

CONSUMPTION SMOOTHING AND EXCESS FEMALE MORTALITY IN RURAL INDIA

Elaina Rose*

Abstract—This paper examines the relationship between consumption smoothing and excess female mortality, by asking if favorable rainfall shocks in childhood increase the survival probabilities of girls to a greater extent than they increase boys' survival probabilities for a sample of rural Indian children. In order to avert the issue of selection bias due to underreporting of births of girls, a methodology is employed that does not require data on births by gender. The results indicate that favorable rainfall shocks increase the ratio of the probability that a girl survives to the probability that a boy survives.

I. Introduction

Substantial evidence suggests that female, relative to male, child mortality is higher in Asia than in most other parts of the world. Sen (1990) has estimated that "over 100 million women are missing." Rosenzweig and Schultz (1982) describe gender differences in child mortality for India, a country in which excess female mortality is particularly high, as the result of the process of intrahousehold resource allocation. The mechanisms through which excess female mortality arises include differential allocation of food and health care by gender,¹ and there is anecdotal evidence that female infanticide is practiced as well.²

There is also evidence that a child's well-being varies with fluctuations in income and prices (e.g., Foster, 1994), and that the well-being of girls is more sensitive to these fluctuations than the well-being of boys. Behrman (1988) finds that girls' nutrition suffers more than boys' nutrition in the lean, as opposed to the peak, agricultural season. Also, Behrman and Deolalikar (1990) report that price changes affect girls' consumption more than boys' consumption. Apparently, the allocation of family resources towards daughters is one margin on which households adjust to transitory events that they are unable to smooth through alternative means, such as credit markets.

This paper addresses two questions that remain unanswered regarding the role of a child's gender in consumption smoothing. First, can variations in female relative to male mortality, as well as consumption, be traced to transitory events? Second, are there critical ages at which girls are most vulnerable to shocks to their families' incomes?³ Using data from a national probability sample of over 4,000 households throughout India, the ratio of the probability that

a girl survives until school age to the probability that a boy survives to this age is related to rainfall shocks in childhood. Each rainfall shock is computed as the deviation of rainfall from its 21-year mean for each district; therefore, it reflects the purely transitory component of rainfall. The results show that survival is higher for girls relative to boys for cohorts experiencing favorable rainfall shocks during the first two years of life. The effect is most pronounced for landless households, which have less access to alternative means to smooth consumption and whose existence is most likely at the margin of survival.

Additionally, this study assesses the impacts on female relative to male mortality of other exogenous variables that would be suggested by an economic model of intrahousehold resource allocation. Mother's education, head's education, landholdings, and the availability of an educational institution in the village are all found to affect the differential between female and male mortality.

The methodology employed in this analysis accounts for the underreporting of female births in areas noted for having high degrees of gender preference. Evidence presented in this paper suggests that the underreporting is systematic; failure to account for the reporting bias would produce biased estimates of the effects of exogenous variables on female relative to male mortality.

The paper is structured as follows. Section II describes the data sets used. Section III discusses the implications of reporting bias, and the methodology employed in the analysis is described in section IV. The results are presented in section V. Section VI concludes.

II. The Data

The data for this study are obtained from the *Additional Rural Incomes Survey* (ARIS), produced by the National Council of Applied Economic Research in Delhi, India.⁴ This data set reports a variety of household and village characteristics for 4,118 households in 16 states in India from 1969 to 1971. In the final year, a retrospective fertility survey was undertaken for the women in the households; the survey asked the number of pregnancies as well as births and surviving children by gender. For each birth, the year of birth, whether the child survived until 1971, and the gender of the surviving child is reported. The ARIS is unique because it is a microlevel data set that spans a wide geographical area in India, and because it can be merged with district level rainfall data.

⁴ The means and standard deviations of the variables used in the analysis are reported in appendix A.

Received for publication August 14, 1995. Revision accepted for publication January 12, 1998.

* University of Washington.

I would like to thank Jere Behrman, Anil Deolalikar, Andrew Foster, Jutta Joesch, Shelly Lundberg, Mead Over, Robert Pollak, Dick Startz, and three anonymous referees for helpful comments and discussions. Any remaining errors are my own. This study was supported in part by the Boeing Endowment at the University of Washington.

¹ Muhuri and Preston (1991) summarize this literature.

² For example, see *Los Angeles Times* (Dahlburg, 1994).

³ Razzaque et al. (1990) find differential mortality effects by gender of one extreme adverse event: the 1974–1975 famine in Bangladesh.

Data on annual rainfall by district, obtained from *District Level Data* collected by the World Bank⁵ are merged with the ARIS data and used to compute rainfall shocks by year and district. For each year/district, the shock is computed as the difference between actual rainfall and the district's mean rainfall over a 21-year period. Because more rainfall increases the income of landed rural households,⁶ a positive rainfall shock can be interpreted as an income shock. Additionally, since rainfall increases output and wage rates in the district, a positive rainfall shock is associated with higher real earnings for landless rural households.^{7,8}

In order to infer mortality differentials without reliable data on births by gender, the sample is defined to minimize the impact of reporting bias. The sample consists of all surviving children born between 1961 and 1964.⁹ The sample is further limited to children of the 1,716 women aged 20–50 in 1971,¹⁰ and by the absence of rainfall data for four states.¹¹ The final sample consists of 2,297 observations.

While it would be ideal to define the dependent variable in this analysis as survival or mortality for a sample of children born between 1961 and 1964, there are two problems with using this definition given this data set. First, for children who died prior to the survey, the year of birth is reported, but not the gender.

A second problem relates to the underreporting of births and deaths of girls in areas in which gender preference exists. Underreporting may occur if female infanticide is practiced, or if deaths occur due to neglect or natural causes and the existence of a daughter is considered a less significant event than that of a son.¹² The presence of this type of bias can be detected by examining the reported sex ratios at birth (i.e., the number of boys reported to have been born for every 100 girls).

