NBER WORKING PAPER SERIES

BROTHERS AND SISTERS IN THE FAMILY AND THE LABOR MARKET

John Bound

Zvi Griliches

Bronwyn H. Hall

Working Paper No. 1476

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 1984

The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #1476
October 1984

Brothers and Sisters in the Family and the Labor Market

ABSTRACT

This paper investigates the relationship between earnings, schooling, and ability for young men and women who entered the labor force during the late 60s and 70s. The emphasis is on controlling for both observed and unobserved family characteristics, extending a framework developed earlier by Chamberlain and Griliches (1975) to the analysis of mixed-sex pairs of siblings. Using the National Longitudinal surveys of Young Men and Young Women, which drew much of the sample from the same households, we were able to construct a sample containing roughly 1500 sibling pairs. For several reasons, particularly the need to have data on two siblings from the same family, only one third of these pairs had complete data; this fact led us to develop new methods of estimating factor models, which combines the data for several "unbalanced" covariance matrices. We use the data on different kinds of sibling pairs (male-male, female-female, and male-female) together with these new methods to investigate the question of whether family background, ability, or "IQ" are the same thing for males and females, in the sense that they lead to similar consequences for success in schooling and in the market place. With a simple two factor model to explain wages, schooling and IQ scores, we are able to test whether these factors are the same across siblings of different sexes and whether the loadings on the two factors are similar. The conclusion is that the unobservable factors appear to be the same and play the same role in explaining the IQ and schooling of these siblings, while there remains evidence of differences once they enter the labor market.

John Bound
National Bureau of
Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

Zvi Griliches
National Bureau of
Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

Bronwyn H. Hall
National Bureau of
Economic Research
204 Junipero Serra Boulevard
Palo Alto, CA 94305

Revised September 1984

Brothers and Sisters in the Family and Labor Market[1]

John Bound, Zvi Griliches, and Bronwyn H. Hall

1. Introduction

Most of the earlier work on earnings functions and returns to schooling estimation has been done with male data. Much of the more recent work on the analysis of male-female wage differentials has focused on labor-force participation questions and the correct measurement of work experience and has bypassed the family background-ability-schooling debate which had been conducted largely on the basis of data on males. There are a number of stylized facts and conclusions which have emerged from these literatures: In the schooling-ability-family background area the conclusion seemed to be that, at least as far as measured IQ and measured family background variables were concerned, their absence did not bias greatly the estimated schooling coefficients in male earnings functions (see Griliches, 1977, and Hauser and Daymont, 1977). The same conclusion also could be reached as far as unmeasured family background is concerned, but here the results were much more sensitive to potential errors in the data which are magnified when within siblings contrasts are used for estimation (see Behrman, et al. 1980; Griliches, 1979). As far as male-female compairsons were concerned, the estimated schooling coefficients in wage equations appeared to be somewhat higher for females than for males while the estimated age coefficients were lower for women than for men. These differences were greatly reduced but not entirely eliminated when work experience was allowed for. More attention to the quality of the work experience and expectations about labor force attachment reduced the estimated average male-female differentials somewhat further, without eliminating most of the original differential (see Becker, 1983; Mincer and Polachek, 1974; Sandell and Schapiro, 1976; and Shackett, 1981). To the extent that the question of "ability bias" was investigated using female data, the conclusions did not differ greatly from those reached using male data.

These debates neither posed clearly nor resolved the question of whether "family background," "ability," or "IQ" are the same thing for males and females, in the sense that they lead to similar consequences for success in schooling and in the market place. Some of the observed differences in market outcomes could arise from a different distribution of abilities across the sexes, different rewards in the labor market to these abilities, and different investment responses by family and individuals.

It is not clear whether such questions can be pursued successfully with the available data. Ideally we would like to have more detail (a series of different test scores) and a longer horizon (life cycle data) than is usually available in the standard economic surveys. Nevertheless, we would like to open up this question and explore which aspects of it might be answerable with currently available data. We were motivated to pursue this topic by the apparent puzzle thrown up, in passing, in Joyce Shackett's thesis (Harvard, 1981). She found that holding schooling and measured IQ constant, there is still an unaccounted for correlation in wages between brothers and between sisters, indicating the presence of an unmeasured family related component of "ability" or marketable human capital. But when she examined brother-sister pairs in a similar fashion, their wage residuals were essentially uncorrelated, suggesting the possibility that "abilities" are either distributed differentially among males and females or priced differently in the market.

To check such conjectures and to interpret them in a broader context, we have updated Shackett's data and extended the framework developed earlier by Chamberlain and Griliches (1975 and 1977) to the analysis of mixed-sex pairs. Our analysis is based on the NLS Young Men and Young Women tapes which contain information on roughly 1500 sibling pairs (male, female, and mixed) over the 1966-1980 period, including IQ test scores for about two thirds of the individuals. Unfortunately, the data are rarely complete for both members of a sibling pair. Only about one third of the pairs (about 150 to 200 pairs each) have complete data on all the variables of interest. This has led us to adopt and develop new methods of estimating such models, combining data from several "unbalanced" moment matrices, i.e., matrices with rows and columns missing (corresponding to the variables for which data are missing in the particular observational subset).

We cannot really test directly the hypothesis that "abilities" are distributed differently across males and females or that they are priced differently, without having information on a number of different test scores for both men and women. What we can do is, first, to check whether the observed empirical fact persists in a more complete unobserved factors model which allows both IQ and schooling to be measured with error; second, to investigate whether this cross-sex difference appears only in wages or can be traced back to the earlier IQ-schooling relationship; and finally, we can ask whether the data imply the presence of more than one ability factor in the sense that the male and female versions of the ability factor are not perfectly correlated.

The basic approach of this paper is to specify a relatively simple model with two common factors for the observed data (test scores, schooling, and two wages: early and late), one factor reflecting unobserved "ability" and the other measuring common endowments across siblings which are orthogonal to ability, e.g., wealth. This model is estimated on data for brother-brother, sister-sister, and brother-sister pairs, allowing both the factor loadings and the factors themselves to differ across the sexes. Using this framework, it is possible to test whether the factor structure is alike for males and females, in the sense that the estimated factor loadings are similar for the two sexes, and whether the male and female factors are the same, that is, have a correlation of unity.

In implementing our model we have chosen to sweep out all of the other exogenous variables contained in these equations, both to simplify the computations and because our samples of men and women have not been drawn in a completely identical fashion; for example, the survey of men begins in 1966 and that for women in 1968. Accordingly we have removed age, race, region, city residence and the constant freely from all of the dependent variables and separately for males and females. Thus, the main male-female difference in the level of wages is already taken out in the first pass at the data and is not explained by the model. The focus of this paper is on the differences in the structure and influence of the unmeasured family components across the two sexes.

The plan of the paper is as follows: First, we outline briefly a simple model of IQ, schooling, and wages in the context of sibling data and explain what we are after. Second, we describe our data and outline the specific estimation problems caused by the relatively high frequency of missing data for one or both of the siblings. Third, we present the results of estimating the complete model and then discuss the results of testing the equality of the factor structure across siblings. Finally, we venture some conclusions relating to the more general topic of male-female differences in earnings.

2. The Model

Consider the standard earnings equation

$$LW = \alpha + \beta S + \gamma I + \delta X + u \tag{1}$$

where LW is the logarithm of wage rates or earnings per some time unit, S is the level of schooling, I is a score on an "intelligence" test, X represents a set of other variables which we shall not consider explicitly here, snch as age, race, and region; α, β, γ, and δ is a set of parameters to be estimated, and u represents all other unmeasured determinants of wages, including unmeasured but relatively permanent differences in human capital levels across individuals and transitory fluctuations and measurement errors in wages and other variables. The usual discussion in this area (e.g., Griliches, 1977) proceeds to focus on the estimation of β, the "rate of return to schooling," in the presence of a number of potentially complicating circumstances: the lack of a good "ability" variable and/or the use of a particular error-prone test score as a proxy for it; and the possibility both of errors of measurement in achieved schooling levels and of endogeneity, in the sense that schooling may be chosen in anticipation, and with the knowledge, of some of the components of u (which is unobservable to the analyst). As stated, β is unidentified in this model in the absence of additional instrumental variables such as measured background variables which would affect S and I without themselves entering the LW equation directly. In this context, sibling data are interesting because they provide another way of identifying β by using the sibling values of S and I as instruments. Earlier work of this type focused primarily on male siblings (see Griliches, 1979, for a review) and this is one of the first papers to look also at sister and brother-sister pairs (see also Scarr and MacAvay, 1982).

