Primary completion rates across socio-religious communities in West Bengal

Husain, Zakir and Chatterjee, Amrita

Institute of Economic Growth, Delhi, Institute of Development Studies Kolkata

11 April 2009

Primary Completion Rates across Socio-Religious Communities in West Bengal

ZAKIR HUSAIN, AMRITA CHATTERJEE

Primary completion rates of Muslims in West Bengal are substantially lower than that of upper caste communities as well as backward castes, scheduled castes and tribes. Further, analysis of age-specific PCR indicates that differences in PCR between Muslims and other communities may have actually increased in recent years. An econometric analysis reveals that age, gender, household size and expenditure levels, education and gender of decision-maker, etc, are important determinants of these differences in PCR. But use of Census data and District Information System for Education statistics indicates that deficiencies in infrastructural facilities in Muslim-concentrated districts also have a significant role in the low PCRs of Muslim children.

The authors would like to express their thanks to Debdas Banerjee and Achin Chakraborty for their comments and suggestions. Needless to say all deficiencies are the sole responsibility of the authors.

Zakir Husain (*dzhusain@gmail.com*) and Amrita Chatterjee (*chatterjeeamrita07@gmail.com*) are with the Institute of Development Studies, Kolkata.

The importance of education as an important development goal has been widely recognised. It has been pointed out that education has important consequences for growth. By improving technical and cognitive skills it increases the productivity of the workforce and facilitates diffusion of technology and new ideas (Bruns et al 2003). Education also generates substantial externalities, like better natural resource management (Godoy and Contreras 2001) and lower incidence of HIV/AIDS (World Bank 2002). In particular, education of girls improves health outcomes for infants and children, family nutrition levels and (in subsequent generations) educational outcomes (Hadden and London 1996). Empirical studies have also noted the high correlation between education and lower levels of poverty and inequality (Birdsall and Londoño 1998). Not only does education facilitate sustained growth and development, but as pointed out by Sen (1985, 1999), it is important as an independent goal, by improving human capabilities.[1]

1 Introduction

Recognising the importance of education, policymakers initially emphasised increasing access to schooling. Over time, however, it was realised that non-attendance, grade repetition, dropouts and poor retention rates were reducing the value of gross enrolment rates as an indicator of progress on the education front. The 1990 World Conference on Education for All in Jomtien, Thailand, therefore, emphasised on providing universal primary education to ensure the reaping of the substantial benefits associated with education.[2]

Despite the attempt to increase primary completion rates (PCR) after the Jontien Conference, the progress made on this front has been modest – at the global level, PCR has increased from 72% from 1990 to 77% in 2000. South Asia has the lowest regional PCR (70%) after Africa (Bruns et al 2003), while PCR in India is even lower at 66%, according to the 2001 Census. While a concerted attempt is needed to improve PCR, intervention strategies must be based on an understanding of the factors responsible for low PCRs. In particular, sections of communities with low PCR should be identified and an attempt made to address the causes underlying their relative deprivation.

There is a substantial body of literature on factors influencing probability of primary completion. Based on limited dependent regression models,[3] studies have identified factors like family income or wealth, parental education, empowerment and education of mother, credit constraints, age and gender of the child, family size or presence of siblings, caste affiliations, place of

residence and educational infrastructure as determinants of PCR (Akhtar 1996; Deolalikar 1997; Tansel 1998; Brown and Park 2002; Connelly and Zheng 2003; Boissiere 2004; Desai and Kulkarni 2005; SIS/DPP 2005, Das and Mukherjee 2007, 2008; Okumu et al 2008). One limitation of these studies is that they have failed to explore the importance of religious identity in educational decision-making. In developing countries, particularly those in south Asia, religion play an important role in determining educational attainment. In India, for instance, a recent report has shown that PCRs of Muslims is substantially below that of other communities (GOI 2006).

The objective of this paper is to probe deeper into this issue by examining the extent and causes of differences in probability of completing primary level of education between Muslims and other socio-religious communities (SRCs). We consider only one state, West Bengal, as there are considerable differences in the historical and socio-cultural environment across regions. The focus on West Bengal may be justified on the following grounds:

(a) The absolute size of the minority population is higher in West Bengal than in other Indian states (except Uttar Pradesh);

(b) The proportion of minorities in total population is higher in West Bengal than in any other major state. Only some small states like Assam, Jammu and Kashmir, north-eastern states and some Union Territories have a higher proportion of minorities;

(c) Ideologically, the Left Front coalition government is committed to secular principles, and its electoral manifesto emphasises upliftment of marginalised sections of the population; and,

(d) Political stability has given the state government the opportunity to implement its objectives of social justice and progress through long-term plans.

Table 1: Share of Socio-Religious Communities in Population in West Bengal (2004)

Analytical Categories used	Socio-Religious Communities	Rural	Urban
Hindu-upper caste (H-UC)	Hindu – upper caste	27.35	12.64
Hindu-backward castes (H-BC)	Hindu – scheduled castes	27.17	19.32
	Hindu – scheduled tribes	6.37	1.21
	Hindu – other backward castes	6.08	5.95
Muslims	Muslims	31.14	14.00
Others	Other minorities	0.99	1.01
	Religion/caste not stated	0.91	0.36
Total population		100%	100%

The scheme of this paper is as follows: Section 2 states the database and methodology used in the paper. Section 3 presents estimates of PCR among children aged 12-15 years for different analytical categories of the population; this is followed by a discussion of age-specific PCRs. In Section 4 we use standard econometric techniques to identify determinants of probability of PCRs. The next section "decomposes" differences in probability of PCR between "explained" and "unexplained" components.[4] The latter may be due to lack of infrastructure in Muslim-dominated areas or due to lack of demand. Section 6 examines both these issues.

2 Database, Methodology and Hypothesis

This paper is based on unit level data from the National Sample Survey (NSS) 61st round (2004-05). This database has information on the religion and caste of the respondents. These two variables have been combined to create socio-religious communities for analytical purposes. The socio-religious communities are upper caste Hindus, Hindu scheduled caste, scheduled tribes and other backward castes, Muslims, other minorities and a residual group who had not stated their religion. These categories are reclassified into four socio-religious communities, given in Table 1.[5]

The state-level figures are decomposed by gender and place of residence to get four analytical categories – urban males, urban females, rural males and rural females. Our analysis is undertaken for each of these categories as the process of educational decisions may vary across these categories.

The analysis is undertaken based on the following hypotheses:

(1) There is a significant disparity in PCRs between Muslims and other communities;

(2) Within the Muslim community, girls constitute a particularly vulnerable section;

(3) The low current PCR of Muslims is in line with a history of backwardness; and,

(4) Educational backwardness of Muslims may be explained only partially by household and personal characteristics (like household size, per capita income, gender and education level of head of family, gender of the child, his/her age).

