Using mixture models to detect sex bias in health outcomes in Bangladesh

Jonathan J. Morduch [a,*], Hal S. Stern [b]

[a] *Department of Economics, Harvard University, Cambridge, MA 02138, USA*
[b] *Department of Statistics, Iowa State University, Ames, IA 50011, USA*

Abstract

Many interesting economic hypotheses entail differences in behaviors of groups within a population, but analyses of pooled samples shed only partial light on underlying segmentations. Finite mixture models are considered as an alternative to methods based on pooling. Robustness checks using t-regressions and a Bayesian analogue to the likelihood ratio test for model evaluation are developed. The methodology is used to investigate pro-son bias in child health outcomes in Bangladesh. While regression analysis on the entire sample appears to wash out evidence of bias, the mixture models reveal systematic girl–boy differences in health outcomes.

Key words: Mixture model; Switching regression; Gender bias; Health production; Bangladesh
JEL classification: C51; C52; D10; I12; 012

1. Introduction

In most econometric analyses, estimation proceeds under the assumption that all sampling units face similar constraints and behave in similar ways. However, many interesting economic hypotheses entail segmentations or clusters among the sampling units in the population. For example, theories of credit and labor markets under asymmetric information yield bifurcations due to rationing, and, where it exists, gender, ethnic, or racial discrimination is unlikely to be manifested with equal force by all groups in a population. The assumption that all observations can

---

be characterized by a single model is convenient but can mask critical features of the data.

We describe the estimation of mixture models as an alternative approach to methods based on pooling. The methodology is closely related to one variant of the switching regression model (see, e.g., Quandt, 1972; Quandt and Ramsay, 1978). We apply the methodology in investigating sex-based biases in intra-household allocations in Bangladesh. As in other models of selection, identification in mixture models often rests on assumptions about the distribution of regression errors, and this requirement must be balanced against biases associated with pooling. Here, we consider issues surrounding model set-up, robustness to outliers, and methods for judging goodness of fit. Robustness checks are implemented through estimation of models in which regression errors have t-distributions with varying degrees of freedom. Since the regularity conditions which admit standard approaches to inference are not met, we develop a Bayesian analogue to the likelihood ratio test for model evaluation. These approaches are straightforward to implement and can, in themselves, provide considerable insight into the nature of the statistical relationships in the data.

As in similar poor economies, pro-son bias has been cited as a critical factor in explaining child health outcomes in Bangladesh (e.g., Chen, Huq, and D'Souza, 1980; Pitt, Rosenzweig, and Hassan, 1990; Ahmad and Morduch, 1993). However, evidence of sex bias appears far stronger in aggregate statistics than in household-level behavioral analyses. A natural hypothesis is that aggregate measures reveal bias by one segment of society but that evidence on those biases is difficult to find in pooled behavioral studies because of variability among households from different segments of the population. Here, as elsewhere, the problem is complicated by the fact that the nature of segmentation is not outwardly evident to the researcher.

Formally, characterize the behavioral relationship between D_h, a measure of the difference in average health status of sons and daughters in a household h, and the predictor variables x_h and disturbance term μ_{1h} by the regression equation

$$D_h = x_h \gamma_1 + \mu_{1h} \tag{1}$$

for one group of households. For another group the best representation might be

$$D_h = x_h \gamma_2 + \mu_{2h}. \tag{2}$$

The slope coefficients, γ_1 and γ_2, measure the effects of predictors on health status differences in the two hypothesized groups. (We could in general consider any finite number of groups.) In this specification, none of the slope coefficients, γ_1 or γ_2, should be statistically significantly different from zero in the absence of bias. Thus, the null hypothesis of no bias implies a pooled model with $\gamma_1 = \gamma_2 = 0$.

The relationship between group membership ($G_{1h} = 1$ for a household in group 1 and $G_{1h} = 0$ for a household in group 2, with $G_{2h} = 1 - G_{1h}$) and predictors z_h is characterized by a generalized linear model

$$\Pr(G_{1h} = 1) = f(z_h \delta). \tag{3}$$

The predictors z_h may overlap partially or fully with the predictors x_h. The principal difficulty in estimating the slope parameters δ is that since group membership is not observed directly, it must be inferred from observed variables.

In the present study we take $x_h = z_h$ so that there are no variables which are thought *a priori* to determine group membership (Eq. (3)) or group behavior (Eqs. (1) and (2)) but not both. It follows that identification of separate effects necessarily depends on distributional assumptions. At the minimum, however, even if these assumptions are in question, estimation of the mixture model provides a test of the robustness of the exclusion restriction $\gamma_1 = \gamma_2 = 0$. Beyond that, under the assumptions it delivers a description of key determinants of health outcomes. [1]

Even when it is suspected that populations are segmented in preferences and/or constraints, it is sometimes asserted that using standard approaches (like linear regression) with pooled data will yield consistent estimates of population-weighted average effects. Thus, the existence of sex bias should still be detectable without having to impose structure on the problem. However, this is not generally so (Appendix A, Murdoch and Stern, 1995). Estimation of a single linear model for the entire population would be subject to a form of omitted variable bias, where the omitted variables are an indicator of group membership and possible interactions between group membership and other predictors of health outcome differences. When the population consists of different subpopulations, OLS regression provides an unbiased estimate of a matrix-weighted average of the subpopulation regression coefficient vectors. But in a multivariate regression, it is possible for the weighted average of a single coefficient to lie outside the range of the subpopulation coefficients (Chamberlain and Leamer, 1976); the estimated coefficient may even differ in sign from both subpopulation coefficients. Pooled data will yield consistent estimates only when the determinants of group membership, z_h, are uncorrelated with the behavioral determinants of health outcomes, x_h – as in standard analyses of selection bias (Heckman, 1990). Even then, the regression standard error is not estimated consistently, and the

