An ESRC Research Group

Rates of Return to Education by Gender in Pakistan

GPRG-WPS-064

Monazza Aslam

Global Poverty Research Group

Website: http://www.gprg.org/

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was part of the programme of the ESRC Global Poverty Research Group.

Rates of Return to Education by Gender in Pakistan

By
Monazza Aslam
University of Oxford

Abstract

Differential labour market returns to male and female education are one potential explanation for large gender gaps in education in Pakistan. We empirically test this explanation by estimating private returns to education separately for male and female wage earners. This paper contributes to the literature by using a variety of methodologies (Ordinary Least Squares, Heckman correction, 2SLS and household fixed effects) in order to consistently estimate economic returns to education. The latest nationally representative data – the Pakistan Integrated Household Survey (2002) – is used. Earnings function estimates consistently reveal a sizeable gender asymmetry in economic returns to education, with returns to women's education being substantially and statistically significantly *higher* than men's. The return to an additional year of schooling ranges between 7 and 11 per cent for men and between 13 and 18 per cent for women. There are also large, direct returns to women's education at low levels of schooling and the education-earnings profile is more convex for women than men. However, a decomposition of the gender wage gap (into the component 'explained' by differing male and female endowments and the residual component) suggests that there is highly differentiated treatment by employers. We conclude that the total labour market returns are much higher for men, despite returns to *education* being higher for women. This suggests that parents may have an investment motive in allocating more resources to boys than to girls within households.

Correspondence: Department of Economics, University of Oxford, Manor Road, Oxford, OX1 3UQ, United Kingdom, Telephone: +44-1865-271089.
Email: monazza.aslam@economics.ox.ac.uk

JEL Classification: I21, J16.

Keywords: Rates of return, gender, Pakistan.

Acknowledgements: This work was funded by a DFID Research Programme Consortium on 'Outcomes of Education' (RECOUP) whose support is gratefully acknowledged. I would like to thank Geeta Kingdon for the ideas, discussions, comments and suggestions that resulted in this paper. This paper has also substantially benefitted from comments by Harold Alderman, John Knight, Francis Teal, Jim Malcomson, Margaret Stevens, Mary Gregory, and the participants of the Labour Economics Seminar, University of Oxford. I would also like to thank Christopher Colclough and Marcel Fafchamps as well as the participants of the UKFIET conference (2005) at Oxford for helpful suggestions. Any errors are mine.

1. Introduction

While Pakistan's large and persistent gender gaps in education are well-documented, explaining their existence and obstinacy has proven more difficult. This paper tests one - the labour market - explanation for gender gaps in education in Pakistan[1]. Based on the investment motive, it contends that if the labour market rewards men's schooling more than women's or if it more generally discriminates among the two genders, parents may have an incentive to invest more in boys' education. In this study, we test whether the rewards to females are less than to males in Pakistan's labour market, i.e. whether the return to educating females is lower than that for men. We also ask more generally whether there is wider gender differentiated treatment in the labour market, i.e. whether much or all of the gender gap in earnings is explained by differences in male and female characteristics.

Private economic returns to education are estimated using Mincer's semi-logarithmic approach in a regression linking individual earnings with additional years (or levels) of schooling completed (Mincer, 1974). As is well known, establishing a causal relationship between education and earnings is problematic. Among the issues to contend with are biases due to omitted variables, measurement error in reported schooling, distinguishing between homogenous and heterogeneous returns to education, and selection into wage employment. Moreover, while human capital theory hypothesises a concave education-earnings profile and diminishing returns to human capital production, empirical evidence from various countries has challenged the prevailing view (see Behrman and Wolfe, 1984 and Alderman and Sahn, 1988; and more recently Kingdon, 1998; Kingdon and Unni, 2001; Duraisamy (2002); Belzil and Hansen, 2002; Söderbom, Teal, Wambugu and Kahyarara, 2005; Ashraf and Ashraf, 1993a, 1993b; and Nasir, 2002). This finding raises serious policy concerns and warrants further investigation.

Despite these concerns, Mincerian returns remain popular and have been widely estimated (see Psacharopoulos, 1994, and Psacharopoulos and Patrinos, 2004). Estimates of private returns to education, though available for Pakistan are mostly dated and often constrained by data (Hamdani, 1977; Haque, 1977; Guisinger, Henderson and Scully, 1984; Khan and Irfan, 1985; Shabbir, 1991; Shabbir and Khan, 1991; Ashraf and Ashraf, 1993a, 1993b; Shabbir, 1994; Nasir, 1998; Siddiqui and Siddiqui, 1998; Nasir, 1999; Nasir, 2002; Asadullah, 2005, and. Riboud, Savchenko and Tan, 2006). There are two consistent findings from past studies in Pakistan: (i) returns to education are low as compared to other developing

[1] Alternative explanations of gender differentiated parental treatment are: a) pure son preference and b) that the returns accruing to parents from a daughters' education are lower than those accruing from sons' education (maybe because daughters' in-laws reap the benefits of her education upon marriage) and economic necessity (or parental selfishness) potentially increases the likelihood that boys are sent to school compared to girls. However, Alderman and King (1998) note that it is difficult to distinguish empirically between these various explanations.

countries and (ii) returns increase with the level of education. The latter finding challenges the dominant view that the earnings function is concave.

The estimation of returns to education by gender has received less attention in the literature partly because in many countries gender differences are not so large. When estimates are available, the evidence from developing countries is mixed. While some studies find returns to schooling to not differ significantly by gender (Behrman and Wolfe, 1984 and Schultz, 1993), others discover lower returns to women's schooling (Kingdon, 1998) or higher returns (Behrman and Deolalikar, 1995 and Asadullah, 2006). Previous studies in Pakistan mostly compute returns to education for males only and hence, are not able to answer the central question addressed in this study: does the labour market explain lower female schooling in Pakistan? Two recent exceptions are Nasir (2002) and Riboud *et al.* (2006). While the former (Nasir, 2002) implies that the answer to this question is a 'yes', the latter (Riboud *et al.* 2006) finds higher returns to women's education, suggesting otherwise. These contradictory findings generate a puzzle in the literature. However, as neither of these studies addresses various methodological problems, their estimates could be biased, raising some uncertainty about their findings.

The objective of this paper is to estimate returns to education by gender in a consistent manner to determine whether childhood and adolescent education investments are affected by how the labour market rewards adult education. Both 'One Factor' and 'Multiple Factor' models are used. In the former, education is defined as a continuous variable (years of education completed). This is a restrictive specification as it assumes that the return to education is the same for different education levels. The alternative model ('Multiple Factor') specifies education in level form – each level is allowed to have a different effect on earnings. This is clearly more flexible than a quadratic specification that includes education in years and its square (Blundell, Dearden and Sianesi, 2005). Briefly, four main methods of estimation are utilised: (i) Standard Ordinary Least Squares (OLS); (ii) the Heckman two-step procedure which deals with the sample selectivity issues which arise because earnings are only observed for individuals who participate in the waged-labour force and who may therefore form a non-random sub-sample of the population; (iii) 2SLS estimates using family background measures (parental education and spouse's education) as instrumental variables for schooling, to deal with endogeneity and measurement error in schooling; (iv) household fixed effects estimation to control for unobserved family-specific heterogeneity. For this, estimates are based on spouse pairs, sibling pairs and parent-child pairs. In all four methods we allow for the possibility that parameters differ between the two genders, by estimating separate earnings functions by gender[2]. Latest, nationally representative data from the Pakistan Integrated Household Survey (PIHS, 2002) are used for the analysis.

The paper is structured as follows. Section 2 discusses the empirical strategy while Section 3 discusses the data. Section 4 analyses the empirical findings and Section 5 concludes.

2. Empirical Strategy

This study adopts the standard Mincerian approach of estimating earnings functions to compute rates of returns to education by gender. The earnings-schooling relationship can be stated in the form of a semi-logarithmic relationship as follows:

$$\mathrm{LnY}_i = \beta_0 + \beta_1 \mathrm{S}_i + \beta_2 \mathbf{X}_{1i} + \beta_3 \mathbf{X}_2 + \varepsilon_i \tag{1a}$$

$$\mathrm{S}_i = \gamma_0 + \gamma_1 \mathbf{X}_{3i} + \gamma_2 \mathbf{X}_4 + \mu_i \tag{1b}$$

In (1a), LnY_i is the log of earnings[3] of individual *i*, S_i measures years of completed schooling in a 'One Factor' model (or levels of schooling with dummy variables representing various levels of completed schooling in 'Multiple Factor' models), $\mathbf{X}_{1i}$ is a vector of observed characteristics of individual *i*, $\mathbf{X}_2$ is a vector of observed characteristics of the family and ε_i is the individual-specific error. Equation (1b) models determinants of schooling where $\mathbf{X}_{3i}$ is a vector of observed characteristics of individual *i*, $\mathbf{X}_4$ is a vector of household-level covariates and μ_i is a residual term. The coefficient on schooling, β_1, measures the rate of return to each additional year of schooling (or to a particular level of schooling). This formulation assumes that the rate of return estimate is 'homogenous' i.e. identical across all individuals, *i*.

We start by estimating Ordinary Least Squares (OLS) models of earnings functions on male and female wage earners to provide some baseline results. However, OLS estimates of earnings functions potentially suffer from sample selectivity, omitted variables and measurement error biases. On the first, earnings are observed only for individuals participating in the paid labour force. Moreover, most studies focus on waged-workers while many individuals in developing countries, especially men, are self employed rather than in waged work. Consequently, estimates of returns to education of wage-workers are on a potentially non-random draw from the population, resulting in sample selection. In most applied work, Heckman's correction for sample selectivity is used. This entails estimating a waged-work participation equation and the predicted probabilities of waged work from this

[2] In this study we are unable to deal with measurement error (except when using the IV method) and although we are able to sign the bias (downward), its magnitude remains unquantified.
[3] Wages are a better measure of labour productivity as earnings incorporate labour supply decisions and a return to capital. Lack of data on wages often prompts use of earnings.

equation are used to derive the selectivity term, lambda (λ), which is then included in the main earnings function, such as (1a). To identify lambda the participation equation must include exclusion restrictions which are not part of the vector **X** in (1a).

The second problem has to do with omitted variable bias. The coefficient on schooling in the earnings function can only be interpreted as the causal effect of education on earnings if earnings differentials between individuals with varying years of schooling do not reflect differences in unobserved ability that happens to be correlated with education. Unobserved inherent ability is clearly a determinant of schooling attainment as well as of earnings, and generates endogeneity of schooling in the earnings function yielding inconsistent estimates of returns to schooling.

Finally, measurement error (ME) in the schooling variable S_i generates a correlation between the error terms in the earnings and schooling functions inducing attenuation bias in the regression coefficient β_1. This problem is compounded in sibling-studies as differencing within families reduces the true signal-to-noise ratio in schooling.

Various methods have been used in extant literature to address school endogeneity in an earnings function framework. The Instrumental Variables (IV) methodology identifies variables (instruments) that are correlated with schooling and uncorrelated with unobserved ability and measurement errors. This method provides a solution to endogeneity with the advantage that it simultaneously addresses ME issues. The key challenge is finding suitable instruments. Social and natural experiments are useful and many studies using 'institutional variations' in schooling due to such factors as proximity to schools, minimum school-leaving age etc. have been used to instrument for schooling. Card (1995b, 1999 and 2001) provides a summary of some of the recent studies that use this approach and include Angrist and Kreuger, 1991, Butcher and Case, 1994, Card, 1995a, and Harmon and Walker, 1995, among others. The consensus from contemporary research on developed countries is that IV estimates based on natural experiments are as high as and sometimes almost 20 percent higher than corresponding OLS estimates (Card, 2001). The evidence from developing countries is mixed and inconclusive (see Strauss and Thomas, 1995, for a review and Maluccio, 1998 and Duflo, 2001 for returns to education estimates using 'institutional variation' for Philippines and Indonesia).

However, experiment-based IV approaches have exacting data demands and an alternative is to use non-experimental IVs for endogenous schooling. As children's schooling outcomes are to a large extent driven by family background (FB), variables such as father's education and mother's education are sometimes used (Söderbom *et al.*, 2005 and Trostel *et al.* 2002 are examples of two recent studies). FB variables constitute valid instruments if they affect earnings only indirectly through their effect on schooling, i.e. if there is no intergenerational transmission of ability. FB then enters the vector of variables in equation

(1b) which directly influence schooling[4]. Alternatively, a number of studies use FB directly in earnings functions on the grounds that FB proxies omitted ability, school quality and out-of-school learning environment or reflects nepotistic family connections (see Heckman and Hotz, 1989 in Panama; Lam and Schoeni, 1993 in Brazil; Krishnan, 1996 in Ethiopia; and Kingdon, 1998 in India). However, Card (1999, pp. 1825) is critical of the use of FB variables as controls in earnings functions: inclusion of FB in earnings functions may reduce the bias but will still yield an upward biased estimate of rates of return unless all of the unobserved components are completely absorbed in the FB variables (Card, 1999, pp. 1825-1826).

An alternative to the IV technique is to either use repeated observations on the same individual over time or observations from different individuals within the same family to 'difference out' the variables generating correlation in the residuals in a 'fixed effects' approach. Arguably, at good part of the unobserved heterogeneity is common to family members. Consequently, differences in unobserved ability and their impact in determining education should be lower *within* rather than *between* families. Earnings functions can be estimated on twin-samples, siblings, father-son or mother-daughter pairs using a 'fixed effects' or first-differencing approach. By introducing sub-samples of households with at least two individuals of a given gender in wage employment (and more stringently households with brothers/sisters, father-son or mother-daughter pairs in wage employment) the fixed effects method effectively controls for all household variables that are common across these individuals within a given household. A simultaneous advantage of the fixed effects procedure is the elimination of the sample selection problem (Pitt and Rosenzweig, 1990, pp.978 and Behrman and Deolalikar, 1995, pp. 106).

