Marriage-Market Selection and Human Capital Allocations in Rural Bangladesh

Andrew D. Foster
Economics Department
University of Pennsylvania

September 2002

This research supported in part by NIH Grant HD10379.

I. Introduction

In recent years there has been increased recognition of the potential importance of the fact that households are composed of groups of individuals, each with his or her own set of preferences. While collective households can be treated for some purposes as if each were each a single agent with well defined preferences and constraints, it is well known that in some circumstances predictions based on collective behavior differ substantially from those that would arise from a unitary model. Moreover, there is a significant body of recent empirical work reporting results that appear to be inconsistent with the unitary model (e.g., Browning et al 1994; Schultz 1990; Thomas 1990; Udry 1995).

A striking feature of most if not all of these studies is that the composition of the household is taken as given. While this might be a reasonable simplification if the results rested primarily on sources of variation in household characteristics over which members have little choice such as the sex or endowments of their offspring, in fact most of these studies rely primarily on variation in the characteristics of spouses, something over which households have a great deal of control.

The fact that attributes of spouses may be endogenously determined along with other household choices has significant implications for the analysis and interpretation of household behavior (e.g., Lam and Schoeni (1993), Schultz (1990)). A man, for example, with unobservable characteristics predisposing him to want educated children may choose an educated wife; thus the wife's schooling may appear to influence child's schooling net of the husband's observable characteristics even if, for the same man, changes in wife's schooling would not affect the education of the resulting children. Conversely, the effects of male schooling might be overestimated if women with a high taste for schooling tend to choose educated husbands. Not only might these sources of bias lead to inappropriate targeting of educational resources but could lead to misleading inferences about the extent to which household behavior conforms with the unitary model.

The main barrier to understanding how the process of partner selection influences household behavior is that in order to measure and properly control for the endogeneity of the choice of partners it is necessary to know a great deal about the operation of marriage markets at the time and place in which a given marriage takes place.[1] Because marriages relevant to the typical targets of household surveys will have taken place a number of years before the survey and little information is likely to be available regarding the set of potential partners, few data sets have sufficient information to support an analysis of the implications of marriage-market selection for household behaviors of interest.

In this paper I examine the implications of marriage-market selection for human capital formation using a unique data set from rural Bangladesh that contains longitudinal information

---

[1]Schultz (1990) clearly articulates the need to draw identification restrictions explicitly from models of marriage market structure but, given the absence of appropriate data, uses functional form restrictions to determine whether marriage-market selection bias underlies distinct effects of husband's and wife's unearned income on fertility and child health. The relevant work on marriage markets (e.g., Bergstrom and Lam 1993) does not consider the implications of the marriage market for human capital investment. Although Boulier and Rosenzweig (1984) does consider the relationship between schooling investment and marriage market behavior in the presence of unobservables, the focus is on the endogeneity of schooling choice with respect to prospects in the marriage market. While the implications of marriage-market selectivity for fertility behavior are also examined briefly, identification rests on the untested assumption that there is no unobserved heterogeneity in male characteristics. Boulier and Rosenzweig do not explicitly model the structure of the marriage market and the aggregation of individual preferences within the household.

on marriage in a population of approximately 200,000 people over a period of 15 years that can be linked to censuses containing information on the households of origin of both spouses as well as information on schooling aspirations for children resulting from these marriages. I develop a model of the marriage market that incorporates the assumption that human capital investment in children is a public good within marriage and makes use of the equilibrium condition that, given transferable utility, the marriage allocations that take place at a given place and time must maximize total utility.

Because the model does not lend itself to the development of analytic expressions for purpose of estimation but can be readily simulated, a three-stage simulation procedure is used. First, parameter estimates based on an approximate but incorrectly specified econometric model are compared to parameter estimates from the same model applied to simulated data in order to establish that the incorrectly specified model provides a reasonable approximation to the true model. Second, residuals from the simulated data are used to estimate the bias likely to arise in estimates of human capital decision rules as a consequence of the covariances induced by the marriage selection process between observable (to the econometrician) attributes of individuals and the unobservable attributes of their spouses. Third, regression equations relating these residuals to the attributes of the couples participating in each marriage market in the simulated data are used to construct the predicted values of the male and female endowments conditional on a given marriage taking place in a given marriage market. Incorporation of these predicted values into desired schooling decision rules yields selection-corrected estimates of the effects of male and female characteristics on desired schooling. The resulting coefficients along with the bias estimates are shown to be mutually consistent. Simulations indicate that these consistency checks would have been violated under a particular form of model misspecification that results in incorrect predictions regarding the correlation between observables and unobservables. Thus the approach appears to capture the essential aspects of marriage-market selection in the study population.

The results provide evidence that marriage-market selection is significant. I find that the uncorrected estimates tend to overstate the effects of male characteristics on desired schooling by 35-55% while understating the effects of female characteristics by 13-16%, with the higher percentage biases with regard to the male characteristics arising when a limited set of female characteristics are used in the analysis. These results suggest that unobservable variation in female income and/or tastes is the primary source of selectivity bias in this population. Although the finding that selectivity is important indicates that tests rejecting the unitary model that do not control for selectivity bias should be interpreted with caution, the approach underlying these results illustrates clearly the significance of the broader perspective on household behavior offered by the collective model.

II. Model

In order to accurately assess the implications of marriage selectivity for measuring the effects of spouses' characteristics on household behaviors such as human capital investment in children it is necessary to construct a model of the marriage market that explicitly incorporates parental decision making about the allocation of resources. In particular, because in contrast to many other forms of expenditure such as that on calories, human capital investment in children is likely to contribute directly to the utility of both partners,[2] I construct a marriage-market model that focuses on the provision of public goods within marriage similar to that used by Bergstrom and Lam (1993).

Specifically, I assume that each individual i of sex s entering the marriage market may be characterized by his or her income and his or her tastes for schooling. After partners are chosen through the marriage market, each couple has one child[3] whose sex becomes known at birth. The couple then allocates total income to private consumption for the two parents and to human capital acquisition h for the child.

The utility that is realized by each member is assumed to be quadratic in the human capital h provided to the child and linear, given h, in own private consumption c:

$$u(c,h;s,z,g) = (1+\eta h)c + \delta_{sz}h - \frac{\alpha_s}{2}(h-g)^2 \qquad \textbf{(1)}$$

where s = {M,F} (for male and female, respectively) denotes the sex of the parent, z the sex of the child, g is an index of the individual's taste for human capital,[4] and δ_{sz} and $\alpha_s>0$ are taste parameters. A key feature of (1) is that it implies that households exhibit transferable utility

---

[2]In the presence of altruism individual caloric intake may also have a significant public-good component. The main point here is that investment in children is likely to be considered more of a public good from the perspective of the parents than is caloric intake.

[3]Extension of the model to allow for multiple children complicates the notation but does not affect the empirical implications that are used below.

[4]Note that an increase in g increases human capital acquisition: the optimal level of human capital acquisition for a single parent of sex s with taste parameter g and a child of sex z is linear and increasing in g: $h=(\eta y_s+\delta_{sz}+\alpha_s g)/(2\eta p_h+\alpha_s)$.

which implies, “Roughly speaking...that utility within the household can be ‘redistributed’, like apples or bananas. If one distribution of utility is possible then so is any other distribution of utilities where individual utilities sum to the same number.” (Bergstrom 1995, p7). While this functional form is more restrictive than necessary for transferable utility[5] it is analytically convenient and complex enough to capture key features of the choice between public and private consumption faced by multiple person households given heterogeneity in tastes and income.

A significant implication of the fact that this function exhibits transferable utility is that, for given incomes and tastes, all Pareto efficient allocations within the household will yield the same level of human capital investment, with different points on the Pareto frontier corresponding to differences in how private consumption is divided within the household. (Bergstrom 1995) Thus the optimal level of human capital investment h_{ij} for a couple composed of male i, with income y_{Mi} and tastes g_{Mi} and female j with income y_{Fj} and tastes g_{Fj}, who have a child of sex z can be determined by maximizing the sum of the couple's utilities with respect to h and total consumption $c = c_{Mi} + c_{Fj}$.

$$(1+\eta h)c+\delta_{Mz}h-\frac{\alpha_M}{2}(h-g_{Mi})^2+\delta_{Fz}h-\frac{\alpha_F}{2}(h-g_{Fi})^2 \qquad \textbf{(2)}$$

---

[5]Necessary and sufficient conditions for transferable utility in the present context given interior solutions are that preferences may be characterized as u(c,h;s,z,g)=q(h,z)c+r(h,s,z,g) for arbitrary functions q and r (see Bergstrom and Cornes (1981, 1983)). Aside from functional form the primary restrictions that are being made are that z is excluded from q() and only allowed to affect the linear coefficient on h in r(). Note that these restrictions are only necessary for the analysis because equation (1) allows both tastes and income to affect human capital decisions. If η=0 then α_s may be allowed to depend on z and if either α_s=0 or there is no variation in g then η may be allowed to depend on z without changing the basic structure of the model.

subject to the household budget constraint

$$c + p_h h = y_{Mi} + y_{Fj} \tag{3}$$

where p_h is the unit cost of human capital. The optimal level of h if the child is sex z is thus

$$h_{ij} = \frac{k_{Mi} + k_{Fj} + \delta_{Mz} + \delta_{Fz} - p_h}{2\eta p_h + \alpha_M + \alpha_F} \tag{4}$$

where $k_{si}=\alpha_s g_{si}+\eta y_{si}$. Evaluating (2) at the optimal levels of human capital and total consumption yields the sum of the couple's utilities conditional on the sex of their offspring. Taking expectations with respect to the sex of the child yields the total expected marital utility accruing to the couple if i marries j, which will be denoted U_{ij}.

