Pakistan Journal of Applied Economics, (1996), Vol.XII No.1, (85–94)

EVIDENCE ON GENDER WAGE DISCRIMINATION FROM THE 1984-85 *HIES*: A Note

Birjees ASHRAF and Javed ASHRAF*

In a recent article, Ashraf and Ashraf (1993b) calculated the male-female earnings differential for Pakistan using *HIES* data for 1979 and 1985-86. This note supplements that study by estimating the gender earnings gap with data from the 1984-85 *HIES*. In addition, rates of returns to different levels of education for men and women are presented.

I. Introduction

In two recent articles, Ashraf and Ashraf (1993a and 1993b) estimated the gender earnings differential using Pakistani data. In the first of these articles, the authors limited their attention to Rawalpindi, and used data that is now almost two decades old. In their second study, Ashraf and Ashraf (1993b) calculated the gender earnings differential for the entire country, using data from the *Household and Income Expenditure Surveys (HIES)* for 1979 and 1985-86. This note uses data from the 1984-85 *HIES*, and is intended as a supplement to the second of the AA (hereafter referred to as AAII) articles. In conjunction with the AAII, this article provides a complete profile of the gender earnings gap in Pakistan for the years for which data are available. The two AA articles are the only existing published articles of gender wage discrimination in Pakistan. As in the AA articles, this study uses the Oaxaca (1973) procedure as well as the more recently developed Cotton (1988) and Neumark (1988) models to estimate male-female earnings differentials in Pakistan. Both the male and female samples have been corrected for sample-selectivity bias.

*Birjees Ashraf is Assistant Professor of Economics at the Textile University of Pakistan, and is also a Ph.D. Candidate in Economics at the University of Karachi. Javed Ashraf is Professor and Vice-President at the Textile University of Pakistan.

II. The Data and Model

The number of respondents in the *1984-85 Household Income and Expenditure Survey* was well in excess of 100,000 individuals. The useable sample of working individuals (with observations on all variables used in this study) was much smaller, with 21,459 men and 1,489 women. However, the entire sample was used for the calculation of the selectivity-bias variables, since non-working individuals were included in those computations.

The variables used in the model were age, the square of age, a variable identifying residence in the urban area, the four provinces, seven levels of educational attainment and nine industrial groups, with which the respondents were associated.

The regression equation (unadjusted for selectivity bias) used to estimate the gender earnings gap (computed separately for male and female) was:

$$\text{InSalary} = a_0 + \sum_{j=1}^{3} b_j P_{mi} + \sum_{k=1}^{6} c_k ED_{ki} + \sum_{m=1}^{8} d_m IND_m + \sum_{n=1}^{3} e_n OC_{ni} \qquad (1)$$

where the dependent variable was the log of salaries. P_m(m=1, 2) represented the three provinces of Punjab, Baluchistan and Sind (NWFP was the omitted reference group); ED_k(k=1,..., 6) represented six levels of educational attainment; IND_m(m=1,...., 8) were eight industries with which the workers were associated; and OC_n were three "other characteristics" which were hypothesized to impact on salaries. These included a dummy variable for respondents from urban areas since part of the earnings difference could simply be a cost-of-living markup to compensate for the higher expenses of residing in urban areas. The two remaining variables were age and the square of age (both are standard in earnings equations). An important point to bear in mind is that since the salary information provided in the *HIES* data were for a constant period of time, the degree of discrimination against women may have been overstated. This is because it is well known that on average, women put in fewer hours per week than men, thereby exacerbating observed earnings differentials, although an analysis based on *hourly* earnings would likely suggest a smaller gap.

Specifically, the variables used were: Educ1, the omitted base variable among the educational attainment variables represented respondents who reported themselves to be uneducated or had less than a primary level of education. Those who had completed primary school were denoted by a dummy variable, Educ2. Dummy variables for other levels of education were Educ3 (those with "middle" level of schooling), Educ4 for matriculates, Educ5 for those with intermediate diplomas, Educ6 for college graduates, and Educ7 for workers with a masters or higher level degree.

Although the 1984 *HIES* identified several occupational groups, these categorizations were very broad. The mix of workers within each group was so varied that little purpose would have been achieved by including the occupational classifications in the earnings equations. However, the model did make use of the industrial groups with which respondents were associated. These were (i) Agriculture, Forestry and Fishing (AGRIC); (ii) Mining and Quarrying (MINING); (iii) Manufacturing (MANUFACTURING); (iv) Electricity, Gas and Water (ELECTRICITY); (v) Construction (CONSTRUCTION); (vi) Wholesale and Retail Trade and Restaurants and Hotels (WHOLESALE); (vii) Transport, Storage and Communications (TRANSPORT); (viii) Financing, Insurance, Real Estate and Business Services (FINANCE); and (ix) Community, Social and Personal Services (COMMUNITY).

