Marital Status and Earnings of Young Men: A Model with Endogenous Selection

Author(s): Robert A. Nakosteen and Michael A. Zimmer

Source: *The Journal of Human Resources*, Spring, 1987, Vol. 22, No. 2 (Spring, 1987), pp. 248-268

Published by: University of Wisconsin Press

Stable URL: https://www.jstor.org/stable/145904

Marital Status and Earnings of Young Men
A Model with Endogenous Selection

Robert A. Nakosteen and Michael A. Zimmer

ABSTRACT

It is a matter of course in applied labor economics to presume that marital status is a predetermined contributor to earnings among young males. We find no evidence to support that proposition. We estimate a model of earnings determination that permits endogenous selection of marital status. Our estimates are based on a sample of young employed males from the Michigan Panel Survey of Income Dynamics. They indicate that marital status, viewed in this manner, does not significantly shift the mean earnings profile. Our sensitivity tests indicate that this conclusion is robust with respect to a variety of alternative specifications.

I. Introduction

In recent years labor economists have maintained vigorous interest in the determinants of individual wages and earnings. Areas of interest, and typical examples, include the role of schooling and experience (Mincer 1974); life-cycle earnings profiles (Rosen and Taubman 1982); earnings differences by race and gender (Borjas 1983); union wage premiums (Lee 1978; Duncan and Leigh 1985); earnings effects of language characteristics (Grenier 1984); and religious denomination (Tomes 1984); and the earnings impact of labor mobility between regions and industries (Nakosteen and Zimmer 1982). These studies proceed from a common

Nakosteen is an associate professor of general business and finance at the University of Massachusetts, Amherst. Zimmer is an associate professor of economics at the University of Evansville, Indiana. *The authors wish to acknowledge benefits from productive insights of anonymous referees.*
[Submitted July 1985; accepted October 1986]

framework in which the natural logarithm of wages or earnings is regressed on a set of explanatory variables, the latter consisting in part of a core set of variables consistent with human capital theory. Many studies augment the set with additional control measures and other variables to serve the purposes of the research at hand. The controls often include conventional dummy variables for gender, race, and marital status. Implicit in this approach is the assumption that each control variable is predetermined, or at least generated independently of the error term. While this assumption does not arouse serious opposition in the case of race or gender, the issue of marital status warrants additional attention. It is worthwhile to question the appropriateness of treating marital status as though it is determined exogenously in a manner analogous to race and gender.[1]

A possibility exists that marital status is determined stochastically by a process whose random unobservable component is correlated with unobservables in the wage/earnings function. In such a case conventional least squares estimates of the wage function, in particular the marital status coefficient and its standard error, are biased and inconsistent. Since most studies have reported positive earnings premiums associated with married status, this is an empirical issue of some significance.[2]

The purpose of this study is to reconsider the role of marital status in a conventional log-earnings model. Based on a sample of young males extracted from the 1977 wave of the Michigan Panel Survey of Income Dynamics, the model consists of equations describing marital status and earnings. The first equation represents observed marital status as a function of explanatory variables along with a random disburbance term representing unobservable factors in the determination of marital status. The second equation specifies the logarithm of annual earnings as a linear combination of human capital and control variables and a random disturbance term. Included among the controls is an indicator of marital status. Previous studies of earnings have implicitly supposed that the random error terms in

1. Examples of such studies are numerous. See, for example, Grenier (1984), Tomes (1984), Rosen and Taubman (1982), and Duncan and Leigh (1985).

2. The studies cited in Footnote 1 are representative. Tomes (1984), using the 1973–80 N.O.R.C. General Social Surveys and log-earnings as the dependent variable, reports a marriage coefficient of −.2068 for single U.S. Protestant males. Grenier (1984), using the 1976 Survey of Income and Education and log-wage as the dependent variable, reports a coefficient of .2798 for married U.S. non-Hispanic white males. Rosen and Taubman (1982), using the 1973 CPS-SSA exact match sample and a Tobit regression of log-earnings, report a coefficient of .2309 for U.S. white males. Duncan and Leigh (1985), based on a sample of mature males from the 1971 National Longitudinal Surveys and using an instrumental variable log-wage regression, obtain an estimate of −.259 for unmarried nonunion workers. All of these estimates differ significantly from zero at the 1 percent level or better, although Tomes fails to obtain significant estimates for other religious denominations and Duncan and Leigh find an insignificant coefficient for union members.

the equations are uncorrelated, thus facilitating conventional least squares estimation of the earnings equation. This study examines the consequences of estimating the equations without the zero-correlation restriction. Of principal interest is the resulting change in the magnitude and statistical significance of the marital status coefficient, and changes in estimates of the remaining parameters.

Elementary principles of human capital suggest at least two reasons that married males might be observed to possess higher earnings than their unmarried counterparts. First, since the marriage arrangement facilitates specialization in labor, and traditionally has done so in a manner that assigns the male's labor to be relatively market intensive, a marriage incentive exists for males possessing above-average market productivity. To the extent that this increased specialization enhances productivity that is then translated into higher wages, positive selection of high-wage males into marriage will tend to create the appearance in observed earnings of an advantage to marriage per se. Second, the marital arrangement itself may create conditions under which either party may accumulate human capital more efficiently than he/she could attain as a single worker. This would be true in cases where either partner may finance training by borrowing from his/her spouse's past or current earnings at rates more favorable than those available in "outside" capital markets. To the extent that the marriage-induced marginal accumulation of human capital translates into higher wages, married males will tend to exhibit premiums in observed earnings relative to unmarried males.