Table 1 shows the reported sex ratio at birth computed over all women in the ARIS fertility survey.¹³ The ratio for this sample is 122, in contrast to the biological ratio of approximately 105.¹⁴ The extent of underreporting varies

⁵ See Khandker (1987).

⁶ Rose (1995) using this data set. Additionally, Paxson (1992, 1993) finds that rainfall increases incomes of rural Thai households.

⁷ Extremely high rainfall, that is, a flood, may be an adverse shock. This issue is addressed in section V.

⁸ By merging the data on rainfall with the district in which the household is currently located, I am assuming that there is little migration of these households between 1961 and 1971.

⁹ These years were chosen because the objective is to study survival until school age (the last birth year that can be used is 1964), and because 1961 is the first year for which rainfall data are consistently available.

¹⁰ These women can plausibly have children aged 7–10 in 1971.

¹¹ The states included in the analysis are listed in table 1.

¹² This potential for bias in the reporting of births and deaths in areas where gender preference exists has been noted previously, by, for example, Visaria (1969). It has been addressed with aggregate data by comparing sex ratios by district over time (Miller, 1989) or by comparing the sex ratio in a population with the sex ratio in a comparable population in which no gender preference exists (Sen, 1990). To my knowledge, it has not been addressed at the microlevel.

¹³ This number is generated from a question in which each woman is asked the total number of live births by gender.

¹⁴ Johansson & Nygren, 1991.

TABLE 1.—REPORTED SEX RATIO AT BIRTH SAMPLE: ALL BIRTHS TO ALL WOMEN ON THE FERTILITY FILE

	Number of Women	Number of Births	Reported Sex Ratio at Birth
Entire Sample	5,038	19,149	122
By Education Category:			
Mother is literate or has some education	901	3,113	114
Mother illiterate and has no education	4,137	16,036	124
By State:			
Andhra Pradesh	279	847	120
Assam*	175	773	99
Bihar	208	686	117
Gujarat	419	1,646	131
Haryana	204	921	112
Jammu & Kashmir*	70	329	122
Karnataka	343	1,159	113
Kerala	344	1,212	101
Madhya Pradesh	460	1,853	116
Maharashtra	314	1,141	106
Orissa*	264	783	128
Punjab	241	1,054	145
Rajasthan	459	2,171	118
Tamilnadu	354	1,347	130
Uttar Pradesh	670	2,472	141
West Bengal*	234	755	165

Note: * Indicates that rainfall data were not available for that state between 1961 and 1964; therefore, the state was excluded from the logit estimation.

with the mother's education. Mothers who are literate or who have some education report births of 114 boys for every 100 girls born, while mothers who are both illiterate and have no formal education report births of 124 boys for every 100 girls.

The reported sex ratios at birth by state are listed in table 1. Several points should be noted. First, the reported sex ratio varies widely by state, ranging from 99 in Assam to 165 in West Bengal. Second, there is a geographical pattern. The three states with the highest sex ratios—Punjab, Uttar Pradesh and West Bengal—are all in the north of India, where the extent of gender preference is believed to be the most acute.¹⁵ One of the lowest ratios is for Kerala, a state in which the status of women is viewed to be one of the highest in India. While some states, such as Assam, do not conform entirely to the north/south distinction, the pattern suggests that there is an underreporting of female births relative to male births, and that the extent of underreporting is related to the degree of gender preference.

If births of daughters are underreported, and if the extent of underreporting is related to the degree of gender preference, then a standard estimation of the probability of survival on a set of exogenous variables for a sample of girls reported to have been born will produce biased coefficient estimates. The implications of the underreporting are discussed in terms of an empirical model in section III. The methodology proposed in section IV is designed to circumvent this problem.

¹⁵ Miller, 1989.

III. Implications of Reporting Bias

Consider a woman of childbearing age in the years 1961–1964, and assume that she can have only one child in each year. Figure 1 illustrates the five possible outcomes for this woman in each year: she can have no child (*NF*), she can have a daughter who survives until 1971 (*S_G*), she can have a daughter who dies by 1971 (*M_G*), she can have a son who survives until 1971 (*S_B*), or she can have a son who dies by 1971 (*M_B*).

The following events can also be defined:

- G = girl born ($G = S_G \cup M_G$)
- B = boy born ($B = S_B \cup M_B$)
- F = boy or girl born ($F = G \cup B$)
- S = any child survives until 1971 ($S = S_G \cup S_B$)

A reduced-form survival production function¹⁶ for a child of sex $s \in \{B, G\}$ born in year y is a function of a vector X of K potentially observable exogenous variables, such as exogenous income, prices, and parents' education, and inherently unobservable variables such as preferences and health endowments of family members which are captured in ϵ_s . Formally,

$$S_s^* = S_s^*(X, \epsilon_s), \tag{1}$$

where S_s^* is a latent variable associated with the outcome of survival. Linearizing equation (1) and allowing for gender-specific coefficients yields

$$S_B^* = X\beta + \epsilon_B, \text{ and} \tag{2a}$$

$$S_G^* = X\gamma + \epsilon_G. \tag{2b}$$

Assuming, for now, that the ϵ 's are independently and identically distributed, and that births of boys and girls by gender are not systematically underreported, β and γ can be consistently estimated using data on survival by gender until school age using a standard discrete choice framework:

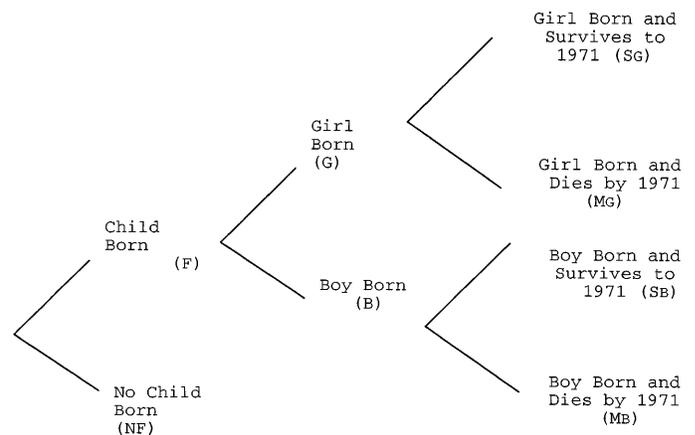
$$Pr(S|X, B) = Pr(S_B^* > \bar{S}) = Pr(X\beta > -\epsilon_B) \tag{3a}$$

$$Pr(S|X, G) = Pr(S_G^* > \bar{S}) = Pr(X\gamma > -\epsilon_G), \tag{3b}$$

where $Pr(S|X, s)$ is the probability that a child of sex s survives until school age, conditional on the exogenous variables. A test for differential consumption-smoothing behavior with respect to sons and daughters leading to differential survival probabilities is a test of the hypothesis that the marginal effects of shocks to income or prices on survival probabilities differ for girls and boys. Because data