In work that focuses on male-female wage differentials, the question is often whether the estimated differences in α and β can be explained by incorrect measurement of the components of X (such as different meanings of work experience for the two sexes) or by different components of u -- the omitted factors (see Mincer and Pollachek, 1974; and Becker, 1983). While sibling data cannot be used to identify and interpret what these unobservable components "really" are, they can be used to ask whether the family components are, to any extent, sex specific.

Consider the following simplified factor model for IQ, schooling, and wage:

$$(2) \quad \begin{aligned} I &= \gamma_1 A + u_1 \\ S &= \gamma_2 A + \eta W + u_2 \\ LW &= \beta S + \gamma_3 A + u_3 = (\beta\gamma_2+\gamma_3)A + \beta\eta W + \beta u_2 + u_3 \end{aligned}$$

where the story differs from the earlier one [eq. 1] in having "swept-out" in an unconstrained fashion the other X variables to simplify both exposition and computation. The model contains an unobservable ability factor A, for which I (an IQ-type score) is an error prone proxy. "Ability" affects achieved schooling levels and may also enter the wage equations directly, above and beyond its indirect effect via schooling. In addition there is a "wealth" factor W, which affects only schooling directly. The following notational definitions and no-correlation assumptions are made:

$$(3) \qquad EA^2 = a^2 \qquad EW^2 = w^2 \qquad Eu_j^{\ 2} = \sigma_j^{\ 2}$$

$$EAW = EAu_j = EWu_j = 0$$

$$Eu_1u_j = 0 \quad \text{for } j \neq 1$$

$$Eu_2u_3 = \sigma_{23}$$

The statements above reflect the following assumptions: The A and W factors are orthogonal, i.e., W is the "wealth" component that is above and beyond that part of wealth that is already correlated with the ability factor. (That A and W are orthogonal is a convenient normalization. Some such rotational assumption is required for the separate identification of the factor coefficients.) These factors are assumed to be independent of all the equation specific disturbances. The error in the test score u_1 is a pure measurement error untransmitted to other equations and uncorrelated with the other disturbances. Because S may be measured with error in (2), or may be chosen endogenously, u_2 is allowed to be freely correlated with u_3.

As written, and in the absence of additional instrumental variables or restrictions, this model is heavily underidentified. This can be most easily seen by counting the number of unknown parameters -- nine, relative to the number of the observed variances and covariances, which is only six.

It is the availability of sibling data which allows us to identify the parameters of such a model. Denoting pair members by a and b or m and f subscripts, and treating them symmetrically (i.e., we assume that siblings have the same variances and coefficients, at least as long as they are of the same sex), we make the following additional assumptions:

(4) $A = f + g, \quad Ef^2 = 1, \quad Eg^2 = \tau, \quad W_1 = W_2, \quad w^2 = 1$

$$Eu_{1a}u_{1b} = Eu_{2a}u_{2b} = Eu_{1a}u_{2b} = Eu_{1a}u_{3b} = 0$$

$$Eu_{2a}u_{1b} = Eu_{2a}u_{3b} = Eu_{3a}u_{2b} = 0$$

$$EWu_{jk} = 0 \qquad \text{for } j = 1,\ldots,3 \text{ and } k = a,b$$

$$Eu_{3a}u_{3b} = \sigma^2_{ab}$$

which imply the following: A is a factor with a family variance components structure with f representing the family component and g the individual one. We normalize so that the variance of f is one and the variance of g is τ. W, on the other hand, is a pure family factor with no individual components and is normalized to have a variance of one. All of the cross-sibling correlations in I and S are assumed to be captured by the two family components f and W, and hence u_2 is not correlated with the other sibling's u_3, though it is allowed to be freely correlated with its own. The residuals in the wage equations are allowed, however, a free family structure.

Note that, under the condition that we do not distinguish between siblings, we are adding six covariances but only two parameters and the model is now identified. Figure 1 makes clear where identification comes from. Factor loadings, the schooling coefficient β and the cross-sib wage covariance are all identified within the cross-sib matrix with the own-sib covariances then identifying the individual residual variances and covariances.[2] The model is recursive with the cross-sib IQ covariance identifying γ_1, the schooling covariances identifying γ_2 and η, and then the wage covariances identifying γ_3, β and σ_{3ab}.

The above is a variant of the standard way of identifying the schooling coefficient in a wage equation, using a proxy for ability and instrumenting both schooling and the proxy with family background variables. An advantage of setting up the model in terms of covariance matrices rather than a standard IV setup is that then it can be easily generalized to allow for another index for the sex of the sibling. We assume that the model specified above applies to each sex separately but that there may be a sex specific component to each factor. This implies that the factors will be less than perfectly correlated, and introduces two additional parameters, ρ_A and ρ_W representing the correlation between the male and female version of each factor. We also allow for free correlations across the brothers' and sisters' wages. The bottom panel of Figure 1 shows the cross-sex cross-sib covariance matrix implied by this model. Note that the factor loadings are also assumed to be different for the two sexes.

To test the hypotheses mentioned in the introduction, we ask, essentially, how well a factor structure identified within the brother and sister pairs separately can rationalize the cross-sex cross-sib covariance matrix. With the model as specified, it is not too difficult to fit the same-sex covariance matrices since we are fitting 12 covariances with 11 parameters but the test on the cross-sex matrix is more stringent. We add nine covariances but only three parameters

(ρ_A, ρ_W and σ_{3mf}). The sequence of tests we will use is the following: First, the test of equality across the sexes of the wage covariances is a test of whether there are still significant differences in the family effect after controlling for ability. Second, we test whether ability is priced differently for men and women by testing the equality of the factor loadings. Finally the test that ρ_A and ρ_W are unity is a test that the factors have no sex specific structure.

Before we turn to a more detailed description of our data and estimation procedures, several additional points should be mentioned: the use of age instead of experience in our list of predetermined variables and the interpretative differences this implies, the use of two wage variables, and the non-use of measured family background variables. Most of the work in this area (e.g., Griliches 1977 and Mincer 1974) uses accumulated work experience as a variable in the wage equation and defines the schooling coefficient as estimating the effect of schooling "holding work experience constant." Experience is usually entered in a non-linear fashion and is a function of age, schooling, and other factors which determine the post-school labor force participation and employment experience of an individual. From our point of view this interpretation of experience is endogenous to the achievement model. Given the potential nonlinearity of its effect, it would be rather difficult to extend our models to incorporate it explicitly. We can think then of our model as one in which this variable has been solved out, leaving one of its determinants, age, among the predetermined X variables. But since the usual schooling coefficient estimates are based on equations of the form $bS + d(Age - S - 6)$, our results are to be interpreted as estimating $(b-d)S + dAGE$. Thus, to compare our estimated schooling coefficients β to earlier estimates in the literature requires the addition of the estimated age coefficient to them.

This paper differs from our earlier efforts (Chamberlain-Griliches, 1975 and 1977) by including two wage variables in the model, early and late. We do not focus ,however, on the wage or earnings growth profiles explicitly (on that, see Chamberlain, 1978, for example). Moreover, since we do not include work experience in the wage equations directly, we do not constrain either the schooling coefficients or the ability coefficients to be the same in the two wage equations. Implicitly, this allows for an age-schooling interaction in the wage equation, which we could not allow for explicitly.

It also differs from some of the other papers in this area by not including measured family variables such as father's occupation and mother's education in the equations to be estimated. Using sibling data they are subsumed instead in the unobservable family factors f and W. One might be tempted to use them also in a more elaborate MIMIC type model, but the model to be used by us is already straining our computational resources and the ability of the data to discriminate between its various slightly different versions.