In either case there is clear potential for correlation between factors that influence both marital status and earnings but that remain unobserved by the analyst, and hence become impounded into the error term of the earnings equation. Kenny (1983) provides evidence in support of the second hypothesis, suggesting that after ten years of marriage males enjoy a wage premium of 17 to 20 percent over comparable single males attributable to their more rapid accumulation of human capital during marriage. Thus there is reason to suspect correlation among unobservables, giving rise to the need for a framework in which the earnings equation may be estimated in a manner that permits improved inferences about the estimated coefficient of marital status.

II. Framework

The model consists of equations in which the individual's marital status and subsequent earnings are determined in a manner that permits endogeneity among outcomes. A distinguishing feature of the model is that it includes marital status in a separate equation, the outcome of which is permitted (under the alternative hypothesis) to alter the structure of the earnings equation. In order to specify the model, let I_i^* denote an unobservable index of the propensity of individual i to be observed as married at a point in time; let Y_i denote a vector of predetermined explanatory variables, hypothesized to be correlated with marital status; let E_{1i} and E_{2i} denote measured earnings of individual i in the married and single states, respectively; and let X_i denote a vector of explanatory human capital and control variables in the earnings equation. The latent index is represented as a linear function of predetermined explanatory variables, a random error term, and log-earnings in the alternative marriage states:

$$(1) \quad I_i^* = \beta' Y_i + \delta_1 \ln E_{1i} + \delta_2 \ln E_{2i} + u_i^*.$$

In this formulation β represents an unknown vector of coefficients, while δ_1 and δ_2 denote inducements toward marriage of earnings in the two married states. The error term u_i^* is assumed to possess a normal distribution with zero mean and variance equal to σ_u^{*2}. Individual i is married if $I_i^* \geqq 0$ and is not married if $I_i^* < 0$. The latent variable cannot be observed. Instead the analyst observes

$$I_i = 1 \text{ if } I_i^* \geqq 0$$

$$I_i = 0 \text{ if } I_i^* < 0.$$

When (1) is viewed in this fashion, its parameters may be estimated by maximum likelihood probit methods.

Additional equations describe the logarithm of earnings in each of the mutually exclusive marriage states:

$$(2) \quad \ln E_{1i} = \alpha_1' X_i + \epsilon_{1i} \qquad \text{if } I_i = 1$$

$$(3) \quad \ln E_{2i} = \alpha_2' X_i + \epsilon_{2i} \qquad \text{if } I_i = 0,$$

where α_1 and α_2 denote unknown parameter vectors, and ϵ_1 and ϵ_2 denote normal random variates with zero means and variances σ_1^2 and σ_2^2, respectively.

The model is completed by acknowledging the unknown covariances between error terms in the earnings and marital status equations, denoted σ_{1u} and σ_{2u}. If these covariances are zero, then the earnings equations may be estimated by ordinary least squares. However, it is well known that the presence of nonzero covariation renders least squares estimates of the earnings equations biased and inconsistent. Least squares estimates are also biased and inconsistent in this model if $\delta_1 \neq 0$ or $\delta_2 \neq 0$, even if $\sigma_{1u} = \sigma_{2u} = 0$. This model has been discussed by Lee (1979), who considers issues of identification and alternative approaches to estimation.

The approach implicit in previous studies has been to assume $\delta_1 = \delta_2 = \sigma_{1u} = \sigma_{2u} = 0$, and $\alpha_1 = \alpha_2$ apart from a constant term. The conventional approach is to estimate the distinct constants in the earnings equation by means of an ordinary marital status dummy variable. This study offers a framework in which these restrictions may be more carefully assessed. The model described here is one of a class of models characterized by endogenous selection. Endogeneity arises in the sense that each individual self-selects into a chosen status, in accordance with some stochastic sorting mechanism. Models in this class are characterized by: (1) equation(s) describing the selection mechanism, often in reduced form, and (2) one or more structural equations describing outcomes of the sorting process. Endogenous sorting manifests itself in a manner that cannot be casually observed, i.e., covariance between unobservables in the determination of status and outcome, the consequence of which is to preclude least squares estimation of the outcome equation. Several procedures exist as alternatives for estimating the earnings equations without selection bias. The approaches employed here are discussed by Heckman (1978), Lee (1979), Maddala and Lee (1976), and Maddala (1983). The following sections address these issues in the context of models that may be viewed as special cases of the model described in this section.

III. Alternative Formulations

A preliminary item of interest in estimation of Equations (1) through (3) concerns the possible equality of the coefficient vectors α_1 and α_2. This problem may be addressed by viewing the model as a switching regression model with endogenous switching. Substituting Equations (2) and (3) into the structural probit equation (1) yields a reduced form probit:

$$(4) \quad I_i^* = \beta Y_i + (\delta_1 \alpha_1' + \delta_2 \alpha_2')X_i + (\delta_1 \epsilon_{1i} + \delta_2 \epsilon_{2i} + u_i^*)$$

Identification conditions that permit estimates of the structural parameters to be retrieved from the reduced form coefficients require the X vector to contain at least one element not found in Y. Rewriting the reduced form more compactly:

$$(4)' \quad I_i^* = \gamma^{*\prime} Z_i - \nu_i^*,$$

where ν_i^* is normally distributed with zero mean and variance σ_ν^{*2}, the switching regression model may be expressed as a two-regime model of log earnings:

$$(5) \quad \ln E_{1i} = \alpha_1' X_i + \epsilon_{1i} \qquad \text{if} \qquad \gamma' Z_i \geqq \nu_i$$

$$(6) \quad \ln E_{2i} = \alpha_2' X_i + \epsilon_{2i} \qquad \text{if} \qquad \gamma' Z_i \geqq \nu_i,$$

where $\gamma = \gamma^*/\sigma_\nu^*$ and $\nu = \nu^*/\sigma_\nu^*$.