Card (1999) provides an excellent summary of findings from twin and sibling studies in developed countries. In almost all instances, fixed effects estimates of the return to education are smaller than naïve OLS estimates suggesting an upward bias in the latter. However, data differencing exacerbates ME problems in sibling studies as part of the true signal is differenced out within families and the return to education is biased towards zero (Griliches, 1979). The finding of smaller estimated returns in sibling studies gives credence to the suspicion that these studies suffer potentially severe attenuation bias. However, research in recent years overcomes measurement error problems and concludes that fixed effects estimates corrected for measurement error are still smaller than OLS estimates (Ashenfelter and Rouse, 1998, Rouse, 1999 Hertz, 2003).

[4] Such that if individual ability is an unobservable in the error term in earnings functions (ε_i), family background instruments (Z_i) must not be correlated with the error i.e. Corr (Z_i, ε_i) = 0.

3. Data and Variable Specification

The PIHS (2002) data set is used in the analysis. This data set collected information on employment and earnings of all males and females aged 10 and above. Earnings information was collected for the past month from those able to report on a monthly basis, and yearly earnings information for annual reporters. We restrict analysis to adults aged 15 to 65 reporting waged-work employment. Consistent with previous literature, full-time students ('currently enrolled in school') are excluded from the sample. This yields a total of 13,519 adult males and females aged 15-65 reporting participation in waged employment.

Table 1 shows the distribution of the labour force by gender in Pakistan. There are striking gender differences in labour force participation rates: whereas 88 per cent males participate in the labour force, only 26 per cent of females do so. A relatively large proportion of males and females are engaged in self employment: 42 and 16 per cent, respectively. Gender differences in *waged*-work participation are particularly striking - 42 per cent males and only 7 per cent females are engaged in some form of waged employment.

Earnings functions are fitted on the sub-sample of waged-workers i.e. 11,501 males and 2,018 females. Selectivity-corrected earnings functions are fitted on wage-work participants with the reference category (or non-participants) including all other individuals (i.e. the unemployed, self-employed and the non-workers). The IV estimates are based on sub-samples of waged-workers who 1) report information on parental education and 2) who are married and report spouse's education. Finally, the household fixed effects methodology estimates earnings functions on sub-samples of households where at least one individual of each gender (male/female) is in waged employment (any relation, sibling pairs or father-son/mother-daughter pairs)[5].

The dependent variable in the participation equation is wage or salaried employment (PAID_EMPLOY) and that in the earnings functions is the natural log of monthly earnings (LN_MONTHLY_Y). The definitions of the variables used in the participation equation and earnings functions are given in Table 2. The education variable has been specified in two different ways, as years of completed education (EDU_YRS) and as education dummy variables representing various levels of education (LESS_PRIMARY, PRIMARY, MIDDLE, MATRIC, INTER, BACHELORS and MA_MORE). The reference category for the dummy-variables specification is individuals with no education. The vector of exclusion restrictions in the work-participation equation includes the following demographic variables: CHILD7 =

[5] The divisions are based on the notion that 'any' relation may not have the same genetic ties as blood relations. Although our sample of 'All' relations in household fixed effects estimates excludes non-blood relations such as parents-in-law and any servants residing in the household, it includes grandchildren. To increase the robustness of the estimates, we divide individuals into tighter groupings: sibling pairs (brothers and sisters) and parent-child pairs. The father-son pairs are, more specifically, male children of the household head. The mother-daughter pairs are the female children of spouses of the household head.

number of children aged 7 or less in the households; ADULT60 = number of adults aged 60 or more in the household; MARRIED = 1 if individual is married, 0 otherwise and HEAD = 1 if individual is the household head, 0 otherwise.

The earnings functions include experience and its quadratic (EXP and EXP2). This variable is often computed as: (Age – years of schooling – 5) on the belief that individuals start schooling at the age of 5 and enter the labour market upon completing schooling. This computation can be misleading in developing countries for at least two reasons. Firstly, individuals do not necessarily enter school at the age of 5 and secondly, a large majority of the labour force is illiterate and may not have attended any formal schooling. The PIHS is unique in that it asks individuals who attended formal schooling, the age at which they entered school. For individuals with positive years of schooling, EXP is computed as (Age – Years of schooling – Age entered school). For individuals with zero schooling, EXP = (Age – 14)[6].

Table 3 presents the means and standard deviations of the variables used in the participation equation by gender separately for wage-work participants and non-participants. Table 4 shows the means and standard deviations of variables used in the earnings functions. The last column shows a t-test of the gender difference in the means. Table 4 suggests that in waged employment males earn very substantially more than females. In logs, male earnings are 24 per cent higher than female earnings. The disparity in earnings is more apparent from Table 5 which shows average monthly earnings of waged employees by gender and education level. At *all* education levels, male earnings are significantly greater than female earnings. The gaps in earnings are similar across education levels but are highest at BACHELORS. In absolute terms, on average male earnings are a massive 113 per cent higher than female earnings.

Men in waged employment are not significantly more experienced than women but have completed significantly more years of education. Female workers report greater father's education than male workers (4.4 versus 2.9 years) suggesting that women wage-workers are a select group in the population with more educated parents and possibly with different aspirations and motivations as compared to non-workers. The proportion of women workers who have attended private schools (PRIVATE) is also marginally significantly greater than those of males. As private schools in Pakistan are believed to be of better quality than government schools, the descriptive statistics highlight the importance of controlling for sample selectivity into waged-employment, especially for female workers.

[6] The calculation of EXP tacitly assumes no grade repetition.

4. Econometric Results

Earnings functions are estimated using four methods: 1) OLS, 2) the Heckman two-step, 3) 2SLS and 4) household fixed effects. The results are divided into three sections. The first section reports probit estimates of waged-work participation by gender. The second presents OLS, Heckman, 2SLS and fixed effects estimates of earning functions. The final section extends the analysis by relaxing the restrictive assumption of linearity in the 'years' specification, introducing occupation and industry controls and, finally, decomposing the gender wage gap using Oaxaca's (1973) method. Unless stated otherwise, equations are fitted separately on males and females aged 15-65 in wage-employment.

4.1 Wage Work Participation (WP)

Table 6 presents the results of probit estimation of waged-work participation. It is clear that factors determining male and female Wage-work Participation (WP) differ significantly. For both genders, education has a largely U-shaped relationship with WP – the coefficients decrease and then increase in magnitude with higher levels of education. The effect of education on WP is stronger for women than for men, with a larger number of education splines significant for females as compared to males.

Being married has a significantly positive association with male WP, increasing the probability of waged work by 4 percentage points. Marriage has a significantly negative association for females, reducing their probability of WP by almost 3 percentage points. This gendered association is reflective of economic responsibility: for males, marriage increases financial responsibility while for females the reverse is true. Marriage and WP may also be jointly determined. Unearned income has a small but statistically significant negative effect on WP for males and females indicating a reduced need to work with alternative sources of income. *A Priori*, household demographics are expected to have a significantly larger effect on female WP as cultural norms in Pakistan delegate the role of 'carers' (of young children and old people) to women. However, contrary to popular expectation, a larger proportion of CHILD7 and ADULT60 has a significantly negative association with both male and female WP with the marginal effects somewhat larger for men than for women[7].

Finally, there are interesting regional and provincial disparities in WP across genders. Although both males and females are significantly more likely to be waged-workers in urban areas than in rural, the effect is large for men (14 per cent) and tiny for women (less than 1 per cent) suggesting that the types of employment available to men and women differ between urban and rural areas.

[7] The male-female coefficients on CHILD7 are significantly different (computed Wald value = 131.6) but insignificant for ADULT60 (computed Wald value = 1.78).

4.2 Earnings Functions

OLS and Sample Selectivity Bias (SSB) Estimates

Ordinary Least Squares estimates of returns to education are presented in Table 7. Columns (a) and (c) report findings for 'Years' of education (EDU_YRS). Columns (b) and (d) depict results for education 'Levels'. Focus on columns (a) and (c) first. The key parameter of interest is the point estimate on EDU_YRS – the rate of return to an additional year of schooling. The marginal rate of return to schooling is 7.2 per cent for males and 16.6 per cent for females. The return to education for women is more than double that for men in Pakistan. A Wald test confirms that the two coefficients on EDU_YRS in (a) and (c) are statistically very significantly different[8]. This baseline result implies large and significant gender differences in returns to education in Pakistan.

Turn now to columns (b) and (d) for the 'Levels' specification. This model relaxes the assumption of linearity of education implicit in columns (a) and (c). Some striking findings emerge. Firstly, the coefficients on education levels are positive and progressively increasing with higher levels of education for both genders, indicating a convex relationship between education and earnings. Secondly, the coefficients at all education-levels are significantly higher for females than for males[9]. The returns to additional years of education at various levels (Table 7a) show that returns to female education are always higher than returns to male education[10]. However, while returns increase for both males and females till INTER, they decline and then increase again at higher education levels for both genders, but more for females than for males.

Thirdly, the increase in coefficients with education levels is much sharper for women than for men, suggesting that the earnings profile is more convex for females as compared to males[11]. Finally, there is a premium in returns from PRIMARY to MIDDLE for females (coefficients increase from 0.34 to 0.96) with the increase being substantially smaller for males (0.14 to 0.27).

[8] The computed Wald statistic is 185.90 which is significant at the 1 % level.

[9] Wald tests for difference in male and female coefficients on the levels of education in (b) and (d) result in the following computed values of the chi-square statistic: 5.1 (LESS_PRIMARY), 2.5 (PRIMARY), 23.6 (MIDDLE), 63.7 (MATRIC), 76.4 (INTER), 97.9 (BACHELORS) and 104.2 (MA_MORE). Except PRIMARY, all values are significant at the 5% level (critical chi2 value is 3.84).

[10] The coefficients in the 'Levels' specification in Table 6 have to be transformed to arrive at the 'returns' as the number of years of education is different for the various levels of education indicated by the dummy variables and as measured here, the wage premia for a graduate of a higher level include the premium from a lower level of education.

[11] Interestingly, in column (d), the return to LESS_PRIMARY and PRIMARY for women is almost identical (34 per cent), rising to almost 96 per cent for MIDDLE schooling suggesting that for women, the return to acquiring *any* education below or up till primary schooling has a return in the labour market.

However, OLS estimates may be biased due to sample-selection and endogenous schooling. We turn next to the SSB estimates (Table 8) which correct for selection bias by using the Heckman two-step procedure and incorporate LAMBDA into earnings function estimates. The selectivity-corrected earnings functions reported in Table 8 include the standard variables – education, experience and its square and the provincial and regional dummies. Household demographic variables (CHILD7, ADULT60) and LNUNEARNED_Y are used as exclusion restrictions. These variables are believed to determine participation in waged-work but do not directly affect labour market earnings. All are individually statistically significant.

The LAMBDA term is large and statistically significantly negative for males (in both Years and Levels specifications) and significant for females only in the Years specification. A comparison across columns (a) and (c) and across columns (b) and (d) in Tables 7 and 8 reveals the effect of correcting for sample selection. Inclusion of the LAMBDA term reduces the point estimates on years of education from 7.2 per cent to 6.4 per cent for males and 16.6 per cent to 14.2 per cent for females.

These differences are statistically significant[12]. In the Levels specification, the inclusion of the LAMBDA term has no significant attenuating effect on the education-coefficients in the female sample (consistent with LAMBDA being insignificant) but in the male sample inclusion of LAMBDA has a significantly attenuating effect on some education-level coefficients (BACHELORS and MA_MORE)[13]. Finally, specifying education in levels rather than as EDU_YRS has an attenuating effect on the point estimate of LAMBDA which falls by a larger absolute value for females as compared to males. The change in LAMBDA coefficients is significant for males but insignificant for females[14]. Overall, these findings suggest that OLS overestimates the return to education (especially in the Years specification).

Importantly, however, the return to female education remains significantly greater than that for males even after controlling for selection bias - the marginal return to schooling is 6.4 per cent for males and 14.2 per cent for females (columns (a) and (c) in Table 8). As with simple OLS, the return to education for women is more than double that for men in Pakistan. This difference is statistically significant[15]. Experience and its square have a fairly standard relationship with earnings for both genders, increasing albeit at a diminishing rate.

[12] The computed Wald statistic for the difference in EDU_YRS coefficients across the OLS and SSB specifications in columns (a) and (d) in Tables 4.7 and 4.8 is 13.69 for males and 4.95 for females suggesting we can reject the null hypothesis that the change in coefficients is equal to zero.

[13] The computed Wald statistics comparing the BACHELORS and MA_MORE coefficients across the OLS and SSB estimates for the males sample are 6.93 and 17.34 suggesting we cannot accept the null hypothesis of equality of coefficients at the 5 % level.

[14] The computed Wald statistic for males is 24.36 and for females is 3.53. The critical chi2 value at the 5% level is 3.84 suggesting that we can reject the null hypothesis that the change in coefficients between the two specifications is equal to zero for males but not for females.

[15] The Wald test results in a computed value of 1183.0 which is statistically significant at the 5% level.

Earnings peak at 25 years and 31 years of experience for males and females respectively. Both genders earn more in urban than in rural regions. The results in the 'Levels' specification reported in columns (b) and (d) are consistent with the OLS findings[16]. Note that the jump in coefficients from PRIMARY to MIDDLE remains even after controlling for selection into waged work.

To conclude this section, we find that the return to women's education is significantly higher than to men's, both in 'Years' and 'Levels' specifications and that the education-earnings profile is convex for males and females in Pakistan. The first result corroborates some studies in Pakistan and diverges from others. The convexity result is corroborated by recent evidence from a number of studies from an array of countries - Kingdon (1998) and Kingdon and Unni (2001) on India, Belzil and Hansen (2002) for USA and Söderbom *et. al.* (2005) in Kenya and Tanzania.