Two sets of methodologies are used – descriptive analysis and econometric. Firstly, the NSS data is used to estimate the proportion of children completing primary education out of children aged 12-15 years.[6] Following the Bruns et al (2003) definition, the PCR is defined as the ratio of number of children completing primary level of education in the appropriate age group to number of children in the age group.[7] This is followed by estimation of age-specific primary completion rates as a proxy for time trends in primary completion rates. The methodology is explained in Section 3.2.1.

Secondly, we attempt to identify the factors influencing enrolment, completion of primary education and choice of schools based on econometric methods. The analysis is undertaken using logit models. Subsequently, the results are decomposed by the extension of the Oaxaca method (Oaxaca 1973) applicable to logit models suggested by Bauer and Sinning (2008) and Fairlie (1999, 2003, 2005).

3 Main Findings

This section discusses the main findings of the analysis.

3.1 Status of Primary Education Completion

The PCR among children aged 12 to 15 years is 70% in West Bengal. However, minority communities have a substantially lower PCR – it is 37% for Muslim children and 35% for children from others. In contrast, PCR is higher among not only H-UC (51%), but also H-BC (45%).

Table 2: Primary Completion Rates among Children Aged 12-15 Years (2004)

Socio-Religious Communities	Urban			Rural		
	Total	Boy	Girls	Total	Boy	Girls
Muslim	65.8	61.2	70.5	65.0	61.8	68.1
H-UC	86.8	88.4	85.1	80.5	80.9	80.0
H-BC	72.8	71.4	74.2	69.4	73.9	64.6
Others	100.0	100.0	100.0	51.4	58.8	44.4
Total	78.8	78.3	79.3	70.0	70.8	69.3

Disaggregation of state-level figures by place of residence and gender reveal a similar picture of educational backwardness of Muslims (Table 2, p 60). Muslims have a lower PCR than both H-UC and H-BC in both urban and rural areas. The gap is relatively wider in urban areas. This may be because of greater poverty among urban Muslims (GOI 2006) and the presence of a substantial non-Bengali population, who are believed to attach lesser value to education.

The observed differences in PCR across gender go against the well-known belief that Muslim parents are only interested in education of their boy child – primary completion rates are higher among Muslim girls than among Muslim boys. In rural areas, in particular, PCR of Muslim girls is higher than that of H-BC and other minorities. Overall, however, gender differences are not marked.

3.2 Trends in Primary Education Completion Rates

First the methodology and then the findings of the analysis:

3.2.1 Methodology

Although this paper focuses on the educational status of SRCS in 2004, it is necessary to place the findings in a temporal context. This requires us to examine *changes* in status of education of SRCS *over the years*. Unfortunately, NSS neither provides time series data nor does it have sufficient large number of rounds containing religion-wise data.[8] Therefore to create a time series on educational attainment we have followed the methodology suggested by Shariff (1999).

As NSS provides data on age and education level of respondents, we may estimate the percentage of persons completing primary education in different age groups. Such age-specific primary education completion rates may be used to estimate the primary education completion rates (PCR) at different points of time as follows. We assume that a person will complete his/her primary level within 12 to 15 years. Given that a person aged 16 to 20 years in 2004, when the NSS 61st round survey was conducted, it is an easy task to compute the year (2000) when that person would have aged 12 to 15 years, and should have completed his primary education. Proceeding in this way, the corresponding year in which each age group had attained 12 to 15 years, can be calculated. This is illustrated in Table 3.

Table 3: Conversion of Age Groups into Years

Age Group (in 2004)	Year	Age Group (in 2004)	Year
12-15	2004	41-45	1975
16-20	2000	46-50	1970
21-25	1995	51-55	1965
26-30	1990	56-60	1960
31-35	1985	61 and above	1958
36-40	1980		

Since the sample size of others may be small in some age groups, results for this community may be misleading. So the analysis is undertaken only for H-UC, H-BCS and Muslims.

3.2.2 Findings

Figure 1a presents the trend among urban males. The PCR for H-UC has remained steady at above 80%. On the whole, PCR for H-BC exhibits an upward rising trend, although there are sharp fluctuations in some years. In the case of Muslims, PCR had increased from the 1970s, but has decreased from the mid-1990s. It should also be noted that PCR of Muslims has generally remained lower than even that of H-BCS.

Among urban females, H-UC historically has the highest PCRS. Among Muslims, PCR has increased steadily from the late 1980s, but has decreased thereafter. Currently, it is about 71%, which is lower than even the PCR of H-BCS. The rate of increase in PCR among H-BCS is sharply pronounced from the 1980s; as a result their PCR has converged with that of H-UCS.

Figure 1a: Trends in PCR in West Bengal – Urban Males
(% population completing primary level)

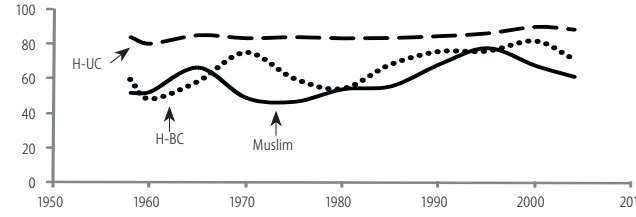
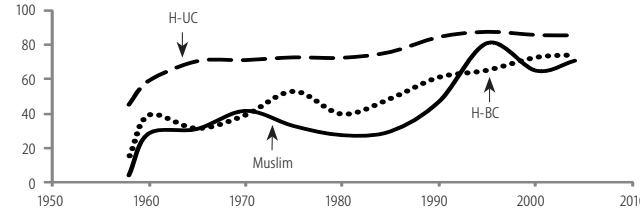


Figure 1b: Trends in PCR in West Bengal – Urban Females
(% population completing primary level)



A similar picture of elative deprivation is observed for Muslim males in rural areas. While PCR has increas ed for all the three SRCS, it has remained highest for H-UC and lowest for Muslims. While H-BCS were educationally backward in the 1950s, they have almost managed to converge with H-UC in recent years. Again, the PCR of Muslims is significantly lower than for other communities. In fact, after 2000, PCR has actually declined.[9]

The picture for rural females is interesting. While PCR has increased for all SRCS, Muslim females have performed quite well. By and large, Muslim females have performed better than H-BC. Although their PCR is lower than that of H-UC, the difference seems to have narrowed somewhat over the last two decades.

Thus, age-specific analysis of PCRS indicates that low PCR among Muslims is not a recent phenomenon. Statistical tests (based on the standard normal statistic, τ) clearly show that such disparities have been a persistent feature of West Bengal society since Independence. After Partition the more educated among the Muslims migrated to East Pakistan so that the education level of the entire community fell below that of even H-BCS. Unfortunately, while PCR has risen for all communities, it has risen at a slower rate among Muslims. The consequent divergence between PCR of Muslims and PCR of other communities constitutes an important area for policy intervention.