It is worth noting that equation (4) yields the standard neoclassical prediction that has been used to test the unitary model (e.g., Schultz 1990), that male and female income effects on household behavior should be equal. Thus, even in a setting in which collective behavior is important one might not find evidence inconsistent with unitary behavior to the extent that transferable utility holds and the outcome in question relates to a public good. Equation (4) illustrates as well a second problem with tests of the unitary model using unearned income. Given that family background characteristics that determine unearned income such as parental wealth may affect tastes as well as income, a test of the equality of male and female income effects that does not adequately control for heterogeneity in tastes g_{si} could lead one to incorrectly reject the predictions of the unitary model.

Under the assumed model structure, efficient allocations in the marriage market require the total sum of utilities across all couples in the marriage market to be maximized (Becker 1974).[6] Henceforth assume that men and women in a given marriage market are ordered so that man i marries woman i. Then, efficient allocation of marital partners dictates that for any two couples, i and j, that are part of the same marriage market,

$$U_{ii}+U_{jj}-U_{ij}-U_{ji}\geq 0 \qquad \textbf{(5)}$$

That is, total utility cannot be increased by having these two couples exchange partners. By substituting the expressions for the U_{ij} into (5) and solving one obtains

$$k_{Mi}k_{Fi}+k_{Mj}k_{Fj}-k_{Mi}k_{Fj}-k_{Mj}k_{Fi}=\Delta_{ij}k_M\Delta_{ij}k_F\geq 0 \qquad \textbf{(6)}$$

where $\Delta_{ij}k_s=k_{si}-k_{sj}$. Equation (6) simply states that the marriage market should yield positive assortative mating on the index k, which is a linear combination of income and tastes: within a given marriage market the male with the highest value of this index should be matched with the female with the highest value of this index and so forth.[7] The intuition for the result is straightforward: because not providing children the desired level of human capital investment from the perspective of each spouse is costly and because each member of the couple must consume the same amount of this public good, welfare is maximized by coupling individuals

---

[6]In particular, this property follows from the assumption that, given the level of the public good, utility is linear in consumption. It is worth noting, however, that the key empirical implication derived using this property, that there is positive assortative mating on the index k, does not depend on this particular form: in the presence of transferable utility there is a unique efficient set of marriage-market allocations with this property (Bergstrom 1995)

[7]Note that if $\alpha_s=0$ then $k_{si}=y_{si}$ and thus there will be positive assortative mating on income as in Lam (1988) while if $\eta=0$ there will be positive assortative mating on tastes.

with similar incomes and tastes for human capital investment.

Equations (4) and (6) can be used to characterize marriage-market selection bias in a setting in which, as one might one expect in general, one does not fully observe the set of characteristics that contribute to variation in income and tastes. Specifically, for each sex s, let $k_{si}=\gamma_s\beta_s' x_{si}+\gamma_s\epsilon_{si}$ where ϵ_{si} is a mean zero unobserved component of the male taste index that is uncorrelated with x_{si}, γ_s is a scalar that determines the variance of the residual term,[8] β_s is a vector of known coefficients, and x_{si} denotes the vector of observable (to the econometrician) characteristics of the i[th] individual of sex i. Thus the human capital allocation equation (4) for offspring of couple i (i.e., of male i and female i) is

$$h_{ii}=\tilde{\alpha}_M\beta_M' x_{Mi}+\tilde{\alpha}_F\beta_F' x_{Fi}+\tilde{\delta}_z+\tilde{\alpha}\epsilon_{Mi}+\tilde{\alpha}\epsilon_{Fi} \quad \textbf{(7)}$$

where $\tilde{\alpha}_s=\gamma_s/(2\eta p_h+\alpha_M+\alpha_F)$ and $\tilde{\delta}_z=(\delta_{Mz}+\delta_{Fz}-p_h)/(2\eta p_h+\alpha_M+\alpha_F)$. It is evident that OLS estimation of equation (7) will yield biased estimates of $\tilde{\alpha}_s$ to the extent that the observable components of the income/taste indices, $\beta_s' x_{si}$, are correlated with the unobservable components, ϵ_{si}. Although by assumption $\sigma(\beta_M' x_{Mi},\epsilon_{Mi})=\sigma(\beta_F' x_{Fi},\epsilon_{Fi})=0$, it is likely that $\sigma(\beta_M' x_{Mi},\epsilon_{Fi})>0$ and $\sigma(\beta_F' x_{Fi},\epsilon_{Mi})>0$ as a consequence of the sorting in the marriage market: positive assortative mating on the index k will ensure that between, for example, two women with the same x_{Fi} the woman with the higher unobserved component ϵ_{Fi} will marry, given two potential grooms, that groom with the higher k_{Mi}, which is likely to be the man with the higher $\beta_M' x_{Mi}$.

---

[8]Differences in residual variance are introduced in this way to emphasize the fact that, as with other limited dependent variable techniques, the first stage of the analysis discussed below does not permit separate identification of the residual variance and the latent variable coefficients.

With sufficient information on the above covariances it is in general possible to assess the magnitude of and to adjust for the bias associated with marriage-market selection. In particular, let the column vectors $\mathbf{k_i} = [\beta_M{'}x_{Mi}, \beta_F{'}x_{Fi}, \epsilon_{Mi}, \epsilon_{Fi}]$, $\tilde{\boldsymbol{\alpha}} = [\tilde{\alpha}_M, \tilde{\alpha}_F]$, and $\hat{\boldsymbol{\alpha}} = [\hat{\alpha}_M, \hat{\alpha}_F]$ where $\hat{\alpha}_M$ and $\hat{\alpha}_F$ are the estimates of $\tilde{\alpha}_M$ and $\tilde{\alpha}_F$ obtained by regressing h_{ii} on the predicted taste indices $\beta_M{'}x_{Mi}$ and $\beta_F{'}x_{Fi}$ and sex.[9] Further, let

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{x\epsilon}^{\prime} \\ \Sigma_{x\epsilon} & \Sigma_{\epsilon\epsilon} \end{bmatrix} \qquad \textbf{(8)}$$

denote the variance-covariance matrix of the $\mathbf{k_i}$, where Σ_{xx} is the 2x2 matrix covariance matrix of the predicted indices for men and women, and so forth. The bias then may be characterized using the following expression

$$\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}} + \Sigma_{xx}^{-1}\Sigma_{x\epsilon}\tilde{\boldsymbol{\alpha}} \qquad \textbf{(9)}$$

Solving for $\tilde{\boldsymbol{\alpha}}$ yields

$$\tilde{\boldsymbol{\alpha}} = (I + \Sigma_{xx}^{-1}\Sigma_{x\epsilon})^{-1}\hat{\boldsymbol{\alpha}} \qquad \textbf{(10)}$$

where I is the 2x2 identify matrix. This formula shows how a consistent estimate of Σ may be used to construct consistent estimates of $\tilde{\boldsymbol{\alpha}}$ using biased OLS coefficients.

Further restrictions are required to obtain predictions about the signs of these biases. In the special case, for example, in which there is no variance in the unobservable for men $\Sigma_{x\epsilon}$ will,

---

[9]For notational simplicity the following expressions abstract from the fact that the constant $\tilde{\delta}$ may depend on the sex of the child. Assuming the sex of children is orthogonal to the k_i the bias correction would not be affected if this assumption were relaxed.

asymptotically, have only one non-zero element, the covariance of the male predicted index and the female unobservable. Under these circumstances it may be shown that the biases will have opposite signs: $\hat{\alpha}_M$ will overestimate $\tilde{\alpha}_M$ but $\hat{\alpha}_F$ will underestimate $\tilde{\alpha}_F$.

III. Estimation

While the fact that the regressors in the human capital allocation equation (7) are correlated with the residual suggests that estimation of this equation might take place using instrumental variables, the structure of the model indicates that no appropriate instruments are likely to be available. In particular, although the fact that the distribution of potential partners influences the choice of spouse might lead one to believe that measures of this distribution could serve as instruments, this is not the case because the unobserved components are known by the marriage-market participants at the time of marriage--as the marriage-market does not distinguish the observed and unobserved components of tastes, any variable that is correlated with the former will also be correlated with the latter and thus with the residual in (7).[10] I consequently make use of an alternative procedure, which in a manner analogous to the Heckman (1979) two-step selectivity correction, includes as a regressor in (7) an estimate of the conditional expectation of the residual term.