The reduced form probit provides a descriptive measure of the probability that $I_i = 1$ or $I_i = 0$ by means of the standard normal distribution. Accordingly, the probability that $I_i = 1$ can be written:

$$(7) \quad P(I_i = 1) = \Phi(\gamma' Z_i),$$

where ϕ denotes the standard normal cumulative distribution function. Normalization of the parameters in (4)′ results in a normally distributed error term with zero mean and unit variance.

For a single individual selected at random, unconditional expected log earnings are expressed as

$$(8) \quad E(\ln E_i) = E(\ln E_i \mid I_i = 1) \cdot P(I_i = 1) + E(\ln E_i \mid I_i = 0) \cdot P(I_i = 0)$$
$$= \alpha_1' X_i[\Phi(\gamma' Z_i)] + \alpha_2' X_i[1 - \Phi(\gamma' Z_i)] + \phi(\gamma' Z_i)(\sigma_{2u} - \sigma_{1u}),$$

where Φ denotes the standard normal density and σ_{1u} and σ_{2u} denote, respectively, covariances between the error term in Equation (4) and error terms in Equations (5) and (6). Since the X vector is defined identically across regimes, Equation (8) may be simplified:

$$(9) \quad E(\ln E_i) = \alpha_2' X_i + (\alpha_1' - \alpha_2') X_i[\Phi(\gamma' Z_i)]$$
$$+ (\sigma_{2u} - \sigma_{1u})[\phi(\gamma' Z_i)].$$

This model is discussed by Maddala (1983, 223–28), who suggests a procedure for testing the equality of coefficient vectors between regimes. Equation (9) contains three sets of explanatory variables: the conventional vector of explanatory variables in the earnings equation; a vector consisting of interactions between these variables and the expected probability of married status [based on Equation (7)]; and a single term in which the standard normal density is evaluated at the argument in Equation (7). A test for equality between regimes is equivalent to a test of the null hypothesis that each of the coefficients of the interaction terms is zero. Maddala suggests that the test can be conducted by first estimating the parameters of (7) by probit maximum likelihood. The resulting estimates are used to estimate the normal distribution and density terms, $\Phi(\gamma' Z_i)$ and $\phi(\gamma' Z_i)$, for each individual in the sample. Results of these computations are inserted into (9), which is estimated by least squares. Standard tests of significance of the interaction terms then serve as evidence on the equality of earnings coefficients.

If this test provides strong evidence of structural earnings differences between regimes then the appropriate formulation of the structural model is Equations (1)–(3). On the other hand if the test supports equality among coefficients then the appropriate model calls for a single earnings regime, and the corresponding marital status equation includes the level of earnings as an endogenous variable:

$$(10) \qquad I_i^* = \beta' Y_i + \delta \ln E_i + u_i^*$$

$$(11) \quad \ln E_i = \alpha' X_i + \tau I_i + \epsilon_i.$$

The principal item of interest in this model is τ, the coefficient of marital status in the earnings equation. The following sections describe specification and estimation of the model in these alternative formulations.

The model is completed by stating the specifications in Equations (1) through (3). Variables included in the earnings equations are schooling, experience, a quadratic experience term, and dummy variables for veteran status and race. The marital status equation includes, in addition to the log earnings variables, a set of predetermined family background variables: each parent's educational attainment, number of siblings, and a dummy indicator for the presence of older siblings, along with dummy variables for religion, race, and urban upbringing.

Variables in the marital status equation attempt to capture potential influences on a young man's chosen marital status between the ages of 18 and 24. Education of parents, in addition to measuring their respective accumulations of formal training, conveys some parental values regarding acquisition of schooling which may be imparted to offspring. To the extent that offspring may postpone marriage pending attainment of schooling goals, education of parents may act as a deterrent to marriage in the 18 to 24 age group.[3] Presence of siblings may play a role in marriage decisions of young men, but its sign is difficult to specify a priori. Similarly, the dummy variables for religion, race, and city size, included as demographic controls, possess no clear economic basis upon which they may be signed a priori.

The earnings specification is consistent with a large number of empirical studies. Confinement of the sample to young men precludes significant sample variation in age. Consequently age is excluded, leaving experience and its square to capture the effects of added years of work associated with older individuals in the sample. In Section VI and the Appendix, we test the sensitivity of the model to alternative specifications by restricting various subsets of these variables from the model.

IV. Data and Specifications

Estimation is based on a sample of 576 employed males, 18 to 24 years of age, extracted from the 1977 wave of the Michigan Panel Survey of Income Dynamics. The age restriction is intended to capture a sample in

3. Inclusion of these variables is motivated in part by the results of Behrman and Wolfe (1984), who report significant impacts of parental schooling on socioeconomic outcomes (income and social status) for female offspring in a developing country.

the formative years of human capital accumulation, and presumably the age interval in which most young males contemplate a first marriage.