Another finding with important policy implications is that, *from the late 1990s onwards, there has actually been a decline in the PCR of Muslim males*. This decline is remarkable in urban areas, where the decline is about 17%! Unless there has been a sharp

change in attitude of Muslims,[10] the explanation has to be sought in supply side factors.

4 Econometric Analysis

This section carries out an econometric analysis on the factors determining the probability of completing primary education.

4.1 Variables

We next undertake an econometric analysis to identify the factors determining probability of completing primary education. Based on the literature referred to in Section 1, we regress probability of completing primary education on the following independent variables:

(1) Monthly family expenditure (MFE): Higher levels of MFE mean that the household has the resources to invest in education of children. This is important, given the high costs of the supposedly "free" primary education (PROBE 1999; NSS 1998). Thus, we expect that MFE will have a positive effect on enrolment rate and completion rate of primary education. Correspondingly, dropout should be lower with high levels of MFE. Given the high values of MFE with respect to the values of dependent variables, we have taken the logarithmic transformation of MFE (LMFE).

Figure 1c: Trends in PCR in West Bengal – Rural Males
(% population completing primary level)

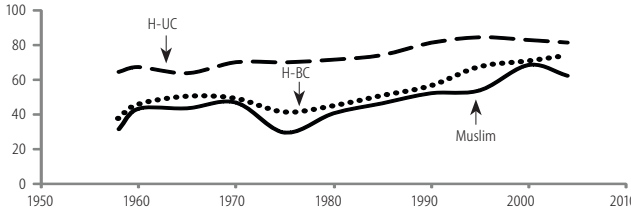
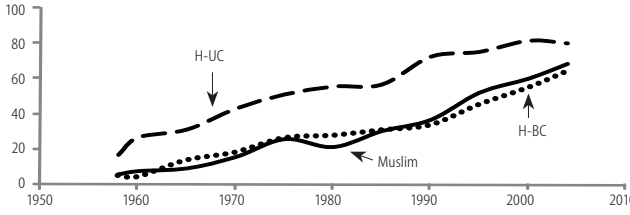


Figure 1d: Trends in PCR in West Bengal – Rural Females
(% population completing primary level)



(2) Household size (HHSIZE): The household size is likely to affect educational status adversely, unless the dependency ratio is low.[11]

(3) Gender (FML): There is a vast body of literature (PROBE 1999) arguing that families tend to invest less in the education of the girl child. To verify this hypothesis, we have constructed a gender dummy, FML equal to unity if the child is female and 0 otherwise.

(4) Age (AGE): As a child becomes older it becomes more difficult for him/her to join school. This is partly because of incompatibility with other children of the same class, or because of loss of interest in schooling, or because parents tend to engage him/her in productive or household tasks, rather than keep him/her idle (increasing opportunity costs of schooling). After getting enrolled, the probability of dropout increases as the child grows older. This may be because parents have a target level of education that is typically low, or because they want to divert the financial resources for schooling to younger siblings (Husain 2005). However, the probability of his/her completing primary level should be higher than younger children.

(5) Educational level of head of household (EDU _ DM): Literature shows that if the decision-maker is him(her)-self educated then he(she) is more likely to educate his(her) children. While studies generally focus on literacy we have taken different levels of education as defined in NSS.[12]

(6) Gender of the head of the household (SEX _ DM): Gender of the head of the household also affects education-related decisions. If a female is the decision-maker, it may be expected that she is more responsive to the needs of educating a child, particularly the girl child.[13] On the other hand, if she is also the main income earner then she may have to remove girls (to undertake household chores) and even boys (if gender discrimination in the labour market results in her getting relatively lower wages, then family income may have to be supplemented).

(7) Socio-Religious Community: Recent literature has shown the presence of inequality in educational attainments at the all-India level between socio-religious communities (Desai and Kulkarni 2005; Sengupta and Guha 2006; GOI 2006; Alam and Raju 2007). This is also supported in our findings for West Bengal. We have therefore constructed dummies to capture differences in educational attainments between SRCS, taking Muslims as the reference category:

H-UC = 1 if respondent belongs to Hindu upper caste, = 0 otherwise

H-BC = 1 if respondent belongs to Hindu backward caste,[14] = 0 otherwise

Others = 1 if respondent belongs to non-Muslim minorities and others, = 0 otherwise

4.2 Econometric Model

The dependent variable – whether the child has completed primary level or not – is a binary variable, assuming values of 1 (if the child has completed primary level) or 0 (if the child has failed to complete primary education). In this situation, logit or probit models are commonly used, estimated using the maximum likelihood method. These models differ with respect to specification of the error term – if the error term follows a logistic distribution *logit* model is used, while *probit* model is used if the error term follows a normal distribution. Since the cumulative normal and logistic distributions are very close to each other except at the tails, we are not likely to get very different results for the logit and probit models. However, while the results of the non-linear decomposition carried out subsequently has been theoretically verified for the logit model, they have been only observed as an empirical regularity for probit models (Fairlie 2005). We have therefore used the logit model in our analysis. The results of this econometric exercise are given in the appendix, while a discussion of the results is undertaken in the subsequent sub-section.

4.3 Results

Three models are estimated – for the entire state (incorporating a dummy for urban areas), the rural population and the urban population.

The results of the econometric exercise are, by and large, as hypothesised (Table 4). It can be seen that family expenditure levels, household size, age of child and educational level of decision-maker are statistically significant and affect PCR as hypothesised. PCR is not statistically lower for girls,[15] or if the decision-maker is a female. The former finding is consistent with the low difference in PCR of boys and girls in both rural and urban areas, observed in Table 1 earlier. The place of residence is not significant in Model 1, despite sharp differences in rural and urban PCRs (Table 1). The reason for differences observed in Table 1 is mainly due to differences in household expenditure level. Controlling for this difference, the rural-urban difference in PCR is eliminated.

Table 4: Determinants of PCR

Variables	Hypothesised	Model 1: West Bengal	Model 2: Rural	Model 3: Urban
Family income (log)	+	+	+	+
Household size	-	-	-	-
Female child	-	Insig	Insig	Insig
Age of child	+	+	+	+
Education of decision-maker	+	+	+	+
Female decision-maker	+	Insig	Insig	Insig
Urban	+	Insig	NA	NA
H-UC	+	+	+	+
H-BC	-	+	Insig	Insig
Others	-	-	-	No variation

The results also show that socio-religious identity is important in determining PCR. Upper caste Hindu children have a higher probability of completing primary level than Muslim children. While H-BC have a higher probability of completing primary level than Muslims in model 1, the difference is statistically insignificant when regression models are run for rural and urban areas separately. Children from non-Muslim minority communities, however, have an even lower probability of completing primary education in rural areas than Muslim children.