The first stage of the analysis is to estimate the equations determining the structure of the marriage market. These equations play the role of the first-stage equation in standard selectivity correction methods: in effect the marriage-market equations determines which among all possible

[10]Boulier and Rosenzweig (1984) do use marriage-market characteristics to instrument the husbands' wage in a fertility decision rule. This process works for their model because, as noted, the unobserved heterogeneity is assumed to be one sided (i.e., the only husband's attribute that is assumed to be valued in the marriage market is his wage).

spouse combinations at a given time and place actually result in marriages. Consider first a marriage market consisting of only two men and two women. The probability that the first and second men will marry the first and second women, respectively, conditional on the observables $\beta_s' x_{si}$, and assuming that the ϵ_{si} are drawn independently from an extreme-value distribution, is

$$\begin{aligned}\text{Prob}(1 \text{ marries } 1,\ 2 \text{ marries } 2) &= \text{Prob}((\beta_M'\Delta_{12}x_M+\Delta_{12}\epsilon_M)(\beta_F'\Delta_{12}x_F+\Delta_{12}\epsilon_F)>0)\\ &= \text{Prob}(\Delta_{12}\epsilon_M \le -\beta_M'\Delta_{12}x_M\ ,\ \Delta_{12}\epsilon_F \le -\beta_F'\Delta_{12}x_F)\\ &\quad + \text{Prob}(\Delta_{12}\epsilon_M > -\beta_M'\Delta_{12}x_M\ ,\ \Delta_{12}\epsilon_F \ge -\beta_F'\Delta_{12}x_F)\end{aligned} \tag{11}$$

$$= \frac{1}{(1+\exp(-\beta_M'\Delta_{12}x_M))(1+\exp(-\beta_F'\Delta_{12}x_F))} + \frac{1}{(1+\exp(+\beta_M'\Delta_{12}x_M))(1+\exp(+\beta_F'\Delta_{12}x_F))} \tag{12}$$

where $\Delta_{12}x_M = x_{M2} - x_{M1}$, and so forth. Note that (12) has two additive components reflecting the fact that the "ones" will marry each other if either both of them have higher tastes for schooling than do the "twos" or if they both have lower tastes than do the "twos". Note also that because the x_{si} variables only enter in differenced form the specification is robust to the inclusion of marriage-market specific fixed effects that might arise, for example, if the mean of the distribution of the ϵ_{si} varied across markets.

If there were a large number of distinct separate markets with two men and women each, maximum likelihood estimates of β_s could be obtained by taking logs of (12), summing over markets, and optimizing. In practice, given the criterion used to define marriage markets, as discussed below, there are likely to be more than two marriages taking place in each market. Unfortunately, however, generalizing equation (12) to the presence of more than two potential marriage partners is not straightforward because, when there are N_t couples in a marriage market, there are N_t! ways that a given set of marriage allocations can arise.[11] Therefore, in the analysis that follows I make use of an approximate procedure in which each of the binomial logit probabilities (e.g., $1/(1+\exp(-\beta_M{}'\Delta_{12}x_m)))$ in (12) is replaced with its multinomial logit analog (e.g., $1/(1+\Sigma\exp(-\beta_M{}'\Delta_{12}x_m))$)[12]. Thus the criterion function that is maximized to obtain coefficient estimates for the taste parameters is:

$$\sum_{t=1}^{T}\ln(1+\sum_{i=2}^{N_t}\exp(-\beta_M\Delta_{1i}x_M)\sum_{i=2}^{N_t}\exp(-\beta_F'\Delta_{1i}x_F)) \\ -\ln(1+\sum_{i=2}^{N_t}\exp(-\beta_M'\Delta_{1i}x_M))-\ln(1+\sum_{i=2}^{N_t}\exp(-\beta_F'\Delta_{1i}x_F)) \tag{13}$$

where N_t denotes the number of marriages taking place in marriage market t and T is the total number of marriage markets observed. Although this criterion function does not conform exactly to the structure of the model, it yields reasonable parameter estimates, as shown below, in the sense that estimated parameters obtained using simulated data do not differ significantly from

[11]This follows from the facts that in equilibrium the rank order of the men will equal the rank order of their wives and that n items may be ranked in n factorial different ways.

[12]The source of the error in the approximation here is that it asserts incorrectly that man one will marry woman one whenever there exists some i>1 and j>1 such that $k_{M1}<k_{Mi}$ and $k_{F1}<k_{Fj}$. This implication will of course only hold when there is just one other couple in the marriage market.

those used to generate the simulated data.[13]

The simulated data were constructed to correspond as closely as possible to the actual data set given the assumed model. Estimates of β_s obtained through the maximization of (13), actual values of the x_{si}, and errors ϵ_{si} drawn from an extreme-value distribution were used to assign to each man and woman in a given marriage market measures of his or her taste cum income index under the assumption that $\gamma_s=1$: $k_{Mi}=\beta_s'x_{si}+\epsilon_{si}$. Men and women in each marriage market were then matched by ranking members of each sex by their respective simulated k_{si} values so that the man with the highest k_{Mi} was matched with the woman with the higest k_{Fj} and so forth.[14]

The third stage of the analysis involves estimation of equation (7) by first constructing predicted values of the ϵ_{si} given marriage market structure and the actual matching that takes place. While analytic expressions for these measures are, to my knowledge, unavailable, it is possible to generate numerical estimates using the simulated data. As with the bias correction discussed above, the idea is that the relationship between the observables and unobservables in the actual data can be understood by examining the relationships in the simulated data. In particular, let $\mathbf{w_i}$ denote a vector consisting of summary statistics of the predicted indices ($\beta_s'x_{si}$)

---

[13]The approach could be made to yield consistency by (1) restricting analysis to two-couple marriage markets (2) selecting the parameters used to generate the simulated data to minimize the distance between the estimates from the actual and simulated data. This latter approach, which is known as indirect inference, has been formally characterized by Gourieroux et al. (1983). Because, as will be shown, the distance is small when the actual parameters are used to to simulate the data, any changes due to adoption of this procedure would likely be small.

[14]Note that in the context of the present model this approach is identical to that that would be obtained by solving a linear program in which the objective function coefficients are the marital surplus generated when male i marries female j.

in a given market (mean and standard deviations separately for women and men and the number of marriages) , the predicted indices of male and female i, all second order interactions. By regressing the vector $\mathbf{k_i}$ on $\mathbf{w_i}$ in the simulated data and applying the resulting vector of coefficients to the vector $\mathbf{w_i}$ in the actual data one obtains a predicted value, $\hat{\mathbf{k}}_\mathbf{i}$ of the vector $\mathbf{k_i}$. Because the $\beta_s'x_{si}$ are included in the list of instruments, the first two elements of $\hat{\mathbf{k}}_\mathbf{i}$ are simply the $\beta_s'x_{si}$. The second two elements, call them λ_{Mi} and λ_{Fi} , are predicted values of ϵ_{Mi} and ϵ_{Fi}., respectively. Regression of h_{ii} on $\hat{\mathbf{k}}_\mathbf{i}$, that is estimation of the equation,

$$h_{ii}=\tilde{\alpha}_M\beta_M^{\prime}x_{Mi}+\tilde{\alpha}_F\beta_M^{\prime}x_{Fi}+\tilde{\tilde{\delta}}_z+\tilde{\alpha}_M\lambda_{Mi}+\tilde{\alpha}_F\lambda_{Fi}+v_{ii} \qquad \textbf{(14)}$$

yields consistent estimates of the vector $\tilde{\alpha}$, as shown in Appendix A.[15] This procedure will subsequently be identified as the selection-correction procedure, as distinguished from the bias correction procedure (equation (10)).

;IV. Data

The data used in this paper are constructed from the records of the Demographic Surveillance System (DSS) of the International Centre for Diarrhoeal Disease Research in Bangladesh (ICDDRB) as well as a survey of educational attainment and aspirations in 1990. The population consists of 164,000 people (as of 1974) in 149 spatially contiguous villages in

---

[15]The use of the expression λ_{si} is deliberately evocative of the inverse Mill's ratio used in standard selection corrections that assume normal errors (Heckman 1979). Identification of this term comes from the fact that the distribution of marriage-market characteristics does not affect h_{ii} except through the allocation of partners, which is in turn a result of the transferable utility assumption. Note, however, that in contrast to the usual approach, the λ_{si} here are not in general consistent estimates of conditional expectations of the shocks and the actual and predicted errors do not have a joint normal distribution. Thus the v_{ii} are uncorrelated but not in general independent of $\mathbf{w_i}$. By implication, for example, using non-uniform weights to estimate (14) could yield inconsistent parameters estimates.

Matlab Upazilla, which is a rural riverine area. Individual-level, longitudinally collected birth, death and migration registration data are available for all residents of the villages starting in 1966, and all marriage information is available after 1975. Censuses were carried out in 1974 and 1982 and these provide information on household structure, education, and resource availability that can be linked to the individual-level vital events using a permanent individual identification number.