Instrumental Variable Estimates

As described above, OLS estimates of returns to education will be biased if years of schooling (EDU_YRS) are correlated with the error term, or if reported schooling is measured with error (endogeneity bias). The SSB estimates reported above may be biased upwards due to the classic 'ability bias'. IV estimates are more robust than the estimates from the previous section on two counts: 1) they control for endogeneity of EDU_YRS, thereby correcting for any upward 'ability biases' and 2) they are unaffected by measurement error so that the reported findings should be purged of any attenuating effects. However, IV estimates are based on selected samples (earnings functions are estimated on subsets of individuals reporting earnings in waged work). Sample selection issues may be further compounded as a small sub-sample of the population reports parental education and spouse's education, the instruments used here (see below). Controlling simultaneously for both sample selection effects and endogeneity of schooling in earnings function estimates would require an additional set of instruments that didn't directly affect either earnings or participation into waged work, a condition often very hard to meet given data constraints (Wooldridge, 2002, pp. 567).

We discussed above how previous literature has used family background variables to instrument for endogenous schooling[17]. We use parental education as instruments for the

---

[16] Wald tests for difference in male and female coefficients on the levels of education in (b) and (d) result in the following computed values: 5.7 (LESS_PRIMARY), 2.5 (PRIMARY), 23.2 (MIDDLE), 90.7 (MATRIC), 52.7 (INTER), 50.9 (BACHELORS) and 35. 3 (MA_MORE). Except PRIMARY, all values are significant at the 5% level (critical chi2 value is 3.84).

[17] As mentioned previously, a number of studies directly control for family background in earnings functions on the premise that family background impacts earnings either directly through nepotism or indirectly through school quality or out of school learning, in which case failure to account for it may subject estimates to 'family background bias'. However, this sub-section uses family background

subset of individuals reporting fathers' and mothers' education and alternatively spouse's education as an instrument for another subset of married wage-workers[18]. Education is instrumented using three variables: FEDYRS (years of education completed by worker's father), MEDPRIM (equals 1 if mother has completed any year of primary education, 0 otherwise) and MEDPRIMORE (equals 1 if mother has completed more than primary education, 0 otherwise)[19]. Mothers reporting no education (MEDNONE) are the omitted category. Parental education may be good instruments for own-schooling if parent's education positively affects schooling but is not correlated with child ability (which is in the error term of the earnings function), i.e., assuming no intergenerational transmission of ability[20]. In the sample of wage-workers aged 15-65 from the PIHS (2001-2002), between 24 and 26 per cent of the variation in education of males and females, respectively, is explained by father's education[21].

The IV analysis using parental background is augmented using spouse's education (SPOUSE_EDU = years of education completed by spouse) as another instrument to compare the findings with parental-education estimates. This draws on the theory of assortative mating (Weiss, 1999): individuals with common social backgrounds, religion, race and caste are more likely to bond together in marriage. This is accentuated by the high correlation between spouse's education and own education in Pakistan (0.29 for males and 0.51 for females using the PIHS, 2002). Family background variables such as parental or spousal education have been used as instruments in previous work on rates of returns estimation. However, given the criticism of such variables as valid instruments, i.e. due to their possible endogeneity, the findings from this section will be interpreted with caution.

---

variables as instruments for the worker's schooling under the assumption that there is no intergenerational transmission of ability and that family background affects earnings only indirectly through its effect on worker's schooling. Hansen's J statistics test confirms the relevance of the instruments used.

[18] These subsets may differ from the population in that individuals reporting parental or spousal education may belong to subsets of the population which are not random draws. This suggests a potential sample selection issue which we are, unfortunately, unable to deal with here.

[19] We experimented with a number of instruments for mother's education. The set of instruments that satisfied the over-identification test and seemed justifiable and theoretically plausible was chosen. For example, in Pakistan, with a large number of mother's reporting no education (90 per cent in the IV sub-sample), it made more sense to define mother's education in terms of dummy variables capturing critical levels of education rather than completed years.

[20] Even assuming no intergenerational transmission of ability, these instruments cans still be criticised on the grounds that parental education may either have a direct effect on individual earnings in the labour market (nepotism and family connections) or indirectly through its effect on school quality. These arguments make a case for including parental education as control variables in earnings functions rather than using them as instruments for schooling.

[21] We also have data on maternal education for a subset of workers. However, the correlation between mother's education and own-education is relatively low (7 and 15 per cent for males and females respectively).

It was stressed above that instruments must be good and valid: variables that are correlated with education and uncorrelated with the residual in the earnings function. To determine the empirical 'goodness' of the instruments, turn to the first-stage estimates reported in columns (b), (d), (f) and (h) in Table 9. For both genders, the first stage equations reveal that FEDYRS, MEDPRIM, MEDPRIMORE and SPOUSE_EDU almost always have large, very precisely determined coefficients with the expected signs. The 'relevance' of the instruments can be assessed by examining the significance of the excluded instruments in the first-stage IV regressions. The objective of this test as suggested by Bound, Jaegar and Baker (1995) is to determine that education is correlated with the instruments, controlling for all other variables. The p-values of the F tests in the first-stage regression indicate that the instruments satisfy the 'relevance' condition very well. If the instruments used are not 'valid' i.e. if $Corr(Z_i,\varepsilon_i) \neq 0$, the IV estimates will be inconsistent. The only way to assess the validity of the instruments is to have a surfeit of instruments and use an over identification (OID) test. This is only possible for the IV-sample using parental education since we have both mother's and father's education. The Hansen's J test of over-identifying restrictions is used. In both the male and female samples using parental education instruments, the p-value of the OID test does not reject the null hypothesis, confirming the validity of instruments used.

2SLS estimates are reported in Table 9. Education is specified as a continuous variable[22]. The first four columns report IV estimates using parental education and SPOUSE_EDU for males and the latter four report those for females. Columns (b), (d), (f) and (h) report first-stage results. Focus first on columns (a), (c), (f) and (g), the earnings functions estimates. The summary statistics reveal a fairly good fit with the R^2 ranging from 0.28-0.30 for males and between 0.47-0.48 for females. The rate of return to an additional year of schooling is between 10 and 11 per cent for males and between 17 and 18 per cent for females using either instrument. The main findings are: 1) as before, the rate of return to education is always higher for females as compared to males and 2) consistent with the findings from numerous other studies, β_{IV} (10-11 per cent for males and 17-18 per cent for females) is larger than β_{OLS} (7 per cent and 17 per cent for females).

Fixed Effects Estimates

We turn now to fixed effects estimates of returns to education. The results are based on sub-samples of at least one male and one female wage-worker within a household who are related in any way (e.g. father-daughter, mother-son, brother-sister or husband-wife) or are siblings (only brother-sister pairs). Table 10 depicts the fixed effects estimates: column (a) for 'All' relations and (b) for sibling pairs. As before, education is measured in 'Years' and

[22] With few instruments per sub-sample, the 'Levels' specification cannot be used in 2SLS.

'Levels'. The set of independent variables remains unchanged with two exceptions. A gender dummy (MALE) is added as the sample includes individuals of both sexes, and interaction terms (EDU_YRS_MALE in 'Years' and LESS_PRIMARY_MALE etc. in 'Levels') are included. These capture the effect of gender on the return to education. Finally, in (a) and (b), OLS estimates are reported along with FE estimates for comparative reasons.

Focus first on the 'Years' specification in columns (a) and (b). The returns to male education have been computed as the sum of the coefficients on EDU_YRS and EDU_YRS_MALE (for example, in 'All', the coefficient on EDU_YRS = 0.14 while that on EDU_YRS_MALE = -0.082. The overall return for males is the sum of 0.14 and -0.082 which equals 0.058, or approximately 6 per cent). Thus, using FE estimates, the return to female education is clearly substantially higher than that to male education for 'All' individuals (14 per cent for females compared to 6 per cent for males) and for 'Siblings' (15 percent versus 11 per cent respectively).

Note that the FE point estimates on EDU_YRS and EDU_YRS_MALE are lower than the OLS estimates, a finding consistent with previous literature[23]. However, although the FE estimates are lower, they do not collapse and are reasonably close to the OLS estimates. Part of the decline in estimates could be because of an upward bias in the OLS estimator due to omitted variables. Some part (albeit unmeasured) of the attenuation could be attributable to ME. Although studies such as Hertz (2003) which correct for ME in within-household estimators *still* find the within-household estimate of the return to education to be smaller than the corresponding OLS estimate, the correction for ME causes a rise from the un-corrected estimate. Data constraints prevent such a correction in the current study[24].

We turn now to the 'Levels' specifications in (a) and (b). Firstly, it is clear that, except for primary education, the FE returns to female education are higher than those for males at all levels of education in (a) and (b). This suggests larger labour market incentives for females (than males) to acquire education. Secondly, the FE findings confirm that the convexity of the education-earnings profile in previous sections is not an artefact of heterogeneity. Thirdly, although evidence points to $\beta_{FE} < \beta_{OLS}$ for a majority of cases in 'All', this is not true for 'Siblings'. Finally, in both 'All' and 'Siblings' samples, the jump in returns from PRIMARY-MIDDLE remains for females – for example, in the 'All' sample, the coefficients for females increase from 0.126 (PRIMARY) to 0.822 (MIDDLE) while those for males are roughly the same for both levels of education (between 0.18 and 0.19).

[23] Hertz (2003) in a recent study using data from South Africa, finds that whereas OLS estimates yield returns of about 13 per cent, the return to education is 3 per cent when using FE. ME correction causes the estimates to rise to 5 per cent. Behrman and Deolalikar (1995) also find FE estimates for male and female workers to be significantly lower than corresponding OLS estimates.
[24] Hertz (2003) corrects for measurement error in schooling using two observations of schooling on the same individual. This was made possible as 13 per cent of the individuals in the sample were re-surveyed to obtain measures of reliability of measured schooling.

Section Summary

Regardless of the empirical methodology adopted, there are a number of consistent findings. Firstly, the estimated marginal returns to additional years of schooling are significantly higher for females than for males. Secondly, returns increase with higher levels of education, pointing to convex education-earnings profiles. Finally, the labour market differentially rewards males and females with relatively low education levels (primary and middle) – women with middle schooling are rewarded substantially more as compared to women with primary education. This is not true for males.

However, a number of questions warrant investigation. The following subsections deal with three further issues:

1) The convexity of education-earnings profiles evidenced so far could be an artefact of endogenous schooling as it is driven by results from the 'Levels' specification where we are unable to control for endogenous schooling levels. This is because of a lack of instruments. It is of interest to see whether the earnings profile remains convex or collapses into the more conventional concave shape after controlling for endogeneity;
2) Given evidence of a 'collapse' in returns to primary and low levels of schooling across Africa and India (Moll, 1996; Schultz, 2004), the finding of large and significant returns for primary and middle schooling for both genders in Pakistan is striking. One wonders why the returns to low-schooling-levels are so high, especially for women. Moreover, what drives the jump in returns from primary-middle schooling for females and the lack thereof for males? We investigate this issue further.
3) Table 5 revealed that at all levels of education male earnings were greater than female earnings. The FE model in Table 10 estimated on the pooled sample also showed a large coefficient on the MALE dummy suggesting that one way labour markets favour males is by providing a wage premium to being male i.e. there could be differential treatment in the labour market other than through differential rewards to education. In order to investigate this, we decompose the total gender wage gap into the 'explained' and 'unexplained' portions using Oaxaca's (1973) methodology.

4.3 Extended Earnings Functions

Non-Linear Earnings Functions

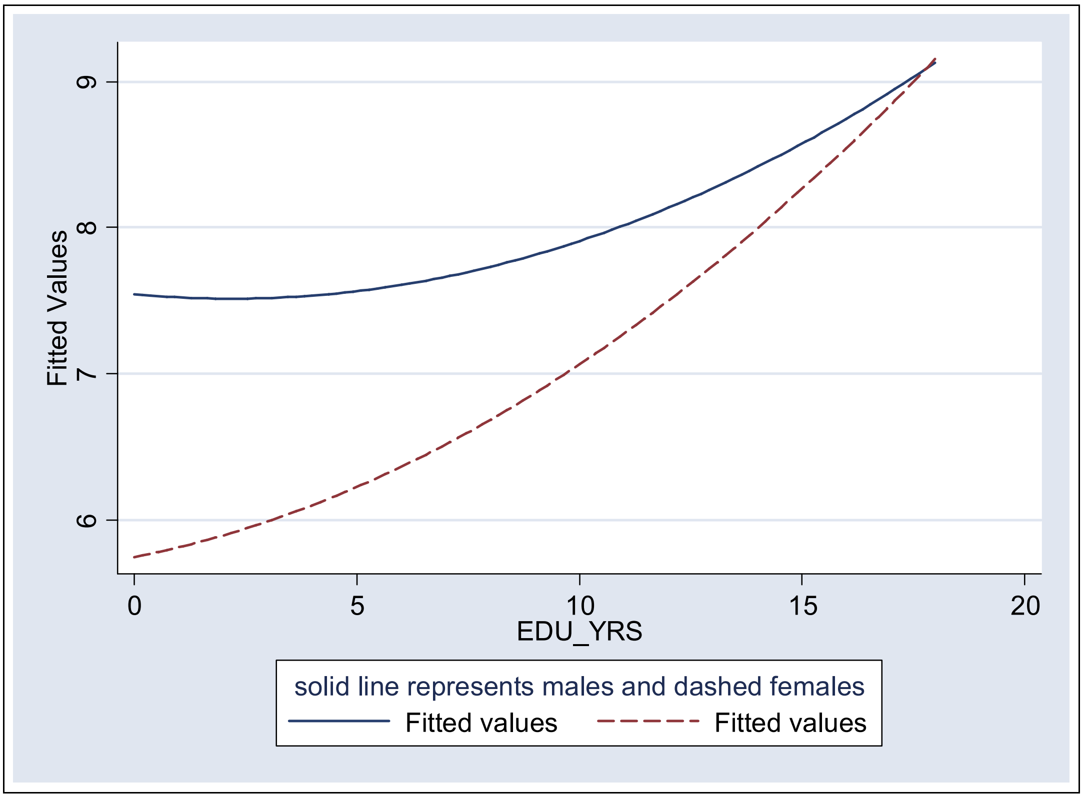
So far, the continuous specification of the earnings function ('Years') has assumed linearity in schooling (captured in EDU_YRS). The theoretical literature suggests that in fact the relationship between education and earnings may be concave due to diminishing returns to education. There is empirical support for this concavity (Psacharopoulos and Patrinos, 2004).