5 Decomposing Differences in PCRs across Communities

Now the differences in PCRs between communities observed in this analysis may be explained in terms of household and individual characteristics. It may also be attributed to "unexplained" factors – lack of demand for education or inadequate educational infrastructure in areas with high concentration of Muslims. This section estimates the contribution of household and individual characteristics in explaining differences in PCR between Muslims and H-UC.

The issue of unravelling the relative contribution of explained (that is, dependent variables) and unexplained variables was first addressed by Blinder (1973) and Oaxaca (1973) in the context of discrimination in wages in the labour market. While a technical discussion is undertaken in Appendix B, an intuitive explanation of the issues is presented here. Starting with the proposition that differences in wages could be attributed to differences in endowments (given identical regression coefficients) and to differences in regression coefficients (given identical endowments), they attempted to estimate the relative contribution of each of these factors. An important problem that arises in this context is which coefficient vector to use in the decomposition process. While it was initially suggested (Oaxaca 1973) that the coefficient vector of either of the two groups be taken, there have been alternate suggestions. For instance, while Remiers (1983) proposes that the simple average of the two groups be used, Cotton (1988) suggests using the relative sample size of the majority group as weight. On the other hand, Neumark (1988) and Oaxaca and Ransom (1994) derive the coefficient vector based on the pooled sample.

The Oaxaca decomposition is, however, not appropriate in the non-linear case. There has been several attempts to develop appropriate decomposition techniques in the non-linear case (Gomulka and Stern 1990; Yun 2004; Fairlie 1999; 2005; Bauer and Sinning 2008), and corresponding decomposition packages.

Differences in probability of completing primary level of education between Muslims and H-UC are decomposed using the *nldecompose* package (after minor modifications) in STATA.[16] The explained proportion and the residual are given in Table 5.

Table 5: Decomposition of Differences in Probability of Completing Primary Level

Reference Category	Components	Rural	Urban
'Backward' group	Explained	69.94	69.59
	Residual	30.06	30.41
'Advanced' group	Explained	71.71	61.41
	Residual	28.29	38.59
Reimers (1983):	Productivity*	71.26	66.24
Simple average	Advantaged**	14.86	15.02
	Disadvantaged***	13.88	18.74
Cotton (1988):	Productivity	71.3	63.59
Average, weighted by relative	Advantaged	14.26	6.18
sample size of groups	Disadvantaged	14.44	30.24
Neumark (1988):	Productivity	73.84	66.81
Pooled sample	Advantaged	13.6	6.9
	Disadvantaged	12.56	26.28

* Corresponding to variation explained by endowments.
** Represents the unexplained advantage of the superior group.
*** Represents disadvantage of the inferior group.

One of the disadvantages of the *nldecompose* package is that it fails to identify the contribution of individual variables to the explained difference. This deficiency is covered in the *fairlie* package, based on Fairlie (1999, 2005), developed by Ben Jann.[17] We have used this package to identify contribution of individual variables for pooled sample – corresponding to the method suggested by Neumark (1988). The results are stated in Appendix A (p 67).

The results show that factors like expenditure levels, household size, age of child and the educational level and gender of the decision-maker operate to create a gap in PCR between Muslims and H-UC. However, the contribution of gender of the child is interesting.[18] In rural areas, it does not contribute to differences in PCR; in urban areas it actually *reduces* differences in PCR! In other words, after controlling for other factors, a girl child has a higher probability of completing the primary level if she is a Muslim, than if she belongs to H-UC community. While this does not support widely held notions about gender discrimination within the Muslim community, the finding is consistent with the earlier sections and also with field observations during surveys of Kolkata slums in 2003 and 2008.

6 Alternative Explanations of the 'Residual' Component

The results of Section 5 show that a significant proportion (about 30%) of the differences in probability of completing primary level remains unexplained by household and individual

characteristics. This raises the question, why is PCR lower among Muslims, vis-à-vis other communities, *even after controlling for factors* like expenditure, household size, education and gender of family head, and age and gender of the child.

Table 6: Educational Infrastructure in Minority Concentrated and Other Districts of West Bengal (2001)

District Name	% Minority Population	Schools		Population ('000 Covered by Schools)	
		Govt Schools	Total Schools	Govt Schools	Total Schools
Non-minority dominated districts					
Puruliya	7.5	2,985	3,313	0.8	0.85
Bankura	7.7	3,914	3,914	0.8	0.82
East Medinipur	11.6	3,490	3,878	2.5	2.75
Hooghly	15.3	3,393	3,565	1.4	1.49
Jalpaiguri	16.6	2,233	2,302	1.5	1.52
Bardhaman	20.4	4,138	4,800	1.4	1.67
Kolkata	22.1	1,532	2,007	2.3	2.98
Darjeeling	22.7	1,371	1,388	1.2	1.17
West Medinipur	23.7	4,831	5,392	-	0.03
Total		27,887	30,559	1.3	1.5
Minority dominated districts					
Cooch Bihar	24.4	1,872	2,070	1.2	1.32
North 24 Parganas	24.7	4,049	4,644	1.9	2.21
Howrah	24.9	2,148	2,627	1.6	1.99
South Dinajpur	25.5	1,351	1,351	1.1	1.11
Nadia	26.1	2,787	3,063	1.5	1.65
South 24 Parganas	34.1	3,604	4,396	1.6	1.92
Birbhum	35.4	2,371	2,774	1.1	1.27
North Dinajpur	48.0	1,438	1,632	1.5	1.70
Maldah	50.0	1,909	2,221	1.5	1.72
Murshidabad	64.0	3,170	3,682	1.6	1.85
Total		24,699	28,460	1.5	1.7

Source: Estimated from Census 2001 data: Table C1 (Population by Religious Community – Blockwise) and Table on Educational Amenities in West Bengal (http://www.wbcensus.gov.in/DataTables/08/FrameTables-e_1.htm).

Two alternative explanations may be examined. Traditionally, it is argued that Muslims do not recognise the value of education, particularly mainstream education (Hunter 1869; Baig 1974; Sharma 1978; Jehangir 1991; Salamatullah 1994; Sengupta and Guha 2006). This is manifested in low rates of enrolment in mainstream schools and preference for Madrasahs (Jehangir 1991; Ruhela 1998; Salamatullah 1994; Bandhopadhyay 2002).[19] This view has been questioned in recent years (Alam and Raju 2007).[20] The High Level Committee Report (GOI 2006) has also identified the lack of an adequate number of schools and infrastructural facilities in Muslim-dominated areas as a major cause for educational backwardness of this community.

Analysis of census data reveals that population coverage of schools (population in thousands, divided by number of schools) is lower in Muslim-dominated districts (Table 6). Moreover, in Muslim-dominated areas, the availability of basic infrastructure (like benches, boards, chalk, duster, toilets) and personnel is often poor (GOI 2006).

