

## Ethnicity and Earnings: An Analysis of Data for Karachi

JAVED ASHRAF and BIRJEES ASHRAF

### I. INTRODUCTION

Pakistan is today, a case study in ethnic strife. Carved out of pre-Independence India in 1947, the country is home to five broad ethnic groups: *Punjabis* (the dominant segment of the population), *Balochis*, *Pathans* and *Sindhis*. Each has a distinctive culture and language. In spite of broad similarities stemming from a common religion (more than 98 percent of the population is Muslim), several factors combine to create dissension and discord. The fifth ethnic group in Pakistan are the *Muhajirs* (which literally translates to "Immigrants") who migrated to Pakistan in large numbers from India in 1947. Most of them settled in Karachi. Pakistan's largest city, a teeming port city with a current population of around 12 million. Charges of favouritism in the award of jobs abound, and the *Muhajirs* in particular allege that, not being "sons of the soil," they are singled out for discriminatory treatment.

Given that this is the case, it is surprising that no study has been carried out to ascertain if indeed, earnings differences originate from ethnic differences in Pakistan. This study attempts to examine this issue, and thus settle questions that have led to many a pitched argument.

This study uses the Oaxaca (1973) procedure as well as modifications of that model developed by Cotton (1988) and Neumark (1988) to estimate earnings differentials between different ethnic groups residing in Karachi. The data sources was the 1987 Socio-economic Survey conducted by the Applied Economics Research Centre and the Karachi Development Authority. The advantage of drawing samples of different ethnic groups from one urban location has a major advantage: earnings data need not be adjusted for differential cost of living that would otherwise be necessary if the respondents had been from different parts of the country.

### II. THE DATA AND THE MODEL

Data for this study are from the AERC/KDA *Socio-Economic Survey* of 1987. The data included 9,368 observations. The data are unique in that they had a question about each respondent's mother tongue. Nine languages were identified in the data:

Javed Ashraf is Director, Institute of Business Administration, Karachi. Birjees Ashraf is Lecturer in Economics at the University of West Florida.

*Authors' Note:* We are indebted to several colleagues for their insightful comments on earlier drafts of this paper, as well as to the participants of the 16th Annual Conference of the Pakistan Society of Development Economists, Islamabad, 2001.

Urdu, Sindhi, Punjabi, Pushto, Balochi, Gujarati, English, Kuchi and Hindko. Only observations for the first five were retained for this study. What might not be immediately obvious to the non-Pakistani reader is that mother tongue and ethnic origin are uniquely tied. The identification of ethnic origin from the mother tongue reported by respondents is not controversial, and is the basis for this study's determination of ethnic groups. No other economic data-set in Pakistan provides this particular information, and thus distinguishing ethnic groups from one another has not hitherto been possible. This explain why a study of this kind has not been undertaken previously.

The explanatory variables used in the regression model were age, the square of age, gender, seven levels of educational attainment ("less than primary" was the missing reference group), and four dummy variables to capture ethnic origin. The log of hourly wages was the dependent variables. The wage equation (unadjusted for selectivity bias) used to compute the estimates (with separate regressions for respondents from each of the five ethnic groups) was:

$$\begin{aligned} \ln HW = & a_0 + a_1 AGE + a_2 AGE-SQUARED + a_3 PRIMARY + a_5 MIDDLE \\ & + a_5 MATRIC + a_7 INTERMEDIATE - a_8 BACHELORS \\ & + a_9 MASTERS + a_{10} PROFESSIONAL + a_{11} PUNJABI \\ & + a_{12} BALOCHI + a_{13} PATHAN + a_{14} SINDHI \quad \dots \quad \dots \end{aligned} \quad (1)$$

The variables PUNJABI, BALOCHI, PATHAN and SINDHI identify ethnic background of the respondents (MUHAJIR is the missing base variable). The variables distinguishing different levels of educational attainment are self-explanatory. PROFESSIONAL represented respondents with a degree in medicine, law, etc. This study recognises that the variables used in the earnings equations are quite limited. It would have been ideal to have had information on such additional attributes as "kind of job", experience, factors that lead to compensating differentials such as pollution and job-hazards, etc. however, such information was not available from the data. Although there was information on occupations, it was so narrow that little purpose would have been achieved including them in the regression equations.