However, Schultz (2004, pp. 123) argues that "…there is nothing in the human capital framework which prescribes this [concave] pattern; returns will vary according to supply and demand conditions in the labour market." In recent years, the implicit concavity of education-earnings profiles has been challenged in studies which find returns increasing with higher levels of education i.e. convex profiles (see Card, 1999; Kingdon, 1998; Belzil and Hansen, 2002; Söderbom *et al.*, 2001; and Schultz, 2004). For the purposes of this chapter, firstly, *if* the education-earnings profile is indeed concave, imposing linearity as we have done in the continuous specification is too restrictive. Secondly, documenting the shape of the education-earnings profile is important because of its potential effect on the estimation of the relative rates of return to schooling for women and men.

Using the 'Levels' specification, we have already found some evidence of convex earnings profiles for males and females with sharper convexities for the latter rather than the former. However, OLS and SSB results in Tables 7 and 8 do not control for the possibility that unobserved ability may be correlated with both earnings and individual schooling within the 'Levels' framework. Hence, one wonders whether the finding of convexity is an artefact of endogeneity. Although the FE estimates also reveal sharply convex earnings profiles for males and females, they are based on smaller samples. Hence, this section introduces a quadratic term for years of education (EDU_YRS2) within the 'Years' specification to relax the linearity constraint previously imposed. Figure 1 below graphically illustrates the education-earnings profiles for males and females fitted using OLS on the sample of waged workers[25]. Clearly, female waged workers earn less than males at each level of education and the education-earnings profiles for both genders are convex. However, EDU_YRS and EDU_YRS2 underlying these depictions are potentially endogenous and we turn next to controlling for endogeneity of schooling in a non-linear setting.

[25] The regression results underlying Figure 1 (also used for the Oaxaca decomposition) are suppressed due to space limitations. OLS functions were fitted for males and females with the standard regressors used so far as independent variables but EDU_YRS2 was incorporated to allow for non-linearity. The estimates were robust and corrected for clustering at the population sampling unit level.

Figure 1.1: Education-Earnings, Profiles Males and Females (15-65)



The endogeneity of EDU_YRS and EDU_YRS2 is tackled using a two-stage control function approach. In the first stage EDU_YRS is regressed on a set of instruments (which include FEDYRS, MEDPRIM and MEDPRIMORE). Two equations are fitted, one each for the male and female sub-samples and based on these regressions, residuals are estimated. In the second stage, earnings functions are estimated using the residuals from the first stage as control variables for unobserved ability. There are two main advantages of the control function approach. On the one hand it allows us to control for the endogeneity of a non-linear variable while on the other hand it allows an identification of the correlation (if any) of the unobserved variables and the potentially endogenous variable. The latter is a test of endogeneity – if the residual term is significant, it implies that the unexplained variation in EDU_YRS also affects variation in earnings. If the residual term is insignificant, one can accept the hypothesis that the schooling variable is not endogenous.

Table 11 reports findings from the first and second step of the control function estimates. Control function models can only be estimated on sub-samples of individuals reporting parental education and for comparative purposes, corresponding OLS estimates are reported on the same sub-samples. Turning to the first stage of control function estimation (see columns b for the males and females respectively): parental education significantly positively determines the years of schooling completed by an individual. However, the effect is stronger for females – mother's education (MEDPRIM and MEDPRIMORE) is significantly positive only in the female sample and the size of the coefficient is double that in the male sample.

Column (c) reports the second stage estimates: earnings functions incorporating the residuals from the first stage. Whereas the residual term is negative and significant at the 1 per cent level for males, it is insignificant for females. This suggests that although we can accept the null hypothesis of exogeneity of schooling for the female sample, we cannot do so for the male sample. Moreover, even after controlling for potential endogeneity of schooling, the education-earnings profiles are convex. Table 12 computes the marginal return to schooling for various years of completed schooling using the OLS and CF estimates and confirms the convexity of the education-earnings profiles. Finally, a comparison across columns (a) and (c) in Table 11 and in Table 12 (OLS and CF estimates), reveals that in most cases they are not significantly different from each other[26]. Consequently, in the extensions that follow, we report only OLS and/or SSB estimates.

What explains the premium to women with middle education?

We noticed in Tables 7, 8 and 10 that: 1) the returns to low levels of education were high especially for women and 2) there was a large premium to women for possession of middle level education and the premium was greater than that for men. We wish to investigate the labour market realities underlying this result. One possibility is that women's higher economic benefits from education are realised through better occupational attainment or better industry attachment. In order to test this, we include industry and occupation dummies in earnings functions specifications. Findings are reported in Appendix Tables A1 and A2 and OLS and SSB models fitted on 'Levels' specifications.

The results suggest that indeed the effect of education on earnings occurs partly by permitting better occupational attainment. The coefficients on OLS/SSB with controls decline for both males and females when occupation/industry dummies are included. However, this is not the only mechanism as there are large and significant direct returns to certain education levels (especially middle) even within occupations and industries. In particular, whereas for men education impacts their earnings via occupation/industry at all education levels, for women the effect of occupation and industry association only operates after matric[27].

---

[26] Wald tests comparing the male and female coefficients on EDU_YRS and EDU_YRS2 reveal the following computed values of the chi-square: 7.04 (EDU_YRS) and 0.125 (EDU_YRS2) for males and 7.12 (EDU_YRS) and 0.00 (EDU_YRS2) for females. The null hypothesis that the coefficients across OLS and CF models are not significantly different is accepted for the quadratic term but rejected for the linear term (EDU_YRS) for both males and females.

[27] This is reflected by a significant Wald statistic for the OLS and SSB specifications without and with occupation/industry dummies for males at all education levels and the significant value for females only from Matric onwards. The Wald statistics comparing OLS with and without industry/occupation dummies for males are: 4.3 (Primary), 5.5 (Middle), 7.8 (Matric), 4.6 (Inter), 6.5 (Bachelors) and 7.4

Therefore, there are large and significant direct returns to women's education at primary and middle levels which are not explained by occupation and industry attainment. This is corroborated by the raw data which shows the large increase in earnings in the occupations across various education levels. One plausible explanation for this finding is that there is a scarcity premium to women's education in Pakistan. For example, 58 per cent of the wage-working women in our sample in Table 4 report having no education. Moreover, Table 4 shows that the proportion of wage-working women with middle schooling is 3.9 per cent (compared to 5.2 per cent women with primary education and 12.2 per cent men with middle education). If there is non-substitutability of jobs (i.e. certain jobs can only be performed by women for example teachers in single-sex girls' schools) or job-reservation quotas in the government sector, the returns to women meeting minimum education qualifications will be high[28].

Does the labour market discriminate against women?

This study has so far found that returns to education are significantly greater for women than men. This raises something of a puzzle as to why then parents allocate lower education to girls than boys. In this section, we investigate whether the total labour market return to boys is greater than to girls as a way of probing parental investment motives further.

We decompose the male-female wage gaps using the technique proposed by Oaxaca (1973). OLS and selectivity-corrected earnings functions (incorporating EDU_YRS2) are estimated to predict earnings. The wage gap is decomposed into two components: 1) that explained by differences in individual characteristics and 2) the residual, unexplained portion, reflecting differences in earnings structure. Assume that the mean earnings of females (f) are Y_f and those of males (m) are Y_m. Mean earnings will be determined by:

$$Y_i = b_i X_i \text{ where } i=m,f \tag{3}$$

(MA_MORE). For females the corresponding values are: 3.6 (Less_primary), 0.1 (Primary), 0.8 (Middle), 15.6 (Matric), 9.8 (Inter), 28.6 (Bachelors) and 27.5 (MA_MORE). Similarly, the Wald statistics for the SSB estimates with and without occupation/industry dummies are: 4.2 (Primary), 5.4 (Middle), 8.4 (Matric), 5.2 (Inter), 7.9 (Bachelors), 7.1 (MA_MORE) and for females: 0.24 (Less_primary), 0.31 (Primary), 1.0 (Middle), 17.9 (Matric), 10.9 (Inter), 9.1 (Bachelors), 3.4 (MA_MORE).

[28] One way to test for the reservation argument would be to include a dummy variable capturing whether the individual is a government or private sector employee. This variable was not available in the dataset. One explanation for why the jump may occur at the middle level could be because the teaching occupation becomes viable for women immediately after completion of middle schooling (the minimum stipulated requirement by the government). To test this, we defined occupations by including a dummy variable equalling 1 if the female reported employment as a teacher and 0 otherwise. However, the inclusion of the 'TEACHER' dummy didn't cause a significant decline in the coefficient on the middle dummy. The results including TEACHER dummies are not reported.

where X is the vector of average characteristics of i and b_i is the vector of estimated parameters for i. Standardising by male means, the total wage gap in mean earnings (T) can be divided into the explained (E) component and the unexplained (D) component as follows:

$T = Y_m - Y_f$

$T = b_m X_m - b_f X_f$

$T = \{X_m (b_m - b_f)\} + \{b_f (X_m - X_f)\}$

$$T = E + D \qquad (4)$$

Similar standardising can be achieved using female means. Using this method, one can decompose the total gender wage gap into the explained and unexplained components. The unexplained component could be seen as the extent of 'discrimination' in the labour market. However, if there are important differences in the unobserved or unmeasured characteristics of males and females, then the residual component cannot so validly be termed 'discrimination'. The Oaxaca decomposition is initially conducted using OLS. As a robustness check, we repeat the exercise using a household fixed-effects (FE) model on sub-samples of at least 2 waged-workers of each gender in a given household.

The results of the decomposition exercise are reported in Table 13. Using OLS, expressed in natural logs, the gross gender wage difference is 1.49. Standardising by male means, 0.25 of the 1.49 gender wage difference is explained by better male characteristics (such as higher educational attainment) while 1.25 of the gender wage gap remains 'unexplained'. Consequently, almost 84 per cent of the gender wage gap is unexplained. Standardising by female means suggests an even greater unexplained proportion (95 per cent). The fixed effects estimates show that an even larger proportion of the gender wage-gap is 'unexplained'. Fixed effects estimates provide a cleaner test since unobserved or unmeasured characteristic differences between males and females *within* the family are likely to be much lower than across families and the large 'unexplained' portion in the FE sample is indicative of high discrimination in the labour market.

However, since male and female hours worked may partly account for the large unexplained component of the gender wage-gap, we should ideally perform the decomposition including hours worked in the earnings function. Although we do not have information on hours, we do have data on days worked in the past month. However, mean days worked is very similar for males and females and including that variable in the OLS and FE regressions causes the 'unexplained' component to remain virtually unchanged (the average estimate of discrimination is 88 per cent in the OLS and 94 per cent in the FE sample).

These estimates of 'discrimination' in Pakistan are large in comparison to international estimates. For example, estimates of the 'unexplained' portion of the wage gap range between 1-5 per cent in the UK (Zabalza and Arrufat, 1983), 12 per cent in the US

(Choudhury, 1993) and 35-45 per cent in India (Kingdon and Unni, 2001). However, previous estimates in Pakistan range from 63 per cent in 1979 (Ashraf and Ashraf, 1993a), 33 per cent in 1985-1986 (Ashraf and Ashraf, 1993b) and between 86-96 per cent and 55-77 per cent in 1993-1994 (Siddiqui and Siddiqui, 1998). Clearly, our findings are closest to those of the last study, which is also the latest past study. The findings are suggestive of a pernicious increase in discrimination in the labour market over time. The large apparent 'discrimination' would not only explain the low participation of women in Pakistan's labour markets but also the large differentials in intra-household education expenditure allocations within households (Aslam and Kingdon, 2007). However, these conclusions are subject to an important caveat – the decomposition of the male-female earnings gap is based only on wage-earners in a conditional equation (conditional on being a wage earner)[29].

6. Conclusion

This study seeks an answer to the following question: does the labour market explain lower education of girls than boys in Pakistan? If the labour market rewards women less than men, scarce resources may be allocated efficiently though inequitably within the household. This question is addressed by estimating returns to education for males and females in wage-employment in Pakistan using household data from 2002. Four methods are used in an attempt to overcome limitations faced in conventional earnings function analyses: 1) OLS, 2) Heckman two-step, 2) 2SLS and 3) household fixed effects. The findings from all four methods consistently point to a sizeable gender asymmetry in returns. Females have significantly higher economic incentives to invest in education than males. The estimated return to additional years of education (EDU_YRS) ranges between 7 and 11 per cent for men and between 13 and 18 per cent for women. By this consideration, the labour market does not explain lower female schooling in Pakistan. If anything, it suggests there should be a pro-female bias in the household decision to educate. However, the Oaxaca decomposition suggests a large element of potential gender discrimination in the Pakistan labour market. While the return to education is considerably lower for men than women, total earnings are dramatically higher for men than women. While a large part of the male-female earnings differential is not explained by men and women's differing productive characteristics, one must be cautious in interpreting the residual unexplained earnings differential as labour market 'discrimination' since certain unobserved but relevant characteristics of men and

[29] The large 'unexplained' component in such conditional equations could be partly due to the fact that women's participation is constrained by cultural factors. A decomposition of the male-female earnings gap based on an 'unconditional' sample would presumably yield a larger 'gender gap' with the likelihood that productive characteristics (education and experience etc.) would explain a greater proportion of the gap.

women may not be controlled, such as their quality of schooling (men are more likely than women to have attended private schools) and certain variables may be measured with error, such as the years of labour market experience.