¹⁶ See Behrman and Deolalikar (1988) for discussion of the structural model of intrahousehold resource allocation which, when augmented to include year-specific prices and income to capture the intertemporal dimension required to accommodate issues of consumption smoothing, underlies this reduced-form specification.

FIGURE 1.—POSSIBLE OUTCOMES FOR WOMAN OF CHILDBEARING AGE IN 1961–1964, FOR EACH YEAR



on income and price shocks are not consistently available for the necessary period of study, I proxy for these shocks with rainfall shocks, which were described in section II.¹⁷

The examination of the reported sex ratios at birth undertaken in section II suggests that the births of girls are in fact systematically underreported. In terms of figure 1, this means that observations for which the true outcome is M_G are misclassified as NF , and estimates of γ generated from the above framework are likely to be biased. To illustrate this problem, consider the simple case in which births of girls are underreported in the following bivariate probit model,

$$S_G^* = X\gamma + \epsilon_{GS} \tag{4a}$$

$$R^* = X\alpha + \epsilon_R, \tag{4b}$$

where S^* and R^* are latent variables associated with the outcomes of survival and the reporting of the birth, respectively.¹⁸ Then, assuming that ϵ_{GS} and ϵ_R are distributed bivariate normally, with zero means, standard deviations equal to one, and correlation ρ , the marginal effect of an exogenous variable, X_k , on the probability that a girl survives, given her birth is reported, is

$$\frac{\partial Pr(S = 1 | R = 1)}{\partial X_k} = \gamma_k E \left[\phi \left(\frac{X\gamma + \epsilon_R \rho}{\sqrt{1 - \rho^2}} \right) \middle| R = 1 \right] + \alpha_k \lambda(X\alpha) [.] \tag{5}$$

¹⁷ The use of rainfall shocks rather than income shocks has the additional advantage that the potential bias of the parameter estimates due to endogeneity of the shock is averted. For instance, if the entire family experiences an illness episode in year y , then the family will likely receive an adverse income shock in that year, and the probability that the child survives will be lower. This would result in a positive correlation between the income shock in year y and child survival, which would be unrelated to consumption-smoothing behavior.

¹⁸ The same sets of exogenous variables are used in equations (1) and (2) for ease of exposition; this need not be the case.

where $\phi(\cdot)$ is the standard normal pdf, $\lambda(X\alpha)$ is the inverse mills ratio, and $[\cdot]$ is an expression which is described in appendix B. Appendix B shows that $[\cdot]$ must be of the opposite sign of ρ , and is zero when ρ is zero.

The first term in equation (5) is comparable to the expression for the marginal effect of X_k in a probit model without sample selection. The second term results from selection due to underreporting of births. For instance, interpreting X_k as years of the mother's education,¹⁹ plausible signs for the parameters in equation (5) are $\gamma_k > 0$ (education improves daughter's survival), $\rho > 0$ (mothers who tend to underreport births also tend to have lower survival of daughters), and $\alpha_k > 0$ (mother's education is associated with higher levels of reporting of births). In this case, the true (causal) effect of education on survival is positive, but the second term, reflecting the selection due to underreporting, is negative.²⁰ The estimated effect of mother's education on survival would be biased downward. This bias would likely be negative and could, in principle, outweigh the true effect, yielding a negative coefficient. Intuitively, a mother's education raises her daughter's survival probability, but the selection bias reflects the fact that as the mother's education increases, the sample includes a larger pool of otherwise unreported girls who died.²¹

However, if equation (4a) is estimated as a probit model for a sample of boys, (whose births are not underreported), then the marginal effect of the mother's education on survival will be unbiased. Therefore, conventional estimates of the effect of the mother's education on girls' survival relative to boys' survival will be biased, and may indicate that the mother's education reduces female relative to male survival when the true effect is positive. A similar exercise, interpreting X_k as a favorable income shock, would also likely indicate that, while the true effect of the shock on boys' relative to girls' survival is positive, and the estimated effect would likely be biased downward. Therefore, in the case of reporting bias, the standard approach for estimating the effects of exogenous variables on the survival probabilities of girls relative to boys would lead to inconsistent parameter estimates. The methodology outlined in the

¹⁹ This discussion parallels Pitt and Rosenzweig's (1990) analysis of selection bias in estimating the effects of the mother's education on human capital outcomes using data from Malaysia. They also consider (non-gender-specific) mortality selection.

²⁰ In the case of a model in which the main equation such as equation (4a) has a continuous dependent variable and one regressor, then the bias in estimating β_k without correcting for sample selection equals the difference between the marginal effect of the change in X_k on the dependent variable and the coefficient. In the case in which the dependent variable is binary and the model contains several explanatory variables, estimates of equation (4a) with a probit model will be inconsistent, but the inconsistency will not in general take the form of this difference. Examining the expression for the marginal effect provides the basis for gaining intuition regarding the likely direction of the inconsistency due to selection in the probit model.

²¹ While the parents' education is treated as exogenous because it is predetermined at the time of the child's birth, the education variables may be endogenous in the sense that they are correlated with unobserved parental preferences and endowments.

following section is designed to eliminate this potential source of inconsistency.