Marriage records are filed for anyone living in the study area at the time of marriage. These records report individual identification numbers on both partners as well as information on age, education and occupation. In this study I make use of the marriage records between 1975 and 1990. During this period there were 44,886 unique marriage records reported,[16] and of these marriages 45% of husbands and 58% of wives could be linked to their original households.[17]

The 1990 survey consists of a 30% sample of women from 1/3 of the 149 villages in the study area. In addition to collecting standard demographic information, the survey obtained information on the level and desired schooling for each of a woman's children under 15.

---

[16]Over the 1974-1990 period there are a number of instances in which multiple marriages are reported for the same individual. These records likely reflect (1) cases in which marriage records were available from both the husband's and the wife's households but were reported in somewhat different time periods (2) some cases in which marriage was never actually consummated (as a result, for example, of the failure of one party to meet their terms of the dowries) (3) remarriage, primarily by divorced or widowed men. In order to avoid the complexity associated with modeling these outcomes, it was decided to focus only on the first reported marriage for each woman. Marriages other than the first were included for men if they occurred more than two years following the initial marriage.

[17]The specific criterion used was that the individual recorded in a given marriage record must have appeared as a single individual in some household in the study area in either the 1975 or 1982 census before the time of marriage. This is reasonable criterion because single individuals are likely to remain resident with parents or older siblings until marriage (see, e.g., Foster 1990).

Although this follow-up survey on a population with detailed marriage information is unique with regard to its ability to address issues of marriage selectivity, it is not ideal. In particular, to estimate the model one needs a measure of human capital investment such as schooling or body size of children for all married women. Unfortunately, no anthropometric information was collected in the survey and over half of the sample of women married over the vital registration period did not have a school-aged child by the time of the survey. I therefore use as the human capital measure data on desired schooling, which are reported for all of a woman's children regardless of age.[18]

The initial stage of data preparation involved grouping the marriage records for the 149 villages over the 15-year period into distinct marriage markets. In a setting in which men and women marry others in the same village at a fixed age this would be straightforward: one could take all marriages in a given village and year as constituting a single marriage market. In fact, however, although a significant fraction of individuals marry others in the same region, men and women from the same village are unlikely to marry. Not only does this imply that the availability of marriage partners in one village may importantly affect outcomes in other villages, but given that household data are only available for residents of the study area, it also means that information is incomplete on a significant fraction of spouses.

Because the extent to which marriage-market selection bias afflicts analyses of household behavior depends, among other things, on the extent to which men and women have a choice

[18]I will assume, for purpose of exposition, that "desired schooling" can be substituted for "achieved schooling" in the estimation of equation (4). An alternative and perhaps more reasonable assumption is that desired schooling corresponds to the level of schooling that a child would receive if schooling were costless (p_h=0). Under this assumption the interpretation of the ratios of the $\tilde{\alpha}_s$ would be unaffected.

over the characteristics of their spouses, there are costs to choosing both too broad and too narrow a definition of a single marriage market. If one includes in the marriage market too few individuals relative to the relevant choice set then estimates of the correlation between observables and unobservables will be too small (e.g., it will be zero if there are no alternative potential spouses) and thus the bias computed using equation (10) will in general also be too small. On the other hand, if one includes in the marriage market potential partners that were not, in fact, part of the choice set one will tend to overestimate the extent of bias.

Given the structure of the available data and because the primary purpose of this paper is to establish whether marriage-market selection is present at all it is desirable to err on the side of underestimating the size of the relevant choice set. Therefore it is assumed that a marriage market consists of all single women from the same village who marry an individual from within the study area in a given year and their grooms. On this basis the 149 villages over a 15 year period could have had up to 2235 distinct marriage markets; in fact there were 1892. On average there were 4.2 couples per marriage market with 90% of the marriage markets having less than 10 couples.[19] Table 1 presents summary statistics from the sample of couples included in the marriage-market analysis of the education and ages of the partners and the land ownership and head's education for the households in which they resided prior to marriage. Summary statistics on the characteristics of the under-15 children from the sub-sample of the 1990 survey women

[19]Note that the inclusion in the marriage market of only those couples in which both partners are from the study area substantially reduces the average size of the marriage market. If the sample were restricted to those couples for which only the bride was from the study area there would have been 11.7 marriages per market. However, if this latter criterion had been used analysis of marriage market allocations would have been limited to consideration of the variables available on the marriage-market records.

that was also included in the marriage-market sample are also presented.

Closely related to the question of what constitutes a marriage market is the question of the process by which individuals enter the marriage market. Again, the focus of the paper dictates the use of a simplifying assumption that will tend to err on the side of minimizing the possibility of finding selection bias. In particular, it is assumed that conditional on the observables for each individual and a fixed village-level effect, allocation to the marriage market is random so that the distribution of the unobserved part of the k_{si} is not influenced by the characteristics of other marriage market entrants. This assumption minimizes the likelihood of finding selectivity bias by assuming away the possibility, for example, that men with especially high ϵ_{Mi} are more likely to marry into a given village in a year when the eligible women in that village have high desired levels of schooling (i.e., high values of k_{Fi}). It also rules out the possibility that an individual, having entered a particular marriage market and observed the attributes of the other entrants, chooses not to marry and then enters a subsequent marriage market.[20] While these restrictions are non-trivial and the resulting coefficients, particularly those relating to the effects of the characteristics x_{si} on k_{si} , must therefore be interpreted with caution, they yield an estimable model that appears to capture the essence of the problem of marriage-market selection. Moreover, as shown below, the assumptions do not appear to do much violence to the data in the sense that estimates of the extent of selection bias are internally consistent.

V. Results

---

[20] Allowing for entry and exit from the marriage market would have introduced more than the usual additional complexity because it would have been necessary to characterize how the distribution of utility within the household is influenced by marriage market structure, something that can be avoided if marriage market entry is taken as given.

a. Marriage-market allocations

The first stage of the analysis involves the estimation of the β_s coefficients relating observable characteristics of brides and grooms marrying at particular places and times to the human-capital tastes indices using equation (13). Table 2 presents the results for two sets of specifications, each of which is applied to both the actual marriage data and to corresponding simulated data. The first specification uses the same set of characteristics for men and women: own education and the landholding and education of the head in the individual's household of origin. The coefficients are estimated with a reasonable degree of precision and, for the most part, are significantly different from zero, indicating that both own characteristics and the characteristics of one's household of origin substantially affect one's choice of spouse. The results are also remarkably comparable between men and women. For example, for both sexes land ownership by head of the household in which one resided prior to marriage has a concave effect on the income-cum-taste, with the effect of a .1 decimal increase in land ownership evaluated at mean land being .75 and 1.01 for brides and grooms, respectively.

Interpretation of these coefficients is difficult because the variance terms γ_s are not identified and because the index k_{si} combines both income and tastes. It would be inappropriate for two reasons to infer on the basis of these coefficients, for example, that land ownership of brides and grooms contribute almost equally to income y_{si} . First, the total effect of land ownership on income depends also on the coefficient γ_{si}, which has been normalized out of the estimated equations. Second, the effect of land ownership on the index k_{si}, especially for brides, may operate primarily through its implications for the bride's taste (or her parents' tastes given arranged marriage) for schooling.

Some insight into the magnitudes of these effects may nonetheless be gained by considering a hypothetical marriage market in which there are two observationally equivalent (to the econometrician) males who therefore each have a 50% probability of being matched to the woman in that market with the higher k_{Fi}, depending on the values of their unobservable components. Under these circumstances providing one of the males with an increase from the mean land owned to the mean plus one standard deviation would result in him having an 88% of being matched with the high-index woman. The schooling effects are of a similar magnitude: providing one of the males with incomplete (complete) primary schooling would give him a 79% (97%) probability of being matched with the higher-index woman. Similarly, a one standard deviation increase in land owned would raise the probability of an observationally equivalent woman being matched with the higher taste man from 50% to 79.6%, while providing a woman with incomplete (complete) primary schooling would yield a 75.8% (89.3%) probability of attracting that man.

The only substantial difference between the men's and women's estimated β_s coefficients appears to be in how the head's schooling is evaluated in the marriage market. In particular, men with high income/tastes seek out both women who are educated and women who come from households where the head is educated, while women with a high taste for the public good are primarily concerned with the schooling of their potential spouse not with that of the head of his household.[21]

---

[21]The finding that high index men are likely to seek out women from households with educated heads supports the conclusion of Lam and Schoeni (1993) that the strong effects of a man's father-in-law's schooling on his wage reflects assortative mating. Unfortunately they do not consider the effects of in-law's schooling on women's wages; the findings here suggest that these effects are likely to be small net of husband's education.

Because a significant approximation had to be made in order to derive equation (13), it is important to establish whether the procedure does indeed produce reasonable parameter estimates. This was done by applying the procedure used to estimate the coefficients in column 1 to the simulated data set constructed based on these coefficients. The resulting coefficient estimates are presented in the second column of Table 2 and indicate that the procedure is quite effective in the sense that the estimates correspond closely to the values of the underlying parameters that they estimate (i.e., the column 1 parameter estimates). Not only are the point estimates and t-ratios quite similar in the two columns, but an overall test that of the hypothesis that the column (2) parameter estimates are equal to the values used to generate the simulated data indicates the two are not statistically different (χ^2_{12}=11.3; P-value=0.5).