The coexistence of high returns to education for women and gender bias against them in household education decisions is a puzzle that demands explanation. One potential explanation is that even if the return to girls' education is higher than that to boys' education, the part of the return to daughters' education accruing to parents may be much lower than that accruing from a sons' education. The PIHS (2002) shows that only 6 per cent of adult daughters aged over 21 reside in their parental homes, suggesting that a majority are married and living with in-laws/husbands. Any returns from these daughters' education would accrue to the in-laws or the husband rather than to the parents. In order to investigate this explanation further, one would need data on transfers received by parents from their male and female offspring. Such data are, to our knowledge, not available.

A second potential explanation is that our estimate of the return to education is misleadingly high because it is estimated on the small wage employment sector whereas a relatively large proportion of women in Pakistan are self employed. Estimating the return to education accurately in self employment is difficult because earnings in self employment contain a return to physical capital as well and we do not have a good measure of physical capital in order to enable us to isolate the pure return to human capital.

This study also finds sharply convex education-earnings profiles for males and females. These findings are robust to control functions estimates. There are several policy implications of convexity of the education-earnings profile. Firstly, the 'higher returns at lower education levels' argument has often been used to justify allocating funds to expand primary education. If indeed the returns are greater at higher education levels, the economic efficiency rationale for channelling these funds to primary education may be diluted. However, this is not to say that all rationales for funding primary schooling are eliminated: there is a strong case for primary education in terms of its non-market returns and also in a rights-based perspective. In any case, the return to primary education includes the benefit that it permits access to further, more lucrative, levels of education. Although the education-earnings profiles in Pakistan are convex, the returns to primary schooling are high compared to other developing countries. This may reflect un-met demand within industry-sectors that need low-skilled labour and policy-makers may need to promote low-level education as well as adopt policies which encourage these individuals to participate in the labour market (especially women).

Secondly, and linked to the first, convexity has implications for increasing education inequality. If private returns to schooling increase with higher education, poorer families who educate their children till only say primary education will face lower returns while richer families who educate children till higher education will reap higher returns. Consequently, the poor are motivated to educate their children less and may also send only the more able children to school for whom returns are higher. Consequently, education and earnings differentials may widen both across families and within families (Schultz 2004).

We also find evidence of high wage premia to low education levels, especially for women in Pakistan. These large and significant direct returns to women's education at primary and middle levels are not fully explained by occupation and industry attainment and are interpreted to reflect scarcity premia in labour markets.

References

Alderman, H. and Sahn, D.E. (1988), 'The Effects of Human Capital on Wages, and the Determinants of Labor Supply in a Developing Country', *The Journal of Development Economics*, 29, pp. 157-183.

Alderman, H. and King, E.M. (1998), 'Gender Differences in Parental Investment in Education', *Structural Change and Economic Dynamics*, 9, pp. 453-468.

Alderman, H., Behrman, J.R., Ross, D. and Sabot, R. (1996), 'The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market', *Oxford Bulletin of Economics and Statistics*, 58 (1), pp. 29-55.

Angrist, J.D. and Krueger, A.B. (1991), 'Does Compulsory Schooling Attendance Affect Schooling and Earnings', *Quarterly Journal of Economics*, 106 (4), pp. 979-1014.

Asadullah, M.N (2006), 'Returns to Education in Bangladesh', *Education Economics* (forthcoming), also as QEH Working Paper 130, University of Oxford.

Asadullah, M.N. (2005), 'The Effectiveness of Private and Public Schools in Bangladesh and Pakistan', University of Oxford (Mimeo).

Ashenfelter, O. and Krueger, A. (1994), 'Estimating the Returns to Schooling Using a New Sample of Twins', *American Economic Review*, LXXXIV, pp. 1157-1173.

Ashenfelter, O. and Rouse, C. (1998), 'Income, Schooling and Ability: Evidence from a New Sample of Identical Twins', *Quarterly Journal of Economics*, 113 (1), pp. 253-284.

Ashenfelter, O. and Zimmerman, D. J. (1997), 'Estimates of the Returns to Schooling from Sibling Data: Fathers, Sons and Brothers', *Review of Economics and Statistics*, 79 (1), pp. 1-9.

Ashraf, J. and Ashraf, B. (1993a), 'Estimating the Gender Wage Gap in Rawalpindi City', *The Journal of Development Studies*, 29 (2), pp. 365-76.

Ashraf, J. and Ashraf, B. (1993b), 'An Analysis of the Male-female Earning Differential in Pakistan', *Pakistan Development Review*, 32 (4).

Aslam, Monazza and Kingdon, G.G. (2007), 'Gender and Household Education Expenditure in Pakistan', forthcoming in *Applied Economics*.

Behrman, J.R. and Deolalikar (1995), 'Are there Differential Returns to Schooling by Gender? The Case of Indonesian Labour Markets', *Oxford Bulletin of Economics and Statistics*, 57 (1), pp. 97-117.

Behrman, J.R. and Wolfe, B. (1984), 'The Socio-economic Impact of Schooling in a Developing Country', *Review of Economics and Statistics*, 66 (2), pp. 296-303.

Behrman, J.R., Rosenzweig, M.R. and Taubman, P. (1994), 'Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment', *Journal of Political Economy*¸102 (6), pp. 1131-1174.

Behrman, J.R. and Rosenzweig, M.R. (1999), ' 'Ability' Biases in Schooling Returns and Twins: A Test and New Estimates', *Economics of Education Review*, 18 (2), pp. 159-67.

Belzil, C. and Hansen, J. (2002), 'Unobserved Ability and the Return to Schooling', *Econometrica*, 70 (5), pp. 2075-91.

Blundell, R., Dearden, L. and Sianesi, B. (2005), 'Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS', *Journal of the Royal Statistical Society, Series A*, 168 (3), pp. 473-512.

Boissiere, M., Knight, J.B. and Sabot, R.H. (1985), 'Earnings, Schooling, Ability and Cognitive Skills', *American Economic Review*, 75(5), pp. 1016-1030.

Bound, J., Jaegar, D.A., Baker, R. (1995), 'Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak', *Journal of the American Statistical Association*, 90 (430), pp. 443-50.

Butcher, K. and Case, A. (1994), 'The Effect of Sibling Sex Composition on Women's Education and Earnings', *Quarterly Journal of Economics*, 109 (3), pp. 531-63.

Card, D. (1995a), 'Using Geographic Variation in College Proximity to Estimate the Return to Schooling', in L.N. Christofides, E. Kenneth Grant and R. Swidinsky (eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vandercamp*, Toronto: University of Toronto Press, pp. 201-222.

Card, D. (1995b), 'Earnings, Schooling and Ability Revisited', in: Solomon P. (ed.), *Research in Labor Economics*, Vol. 14, JAI Press, Greenwich, CT, pp. 23-48.

Card, D. (1999), 'The Causal Effect of Education on Earnings', in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3A, Amsterdam, North Holland, pp. 1801-63.

Card, D. (2001), 'Estimating the Returns to Schooling: Progress in some Persistent Econometric Problems', *Econometrica*, 69 (5), pp. 1127-60.

Choudhury, S. (1993), 'Reassessing the Male-female Wage Differential: A Fixed-effects Approach', *Southern Economic Journal*, 60 (2), pp. 327-40.

Duflo, E. (2001), 'Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment', *The American Economic Review,*91 (4), pp. 795-813.

Duraisamy, P. (2002), 'Changes in Returns to Education in India: 1983-1994: By Gender, Age-cohort and Location', *Economics of Education Review*, 21 (6), pp. 609-22.

Glewwe, P. (1996), 'The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment', *Journal of Development Economics*, 51, pp. 267-290.

Griliches, Z. (1977), 'Estimating the Returns to Schooling: Some Econometric Problems', *Econometrica*, 45 (1), pp. 1-22.

Griliches, Z. (1979), 'Sibling Models and Data in Economics: Beginnings of a Survey', *Journal of Political Economy*, 87 (5), pp. S 37-S64.

Guisinger, S.E., Henderson, J.W. and Scully, G.W. (1984), 'Earnings, Rate of Returns to Education and Earning Distribution in Pakistan', *Economics of Education Review*, 3 (4).

Hamdani, K. (1977), 'Education and the Income Differential: An Estimation for Rawalpindi City', *Pakistan Development Review*, 16 (2), pp. 144-64.

Haque, Nadeemul (1977), 'An Economic Analysis of Personal Earnings in Rawalpindi City', *Pakistan Development Review*, 16 (4).

Harmon, C., Oosterbeek, H. and Walker, I. (2000), ' The Returns to Education : A Review of Evidence, Issues and Deficiencies in the Literature', London School of Economics (Mimeo).

Harmon, C. and Walker, I. (1995), 'Estimates of the Economic Return to Schooling for the United Kingdom', *The American Economic Review*, 85 (5), pp. 1278-86.

Heckman, J.J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica*, 47 (1), pp. 153-161.

Heckman, J.J. and Hotz, V.J. (1989), 'An Investigation of the Labour Market Earnings of Panamanian Males: Evaluating the Sources of Inequality', *The Journal of Human Resources*, 21, pp. 507-63.

Hertz, T. (2003), 'Upward Bias in the Estimated Returns to Education: Evidence from South Africa', *The American Economic Review*, 93 (4), pp. 1354-68.

Khan, S.R. and Irfan, M. (1985), 'Rate of Returns to Education and Determinants of Earnings in Pakistan', *Pakistan Development Review*, 34 (3).

Kingdon, G.G. (1998), 'Does the Labour Market Explain Lower Female Schooling in India?', *Journal of Development Studies*, 35 (1), pp. 39-65.

Kingdon, G.G. and Unni, J. (2001), 'Education and Women's Labour Market Outcomes in India', *Education Economics*, 9 (2), pp. 173-195.

Knight, J. and Sabot, R. (1990), *Education, Productivity and Inequality: The East African Natural Experiment*, Oxford University Press, Oxford.

Krishnan, P. (1996), 'Family Background, Education and Employment in Urban Ethiopia', *Oxford Bulletin of Economics and Statistics*, 58 (1), pp. 167-183.

Lam, D. and Schoeni, R.F. (1993), 'Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil', *Journal of Political Economy*, 101 (4), pp. 710-740.

Maluccio, J. (1998), 'Endogeneity of Schooling in the Wage Function: Evidence from the Rural Phillippines', Food Consumption and Nutrition Division (IFPRI), Working Paper No. 54.

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

Moll, P.G. (1996), 'The collapse of primary schooling returns in South Africa, 1960-1990', *Oxford Bulletin of Economics and Statistics,* 58, pp.185-210.

Nasir, Z.M. (1998), 'Determinants of Personal Earnings in Pakistan: Findings from the Labour Force Survey 1993-1994', *Pakistan Development Review*, 37, pp. 251-74.

Nasir, Z.M. (1999), 'Do Private Schools Make Workers More Productive', *Pakistan Development Review*, 38, pp. 937-964.

Nasir, Z.M. (2002), 'Returns to Human Capital in Pakistan: A Gender Disaggregated Analysis', *Pakistan Development Review*, 41 (1).

Oaxaca, R. (1973), 'Male-female Differentials in Urban Labour Markets', *International Economic Review*, 3, pp. 603-709.

Psacharopoulos, G. (1985), 'Returns to Education: A Further International Update and Implications', *Journal of Human Resources*, 20 (4), pp. 583-604.

Psacharopoulos, G. (1994), 'Returns to Investment in Education: A Global Update', *World Development*, 22 (9), pp. 1325-1343.

Psacharopoulos, G. and Patrinos, H. (2004), 'Returns to Investment in Education: A Global Update', *Education Economics*, 12 (2), pp. 111-134.

Pitt, M. and Rosenzweig, M.R. (1990), 'Estimating the Intrahousehold Incidence of Illness: Child Health and Gender Inequality in the Allocation of Time', *International Economic Review*, 31 (4), pp. 969-989.

Riboud, M., Savchenko, Y. and Tan, H. (2006), 'The Knowledge Economy and Education and Training in South Asia: A Mapping Exercise of Available Survey Data', World Bank Working Paper, South Asia Region, 2006.

Rouse, C. E. (1999), 'Further Estimates of the Economic Return to Schooling from a New Sample of Twins', *Economics of Education Review*, 18, pp. 149-157.

Schultz, P. (1988), Education Investments and Returns, in: H. Chenery and T.N. Srinivasan (eds.), *Handbook of Development Economics*, Vol. 1 (North Holland, Amsterdam), pp. 543-60.

Schultz, T.P. (1993), 'Returns to Women's Education' in Elizabeth M. King and M. Anne Hill (eds.), *Women's Education in Developing Countries: Barriers, Benefits, and Policies*. Baltimore, MD: Johns Hopkins University Press, Chapter 2.

Schultz, T.P. (2004), 'Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education', *Journal of African Economies*, AERC Supplement 2, pp. 95-148.

Shabbir, T. and Khan, A.H. (1991), 'Mincerian Earnings Functions for Pakistan: A Regional Analysis', *Pakistan Economic and Social Review*, 29 (2).

Shabbir, T. (1991), 'Sheepskin Effects in the Returns to Education in a Developing Country', *The Pakistan Development Review*, 30 (1).

Shabbir, T. (1994), 'Mincerian Earnings Functions for Pakistan', *Pakistan Development Review*, 33 (1).

Siddiqui, R. and Siddiqui, R. (1998), 'A Decomposition of Male-female Earnings Differentials', *Pakistan Development Review*, 37 (4), pp. 885-898.

Strauss, J. and Thomas, D. (1995), 'Human Resources: Empirical Modelling of Household and Family Decisions', In: Behrman, J. and Srinivasan, T.N. (Eds.), Handbook of Development Economics, Vol. 3, Elsevier Science B.V., pp. 1883-2005.

Söderbom, M., Teal, F., Wambugu, A. and Kahyarara, G. (2005), 'The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing', Global Poverty Research Group Working Paper Series, No. 017.

Trostel, P., Walker, I. and Woolley, P. (2002), 'Estimates of the Economic Return to Schooling for 28 Countries', *Labour Economics*¸ 9, pp. 1-16.