The technique used in this study to compute ethnic earnings differentials is very similar to the procedure used by Ashraf and Ashraf (1993) in a study of the gender earnings gap for Rawalpindi. Following Heckman (1979), variables to correct for sample-selectivity bias were developed, and used in the wage equations. They are defined as  $[-f(EMP_i)/F(EMP_i)]$  and  $f(EMP_i)/\{1-F(EMP_i)\}$  where  $F$  and  $f$  are the cumulative and density functions of a standard normal variable.  $EMP_i$  is the predicted employment status of an individual, obtained from probit estimates of the reduced-form equation determining employment status. Given the likelihood of different factors not influencing the work decision of different ethnic individuals in the same manner, the selectivity variables were computed from separate probits for each ethnic group. The joint determination of participation and earnings is given by:

$$\ln W_i = \beta_0 + \beta_1 X_i + \beta_2 [-f(Emp_i)/f(Emp_i)] + u_i \quad \dots \quad \dots \quad (2)$$

$$Emp_i = Z' \delta_i + e_i \quad \dots \quad \dots \quad \dots \quad \dots \quad (3)$$

for respondents from each group.

The  $X_i$  represent characteristics of the respondents that impact on earnings as represented in Equation (1).  $Z$  is a subset of  $X$ , and represents those worker characteristics that are instrumental in determining whether an individual will be in the work force. In Equation (2), the use of the selectivity variables leads to consistent estimation of the coefficients of the equation. Drawing from Oaxaca (1973), a model for calculating the ethnic wage gap is:

$$\ln W_i - \ln W_j = X'_i B_i - X'_j B_j \quad \dots \quad \dots \quad \dots \quad \dots \quad (4)$$

where  $X_i$  and  $X_j$  are vectors containing means of the variables for the  $i$ th and  $j$ th ethnic groups, while  $B_i$  and  $B_j$  are vectors with the OLS coefficients estimates for the  $i$ th and  $j$ th ethnic group, respectively. The log wage differential represented in Equation (4) can be expressed as:

$$\ln W_i - \ln W_j = (X'_i - X'_j) B_i + X'_j (B_i - B_j) \quad \dots \quad \dots \quad \dots \quad (5)$$

$$\text{or as } \ln W_i - \ln W_j = (X'_i - X'_j) B_j + X'_i (B_i - B_j) \quad \dots \quad \dots \quad (6)$$

Equations (5) and (6) have different interpretations. Equation (5) implies that in the absence of discrimination, the wage structure of the  $i$ th group would prevail in the market, while Equation (6) assumes that it is the wage structure of the  $j$ th group that would obtain in a non-discriminatory environment. The two assumptions do not yield the same estimate for the discrimination component, and sometimes these estimates can be significantly different from each other.

Cotton (1988) has argued that the Oaxaca decomposition procedure is flawed, since the wage structures of neither the  $i$ th nor  $j$ th group would prevail in a discrimination-free setting. Instead he suggests that members of the  $i$ th group will be paid more than the non-discriminatory wage and those from the  $j$ th group will be paid less. Hence, the discrimination component should comprise of two parts: one representing the amount by which characteristics of the  $i$ th group are overcompensated relative to their marginal product (hereafter referred to as the “favoured group advantage”) and the other representing the amount by which characteristics of the  $j$ th group are under compensated (hereafter referred to as the “non-favoured group disadvantage”). The true non-discriminatory wage would therefore lie somewhere between the wage structures of the  $i$ th and  $j$ th groups used by Oaxaca. Specifically, Cotton’s log wage differential is given by:

$$\ln W_i - \ln W_j = \beta^*(X_i - X_j) + X_j(\beta^* - \beta_j) \quad \dots \quad \dots \quad \dots \quad (7)$$

where  $\beta^*$  is a vector containing the weighted averages of the OLS coefficients from the regressions for the  $i$ th and  $j$ th groups. The weighted average uses the proportion of respondents from the two groups as the relevant weights. In Equation (7) then, the first component on the right-hand-side is the skill or productivity advantage of members of the  $i$ th group over members of the  $j$ th group in the absence of discrimination. The second term is the “favoured group disadvantage” or the amount

by which members of the  $j$ th group trail the wage-rate that would prevail in the absence of discrimination.