In order to evaluate the extent of selectivity bias that would arise if only a more limited set of covariates were available, a second specification, henceforth called the "reduced specification", was estimated that excludes the characteristics of the brides' households. A second simulated data set, based on the resulting coefficient estimates, was also constructed. The resulting estimates are presented in columns (3) and (4) of Table 2 and are reasonably comparable to those provided by the full specification. As before the estimates based on the actual and simulated data are not significantly different (χ^2_8=9.8; P-value=0.28). As might be expected given that schooling is positively associated with land ownership the coefficients on bride's schooling increase when schooling is the only available predictor of the brides' incomes and tastes. While the increase in the primary schooling coefficient for brides appears reasonable, the estimated coefficients for secondary schooling are less satisfactory: both increase dramatically relative to that in the previous specification and the coefficients are estimated very imprecisely. This instability likely reflects the relatively low degree of completed primary schooling in this population (see Table 1) coupled with the increased importance of education as an explanatory variable when the brides' household characteristics are excluded.

b. Simulated data covariances

The next step in the analysis is to use the predicted indices $\beta_s{'}x_{si}$, errors ϵ_{si}, and allocations from the simulated model to examine the extent of bias likely to arise in the estimation of household decision rules given the structure of the model. As noted, this bias arises because the process of selection in the marriage market induces a positive correlation between the male observable and the female unobservable and vice versa. This pattern is clearly evident in Table 3, which presents the variances and covariances of the relevant variables for the two simulated data sets.[22] As expected the correlation between the male and females observables and their respective unobservable components is not significantly different from zero. By contrast the cross covariances are strongly significant, yielding, for the full specification, a correlation of .335 between the female residual and the male predicted index and a correlation of .243 between the male residual and the female predicted index. As expected the reduced specification, by decreasing the fraction of variation in the index k_{Fi} attributable to observable characteristics of the brides, increases the relative importance of the bride's unobservable as a source of correlation between bride's and groom's indices. In particular, the correlations between the female residual and the male predicted index in the reduced specification is higher (.344 vs. .335) and the

---

[22]In order to make these estimates relevant to those presented in Table 4, these covariances are estimated net of village-level fixed estimates. The consequent removal of the mean shock by village explains the fact that the estimated variance of the shocks is substantially less than the true variance of residuals generated from an extreme value distribution (1.64).

opposite cross-correlation is lower (.121 vs. .243).

The bias adjustment equations based on the respective covariances for the two simulated data sets and equation (10) are presented in the lower panel of Table 3. As noted these equations indicate how one can obtain consistent estimates of the underlying parameters given the biased estimates that will be obtained using OLS. Suppose, for example, that the OLS coefficients happen to be equal, that is $\hat{\alpha}_M=\hat{\alpha}_F$. Then the equations for the first panel indicates that the OLS coefficient estimate $\hat{\alpha}_M$ overstates the true value of $\tilde{\alpha}_M$ by 40% while the coefficient $\hat{\alpha}_F$ understates the true value by 12%. Thus even if the effects of the predicted index β_s'x_{si} for mothers substantially exceeds that for fathers, it may appear that the coefficients are the same if selectivity is ignored.

The bias-correction formulae for the reduced specification do not differ qualitatively from those derived for the full specification. In particular assuming equal estimated coefficients as before we may infer that the OLS estimate $\hat{\alpha}_M$ overstates the true value of $\tilde{\alpha}_M$ by 13% while the coefficient $\hat{\alpha}_F$ understates the true value by only 2%. While it is interesting that these biases are small, it is worth noting that this conclusion depends significantly on the assumption that the uncorrected coefficient estimates are equal. If the OLS estimate $\hat{\alpha}_M$ were only half that of $\hat{\alpha}_F$ then these equations imply that $\hat{\alpha}_M$ would overstate the true value by 35%.

c. Estimates of desired schooling decision rules

Table 4 presents estimates of the desired schooling decision rules (equation (7)) with and without corrections for marriage-market selectivity.[23] The first column of numbers presents the

[23]Each of these specifications is estimated using fixed effects for the brides' villages and includes maternal age and age squared. Hausman tests rejected the random effects specification in favor of village fixed effects for the selectivity corrected estimates using the full specification

uncorrected estimates obtained by regressing desired schooling on the predicted indices for the full specification, where the latter were constructed using the corresponding marriage-market allocation coefficient estimates from Table 2. Interestingly, the coefficients on the male and female indices are comparable. Indeed, they are not significantly different (P-value=0.51). Because the values of the β_s obtained from the marriage-market allocation part of the estimation are also comparable this result seems to suggest that male and female characteristics have similar effects on desired schooling. For example, taking the marriage market entered by a woman and her choice of spouse as given, increasing a woman's schooling from zero to incomplete primary results in a .480x1.14=.55 year increase in desired schooling; the comparable figure for men is also .55.

The exercise carried out in the previous sub-section, however, suggests that these estimates may be substantially biased. Indeed, application of the formulae presented in Table 3 yields a rather different picture (column 2). While the male coefficient estimate falls relative to that in column 1 by 35% to .267 the female coefficient estimate rises by 16% to .559. Thus it appears that the effect of female incomplete primary schooling on desired schooling of children is about twice that of the effect of male incomplete primary schooling.

While it is instructive to show that the model, if correct, predicts that there will be substantial bias in the estimated male and female coefficients, there is reason to question whether the model conforms sufficiently to the true underlying process that the covariation between the

---

(χ_7^2=15.5, P-value=0.03) but not the reduced specification (χ_7^2=9.14, P-value=0.24) . The maternal age (at the time of the 1990 survey) effects are included to allow for life-cycle effects in reported desired schooling that do not affect marriage allocations. These effects may also capture trends in schooling aspirations across cohorts.

predicted indices and residuals in the simulated data, which may be observed directly, is a reasonable estimate of that in the actual data. Moreover, even if the resulting covariances are correctly estimated, the estimated biases may be incorrect if the respective (by sex) residuals and predicted indices do not have the same coefficient as predicted by the model (equation (7)).

An alternative set of estimates were therefore obtained based on equation (14). The idea, as noted, is to regress the male and female residuals from the simulated data on a set of marriage market characteristics given the allocations arising in those data [24] and then to apply the resulting coefficients to the same set of statistics in the actual data to obtain the λ_{si}. The third column in Table 4 presents these estimates.

The results conform strongly to those presented in the second column. The estimate of the male index effect is .246, which is substantially closer to the .267 obtained using the simulated data covariances than the uncorrected figure of .412. Similarly, the female index effect is .608 which corresponds to the estimate of .559 obtained using the simulated data covariances. Moreover, in contrast to the result for the uncorrected estimates, the hypothesis that the two sets of coefficients are equal is rejected (P-value=0.03).

There are several other features of this specification that provide support for the model. First, the selection-term coefficients are significantly different from zero (P-value=0.06) indicating that the marriage-market characteristics that influence the allocation of partners in the simulated data are significantly correlated with the residuals in the desired schooling equation as they should be given marriage-market selection. Second, the hypothesis that the coefficients on the respective (by sex) predicted index and selection terms are equal as predicted by the model

[24] The estimates are presented in the Appendix B.

(equation (14)) is not rejected (P-value=0.61). By imposing this constraint and reestimating the model new estimates of the $\tilde{\alpha}_s$ are obtained (column 4) that conform even more closely to those in column (2). Third, the model implies that the relative values of the β_s coefficients for a given sex from the marriage-market specification should equal the relative values of the $\tilde{\alpha}_s\beta_s$ that would obtain if the corresponding x_{si} were included as regressors in equation (14) instead of the predicted indices. This latter model was therefore estimated and the hypothesis tested taking the β_s from the marriage-market specification as given. The implied constraints were not rejected (χ^2_{10}=14.31; P-value=.16).

Given that both the bias correction and selectivity approach make use of information on simulated residuals it is reasonable to question whether the two approaches can really distinguish the proposed model from plausible alternatives. Of particular interest is whether misspecification in the form of incorrect predictions about the correlations between observables and unobservables, arising, for example from incorrect characterization of the marriage market, will differentially affect the two measures.

Evidence on this point was obtained by applying the estimation procedures to simulated data that reflect this potential form of misspecification. In particular, a hypothetical set of marriage markets, each with only two pairs of partners, were simulated. Couples in a prespecified fraction of the marriage markets were randomly matched, while those in the remainder of the markets were sorted according to their respective k_{si} indices. Measures of h_{ii} were then simulated with, $\tilde{\alpha}_M$ and $\tilde{\alpha}_F$ set to .2 and .5, respectively.[25] . OLS as well as the two forms of corrected

[25] Assumptions were designed to conform roughly to the parameter estimates in Table 4. Fifty simulations were run for each level of the share of random matching, with each of the simulations having 4000 marriage markets consisting of two men and two women. Coefficient

estimates were then obtained for the simulated data using only the information on the observable component of the k_{si} for each individual in each marriage market, the predicted marriages, and the predicted h_{si}.[26] Because the correction procedures incorporates the perfect assortative mating result, they are misspecified whenever some fraction of the marriage markets exhibit random matching.