---

[1] Related problems of identification in selectivity models are discussed by Heckman (1990) and Manski (1990). Since here the classification is based on an unobserved variable, identification of the causal effect is even more difficult than in standard selection problems – e.g., the effect of unionization in Heckman (1990) and Lewis (1986) or female labor force participation in Newey, Powell, and Walker (1990) and Mroz (1987). Recent work in standard selection problems attempts to relax distributional assumptions through semi-parametric methods. Lee and Porter (1984) address estimation when sample separation information is available but imperfect.

upward bias leads to a significant loss of power for detecting nonzero slopes. With large samples, generalized heteroscedasticity correction methods might be used to address the nonconstant variance that is a symptom of a mixture of regressions, but they do not directly address the presence of subpopulations. In small samples, mixture models have the advantage of putting structure on the nature of heterogeneity and thus offering the possibility of greater precision in estimates.

In the present case, under our distributional assumptions, we find that, unlike the results from the pooled regression, evidence of pro-son bias is clear in the slope estimates obtained under the mixture model, suggesting that unobserved heterogeneity helps to explain why so little evidence on bias has been found in pooled regressions. Our relatively small sample, however, limits the precision of the estimates.

In the next section we describe recent evidence on health outcomes of children in Bangladesh. Section 3 describes the mixture model. In Section 4, we describe our results and tests for robustness using t-regressions, and in Section 5 a Bayesian approach to inference about the number of subpopulations is developed and applied.

2. Gender-based child health outcomes in Bangladesh

We focus attention on differences in the health of male and female children as reported in the 1988–89 Child Nutrition Survey (CNS) conducted by the Bangladesh Bureau of Statistics. A variety of household-level socio-economic data were collected, as well as anthropometric data and information on health practices for each child between the ages of six months and six years. The survey covers 1555 households, which are drawn from a re-survey of 50% of the random sample initially included in the 1988 Household Expenditure Survey. [2]

Evidence of gender bias in mortality patterns is clear in the survey. There are 107 males for every 100 females age three and below. Between the ages of four and six years, the ratio rises to 113 to 100. This finding suggests substantially higher mortality rates for girls than for boys (a phenomenon which counters the biological tendency toward parity). While under-counting of females can also explain some of the difference, under-counting is likely to be minor given the nature of the survey. Moreover, this evidence is consistent with previous studies which have noted sex bias in health outcomes and treatment in Bangladesh.

---

[2] Given that households were less likely to be surveyed if children died or fell ill between the Household Expenditure Survey and the CNS, there is a selection bias which will likely work against finding evidence of discrimination.

Clearly, gender seems to have a role in explaining mortality patterns and some acute health outcomes. However, when we turn to econometric analyses of the Bangladesh Child Nutrition Survey, the evidence of bias is not clear. We consider three anthropometric measures of health status: height-for-age (a measure of chronic malnutrition), weight-for-height (a measure of current and acute malnutrition), and weight-for-age (a second measure of principally short-term deprivation).[3] Since our definition of the dependent variable, D_h, is the difference between the logarithm of the mean health status for male children and the logarithm of the mean health status for female children, we restrict attention to 307 households with at least one male and one female child. Across the 307 households, the mean D_h for height-for-age is 0.0067 with a standard error of 0.0042, the mean difference for weight-for-height is −0.0008 with a standard error of 0.0061, and the mean difference for weight-for-age is 0.011 with a standard error of 0.0085. Males do appear to have an advantage of roughly 1% as measured by height-for-age and weight-for-age although there is considerable variability across households.

We first investigate the correlates of gender differences in health outcomes using pooled regressions, estimating Eq. (1) under the assumption that the coefficients are constant for the entire sample.[4] The dependent variable D_h controls for fixed effects like genetic and environmental factors by differencing male and female averages within each household. Such factors may be critical in explaining child health outcomes. For example, tall parents are more likely to have tall children, and this should be controlled to reduce bias. The disturbances μ_h are heteroscedastic, with variance depending on the number of male and female children, and this is addressed using weighted least squares (see Section 3). As above, none of the coefficients will be significantly different from zero in the absence of bias.

The basic regression results are shown in Table 1. The independent variables are among those household, community, and child-specific variables most likely to be related to discrimination, and all continuous variables have been centered at their means. Thus, the constant will reflect the average advantage of boys over girls in households with zero values for all dummy variables and mean values for all continuous variables; the other coefficients will reflect ways in which boys receive different treatment than girls.

In the three equations in Table 1, the intercepts are all positive but small and not significant, suggesting slight or no advantages for boys over girls as a base-line. The Hindu dummy, the log of family size, and the dummy variable

---

[3] The reference standards for the anthropometric measure are given by the United States National Center for Health Statistics, as employed by the Bangladesh Bureau of Statistics and used frequently in international comparisons.

[4] Useful discussions of issues raised in estimating health relationships are found in Strauss and Thomas (1995) and Bhargava (1994).