Population-school ratio by schools, however, is not enough to refute the value theory. If Muslims really do not acknowledge the value of education, they will not enrol their children in schools. Given scarce resources, a geographical distribution of schools biased against Muslim-concentrated areas may be optimal from the utilisation point of view.[21]

Unfortunately, there is no reliable method by which we can measure demand for education, given a possible gap between actual demand for education and the *revealed* demand for education. A distinction between two possible situations becomes relevant in this context:

(a) Muslims may not be availing of educational facilities existing in this area as they feel that education is not important. In such a situation, revealed demand is an acceptable measure of actual demand.

(b) Muslims recognise the value of education, and demand more facilities than existing currently. In this case, demand revealed through enrolment figures will be much less than actual demand.

In other words, we may have a *truncated* demand curve for education when we consider enrolment-based figures.

In this situation, figures relating to indicators like percentage of schools with more than 60 students per classroom, or with student-teacher ratio exceeding 100, etc. may provide a more reliable sign of whether there is an infrastructural deficiency compared to demand for education. Estimates from district information for school education (DISE) statistics for the year 2005-06 (Table 7) reveal that Muslim-concentrated districts perform poorly relative to other districts for all these indicators.

This indicates that while population growth led to increase in demand for more educational facilities, the State government has failed to respond adequately, particularly in Muslim-concentrated districts. Consequently, the gap between demand for educational facilities and infrastructure available is greater in

Table 7: Indicators of Deficiency between Educational Infrastructure and Demand (2005-06)

District	% Schools with More Than 60 Students in a Class Room	% Schools with Student Teacher Ratio > 101	Student Teacher Ratio (in Govt Schools)	Percentage of Enrolled Childern Studying in Single Teacher Schools	Percentage of Enrolled Childern Studying in Schools without Building	% Minority Population
Murshidabad	73.6	13.15	57.2	0.1	0.3	63.92
Maldah	75.3	25.56	69.3	0.5	1.4	49.99
North Dinajpur	85.1	22.11	65.2	1.0	0.0	47.93
Birbhum	47.8	2.74	39.9	1.2	0.2	35.35
South 24 Parganas	67.9	31.91	74.1	9.3	0.4	34.06
Zone A	69.9	19.1	61.1	2.4	0.5	46.2
Nadia	58.1	3.44	47.9	0.5	0.3	26.09
South Dinajpur	45.7	6.14	50.9	1.6	0.0	25.51
Howrah	48.7	4.33	40.4	0.4	0.0	24.70
North 24 Parganas	51.9	9.21	54.3	1.1	0.4	24.63
Cooch Bihar	50.6	8.01	49.3	0.1	0.2	24.36
Darjeeling	12.2	5.97	34.3	10.9	1.3	22.63
Bardhaman	29.7	1.23	41.9	0.5	0.3	20.36
Zone B	42.4	5.5	45.6	2.2	0.4	24.0
Jalpaiguri	57.0	6.74	55.8	1.0	0.3	16.53
Hooghly	28.5	3.89	48.8	1.2	0.1	15.30
West Medinipur	24.4	3.71	41.7	2.9	0.4	13.41
East Medinipur	20.7	18.94	62.3	0.5	0.2	10.07
Bankura	28.3	1.97	40.5	3.6	0.3	7.61
Puruliya	29.3	5.39	45.1	7.8	0.9	7.43
Zone C	31.4	6.8	49.0	2.8	0.3	11.7
Kolkata	11.3	2.81	36.9	1.1	0.0	21.63

Source: Estimated from District Information for School Education, District Report Cards, 2005-06. Accessed at http://www.dise.in// on 1 October 2007.

Muslim-concentrated districts. This may have resulted in a lowering of primary completion rates of Muslims below levels warranted by their economic status and demographic characteristics.

7 Conclusions

The analysis undertaken in this paper provides useful insights into the relative status of Muslims and non-Muslim minorities in West Bengal with regard to primary educational attainments. In Section 3 we had seen that PCR of Muslims is lower than that of not only H-UC, but also of H-BCs in both rural and urban West Bengal. Statistical tests clearly show that these rates have remained significantly lower than that within H-UC and H-BC communities over time in West Bengal.[22] While Section 4 shows that such differences are due to household and personal characteristics like lower income levels, larger household size, lower educational level of household head, etc, such differences account for only about 70% of variations in probability of primary completion between Muslims and upper caste Hindus.

The relative importance of demand for education and infrastructural deficiencies are examined in section 6. Since the demand revealed through enrolment is a truncated version of actual demand for education, we look at indicators reflecting infrastructural gap vis-à-vis revealed demand. The proportion of schools in Muslim dominated districts with high student-classroom ratio and student-teacher ratio, or proportion of children enrolled in single classroom schools or schools without buildings in such areas indicate government lacunae in addressing the educational needs of Muslims. While the Central and State governments have emphasised on schemes like the merit-cum-means scholarship and on expanding madrasa education, such measures are inadequate.[23] As a result, Muslims are slipping behind other communities in education. The lack of human capital is getting reflected in their inability to shift to the formal sector and to more remunerative jobs, and is a major factor underlying the high levels of poverty (GOI 2006; Husain 2008) and overall economic backwardness found in both rural and urban West Bengal.

NOTES

1 Capability refers to the power to reflect, make choices, seek to participate in the civil society and enjoy a better life.

2 Primary education corresponds to at least four or six years of education. In this paper, we have defined primary education in terms of the first five years of education.

3 Logistic models are generally used. Akhtar (1996) uses hazard rate analysis.

4 By "explained" we mean the proportion of difference that may be attributed to differences in "endowments" – the determinants of PCR used in the regression method.

5 The small proportion of other minorities (about 1% in both rural and urban areas in West Bengal) and others (which is even lower, see Table 1) means that any meaningful analysis of these categories separately will not be possible. We have therefore clubbed them together, along with those who did not reveal their socio-religious identity, as "others". Similarly, Hindu scheduled castes, tribes, and other backward castes have been clubbed together to form a category called "Hindu backward castes" (H-BC).

6 The age group taken in this case is 12 to 15 years, as children aged 6 or slightly more have no chance of completing primary education. Assuming enrolment at 6 to 10 years and 5 years of education, a child should have completed primary education by 15 years. Incidentally, there are no failures in Government schools in the West Bengal. All students are promoted irrespective of academic performance, thereby ruling out grade repetition.

7 Alternatively, we could have taken only the number of enrolled children as the denominator. However, this implies overlooking the considerable number of children without access to schooling.

8 Religion-wise data is being collected in NSS surveys only since 1987-88.

9 The τ-value (-2.0507) of the difference between proportion of Muslims completing primary level in 2000 and 2004 is statistically significant.

10 This may have been caused by greater alienation from the mainstream following the 1992 riots, subsequent communalisation of Indian politics and increasing scepticism about the sincerity of the Left Front Government to improve the conditions of Muslims. During primary surveys of Muslim-dominated Kolkata slums (2003, 2007-08), a sense of frustration about the Left Front within the community could be perceived. These factors may have led to a reduction in *perceived* benefits to education (in terms of increased probability of securing employment).