Averages of the resulting coefficient estimates appear in Figure 1. When the simulated data contain no random matching (i.e., share=0), the OLS estimate for males is quite distant from its true value, while the bias corrected estimate, for both sexes are very close to their true values (panel 1). The selection correction estimates (panel 2), show that this procedure also works well under this scenario--the average estimates of the index and the λ_{si} parameters for both sexes correspond closely to their true values.

As the share of random matching increases, however, substantial bias afflicts both the bias corrected estimate for males in panel 1 and the λ_{si} coefficients in panel 2 for both sexes but the male and female index coefficients from the selection correction remain largely unaltered.

---

estimates are averages across the fifty simulations. For each simulation, individual k_{si} indices were constructed for each participant by adding together two scaled extreme value distributions, corresponding to the observable and unobservable components. The extreme value shocks for the female unobservable was scaled by a factor of .3, while the male unobservable and the observable for both sexes were scaled by a factor of .1.

[26]For the purpose of this simulation it is assumed that the β_{si} are known and not estimated. This assumption aids clarity but is potentially misleading because one might expect the β_{si} estimates in practice to be affected by the presence of random matching. Simulations comparable to those presented here allowing for estimation of the β_{si} yielded similar results for random matching probabilities in the range of 0-40%. For higher levels of random matching, however, imprecision in the β_{si} estimates substantially reduced the power of the test. The fact that the β_{si} from the actual data are well measured indicates that random matching, if present at all in marriage markets in rural Bangladesh, is in the lower range.

For a sample size of 8000 couples in 4000 marriage markets with these parameter values, the 50 simulations indicate that the difference between the male index and selection terms with 10 and 20% random matching are .0737 (s.d.=.0447) and .117 (s.d.=.0428), respectively. These imply probabilities of 95% and 99.7%, respectively, of rejecting the null of perfect assortative mating according to the model when the alternative is true.

Note also that in n the case of 100% random matching, the male bias corrected estimate and the λ_{si} for both sexes yield an average of zero, which is substantially below the true values of .2 and .5. Not suprisingly, the OLS coefficients are close to .2 and .5 for men and women respectively under 100% random matching--in this case no selectivity will be present.

The fact that the parameter estimates from the two approaches are differently affected by the presence of random matching indicates that the comparison of correction estimates from the two estimates, and indeed, comparison of the two sets of estimates for the selectivity correction approach, has some ability to identify misspecification due to inappropriate characterization of the structure of marriage markets that affects the correlation between observables and unobservables. The facts that the two approaches in Table 4 yield similar estimates and that the two components of the selection correction are not significantly different suggests that the correlations and biases predicted by the developed model are reasonable approximations.[27]

The results for the reduced specification, while in some respects less satisfactory, provide

[27]It is easily seen that similar patterns emerge if the unobservable components affected matching but, in contrast to the model, did not affect human capital allocations--that is the coefficients on the ϵ_{si} in (7), but not those on the $\beta' x_{si}$, are 0. Although there would in that case be no selection bias associated with OLS and equation (14) would correctly identify the relevant parameter values, the bias correction (equation (10)) would still correct for the bias dictated by the model and thus yield inconsistent estimates of the coefficients on the $\beta' x_{si}$. Thus, the comparison has power against this form of alternative as well.

further support for the notion that marriage-market selection effects are significant. They also provide insight into the extent of bias likely to arise under circumstances in which more information is available on the pre-marriage household characteristics for one spouse (typically the husband in a patriarchal society) than the other.

It is first of all evident from columns (5)-(8) in Table 4 that limiting information on the women to schooling substantially changes the values of the estimated coefficients relative to those for the full specification. This difference should be interpreted with care, however. As a consequence of the fact that the marriage-market estimation does not identify the variance of the residual, the coefficient $\tilde{\alpha}_s$ on the predicted index $\beta_s'x_{si}$ increases proportionately with γ_s.[28] Because less information is available in the reduced specification to predict variation in the women's index, k_{Fi}, the coefficient γ_F that captures the variation in the women's residual should increase relative to that in the full specification as, therefore, should $\tilde{\alpha}_F$.[29]

With regard to the extent of bias introduced by the correlation of the regressors with the residuals, the picture that emerges from the reduced specification is again similar to that in the full specification. First, based on a comparison of columns (5) and (6) it appears that the uncorrected estimate overestimates the value of the male coefficient by 55% and underestimates the value of the female coefficient by 13%. Second, the coefficient estimates obtained using the selection terms (column 7) are comparable to those obtained using the simulated data covariances (column 6). Third, the coefficients on the selection terms are both significantly different from

---

[28]See the expression following equation (7).

[29]Decreases in the male coefficient are, of course, not as easily explained as there is no reason to expect γ_M to fall when fewer characteristics on the wife are available.

zero.

The only part of the reduced specification that is problematic in view of the model is that, in contrast to the results of the full specification, the test that the respective index and selection term coefficients are equal is rejected (χ_2^2=8.28; P-value=0.02). Apparently the simulated data do not correctly predict the degree of selectivity relative to that which actually obtains in the desired schooling data. While the exact reason for this misspecification cannot be identified without further exploration, one possible source is the assumption that, conditional on the observables, individuals are randomly allocated to marriage markets. Reduction in the observable information available for some market participants presumably increases the extent to which this assumption is violated in practice and thus may lead to an underestimate of the degree of selection operating in the marriage market.

In summary, while there are some differences in terms of percentage bias in the two cases, the qualitative implications are the same: marriage-market selection results in overestimates of the male coefficient $\tilde{\alpha}_M$ and underestimates of the female coefficient $\tilde{\alpha}_F$. Indeed, given the estimated covariances in Table 3 it may be shown using the bias correction formulae (10) that the ratio $\hat{\alpha}_F/\hat{\alpha}_M<\tilde{\alpha}_F/\tilde{\alpha}_M$ when $\hat{\alpha}_F/\hat{\alpha}_M>0.49$ for the full specification and when $\hat{\alpha}_F/\hat{\alpha}_M>0.70$ for the reduced specification. This result is of potentially significance for the interpretation of unitary-model tests of the equality of male and female income effects. If, for example, in an uncorrected regression one finds evidence that the mother's income effect exceeds that of the father, as is the case for the health equations estimated in Thomas (1990), then these patterns of bias suggest that correcting for selectivity will yield an even greater difference.[30]

Two significant caveats must be mentioned, however. (1) The relative magnitudes of the $\tilde{\alpha}_s$ estimates in this study for men and women do not necessarily correspond to the relative magnitudes of the coefficients on the x_{si} variables, that is $\tilde{\alpha}_s \beta_s$, which is what is being estimated in a regression of child health or schooling on parental characteristics.[31] (2) The covariances used to calculate the bias apply to a particular marriage market structure and set of observable covariates and thus are likely to vary across regions and across specifications that include different observable characteristics.

Although the model does not permit direct identification of the values of the γ_s parameters and thus of the sex-specific variances of the residual components of the k_{si} indices, the patterns of bias for the two specifications suggest that the dominant source of unobserved heterogeneity in this model is with respect to the female unobservable. In particular, the finding of a positive male bias and a negative female bias is consistent with equation (9) under the assumption of no variance in the male residual discussed above. The fact that the magnitudes of the bias in percentage terms[32] increase in the reduced specification, which effectively increases the variance of the female unobservable, further supports this conclusion. The fact that higher

---

[30]I assume here for purpose of exposition that the only source of bias is marriage-market selection, ignoring, for example, the possibility that an individual's unearned income may be correlated with his or her tastes.

[31]This caveat would seem to be of particular relevance in models comparable to the reduced specification where, for example, the ratio β_F/β_M for incomplete schooling is 1.9. In this case it is possible that $\hat{\alpha}_F \beta_F / \hat{\alpha}_M \beta_M > 1$ but that $\tilde{\alpha}_F \beta_F / \tilde{\alpha}_M \beta_M < 1$.

[32]The absolute bias for men is, of course, smaller for the reduced specification. The percentage bias is the relevant figure because it is not influenced by the incidental parameter γ_s.

female variance should yield greater upward bias for the male coefficient is also evident in Figure 1 which, as noted, assumed built in a higher unobservable standard deviation for women relative to men by a factor of 3:1.

Finally, the fact that the estimates of the correlation of observables and unobservables across couples appear to be internally consistent does not necessarily imply that the notion of marriage market used here, i.e., those men and women marrying in a particular village in a particular year, is an accurate characterization of the set of potential partners open to each spouse. As the selection-correction approach makes clear, the correlation between a given male observable and a female unobservable and vice-versa is importantly affected by the mean and variance of the characteristics of other couples in the marriage market. It appears at least possible that the procedure works because variation in these characteristics at the time and village of marriage reasonably captures variation in the true marriage market faced by a given man or woman.