IV. Estimation Methodology

This section describes the methodology for calculating the effects of exogenous variables on the ratio of survival probabilities (RSP), which is the ratio of the probability that a girl survives until school age to the probability that a boy survives to the same age. The proposed procedure eliminates the potential inconsistency that is due to the systematic underreporting of births. Data are not required on births, only on surviving children by gender, along with the set of exogenous variables. Because the gender of a child at birth is essentially a random event, any variable that is found to affect the probability that a surviving child is a girl ($Pr(G|S)$) must affect the girls' survival probabilities differently than it affects the boys' survival probabilities.

With data on surviving children only, the events M_G , M_B , and NF are obviously indistinguishable in the data. However, using Bayes Theorem:

$$Pr(G|S) = Pr(G)Pr(S|G)/Pr(S) \quad (6a)$$

$$Pr(B|S) = Pr(B)Pr(S|B)/Pr(S) \quad (6b)$$

The probabilities that girls or boys are born are given by

$$Pr(G) = q[Pr(B) + Pr(G)] \quad (7a)$$

$$Pr(B) = (1 - q)[Pr(B) + Pr(G)], \quad (7b)$$

where q is the probability that a given birth is a girl. Because the probability that a given birth is a girl is exogenous and assumed to be the same for all mothers,²² q is constant. Dividing equation (6a) by (6b), (assuming all probabilities are positive), substituting equation (7a) and (7b), and rearranging yields the RSP:

$$\begin{aligned} \frac{Pr(S|G)}{Pr(S|B)} &= \frac{Pr(G|S)(1 - q)}{Pr(B|S)q} \\ &= \frac{Pr(G|S)}{[1 - Pr(G|S)]} \frac{(1 - q)}{q}, \end{aligned} \quad (8)$$

since $Pr(B|S) + Pr(G|S) = 1$.

The advantages and drawbacks of this approach can be illustrated in terms of a multinomial logit model in which the three outcomes are S_B , S_G , and (NS means no surviving child, which includes NF , M_G , and M_B). The outcomes can be

²² Actually, female fetuses are less frail than male fetuses, so that the nutritional status of a mother during pregnancy may have a small effect on the probability that a child born is a girl (Hassold et al., 1983). Therefore, it may be more appropriate to say that the probability that a given conception, rather than birth, is exogenous and does not vary by mother.

modeled as

$$S_B^* = X\beta + \epsilon_B, \tag{9a}$$

$$S_G^* = X\gamma + \epsilon_G, \text{ and} \tag{9b}$$

$$NS^* = X\theta + \epsilon_N, \tag{9c}$$

where ϵ_B , ϵ_G , and ϵ_N are independently and identically distributed with the Type I extreme-value distribution. The probability that a girl is born and survives, given that either a boy or a girl is born and survives, can be expressed as

$$\begin{aligned} Pr(G|S) &= Pr(S_G|S_G + S_B) \\ &= \exp(X\gamma) * [\exp(X\beta) + \exp(X\gamma)]^{-1} \\ &= \{1 + [\exp(-(X(\gamma - \beta)))]\}^{-1} \\ &= \Lambda[X(\gamma - \beta)]. \end{aligned} \tag{10}$$

This conditional probability can be estimated as a logit model with the outcome being the survival of a girl until 1971 for a sample of surviving children. Defining δ as $\gamma - \beta$, the change in the probability that a surviving child is a girl due to a change in a (continuous) exogenous variable X_k is

$$\partial Pr(S_G|S_G + S_B)/\partial X_k = \delta_k * \text{FACTOR1}, \tag{11}$$

where $\text{FACTOR1} = [\Lambda(X\delta)][1 - \Lambda(X\delta)] > 0$.

The RSP can be expressed in terms of the logit coefficients and q as

$$Pr(S|G, X)/Pr(S|B, X) = \exp(X\delta) * [(1 - q)/q] \tag{12}$$

The marginal effect of a change in X_k on this ratio is

$$\partial [Pr(S|G, X)/Pr(S|B, X)]/\partial X_k = \delta_k * \text{FACTOR2}, \tag{13}$$

where $\text{FACTOR2} = \exp(X\delta) * [(1 - q)/q] > 0$. The expression $[(1 - q)/q]$ can be set to 1.05, consistent with a population sex ratio at birth of 105. Thus, the effect of an exogenous variable on the RSP can be estimated with a dichotomous logit model, and requires estimation of K parameters, rather than $2K$ parameters, which would be required if equation (9) were estimated directly.

An additional advantage of this approach is that conditioning the outcome that a girl survives on the probability that any child survives provides the means for eliminating household-specific heterogeneity which is not gender specific. Until this point, I have not specified a structure for the ϵ 's, which capture unobservables such as child health endowments, parental preferences, and unobservable shocks. A plausible structure for each of the error terms is

$$\epsilon_{smy} = \mu_y + \mu_m + \mu_s + \mu_{sm} + \mu_{my} + \mu_{sy}, \tag{14}$$

where the subscript s represents the child's sex, y represents the year of the child's birth, and m represents mother. μ_y captures age effects and the effects of aggregate (national level) shocks. The effects of preferences and health endowments are viewed as time invariant and consist of a gender-specific effect (μ_s), a mother-specific effect (μ_m), and a gender/mother-specific effect (μ_{sm}). Mother- and gender-specific year effects are captured by μ_{my} and μ_{sy} , respectively. The error term associated with the model described in equation (10) is the difference between ϵ_{Bmy} and ϵ_{Gmy} , which reduces to

$$\epsilon_{Bmy} - \epsilon_{Gmy} = (\mu_B - \mu_G) + (\mu_{Bm} - \mu_{Gm}), \tag{15}$$

when year-specific dummy variables are included as regressors to capture the effect of ($\mu_{By} - \mu_{Gy}$). I remove from the error term any heterogeneity that is not gender specific (and therefore not reflected in $\mu_y + \mu_{my} + \mu_m$) nor arising from endowments that are not gender specific, preferences towards children that are not gender specific, and shocks that do not affect boys and girls differently. The only terms remaining are those reflecting gender-specific endowments and preferences, or the impacts of shocks that affect girls differently from boys, at the aggregate level and mother level respectively.