Weiss, Y. (1999), 'The formation and dissolution of families: why marry? Who marries whom? And what happens upon marriage and divorce? In: Rosenzweig, M.R., Stark, O. (Eds.), Handbook of Population and Family Economics, North Holland, Amsterdam.

Willis, R. (1986), 'Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions', in O. Ashenfelter and R. Layard (eds.), *Handbook of Labour Economics*, Amsterdam, North Holland.

Zabalza, A. and Arrufat, J. (1983), 'Wage Differentials between Married Men and Women in Great Britain: the Depreciation Effect of Non-participation', Working Paper No. 382, London School of Economics.

Tables

Table 1: Distribution of the Labour Force in Pakistan by Gender (Persons aged 15-65)

Labour Force Status		Male		Female		Total	
		N	%	N	%	N	%
Unemployed (seeking work)	(a)	962	3.51	949	3.23	1911	3.37
Employed	(b = c + d)	23095	84.34	6800	23.14	29895	52.80
Self Employed	(c)	11594	42.34	4782	16.27	16376	28.85
Wage Employed	(d)	11501	42.00	2018	6.87	13519	23.81
Total labour Force	(e = a + b)	24057	87.85	7749	26.37	31806	56.02
Out of labour force	(f)	3328	12.15	21638	73.63	24966	43.98
All Persons	(g = e + f)	27385	100	29387	100	56772	100

Calculated from the PIHS (2002). Our definition of: Unemployed includes everyone reporting being jobless but seeking work, self employed includes all defined as: employers employing individuals, unpaid family workers, owner cultivators, share croppers, cultivators and livestock owners, wage employed includes all defined as paid employees while non labour force participants (out of labour force) are those reported as jobless and not seeking work.

Table 2: Definition of variables used in wage-work participation and earnings functions

Variable	Description
PAID_EMPLOY	Participation in salaried/waged work during the past month
AGE	Age in years
AGE2	Square of age
HEAD	Head of the household? Yes=1, No=0
MARRIED	Married? Yes=1, No=0
LNUNEARNED_Y	Natural Log of Unearned Income (income from boarders/lodgers, zakat, remittances, pensions, gifts and insurance etc.)
CHILD7	Number of children aged 7 or less in the household
ADULT60	Number of adults aged 60 or above in the household
NO_EDUCATION	Equals 1 if individual reports 0 years of education, 0 otherwise
LESS_PRIMARY	Individual has completed less than 5 years of education (katchi class, 1, 2,3 or 4 years), equals 1 if has completed less than primary and equals 0 otherwise
PRIMARY	Equals 1 if individual has completed 5 years, 0 otherwise
MIDDLE	Equals 1 if individual has completed 6, 7 or 8 years, 0 otherwise
MATRIC	Equals 1 if individual has completed 9 or 10 years, 0 otherwise
INTER	Equals 1 if individual has completed 11or 12 years, 0 otherwise
BACHELORS	Equals 1 if individual has completed 13 or 14 years, 0 otherwise
MA_MORE	Equals 1 if individual has completed 15 years of education or more, 0 otherwise
SINDH	Province is Sindh, Yes=1, No=0
NWFP	Province is NWFP, Yes=1, No=0
BALOCHISTAN	Province is Balochistan, Yes=1, No=0
AJK	Province is AJK, Yes=1, No=0
NORTH	Northern Areas, Yes=1, No=0
FATA	Federally Administered Tribal Areas, Yes=1, No=0
URBAN	Region is urban, Yes=1, No=0
LAMBDA	Selectivity term, inverse of Mill's Ratio
LN_MONTHLY_Y	Natural log of monthly earnings (Rupees) of individuals in paid employment in the labour market
EXP	Experience (years)
EXP2	Square of Experience
EDU_YRS	Number of years of education acquired

EDU_YRS2	Square of years of education
FEDYRS	Father's education (years)
MEDPRIM	Mother's education primary or less equals 1 if mother has positive but less than or equal to primary education, 0 otherwise
MEDPRIMORE	Mother's education more than primary equals 1 if mother has more than primary education, 0 otherwise
SPOUSE_EDU	Spouse's (husband's/wife's) education (years)
READ	Equals 1 if individual can 'read in any language with understanding', 0 otherwise
WRITE	Equals 1 if individual can 'write in any language with understanding', 0 otherwise
MATHS	Equals 1 if individual can 'solve simple (plus minus) sums', 0 otherwise
PRIVATE	Equals 1 if individual attended private school in the past, 0 otherwise

Table 3: Descriptive Statistics of Variables Used in the PAID_EMPLOY Participation Function

Variable	Mean Characteristics of Males			Mean Characteristics of Females		
	Participants	Non Participants	All	Participants	Non Participants	All
PAID_EMPLOY*	1.000	0.000	0.420	1.000	0.000	0.069
	(0.00)	(0.00)	(0.49)	(0.00)	(0.00)	(0.25)
AGE	33.378	34.928	34.277	32.452	33.174	33.124
	(12.14)	(15.18)	(14.01)	(12.01)	(13.81)	(13.70)
AGE2	1261.457	1450.359	1371.025	1197.274	1291.376	1284.914
	(906.65)	(1188.09)	(1082.88)	(869.53)	(1045.04)	(1034.20)
HEAD*	0.529	0.462	0.491	0.053	0.041	0.042
	(0.50)	(0.50)	(0.50)	(0.224)	(0.20)	(0.20)
MARRIED*	0.671	0.613	0.637	0.629	0.706	0.701
	(0.47)	(0.49)	(0.48)	(0.48)	(0.46)	(0.46)
LNUNEARNED_Y	2.402	2.832	2.652	2.900	3.438	3.401
	(4.09)	(4.44)	(4.31)	(4.39)	(4.73)	(4.70)
CHILD7	1.785	1.951	1.881	1.654	2.011	1.987
	(1.79)	(1.98)	(1.91)	(1.84)	(1.95)	(1.95)
ADULT60	0.435	0.559	0.507	0.465	0.548	0.542
	(0.65)	(0.71)	(0.69)	(0.66)	(0.71)	(0.71)
LESS_PRIMARY*	0.079	0.087	0.084	0.039	0.036	0.036
	(0.27)	(0.28)	(0.28)	(0.19)	(0.19)	(0.19)
PRIMARY*	0.097	0.104	0.101	0.052	0.070	0.068
	(0.30)	(0.31)	(0.30)	(0.22)	(0.25)	(0.25)
MIDDLE*	0.122	0.134	0.129	0.039	0.052	0.051
	(0.33)	(0.34)	(0.33)	(0.19)	(0.22)	(0.22)
MATRIC*	0.168	0.167	0.167	0.100	0.066	0.068
	(0.37)	(0.37)	(0.37)	(0.30)	(0.25)	(0.25)
INTER*	0.065	0.051	0.057	0.062	0.025	0.027
	(0.25)	(0.22)	(0.23)	(0.24)	(0.16)	(0.16)
BACHELORS*	0.058	0.029	0.041	0.070	0.015	0.019
	(0.23)	(0.17)	(0.20)	(0.25)	(0.12)	(0.14)
MA_MORE*	0.050	0.016	0.030	0.058	0.004	0.007
	(0.22)	(0.13)	(0.17)	(0.23)	(0.06)	(0.09)
SINDH*	0.302	0.241	0.267	0.353	0.233	0.241
	(0.46)	(0.43)	(0.44)	(0.48)	(0.42)	(0.43)
NWFP*	0.137	0.171	0.157	0.075	0.193	0.185
	(0.34)	(0.38)	(0.36)	(0.26)	(0.39)	(0.39)
BALOCHISTAN*	0.165	0.164	0.140	0.085	0.128	0.125
	(0.37)	(0.37)	(0.35)	(0.28)	(0.33)	(0.33)
AJK*	0.032	0.028	0.029	0.027	0.042	0.041
	(0.18)	(0.16)	(0.17)	(0.16)	(0.20)	(0.20)
NORTH*	0.011	0.034	0.024	0.005	0.029	0.027
	(0.10)	(0.18)	(0.15)	(0.07)	(0.17)	(0.16)
FATA*	0.012	0.020	0.017	0.000	0.019	0.017
	(0.11)	(0.14)	(0.13)	(0.00)	(0.14)	(0.13)
URBAN*	0.473	0.323	0.386	0.501	0.355	0.366
	(0.50)	(0.47)	(0.49)	(0.50)	(0.48)	(0.48)
N	11501	15884	27385	2018	27369	29387

Note: The variables with superscript (*) are binary 0/1 variables and their means represent the proportions of ones in the sample. Standard Deviations are reported in parentheses.

Table 4: Descriptive Statistics of Variables used in Earnings functions, Aged 15-65 in Waged Work

Average value of Variable:	Males	N	Females	N	t-test (M – F)
LN_MONTHLY_Y	7.783 (0.01)	11501	6.284 (0.03)	2018	66.68
EXP	20.492 (0.12)	11501	20.097 (12.79)	2018	1.30
EXP2	577.429 (5.98)	11501	567.414 (14.09)	2018	0.65
EDU_YRS	5.666 (0.05)	11501	4.326 (0.13)	2018	10.32
EDU_YRS2	60.295 (0.70)	11501	51.925 (1.82)	2018	4.53
NO_EDUCATION	0.361 (0.00)	11501	0.581 (0.01)	2018	-18.92
LESS_PRIMARY	0.079 (0.00)	11501	0.039 (0.00)	2018	6.42
PRIMARY	0.097 (0.00)	11501	0.052 (0.00)	2018	6.57
MIDDLE	0.122 (0.00)	11501	0.039 (0.00)	2018	11.12
MATRIC	0.169 (0.00)	11501	0.100 (0.01)	2018	7.78
INTER	0.065 (0.00)	11501	0.062 (0.01)	2018	0.58
BACHELORS	0.058 (0.00)	11501	0.070 (0.01)	2018	-2.16
MA_MORE	0.050 (0.00)	11501	0.058 (0.01)	2018	-1.58
FEDYRS	2.865 (4.19)	4155	4.380 (4.91)	493	-7.78
MEDPRIM	0.057 (0.00)	4155	0.075 (0.01)	493	-1.65
MEDPRIMORE	0.040 (0.00)	4155	0.061 (0.01)	493	-2.21
SPOUSE_EDU	2.059 (0.05)	5638	4.559 (0.17)	943	-16.68

Notes: 1) Standard errors are in parentheses beneath the mean values of the variables, 2) Descriptive statistics computed excluding any individuals in paid employment who are currently enrolled in school. NO_EDUCATION is the reference category for education splines.

Table 5: Average Monthly Earnings of Wage Employees, by Education Level and Gender

Education Level/Gender	MALE (a)	FEMALE (b)	GAP (M-F) (c = a – b)	t-test (M-F) (d)	F/M (e = b/a)
NO_EDUCATION	2271.5 (19.04)	581.3 (23.14)	1690.2	-44.62	0.26
LESS_PRIMARY	2258.7 (46.52)	732.1 (118.16)	1526.6	-9.40	0.32
PRIMARY	2539.7 (52.28)	709.0 (91.71)	1830.7	-10.54	0.28
MIDDLE	2599.0 (43.23)	1054.0 (154.01)	1545.0	-8.26	0.41
MATRIC	3242.5 (45.00)	2127.5 (127.57)	1115.0	-7.67	0.66
INTER	4109.6 (100.11)	2512.9 (163.71)	1597.7	-6.27	0.61
BACHELORS	5845.0 (165.40)	3818.5 (245.64)	2026.5	-5.39	0.65
MA_MORE	8521.9 (282.55)	6518.3 (301.11)	2003.6	-3.13	0.76
ALL	3136.3 (26.41)	1456.8 (48.50)	1664.7	-25.25	0.46

Notes: Standard errors are in parentheses beneath the mean values of the variables

Table 6: Binary Probit Estimates of Waged Work Participation (15-65), by Gender

Variable	Males				Females			
	Coefficient	t-value		Marginal Effect	Coefficient	t-value		Marginal Effect
CONSTANT	-1.485	-20.88	***	-	-2.281	-22.14	***	-
AGE	0.073	17.10	***	0.029	0.066	10.18	***	0.006
AGE2	-0.001	-21.56	***	0.000	-0.001	-10.35	***	0.000
LESS_PRIMARY	-0.001	-0.03		0.000	0.057	0.91		0.006
PRIMARY	-0.035	-1.26		-0.014	-0.159	-3.11	***	-0.014
MIDDLE	-0.046	-1.77	*	-0.018	-0.150	-2.53	**	-0.013
MATRIC	-0.031	-1.30		-0.012	0.201	4.58	***	0.023
INTER	0.054	1.53		0.021	0.421	7.21	***	0.056
BACHELORS	0.256	6.17	***	0.101	0.733	11.69	***	0.123
MA_MORE	0.494	10.02	***	0.195	1.504	16.81	***	0.380
HEAD	0.204	8.59	***	0.079	0.169	2.82	***	0.019
MARRIED	0.113	4.58	***	0.044	-0.267	-8.55	***	-0.029
LNUNEARNED_Y	-0.004	-2.10	**	-0.002	-0.011	-3.71	***	-0.001
CHILD7	-0.041	-9.00	***	-0.016	-0.014	-2.02	*	-0.001
ADULT60	-0.069	-5.59	***	-0.027	-0.040	-2.21	**	-0.004
SINDH	0.234	11.45	***	0.092	0.096	3.41	***	0.010
NWFP	0.001	0.05		0.000	-0.535	-12.75	***	-0.040
BALOCHISTAN	0.312	12.38	***	0.123	-0.276	-6.58	***	-0.023
AJK	0.249	5.19	***	0.099	-0.342	-4.76	***	-0.026
NORTH	-0.630	-10.60	***	-0.217	-0.771	-6.22	***	-0.041
FATA	-0.024	-0.37		-0.009	-	-	-	-
URBAN	0.363	21.44	***	0.142	0.080	3.02	***	0.008
Log L	-17398.48				-6875.81			
Pseudo-R^2	0.071				0.080			
N (Un-Censored)	27385				29387			

Note: *, ** and *** represent significance at the 10%, 5% and 1% levels respectively. The dependent variable is PAID_EMPLOY which equals 1 if individual is in paid employment and 0 otherwise. (-) indicates no observations. NO_EDUCATION is reference category for education splines, PUNJAB for provinces.