Table 1
Pooled sample: Weighted regression estimates of male–female differences in three anthropometric measures (estimates and standard errors in parentheses obtained via maximum likelihood)

Independent variables	Dependent variable: Male average less female average		
	Logarithm of height-for-age	Logarithm of weight-for-age	Logarithm of weight-for-height
Log age of woman head of household	−0.015 (0.021)	−0.037 (0.042)	0.002 (0.030)
Log per capita income	−0.007 (0.010)	−0.004 (0.021)	0.002 (0.015)
Hindu dummy variable	−0.005 (0.016)	−0.010 (0.032)	−0.006 (0.023)
Rural dummy variable	0.005 (0.012)	0.007 (0.023)	−0.009 (0.017)
Log household size	0.008 (0.014)	0.015 (0.029)	0.001 (0.021)
Mother has primary education only	−0.007 (0.011)	0.005 (0.022)	0.024 (0.016)
Mother has secondary education or more	−0.019 (0.015)	−0.042 (0.031)	−0.007 (0.023)
Log distance to medical facility	−0.006 (0.005)	−0.021 (0.009)	−0.012 (0.007)
Log distance to regional center	0.009 (0.006)	0.016 (0.012)	0.005 (0.009)
Girl is oldest dummy variable	0.012 (0.008)	−0.021 (0.017)	−0.017 (0.012)
Intercept	0.0006 (0.010)	0.023 (0.020)	0.011 (0.014)
Log likelihood	657.3	439.3	538.6
Number of households	307	307	307
Regression standard error	0.073	0.149	0.108

for secondary education of the head woman take the expected signs in all three regressions but none are significant. Interestingly, the coefficient of distance from the medical center is consistently negative, and it is significant in the equation predicting weight-for-age. Being further from a health facility lowers the

advantage sons have over daughters. Distance from the regional center apparently increases the advantage.[5] However, the general fit of the regressions is poor; the R^2 is 0.03 for each regression. The F-statistics $(10, 296)$ are 1.026, 0.980, and 0.983 for height-for-age, weight-for-age, and weight-for-height, respectively, with resulting p-values of 0.42, 0.46, and 0.46. Thus, overall, the absence of bias cannot be rejected for any of the three health measures.

There are a number of explanations for these weak results on discrimination. First, there may in fact be very little bias in health outcomes, perhaps because the most disadvantaged children are the least likely to survive to show up in the sample. Second, the regression specification may not be flexible enough to capture evidence of bias.[6] Third, measurement error (possibly exacerbated by the use of a dependent variable which is the difference of two random outcomes) may be large enough that there is little power for detecting nonzero coefficients. Below, however, we focus on the potential consequences of heterogeneity. To the extent that the alternative explanations have force, findings of biases using mixture models will be strengthened.

3. Mixture model

We have hypothesized that the behavioral relationship between the difference in average male and female health outcomes in a household is described by Eqs. (1) or (2), $D_h = x_h \gamma_i + \mu_{ih}$, where $i = 1, 2$ indexes subpopulations and $h = 1, \dots, n$ indexes households. More generally, the population may be thought to consist of a mixture of M subpopulations each with its own set of regression coefficients. Below we assume $M = 2$, but the model could be extended so that M is treated as a parameter and estimated along with other model parameters. Alternatively, several M's could be used and the resulting model fits compared (Rubin and Stern, 1994). Here, a mixture of three groups seems to provide only a small increase in the likelihood. Each subpopulation is described by a standard normal linear model, and the variance parameter σ^2 is assumed to be the same in the two

---

[5]The distance to the medical and regional centers will raise the male–female health gap if, by increasing the price of treatment, girls do not receive adequate care. However, if the distance raises the price so that boys also receive inadequate care, it could lower the gap. Note that the two distance variables are highly correlated, which may explain why they take opposite signs in the equations. Ahmad and Morduch (1993) discuss the other predictor variables in greater detail.

[6]For example, pro-son bias may not be a general phenomenon. Instead, it may be that girls born earlier receive better treatment than girls born later (Das Gupta, 1987) or girls with brothers may be treated better than those with sisters (Muhuri and Preston, 1991). However, neither of these explanations appears to have much power in this data (Ahmad and Morduch, 1993). Bardhan (1984) argues that the strength of the health sector can be critical in determining child health outcomes, and he critically addresses economic and social explanations of sex bias in South Asia.

subpopulations. Separate variances in each group are permissible but considerably complicate estimation – we address this point below.

A central feature of the model is that the composition of subpopulations cannot be explicitly identified *a priori*. Let $\lambda_{ih} = \Pr(G_{ih} = 1)$ denote the probability that household h belongs to subpopulation group i. In this application the model for group membership (Eq. (3)) is taken to be a logistic regression model with predictors x_h just as in the linear model. The logistic regression model is defined in terms of λ_{1h},

$$\log\left(\frac{\lambda_{1h}}{1 - \lambda_{1h}}\right) = x_h\delta, \tag{4}$$

with $\lambda_{2h} = 1 - \lambda_{1h}$. Such a model might result from considering group membership to be determined by a continuous measure $I_h = x_h\delta + v_h$ where v_h has the logistic distribution and is assumed independent of the linear model disturbances μ_h. Under these assumptions the mixture model likelihood is given by

$$L = \prod_h \left[\sum_{i=1}^{2} \lambda_{ih} \frac{1}{\sqrt{2\pi}\sigma_h} \exp\left(-0.5\frac{(D_h - x_h\gamma_i)^2}{\sigma_h^2}\right)\right], \tag{5}$$

where the dependence of λ_{ih} on x_h and δ is suppressed.