3. Data and Variables

Our data come from the National Longitudinal Survey of Young Men (1966-1980) and Young Women (1968-1980). (See Center for Human Resource Research, 1979, for a detailed description.) These surveys started with about 5000 respondents each, and were down to about 4000 interviewees each by the end of the last decade (the attrition is for such reasons as death, inability to locate, and refusal to answer). When these surveys were originally designed (including the Older Men and Mature Women panels), they were chosen in a stratified random fashion from a larger underlying household sampling frame. This has led to the presence of a number of same household members within and across different panels. In particular, it is possible to identify approximately 703 households with at least two brothers, 668 households with at least two sisters, and 1075 with at least one brother-sister pair. The cohorts covered were originally 14 to 24 years old in 1966 for males, and 14 to 24 years old in 1968 for females. The latest surveys available to us at the time this analysis was initiated followed them through 1980 with the age of respondents ranging from 28 through 38 for males and from 26 through 36 for females.

We have tried to use the data for all the individuals who finished schooling before or during the survey periods and for whom we could construct the requisite data. We use data from three points in these surveys: (1) First interview data (1966 for men, 1968 for women) for age, race, and IQ test scores collected from the respondents' high schools[3] (missing for about one-third of the sample). (2) Schooling level achieved at completion of school (in years) and wage received on an "early job" (after leaving school, not before age 18 and around age 22 if data are available, later if the school leaving age was higher) and other associated variables at that juncture (age, region, city size, and marital status). And (3) a "later" wage (around age 28, but at least three years later than the early wage) with the same set of associated variables as of that date. The rules we followed in selecting our observations and constructing our variables are described in greater detail in Appendix B.

Table 1 shows the sample sizes which resulted when we made various cuts on good data and gives some idea of the relatively small fraction of our observations which contains data on siblings. Among the original 10,000 or so respondents, it was possible to identify about 1600 pairs or roughly 3000+ individuals who had a sibling in one of these surveys. By the time we ask that both siblings should have completed school, had observations on both an early and later wage and data on IQ scores, we are down to less than one third of the original number: about 520 pairs or 1040 individuals (see the first line of the bottom panel of Table 1). The major attrition occurs due to missing IQ scores and missing late wage (due to attrition from the sample, late school leaving, or non labor force participation). Overall attrition is slightly higher for males than females.

From the point of view of our model, we are missing data for two quite different reasons: first, because of the usual problems with sample attrition and nonresponse, many observations have missing values for one or more variables. Second, each male or female in the sample may or may not have both a brother and a sister from which we can obtain a full set of covariances. It turns out that both these problems can be solved in the same way, enabling us to use the maximal amount of the available data, rather than restricting the estimation to the subsample which is complete. We describe the methodology for obtaining such estimates in the next section of the paper, and focus here on more general data selection problems and sample description.

Table 1 shows that we are relatively short on complete data and on data for same-sexed pairs. Our data selection strategy was designed around this. First, for families with only one or two individuals in the original sample (most of our data) the assignment to a particular matrix was unambiguous. For families with three or more siblings, however, we were forced to make selections to avoid using individuals more than once. We ordered sibs by data availability and then assigned all the complete data pairs we could to the brother-brother and sister-sister complete data pairs. The remainder of the complete data pairs were assigned to the cross-sex matrix. All the remaining siblings were either assigned to a pair with some data missing, or if no data remained on their sibling, they were placed with the residuals and treated as individuals. The consequence of this procedure was to leave us with a nearly balanced design in terms of the number of brother, sister, and brother-sister pairs in the data. Families are sometimes represented more than once, but for the vast majority this means that a non-matched individual rarely has sibs in the sib-pair matrices.[4]

This process yielded 24 different moment matrices with the observations and data patterns given in the bottom panel of Table 1. Each person from the original sample who has a good observation on completed schooling has been placed in one of these matrices. In section 4 we describe how we combined the information in these different matrices when estimating the model.

Table 2 gives the means of the variables in our data. There are no surprises in the male-female differences: the average male wage is higher, and seems to grow somewhat faster (with a caveat due to the changing sample) and the male variances are higher for our key endogenous variables. Because the original surveys oversampled blacks, our samples have a significantly larger non-white proportion (.29) and more respondents in the South (.36) than is true of the general U.S. population. Given that non-whites tend to have larger families, this is even more so for our sibling data. Except for including race and region as conditioning variables we have made no further adjustments for this discrepancy from national representativeness.

The table also shows that the average age of our respondents is 23 at the early wage date and 27 at the later one. This is still quite early in their labor force careers and just before or approaching Mincer's (1974) "overtaking" point. Thus, our results have to be interpreted remembering the relative youth of these respondents.

In the next section, we describe the method of estimation which we used; it essentially involves fitting our model to several matrices of variances and covariances of the data simultaneously. Because of this, each additional variable we include tends to be rather expensive in terms of computational costs. This has led us to preprocess the variables of interest by regressing each of them on a set of exogenous variables and using the residuals from these regressions to form the covariance matrices from which we estimate the parameters of interest. From MaCurdy (1981) we know that the estimates of the parameters of the covariance matrix (including the structural coefficient β) which are obtained conditional on these regression estimates are consistent and asymptotically normally distributed with a covariance matrix which does not depend on the fact that we preprocessed the variables in this way. We give the details of these first stage regressions in Appendix B; briefly, the variables we removed were the appropriately dated race, age, and region of residence variables (at the initial survey date for schooling and IQ, at the date of the observation for the wages) and dummies corresponding to the data sample (that is, the covariance matrix) into which an observation falls. These dummies adjust for missing data which may be randomly missing conditional on the unobservables but still not randomly missing unconditionally.[5]

4. Econometric Methodology

The model we are estimating can be thought of as consisting of eight equations (four "dependent" variables -- I, S, LW1 and LW2, for each of the two siblings). A version of this model with only one wage variable is depicted in Figure 1. If one assumes that conditional on the exogenous X's (which have been swept out freely by the preprocessing) the observed variables are distributed according to a multivariate normal distribution, then the observed moment matrix is a sufficient statistic. Figure 1 gives the expected values for the components of this matrix conditional on the correctness of our assumed model.

Many econometric models can be written in the form $\Omega(\theta)$, where $\Omega(\theta)$ is the true population covariance matrix associated with the assumed multivariate normal distribution, and θ is a vector of parameters of interest. Denote the observed covariance matrix by S. Then maximizing the likelihood function of the data with respect to the model parameters comes down to maximizing

$$(5) \qquad \ln L(\Omega \mid S,\theta) = k - (n/2)[\ln |\Omega(\theta)| + \mathrm{tr}\ \Omega(\theta)^{-1}\ S]$$

with respect to θ. If θ is exactly identified, the estimates are unique and can be solved directly from the definition of Ω and the assumption that S is a consistent estimator of it. If $\Omega(\theta)$ is overidentified, then the maximum likelihood procedure "fits" the model $\Omega(\theta)$ to the data S so as to maximize the likelihood. This can be done either using the LISREL program (Joreskog and Sorbom 1981) or the MOMENTS program (B. H. Hall 1979). If the observed variables are multivariate normal this estimator is the full information maximum likelihood estimator for this model. Even if the data are not multivariate normal but follow some other distribution satisfying mild regularity conditions[6] with $E(S|\theta) = \Omega(\theta)$, this is a pseudo- or quasi-maximum likelihood estimator yielding a consistent estimator of θ. In this case, however, the asymptotic variance of the estimator is somewhat more complicated to compute and the standard programs do not produce the correct answer. A later version of this paper will contain estimates of the standard errors which are robust to nonnormality of the data.

This is fine for a random sample from the underlying population with all the variables present. But what is to be done if for one-third of the sample one is missing measurements on one of the variables (say I) or with observations which have no sibling data at all? In such situations one can think of the observed matrix S for one or more of the relevant sub-samples as missing one (or more) rows and columns.