11 It may be better to take the dependency ratio, rather than household size. However, dependency ratio will have to be derived manually – making it an extremely cumbersome, though not impossible, task. Hence we have taken household size, rather than dependency ratio.

12 The variable is therefore a continuous variable, representing illiterate (1), literate without formal schooling: EGS/NFEC/AEC (2), TLC (3), literate: below primary (5), primary (6), middle (7), secondary (8), higher secondary (10), diploma/certificate course (11), graduate (12), postgraduate and above (13).

13 For instance, Das and Mukherjee (2007, 2008) argue, based on all-India NSS data, that empowerment of women will increase enrolment.

14 Defined as SC, ST and OBC.

15 Similar findings questioning the popular belief of gender discrimination in education within Muslim communities have been observed in the Islamic country of Pakistan (Akhtar 1996).

16 Available in Mathias Sinning's homepage: http://econrsss.anu.edu.au/~sinning/Stata%20Files.html. Accessed on 17 November 2008.

17 Fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials. Available: http://ideas.repec.org/c/boc /bocode/s456727.html. Accessed on 17 November 2008.

18 Gender of child is defined as =0 for the boy child, and = 1 for the girl child.

19 Based on a study of slum-dwellers in Kolkata, Husain (2005) argued that Muslims do recognise the importance of education, but *perceived* discrimination in the labour market reduces expected employability and/or wages paid to Muslims relative to other communities. This lowers perceived returns to education, and hence *effective* demand. Interestingly, NSS data reveals that wages paid to Muslims is substantially lower than that paid to upper caste and even backward caste communities for nearly all educational levels in urban West Bengal. Only among illiterates and matriculates were differences marginal, while among postgraduate Muslims earned higher wages (Husain 2008: pp 42). Coupled with the fact that Muslims are underrepresented in the formal sector (particularly in the government sector) there seems to be prima facie grounds for this perception.

20 Analysis of NSS data reveals that parents of only 2.6% children aged 6 to 12 years believed that education was unimportant. In 70% of such cases, parents were illiterates.

21 For instance, during deliberations on the Annual Plan for 2007-08 for West Bengal, officials of the school education board argued before the State Planning Board that there was no point in setting up schools in Muslim-dominated areas as Muslim parents were not interested in sending their children to government schools. This contrasts interestingly to the finding that more than 80% of rural Muslims are enrolled in government schools.

22 The case of West Bengal is not isolated. Increasing disparities in primary completion rates between Muslims and other SRCs (and even completion rates for higher levels of education) have been observed at the all-India level and most of the major states (GOI 2006).

23 Officials of the minority affairs department, government of West Bengal, revealed, during the State Planning Board meetings to finalise allocations for 2008-09, that in 2008 about 7.5 lakh minority candidates (nearly all of whom were Muslims) applied for the pre-matric merits-cum-means scholarship. Only about 44,460 (representing less than 6% of

applicants) would be awarded scholarships, amounting to Rs 1,100 per annum. It should also be noted that the amounts will be released only after about a year from the date of application. The dubious utility of expanding madrasa education may be seen from the fact that only 4.6% of Muslim children aged 6-19 years are enrolled in such institutions (GOI 2006).

REFERENCES

Akhtar, S (1996): "Do Girls Have a Higher School Drop-out Rate Than Boys? A Hazard Rate Analysis of Evidence from a Third World City", *Urban Studies*, 33(1): 49-62.

Alam, M S and Saraswati Raju (2007): "Contextualising Inter-, Intra-Religious and Gendered Literacy and Educational Disparities in Rural Bihar", *Economic & Political Weekly,* Vol 42 (18), pp 1613-22.

Bandyopadhyay, D (2002): "Madrasa Education and the Condition of Indian Muslim", *Economic & Political Weekly*, 37(16): 1481-84.

Baig, M R A (1974): *Muslim Dilemma in India* (Delhi: Vikas Publication).

Bauer, T K and M Sinning (2008): "An Extension of the Blinder-Oaxaca Decomposition to Non-Linear Models", *Advances in Statistical Analysis*, 92: 197-206.

Birdsall, N and J L Londoño (1998): "No Trade-off: Efficient Growth via More Equal Human Capital Accumulation in Latin America" in N Birdsall, C Graham and R Sabot (ed.), *Beyond Trade-offs: Market Reform and Equitable Growth in Latin America (*Washington DC: Brookings Institution and Inter-American Development Bank).

Blinder, A S (1973): "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, 8: 436-55.

Boissiere, M (2004): "Determinants of Primary Education Outcomes in Developing Countries", Background paper for evaluation of the World Bank's support to primary education, The World Bank, Washington.

Brown, P H and A Park (2002): "Education and Poverty in Rural China", *Economics of Education Review*, 21(6): 523-41.

Bruns, B, A Mingat and R Rakotomalala (2003): "Achieving Universal Primary Education by 2015: A Chance for Every Child", The World Bank, Washington.

Connelly, R and Z Zheng (2003): "Determinants of School Enrolment and Completion of 10-18 Year Ilds in China", *Economics of Education Review*, 22(4): 379-88.

Cotton, J (1988): "On the Decomposition of Wage-Differentials", *The Review of Economics and Statistics*, 70: 236-43.

Das, S and D Mukherjee (2007): "Role of Women in Schooling and Child Labour Decision: The Case of Urban Boys in India", *Social Indicators Research*, 82(3): 463-86.

– (2008): "Role of Parental Education in Schooling and Child Labour Decision: Urban India in the Last Decade", *Social Indicators Research*, forthcoming.

Deolalikar, A (1997): "The Determinants of Primary School Enrolment and Household School Expenditures in Kenya: Do They Vary by Income?" Mimeograph, Department of Economics, University of Washington.

Desai, S and V Kulkarni (2005): "Changing Educational Inequalities in India in the Context of Affirmative Action", Mimeograph, Deparmentt of Sociology, University of Maryland.

Fairlie, Robert W (1999): "The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment", *Journal of Labour Economics*, 17(1): 80-108.

– (2003): *An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models*, Discussion Paper No 873, Economic Growth Centre, Yale University.

– (2005): "An extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models", *Journal of Economic and Social Measurement*, 30: 305-16.

Godoy, R and M Contreras (2001): "A Comparative Study of Education and Tropical Deforestation among Lowland Bolivian Amerindians: Forest Values, Environmental Externality, and School Subsidies", *Economic Development and Cultural Change*, 49(3): 555-74.

GOI (Government of India) (2006): *Social, Economic and Educational Backwardness of Muslims in India: A Report*, Report of the Prime Minister's High Level Committee, New Delhi.

Gomulka, Joanna and Nicholas Stern (1990): "The Employment of Married Women in the United Kingdom 1970-83", *Economica*, 57: 171-99.