Heteroscedasticity arises naturally in the definition of D_h as the difference between average male and average female health outcomes. Define σ^2 to be the residual variance in households with one male child and one female child. If n_{hb} and n_{hg} are the number of boys and girls respectively in household h, then an approximation to the variance associated with the residual in the hth household is $\sigma_h^2 = 0.5\sigma^2(1/n_{hg} + 1/n_{hb})$. The corresponding weight for the hth household is $w_h = 2n_{hg}n_{hb}/(n_{hg} + n_{hb})$. This variance is approximate because it assumes the disturbances specific to individual children are independent when any gender bias is likely to induce some correlation.

A pooled regression using individual children as the unit of analysis but with household fixed effects would provide an alternative means of addressing heterogeneity. However, it only addresses household-level heterogeneity in levels of child health outcomes, not in regression coefficients detailing the size of gender bias and the effects of covariates. Using only cross-sectional data, it is not possible to estimate separate household-specific regression slopes because only one value of the household level predictor is observed (see the overview on variable-coefficient models in Chapter 6 of Hsiao, 1986). The mixture model approach provides a compromise in that heterogeneity in both intercepts and variable coefficients is addressed by assuming a discrete number of subsamples in the population, each with its own behavioral responses and constraints, and by removing household fixed effects through the definition of the dependent variable.

The survey provides information on 1206 individual children, 675 of these children reside in households having both male and female children in the age

range of interest. There would seem to be a loss of efficiency in restricting atten-
tion to the 307 households in place of the 675 children. We considered a mixture
model applied directly to the 675 children in these households where standard
fixed-effect methods are used to estimate separate household effects in each of
the two mixture components. However, this requires over 600 parameters since
$M = 2$ fixed effects must be estimated for each household, leaving few degrees
of freedom for assessing variability. Random-effects models would replace the
household terms in the model by an assumed normal population of household
effects with possibly different means and variances in the two subpopulations.

4. Results

After allowing for the possibility that there may be several internally homo-
geneous groups in the sample, the sample divides into two clear groups for the
equations which describe the correlates of height-for-age. Here, the largest group
includes 72% of the sample, leaving 85 of 307 households in the smaller group.
The evidence, however, is not strong for a mixture of groups in investigating
the two measures of shorter-term health: weight-for-height and weight-for-age. In
the case of weight-for-age only a small increase in the likelihood is obtained by
allowing for two subpopulations. In the case of weight-for-height, the maximum
likelihood solution obtains a small group (25 households) and a large group. A
larger sample than the 307 households would be needed for accurate inference
for a subpopulation thought to consist of less than 10% of the population, es-
pecially given the large number of predictors. Section 5 addresses methods for
assessing the increase in the likelihood and whether these groupings are likely to
have occurred just by chance clustering of the data.

The results for the height-for-age model are presented in Table 2. The first
column repeats the pooled regression results of Table 1. The second and third
columns correspond to the special restricted model in which no predictors are
assumed available for identifying group membership (these results are discussed
later). We focus on the results of the full model, appearing in the final three
columns of Table 2, which incorporates separate linear regressions within each of
two subpopulations and a logistic regression for explaining subpopulation mem-
bership. The estimates in Table 2, obtained using the EM algorithm of Dempster,
Laird, and Rubin (1977), are those obtaining the highest mode of the likelihood
function (see the Appendix).

In this section, we concentrate on interpreting the estimated model parameters,
deferring tests of the overall model until the final section. The centered intercepts
give the advantage of boys over girls in households with zero values for all
dummy variables and mean values for all continuous variables. Thus, in the
smaller group boys have a 7% advantage over girls in urban, Muslim, uneducated
households with younger girls. The larger group shows the reverse, girls have a

Table 2
Pooled model versus two class (normal) mixture model: Estimates of male–female differences in height-for-age (approximate asymptotic standard errors in parentheses)

Independent variables	Dependent variable: Male average less female average logarithm of height-for-age					
	Pooled sample	Restricted		Full mixture		
		Group 1	Group 2	Group 1	Group 2	Logit
Log age of woman head of household	−0.015 (0.021)	0.038 (0.042)	−0.032 (0.023)	0.008 (0.037)	−0.004 (0.025)	−1.05 (1.59)
Log per capita income	−0.007 (0.010)	0.003 (0.021)	−0.009 (0.012)	0.030 (0.018)	−0.008 (0.012)	−1.34 (0.87)
Hindu dummy variable	−0.005 (0.016)	−0.014 (0.031)	−0.007 (0.017)	−0.016 (0.023)	−0.039 (0.020)	2.16 (0.98)
Rural dummy variable	0.005 (0.012)	−0.063 (0.031)	0.024 (0.013)	0.034 (0.022)	0.044 (0.015)	−2.81 (0.96)
Log household size	0.008 (0.014)	−0.019 (0.033)	0.022 (0.017)	−0.140 (0.035)	−0.043 (0.021)	6.00 (1.47)
Mother has primary education only	−0.007 (0.011)	−0.064 (0.025)	0.013 (0.012)	0.023 (0.043)	0.027 (0.010)	−2.97 (1.76)
Mother has secondary education or more	−0.019 (0.015)	−0.097 (0.047)	0.005 (0.021)	−0.004 (0.047)	0.019 (0.015)	−3.03 (2.04)
Log distance to medical facility	−0.006 (0.005)	0.023 (0.010)	−0.016 (0.006)	0.018 (0.009)	−0.012 (0.005)	−0.39 (0.41)
Log distance to regional center	0.009 (0.006)	0.038 (0.015)	0.003 (0.007)	−0.008 (0.010)	0.003 (0.007)	1.05 (0.60)
Girl is oldest dummy variable	0.012 (0.008)	0.048 (0.020)	0.003 (0.009)	0.038 (0.017)	0.010 (0.009)	−0.46 (0.60)
Intercept	0.0006 (0.010)	0.085 (0.022)	−0.028 (0.011)	0.069 (0.017)	−0.059 (0.015)	0.73 (0.66)
Log likelihood	657.3	672.0		681.1		
Proportion of sample	1.0	0.26	0.74	0.28	0.72	
Number of households	307	80.3	226.7	85.5	221.5	
Regression standard error	0.073	0.059	0.059	0.058	0.058	