Variables on each male's record include age; education, number of years (ED); experience, number of years work experience since age 18 (EXP); a quadratic term for experience (EXPSQ); a dummy variable equal to zero if white, one if nonwhite (DRACE); a dummy variable equal to one if veteran, zero otherwise (DVET); education of father and mother (FED1–FED7 and MED1–MED7), each measured as a set of seven dummy variables for intervals extending from six to eight years to college plus advanced degree (the omitted category is zero to five years); number of siblings (NSIB); a dummy variable equal to one if the respondent had older siblings, zero otherwise (OSIB); a dummy variable equal to zero if religious preference was Catholic, one otherwise (DREL); a dummy variable equal to one if the respondent grew up in a large city, zero otherwise (DGREW); a dummy variable equal to one if married, zero otherwise (I); and total income from wages in 1976 (E). A summary of variable definitions and descriptive statistics is given in Table 1.

A representative member of the sample is seen to be about 22 years old, possesses the equivalent of a nearly completed high school education, and has worked more than three years since the age of 18. The sample consists of more than one-third nonwhites and 11 percent veterans. Parents' educational attainments appear to be relatively concentrated at the level of high school completion or less. The proportion of marrieds equals 66 percent. Average earnings total $7,307.28, measured as income in dollars from wages during the survey year.

V. Preliminary Estimates and Alternative Specifications

A useful basis for comparison of estimates is a conventional log earnings equation wherein marital status is represented as a simple dummy variable, similar to the manner in which it commonly appears in other studies. Estimates of the simple model are presented in the first column of Table 2. The results indicate a very rapid return on work experience among young males, although at the customary declining rate. Schooling appears to have no significant impact, reflecting the fact that educational attainment does not strongly exert its influence on earnings before the age of twenty-four. In this sample, most of the more educated are still quite young. Although for young adults the effect of education is to enhance earnings, it almost certainly reduces immediate postschool earnings due to experience foregone in the acquisition of schooling. Veteran status appears not to

Table 1
Descriptive Statistics

Variable	Description	Mean	Standard Deviation
N		576	
AGE	Years	22.188	1.55
ED	Number of years of education	11.984	1.642
EXP	Number of years work experience since age 18	3.304	1.566
DRACE	= 0 if white; = 1 if nonwhite	0.375	0.485
DVET	= 0 if nonveteran; = 1 if veteran	0.111	0.315
FED1–FED7	Father's maximum education, years		
	FED1 = 1 if 6–8, 0 otherwise	0.311	0.463
	FED2 = 1 if 9–11, 0 otherwise	0.156	0.363
	FED3 = 1 if 12, 0 otherwise	0.253	0.435
	FED4 = 1 if 12 + technical training, 0 otherwise	0.024	0.154
	FED5 = 1 if college but no degree, 0 otherwise	0.061	0.239
	FED6 = 1 if college degree, 0 otherwise	0.050	0.218
	FED7 = 1 if college + advanced degree, 0 otherwise	0.021	0.143
MED1–MED7	Mother's maximum eduction, years (Same definitions as FED1–FED7)		
	MED1	0.191	0.393
	MED2	0.199	0.400
	MED3	0.378	0.485
	MED4	0.035	0.183
	MED5	0.066	0.248
	MED6	0.039	0.179
	MED7	0.016	0.124
NSIB	Number of siblings	4.139	2.372
DSIB	= 1 if respondent had older siblings; = 0 otherwise	0.752	0.432
DGREW	= 1 if respondent grew up in large city; = 0 otherwise	0.424	0.495
DREL	= 0 if Catholic; = 1 otherwise	0.675	0.469
DMAR	= 0 if not married; = 1 if married	0.658	0.475
E	Income from wages in 1976	$7,307.08	$4,640.04

Table 2
Preliminary Log Earnings Estimates[a]

	Conventional Model		Equation (9)	
CONSTANT	7.575	(12.53)	23.224	(0.51)
ED	−0.047	(−1.06)	−0.649	(−1.01)
EXP	0.606	(3.04)	4.925	(1.92)
EXPSQ	−0.059	(−2.02)	−0.447	(−1.16)
DVET	−0.065	(−0.24)	−2.445	(−0.71)
DRACE	−0.536	(−2.95)	−0.610	(−0.19)
I	0.370	(1.99)		
$\hat{\Phi}$ ED			0.790	(0.93)
$\hat{\Phi}$ EXP			−6.245	(−1.77)
$\hat{\Phi}$ EXPSQ			0.582	(1.14)
$\hat{\Phi}$ DVET			3.293	(0.68)
$\hat{\Phi}$ DRACE			0.211	(0.05)
$\hat{\phi}$			−12.616	(−0.60)
$\hat{\Phi}$			−13.901	(−0.26)
SEE	2.033		2.124	
R square	0.058		0.061	

a. Figures in parentheses are *t*-statistics.

significantly affect earnings, and nonwhite status exerts a substantial downward shift in the earnings profile.

Of particular interest is the marital status coefficient. Note from Table 1 that the sample proportion of married males is .658. Representation of marital status by a conventional dummy is equivalent to assuming that a simple exogenous mechanism (say, a hypothetical loaded coin) randomly sorts some population members into marriage. The coefficient estimate in Table 2, .370 and significant, implies that these individuals are fortunate from the standpoint of earnings, since the coefficient translates, ceteris paribus, into a 45 percent premium in earnings. Although this estimate is somewhat larger than those obtained in other studies, it does not render them any more palatable, for these results all contain an unsettling policy implication: a young single black, who evidently may expect to suffer a large earnings disadvantage with respect to a comparable white, may offset a substantial portion of the deficit by entering into marriage. Moreover the model implies that this remedy pertains to any randomly selected black male, and presumably to any of his potential marital arrangements. Clearly there exists a need for an alternative framework that is free of such interpretations.