The drawback of this approach is that the assumption of independence of irrelevant alternatives (IIA)—which is a potential source of inconsistency in the multinomial logit model—is inherent in this procedure as well. This assumption can be thought of in two ways in terms of the model in equations (9a–c). First, the assumption means that the ratio of any two of the probabilities is independent of the probability of the third event. For instance, the probability that a girl survives relative to the probability that a boy survives is independent of the probability that no child survives. Second, the ϵ 's in (9) are assumed to be independent under the IIA assumption. This would be violated, for example, if there are unobservable health endowments or preferences toward allocation of resources to children that are correlated with the unobservables affecting fertility behavior.²³

V. Results

Table 2 reports the results of logit regressions in which the outcome is whether a surviving child is a girl, and the set of explanatory variables includes measures of rainfall shocks. The results for the full sample are reported in column (1). A positive rainfall shock in the first year of life significantly

²³ The IIA assumption could be avoided by estimating equation (9) using a multinomial probit model. However, because the set of X 's is identical for each of the three outcomes, the identification of the parameters would be tenuous (Keane, 1992). This identification issue does not arise in the approach used in this paper. Equation (8) shows that the effect of an exogenous variable X_k on the RSP is the same sign as the estimated effect of that variable on $Pr(G|S)$. This identification is achieved prior to assuming IIA, which amounts to a choice of functional form for equation (8).

TABLE 2.—LOGITS DEPENDENT VARIABLE: GIRL^a

Sample	Pooled		Landed		Landless	
	All (1)	No Floods (2)	All (3)	No Floods (4)	All (5)	No Floods (6)
RF Shock (year of birth)	0.26 (2.6)	0.39 (2.4)	-0.02 (0.1)	0.17 (0.9)	0.89 (3.5)	0.92 (2.0)
RF Shock (age 1)	0.39 (2.6)	0.57 (3.5)	0.40 (2.2)	0.54 (2.5)	0.11 (0.23)	0.63 (1.5)
RF Shock (age 2)	0.27 (1.5)	0.09 (0.4)	0.09 (0.4)	-0.22 (0.9)	0.73 (1.8)	0.75 (1.4)
RF Shock (age 3)	0.11 (0.7)	0.03 (0.2)	0.10 (0.5)	-0.06 (0.3)	0.70 (1.8)	0.60 (1.4)
RF Shock (age 4)	0.20 (1.0)	0.29 (1.4)	0.28 (1.3)	0.13 (0.5)	0.33 (1.0)	0.59 (1.4)
RF Shock (age 5)	0.11 (0.8)	0.17 (1.1)	0.11 (0.6)	0.06 (0.3)	0.13 (0.4)	0.40 (1.2)
Land Owned	0.001 (1.5)	0.002 (1.8)	0.002 (1.5)	0.002 (1.9)		
Mother Educated	0.25 (2.1)	0.31 (2.1)	0.23 (1.5)	0.27 (1.6)	0.39 (1.5)	0.44 (1.3)
Head Educated	-0.28 (2.5)	-0.38 (3.3)	-0.32 (2.3)	-0.50 (3.2)	-0.16 (0.7)	-0.16 (0.7)
Educational Institution	-0.28 (2.5)	-0.30 (2.5)	-0.13 (1.1)	-0.17 (1.3)	-1.2 (3.4)	-0.99 (3.0)
Health Institution	0.19 (1.6)	0.21 (1.6)	0.14 (0.9)	0.22 (1.3)	0.38 (2.1)	0.25 (1.2)
FACTOR1	0.25	0.25	0.25	0.25	0.25	0.25
FACTOR2	0.83	0.82	0.82	0.80	0.85	0.86
N	2,297	1,926	1,722	1,418	575	508

Notes: ^a T-statistics computed from Huber (1967) standard errors reported in parentheses. Controls for the mean and standard deviation of rainfall and dummy variables for state, year of child's birth, mother's age and missing mother's education (33 observations) are included. Multiply coefficients by FACTOR1 to obtain marginal effect evaluated at the sample mean of a change in the exogenous variable on the probability that a surviving child is a girl; multiply by FACTOR2 for the marginal effect on the ratio of the survival probability of girls to the survival probability of boys.

increases the probability that a surviving child is a girl. The coefficient for the subsequent year of life is fifty percent larger. Clearly, more rainfall in the first two years of life increases the likelihood of a girl's surviving relative to a boy's surviving. This is consistent with the consumption-smoothing hypothesis: in good years, overall survival is higher, and girls benefit disproportionately from the shock. They also suffer disproportionately from adverse shocks. If households were able to use alternative means to smooth consumption between years, there a weather shock would have no effect. There are similar positive effects of rainfall in years two to five following birth, but they are not significant for this sample.

An alternative explanation for the results must be considered. Because more rainfall is associated with the increased risk of certain illnesses—and because girls may be more resistant to these diseases—a positive effect of rainfall on the outcome may be due to girls' relative strength in combating illnesses associated with rainfall. To see if this biological explanation is responsible for the pattern, all observations with flooding are dropped from the sample in the specification reported in column (2) of table 2.²⁴

Under the consumption-smoothing hypothesis, rainfall increases the probability that a surviving child is a girl because a positive rainfall shock is interpreted as a favorable shock to income or purchasing power, and the resources, that

are allocated to girls respond more to these shocks than the resources allocated to boys. However, more rainfall is not a favorable shock at extremely high levels of rainfall; that is, floods are an adverse shock to income. Therefore, removing observations associated with flooding would increase the estimated impact of rainfall. However, with the biological explanation, if deaths from illness associated with rainfall are more likely to occur at the highest levels of rainfall, then the estimated effect of rainfall would decrease when observations associated with flooding are omitted.