Hadden, K and B London (1996): "Educating Girls in the Third World: The Demographic, Basic Needs, and Economic Benefits", *International Journal of Comparative Sociology*, 37:1-2.

Hunter, W W (1869): *Indian Musalmans* (New Delhi: Indological Book).

Husain, Z (2005): "Analysing Demand for Primary Education: Muslims Slum Dwellers of Kolkata", *Economic & Political Weekly,* Vol XL (2), pp 137-47.

– (2008): *Employment and Socio-Economic Status of Socio-Religious Communities in West Bengal: Some Evidence from NSS Data*, unpublished report, Dept of Minority Affairs and Madrasa Education, Government of West Bengal.

Jehangir, K N (1991): *Muslim Women in Bengal: Socio-Economic and Political Status* (Calcutta: Minerva Publishers).

National Sample Survey Organisation (1998): *Attending Educational Institutions in India – Its Level, Nature and Costs*: *NSS 52nd Round (1997-98)*, Report No 439, Department of Statistics, Government of India, New Delhi.

Neumark, D (1988): "Employers' Discriminatory Behaviour and the Estimation of Wage Discrimination", *Journal of Human Resources*, 23: 279-95.

PROBE (1999): *Public Report on Basic Education in India* (New Delhi: Oxford University Press).

Oaxaca, R L (1973): "Male-Female Wage Differentials in Urban Labour Markets", *International Economic Review*, 14: 693-709.

Oaxaca, R L and M Ransom (1994): "On Discrimination and the Decomposition of Wage Differentials", *Journal of Econometrics*, 61: 5-21.

Okumu, Ibrahim M, Nakajjo, Alex and Isoke, Doreen (2008): *Socioeconomic Determinants of Primary School Dropout: The Logistic Model Analysis.* Munich Personal Repec Archive. Accessed at: http://mpra.ub.uni-muenchen.de/view/people/Okumu,_Ibrahim_M=2E.html on 20 November 2008.

Reimers, C (1983): "Labour Market Discrimination Against Hispanic and Black Men", *The Review of Economics and Statistics*, 65: 570-79.

Ruhela, S P (1998): "Religion, Social Class and Educational Opportunity: Case Studies of Eight Muslim Girls" in S P Ruhela (ed.), *Empowerment of the Indian Muslim Women* (New Delhi: MD Publications), 1-21.

Salamatullah (1994): *Education of Muslims in Secular India* (Chandigarh: Centre for Research in Rural and Industrial Development).

Shariff, Abusaleh (1999): *Indian Human Development Report* (New Delhi: NCAER).

Sharma, K D (1978): *Education of a National Minority: A Reader in Social Class of Indian Muslim* (New Delhi: Kalamkar Prakashan).

Sen, A (1985): *Commodities and Capabilities* (Amsterdam: North Holland).

– (1999): *Development as Freedom* (New York: Knopf).

Sengupta, Piyali and Jaba Guha (2006): "Enrolment, Dropout and Grade Completion of Girl Children in West Bengal", *Economic & Political Weekly*, XXXVII (17): 1621-37.

Sinning M, M Hahn and T K Bauer (2008): "The Blinder-Oaxaca Decomposition for Non-Linear Regression Models", *The Stata Journal*, forthcoming, Accessed at http://econrsss.anu.edu.au/~sinning/files/sj.pdf on 20 November 2008.

SIS/DPP (Strategic Information Section and Division of Policy and Planning) (2005): "Levels, Trends and Determinants of Primary School Participation and Gender Parity" (New York: UNICEF).

Tansel, A (1998): "Differences in School Attainments of Boys and Girls in Turkey", Discussion Paper 789, Economic Growth Center, Yale University.

World Bank (2002): "Education and HIV/AIDS: A Window of Hope" (Washington DC: The World Bank).

Yun, Myeong-Su (2004): "Decomposing Differences in the First Moment", *Economics Letters*, 82: 275-80.

Appendix A: Results of Econometric Analysis

I. Determinants of Primary Completion

Model 1: Logistic regression: West Bengal — Number of obs = 3366; LR chi2(10) = 560.07; Prob > chi2 = 0.0000; Log likelihood = -1694.2421; Pseudo R2 = 0.1418

pcr	Coef	Std Err	z	P>|z|	[95% Conf	Interval]
lmfe	1.008517	.1164784	8.66	0.000	.7802239	1.236811
hhsize	-.1134646	.0203882	-5.57	0.000	-.1534247	-.0735045
fml	-.0573438	.0846445	-0.68	0.498	-.2232438	.1085563
age	.1283985	.0368012	3.49	0.000	.0562695	.2005274
edu_dm	.2227545	.0166581	13.37	0.000	.1901053	.2554038
sex_dm	.1965441	.1542012	1.27	0.202	-.1056847	.4987728
h-uc	.4146125	.1166139	3.56	0.000	.1860535	.6431715
h-bc	.2053415	.0988943	2.08	0.038	.0115121	.3991708
others	-.7324167	.3595497	-2.04	0.042	-1.437121	-.0277124
urban	-.0273242	.1009594	-0.27	0.787	-.225201	.1705526
_cons	-9.297479	.9831879	-9.46	0.000	-11.22449	-7.370466

Model 2: Logistic regression: Rural — Number of obs = 2347; LR chi2(9) = 400.47; Prob > chi2 = 0.0000; Log likelihood = -1233.379; Pseudo R2 = 0.1397

pcr	Coef	Std Err	z	P>|z|	[95% Conf	Interval]
lmfe	1.263685	.1485378	8.51	0.000 .	9725559	1.554813
hhsize	-.128999	.0239839	-5.38	0.000	-.1760066	-.0819913
fml	-.106916	.0988853	-1.08	0.280	-.3007276	.0868957
age	.1209443	.042918	2.82	0.005	.0368266	.205062
edu_dm	.2284959	.0198314	11.52	0.000	.189627	.2673648
sex_dm	.3182691	.1993255	1.60	0.110	-.0724016	.7089398
h-uc	.313782	.1402765	2.24	0.025	.0388452	.5887189
h-bc	.1589375	.1125664	1.41	0.158	-.0616885	.3795635
others	-.9587445	.3923839	-2.44	0.015	-1.727803	-.1896863
_cons	-11.18987	1.220716	-9.17	0.000	-13.58243	-8.797313