Details of the statistical methodology are available in the two AA articles. It suffices to mention here that the data were corrected for sample selectivity bias, as suggested in Heckman (1979). As mentioned previously, such a correction has been rare in most studies of the gender earnings gap for other countries.

The AA articles also provide details of the widely-used Oaxaca technique for calculating male-female earnings differentials. This study follows AAII in not only calculating the gender earnings gap, but reporting the proportion of it which stems from discrimination.

AAII argue that the Oaxaca technique suffers from the drawback that it yields two possible estimates of the gender earnings gap, and that these estimates can be quite different from each other. The dual estimates are based on the assumption of whether the male wage structure would prevail in the absence of discrimination, or whether the female wage structure would be observed in discrimination-free environment. Citing Cotton (1988) and Neumark (1988), AAII make the case that in the absence of discrimination, the wage rate would likely settle between the pure male or pure female wage structure. (See AAII for details).

The modified AAII log wage differential is given by:

$$\ln W_m - \ln W_f = (X_m - X_f)\beta^* + X_m(\beta_m - \beta^*) + X_f(\beta^* - \beta_f) \qquad (2)$$

where β^* is a vector containing the weighted averages of the male and female OLS coefficients, with the proportion of males and females in the sample as the relevant weights. In equation (2), the first component on the right-hand-side is the skill or productivity advantage of males over females in the absence of discrimination. The second term is the "male advantage" or the amount by which males are overcompensated relative to a discrimination-free setting, and the third term is the "female disadvantage," or the amount by which females trail the wage-rate which would prevail in the absence of discrimination.

III. Empirical Results

An aspect of the 1984-85 *HIES* data deserves comment: Of 1,858 working women in that year, 369 reported an income of Rs.50,000, which is an unusually high level of income (only three women reported higher earning levels in that year, the highest being Rs.61,513). This appears to be a coding error. As a result, all women reporting income equal to or above Rs.50,000 were deleted. Since this almost certainly meant that some authentic observations on high-income women were deleted, the estimates of the male-female earnings differential reported in this article are biased upward for that year. For consistency, males reporting an income level of Rs.50,000 were also deleted. However, only 91 out of 21,459 men reported this income level (there were 12 males who reported a higher income, with the highest being Rs.90,233). This meant that while 20 per cent of all working-female observations were deleted, only about 0.4 per cent of working-male observations were not included in the computations. In other words, a very large proportion of high-income females, but a much smaller proportion of similar males were deleted. This of course resulted in an upward bias in the estimates of the gender earnings gap reported in this article. Also inherent in these observations is the broader inference that the HIES data may suffer from other coding errors also. This is suggested, for example, by the results in Table 3 in which the fall in the gender earnings gap from 63.27 per cent to 33.09 per cent between 1979 and 1985-86 defies credulity. This article may also be taken then, as a call for improved data collection in Pakistan.

The data did not provide direct information on work experience of the respondents. Many studies in the West estimate experience as [Age - Schooling - 6] following Mincer (1974). AAll argue that such a formulation would be inappropriate for Pakistan, since it is well-known that there is no uniform age at which children begin schooling in Pakistan, and that the variance of age among beginning schoolgoers is quite large (specially among rural females). Age is thus a more appropriate proxy for experience than is the Mincer (1974) formulation. It must be borne in mind that the use of age instead of experience leads to misleading inferences in the case of individuals with interrupted labor force participation.

Separate regression coefficients of the wage equations for males and females are listed in Table 2. AGE was highly significant in both sets of estimates. As mentioned previously, age is highly correlated with experience, and thus had the expected sign and magnitude. The negative (and statistically significant) coefficient for AGE SQUARED confirmed the concavity of the age-earnings profile. The missing reference variable among the provinces was NWFP. An interesting pattern was observed across the different provinces. Both men and women in the Punjab earned less than their respective counterparts in NWFP. However the coefficient estimates in both of the samples for Sind were statistically insignificant suggesting that earnings of Sindhis were no different from those of workers in the NWFP, after controlling for productivity characteristics. Most notable were the results for

Baluchistan: while Baluchi men earned *less* than those in NWFP, the opposite was the case for women (the coefficient estimates were significant in both cases).

