of the rural areas.

In short, the general conclusion that emerges from the above analysis of inter-provincial variation is that it is important to consider the possibility of segmentation along the inter-provincial dimension, albeit its severity for rural areas is relatively less. This evidence of significant inter-provincial differences provides valuable insights to formulate a correct policy to reduce regional disparities.

5. CONCLUDING REMARKS

In an attempt to fill a void in the relevant literature for Pakistan, this paper

uses a nationally representative sample of male earners to estimate the Mincerian Earnings Function (MEF), both in its 'strict' as well as the 'extended' forms.

In terms of the 'strict' MEF, the main findings are that the marginal rate of return to schooling is 8 percent for males, the experience-earnings profile for Pakistan is consistent with the pattern implied by the human capital theory, and finally, almost 41 percent of the variance in the dependent variable is accounted for by the (rather parsimonious) 'strict' MEF. Besides being important for their policy implications, these results are also significant from a purely analytical point of view, since in the case of Pakistan they provide the only available opportunity for comparison of results between the MEF and the DEF (the Dummies Earnings Function).

The other particularly interesting result relates to our test of an assumption that is implicit in the specification of the 'strict' MEF, namely, that labour market is homogenous. The evidence tends to refute this assumption as it strongly suggests segmentation along the following strata: urban vs rural regions of residence, self vs wage-employment, occupational groups and province-wise sub-samples. Separate ('extended') MEFs for urban vs rural samples and the various provincial sub-samples have been presented, and formal *F*-tests for homogeneity of the parameters of the MEF suggest a fairly strong heterogeneity across both these levels of stratification.

From a policy perspective, the above results reaffirm the significant positive effect of education on earnings, and also suggest ways to reduce the inter-regional (both urban vs rural as well as province-wise) disparities of income.

REFERENCES

Ashraf, Javed, and Birjees Ashraf (1993) An Inter-temporal Analysis of the Male-Female Earnings Differential in Pakistan. *The Pakistan Development Review* 32:4.

Behrman, Jere R. (1990) *Human Resource Led Development? Review of Issues and Evidence*. New Delhi: ILO-ARTEP.

Behrman, Jere R., and Nancy Birdsall (1983) The Quality of Schooling. Quantity Alone is Misleading. *The American Economic Review* 73:5 928–946.

Behrman, Jere R., and Nancy Birdsall (1987) Communication on 'Returns to Education: A Further Update and Implications'. *Journal of Human Resources* (Fall) 22:4 603-606.

Chiswick, Barry (1973) Schooling, Screening and Income. In Lewis Solomon and Paul Taubman (eds) *Does College Matter*? New York: Academic Press.

Fields, G. S. (1978) Analyzing Colombian Wage Structure. Washington, D.C.: The World Bank. (World Bank Studies in Employment and Rural Development No. 46.)

Fleisher, Belton M., and Thomas J. Kniesner (1984) *Labour Economics: Theory,*

Evidence, and Policy. Old Tappan, NJ: Prentice-Hall, Inc.

Guisinger, S. E., J. W. Henderson and G. W. Scully (1984) Earnings, Rates of Return to Education and the Earnings Distribution in Pakistan. *Economics of Education Review* 3:4.

Haque, Nadeem Ul (1977) Economic Analysis of Personal Earnings in Rawalpindi City. *The Pakistan Development Review* 26:4.

Haque, Nadeem Ul (1984) Work Status Choice and the Distribution of Family Earnings. Santa Monica: The Rand Corporation. (Rand Paper Series No. P-4037.)

King, T. (ed) (1980) Education and Income. Washington, D.C.: The World Bank. (World Bank Staff Working Paper No. 402.)

Khan, Shahrukh Rafi, and Mohammad Irfan (1985) Rates of Returns to Education and the Determinants of Earnings in Pakistan. *The Pakistan Development Review* 24:3&4.

Mincer, Jacob (1974) *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

Pindyck, Robert S., and Daniel L. Rubinfeld (1981) *Econometric Models and Economic Forecasts*. New York: McGraw-Hill Book Company.

Psacharopoulos, G. (1977) Schooling, Experience and Earnings: The Case of an LDC. *Journal of Development Economics* 4.

Psacharopoulos, G. (1980) Returns to Education: An Updated International Comparison. In T. King (ed) *Education and Income*. Washington, D.C.: (July) 73–109. (World Bank Staff Working Paper No. 402.)

Sabot, Richard H. (1989) Human Capital Accumulation in Post-green Revolution Pakistan: Some Preliminary Results. *The Pakistan Development Review* 28:4 413–431.

Shabbir, Tayyeb (1991) Sheepskin Effects in the Returns to Education in a Developing Country. *The Pakistan Development Review* 30:1 1–19.

Shabbir, Tayyeb (1993) Misspecification Bias in the Rates of Return to Completed Levels of Schooling. Islamabad: The Pakistan Institute of Development Economics. (Mimeographed.)