Estimation of Equation (9), as a means of testing for the presence of two distinct earnings regimes, proceeds in two steps. Step one entails probit estimation of the marital status reduced form, Equation (7). From these estimates we construct the normal distribution and density terms, $\Phi(\hat{\gamma}Z_i)$ and $\phi(\hat{\gamma}Z_i)$, hereafter $\hat{\Phi}_i$ and $\hat{\phi}_i$. Step two requires insertion of these estimates for their unknown counterparts in Equation (9), which is then estimated by least squares.

Marital status probit estimates are presented in the first column of Table 3. Equation (7), as a reduced form, does not possess structural interpretation. Rather, the estimates must be interpreted as capturing both direct effects on marital status and indirect effects through earnings. Since the partial derivative of the probit dependent variable with respect to any independent variable is identical in sign with the corresponding coefficient, positive (negative) estimates are indicative of inducements (deterrents) to marriage. Examination of the coefficients and associated t-statistics reveals several estimates that attain significance at the .05 level or better. Of the immediate family variables, number of siblings and parents' education emerge as significant deterrents to marriage. The sibling coefficient may reflect the presence of close companionship in larger families, perhaps attenuating one motive, the need for same, that marriage would otherwise provide. Each of the sets of parents' education dummy variables contains one negative significant coefficient. These represent the categories "college plus advanced degree" for fathers and "college but no degree" for mothers. The deterrent effect of parents' schooling may represent values conveyed to offspring that motivate them to similar educational attainment and perhaps postponement of marriage until these goals are realized. The importance of this variable has appeared in other studies. Behrman and Wolfe (1984, 301), in a study of young females in a developing country, describe parental schooling as one of the few familial background variables that consistently exerts significant impact on adult socioeconomic outcomes of offspring.

Dummy variables for the nonwhite racial category and urban upbringing exert strong negative impacts on the propensity to marry. Of particular interest is the apparent higher propensity of whites in this age interval to marry, since there exists an apparent earnings premium, widely documented in other studies, of whites over nonwhites. Although it is tempting to infer that the observed earnings advantage of marriage may reflect the preponderance of whites among married males, this conclusion is qualified somewhat by the reduced form nature of these estimates.

The dummy variable indicating veteran status is a significant deterrent to marriage. Variables that fail to attain significance are education, experience, and dummy variables indicating non-Catholic religious affiliation and

Table 3
Estimates of the Treatment Effects Model[a]

	Probit: Equation (7)		Log Earnings: Equation (11)	
CONSTANT	0.679	(4.17)	7.611	(9.93)
NSIB	−0.020	(−2.13)		
OSIB	0.062	(1.33)		
DREL	0.056	(1.33)		
DRACE	−0.140	(−2.81)	−0.458	(−2.13)
DGREW	−0.091	(−2.20)		
FED1	−0.025	(−0.38)		
FED2	0.069	(0.87)		
FED3	0.020	(0.27)		
FED4	−0.015	(−0.11)		
FED5	−0.015	(−0.14)		
FED6	−0.014	(−1.22)		
FED7	−0.375	(−2.25)		
MED1	−0.082	(−1.00)		
MED2	−0.171	(−0.20)		
MED3	−0.139	(−1.67)		
MED4	−0.167	(−1.29)		
MED5	−0.224	(−2.07)		
MED6	−0.181	(−1.27)		
MED7	−0.095	(−0.50)		
ED	−0.001	(−0.10)	−0.055	(−1.18)
EXP	0.063	(1.42)	0.550	(2.57)
EXPSQ	−0.001	(−0.07)	−0.049	(−1.63)
DVET	−0.137	(−2.25)	−0.012	(−0.04)
I			0.410	(0.59)
h[b]			−0.190	(−0.04)
Chi square	86.47	(23 df)		
SEE			2.103	
R square			0.048	

a. Figures in parentheses are t-statistics.
b. Computed from Equation (19).

presence of older siblings. The chi-square statistic easily rejects the null hypothesis that the vector of coefficients is jointly zero.

Second stage estimates of Equation (9) are presented in the second column of Table 2. As noted in Section III, the earnings-marital status model lends itself to competing formulations, depending upon whether earnings switch between two regimes [Equations (1)–(3)] or conform to a single regime [Equations (10)–(11)]. A suitable test is provided by the estimated coefficients of interaction terms in Equation (9).[4]

Examination of the results reveals little support for the two-regime model. None of the essential interaction terms attains significance at the .05 level or better, nor do the other estimates in the model appear to be particularly stable when compared to their conventional counterparts in the first column. In addition, inclusion of the interaction terms appears to have diminished the significance of the experience terms and the race dummy variable. A null hypothesis that jointly restricts coefficients of the final six terms in Equation (9) to zero can be tested by estimating the equation with and without those terms. A standard F-statistic, based on the sums of squared residuals from the two regressions, serves as the test statistic. For Equation (9) the F-value was 1.05, which is considerably smaller than the appropriate critical value corresponding to a 10 percent level of significance. It appears reasonable, therefore, to specify a single earnings regime in which marital status is portrayed as an endogenously determined shifter of the earnings profile, as in Equations (10) and (11). In this context the reduced form probit, Equation (4)′, is understood to result from substitution of Equation (11) into Equation (10).

The model described by Equations (4)′ and (11) is described by Barnow, Cain, and Goldberger (1978), who name it the treatment effects model and discuss methods of estimation. The model is also discussed by Maddala and Lee (1976, 527–29), who point out that identification of the earnings equation parameters, in particular the coefficient of marital status, requires the marital status equation to contain at least one exogenous variable that is not included in the earnings equation. As noted above, the model includes a set of family background variables that appear only in the marital status equation.