6% advantage over boys. Both findings are significant; it is somewhat surprising that the larger group has a pro-daughter advantage.

` In the smaller group, the advantages of males are highly sensitive to changes distance from the health center, whether the oldest child is a girl, and the size

of household. There, being further from the health center tends to increase the advantage to boys. Surprisingly, larger households tend to decrease the advantage to boys and having a female oldest child increases the advantage. The opposite signs were expected. In the pooled regression no variables have significant coefficients. In the larger group, distance to the health center appears to increase the advantage of females. Rural location and primary education of the mother tend to reduce the advantage of females. Again, larger households tend to favor females more.

In the logit equation, the most important determinant of membership is household size. Larger households are predominantly in the smaller group (that tends to favor boys). Within that group, recall that additional household size is associated with smaller advantages for male children. Also, Hindu households tend be in the smaller group while rural households tend to be in the larger group.

To assess robustness to outliers, we re-estimated the models under the assumption that regression errors have t-distributions, as shown in Table 3. The t-distributions with 4 and 10 degrees of freedom have relatively long tails, so that outlying observations get less weight. No standard errors are supplied as the loglikelihood of the t-regressions are relatively flat so that standard errors are difficult to obtain (and generally too small) in small samples. The results for the t-distribution with 10 d.f. are quite similar to those for the normal regressions. Coefficients differ by at most one-quarter of the normal regression standard error. The results with 4 d.f. indicate some larger differences, especially in the smaller group. The coefficients of age of the female household head and the dummy for education in the smaller group change signs. There are relatively large changes in the logistic regression coefficients, especially those for the education variables and the rural location dummy variable. The results in Tables 2 and 3 suggest that the distributional assumptions are not innocuous. However, the principal findings concerning gender differences in the two subpopulations remain the same. The nonrobustness of the logistic regression is a first indication that the current data set may not be sufficient for estimating both the between-group effects and the within-group effects of the predictor variables.

5. Assessing the mixture model

One chief concern is that, lacking an explicit identifier of group identity, the mixture model may yield groups which occur simply by chance in the data. For example, when the estimation procedure divides the sample into one large nebulous group and one precisely-estimated group with a small number of observations, it could reflect a random clustering rather than true underlying structure. We develop and apply a Bayesian approach to inference which addresses these issues in an analogue to standard loglikelihood ratio tests.

Table 3
Two-class mixture model under *t*-distribution: Estimates of male–female differences in height-for-age

Independent variables	Dependent variable: Male average less female average logarithm of height-for-age					
	4 degrees of freedom			10 degrees of freedom		
	Group 1	Group 2	Logit	Group 1	Group 2	Logit
Log age of woman head of household	−0.027	−0.015	0.02	−0.005	−0.007	−0.70
Log per capita income	0.038	−0.010	−1.59	0.032	−0.011	−1.26
Hindu dummy variable	−0.015	−0.040	1.91	−0.014	−0.040	2.10
Rural dummy variable	0.021	0.033	−1.49	0.032	0.042	−2.60
Log household size	−0.107	−0.064	6.68	−0.132	−0.044	5.93
Mother has primary education only	−0.009	0.024	−0.82	0.018	0.031	−2.75
Mother has secondary education or more	−0.009	0.017	−1.29	−0.002	0.022	−3.05
Log distance to medical facility	0.008	−0.015	0.06	0.015	−0.013	−0.30
Log distance to regional center	−0.004	0.012	0.11	−0.006	0.005	0.87
Girl is oldest dummy variable	0.031	0.022	−0.84	0.038	0.015	−0.55
Intercept	0.067	−0.065	0.24	0.068	−0.061	0.67
Log likelihood	701.9			690.3		
Proportion of sample	0.34	0.66		0.29	0.71	
Number of households	104.5	202.5		87.5	219.5	
Regression scale parameter	0.044	0.044		0.052	0.052	

   While standard methods like maximum likelihood estimation can be used to obtain parameter estimates, it is not appropriate to test the mixture model against the pooled-regression null hypothesis by applying the usual asymptotic chi-square test based on twice the difference in maximum loglikelihood values. The principal reason for this is that under the null hypothesis (that the population is homo-

geneous and a single pooled regression is appropriate) the two component mixture model likelihood is not well approximated by a quadratic surface as required for the likelihood ratio test (see, e.g., Titterington, Smith, and Makov, 1985). Suggestions for approximate degrees of freedom to apply as a correction are found in the statistics and economics literature (Goldfeld and Quandt, 1976; Titterington, Smith, and Makov, 1985). Instead, we follow Rubin and Stern (1994) in implementing a Bayesian posterior predictive evaluation of the propriety of the pooled regression relative to that of the mixture.