There is no conceptual difficulty in generalizing the sample matrix approach to a multiple sample situation where the resulting $\Omega_j(\theta_j)$ may depend on somewhat different parameters. As long as the different matrices can be taken as arising independently, their respective contributions to the likelihood function can be added up, and as long as the θ_j's have parameters in common, there is a return from estimating them jointly. This can be done either utilizing the multiple samples feature of LISRELV (see Allison, 1981), or by extending the MOMENTS program (Hall, 1979) to the connected-multiple matrices case. The estimation procedure combines these different matrices and their associated pieces of the likelihood function, and then iterates across them until a maximum is found. A more detailed description of the mechanics of this approach is given in Appendix C.

The main assumption required for the consistency of this approach in the context of missing data is our ability to treat the various sub-samples as independent pieces of the likelihood function. That is, we have to assume no significant sample selection or self-selection problem, treating our data as if the missing pieces are **missing at random.** This does not mean that the expected value of missing data is the same in all the matrices, only that (in the newer terminology of Rubin 1976 and Little 1982) the data generation process is **ignorable** in the sense that the desired parameters can be estimated consistently from the complete data subsets and that "missing data" methods use the rest of the available data only to improve the efficiency of such estimates.[7]

To be more precise, the distribution of the missing data must be, conditional on the distribution of the available data, independent of the fact that it is missing. This condition justifies integrating the full likelihood over the distribution of the missing data to get a marginal distribution for the partially observed data. The marginal distribution, sharing parameters of the original, can add information to our estimates even when not all would be identified in the partially observed data alone.

While these conditions are unlikely to hold exactly in practice, we do expect them to hold approximately. The presence or absence of siblings is likely to be random with respect to the parameters of interest to us. Attrition and labor force participation (especially for young women) is likely to be non-random with respect to the unobserved wage components, but earlier work on sample selectivity bias in both of these areas (Griliches, Hausman, Hall, 1978; Smith, 1980) has not uncovered a consistent and large biasing effect. While we do know that IQ is not missing randomly in an overall sense, conditionally on our X's and the unobserved factors it too may be missing at random.

We shall proceed assuming that it is indeed legitimate for us to pool these various matrices. It would be possible to investigate the issue further, but we shall not do that here. Under the maintained assumption, our parameter estimates should change little as we include more data. We have estimated the model using various amounts of the incomplete data and have found fe qualitative differences. As an example, results using only the complete dat are reported in an appendix.

5. Results

Before we proceed to examine the full model results it is useful to look briefly at simple least squares estimates on these data and to examine the residual correlation matrices for our main variables, by sex and across siblings, to get an impression of the type of results one may expect to get with these kinds of data and models. As mentioned earlier, all of the estimation in this section has been done with variables from which the mean effects of time, age, urban and southern residence, race, sex, and data presence have been removed using unconstrained reduced form regressions.

Table 3 gives the ordinary least squares and instrumental variable estimates of a standard earnings equation for the brothers and sisters separately. In order to highlight the differences in our estimates which are due to the estimation method and those which are due to the use of IV techniques, we show three different sets of estimates. The first two columns are OLS estimates based on all those observations which had complete data on schooling, IQ, and two wages. The next two columns show OLS estimates obtained by pooling across several matrices containing all our data, including those observations which are missing IQ and/or one or more wages. The point estimates do not change that much, and the standard errors go down by about twenty or thirty percent, which is somewhat less than the forty or fifty percent which would be predicted by the increase in the number of observations alone. The last two columns are instrumental variables estimates obtained with the combined data sample, using the sibling's IQ and schooling as instruments. Since most of our sample do not have siblings, these estimates are effectively based on a much smaller number than the number of observations shown in the table.

The OLS estimates of the schooling coefficients are relatively low, but when they are combined with the age coefficient from the reduced form regression, we obtain more conventional estimates, .061 and .059 for males and 0.096 and .069 for females, similar to those already in the literature (see Shackett, 1981, and Sandell and Shapiro, 1974, among others). Instrumenting both schooling and IQ raises the schooling coefficient by as much as four or five percent in rate of return units but at the price of much larger standard errors on both coefficients, due both to the reduction in effective sample size and the usual increase from IV.

Table 4 gives the correlation matrices for our main variables (net of the previously swept out exogenous variables) for our combined data siblings sample, showing both the individual correlations and the cross-sib ones. These matrices are pairwise combinations of the set of 24 matrices for which we obtain maximum likelihood estimates in Table 5. Taking LW2 as the variable of primary interest, the observed cross-sib wage correlations are quite low: .11, .34, and 0.07 for brother, sister, and brother-sister pairs, respectively. While the general pattern is similar to that observed earlier by Shackett, (.18, .22, and 0.00), we find less of a contrast between same sex and opposite sex cross-sib correlations. The pattern in the male and female matrices appears to be very similar, except for somewhat higher correlations for the females and correspondingly higher variances for the males. In fact, the covariance matrices appear more similar than the correlation matrices. The other difference which can be seen in this table is a higher ratio of individual to family variance for the men, a finding which is confirmed by our estimates later on (compare the diagonals of the two cross-sib matrices).

Table 5 gives the maximum likelihood estimates of our model on all of the available data for each of the sexes, based on the combination of data from 24 matrices. These matrices were created by considering two dimensions of "missing": missing data and missing siblings. First we have individuals that have (1) complete data on all variables, (2) are missing IQ scores, (3) are missing wages, and (4) are missing both wages and IQ scores. Second, we have three types of siblings (male, female, and opposite) with matching data missing patterns and an extra matrix where only one wage of one sibling is missing. The intersection of these two dimensions yields nine matrices for each sex and six for the male-female pairs. The actual distribution of the data across these matrices was given in Table 1. The final results in Table 5 are based on a combination of information from 579 sibling pairs and 3262 additional individuals for males and 557 siblings and 4732 individuals for females.

The model for which estimates are presented in Table 5 is the model given by equations (2)-(4) and Figure 1, with the addition of a second wage variable. Since the coefficients on the wage variables are not constrained and there is a free correlation between wages both within individuals and across siblings, this additional wage variable imposes no new constraints on the model, but merely provides another, later indicator of the individual's lifetime income. In estimating this model in its most general form, we allowed both for different (correlated) female and male factors and for different loadings on these factors across the sexes. The estimated correlations for the two factors were 0.97 (.07) and 0.90 (.16) for the ability and wealth factors respectively and the $X^2(2)$ statistic for a correlation of unity across male and female factors was 0.8; accordingly, we have constrained the factors, but not the factor loadings, to be the same in the results presented. The estimates of the other parameters are not affected materially by this constraint.

The first part of the table gives the estimated coefficients, standard errors, and residual variances while the second part lists the estimated covariances across equations and across siblings. The final panel in this table shows also the estimated wage covariances for the cross-sib pairs. The method of estimation was maximum likelihood and the standard errors reported are the conventional estimates.[8]

There are a number of remarks about these results: (1) The estimated factor loadings for both unobservable factors, A and W, are quantitatively and statistically very similar for males and females ($X^2(5) = 6.6$). The estimated taus (the ratio of individual to family variance components of the ability factor) do seem to be different, implying a higher overall contribution of the ability factor to male success, but also, simultaneously, a relative larger role of the family component for women in this story. These differences, however, are only marginally significant, with an estimated t statistic of 1.5.

(2) The role of the "ability" factor in the wage equation is marginal, both in the sense that its coefficients are not significantly different from zero and in the sense that it contributes little to the explanation of the variance of wages. In fact, the model in general adds little (about .01 out of 0.15) to the explanation of the variance of wages once we have swept out the exogenous variables.

(3) The schooling coefficients are not estimated very precisely. If the relevant age coefficients from Appendix B are added to them, the resulting estimates are 0.094, 0.063 and 0.122, 0.069 for LW1, LW2, and males and females respectively. In spite of the fact that the contribution of the "ability" factor in the wage equation is not well defined, it appears to be multi-collinear with schooling, with the schooling coefficients falling when the estimated factor coefficients are higher. This basic result is the same as what we saw in the OLS-IV contrast in Table 3: using the sibling's IQ and schooling as instruments increases the estimated schooling coefficient but also greatly increases the standard errors on both IQ and schooling since the parts of IQ and schooling which are correlated with the sibling variables are more collinear in the wage equation.