For the treatment effects model Barnow, Cain, and Goldberger and Heckman (1978, 938) suggest a two-stage estimation procedure. Stage one

4. An earnings model such as Equation (9), which includes interactions between the probability of married status and the other determinants of earnings, implies the presence of observed population heterogeneity in gains from marriage. Björklund and Moffitt (1985) discuss a more general model which permits observed and unobserved heterogeneity in both gains and costs associated with marriage. Björklund and Moffitt use the model to present evidence of heterogeneous self-selection among participants in a government manpower training program in Sweden.

consists of probit estimation of the marital status reduced form. The resulting estimates, denoted $\hat{\gamma}$, are used with the arguments in Equation (4)′ to obtain fitted values of the standard normal density and distribution functions for each observation:

$$(18)\quad \hat{\phi}_i = \phi(\hat{\gamma}Z_i),$$
$$\hat{\Phi}_i = \Phi(\hat{\gamma}Z_i).$$

Next define r_i as the ratio

$$r_i = \hat{\phi}_i / \hat{\Phi}_i,$$

and construct the marital status selectivity variable

$$(19)\quad h_i = I_i r_i - (1 - I_i) r_i [\hat{\Phi}_i / (1 - \hat{\Phi}_i)].$$

Stage two entails least squares estimation of the log earnings equation after appending h_i as an additional explanatory variable. The two-step procedure produces a consistent estimate of the treatment effect which is purged of the selection bias.[5]

As a reduced form, Equation (4)′ cannot be interpreted as a model of the decision to marry. However, as Barnow, Cain, and Goldberger (1978, 19) point out, it is not necessary for the treatment status equation to have a causal interpretation; rather it is interpreted as descriptive of the joint distribution between itself and the earnings equation.

VI. Estimates of the Single Regime Model

Table 3 presents nonlinear least squares estimates of the single regime treatment effects model, Equations (4)′ and (11). The first column presents estimates of the marital status reduced form, which were discussed in Section V, where they provided input in the test for distinct earnings regimes. In the single regime model the probit estimates are used to construct the selectivity term, in accordance with Equation (19). The second column presents selectivity-corrected estimates of the earnings coefficients.

Results for the schooling and experience terms and the race dummy are qualitatively similar to their simple least squares counterparts in Table 2, although the quadratic term on experience is now significant only at the .10 level of significance. The veteran status dummy decreases in magnitude and remains insignificant.

The principal difference lies in the marital status coefficient which in-

5. Although the coefficient estimates are consistent, least squares estimates of their standard errors are biased. Estimates in this paper were obtained by means of a routine in the LIMDEP econometric package which produces corrected standard errors. The authors wish to thank William Greene for helpful comments.

creases in magnitude but is no longer significant. The significant effect of married status that was evident in Table 2 disappears when the model is estimated free of selectivity bias, while the remaining estimates continue to be largely unaffected. Thus, despite the apparent lack of significance of the selectivity term itself, these estimates indicate that marital status per se, when estimated consistently as an endogenous treatment effect, fails to significantly shift the earnings profile.

The diminished role of marital status in the treatment effects model invites additional scrutiny regarding specification of the structural model. In particular, we wish to examine analogous results for specifications that accord different treatment to earnings in Equation (10). Our principal focus is on τ, the coefficient of marital status in the earnings equation. Thus we shall concentrate on alternative specifications of the Z vector in Equation (4), the marriage reduced form, measuring their impacts on the marriage earnings coefficient. At one extreme is the restriction $\delta = 0$ in Equation (10), which effectively removes determinants of earnings from the marital status equation.

Estimates of this restricted model are presented in Table 4. The probit estimates are qualitatively similar to those in Table 3, with two notable exceptions; the dummy variables for race and urban upbringing increase substantially as deterrents to marriage while remaining significant. The earnings coefficients are similar to their counterparts in Table 3, indicating lack of significance for both marital status and the marriage selectivity term. Thus the results in Table 3 appear to be reasonably robust with respect to the presence of earnings in the structural marriage equation.

To further assess the stability of these estimates, we attempted a series of intermediate restrictions in which some, but not all, determinants of earnings are permitted to appear in the reduced form marriage probit. The resulting estimates for the log earnings model, Equation (11), are presented in the Appendix. In addition, the Appendix includes a series of restrictions in which some predetermined variables in the structural marital status equation, Equation (10), were excluded from the reduced form probit. A summary of these restrictions is presented below:

Earnings Variables Restricted from the Reduced Form Probit: Appendix Table A1	Family Background Variables Restricted from the Reduced Form Probit: Appendix Table A2
DVET	NSIB
EXP, EXPSQ	MED1–MED7
EXP, EXPSQ, DVET	MED1–MED7, FED1–FED7
ED, DVET	DRACE
ED, EXP, EXPSQ	DGREW
	NSIB, DGREW

Table 4
Estimates of the Treatment Effects Model[a]

	Probit: Equation (7)		Log Earnings: Equation (11)	
CONSTANT	1.056	(3.67)	7.655	(9.09)
NSIB	−0.058	(−2.07)		
OSIB	0.123	(0.88)		
DREL	0.184	(1.50)		
DRACE	−0.470	(−3.15)	−0.468	(−2.09)
DGREW	−0.258	(−2.14)		
FED1	−0.119	(−0.61)		
FED2	0.089	(0.38)		
FED3	0.063	(0.28)		
FED4	−0.029	(−0.07)		
FED5	0.033	(0.11)		
FED6	−0.394	(−1.20)		
FED7	−1.040	(−2.12)		
MED1	−0.174	(−0.73)		
MED2	0.017	(0.07)		
MED3	−0.418	(−1.70)		
MED4	−0.420	(−1.11)		
MED5	−0.708	(−2.28)		
MED6	−0.435	(−1.07)		
MED7	−0.111	(0.18)		
ED			−0.056	(−1.20)
EXP			0.553	(2.68)
EXPSQ			−0.049	(−1.64)
DVET			−0.015	(−0.05)
I			0.350	(0.49)
h[b]			0.019	(0.04)
Chi square	58.18	(19 df)		
SEE			2.103	
R square			0.048	

a. Figures in parentheses are t-statistics.
b. Computed from Equation (19).