The posterior predictive check asks whether the observed improvement in moving from a single regression to the two-component mixture is large relative to the improvement that might be observed in future similar data sets. Of course, this is exactly the question that the likelihood ratio chi-square addresses when the regularity conditions are met. In the present case, a set of 200 replicate data sets are obtained as draws from a Bayesian posterior predictive distribution under the null model, a single pooled regression. We use a standard noninformative prior distribution for the coefficients of the regression γ and regression variance σ^2 (corresponding to families with one male and one female child) in the Bayesian analysis, $p(\gamma, \sigma^2) \propto \sigma^{-2}$. Only a brief summary of the steps required for the Bayesian analysis is presented here; the methods are described more fully in Box and Tiao (1973). Under the assumption that a single pooled weighted least squares regression is appropriate and using the noninformative prior distribution, the posterior distribution of σ^2 can be expressed as

$$\sigma^2 | D \sim \frac{(D - X\hat{\gamma})^t W (D - X\hat{\gamma})}{\chi^2_{296}}, \tag{6}$$

where $\hat{\gamma}$ is the weighted least squares estimate of γ, D is the vector of health outcomes, X is the matrix of predictor variables, W is a diagonal matrix containing the household weights, and χ^2_{296} refers to a random variable with a chi-square distribution. Given σ^2 the conditional posterior distribution of γ is

$$\gamma | \sigma^2, D \sim N(\hat{\gamma}, \sigma^2 (X^t W X)^{-1}). \tag{7}$$

We generate hypothetical replications of the data D by first drawing σ^2 from its posterior distribution, then drawing a simulation of γ from its posterior distribution conditional on σ^2, and finally drawing a vector of new responses $D^{(\text{rep})} = X\gamma + \mu$ where the elements of μ are normally distributed disturbances with variance matrix $\sigma^2 W^{-1}$. The vectors $D^{(\text{rep})}$ represent values of the response variable that might be expected in new observations at the same predictors X if the single pooled regression is correct. For each of the 200 replicate data sets, we fit the pooled regression and the mixture model. The replicate data sets define a reference set for determining whether the observed increase in the likelihood is large relative to what might occur by chance.

In the Bangladesh data, the loglikelihood for the pooled regression is 657.3 and the loglikelihood for a two-class mixture is 681.1 – an increase of 23.8.

For the 200 replicated data sets the average increase just due to chance is 25.6, so the data do not support the need for the mixture model. (Note also that the increase in the loglikelihood is larger than would be expected under the usual chi-squared reference distribution: with 22 degrees of freedom the loglikelihood would be expected to increase by only 11 – twice the loglikelihood would be expected to increase by 22.)

We also fit a mixture model under the assumption that $\lambda_{hi} = \lambda_i$ for all households, i.e., no predictors of group membership are available. Parameter estimates for this restricted mixture model are displayed in Table 2; the loglikelihood is 672.0, an increase of 14.7 over the pooled regression. This is larger than would be expected by chance – only 8% of the 200 replicate data sets had increases as large. The data seem to support this restricted mixture as an alternative to the null, pooled model but do not support the seemingly more realistic logit mixture model. The simpler model is less interesting in that the pooled regression gives unbiased estimates of the average effect in this case, but the restricted model still shows that heterogeneity does play a role in masking important gender differences in health outcomes. The results in the restricted model are consistent with those of the full mixture with some noteworthy differences. As a baseline under the restricted model, boys maintain a large advantage (8.5%) in the smaller 'more discriminatory' group (a quarter of the sample), while girls have a small (less than 3%) advantage in the larger group. The unexpected coefficients on household size are no longer significant. Distance to the medical center remains an important determinant of health outcomes in both models. Moreover, both female education variables strongly reduce pro-son bias in the smaller group, with secondary education having a 50% greater impact than primary education alone.

There is a bit of paradox in comparing the three models: pooled, restricted mixture, and logit mixture. Comparisons with the pooled model suggest that the restricted mixture model is helpful but that the logit mixture model does not add much beyond this. The increase in the loglikelihood between the restricted and logit models is smaller than, but similar to, the increase obtained in moving from the pooled model to the restricted mixture. Thus, direct comparison of the two mixture models provides some support for the logit model. A Bayesian analogue for the this last test would provide more insight but would require a full Bayesian analysis of the mixture model.

6. Concluding comments

The mixture model provides an empirical framework which is consistent with theoretically- and empirically-based concerns about population heterogeneity with regard to gender-based differences in health outcomes. When populations are heterogeneous, running regressions on pooled samples will only yield consistent estimates of average slope coefficients under special conditions – and then it is

likely that uncorrected standard errors will be biased upward (so that the null hypothesis of no discrimination is not rejected frequently enough). With a large sample, generalized heteroscedasticity-consistent estimators can be used to reduce this problem, but the mixture model can yield greater precision of estimates by putting specific structure on the sources of heterogeneity.

While it remains that specific distributional assumptions are required to provide an identifiable model of the hypothesized heterogeneous population, we have demonstrated how the robustness of the mixture model results can be assessed by considering alternative distributional assumptions and model structures. While equivocal, the results in the present data set suggest that addressing behavioral heterogeneity is an important approach to shedding more light on differences in health outcomes in Bangladesh and similar poor economies. Pursuing this path is practical, and the Bayesian approach to model evaluation developed here straightforwardly follows the intuition of standard likelihood ratio tests. However, our results caution that large samples may be needed for precise estimation.