When the observations associated with floods are deleted from the model, the coefficients on the shock variables with significant coefficients (and the associated marginal effects) increase. (See column (2) of table 2.) The hypothesis that excess female mortality results from the household inability to smooth consumption in response to adverse shocks is therefore supported.²⁵

The specifications reported in columns (1) and (2) of table 2 are repeated in columns (3) and (4) for landed households, and in columns (5) and (6) for landless households. Two results are notable in comparing the disaggregated results

²⁵ Additional evidence from Matlab, Bangladesh, in the 1970s weighs against the biological explanation. Intestinal infections may be associated with rainfall and are a major source of mortality for rural South Asia. Chen et al. (1981) find that boys' perceived greater susceptibility to intestinal infections is substantially overstated due to parents' greater tendency to seek treatment for sons relative to daughters. In fact, Waldron (1987) reports that deaths from intestinal infections are actually higher for girls than for boys.

²⁴ Flooding here is defined as rainfall greater than two standard deviations over mean rainfall for the district.

with the pooled results. First, in general, the rainfall effects are much larger for the landless households than the landed households. Because landed households are better able to smooth consumption, the consumption-smoothing hypothesis is again supported. Second, for the disaggregated samples, the marginal effects of rainfall generally increase when observations with floods are eliminated.²⁶

Various theories within economics can explain differential allocation of resources towards sons and daughters. In a common-preference model, there is a single household utility function. Differences in the allocation of resources arise through preferences and are attributed to discrimination against girls if a household prefers sons to daughters, or through the investment component of children if sons generate greater net returns to their parents than do daughters. Alternatively, discrimination can arise in the context of a household bargaining framework if each parent prefers children of his/her own gender, and mothers have less bargaining power within the household than fathers.²⁷

The results in table 2 indicate that landholdings increase the survival probabilities of girls relative to boys. Since landholdings reflect a household's permanent income, this result is consistent with a common-preference approach in which daughters are a normal good.²⁸ This may also be related to consumption-smoothing behavior, since greater wealth enhances a farm household's ability to smooth consumption.²⁹

For landed households, the mother's education is associated with higher survival probabilities of daughters relative to sons, while the head's education has the opposite effect. The same pattern arises for landless households, although the effect of the head's education is insignificant. These

²⁶ The consumption-smoothing hypothesis implies that the presence of a resource that enables households to smooth consumption in response to income variations would reduce the impact of the shock on the RSP. This implication was tested by adding a dummy variable indicating the presence of a moneylender in the village in 1971 (the only year in which this variable is reported) and interacting the dummy variable with the six shock variables. Moneylenders are virtually the sole source of consumption loans outside of friends and relatives for this sample, and their presence promotes consumption smoothing (Rose, 1995). For all subsamples reported in table 3, the coefficients on the dummy variable and the interactions were insignificant.

²⁷ Also, if gender preference is manifested by parents' having a minimum number of sons, then households with more children are likely to have a higher percentage of daughters. If households are credit constrained, girls may suffer more than boys in bad years because they are in households with fewer resources per capita, even if boys and girls within the same household are treated the same.

²⁸ An alternative measure of household permanent income that can be computed with these data is the average of total income received by each household over the three crop years of 1969 through 1971. The specifications reported in table 2 were estimated for landed and landless households including this measure of average income. For landless households, the coefficient on average income was positive and significant. For landed households, the coefficient on average income was positive, but not significant unless landholdings were excluded, due to the high degree of collinearity between landholdings and average income for landed households. The other results were essentially unchanged.

²⁹ Binswanger and Rosenzweig, 1993; Morduch, 1990; Rose, 1996.

results accord with previous findings for both developing and developed countries. Behrman (1988) finds that the preference toward boys in allocation of nutrients increases with the head's education for Indian households. Additionally, Thomas (1994) reports that, in the U.S., Ghana, and Brazil, the long-term nutritional status of girls (measured as height for age) is positively associated with the mother's education, while boys' height for age is associated with the father's education.³⁰ The significant results are consistent with a bargaining framework in which each parent's education is associated with his/her power, and parents prefer children of their own gender. Alternatively, this pattern can arise within a common-preference framework if education by gender is related to relative rewards in the labor market, and households allocate resources towards children who will generate higher returns.³¹

The presence of an educational institution in the village increases the probability that a surviving child is a boy for landless households. In terms of a common-preference framework, this result is consistent with boys' reaping higher returns to education than girls.³² Then, if parents receive at least the same share of the returns to educating their sons as they receive from educating their daughters, the presence of an educational institution will increase parents' returns to sons relative to daughters.

VII. Conclusion

In recent years, two issues have been of particular concern to economists studying the behavior of households in developing countries. First, because incomes of rural households vary widely by season and year, the ability of households to cope with income fluctuations and the measures they use to smooth their consumption in response to variations in income have been examined closely. Second, various authors have identified the magnitude of bias in the allocation of resources in favor of boys relative to girls and have uncovered factors associated with this gender bias, which ultimately leads to the excess mortality of girls in South Asia.

In this paper, the connection between these two issues is examined. The results support a relationship between consumption smoothing and excess female mortality: a favorable rainfall shock increases the likelihood—relative to that of a boy—that a girl survives until school age.

The results of this paper have several implications. In terms of methodology, the procedure employed in this work

³⁰ On the other hand, Das Gupta (1987) finds that the mother's education is associated with increased mortality of daughters in rural Punjab, although she notes that this result is based on a small sample of births.

³¹ This result might also occur if both parents' education tends to increase the welfare of girls relative to boys, but the two education variables are highly colinear. However, the specifications reported in table 2 are all reestimated twice: excluding either the mother's or the head's education each time. The results were essentially unchanged.

³² Rosenzweig (1980) finds that the returns to education in the labor market are higher for males than for females using this data set.

enables inferences to be made regarding the effect of exogenous variables on excess female mortality using microlevel data, even when data on births by gender are unreliable. As tools to diagnose the gender of a fetus become more widely available in South Asia, this methodology can also be employed to assess the household characteristics that are associated with the use of gender-selection abortions by estimating the effects of a set of exogenous variables on observed sex ratios at birth. However, the methodology will not provide a means for disentangling the effects of variables associated with sex-differential mortality and those associated with the use of gender-selection abortions as the latter become more prevalent.