Table 7: OLS Mincerian Earnings Functions, (Males and Females), with years of Education and Levels of Education

Variable	Males (15-65)				Females (15-65)			
	Years (a) Coefficient		Levels (b) Coefficient		Years (c) Coefficient		Levels (d) Coefficient	
CONSTANT	6.223	***	6.357	***	4.188	***	4.307	***
	(0.03)		(0.03)		(0.12)		(0.13)	
EDU_YRS	0.072	***			0.166	***		***
	(0.00)		-		(0.01)		-	
EXP	0.076	***	0.075	***	0.073	***	0.068	***
	(0.00)		(0.00)		(0.01)		(0.01)	
EXP2	-0.001	***	-0.001	***	-0.001	***	-0.001	***
	(0.00)		(0.00)		(0.00)		(0.00)	
LESS_PRIMARY	-		0.011		-		0.334	**
			(0.03)				(0.14)	
PRIMARY	-		0.136	***	-		0.342	**
			(0.02)				(0.13)	
MIDDLE	-		0.271	***	-		0.958	***
			(0.02)				(0.14)	
MATRIC	-		0.534	***	-		1.505	***
			(0.02)				(0.12)	
INTER	-		0.762	***	-		1.843	***
			(0.03)				(0.12)	
BACHELORS	-		1.070	***	-		2.294	***
			(0.03)				(0.11)	
MA_MORE	-		1.371	***	-		2.909	***
			(0.03)				(0.11)	
SINDH	0.205	***	0.188	***	0.281	***	0.269	***
	(0.02)		(0.02)		(0.09)		(0.09)	
NWFP	-0.060	**	-0.078	***	0.494	***	0.479	***
	(0.03)		(0.03)		(0.11)		(0.11)	
BALOCHISTAN	0.423	***	0.386	***	0.664	***	0.643	***
	(0.03)		(0.03)		(0.13)		(0.13)	
AJK	0.165	***	0.174	***	0.711	***	0.681	***
	(0.04)		(0.04)		(0.14)		(0.15)	
NORTH	0.216	***	0.204	***	1.370	***	1.392	***
	(0.05)		(0.05)		(0.32)		(0.32)	
FATA	0.124	**	0.096					
	(0.06)		(0.06)		-		-	
URBAN	0.200	***	0.204	***	0.487	***	0.503	***
	(0.02)		(0.02)		(0.09)		(0.09)	
R^2	0.388		0.408		0.472		0.478	
N	11501		11501		2018		2018	
Mean (Dep. Var)	7.783		7.783		6.284		6.284	

Note: *, ** and *** represent significance at the 10%, 5% and 1% levels respectively. The dependent variable is LN_MONTHLY_Y. Standard errors are in parentheses. (-) indicates no observations. NO_EDUCATION is reference category for education splines, PUNJAB for provinces.

Table 7a: Rates of Returns to additional years of education (Males and Females) at various levels of education

LEVEL OF EDUCATION	Rates of return (%)	
	Males	Females
PRIMARY	2.7	6.8
MIDDLE	4.5	20.5
MATRIC	13.2	27.4
INTER	11.4	16.9
BACHELORS	15.4	22.6
MA_MORE	15.1	30.7

Table 8: Heckman corrected Mincerian Earnings Functions, (Males and Females), Years of Education and Levels of Education

Variable	Males (15-65)		Females (15-65)	
	Years (a) Coefficient	Levels (b) Coefficient	Years (c) Coefficient	Levels (d) Coefficient
CONSTANT	7.295 ***	6.908 ***	5.227 ***	4.310 ***
	(0.08)	(0.08)	(0.29)	(0.46)
EDU_YRS	0.064 ***	- -	0.142 ***	- -
	(0.00)		(0.01)	
EXP	0.049 ***	0.060 ***	0.062 ***	0.068 ***
	(0.00)	(0.00)	(0.01)	(0.01)
EXP2	-0.001 ***	-0.001 ***	-0.001 ***	-0.001 ***
	(0.00)	(0.00)	(0.00)	(0.00)
LESS_PRIMARY	- -	0.020	- -	0.334 ***
		(0.02)		(0.13)
PRIMARY	- -	0.149 ***	-	0.343 ***
		(0.02)		(0.12)
MIDDLE	- -	0.283 ***	- -	0.958 ***
		(0.02)		(0.14)
MATRIC	- -	0.533 ***	- -	1.504 ***
		(0.02)		(0.10)
INTER	- -	0.732 ***	- -	1.842 ***
		(0.03)		(0.15)
BACHELORS	- -	0.991 ***	- -	2.293 ***
		(0.03)		(0.18)
MA_MORE	- -	1.246 ***	- -	2.908 ***
		(0.03)		(0.28)
SINDH	0.085 ***	0.127 ***	0.239 ***	0.269 ***
	(0.02)	(0.02)	(0.06)	(0.06)
NWFP	-0.053 **	-0.070 ***	0.695 ***	0.479 ***
	(0.02)	(0.02)	(0.11)	(0.14)
BALOCHISTAN	0.257 ***	0.303 ***	0.778 ***	0.643 ***
	(0.02)	(0.02)	(0.10)	(0.11)
AJK	0.062	0.117 ***	0.848 ***	0.681 ***
	(0.04)	(0.04)	(0.17)	(0.17)
NORTH	0.567 ***	0.395 ***	1.786 ***	1.393 ***
	(0.06)	(0.06)	(0.37)	(0.40)
FATA	0.160 ***	0.122 **	- -	- -
	(0.06)	(0.05)		
URBAN	0.027 ***	0.111 ***	0.443	0.503 ***
	(0.02)	(0.02)	(0.06) ***	(0.06)
LAMBDA	-0.756 ***	-0.407 ***	-0.472 ***	-0.001
	(0.05)	(0.05)	(0.12)	(0.22)
N_UNCENSORED	11501	11501	2018	2018
WALD_CHI2	3836.09	5144.57	1297.56	1901.26
PVALUE (WALD)	0.000	0.000	0.000	0.000

Note: *, ** and *** represent significance at the 10%, 5% and 1% levels respectively. The dependent variable is LN_MONTHLY_Y. Standard errors are in parentheses. (-) indicates no observations. NO_EDUCATION is reference category for education splines, PUNJAB for provinces.

Table 9: Instrumental Variable Earning Functions Estimates (Males and Females), using years of Education

Variable	Males (15-65)				Females (15-65)			
	IV (PARENTAL EDUCATION) (a) Coefficient	First Stage (b) Coefficient	IV (SPOUSE_EDU) (c) Coefficient	First Stage (d) Coefficient	IV (PARENTAL EDUCATION) (e) Coefficient	First Stage (f) Coefficient	IV (SPOUSE_EDU) (g) Coefficient	First Stage (h) Coefficient
CONSTANT	5.894 ***	7.516 ***	6.216 ***	9.755 ***		4.623	4.432 ***	3.647 ***
	(0.08)	(0.24)	(0.08)	(0.37)		(0.76)	(0.28)	(0.75)
EDU_YRS	0.099 ***	-	0.106 ***	-	0.169 ***	-	0.176 ***	-
	(0.01)		(0.00)		(0.03)		(0.01)	
EXP	0.102 ***	-0.473 ***	0.064 ***	-0.317 ***	0.154 ***	-0.422 ***	0.037 *	-0.232 ***
	(0.01)	(0.03)	(0.00)	(0.02)	(0.03)	(0.10)	(0.02)	(0.06)
EXP2	-0.02 ***	0.010 ***	-0.001 ***	0.003 ***	-0.003 ***	0.008 ***	0.000	0.003 **
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SINDH	0.225 ***	0.381 **	0.188 ***	-0.011	0.230 *	-0.062	0.345 ***	-0.658 ***
	(0.03)	(0.16)	(0.03)	(0.15)	(0.13)	(0.59)	(0.11)	(0.25)
NWFP	-0.103 **	0.941 ***	0.006	0.375 *	-0.077	3.823 ***	0.775 ***	0.919 *
	(0.04)	(0.20)	(0.03)	(0.19)	(0.25)	(0.96)	(0.14)	(0.49)
BALOCHISTAN	0.486 ***	-0.092	0.403 ***	0.174	0.728 ***	1.103	0.656 ***	-0.854 *
	(0.04)	(0.19)	(0.03)	(0.18)	(0.26)	(1.14)	(0.14)	(0.44)
AJK	0.029	1.573 ***	0.215 ***	1.273 ***	0.165	5.366 ***	1.231 ***	1.533 *
	(0.07)	(0.36)	(0.05)	(0.35)	(0.25)	(0.90)	(0.20)	(0.92)
NORTH	0.162 **	0.668	0.208 ***	1.399 **	-	-	1.623 ***	-0.302
	(0.08)	(0.63)	(0.06)	(0.56)			(0.26)	(1.43)
FATA	0.169 *	-1.596 ***	0.142	-0.160	-	-	-	-
	(0.09)	(0.59)	(0.09)	(0.53)				
URBAN	0.096 ***	0.657 ***	0.164 ***	1.396 ***	0.324 *	3.745 ***	0.572 ***	1.919 ***
	(0.03)	(0.13)	(0.02)	(0.13)	(0.19)	(0.59)	(0.11)	(0.25)
FEDYRS	-	0.491 ***	-	-	-	0.360 ***	-	-
		(0.02)				(0.06)		
MEDPRIM	-	0.526			-	2.067 ***		
		(0.29)				(0.69)		
MEDPRIMORE	-	1.397 ***			-	2.486 **		
		(0.35)				(0.99)		
SPOUSE_EDU	-	-	-	0.584 ***	-	-	-	0.587 ***
				(0.02)				(0.02)
R^2	0.282	0.323	0.303	0.383	0.458	0.510	0.478	0.578
N	4155	4155	5590	5590	493	493	903	903
F-Test of Excl. Instruments	-	258.62	-	1305.46	-	23.48	-	610.60
P-value (F test)	-	0.000	-	0.000	-	0.000	-	0.000
Over-identification Test	-	4.69	-	-		2.37	-	
P-value (Over-id Test)	-	0.100	-	-		0.306		

Note: *, ** and *** represent significance at the 10%, 5% and 1% levels respectively. Standard errors are in parentheses and are robust and corrected for clustering at PSU level. The dependent variable is LN_MONTHLY_Y. (-) indicates no observations or not used. PUNJAB is the base category for provinces and MEDNONE (=1 if mother has no education, 0 otherwise) is the base for mothers' educational dummies – MEDPRIM =1 mother has primary or less education (but more than 0) and 0 otherwise, MEDPRIMORE =1 if mother has more than primary education, 0 otherwise.

Table 10: Fixed Effects Estimates of Earnings Functions, Males and Females (15-65), Years and Levels of Education

Variable	All (a)				Siblings (Brother/Sister) (b)			
	Years		Levels		Years		Levels	
	OLS	FE	OLS	FE	OLS	FE	OLS	FE
CONSTANT	4.432	4.879	4.587	5.132	4.057	4.143	4.354	
	(0.08)***	(0.09)***	(0.08)***	(0.08)***	(0.20)***	(0.26)***	(0.21)***	
EDU_YRS	0.165	0.140	-	-	0.155	0.153	-	-
	(0.01)***	(0.01)***			(0.01)***	(0.02)***		
EDU_YRS_MALE	-0.093	-0.082	-	-	-0.049	-0.043	-	-
	(0.01)***	(0.01)***			(0.02)***	(0.02)***		
MALE	1.659	1.630	1.680	1.628	1.138	1.105	1.173	1.245
	(0.05)***	(0.05)***	(0.06)***	(0.06)***	(0.15)***	(0.15)***	(0.17)***	(0.18)***
EXP	0.073	0.072	0.067	0.069	0.179	0.213	0.167	0.222
	(0.01)***	(0.01)***	(0.06)***	(0.01)***	(0.03)***	(0.03)***	(0.03)***	(0.03)***
EXP2	-0.001	-0.001	-0.001	-0.001	-0.004	-0.005	-0.004	-0.01
	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.00)***	(0.001)***	(0.001)***
LESS_PRIMARY	-	-	0.196	0.218	-	-	0.443	0.802
			(0.15)	(0.18)			(0.30)	(0.35)**
PRIMARY	-	-	0.126	0.050	-	-	0.110	0.150
			(0.12)	(0.15)			(0.292)	(0.37)
MIDDLE	-	-	0.822	0.707	-	-	0.522	1.010
			(0.14)	(0.17)***			(0.24)**	(0.32)***
MATRIC	-	-	1.359	1.204	-	-	0.957	1.376
			(0.10)***	(0.13)***			(0.22)***	(0.29)***
INTER	-	-	1.742	1.637	-	-	1.498	1.777
			(0.12)***	(0.16)***			(0.22)***	(0.30)***
BACHELORS	-	-	2.306	2.060	-	-	2.078	2.163
			(0.11)***	(0.15)***			(0.22)***	(0.30)***
MA_MORE	-	-	3.012	2.767	-	-	2.809	2.752
			(0.12)***	(0.19)***			(0.23)***	(0.37)***
LESS_PRIMARY_MALE	-	-	-0.202	-0.170	-	-	-0.528	-0.927
			(0.17)	(0.21)			(0.39)	(0.44)**
PRIMARY_MALE	-	-	0.051	0.139	-	-	0.097	-0.071
			(0.15)	(0.18)			(0.37)	(0.43)
MIDDLE_MALE	-	-	-0.669	-0.532	-	-	-0.060	-0.401
			(0.16)***	(0.18)**			(0.31)	(0.35)
MATRIC_MALE	-	-	-0.888	-0.776	-	-	-0.311	-0.571
			(0.12)***	(0.14)***			(0.28)	(0.31)*
INTER_MALE	-	-	-1.051	-0.956	-	-	-0.728	-0.911
			(0.16)***	(0.18)***			(0.32)**	(0.35)***
BACHELORS_MALE	-	-	-1.260	-1.080	-	-	-0.548	-0.223
			(0.15)***	(0.16)***			(0.32)*	(0.37)
MA_MORE_MALE	-	-	-1.632	-1.546	-	-	-0.976	-0.978
			(0.16)***	(0.16)***			(0.33)***	(0.34)***
SINDH	0.205	-	0.196	-	0.181	-	0.147	-
	(0.04)***		(0.04)***		(0.10)*		(0.10)	
NWFP	0.171	-	0.150	-	-0.145	-	-0.166	-
	(0.071)**		(0.07)**		(0.17)		(0.17)	
BALOCHISTAN	0.400	-	0.369	-	0.481	-	0.397	-
	(0.070)***		(0.07)***		(0.20)**		(0.20)*	
AJK	0.422	-	0.371	-	0.007	-	-0.093	-
	(0.124)***		(0.12)**		(0.21)		(0.21)	
NORTH	0.472	-	0.535	-	-	-	-	-
	(0..324)		(0.32)*					
FATA	-	-	-	-	-	-	-	-
URBAN	0.293	-	0.320	-	0.226	-	0.279	-
	(0.04)***		(0.04)***		(0.10)**		(0.10)***	
N	2423	2423	2423	2423	437	437	437	437
No. Groups	-	948	-	948	-	160	-	160
R^2	0.551	0.522	0.567	0.541	0.522	0.501	0.560	0.528

Note: *, ** and *** represent significance at the 10%, 5% and 1% levels respectively. The dependent variable is LN_MONTHLY_Y. (-) indicates no observations or not used. NO_EDUCATION is reference category for education splines, PUNJAB for provinces.