Workers in urban areas earned more than those in rural parts of the country – a likely reflection of the differences in cost of living. Returns to the seven different levels of education rose monotonically (Educ1 – those with less than a primary level of education was the missing base variable). Among men, those with a primary education earned 17 per cent more than those who were less educated, and rose to as much as 129 per cent for those with a masters degree.[1] Among women the returns started at 24 per cent for those with primary education, and were 209 per cent for those with masters' degrees. The uniformly higher returns to all levels of education for women relative to men appears to be a premium to the relative scarcity of educated females in Pakistan. The returns to males with a primary education (Educ2) of 17.35 per cent is not much different from the 12 per cent return for this group reported by Pasha and Wasti (1989). The latter estimated the rates of return (both private and social) for the same groups as in this study. However, their estimates are not directly comparable with those in this study. This is because while this article compares rates of returns at all levels of education relative to those at the Educ1 (illiterate or less than primary) level, Pasha and Wasti's (1989) returns for each level of education is in relation to the immediately preceding level of educational attainment. Thus Pasha and Wasti (1989) report the returns to Educ6 relative to the returns to an Educ5 level, and for Educ5 relative to Educ4, etc. This paper reports all returns relative to Educ1. Other studies have also estimated the rates of return to education in Pakistan.[2] But since this is not the principal focus of this article, they have not been discussed here.

A complete listing of returns to different levels of education is presented in Table 1.

TABLE 1

Percentage Rates of Return to Different Levels of Education

	Males	Females
Educ2	17.35%	24.61%
Educ3	17.35%	36.34%
Educ4	27.12%	91.55%
Educ5	47.70%	109.59%
Educ6	78.60%	174.56%
Educ7	129.33%	209.57%

[1] The percentages reported in this section were obtained by taking the antilog of the relevant coefficient estimates of the regression equation.

[2] See for example, Khan and Irfan (1985), and Guisinger et al., (1984).

TABLE 2

Regression Coefficients of Wage Equations for Males and Females

	Males		Females	
Constant	5.07	(157.24)	5.32	(26.92)
Age	8.28	(57.90)	7.41	(8.39)
Age Squared	-8.13	(-50.26)	-8.83	(-8.76)
Punjab	-0.11	(-8.17)	-0.13	(-1.64)
Sind	0.02	(1.17)	-0.00	(-0.03)
Baluchistan	-0.06	(-2.69)	0.49	(2.75)
Urban	0.09	(7.91)	0.07	(1.03)
Educ 2	0.16	(6.93)	0.22	(1.11)
Educ 3	0.16	(9.88)	0.31	(2.20)
Educ 4	0.24	(13.45)	0.65	(3.27)
Educ 5	0.39	(22.17)	0.74	(6.42)
Educ 6	0.58	(20.05)	1.01	(6.45)
Educ 7	0.83	(28.19)	1.13	(7.12)
Mining	0.29	(2.51)	-0.63	(-0.92)
Manufacturing	-0.04	(-2.32)	-0.83	(-8.75)
Electricity	-0.09	(-2.04)	-0.08	(-0.11)
Construction	-0.18	(-9.96)	-0.61	(-2.67)
Wholesale	0.08	(4.88)	-0.57	(-4.00)
Transportation	0.04	(1.97)	0.19	(0.54)
Finance	0.18	(3.72)	-0.64	(-1.52)
Communications	-0.17	(-10.52)	-0.61	(-7.44)
Selectivity Variable	-0.06	(-1.04)	0.39	(0.51)
R^2	0.25		0.18	
N	21,459		1,489	

Note: AGE was divided by 100, since the coefficient estimates for this variable and for AGESQ were otherwise so small that they rounded off to zero. Figures in parentheses represent t-statistics.

TABLE 3

Male-Female Earnings Differentials in Pakistan

	1979	1984-85			1985-86
Pakistan	63.27%	47.90%	(21,460;	1,490)	33.09%
Pakistan (urban)	49.55%	50.88%	(10,439;	815)	32.69%
Pakistan (rural)	74.92%	47.84%	(11,021;	675)	38.29%
By Province					
Punjab	75.50%	50.94%	(11,996;	964)	39.12%
Punjab (urban)	53.06%	56.52%	(5,297;	31)	27.24%
Punjab (rural)	90.25%	46.17%	(6,699;	533)	43.54%
Sind	53.23%	52.41%	(5,186;	242)	34.92%
Sind (urban)	41.91%	52.74%	(3,178;	207)	29.71%
Sind (rural)	157.54%	15.04%	(2,008;	35)	66.63%
Baluchistan	9.14%	-22.81%	(1,106;	44)	-34.07%
Baluchistan (urban)	28.66%	-24.91%	(537;	30)	-15.31%
Baluchistan (rural)	-9.45%	-10.00%	(569;	14)	-55.49%
NWFP	28.79%	48.59%	(3,172;	240)	9.04%
NWFP (urban)	28.57%	48.59%	(1,427;	147)	9.04%
NWFP (rural)	28.11%	25.58%	(1,745;	93)	20.96%
By Industry					
Manufacturing	162.43%	171.82%	(3,294;	178)	124.81%
Construction	62.82%	91.00%	(2,101;	23)	148.56%
Wholesale	167.56%	151.16%	(3,750;	65)	94.38%
Transport	29.69%	-0.04%	(1,545;	10)	16.20%
Finance	82.33%	117.84%	(199;	7)	33.23%
Communications	70.26%	48.04%	(3,338;	390)	57.71%
Agriculture	198.56%	221.85%	(5,592;	59)	202.37%

Notes: Estimates are not presented for Mining and Electricity as a result of the very low number of females in those industries. The 1979 and 1985-86 results are drawn from the AAII results. Figures in parentheses are the number of males and females in the sample for 1984-85.