Another implication for studies of gender bias is that, in analyses involving data from areas with excess female mortality, it is inappropriate to treat a child's gender as exogenous. Because of differential mortality selection, the probability that a surviving child is a girl is related to household and village characteristics, and the experiences of the child's cohort. This will result in estimates of the well-being of girls relative to boys that are biased upward, as well as to biased estimates of the effects of exogenous variables on these measures of relative well-being. In particular, the effects of policies intended to improve the well-being of girls relative to boys will tend to be underestimated.

Several authors have noted that household expenditure data from areas of south Asia in which excess female mortality is particularly high fail to exhibit gender bias. For instance, Ahmad and Morduch (1993) find little evidence of gender bias in household expenditures in the 1988 Household Expenditure Survey for Bangladesh, although the sex ratios in the same sample are dramatically skewed. Subramanian (1994) reports no evidence of gender bias in allocation of resource in states in the north of India where excess female mortality is believed to be most severe, while some evidence is found for states in areas of India where mortality differentials are not as high. These apparent paradoxes can be resolved with the recognition that gender is endogenous in samples of data drawn from areas in which excess female mortality exists, and that factors affecting consumption patterns will also tend to affect the gender composition of the household. The results in this paper add to the growing evidence (e.g., Rosenzweig & Stark, 1989; Foster, 1994) that empirical analyses of the determinants of the patterns of intrahousehold resource allocation will lead to misleading conclusions if the endogeneity of family structure in general is not addressed.

The final implication concerns development policy for rural India, and perhaps other areas as well. Because the excess mortality of girls is related to an inability to smooth consumption, promoting institutions that facilitate consumption smoothing may reduce the pressure on rural households to cope with adverse shocks by sacrificing the survival of their daughters.

REFERENCES

- Ahmad, Asif, and Jonathan Morduch, "Identifying Sex Bias in the Allocation of Household Resources: Evidence from Linked Household Surveys from Bangladesh," mimeograph, Harvard University (1993).
- Behrman, Jere R., "Intrahousehold Allocation of Nutrients in Rural India: Are Boys Favored? Do Parents Exhibit Inequality Aversion?" *Oxford Economic Papers* 40:1 (1998), 32–54.
- Behrman, Jere R., and Anil Deolalikar, "Health and Nutrition," Chap. 14 in *Handbook of Development Economics* (1988), 631–711.
- "The Intrahousehold Demand for Nutrients in Rural South India: Individual Estimates, Fixed Effects and Permanent Income," *The Journal of Human Resources* 25: (1990), 655–696.
- Binswanger and Rosenzweig, "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments," *The Economic Journal* 103 (1993), 56–78.
- Chen, Lincoln C., Emdadul Huq, and Stan D'Souza, "Sex Bias in the Family Allocation of Food and Health Care in Rural Bangladesh," *Population and Development Review* 7:1 (1981), 55–70.
- Dahlburg, John-Thor, "The Fight to Save India's Baby Girls," *Los Angeles Times* (February 24, 1994), p. A1.
- Das Gupta, Monica, "Selective Discrimination Against Female Children in India," *Population and Development Review* 13: (1987), 77–101.
- Foster, Andrew, "Prices, Credit Constraints and Child Growth in Rural Bangladesh," *Economic Journal* 105:43 (1994), 551–570.
- "Analysis of Household Behavior when Households Choose Their Members: Marriage Market Selection and Human Capital Allocations in Rural Bangladesh," mimeograph, University of Pennsylvania (1995).
- Hassold, T., S. D. Quillen, and J. A. Yamana, "Sex Ratio in Spontaneous Abortion," *Annales of Human Genetics* 47 (1995), 39–47.
- Huber, P.J., "The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1 (1987), 221–233.
- Johansson, Sten, and Ola Nygren, "The Missing Girls of India," *Population and Development Review* 17:1 (1991), 35–51.
- Keane, Michael, "A Note on Identification in the Multinomial Probit Model," *Journal of Business and Economic Statistics* 10:2 (1992), 193–200.
- Khandker, Shahidur, "Notes on District Level Data from India," mimeograph, The World Bank.
- Miller, Barbara, "Changing Patterns of Juvenile Sex Ratios in Rural India," *Economic and Political Weekly* 24:22 (1989), 1229–1236.
- Morduch, Jonathan, "Risk, Production and Savings: Evidence from Indian Households," mimeograph, Harvard University (1990).
- Muhuri, Pradip K., and Samuel H. Preston, "Effects of Family Composition on Mortality Differentials by Sex Among Children in Matlab, Bangladesh," *Population and Development Review* 17:3 (1991), 415–434.
- Paxson, Christina, "Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand," *American Economic Review* 82:1 (1992), 15–33.
- "Consumption and Income Seasonality in Thailand," *Journal of Political Economy* 101:1 (1993), 39–72.
- Pitt, Mark, and Mark Rosenzweig, "The Selectivity of Fertility and the Determinants of Human Capital Estimates: Parametric and Semiparametric Estimates," Living Standards Measurement Survey Working Paper 89–03 (1990).
- Razaque, Abdur, Nural Alam, Lokky Wai, and Andrew Foster, "Sustained Effects of the 1974–5 Famine on Infant and Child Mortality in Rural Area of Bangladesh," *Population Studies* 44 (1990), 145–154.
- Rose, Elaina, "Ex Ante and Ex Post Labor Supply Responses to Risk in a Low-Income Area," mimeograph, University of Washington, (1995).
- "Gender Bias, Credit Constraints and Time Allocation in Rural India," mimeograph, University of Washington (1996).
- Rosenzweig, Mark R., "Neoclassical Theory and the Optimizing Peasant: An Econometric Analysis of Market Family Labor Supply in a Developing Country," *Quarterly Journal of Economics*, 95 (1980) 31–55.

Rosenzweig, Mark R., and T. Paul Schultz, "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India," *American Economic Review* 72 (1982) 803–815.