(4) There is no significant pattern in the residual covariances reported in the second part of Table 5 except for the own serial correlation between early and late wages, which is estimated at about 0.4. Besides this, the only covariances which appear to be significant are those across the wage residuals of the same-sex siblings. This is the same effect we noted in the data in Table 4; in these estimates about half of the higher late wage covariance between sisters is explained by the stronger family component of the ability factor (both on its own in wages and via schooling) while the remainder appears in the differing estimates of the residual covariance (.025 versus .016). The difference between the estimated cross-sex wage covariance and the same-sex covariances has not been explained by the ability-schooling components of these variances - the estimated covariances are as far apart as in the original correlation matrix. However, a test for the equality of the wage covariances across all the siblings is not rejected due to their small size and fairly large standard errors ($X^2(7) = 6.8$).

All of the tests based on estimates in Table 5 depend on the particular identifying restriction we chose (the second factor appearing only in the schooling equation and not in IQ). We can ask, however: how many common factors are needed to rationalize the cross-sib correlations independently of this restriction or any particular rotation. Depending on whether we include wages or restrict attention to just IQ and schooling, two or three common family factors should be enough to fully rationalize the same sex cross-sib correlations, but if there were sex-specific components of "ability", "wealth", or wages, we would expect to need more than these two or three to fit the brother-sister correlations. Again we find no indication of sex-specific effects. Using the complete data subset only, two factors adequately explain the IQ-schooling correlation ($X^2(3) = 0.14$) and three adequately explain the IQ, schooling, wage correlations ($\chi^2(6)$ =0.88 or 2.28 depending on whether we use early or late wages. Since by allowing free correlation of the wages across the siblings we have effectively allowed for a third family factor in the estimating model, the factor analysis results confirm our finding that the unobserved family factors may be treated as the same across male and female siblings.

Each of these approaches leads us to essentially the same conclusion: At least as far as the IQ-schooling nexus is concerned, the unobservables that we can estimate play similar roles in accounting for the observable data and appear to be the same constructs for males and females. Families and schools treat brothers and sisters symmetrically, as far as we can discern using the rather gross measures of IQ scores and years of schooling completed.

The labor market story is somewhat different, however. We know already that the schooling, age, and race coefficients differ between males and females. Beyond that it is hard to discern other differences in returns to the unobservable, non-schooling and IQ related components of human capital. There is a slight indication of such differences in the asymmetry of the cross-sex cross-sib correlations. A sister's IQ and schooling is more helpful in predicting her brother's wages than vice versa, implying that those components of female IQ and schooling which are correlated with their brother's success in the labor market are less useful in predicting their own success. Nevertheless, these effects are small and not very significant either by statistical or substantive criteria. A difference of 0.1 in correlation can account for little of the overall variance in the difference between male and female experiences in the labor market.

6. Conclusion

The main finding of this paper is that the family effects in the IQ-schooling-wage relationship are essentially sex-blind. This result is particularily strong for the IQ-schooling relationship, where the observed differences in the data can be accounted for by a higher within family variance among men of the single unobserved ability factor. Although we are also able to accept equality of the unobserved factors when fitting the wage equations, conclusions here are much less robust since most of the systematic variation in wages is taken out when exogenous factors are controlled for and our model is able to explain very little of the remaining variance. One of the other questions this paper was designed to answer was whether we could gain precision in our estimates of sibling models by using missing data techniques. In comparing estimates on the complete data (which are given in Appendix A) to those based on the combined sample of 24 matrices containing roughly four times as many observations, we find that the standard erros did go down in many cases by a factor of two. However, for some crucial parameters such as the wage covariances, they did not go down at all. This, of course, should not be too surprising since the wage covariances are free and information on other components of the model should not really help in estimating them. The lesson is that the technology helps only when we have extra data with information on the parameters of interest.

On the substantive issue that motivated this work, whether ability is priced differently in the marketplace for men and women, we have been able to say very little. There are two sources of the problem: (1) Wage correlations across the siblings are very important for answering this question and we have relatively few wage pairs in these data. (2) It is difficult for us to differentiate between the sexes using test scores, since we have only one indicator of ability, IQ, and in designing that indicator attempts were made to minimize the appearance of sex differences. An interesting extension of this work might be to apply this framework to a sample with a variety of test scores,[7] such as the recent High School and Beyond surveys (NORC, 1980) although the within person correlation could be a problem when all tests are of the academic variety; that is, there may be little additional information in them.

Finally, we remind the reader again that the mean wage for the men aged 27 in this dataset is forty per cent higher than the mean wage for women of the same age and that this difference is unaccounted for by anything reported in this paper. The mean IQ and schooling level for the same men and women are equal, and our results indicate that they are getting the same returns from these factors. The cause of the discrepancy must be looked for elsewhere.

Figure 1

Expected Variances and Covariances Implied by the Model in Equations 2-4

	IQ	S	LW
Individual			
IQ	$\gamma_1^2(1+\tau)+\sigma_1^2$	$\gamma_1\gamma_2(1+\tau)$	$\gamma_1(\beta\gamma_2+\gamma_3)(1+\tau)$
S		$\gamma_2^2(1+\tau)+\eta^2+\sigma_2^2$	$\gamma_2(\beta\gamma_2+\gamma_3)(1+\tau)+\beta\eta^2+\sigma_{23}$
LW			$(\beta\gamma_2+\gamma_3)^2(1+\tau)+\beta^2(\eta^2+\sigma_2^2)+\sigma_3^2+2\beta\ \sigma_{23}$

Cross-Sibs (same sex)

IQ	γ_1^2	$\gamma_1\gamma_2$	$\gamma_1(\beta\gamma_2+\gamma_3)$
S		$\gamma_2^2+\eta^2$	$\gamma_2(\beta\gamma_2+\gamma_3)+\beta\eta^2$
LW			$(\beta\gamma_2+\gamma_3)^2+\beta^2\eta^2+\sigma_{3ab}$

Cross-Sex Cross-Sib (male down versus female across)

Q	$\gamma_{1m}\gamma_{1f}\rho_A$	$\gamma_{1m}\gamma_{2f}\rho_A$	$\gamma_{1m}(\beta_f\gamma_{2f}+\gamma_{3f})\rho_A$
	$\gamma_{1f}\gamma_{2m}\rho_A$	$\gamma_{2f}\gamma_{2m}\rho_A+\eta_f\eta_m\rho_W$	$\gamma_{2m}(\beta_f\gamma_{2f}+\gamma_{3f})\rho_A+\beta_f\eta_f\eta_m\rho_W$
W	$\gamma_{1f}(\beta_m\gamma_{2m}+\gamma_{3m})\rho_A$	$\gamma_{2f}(\beta_m\gamma_{2m}+\gamma_{3m})\rho_A+\beta_m\eta_m\eta_f\rho_W$	$(\beta_f\gamma_{2f}+\gamma_{3f})(\beta_m\gamma_{2m}+\gamma_{3m})\rho_A+\beta_f\beta_m\eta_f\eta_m\rho_W+\sigma_{3m}$

Table 1

Data Availability

	Young Men	Young Women	Brother Sample	Sister Sample	Sibs Sample
Original sample	5225	5159	1499	1464	3042
With good schooling	4901	5027	1402	1410	2906
And good IQ	3131	3149	885	874	1737
And an early wage	4291	4060	1253	1162	2498
And both wages	3110	2876	909	814	1728
And both wages and IQ	2098	2016	594	562	1134

Data Arrangement for Estimation

	Pairs			Individuals	
	Brother	Sister	Sibs	Men	Women
Complete data	164	151	204	1616	1604
Missing IQ	127	101	119	892	792
Missing wages for a male	103		59	232	
Missing wages for a female		107	87		278
Missing wages for both	38	40	48		
Residual	147	158	257	112	167
Total	579	557	774	2852	3398

Note: Cell counts are the number of sibling pairs, or number of individuals in the case of the last two columns. Individuals occur only once, but families occasionally occur more than once (one percent in sibling samples, three percent in total sample). The slight discrepancies in observations counts between the top and bottom panels are due to the fact that the bottom panel observations were also required to have good data on the KWW test score.