As noted in the Appendix, results for these models are generally consistent with the two polar cases presented in Tables 3 and 4. Each of the twelve restrictions fails to reject the null hypothesis of a neutral effect of marital status on earnings, despite a persistent lack of precision in the estimates of the selectivity term. In Table A1 the marital status coefficients range from 0.347 to 0.409, with no *t*-statistic exceeding 0.58, while none of the selectivity terms differs significantly from zero; in Table A2 the coefficients are less stable, ranging between 0.247 and 1.002, with a maximum *t*-statistic of 0.94.

Despite the lack of significance of the selectivity terms, evidence from Tables 3, 4, A1, and A2 provides a useful inference about marital status and earnings. When the model is estimated in a manner that addresses the potential for endogenous selection, marital status itself fails to emerge as a *source* of enhanced earnings. The nature of the self-selection process and the observed marriage earnings premium appears to be a useful item for additional research. A considerable body of current evidence suggests that this phenomenon arises in part from marriage-induced changes in labor market behavior of young men. As noted in Section I, Kenny (1983) reports more rapid accumulation of human capital among married males. In addition, Leigh (1978, 109) measures a greater degree of occupational mobility among married respondents to the 1966 and 1969 National Longitudinal Surveys. Weiss (1984) finds lower quit propensities at marginal levels of significance among married workers in a sample of production employees of a U.S. manufacturer, while Gottschalk and Maloney (1985), using a sample from the Michigan Panel Survey, report that married males are significantly more likely to find new jobs within one year of a job separation. Moreover, there is evidence that the stronger job attachment of married males is not confined to the early stages of the labor market experience. Burtless and Moffitt (1985), utilizing a sample of males aged 58–63 from the Longitudinal Retirement History Survey, report that the marriage coefficient in an annual hours of work equation is positive and highly significant. Thus it appears increasingly clear that the labor force attachment of married males differs from that of nonmarrieds. Evidence presented in this section suggests that conventional methods of estimation may conceal the mechanism through which marital status affects earnings.

VII. Summary and Conclusion

It is largely a matter of course in applied labor economics to presume that marital status is a predetermined contributor to earnings of young males. This study finds no evidence to support that proposition, and thus offers potential for fruitful additional research. The process of marital status selection warrants more attention. In addition, based on the premise that marriage offers young men opportunities for accumulation of human capital, the duration of marriage may be found to vary systematically with stocks of human capital in the married population. Among young adults marriages of short duration are likely to end in divorce. An extended model might include structural equations for expected duration and the probability of dissolution. Another useful extension is to include nonemployed males in the sample, allowing the structural earnings equation to be determined by the double selection rule of marital status and employment.

Progress in this area might be enhanced by application of similar models to other data sets that permit analysis of older age cohorts and richer specifications of the marital status and earnings equations. In addition, extension of the reasoning to other determinants of earnings may be fruitful. Economists have recently become interested in the role of health status as a choice variable, and it is conceivable that variables such as language characteristics and religious affiliation may also exhibit some degree of endogeneity. Research in this area will contribute to the continuing refinement of models of earnings determination.

Appendix

As noted in Section VI we attempted two sets of sensitivity tests for the log earnings estimates reported in Tables III and IV. Our purpose was to determine whether the principal conclusion in those tables, that marital status is insignificant in the treatment effects model, is robust with respect to various exclusions in the marital status reduced form.

This Appendix presents two tables summarizing the results of these tests. Table A1 contains results for restrictions in the structural earnings equation [the X-vector in Equation (11)]. Table A2 pertains to restrictions among the family background variables in the structural marital status equation [the Y-vector in Equation (10)]. The column headings indicate which predetermined variables were restricted from the reduced form probit in the course of obtaining the earnings estimates.

The sensitivity tests are motivated in part by the potential for collinearity in models of the type estimated in Section VI. In particular, it is possible that a degree of correlation between the marital status dummy variable (I) and the selectivity term (h) might dilute the significance of both terms. These tests serve to expose the model to varying degrees of collinearity. Results in these tables indicate that the central conclusion of Section VI survives a variety of restrictions in the reduced form. The evidence continues to support the proposition that marital status is not a significant treatment effect in the earnings profiles of young employed males.