Appendix: Estimation procedure

Maximum likelihood estimates for the model parameters can be computed using any maximization procedure. The calculations are especially straightforward using the EM algorithm of Dempster, Laird, and Rubin (1977) for computing maximum likelihood estimates for problems with incomplete data. The data is incomplete here in the sense that the indicators of group membership, $G_h = (G_{1h}, G_{2h})$, are unobserved. The joint distribution of the full data set D_h, G_h is

$$L = \prod_h \prod_{i=1}^{2} \left(\lambda_{ih} \frac{1}{\sqrt{2\pi}\sigma_h} \exp\left(-0.5 \frac{(D_h - x_h\gamma_i)^2}{\sigma_h^2} \right) \right)^{G_{ih}}. \tag{8}$$

The EM algorithm is an iterative maximization procedure with each iteration consisting of two steps, the E-step and the M-step. We introduce the superscript (t) to describe the tth iteration and assume that parameter estimates $\gamma_1^{(t-1)}, \gamma_2^{(t-1)}, \sigma_h^{(t-1)}$, $\delta^{(t-1)}$ obtained during the previous iteration's M-step are available. The E-step computes the expected value of the complete data sufficient statistics which is equivalent in the current case to finding

$$\lambda_{ih}^{(t)} = \mathrm{E}(G_{ih}|D_h, x_h, \gamma_i^{(t-1)}, \sigma_h^{(t-1)}, \delta^{(t-1)})$$
$$= \Pr(G_{ih} = 1 \mid D_h, x_h, \gamma_i^{(t-1)}, \sigma_h^{(t-1)}, \delta^{(t-1)}) \tag{9}$$
$$= \frac{\lambda_{ih}^{(t-1)} \phi(D_h; x_h\gamma_i^{(t-1)}, \sigma_h^{2(t-1)})}{\lambda_{1h}^{(t-1)} \phi(D_h; x_h\gamma_1^{(t-1)}, \sigma_h^{2(t-1)}) + \lambda_{2h}^{(t-1)} \phi(D_h; x_h\gamma_2^{(t-1)}, \sigma_h^{2(t-1)})},$$

where $\phi(x; \mu, \tau^2)$ is the normal density with mean μ and variance τ^2 evaluated at x. The result $\lambda_{ih}^{(t)}$ represents the posterior probability of household h coming

from subpopulation i, and it is not generally equal to 0 or 1 since households might conceivably belong to either subpopulation. In the M-step of the tth iteration the complete data likelihood, Eq. (8), is maximized after fixing G_{ih} at the expected value computed during the E-step. This M-step maximization is achieved by separately maximizing the likelihood of the classification portion of the model, describing the distribution of G_h given x_h, and the likelihood of the two normal regression models, describing the distribution of D_h given x_h and G_h. The classification portion of the likelihood resembles the likelihood from a logistic regression except that the classification variable G_{ih} is not restricted to the values 0 and 1. In the regression likelihoods, each household can contribute to the regression estimates of both mixture components since G_{ih} is not restricted to the values of 0 and 1. Each mixture component requires a weighted least squares regression that now incorporates the weight $\Pr(G_{ih} = 1)$ as well as the weights that adjust for household heteroscedasticity. The E- and M-steps are repeated until further iterations do not modify the estimates. Approximate asymptotic standard errors can be obtained using an extension of the EM algorithm (Meng and Rubin, 1991), or by numerically obtaining the second derivative matrix of the loglikelihood function and inverting it as was done here.

To assess the robustness of model inferences, the models have also been estimated under the assumption that the regression disturbances have a t-distribution with 4, 10, 20, and 100 degrees of freedom (the last is quite close to the model with normally distributed errors). The likelihood for the mixture model with t-distributed regression errors having v degrees of freedom is similar to the normal likelihood equation (5) with normal densities replaced by t_v densities. Following Liu (1994) and Dempster, Laird, and Rubin (1977) the t-regression part of the likelihood maximization is carried out using iteratively reweighted least squares. Specifically, the distribution of D_h for group i is taken to be $N(x_h\gamma_i, \sigma_h^2/\tau_{ih})$ where τ_{ih} has a gamma distribution with shape and scale parameter $v/2$. The τ_{ih} are treated as missing data. Each E-step now also computes the expected value of τ_{ih} as

$$E(\tau_{ih}|D_h, x_h, \gamma_i^{(t-1)}, \sigma_h^{(t-1)}, \delta^{(t-1)}) = \frac{v+1}{v + ((D_h - x_h\gamma_i^{(t-1)})^2/\sigma_h^{2(t-1)})}, \qquad (10)$$

in addition to computing the expected value of the G_{ih} according to Eq. (9). The M-step computes normal regressions in each group, now weighted by the heteroscedasticity correction, the probability of group membership, and the t-weight τ_{ih}. The t-weight serves to downweight outlying or unusual values, taking us in the direction of least absolute deviation estimators.

In practice, mixture model likelihood functions may have more than one mode, corresponding to different descriptions of the underlying population. Only the highest peak of the likelihood function is explored in the discussion in the text. Large samples or less complex models appear to be less likely to have multiple modes. For this analysis, fitting a restricted mixture model (without the logistic

regression but with residual variances in the two mixture components constrained to be equal) yields only a single minor mode in addition to the major mode given in Table 2. The full mixture model, however, yields a number of solutions to the likelihood equation. In small samples, problems of multiple modes might be addressed by constructing the model to incorporate appropriate prior information about plausible sizes of the subpopulations or plausible values of the residual variance. (Kiefer, 1978, shows that for large enough n there is a unique consistent root to the likelihood equation in the restricted model – i.e., where $\lambda_{ih} = \lambda_i$.) More generally, a full Bayesian analysis, averaging over the different modes, might more accurately summarize the data but is computationally difficult given current tools.