Table 2

Summary Statistics

Variable	Young Men			Young Women		
	Number	Mean	Standard Deviation	Number	Mean	Standard Deviation
LW2	3110	6.18	0.49	2876	5.78	0.44
LW1	4291	5.70	0.54	4059	5.46	0.44
SC	4783	12.8	2.75	4728	12.6	2.41
IQ	3131	101.4	15.9	3149	102.3	15.2
WHITE	4783	0.72	0.45	4729	0.71	0.45
AGE68	4783	18.2	3.2	4729	18.8	3.1
REG68	4783	0.41	0.49	4729	0.32	0.47
AGE1	4291	22.6	2.9	4060	23.0	2.8
SMSA1	4291	0.71	0.45	4060	0.78	0.42
REG1	4291	0.39	0.49	4060	0.32	0.47
YEAR1	4291	70.6	3.6	4060	72.2	3.0
MAR1	4?51	0.48	0.50	3396	0.56	0.50
AGE2	3110	27.1	1.5	2876	27.0	2.0
SMSA2	4783	0.46	0.50	4729	0.61	0.49
REG2	4783	0.26	0.44	4729	0.20	0.40
YEAR2	3110	75.1	3.2	2876	75.8	2.3
MAR2	3047	0.67	0.47	2279	0.66	0.48

Variable definitions:

LW1 – an early measure of log hourly earnings.
LW2 – a late measure of log hourly earnings.
SC – years of schooling completed.
IQ – IQ test score.
WHITE– dummy variable, 1 if respondent is white.
AGE – Age in years (at the time of early or late wage).
SMSA – dummy variable, 1 if respondent lives in SMSA.
REG – dummy variable, 1 if respondent lives in the South.
YEAR – calendar year corresponding to early or late wage.
MAR – dummy variable, 1 if respondent married, spouse present.
(This variable was not swept out in reduced form regressions).

Warning: The means for variables indexed with 1's and 2's were taken over those with early or late wages respectively. The changes in these variables should not therefore be interpreted as changes in the underlying population.

Table 3

Individual Earnings Equations

Men

	OLS		OLS with Missing Data		Instrumental Variables	
	LW_1	LW_2	LW_1	LW_2	LW_1	LW_2
SC	.007	.023	.017	.030	.054	.043
	(.004)	(.004)	(.003)	(.003)	(.011)	(.011)
IQ	.0013	.0026	.0006	.0019	-.0066	-.0013
	(.0007)	(.0007)	(.0006)	(.0006)	(.0016)	(.0016)
σ^2	.143	.151	.149	.153	.157	.154
Number	2148	2148	4784	4784	4784	4784

Women

	OLS		OLS with Missing Data		Instrumental Variables	
	LW_1	LW_2	LW_1	LW_2	LW_1	LW_2
SC	.050	.050	.052	.051	.091	.073
	(.004)	(.005)	(.003)	(.003)	(.013)	(.014)
IQ	.0027	.0043	.0024	.0042	.0021	.0050
	(.0006)	(.0007)	(.0006)	(.0006)	(.0017)	(.0019)
σ^2	.116	.134	.116	.128	.119	.132
Number	2110	2110	5286	5286	5286	5286

Note: All equations were estimated on the residuals from equations which included age, urban and southern residence, race, and year dummies (in the case of wages). The number of observations shown is the total number used for estimation in that column.

Table 4

Correlation Matrices of Residuals from Reduced

Form Equations: Pairwise Available Data

Individual

	Young Men (N = 2098-4783)				Young Women (N = 2016-4729)			
IQ	14.233				13.607			
SC	.491	2.452			.452	2.258		
LW1	.059	.112	.383		.241	.367	.365	
LW2	.144	.208	.430	.399	.258	.341	.528	.382

	Brothers (N = 279-611)				Sisters (N = 259-581)			
IQ	.440				.514			
SC	.330	.479			.359	.446		
LW1	.008	.075	.164		.221	.187	.204	
LW2	.099	.117	.109	.112	.198	.223	.171	.336

Brother-Sister

(N = 213-527)

		Brothers			
Sisters	IQ	.480	.338	-.006	.137
	SC	.346	.441	.035	.135
	LW1	.163	.189	.051	.123
	LW2	.104	.108	.022	.074

Note: All variables are residuals from regressions reported in Appendix 3 which sweep our exogenous variables, such as race and age. Numbers on diagonals in uppermost panels are standard deviations. Correlations are computed over all available pairs, or individuals.

Table 5

Joint Maximum Likelihood Estimates of the Full Model*

Dep. Var.	Young Men SC	Young Men A	Young Men W	Young Men σ^2	Young Women SC	Young Women A	Young Women W	Young Women σ^2
IQ		9.86		66.0		9.72		70.7
		(0.42)		(9.3)		(0.39)		(7.9)
SC		1.28	1.16	2.29		1.22	0.94	2.38
		(0.08)	(0.09)	(0.18)		(0.07)	(0.09)	(0.15)
LW1	.040	−.016		0.15	.076	.013		0.11
	(.019)	(.027)		(.004)	(.023)	(.030)		(.003)
LW2	.027	.034		0.15	.050	.060		0.13
	(.019)	(.027)		(.004)	(.024)	(.032)		(.004)

Young Men: τ = .48 (.10)

Young Women: τ = .27(.08)

Estimated Covariances

	Men Individual SC	Men Individual LW1	Men Individual LW2	Men Sibling LW1	Men Sibling LW2	Women Individual SC	Women Individual LW1	Women Individual LW2	Women Sibling LW1	Women Sibling LW2
SC	--									
LW1	−.11	--		.023		−.11	--		.014	
	(.07)			(.008)		(.07)			(.007)	
LW2	−.023	.059	--	.018	.016	−.043	.052	--	.012	.025
	(.07)	(.003)		(.007)	(.009)	(.076)	(.004)		(.006)	(.008)

Estimated Covariances Across Sexes

		Female LW1	Female LW2
	LW1	.011	.003
		(.008)	(.008)
Male	LW2	.005	.001
		(.008)	(.008)

Log Likelihood = −22,129.2

*These estimates are based on the constraint that $\rho_a = \rho_w = 1.0$; the χ^2 (2) for equality of the male and female factors = 0.8.

Notes

1. An earlier version of this paper was presented at the Conference on the Economics of the Family, University of Pennsylvania, April 12-13, 1984. We are grateful to the conference participants for comments and to Mark Watson for helpful discussions. We are also indebted to NSF Grant SOC78-04279 for financial support, to Ted Shi for research assistance, and to Sumanth Addanki and Clint Cummins for assistance with the computation.

2. If we were to allow u_1 and u_3 to be freely correlated, the model would be exactly identified. The restriction that $Eu_{1A}u_{3A} = Eu_{1B}u_{3B} = 0$, i.e. that this covariance is fully captured by the variance component g, is in the spirit of IQ being an error ridden measurement of ability, but is not essential for identification.

3. These "IQ test scores" are in fact from a variety of intelligence tests collected by the high schools and rescaled to standard IQ units by the NLS.

4. Less than one percent of families occur twice among the sib data and less than three percent of unmatched individuals actually have a sib in the sib data.

5. Necessary conditions are given in MaCurdy 1981. Basically the first and second partials of the model must be uniformly continuous and possess finite first and second moments.

6. Data which is not missing randomly may also change the variances and distributions of the observed data. This can be accomodated in estimation by (1) allowing the estimated variances of the unobservables to vary with the samples and (2) by computing robust standard errors for the model. Neither of these have been done in the current version of the paper but we plan to do so in the future.

7. The standard attack, in this context, on the missing data problem, would be to compute a correlation matrix based on pairwise complete data and then base estimates on that. As long as data is missing at random this method should be consistent, but it suffers from two drawbacks. The standard errors computed ignoring the differential data availability will be nonsense. Furthermore the maximum-likelihood technique shades naturally into estimation that at least partially models the sample generating mechanism and is thus robust to a certain amount of non-randomness.