Model 3: Logistic regression: Urban — Number of obs = 1012; LR chi2(8) = 145.42; Prob > chi2 = 0.0000; Log likelihood = -451.98989; Pseudo R2 = 0.1386

pcr	Coef	Std Err	z	P>|z|	[95% Conf	Interval]
lmfe	.5692156	1852953	3.07	0.002	.2060435	.9323876
hhsize	-.0945059	.040426	-2.34	0.019	-.1737393	-.0152724
fml	.0710964	.167399	0.42	0.671	-.2569997	.3991925
age	.1366256	.0724641	1.89	0.059	-.0054013	.2786526
edu_dm	.2050891	.0309318	6.63	0.000	.1444639	.2657143
sex_dm	-.0641106	.2469988	-0.26	0.795	-.5482193	.4199982
h-uc	.7042587	.2183382	3.23	0.001	.2763237	1.132194
h-bc	.3642174	.2137738	1.70	0.088	-.0547715	.7832063
_cons	-5.92646	1.69557	-3.50	0.000	-9.249716	-2.603204

II. Fairlie Decomposition

Model A: Rural West Bengal — Number of obs = 17399; N of obs G=0 = 8354; N of obs G=0 = 9045; Pr(Y!=0|G=0) = .57672971; Pr(Y!=0|G=1) = .36042012; Difference = .21630959; Total explained = .15973376

pcr	Coef	Std Err	z	P>|z|	[95% Conf	Interval]
age	.0058287	.0010711	5.44	0.000	.0037294	.007928
fml	.0000444	.0001587	0.28	0.780	-.0002667	.0003555
hhsize	.0041054	.0005378	7.63	0.000	.0030514	.0051594
lmfe	.0349091	.0016079	21.71	0.000	.0317577	.0380606
sex_dm	.0002753	.0000744	3.70	0.000	.0001295	.000421
edu_dm	.1145495	.0023178	49.42	0.000	.1100067	.1190923

Model B: Urban West Bengal — Number of obs = 9575; N of obs G=0 = 7583; N of obs G=0 = 1992; Pr(Y!=0|G=0) = .71040485; Pr(Y!=0|G=1) = .45180723; Difference = .25859762; Total explained = .17277084

pcr	Coef	Std Err	z	P>|z|	[95% Conf	Interval]
age	.0260339	.00176	14.79	0.000	.0225844	.0294833
fml	-.0017305	.0002695	-6.42	0.000	-.0022586	-.0012023
hhsize	.0074341	.0021433	3.47	0.001	.0032334	.0116349
lmfe	.0148033	.0016871	8.77	0.000	.0114966	.01811
sex_dm	.0063252	.0008171	7.74	0.000	.0047236	.0079267
edu_dm	.1197523	.0035051	34.17	0.000	.1128825	.1266221

Appendix B: Oaxaca Decomposition and Extension to Non-Linear Models*

Considering the linear regression model:

$$Y_{ig} = X_{ig}\,\beta_g + \varepsilon_{ig} \qquad ...[1]$$

for I = 1, 2, 3 .... N_g, and $\Sigma_g\, N_g = N$. Blinder (1973) and Oaxaca (1973) proposes the following decomposition:

$$\Delta^{OLS} = \bar{Y}_A - \bar{Y}_B = (\bar{X}_A - \bar{X}_B)\,\hat{\beta}_A + \bar{X}_B\,(\hat{\beta}_A - \hat{\beta}_B) \qquad ...[2]$$

where $\bar{Y}_g = N_g^{-1} \Sigma_{i=1}^{Ng}\, Y_{ig}$ and $\bar{X}_g = N_g^{-1} \Sigma_{i=1}^{Ng}\, X_{ig}$. The first term on the right hand side of [2] displays the difference in outcome variable between the two groups that is due to differences in observable characteristics, whereas the second term shows the differential that is due to differences in coefficient estimates, and is referred to as an estimate of discrimination by Oaxaca.

In a subsequent paper, Oaxaca and Ransom (1994) generalised [2] as follows:

$$\bar{Y}_A - \bar{Y}_B = (\bar{X}_A - \bar{X}_B)\,\beta^* + \bar{X}_A\,(\hat{\beta}_A - \beta^*) + \bar{X}_B\,(\beta^* - \hat{\beta}_B) \qquad ...[3]$$

In equation [3], β^* is defined as the weighted average of the coefficient vectors β_A and β_B:

$$\beta^* = \Omega\,\beta_A + (I - \Omega)\,\beta_B \qquad ...[4]$$

where Ω is a weighting matrix, and I is an identity matrix.

The above decomposition, however, is not appropriate in the non-linear (NL) case. The reason is that the conditional expectations $E(Y_{ig} \mid X_{ig})$ may differ from $\bar{X}_g\,\beta_g$. In the non-linear case, therefore, [2] has to be modified. There has been several works in this context (Gomulka and Stern 1990; Yun 2004; Fairlie 1999, 2005). In Bauer and Sinning (2008) the decomposition is assumed to take the form:

$$\Delta_A^{NL} = [E_{\beta A}(Y_{iA}|X_{iA}) - E_{\beta A}(Y_{iB}|X_{iB})] + [E_{\beta A}(Y_{iB}|X_{iB}) - E_{\beta B}(Y_{iB}|X_{iB})] \qquad ...[5]$$

where $E_{\beta g}(Y_{ig}|X_{ig})$ refers to the conditional expectation of Y_{ig}, while $E_{\beta h}(Y_{ih}|X_{ih})$ refers to the conditional expectation of Y_{ih} evaluated at the parameter vector β_g, with g, h = (A, B) and g ≠ h. If B is taken to be the reference group, then the decomposition expression becomes:

$$\Delta_B^{NL} = [E_{\beta B}(Y_{iA}|X_{iA}) - E_{\beta B}(Y_{iB}|X_{iB})] + [E_{\beta A}(Y_{iA}|X_{iA}) - E_{\beta A}(Y_{iB}|X_{iB})] \qquad ...[6]$$

As earlier, the first term of the decomposition expression indicates the part of the differential in the outcome variable between the two groups that is due to differences in the covariates X_{ig}, while the second term represents the differential in I_{ig} due to coefficients.

The generalised form of the non-linear decomposition (corresponding to [4] is:

$$\begin{aligned}\Delta^{NL} = \; & [E_{\beta^*}(Y_{iA} \mid X_{iA}) - E_{\beta^*}(Y_{iB} \mid X_{iB})] + \\ & [E_{\beta A}(Y_{iA} \mid X_{iA}) - E_{\beta^*}(Y_{iA} \mid X_{iA})] + \\ & [E_{\beta^*}(Y_{iB} \mid X_{iB}) - E_{\beta B}(Y_{iB} \mid X_{iB})]\end{aligned} \qquad ...[7]$$

The two decomposition expressions, [5] and [6], represent special cases of the generalised equation [7] in which Ω is a null matrix or an identity matrix, respectively. It is also possible to interpret Ω in other ways. For instance, Remiers (1983) proposes that $\Omega = 0.5 * I$, while Cotton (1988) suggests $\Omega = sI$, when s denotes the relative sample size of the majority group. On the other hand, Neumark (1988) and Oaxaca and Ransom (1994) derives the counterfactual coefficient vector β^* based on the pooled sample.

* This section largely draws from Sinning et al (2008).