This study recognises that estimates reported here may actually overstate or underestimate the true level of discrimination. Relevant variables that impact upon productivity may have been omitted. If members of the  $i$ th group, on average possess more of the productivity enhancing attributes not included in our model, and over estimate of discrimination results. Furthermore it has been suggested in the literature that some of the differences in such attributes may stem from discrimination prior to an individual's entry into the labour market. Polacheck (1978) supports this kind of reasoning. In a study of gender earnings differences, he argued that, ".....if discrimination in the labour market causes females to specialise differently in the human capital market, then looking only at the labour market underestimates the far-reaching effects of discrimination". It is in this spirit that this article concedes that similar biases along ethnic lines are likely in Pakistan. One might note, by way of example the heavy number of security guards and truck drivers of *Pathan* origin in Karachi. Similarly, relatively few people of *Muhajir* origin are associated with the armed forces. This paper recognises biases in reported estimates stemming from the inability of the data to capture influences of the kind described. Specifically, reported estimates of discrimination most likely include the effect of pre-market and extra-market forces.

### III. EMPIRICAL RESULTS

Table 3 lists the regression coefficients of the earnings equation for each of the five ethnic groups. Regression estimates are also provided for the entire sample. Unfortunately, the data did not allow determination of the experience level of the individuals in the sample. Following Mincer (1974), it has been common practice in the west to estimate experience as [Age-Schooling-6]. However, 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 school-goers is quite large (specially among rural females), age is a more appropriate proxy for experience than is the Mincer formulation. Clearly however, the use of age instead of experience leads to misleading inferences in the case of individuals with interrupted labour force participation.

AGE (or by proxy, experience) was statistically significant for all ethnic groups. It was not noteworthy however, that the absolute value for the coefficient was much lower for *Muhajirs* than for the other ethnic groups.

This implies that experience had a smaller effect on *Muhajir* earnings than it did on those from the other ethnic groups, AGE-SQUARED was negative and highly significant in all ethnic groups confirming the concavity of the age-earnings profile.

As expected, the variable for MALE was highly significant across all provinces. The existence of considerable discrimination against women in Pakistan has been previously documented elsewhere. The male-female earnings differentials across different ethnic groups are provided in Table 1.

Table 1

*Male-Female Earnings Differentials*

Ethnic Group	Earnings Differential	No. of Males/Females
Muhajir	73.33%	4,919; 399
Punjabi	59.99%	1,291; 55
Balochi	49.18%	498; 16
Pathan	39.10%	694; 7
Sindhi	118.15%	455; 25

The male-female earnings differential reported in Table 1 was calculated as  $\exp(M) - 1$  where  $M$  is the coefficient estimate of the MALE variable in the regression estimates of each province. A high degree of variability in the male-female earnings differential across different ethnic groups is noted. However, relatively little credence can be placed in any of the estimates except for *Muhajirs*, and to degree, the *Punjabi* groups. This is because of the very low number of women in the other groups. As a result valid questions might arise about how representative such small samples are of the population of women at large. The gender earnings gap is noticeably higher among *Muhajirs* than it is among *Punjabis*.

The coefficient estimates for the variable representing different levels of educational attainment were consistent with *a priori* expectations. Earnings levels rose monotonically with the level of educational attainment for most groups. In some cases where this was not case (MASTERS for *Balochis* and *Pathans*, and PROFESSIONALS for *Balochis*) the reason was, as in the case of the gender earnings differential, the extremely small number of observations for some of the higher educational levels. This could have led to perverse results, since the few observations may not have been representative of a wider population.

Highlighting this point, the means of the variables in Table 4 shows that only 0.002 percent of *Balochis* had a masters degree. Since the sample consisted of 514 *Balochis*, this means only one *Balochi* held a masters degree.

The percentage gains from different levels of education (with "less than primary" as the reference group) is listed in Table 2 below.