8. A later version of this paper will contain estimates of the standard errors which are robust to nonnormality of the data.

9. The National Longitudinal Surveys of Young Men and Women also contain scores on a "Knowledge of the World of Work" test which we originally planned to use in this study. Unfortunately, the tests themselves were not the same across the two sexes so that they could not be used as an indicator variable which would provide additional identifying power. We therefore decided not to use these scores in the final version of the model.

References

Allison, P.D., "Maximum Likelihood Estimation in Linear Models When Data Are Missing." Revised version of a paper presented at the 1981 Annual Meeting of the American Sociological Association.

Becker,G.S., "The Allocation of Effort, Specific Human Capital, and Sexual Differences in Earnings and the Allocation of Time," unpublished (1983).

Behrman, J.R., Z. Hrubec, P. Taubman and T.J. Wales, **Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment, and Schooling** (Amsterdam: North Holland, 1980).

Center for Human Resource Research, **The National Longitudinal Survey Handbook** (Columbus, Ohio: Ohio State University, 1979).

Chamberlain, G., "Omitted Variable Bias in Panel Data: Estimating the Returns to Schooling," in, **The Econometrics of Panel Data**, (Paris: Annales de l'INSEE, No. 30/31, pp. 49-82, 1978).

________ and Z. Griliches, "Unobservables with a Variance-Components Structure: Ability, Schooling and the Economic Success of Brothers," **International Economic Review** 16(2) (1975).

________, "More on Brothers," in, P. Taubman, ed., **Kinometrics: Determinants of Socioeconomic Success within and between Families** (New York: North Holland, 1977).

Griliches, Z., "Estimating the Returns to Schooling: Some Econometric Problems," **Econometrica** 45 (1977), 1-22.

________, "Sibling Models and Data in Economics: Beginnings of a Survey," **Journal of Political Economy** 87(5), Part 2 (1979), S37-S64.

________, "Data Problems in Econometrics," in, M. Intriligator and Z. Griliches, eds., **Handbook of Econometrics**, Volume III (Amsterdam: North Holland, forthcoming 1985).

________, B.H. Hall and J.A. Hausman, "Missing Data and Self-Selection in Large Panels," **Annales de L'INSEE** No. 30-31 (1978).

Hall, B.H., **MOMENTS, the Moment Matrix Processor User Manuel**, (Stanford, California: 1979).

Hauser, R.M. and R.N. Daymont, "Schooling, Ability and Earnings: Cross-Sectional Findings 8 to 14 Years After High School Graduation," Soc. Educ. 50 (1977), 182-206.

Jencks, C., **Who Gets Ahead?** (New York: Basic Books, 1979).

Joreskog, K.G. and D. Sorboun, **LISREL: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood** (Chicago: National Resources, 1981).

Little, R.J.A., "Models for Non-Response in Sample Surveys," **Journal of the American Statistical Association**, 77(378) (1982), 237-250.

Mincer, J., **Schooling, Experience, and Earnings** (New York: Columbia University Press for NBER, 1974).

________ and J. Polachek, "Family Investments in Human Capital: Earnings of Women," **Journal of Political Economy** 82(2), Part II (1974), S76-S108.

NORC, **High School and Beyond: A National Longitudinal Survey for the 1980's**, (Chicago: NORC, 1980).

Rubin, D.B., "Inference and Missing Data," **Biometrika** 63(3) (1976), 581-592.

Sandell, S.H. and D. Shapiro, "The Theory of Human Capital and the Earnings of Women: A Reexamination of the Evidence," **Journal of Human Resources** 8 (1978), 103-117.

Scarr, S. and G. McAvay, "Predicting the Occupational Status of Young Adults: A Longitudinal Study of Brothers and Sisters in Adoptive and Biologically Related Families," Department of Psychology, Yale University (unpublished, 1982).

Shackett, J.R., **Experience and Earnings of Young Women**, Ph.D. thesis, Harvard University (unpublished, 1981).

Smith, J.P., **Female Labor Supply: Theory and Estimation** (Princeton: Princeton University Press, 1980).

Appendix C

In this appendix we describe in somewhat more detail the method we use for pooling our estimates across covariance matrices which may be missing one or more rows due to missing data. Although developed independently using a different computer program, our approach is the same as that of Allison (1981). His paper describes a method for using LISRELV to obtain maximum likelihood estimates of linear models when data are missing randomly, conditional on the observed data. We apply the same technique using the program MOMENTS (B. H. Hall 1979); for the models we consider, this turns out to be somewhat faster than using LISREL, although we have used that program to check some of our results.

Assume that we have T_1 observations on a vector of normally distributed random variables y_1 and T_2 observations on a vector y_2. y_1 and y_2 are jointly independent across observations. In our case y_1 and y_2 may include some of the same variables: for example, y_1 may consist of I, S, LW1, and LW2 for each of the two siblings, while y_2 may consist of the same variables, except for LW2 for one of the siblings. That is, we have T_1 observations with complete data, and T_2 observations where one of the late wages is missing. Our model specifies that $y_1 \sim N(0,\Sigma_1(\theta))$ and $y_2 \sim N(0,\Sigma_2(\theta))$ where Σ_1 is an eight by eight covariance matrix and Σ_2 is a seven by seven submatrix of Σ_1. The parameter vector θ is in common across the matrices, although some elements of it may not appear in the second matrix.

The basic likelihood function for data generated by such a process was given in the body of the paper:

$$\ln L(\Sigma | S_j, \theta) = k_i - (T_i/2)\,(\ln|\Sigma_j(\theta)| + tr\Sigma_j(\theta)^{-1}S_j) \text{ for } j=1,2$$

Under the assumption that the data are missing at random, conditional on the X's and the unobserved factors, we can add the log likelihoods for the two datasets:

$$\ln L(\Sigma_1, \Sigma_2 | S_1, S_2, \theta) \sim -\sum_{j=1}^{2} \frac{T_j}{2} \left\{ \ln|S_j(\theta)| + tr\, \Sigma_j(\theta)^{-1} S_j \right\}$$

This likelihood is additively separable in the data, although each term may depend on all or a subset of the parameter vector θ. Hence, to *maximize* the combined likelihood with respect to θ, we need only solve the sum of the first order conditions for the individual pieces:

$$\frac{\partial \ln L}{\partial \theta_i} = \sum_{j=1}^{N} \left\{ -\frac{T_j}{2} tr\left[\left(I - S_j \Sigma_j(\theta)^{-1}\right) \frac{\partial \Sigma_j}{\partial \theta_i} \Sigma_j(\theta)^{-1} \right] \right\} = 0 \;\; \forall\, i, i = 1, \cdots, k$$

Note that we have now generalized this equation from two such matrices to N. Each matrix S_j may be of different order; the only requirement they must satisfy is the requirement of independence across samples conditional on the exogenous variables. The appropriate weighting in the combined likelihood is supplied by T_j, assuming that the underlying process is i.i.d.

To solve the first order conditions and maximize the likelihood function, we use the method of scoring. Under the multivariate normal assumption, the information matrix is given by

$$E\,\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_\ell} = \sum_{j=1}^{N}\left\{-\frac{T_j}{2}\,tr\left[\frac{\partial\Sigma_j(\theta)}{\partial\theta_\ell}\,\Sigma_j(\theta)^{-1}\,\frac{\partial\Sigma_j(\theta)}{\partial\theta_i}\,\Sigma_j(\theta)^{-1}\right]\right\}$$

This is obviously easy to compute once we are computing the gradient given above. The method of scoring is an iterative quasi-Newton method where the change in the parameter vector θ^i at each iteration i is given by

$$\Delta\theta^i = \left[E\,\frac{\partial^2 \ln L}{\partial\theta\,\partial\theta'}\bigg|_{\theta^{i-1}}\right]^{-1}\frac{\partial \ln L}{\partial\theta}\bigg|_{\theta^{i-1}}$$

At convergence, the asymptotic variance-covariance matrix of θ is the inverse of the information matrix.