References

Ahmad, A. and J.J. Morduch, 1993, Identifying sex bias in the allocation of household resources: Evidence from linked household surveys from Bangladesh, Harvard Institute of Economic Research discussion paper no. 1636, April (Harvard University, Cambridge, MA).

Bangladesh Bureau of Statistics, 1991, Report of the child nutrition survey 1989–90 (Ministry of Planning, Dhaka).

Bardhan, P.K., 1984, On life and death questions: Poverty and child mortality in rural India, Ch. 15 in: Pranab Bardhan, Land, labor, and rural poverty: Essays in development economics (Oxford University Press, Delhi).

Bhargava, A., 1994, Modelling the health of Filipino children, Journal of the Royal Statistical Society A 157, part 3, 417–432.

Box, G.E.P. and G.C. Tiao, 1973, Bayesian inference in statistical analysis (Addison-Wesley, Reading, MA).

Chamberlain, G. and E.E. Leamer, 1976, Matrix weighted averages and posterior bounds, Journal of the Royal Statistical Society B 38, 73–84.

Chen, L.C., E. Huq, and S. D'Souza, 1980, A study of sex biased behaviour in the intra-family allocation of food and the utilization of health care services in rural Bangladesh (International Centre for Diarrhoeal Disease Research, Dhaka).

Das Gupta, M., 1987, Selective discrimination against female children in rural Punjab, India, Population and Development Review 13, 77–100.

Deaton, A., 1989, Looking for boy–girl discrimination in household expenditure data, World Bank Economic Review 3, 1–15.

Dempster, A.P., N.M. Laird, and D.B. Rubin, 1977, Maximum likelihood from incomplete data via the EM algorithm (with discussion), Journal of the Royal Statistical Society B 39, 1–38.

Dickens, W.T. and K. Lang, 1985, A test of dual labor market theory, American Economic Review 75, 792–805.

Goldfeld, S.M. and R.E. Quandt, 1976, Techniques for estimating switching regressions, in: S.M. Goldfeld and R.E. Quandt, Studies in non-linear estimation (Ballinger, Cambridge, MA).

Heckman, J.J., 1990, Varieties of selection bias, American Economic Review: Papers and Proceedings 80, 313–318.

Hsiao, C., 1986, Analysis of panel data, Econometric Society monographs no. 11 (Cambridge University Press, Cambridge).

Kiefer, N.M., 1978, Discrete parameter variation: Efficient estimation of a switching regression model, Econometrica 46, 427–434.

Lee, L.-F. and R. Porter, 1984, Switching regression models with imperfect sample separation information, with an application to cartel stability, Econometrica 52, 391–418.

Lewis, H.G., 1986, Union relative wage effects: A survey (University of Chicago Press, Chicago, IL).

Liu, C., 1994, Statistical analysis using the multivariate t distribution, Unpublished Ph.D. dissertation (Department of Statistics, Harvard University, Cambridge, MA).

Manski, C.F., 1990, Non-parametric bounds on treatment effects, American Economic Review: Papers and Proceedings 80, 319–323.

Meng, X.L. and D.B. Rubin, 1991, Using EM to obtain asymptotic variance–covariance matrices: The SEM algorithm, Journal of the American Statistical Association 86, 899–909.

Mroz, T.A., 1987, The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions, Econometrica 55, 765–799.

Muhuri, P.K. and S. Preston, 1991, Effects of family composition on mortality differentials by sex among children in Matlab, Bangladesh, Population and Development Review 17, 415–434.

Murdoch, J.J. and H.S. Stern, 1995, Using mixture models to detect sex bias in health outcomes in Bangladesh, Harvard Institute for Economic Research discussion paper no. 1728 (Harvard University, Cambridge, MA).

Newey, W., J.L. Powell, and J.R. Walker, 1990, Semiparametric estimation of selection models: Some empirical results, American Economic Review: Papers and Proceedings 80, 324–328.

Pitt, M.M., M.R. Rosenzweig, and M.N. Hassan, 1990, Productivity, health, and inequality in the intrahousehold distribution of food in low-income countries, American Economic Review 80, 1139–1156.

Quandt, R.E., 1972, A new approach to estimating switching regressions, Journal of the American Statistical Association 67, 306–310.

Quandt, R.E. and J.B. Ramsay, 1978, Estimating mixtures of normal distributions and switching regressions (with discussion), Journal of the American Statistical Association 73, 730–752.

Rubin, D.B. and H. Stern, 1994, Testing in latent class models using a posterior predictive check distribution, in: A. von Eye and C. Clogg, eds., Latent variables analysis: Applications for developmental research (Sage Publications, Thousand Oaks, CA) 420–438.

Strauss, J. and D. Thomas, 1995, Human resources: Empirical modeling of household and family decisions, in: T.N. Srinivasan and Jere Behrman, eds., Handbook of development economics, Vol. 3 (North-Holland, Amsterdam).

Titterington, D.M., A.F.M. Smith, and U.E. Makov, 1985, Statistical analysis of finite mixture distributions (Wiley, Chichester, MA).