Table 2

*Percentage Earnings Gain from Different Levels of Education*

Education Level	Muhajir	Punjabi	Balochi	Pathan	Sindhi
Primary	18.53%	17.35%	9.41%	20.92%	41.90%
Middle	20.92%	18.53%	19.72%	12.74%	18.53%
Matric	53.72%	61.60%	17.35%	29.69%	78.60%
Intermediate	101.37%	109.5%	61.61%	59.99%	50.68%
Bachelor's	150.92%	138.6%	69.89%	50.68%	99.37%
Master's	242.12%	197.4%	17.35%	46.22%	131.63%
Professional	285.74%	309.5%	76.82%	182.92%	405.30%

As a result of the small number of observations for individuals with higher levels of education in some of the ethnic groups, the results for *Muhajirs* and *Punjabis* are the only ones in which much faith can be reposed. It appears that the returns to education for both groups were fairly similar. Even respondents with only a primary level of education earned 17-18 percent more than the reference group of individuals who had less than a primary level of education. As the table indicates, *Punjabis* with professional qualifications earned as much as 309 percent more than the reference group, while for *Muhajirs*, this figure was 286 percent. These figures appear to be reasonable, and in line with *a priori* expectations.

The coefficient estimates for the selectivity variable were statistically insignificant for all five ethnic groups. This indicates that there is no self-selectivity in the individuals from various provinces who have chosen Karachi as a home.

Table 4 provides the mean of variable is used in the regression equations. It is evident that the observations from each province consist mostly of males. It is also clear that most of the respondents are not highly educated, which appears to be in conformity with patterns in the general society. Most of the individuals hail either from the Punjab (13.6 percent) or are *Muhajirs* (56.80 percent).

#### IV. ETHNIC EARNINGS DIFFERENCES

In Table 3, the regression estimates for the entire sample give coefficient estimates on the dummy variable for respondents from each of the four provinces. It

Table 3

*Regression Coefficients of Wages Equations for Different Ethnic Groups*

	All	Muhajir	Punjabi	Balochi	Pathan	Sindhi
Professional	1.35***	1.35***	1.41***	0.57	1.04***	1.62**
Age-Squared	-1.44***	-1.24***	-5.21***	-8.14***	-6.53***	-6.79***
Primary	0.18***	0.17***	0.16**	0.09***	0.19**	0.35**
Male	0.54***	0.55***	0.47***	0.40	0.33**	0.78***
Intermediate	0.67***	0.70***	0.74***	0.48***	0.47***	0.41***
Master's	1.16***	1.23***	1.09***	0.16	0.38***	0.84***
Middle	0.17***	0.19***	0.17***	0.18***	0.12***	0.17*
Matric	0.43***	0.43***	0.48***	0.16***	0.26***	0.58***
Bachelor's	0.89***	0.92***	0.87***	0.53***	0.41***	0.69***
Age	2.84***	2.79***	5.69***	7.71***	6.21***	6.65***
Constant	5.60***	5.55***	5.23***	5.05***	5.42***	4.83***
Sindhi	0.01					
Balochi	0.02					
Pathan	0.01					
Punjabi	0.05**					
Selectivity	0.02	-0.21	0.34	-0.46	0.21	-0.32
R <sup>2</sup> :	0.36	0.40	0.36	0.19	0.16	0.31
No of Observations	9,368	5,318	1,274	514	701	480

*Note:* AGE was divided by 100, since the coefficient estimates for this variable and for AGE-SQUARED were otherwise so small that they rounded off to zero. \*, \*\* and \*\*\* indicate statistical significance at the 90 percent, 95 percent, and 99 percent levels.

is noticeable that *Sindhis*, *Balochis* and *Pathan* earn 1 percent, 2 percent and 1 percent more than *Muhajirs*, although the coefficients are not significant. *Punjabis* however earn 5 percent more, and this is statistically significant. Using the Cotton-Neumark decomposition technique, the results are little changed (Table 5). *Punjabis* continue to earn 5 percent more than *Muhajirs*. But *Balochis* and *Sindhis* now earn 2 percent and 1 percent more, respectively, than *Muhajirs*, while *Pathan* earn 2 percent less. The small magnitudes of the earnings differentials tell an important story: in spite of much propaganda by various groups about discrimination, the data do not appear to suggest any major earnings discrepancies attributable to ethnic origin.

Conversely, the small differences between different groups can be viewed as evidence that market forces rather than ethnic backgrounds determine earnings—a result that should equality, and damaging to the cause of those who allege victimisation based on one's place of birth.