The models we use for $\Sigma(\theta)$ can all be written as

$$A(\theta)\,y = B(\theta)\,\varepsilon$$

where y is an n by 1 vector of observable variables and ε is and m by 1 vector of unobservables. The variance of ε depends on some of the parameters and is given by $E\varepsilon\varepsilon' = \Omega(\theta)$. ε consists of the components of the two factors A and W as well as the equation-specific disturbances u. If the matrix of structural coefficients $A(\theta)$ is invertible, we can derive an expression for the covariance of the observable variables:

$$Eyy' = \Sigma(\theta) = A(\theta)^{-1}B(\theta)\Omega(\theta)B(\theta)'A(\theta)^{-1}$$

This is the equation we use for estimation. Since the elements of A, B, and Ω are generally single elements of the parameter vector θ, the derivatives of $\Sigma(\theta)$ with respect to the parameters are easy to compute.

Details on the implementation of this estimation scheme for structural models with unobservables are given in Chapter 4 of the MOMENTS manual. To do estimation with more than one matrix, we take advantage of the multiple matrix feature of MOMENTS, which is described in Chapter 1 of the manual. When loading more than one moment matrix, some of which may be missing data, we need to fill in the rows and columns corresponding to the missing data with zeroes. For example, if row 4 contains the second wage LW2 in the first matrix, row 4 will also contain the second wage in matrix 2, 3, etc., even if the second wage is missing for one of these matrices. In this case, you simply fill in the row and column with zeroes. When a row is missing in the data matrix, the corresponding row and column of $\Sigma_j(\theta)$ should also be set to zero, so that the program will compute the correct likelihood for the subset of observed data. MOMENTS has been modified so that zero rows and columns in Σ_j and S_j cause the following to happen during computation: 1) the determinant of the largest submatrix of Σ_j which is of full rank is computed when some rows are zero, and 2) the inverse of the largest submatrix of Σ_j which is of full rank is computed and the rows and columns of the inverse corresponding to the zeroes on the diagonal of the original matrix are set to zero. When the product of Σ_j^{-1} and S_j is taken this will obviously lead to zeroes on the diagonals corresponding to the missing data and therefore they will not contribute to the trace. Inspection of the formula for the log likelihood should convince you that this technique is equivalent to computing the likelihood for a data matrix which contains only the subset of S_j which is observed.

The advantage of setting up the problem so that all the matrices are of the same size comes in programming the model and its derivatives in the MODEL (or MDLSTR) subroutine - the computations will be the same no matter which data matrix is being estimated and the only special programming involved is the bookkeeping which keeps track of which row or rows are missing.

In the MOMENTS program, the maximum likelihood procedure MAXLIK will loop over as many covariance matrices as are specified by the MATRIX command or supplied on the MAXLIK command itself. The MODEL subroutine is called to evaluate the likelihood function and its derivatives for each of the matrices. A special version of MODEL contains the structural model described here and it calls MDLSTR to compute the A, B, and Ω matrices as a function of the parameter vector θ for each data matrix IMAT in turn. This allows any of the elements of A, B, or Ω to be different across matrices, or, if you wish, they can be the same. To zero the rows of any matrix Σ, it is sufficient to zero the corresponding rows of A and B.

Appendix D

Robust Standard Errors for Quasi-Maximum Likelihood Estimates of Covariance Parameters

The derivation and equations in this section are drawn from MaCurdy (1981), "Asymptotic Properties of Quasi-Maximum Likelihood Estimators and Test Statistics." The purpose of the present appendix is merely to derive from the general formulas the exact equations we use to estimate standard errors for our model.

A key result in MaCurdy's paper is that the limited information estimator of a set of variance parameters ω which is obtained conditional on a set of regression parameter estimates $\hat{\gamma}$ is consistent for the true parameter values ω_o and is asymptotically normally distributed with a covariance matrix which does not depend on the fact that $\hat{\gamma}$ are the estimated rather than the true values of γ_o. This result holds for the general nonlinear multivariate regression model with independently, identically, but not necessarily normally distributed disturbances. Our model is a linear structural model in which the structural parameters appear explicitly in the covariance matrix, so that we can think of it as a linear multivariate regression model. The estimated parameter vector $\hat{\gamma}$ corresponds to our set of reduced form coefficients which are swept out of all the variables before forming the moment matrices of residuals. The parameters ω for which we want standard errors are all of the parameters in the structural model, both the slopes and the covariance parameters.

If all the disturbances in our model were normally distributed, the estimator we are using would be a conventional maximum likelihood estimator and the standard errors would be computed as described in Appendix C, as the inverse of the expectation of the information matrix evaluated at the maximum of the likelihood. However, when the disturbances are nonnormal, the estimator we are using can be interpreted as a quasi-maximum likelihood estimator; the estimates of ω thus obtained have the same properties as ML estimates under normality but with a covariance matrix given by

$$V(\hat{\omega}) = \frac{1}{N} H^{-1}_{\omega\omega}(\hat{\gamma},\hat{\omega})\, G_{\omega\omega}(\hat{\gamma},\hat{\omega})\, H^{-1}_{\omega\omega}(\hat{\gamma},\hat{\omega})$$

When the disturbances are normal, this reduces to $V(\omega)=H^{-1}=G^{-1}$. The matrices G and H are the outer product of the gradients of the individual quasi-likelihood functions with respect to the parameter vector and the matrix of second partials of the sum of the individual functions with respect to the parameters. Both of these are to be evaluated at the estimated parameters $\hat{\gamma}$ and $\hat{\omega}$.

In the notation of Appendix C, the contribution of the ith individual to the quasi-likelihood function Q_N is

$$q_i = -\frac{1}{2}\left\{ \ln|\Sigma(\theta)| + \operatorname{tr}\Sigma(\theta)^{-1} S_i \right\}$$

where we have suppressed the j subscript which specified which moment matrix was under consideration. $\Sigma(\theta)$ depends only on the parameters of the covariance matrix, w, and S_i is the observed cross product matrix of the estimated residuals for the ith individual. To estimate the standard errors of our estimated $\hat{\theta}$, we need expressions for $G(\hat{\theta})$ and $H(\hat{\theta})$:

$$G(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\partial q_i}{\partial \theta} \frac{\partial q_i}{\partial \theta'} \right]_{\hat{\theta}}$$

$$H(\theta) = - \left. \frac{\partial^2 Q_N}{\partial \theta \, \partial \theta'} \right|_{\hat{\theta}}$$

To derive these for our model, first rewrite q_i as

$$q_i = -\frac{1}{2} \left\{ \ln|\Sigma(\theta)| + \hat{s}_i{}' \, \Sigma(\theta)^{-1} \, \hat{s}_i \right\}$$

where $\hat{s}_i$ are the estimated residuals. Then

$$\frac{\partial q_i}{\partial \theta_k} = -\frac{1}{2} \operatorname{tr} \left\{ \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} \left[I - S_i \Sigma^{-1} \right] \right\}$$

From this we can compute a typical element of plim $G(\theta)$ as

$$G_{kl} = \frac{1}{4} \left[\operatorname{tr} \left\{ \frac{\partial \Sigma}{\partial \theta_k} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-4} \left[\frac{1}{N} \sum_{i=1}^{N} S_i S_i \right] \right\} - \left[\operatorname{tr} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_k} \right] \left[\operatorname{tr} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_l} \right] \right]$$

and a typical element of plim $H(\theta)$ as

$$H_{kl} = -\frac{1}{2} \operatorname{tr} \left\{ \frac{\partial \Sigma}{\partial \theta_k} \frac{\partial \Sigma}{\partial \theta_l} \Sigma^{-2} \right\}$$

Here we have used the fact that $E\, S_i = \Sigma$. Note that if

$$E \frac{1}{N} \sum_{i=1}^{N} S_i^2 = 3\Sigma^2,$$

the normal case, then $G(\theta) = -H(\theta)$ and we are back to the maximum likelihood case described in Appendix C.