Male-female earnings differentials for the entire sample, as well as for smaller sub-groups are listed in Table 3. The gender earnings gap was 47.90 per cent in 1984-85. It was slightly higher in urban parts of the country (50.88 per cent) than in rural areas (47.84 per cent).

Viewing the results by province, it is seen that considerable variability existed in different parts of the country in terms of the gender earnings gap. Rural Sind was particularly notable with an enormous 115 per cent earnings differential, though with only 35 women in that group, results should be viewed cautiously. On the other extreme, the gender gap was *negative* 23 per cent in Baluchistan. In other words, after holding constant for productivity characteristics, women in that province actually earned more than their male colleagues. This was true across both the urban and rural parts of the country, though it was substantially higher in the former.

Among the industrial subgroups, the highest differential (222 per cent) was in agriculture. This is in line with anecdotal evidence that suggests the existence of

TABLE 4

Male-Female Earnings Differentials
Using Cotton/Neumark and Oaxaca Models

	Characteristics		Male Advantage		Female Disadvantage		Total
1984-85	14.58%	+	14.31%	+	19.01%	=	47.90%
	Oaxaca Model 1*						
			Characteristics		Discrimination		Total
1984-85			20.58%	+	27.32%	=	47.90%
	Oaxaca Model 2*						
			Characteristics		Discrimination		Total
1984-85			18.57%	+	29.32%	=	47.90%

*Oaxaca Model 1 was estimated on the assumption that male earnings would be observed in the absence of discrimination. Oaxaca Model 2 was estimated on the alternative assumption that female earnings would prevail in the absence of discrimination.

considerable suppression of women in that sector. On the other hand, it appeared that women earned no less than men in the transportation sector (–0.04 per cent). With only ten women in that sector however, great faith cannot be attached to this estimate.

Comparison with the AAII Results

By and large, the estimates of the gender earnings gap for Pakistan as a whole as well as the individual provinces are higher than reported by AA for 1985-86.[3] The reader is reminded again that the exclusion of a large number of high-income women from the sample biased the estimates of the gender gap upwards for the sample in this study. This is what has, perhaps, led to the earnings differences for 1984-85 being almost consistently higher than those for 1985-86. The percentages for 1979 indicate that there has been secular decline in the gender earnings gap over the period studied.

IV. Concluding Remarks and Summary

This study provides separate estimates of the male-female earnings differential in Pakistan for each of the four provinces, and for nine major industrial groups, using data from the 1984-85 *Household Income and Expenditure Survey*. A modified version of the Oaxaca (1993) model, recently suggested by Cotton (1988) and Neumark (1988) was used to derive the estimates of male-female earnings differentials. Corrections for selectivity bias in the data constituted a methodological improvement in the model not normally found in such studies.

Apparent coding errors suggest that data collection in Pakistan is far from satisfactory. Steps must be taken to ensure the authenticity of all data collected since all analyses must be considered suspect if this is not the case.

Textile University of Pakistan
Karachi

References

Ashraf, J., and B. Ashraf, 1993a, Estimating the gender wage gap in Rawalpindi city, Journal of Development Studies, 24(2).

Ashraf, J., and B. Ashraf, 1993b, An analysis of the male-female earnings differential in Pakistan, The Pakistan Development Review, 32(4).

Cotton, J., 1988, On the decomposition of wage differentials, The Review of Economics and Statistics, 70(2).

Guisinger, S.E., J.W. Henderson, and G.W. Scully, 1984, Estimating rates of return to education and the earnings distribution in Pakistan, Economics of Education Review, 3(4).

Heckman, J., 1979, Sample selection bias as a specification error, Econometrica, 47(1).

Khan, S.R., and M. Irfan, 1985, Rates of return to education and the determinants of earnings in Pakistan, The Pakistan Development Review, 24(3).

Mincer, J., 1974, Schooling, experience and earnings, New York: Columbia University Press.

Neumark, D., 1988, Employers' discriminatory behavior and the estimation of wage discrimination, Journal of Human Resources, 23(3).

Oaxaca, R., 1973, Male-female wage differentials in urban labor markets, International Economic Review, 14(3).

Pasha, H. and A. Wasti, 1989, Unemployment and rates of return to education, Singapore Economic Review, 34(2).