VI. Conclusions

In this paper I have developed a methodology for looking at the implications of marriage market structure for analyses of the effects of parental characteristics on desired schooling. This approach has been implemented using a unique data set from rural Bangladesh that contains longitudinal information on marriage in a contiguous area that can be linked to censuses containing information on parental characteristics as well as information on schooling aspirations. The results show that marriage-market selection has a substantial influence on estimates of the effects of parental characteristics on desired schooling. In addition to supporting the notion that human capital investment in children plays an important role in the process of marriage allocations in this population, the results also indicate that there is substantial systematic variation in income and/or tastes for human-capital investment that is not captured by the observed individual and household characteristics.

This paper also has broader implications for analyses of household behavior. As noted in the introduction, the primary motivation for this particular paper was a concern that the existing literature testing the unitary model of household behavior has not addressed the potential selectivity problems that arise if household membership is a choice variable, a problem that seems especially salient in the context of marriage. The fact that selection effects importantly influence the estimated desired schooling decision rules suggests that many of the results of this literature must be interpreted with caution, although it appears that, at least in this population, biases operate to reduce the probability of finding evidence against the unitary model.

Nonetheless, it would be inappropriate for this paper to be read as a defense of the unitary model. The structure of the model examined in this paper is that of the collective household. While expenditure on human capital is not influenced by the structure of the marriage market except through its effects on the allocation of partners, the model implies that the distribution of consumption within the household is constrained by the structure of the marriage market (see Becker 1973). While the limits of the currently available data in this population[33] do not permit implementation of a test of the effects of marriage market structure on household allocations, the fact that the model as specified appears to do a reasonable job of characterizing the operation of the marriage market not only suggests that such effects may well be present but also provides a

[33]This constraint will be removed shortly as an intensive data collection exercise in this population is currently being carried out by the author in collaboration with Jane Menken (U of Pennsylvania), Omar Rahman (Harvard University), and Paul Gertler (RAND).

framework for testing for such effects using micro-level information on marriage-market structure. Indeed, the model presented in this paper may provide a useful framework for examining a variety of aspects of household behavior that interact importantly with the process of household formation.

Appendix A

Selectivity Corrected Simulation Estimator

Let $\boldsymbol{k_i}=[\beta_M^{\prime}x_{Mi},\beta_F^{\prime}x_{Fi},\epsilon_{Mi},\epsilon_{Fi}]$ and $\mathbf{w}_i$ be a vector consisting of the $\beta_s^{\prime}x_{si}$ for s=[M,F], the number of couples, the mean and standard deviation of the $\beta_s^{\prime}x_{si}$ indices for all men and women in the marriage market of individual i, and all second order conditions. Further, let **K** and **W** denote matrices of $\mathbf{k}_i$ and $\mathbf{w}_i$ from the actual data and **K*** and **W*** the corresponding matrices in the simulated data. Finally let $\boldsymbol{a}=[\tilde{\alpha}_M,\tilde{\alpha}_F,\tilde{\alpha}_M,\tilde{\alpha}_F]$ and **h** denote the vector of human capital allocations in the actual data. Then, if **K** were fully observable, consistent estimates of ***a*** could be obtained using OLS:

$$\hat{\boldsymbol{a}}^{OLS}=[\boldsymbol{K}^{\prime}\boldsymbol{K}]^{-1}\boldsymbol{K}^{\prime}\boldsymbol{h} \qquad \textbf{(15)}$$

or, assuming the ϵ_{si} are correlated with the $\mathbf{w}_i$ as will be the case given assortative mating, a less efficient instrumental variables procedure could be used:

$$\hat{\boldsymbol{a}}^{IV}=[\boldsymbol{K}^{\prime}\boldsymbol{W}(\boldsymbol{W}^{\prime}\boldsymbol{W})^{-1}\boldsymbol{W}^{\prime}\boldsymbol{K}]^{-1}[\boldsymbol{K}^{\prime}\boldsymbol{W}(\boldsymbol{W}^{\prime}\boldsymbol{W})^{-1}\boldsymbol{W}^{\prime}\boldsymbol{h}] \qquad \textbf{(16)}$$

As the matrix **K** is not fully observable, a selectivity corrected simulation estimator is proposed that involves regressing the **K*** on **W***, constructing predicted values by applying the resulting coefficients to **W**, and then regressing **h** on the resulting predicted values. As the predicted values may be written $\boldsymbol{W}(\boldsymbol{W}^{*\prime}\boldsymbol{W}^{*})^{-1}\boldsymbol{W}^{*\prime}\boldsymbol{K}^{*}$, the proposed estimator is:

$$\boldsymbol{a}^{IV*}=[\boldsymbol{K}^{*\prime}\boldsymbol{W}^{*}(\boldsymbol{W}^{*\prime}\boldsymbol{W}^{*})^{-1}\boldsymbol{W}^{\prime}\boldsymbol{W}(\boldsymbol{W}^{*\prime}\boldsymbol{W}^{*})^{-1}\boldsymbol{W}^{*\prime}\boldsymbol{K}^{*}]^{-1}[\boldsymbol{K}^{*\prime}\boldsymbol{W}^{*}(\boldsymbol{W}*^{\prime}\boldsymbol{W}*)^{-1}\boldsymbol{W}^{\prime}\boldsymbol{h}] \qquad \textbf{(17)}$$

Because $\text{plim}\ (\boldsymbol{K}^{*\prime}\boldsymbol{W}^{*}/N-\boldsymbol{K}^{\prime}\boldsymbol{W}/N)=0$ $\text{plim}\ (\boldsymbol{W}^{*\prime}\boldsymbol{W}^{*}/N-\boldsymbol{W}^{\prime}\boldsymbol{W}/N)=0$ where $N=\sum N_t$., $\hat{\boldsymbol{a}}^{IV}$ and

$\hat{\boldsymbol{a}}^{IV*}$ are easily seen to have the same limiting distribution. Thus $\hat{\boldsymbol{a}}^{IV*}$ consistently estimates $\boldsymbol{a}$.

References

Becker, G. (1973) "A Theory of Marriage, Part I", *Journal of Political Economy*, 81: 813-47.

Becker, G. (1974) "A Theory of Marriage, Part II", *Journal of Political Economy*, 82: S11-S26.

Bergstrom, T. (1994) "A Survey of Theories of the Family", Manuscript.

Bergstrom, T. And R. Cornes (1981) "Gorman and Musgrave Are Dual-an Antipodean Theorem on Public Goods", *Economi Letters,* 371-378.

Bergstrom, T. And R. Cornes (1983) "Independence of Allocative Efficiency from Distribution in the Theory of Public Goods", *Economi Letters,* 371-378.

Bergstrom, T. And D. Lam (1993) "The Effects of Cohort Size on Marriage-markets in Twentieth-century Sweden", *the Family, the Market and the State of Ageing Societies*, Ermisch and Ogawa, Eds., Oxford University Press.

Boulier, B. and M. Rosenzweig (1984) "Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior", *Journal of Political Economy,* 92:712-732.

Browning, M., F. Bouruignon, P.A. Chiappori, V. Lechene (1994) "Incomes and Outcomes: a Structural Model of Intrahousehold Allocation", *Journal of Political Economy*, 102 (6): 1067-96.

Foster, A. D. (1993) "Household Partition in Rural Bangladesh", *Population Studies*, 47(1): 97-114, March.

Gourieroux, C., A. Monfort, and E. Renault, "Indirect inference", *Journal of Applied Econometrics,* Dec 1993, v8(Supplement): S85-S118 .

Heckman, J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, 47 : 153-

161.

Lam, D. (1988) "Marriage Markets and Assortative Mating with Household Public Goods", *Journal of Human Resources*, 234: 462-87.

Lam, D. and R. F. Schoeni (1993) “Effects of Family Background on Earnings and Returns to Schooling : Evidence from Brazil”, *Journal of Political Economy,* 101(4):710-740.

McElroy, M. (1990) “The Empirical Content of Nash-Bargained Household Behavior”, *Journal of Human Resources*, 25: 559-83.

Schultz,T.P. (1990) "Testing the Neoclassical Model of Family Labor Supply and Fertility, *Journal of Human Resources,* 25(4): 599-634.

Strauss, J. and D. Thomas (1995) “Human Resources: Empirical Modeling of Household and Family Decisions”, Jere R. Behrman and T.n. Srinivasan (Eds.), Handbook of Development Economics, Volume 3. Amsterdam: North Holland.

Thomas, D. (1990) “Intra-Household Resource Allocation : An Inferential Approach”, *Journal of Human Resources*. 25(4): 635-664.

Udry, C. (1995) “Gender, Agricultural Production and the Theory of the Household”, Manuscript.