Rosenzweig, Mark R., and Oded Stark, "Consumption Smoothing, Migration and Marriage: Evidence from Rural India," *Journal of Political Economy* 97 (1989).

Sen, Amartya, "More than 100 Million Women are Missing" *New York Review of Books* (December 20, 1990), 61–66.

Subramanian, Shankar, "Gender Discrimination in Intra-Household Allocation in India," mimeograph, Cornell University (1994).

Thomas, Duncan, "Like Father, Like Son; Like Mother, Like Daughter: Parental Resources and Child Height," *The Journal of Human Resources* 29:4 (1994) 950–988.

Visaria, Pravin M., *The Sex Ratio of the Population of India, Census of India 1961, Vol I, Monograph 10* (New Delhi, India: Office of the Registrar General, 1969).

Waldron, Ingrid, "Patterns and Causes of Excess Female Mortality among Children in Developing Countries," *World Health Statistics Quarterly* 40 (1987) 194–210.

$$Pr(S = 1 | R = 1) = Pr(\epsilon_S > -X\gamma, \epsilon_R > -X\alpha) = \frac{Pr(\epsilon_S > -X\gamma, \epsilon_R > -X\alpha)}{Pr(\epsilon_R > -X\alpha)} \tag{A1-3}$$

Evaluating the numerator of equation (A1-3) yields

$$Pr(\epsilon_S > -X\gamma, \epsilon_R > -X\alpha) = \int_{-X\alpha}^{\infty} \int_{-X\gamma}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{\epsilon_S^2 - 2\rho\epsilon_S\epsilon_R + \epsilon_R^2}{-2\sqrt{1-\rho^2}}\right) d\epsilon_S d\epsilon_R \tag{A1-4}$$

$$= \int_{-X\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\epsilon_R^2}{-2}\right) \int_{-X\gamma}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \tag{A1-5}$$

$$\times \exp\left(\frac{(\epsilon_S - \rho\epsilon_R)^2}{-2(1-\rho^2)}\right) d\epsilon_S d\epsilon_R = \int_{-X\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\epsilon_R^2}{-2}\right) \Phi\left(\frac{X\gamma + \rho\epsilon_R}{\sqrt{1-\rho^2}}\right) d\epsilon_R \tag{A1-6}$$

Also, note that the denominator of (A1-3) is

$$Pr(\epsilon_R > -X\alpha) = \Phi(X\alpha) \tag{A1-7}$$

Substituting equation (A1-6) and (A1-7) into equation (A1-3), and differentiating with respect to X_k (applying the quotient rule and Leibniz's rule) yields

$$\frac{\partial Pr(S = 1 | R = 1)}{\partial X_k} = \gamma_k \int_{-X\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\epsilon_R^2}{-2}\right) \times \left[\frac{1}{\sqrt{1-\rho^2}} \phi\left(\frac{X\gamma + \rho\epsilon_R}{\sqrt{1-\rho^2}}\right) [\Phi(X\alpha)]^{-1} \right] \times d\epsilon_R + \alpha_k \lambda(X\alpha) [\dots] \tag{A1-8}$$

where $\phi(\cdot)$ is the standard normal pdf, $\lambda(X\alpha)$ is the inverse mills ratio, and

$$[\dots] = \Phi\left(\frac{X\gamma - \rho X\alpha}{\sqrt{1-\rho^2}}\right) - \int_{-X\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{\epsilon_R^2}{-2}\right) \times \Phi\left(\frac{X\gamma + \rho\epsilon_R}{\sqrt{1-\rho^2}}\right) d\epsilon_R [\Phi(X\alpha)]^{-1} \tag{A1-9}$$

When $\alpha_k = 0$, the second term in equation (A1-8) is zero, so equation (A1-8) reduces to

$$\gamma_k E \left[\Phi\left(\frac{X\gamma + \rho\epsilon_R}{\sqrt{1-\rho^2}}\right) \middle| S = 1 \right] \tag{A1-10}$$

When $\rho = 0$, equation (A1-8) reduces to:

$$\gamma_k \phi(X\gamma). \tag{A1-11}$$

Equations (A1-10) and (A1-11) are comparable to the expression for the marginal effect of X_k on the probability of survival when there is no sample selection.

Also note that, since the inverse mills ratio is positive, the sign of the second term in equation (A1-8) depends upon the sign of α_k and the sign of $[\dots]$ from equation (A1-9). Since $\epsilon_R > -X\alpha$ for all reported births, $\rho > (<) 0 \Rightarrow (X\gamma - \rho X\alpha) < (>) (X\gamma + \rho\epsilon_R)$ for all reported births. Therefore, $\text{sgn}([\dots]) = -\text{sgn}(\rho)$.

APPENDIX A.—VARIABLES AND DEFINITIONS

Variable	Mean (Std. Deviation)
Child is a girl	0.44
Deviation of district rainfall from its 21-year mean in year that the child was born ^a	0.12 (0.40)
Mean rainfall for the district	1.2 (1.5)
Standard deviation of rainfall for the district	0.32 (0.19)
Household's landholdings	37 (50)
Mother's Education: = 1 if mother is literate or has some education, = 0 otherwise	0.18
Head's Education: = 1 if head is literate or has some education, = 0 otherwise	0.78
Educational Institution: = 1 if education institution was in village in 1971, = 0 otherwise	0.92
Health Institution: = 1 if health institution was in village in 1971, = 0 otherwise	0.29
Mother's age in 1971.	35 (7.0)

^a Rainfall is measured in meters.

Appendix B

Selection Effect in Probit Model

Consider the following bivariate probit model:

$$S^* = X\gamma + \epsilon_S \tag{A1-1}$$

$$R^* = X\alpha + \epsilon_R \tag{A1-2}$$

$(\epsilon_S, \epsilon_R) \sim \text{BVN}(0, 0, 1, 1, \rho)$,

where S^* is a latent variable associated with the outcome of survival, and R^* is a latent variable associated with the outcome of the child's birth being reported. Then, the probability that a child dies, given that the birth was reported is