Table 11: OLS and Control Function Estimates of the Returns to Schooling, Males and Females (15-65)

VARIABLE	MALES (15-65)			FEMALES (15-65)		
	OLS	CONTROL FUNCTION		OLS	CONTROL FUNCTION	
	LN_MONTHLY_Y	STEP 1: EDU_YRS	Step 2: LN_MONTHLY_Y	LN_MONTHLY_Y	Step 1: EDU_YRS	Step 2: LN_MONTHLY_Y
	(a)	(b)	(c)	(a)	(b)	(c)
	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
CONSTANT	6.271	7.516	6.114	4.284	4.576	4.275
	(0.05) ***	(0.28) ***	(0.08) ***	(0.20) ***	(0.76) ***	(0.26) ***
EDU_YRS	-0.001	-	0.019	0.044	-	0.045
	(0.01)		(0.01) *	(0.04)		(0.05)
EDU_YRS2	0.006	-	0.005	0.007	-	0.007
	(0.00) ***		(0.00) ***	(0.00) ***		(0.00) ***
EXP	0.086	-0.473	0.096	0.128	-0.419	0.129
	(0.01) ***	(0.03) ***	(0.01) ***	(0.02) ***	(0.10) ***	(0.03) ***
EXP2	-0.002	0.010	-0.002	-0.002	0.009	-0.002
	(0.00) ***	(0.00) ***	(0.00) ***	(0.00)	(0.00) ***	(0.00) ***
SINDH	0.202	0.381	0.198	0.202	-0.097	0.201
	(0.03) ***	(0.24)	(0.03) ***	(0.13)	(0.59)	(0.13)
NWFP	-0.110	0.941	-0.116	0.040	3.786	0.035
	(0.04) **	(0.27) ***	(0.04) ***	(0.25)	(0.96) ***	(0.25)
BALOCHISTAN	0.424	-0.092	0.439	0.753	1.076	0.751
	(0.04) ***	(0.31)	(0.04) ***	(0.28) ***	(1.14)	(0.29) ***
AJK	0.079	1.573	0.055	0.325	5.313	0.317
	(0.07)	(0.42) ***	(0.07)	(0.21)	(0.91) ***	(0.25)
NORTH	0.148	0.668	0.153	2.554	-6.980	2.560
	(0.08) *	(0.74)	(0.08) **	(0.22) ***	(0.95) ***	(0.26) ***
FATA	0.076	-1.596	0.110	-	-	-
	(0.08) ***	(0.67) **	(0.08)			
URBAN	0.134	0.657	0.108	0.455	3.769	0.449
	(0.03)	(0.20) ***	(0.03) ***	(0.15) ***	(0.59) ***	(0.20) **
FEDYRS	-	0.491		-	0.364	
		(0.02) ***	-		(0.06) ***	-
MEDPRIM	-	0.526	-	-	2.066	
		(0.33)			(0.69) ***	-
MEDPRIMORE	-	1.397	-	-	2.454	
		(0.32)			(0.99) **	-
RESIDUAL	-	-	-0.020	-	-	-0.002
			(0.01) ***			(0.03)
N	4155	4155	4155	493	493	493
R^2	0.329	0.323	0.331	0.484	0.510	0.484

Note: Significance at the 1 per cent, 5 per cent and 10 per cent level is indicated by *** , ** and * respectively. All other variables are as defined before. The instrument variables are FEDYRS,MEDPRIM and MEDPRIMORE.

Table 12: Marginal returns to education (OLS and CF)

Marginal Return to education at --- years of education:	Males		Females	
	OLS	CF	OLS	CF
6	7.1	7.9	12.8	12.9
10	11.9	11.9	18.4	18.5
12	14.3	13.9	21.2	21.3
14	16.7	15.9	24.0	24.1

Table 13: Oaxaca Decomposition (OLS and FE)

Based on the OLS Estimates	**Males**			**Females**		
	Standardising by Male Means			**Standardising by Female Means**		
	Characteristics (E)	**Coefficients (D)**	**Combined (T)**	**Characteristics (E)**	**Coefficients (D)**	**Combined (T)**
INTERCEPT	0.000	2.062	2.062	0.000	2.062	2.062
EDU_YRS	0.123	-0.479	-0.356	0.010	-0.366	-0.356
EDU_YRS2	0.044	-0.030	0.014	0.040	-0.026	0.014
EXP	0.027	0.105	0.132	0.029	0.103	0.132
EXP2	-0.009	-0.096	-0.105	-0.011	-0.094	-0.105
SINDH	-0.014	-0.024	-0.038	-0.009	-0.028	-0.038
NWFP	0.030	-0.077	-0.047	-0.005	-0.042	-0.047
BALOCHISTAN	0.052	-0.044	0.009	0.031	-0.023	0.009
AJK	0.004	-0.018	-0.014	0.001	-0.015	-0.014
NORTH	0.009	-0.014	-0.005	0.001	-0.006	-0.005
FATA	0.000	0.001	0.001	0.001	0.000	0.001
URBAN	-0.014	-0.139	-0.153	-0.006	-0.147	-0.153
TOTAL	**0.253**	**1.247**	**1.499**	**0.083**	**1.417**	**1.499**
Explained by Coefficients	83 %			95 %		
Average Estimate of discrimination 89 %						
Based on the FE Estimates	**Males**			**Females**		
	Standardising by Male Means			**Standardising by Female Means**		
	Characteristics (E)	**Coefficients (D)**	**Combined (T)**	**Characteristics (E)**	**Coefficients (D)**	**Combined (T)**
INTERCEPT	1.403	0.000	1.403	1.403	0.000	1.403
EDU_YRS	0.277	-0.097	0.179	0.198	-0.018	0.179
EDU_YRS2	-0.311	0.095	-0.216	-0.254	0.038	-0.216
EXP	0.384	0.040	0.424	0.362	0.062	0.424
EXP2	-0.108	-0.015	-0.123	-0.104	-0.019	-0.123
TOTAL	**1.645**	**0.023**	**1.668**	**1.606**	**0.062**	**1.668**
Explained by Coefficients	98 %			96 %		
Average Estimate of discrimination 97 %						

Appendix

Appendix Table A1: Earnings Functions with Occupation/Industry, Males (15-65)

VARIABLE	**OLS W/O** (a) Coefficient	**OLS WITH** (b) Coefficient	**SSB W/O** (c) Coefficient	**SSB WITH** (d) Coefficient
CONSTANT	6.357 ***	6.608 ***	6.908 ***	7.214 ***
	(0.03)	(0.04)	(0.08)	(0.08)
EXP	0.075 ***	0.072	0.060 ***	0.056
	(0.00)	(0.00) ***	(0.00)	(0.00)
EXP2	-0.001 ***	-0.001	-0.001 ***	-0.001 ***
	(0.00)	(0.00) ***	(0.00)	(0.00)
LESS_PRIMARY	0.011	-0.028	0.020	-0.019
	(0.03)	(0.03)	(0.02)	(0.02) ***
PRIMARY	0.136 ***	0.077	0.149 ***	0.091
	(0.02)	(0.02) ***	(0.02)	(0.02) ***
MIDDLE	0.271 ***	0.205	0.283 ***	0.217
	(0.02)	(0.02) ***	(0.02)	(0.02) ***
MATRIC	0.534 ***	0.455	0.533 ***	0.451
	(0.02)	(0.02) ***	(0.02)	(0.02) ***
INTER	0.762 ***	0.671	0.732 ***	0.635
	(0.03)	(0.03) ***	(0.03)	(0.03) ***
BACHELORS	1.070 ***	0.962	0.991 ***	0.872
	(0.03)	(0.03) ***	(0.03)	(0.03) ***
MA_MORE	1.371 ***	1.255	1.246 ***	1.113
	(0.03)	(0.04) ***	(0.03)	(0.04) ***
OCLERICAL	-	-0.090	-	-0.099
		(0.02) ***		(0.02) ***
OAGRICRAFT	-	-0.042	-	-0.050
		(0.03)		(0.02) **
OELEMENTARY	-	-0.176	-	-0.186
		(0.03) ***		(0.02) ***
OOTHER	-	0.054	-	0.041
		(0.03) **		(0.02) *
IAGRI	-	-0.281	-	-0.283
		(0.03) ***		(0.02) ***
ICONSTRUCT	-	-0.068	-	-0.069
		(0.02) ***		(0.02) ***
ISOCIAL	-	-0.060	-	-0.061
		(0.02) ***		(0.01) ***
LAMBDA	-	-	-0.407 ***	-0.439
			(0.05)	(0.05) ***
PROVINCIAL/REGIONAL FE	YES	YES	YES	YES
N	11501	11501	11501	11501
R^2	0.408	0.433	-	-
WALD_CHI2	-	-	5144.57	5753.03
P_VALUE (WALD)	-	-	0.000	0.000

Note: Figures in columns (a) and (c) replicated from Tables 4.7 and 4.8. Standard errors in parentheses. All t-values are robust and corrected for clustering at the PSU-level. * denotes significance at 10 %, ** at 5 % and *** at 1 % or better.

Appendix Table A2: Earnings Functions with Occupation/Industry, Females (15-65)

VARIABLE	**OLS W/O** **(a)** **Coefficient**	**OLS WITH** **(b)** **Coefficient**	**SSB W/O** **(c)** **Coefficient**	**SSB WITH** **(d)** **Coefficient**
CONSTANT	4.307 ***	5.224 ***	4.310 ***	4.861 ***
	(0.13)	(0.19)	(0.46)	(0.45)
EXP	0.068 ***	0.060 ***	0.068 ***	0.063 ***
	(0.01)	(0.01)	(0.01)	(0.01)
EXP2	-0.001 ***	-0.001 ***	-0.001 ***	-0.001 ***
	(0.00)	(0.00)	(0.00)	(0.00)
LESS_PRIMARY	0.334 **	0.235 *	0.334 ***	0.244 *
	(0.14)	(0.13)	(0.13)	(0.13)
PRIMARY	0.342 **	0.277 **	0.343 ***	0.252 **
	(0.13)	(0.13)	(0.12)	(0.11)
MIDDLE	0.958 ***	0.784 ***	0.958 ***	0.764 ***
	(0.14)	(0.13)	(0.14)	(0.13)
MATRIC	1.505 ***	0.804 ***	1.504 ***	0.843 ***
	(0.12)	(0.13)	(0.10)	(0.12)
INTER	1.843 ***	1.021 ***	1.842 ***	1.093 ***
	(0.12)	(0.14)	(0.15)	(0.16)
BACHELORS	2.294 ***	1.384 ***	2.293 ***	1.502 ***
	(0.11)	(0.13)	(0.18)	(0.19)
MA_MORE	2.909 ***	1.975 ***	2.908 ***	2.183 ***
	(0.11)	(0.14)	(0.28)	(0.28)
OCLERICAL	-	-0.323	-	-0.316 ***
		(0.12) ***		(0.11)
OAGRICRAFT	-	-1.018	-	-1.015 ***
		(0.13) ***		(0.11)
OELEMENTARY	-	-0.498	-	-0.493 ***
		(0.16) ***		(0.12)
OOTHER	-	-0.124	-	-0.116
		(0.34)		(0.31)
IAGRI	-	-0.418	-	-0.421 ***
		(0.17) **		(0.09)
ICONSTRUCT	-	-0.787	-	-0.784
		(0.90)		(0.48)
ISOCIAL	-	0.334	-	0.336 ***
		(0.10) ***		(0.07)
LAMBDA	-	-	-0.001	0.176
			(0.22)	(0.21)
PROVINCIAL/REGIONAL FE	YES	YES	YES	YES
N	2018	2018	2018	2018
R^2	0.478		-	-
WALD_CHI2	-	-	1901.26	2318.85
P_VALUE (WALD)	-	-	0.000	0.000

Note: Figures in columns (a) and (c) replicated from Tables 4.7 and 4.8. Standard errors in parentheses. All t-values are robust and corrected for clustering at the PSU-level. * denotes significance at 10 %, ** at 5 % and *** at 1 % or better.