Table 1
Descriptive Statistics

	Means	S.D.
Data from Marriage Sample		
Bride's HH Land (decimals)	11.57	15.77
Bride's Schooling		
1-5	.345	.475
6+	.012	.107
Bride's HH Head's Schooling		
1-5	.312	.464
6+	.129	.334
Bride's Age	18.4	3.08
Groom's HH Land (decimals)	11.41	17.10
Groom's Schooling		
1-5	.377	0.485
6+	.154	0.360
Groom's HH Head's Schooling		
1-5	.285	.451
6+	.111	0.314
Groom's Age	24.75	4.66
Data from 1990 KAP Survey		
Achieved Schooling of Children<15	0.236	0.799
Desired Schooling of Children<15	9.461	3.751
Girls	.477	0.500
Age	4.899	3.276

Table 2
Maximum Likelihood Estimates of Human Capital Taste Indices Using Actual and Simulated Marriage Data[a]

	Full Specification		Reduced Specification	
	(1) Estimates Based on Actual Marriages	(2) Estimates Based on Simulated Marriages[b]	(3) Estimates Based on Actual Marriages	(4) Estimates Based on Simulated Marriages[b]
Bride's HH Land (decimals)	0.834	.716	--	--
	(3.57)[c]	(4.86)		
Bride's HH Land ^2	-.035	-.030	--	--
(decimals)	(0.79)	(3.18)		
Bride's Schooling				
1-5	1.14	0.894	1.71	1.52
	(4.35)	(3.76)	(6.41)	(5.71)
6+	2.12	2.12	13.7	6.03
	(2.16)	(1.57)	(0.05)	(0.02)
Bride's HH Head's Schooling				
1-5	.552	.745	--	--
	(2.28)	(3.16)		
6+	1.28	2.24	--	--
	(2.98)	(5.21)		
Groom's HH Land (decimals)	1.12	1.27	.604	.569
	(4.20)	(4.26)	(2.71)	(2.55)
Groom's HH Land ^2	-.050	-.077	-.024	-.009
(decimals)	(3.86)	(3.28)	(1.31)	(0.49)
Groom's Schooling				
1-5	1.34	1.81	0.899	1.64
	(3.73)	(3.83)	(2.57)	(4.70)
6+	3.58	6.70	10.8	6.75
	(3.62)	(2.77)	(0.23)	(0.14)
Groom's HH Head's Schooling				
1-5	.004	-.427	-.122	-.821
	(0.01)	(1.12)	(0.32)	(2.16)
6+	0.094	-.074	-.934	-.802
	(0.16)	(0.11)	(1.58)	(1.35)
Test of equality of actual and simulated coefficients	$\chi_{12}^2=11.3$ P=0.50		$\chi_8^2=9.8$ P=0.28	

a. Based on 7892 marriages in 1893 marriage markets.
b. Simulations based on parameter estimates from previous column
c. Asymptotic t-ratios in parentheses

Table 3
Covariances of Predicted Taste Indices and Residuals by Sex and Correction Formulae Based on Simulated Data Sets[a]

	Full Specification				Reduced Specification			
	(1) $\beta_M' x_{Mi}$	(2) $\beta_F' x_{Fi}$	(3) ϵ_{Mi}	(4) ϵ_{Fi}	(5) $\beta_M' x_{Mi}$	(6) $\beta_F' x_{Fi}$	(7) ϵ_{Mi}	(8) ϵ_{Fi}
Male Predicted Taste Index ($\beta_M' x_{Mi}$)	3.76				14.97			
Female Predicted Taste Index ($\beta_F' x_{Fi}$)	1. 27 (26.77)[b]	1.55			1.20 (20.72)	.794		
Male Residual from Taste Index (ϵ_{Mi})	.078 (1.57)	.198 (6.34)	1.36		.085 (0.92)	.126 (7.01)	1.36	
Female Residual from Taste Index (ϵ_{Fi})	.752 (17.24)	.015 (0.52)	.367 (13.24)	1.34	1.55 (15.85)	.005 (0.24)	.440 (15.94)	1.36

Formulae for corrected estimates given respective covariances

	Full Specification	Reduced Specification
$\tilde{\alpha}_M$	$1063\hat{\alpha}_M - .590\hat{\alpha}_F$	$1033\hat{\alpha}_M - .146\hat{\alpha}_F$
$\tilde{\alpha}_F$	$-.214\hat{\alpha}_M + 1340\hat{\alpha}_F$	$-.213\hat{\alpha}_M + 1236\hat{\alpha}_F$

a. Based on 1185 marriages to women from 141 villages.
b. T-ratios based on heteroskedasticity-consistent standard errors in parentheses.

Table 4
Estimates of Desired Schooling Decision Rules
With and Without Corrections for Marriage Market Selectivity[a]

	Full Specification				Reduced Specification			
	(1) Uncor-rected	(2) Corrected Using Simulated Data Covariance	(3) Corrected Using Estimated Selection Term	(4) Index and Selection Coefficients Set Equal	(5) Uncor-rected	(6) Corrected Using Simulated Data Covariance	(7) Corrected Using Estimated Selection Term	(8) Index and Selection Coefficients Set Equal
Male Index (β_M'x_{Mi})	.412 (8.27)[b]	.267 (3.74)	.246 (2.81)	.252 (3.77)	.199 (8.89)	.128 (4.34)	0.084 (2.80)	.117 (4.48)
Male Selection Term (λ_{Mi})			.032 (0.13)				.707 (2.80)	--
Female Index (β_F'x_{Fi})	.480 (6.68)	.559 (5.46)	.608 (6.19)	.581 (6.20)	.534 (5.29)	.617 (4.92)	. 597 (5.93)	.715 (7.89)
Female Selection Term (λ_{Fi})			.630 (2.36)				1.07 (6.24)	--
Sex of Child (Girl=1)	-1.17 (8.10)		-1.17 (8.10)	-1.17 (8.10)			-1.18 (8.13)	-1.19 (8.17)
Male and Female Index Coefs. Same?	χ_1^2=0.42 P=0.51		χ_1^2=4.46 P=0.03	χ_1^2=4.71 P=0.03	χ_1^2=9.75 p=0.00		χ_1^2=20.88 P=0.00	χ_1^2=31.17 P=0.03
Index and Selection Term Coefs. Same?			χ_2^2=1.00 P=0.61				χ_2^2=8.28 P=0.02	
Selection Coefs. Zero?			χ_2^2=5.58 P=0.06				χ_2^2=40.54 P=0.00	

a. Based on 2404 children from 1185 marriages to women from 141 villages. Each specification also includes maternal age and age squared and women's village dummies.
b. T-ratios derived from heteroskedastic consistent standard errors accounting for multiple observations per woman are reported in parentheses.

Appendix B

OLS Estimates of Residual Conditional on Marriage
Given Own and Marriage Market Characteristics

		Male Residual		Female Residual	
		Coefficient	Abs. T-ratio	Coefficient	Abs. T-ratio
1	$\beta_M' x_{Mi}$	0.240	2.984	0.747	10.015
2	$\beta_F' x_{Fi}$	0.116	0.751	-0.228	1.596
3	Mean($\beta_M' x_{Mi}$)	0.068	0.705	-0.561	6.250
4	Mean($\beta_M' x_{Mi}$)	-0.217	1.261	0.275	1.726
5	S.D.($\beta_M' x_{Mi}$)	-0.320	2.856	0.148	1.421
6	S.D.($\beta_F' x_{Fi}$)	0.288	1.486	-0.410	2.279
7	N	0.013	1.682	-0.002	0.258
	1 x 1	0.046	3.291	0.017	1.329
	1 x 2	-0.011	0.506	-0.061	3.111
	1 x 3	-0.003	0.114	0.030	1.204
	1 x 4	-0.126	3.324	-0.008	0.219
	1 x 5	-0.099	2.555	-0.198	5.536
	1 x 6	-0.039	1.249	0.010	0.357
	1 x 7	-0.001	0.342	-0.001	0.457
	2 x 2	-0.133	3.603	0.051	1.491
	2 x 3	-0.057	1.237	0.096	2.246
	2 x 4	0.233	2.643	-0.097	1.182
	2 x 5	-0.006	0.109	-0.095	1.791
	2 x 6	0.176	2.490	0.098	1.488
	2 x 7	0.003	0.650	0.001	0.116
	3 x 3	-0.072	2.895	-0.069	2.975
	3 x 4	0.205	3.549	-0.022	0.402
	3 x 5	0.161	3.267	0.226	4.953
	3 x 6	-0.015	0.240	-0.078	1.360
	3 x 7	-0.022	4.424	0.016	3.551
	4 x 4	-0.101	1.411	0.033	0.493
	4 x 5	-0.058	0.831	0.054	0.836
	4 x 6	-0.201	1.921	-0.011	0.117
	4 x 7	0.014	1.757	-0.003	0.475
	5 x 5	0.028	0.644	-0.087	2.169
	5 x 6	0.016	0.225	0.131	1.994
	5 x 7	0.001	0.112	-0.015	2.419
	6 x 6	0.154	2.069	0.039	0.567
	6 x 7	-0.009	0.724	0.008	0.739
	7 x 7	0.001	2.129	0.000	0.102
	Constant	0.256	2.401	0.227	2.294
		$F_{35, 2368}=4.38$; $p = 0.000$ $R^2 = 0.06$		$F_{35,2368}=14.96$; $p = 0.000$ $R^2= 0.18$	

Figure 1
Simulated Effects of Model Misspecification: Average OLS, Bias Corrected, and Selectivity Corrected Estimates of Parental Effects on Child Human Capital By Share of Marriage Markets Assumed to Exhibit Random Matching. True effects are .2 for Males and .5 for Females.