Table 4  
*Means of Variables*

Variable	All	Muhajir	Punjabi	Balochi	Pushto	Sindhi
Professional	0.042	0.059	0.037	0.002	0.006	0.130
Age-squared	0.141	0.142	0.147	0.143	0.143	0.130
Primary	0.042	0.035	0.056	0.060	0.044	0.056
Male	0.939	0.925	0.957	0.969	0.990	0.948
Intermediate	0.094	0.123	0.084	0.035	0.029	0.033
Master's	0.033	0.048	0.014	0.002	0.003	0.035
Matric	0.168	0.179	0.197	0.101	0.107	0.077
Bachelor's	0.150	0.201	0.101	0.047	0.036	0.083
Age	0.351	0.352	0.362	0.354	0.353	0.334
Balochi	0.055					
Punjabi	0.136					
Pathan	0.075					
Sindhi	0.051					
No. of Observations	9,368	5,318	1,274	514	701	480

Table 5  
*Ethnic Earnings Differences*

Skill Advantage	Favoured Group Advantage	Non-favoured Group Disadvantage		Total
		$X_i (\beta^* - \beta_j)$	$X_j (\beta^* - \beta_j)$	
$\beta^* (X_i - X_j)$	$X_i (\beta^* - \beta_j)$			
Muhajir – Punjabi	0.14	–	0.08	–
Muhajir – Balochi	0.27	–	0.10	–
Muhajir – Pathan	0.25	–	0.18	–
Muhajir – Sindhi	0.33	–	0.27	–
				0.12 = –0.05
				0.05 = –0.02
				0.05 = 0.02
				0.07 = –0.01

## V. CONCLUDING REMARKS AND SUMMARY

This study examined earnings differentials between different ethnic groups in Pakistan. Specifically, differences were estimated between *Muhajirs* (the reference groups) and *Punjabis*, *Balochis*, *Sindhis* and *Pathan*. Recently developed models to estimate such differences suggested by Cotton and Neumark were used for the estimation. The results do not support the contention that ethnic background plays a major role in determining earnings in Pakistan.

## REFERENCES

- Ashraf, Javed, and Birjees Ashraf (1993) Estimating the Gender Wage Gap in Rawalpindi City. *Journal of Development Studies* 29, 365–76.
- Cotton, Jeremiah (1988) On the Decomposition of Wage Differentials. *The Review of Economics and Statistics* 70, 236–43.
- Heckman, James (1974) Sample Selection Bias as Specification Error. *Econometrica* 47, 153–61.
- Mincer, Jacob (1974) *Schooling Experience and Earnings*. New York: Columbia University Press.
- Neumark, David (1988) Employers Discriminatory Behaviour and Estimation of Wage Discrimination. *Journal of Human Resources* 23, 279–095.
- Oaxaca, Ronald (1973) Male-Female Wage Differentials in Urban Labour Markets. *International Economics Review* 14, 693–709.
- Polacheck, S. (1978) Sex Differences in College Major. *Industrial and Labour Relations Review* 31, 498–508.

## **Comments**

This paper touches upon a very sensitive issue of discrimination due to ethnicity in the labour market. Authors have used earnings functions for the analysis for different ethnic groups and used Oxaca technique on the estimates for decomposition of earnings to detect discrimination. As expected, authors could not find support for the hypothesis that Mohajir (dominant ethnic group in Karachi city) face discrimination in the labour market.

In my opinion selection of Karachi city for this purpose is not very suitable because most of the businesses are owned by the Mohajirs who may not like to discriminate against people of their own ethnic group rather they may discriminate against other minorities. A nationally representative data would have been more useful for such type of analysis. Another problem of the analysis is to gauge the extent of discrimination through earnings, which may not reflect the actual situation. The use of other measure such as unemployment rates of different ethnic groups may have provided stronger evidence of discrimination.

As mentioned by the authors themselves, the female sample for different ethnic group is so small that no conclusive statement can be made on the basis of results obtained from this data.

Secondly the assumption that mother tongue and ethnic origin are tied may not be true. Most of the families residing in Urdu speaking areas for long may have adopted the language and other customs of the Urdu speaking communities, i.e. assimilation of immigrants. So tying mother tongue with ethnic origin may not be good assumption. In these situations, results may bear no credibility and have no use for any policy formulation.

**Zafar Mueen Nasir**

Pakistan Institute of Development Economics,  
Islamabad.