Table A1
Estimates of Equation (11): Log Earnings[a]

Variable	DVET	EXP, EXPSQ	ED	EXP, EXPSQ, DVET	ED, DVET	EXP, EXPSQ, ED
CONSTANT	7.651	7.638	7.611	7.650	7.651	7.640
	(9.95)	(9.55)	(9.96)	(9.59)	(10.01)	(9.01)
ED	−0.056	−0.056	−0.055	−0.056	−0.056	−0.056
	(−1.19)	(−1.21)	(−1.18)	(−1.21)	(−1.20)	(−1.19)
EXP	0.555	0.553	0.550	0.553	0.555	0.553
	(2.61)	(2.68)	(2.58)	(2.68)	(2.62)	(2.68)
EXPSQ	−0.049	−0.049	−0.049	−0.049	−0.049	−0.049
	(−1.64)	(−1.63)	(−1.63)	(−1.64)	(−1.64)	(−1.63)
DVET	−0.015	−0.016	−0.012	−0.015	−0.015	−0.017
	(−0.05)	(−0.06)	(−0.04)	(−0.53)	(−0.05)	(−0.06)
DRACE	−0.468	−0.465	−0.458	−0.467	−0.468	−0.465
	(−2.13)	(−2.13)	(−2.13)	(−2.12)	(−2.14)	(−2.10)
I	0.348	0.369	0.409	0.353	0.347	0.368
	(0.49)	(0.54)	(0.58)	(0.51)	(0.49)	(0.52)
h[b]	0.021	0.007	−0.019	0.018	0.021	0.008
	(0.05)	(0.02)	(−0.04)	(0.04)	(0.05)	(0.02)
SEE	2.103	2.103	2.103	2.103	2.103	2.103
R squared	0.059	0.048	0.048	0.048	0.048	0.048

a. Figures in parentheses are *t*-statistics. Column headings refer to variables restricted from the probit reduced form.
b. Computed from Equation (19).

References

Barnow, Burt S., Glen G. Cain, and Arthur S. Goldberger. 1978. "Issues in the Analysis of Selection Bias." Mimeo.

Behrman, Jere R., and Barbara L. Wolfe. 1984. "The Socioeconomic Impact of Schooling in a Developing Country." *Review of Economics and Statistics* 66(2):296–303.

Björklund, Anders, and Robert Moffitt. 1985. "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models." Mimeo.

Borjas, George J. 1983. "Measurement of Race and Gender Wage Differences: Evidence from the Federal Sector." *Industrial and Labor Relations Review* 37(1):79–91.

Burtless, Gary, and Robert A. Moffitt. 1985. "The Joint Choice of Retirement Age and Postretirement Hours of Work." *Journal of Labor Economics* 3(2):209–36.

Table A2
Estimates of Equation (11): Log Earnings[a]

Variable	NSIB	MED1–MED7	MED1–MED7, FED1–FED7	DRACE	DGREW	NSIB DGREW
CONSTANT	7.553	7.656	7.220	7.705	7.565	7.495
	(9.64)	(9.55)	(7.75)	(10.66)	(9.65)	(9.36)
ED	−0.054	−0.056	−0.047	−0.057	−0.054	−0.053
	(−1.15)	(−1.19)	(−0.98)	(−1.22)	(−1.15)	(−1.12)
EXP	0.543	0.556	0.500	0.565	0.544	0.535
	(2.53)	(2.58)	(2.23)	(2.63)	(2.54)	(2.48)
EXPSQ	−0.049	−0.049	−0.047	−0.050	−0.049	−0.048
	(−1.62)	(−1.63)	(−1.55)	(−1.65)	(−1.62)	(−1.61)
DVET	−0.001	−0.020	0.063	−0.341	−0.003	0.010
	(−0.00)	(−0.07)	(0.20)	(−0.02)	(−0.01)	(0.03)
DRACE	−0.444	−0.469	−0.365	−0.471	−0.447	−0.465
	(−2.03)	(−2.11)	(−1.47)	(−2.45)	(−2.04)	(−2.10)
I	0.498	0.340	1.002	0.247	0.479	0.585
	(0.67)	(0.43)	(0.94)	(0.37)	(0.64)	(0.75)
h^{b}	−0.076	0.256	−0.386	0.086	−0.063	−0.130
	(−1.64)	(0.05)	(−0.59)	(0.21)	(−0.13)	(−0.27)
SEE	2.103	2.103	2.102	2.103	2.103	2.103
R squared	0.059	0.059	0.060	0.059	0.059	0.059

a. Figures in parentheses are *t*-statistics. Column headings refer to variables restricted from the probit reduced form.
b. Computed from Equation (19).

Duncan, Gregory M., and Duane E. Leigh. 1985. "The Endogeneity of Union Status: An Empirical Test." *Journal of Labor Economics* 3(3):385–402.
Gottschalk, Peter, and Tim Maloney. 1985. "Involuntary Terminations, Unemployment, and Job Matching: A Test of Job Search Theory." *Journal of Labor Economics* 3(2):109–23.
Grenier, Gilles. 1984. "The Effects of Language Characteristics on the Wages of Hispanic American Males." *Journal of Human Resources* 19(1):35–51.
Heckman, James J. 1978. "Dummy Endogenous Variables in Simultaneous Equation System" *Econometrica* 46(4):931–59.
Kenny, Lawrence W. 1983. "The Accumulation of Human Capital During Marriage by Males." *Economic Inquiry* 21(2):223–31.
Lee, Lung-fei. 1978. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19(2):415–34.

———. 1979. "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables." *Econometrica* 47(4):977–96.

Leigh, Duane E. 1978. *An Analysis of the Determinants of Occupational Upgrading*. New York: Academic Press.

Maddala, G. S. 1978. "Selectivity Problems in Longitudinal Data." *Annales de l'insee* 30–31:423–50.

———. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maddala, G. S., and Lung-fei Lee. 1976. "Recursive Models with Qualitative Endogenous Variables." *Annals of Social and Economic Measurement* 5(4):525–45.

Mincer, Jacob. 1974. *Schooling, Experience and Earnings*. New York: Columbia University Press.

Nakosteen, Robert A., and Michael A. Zimmer. 1982. "The Effects on Earnings of Interregional and Interindustry Migration." *Journal of Regional Science* 22:325–41.

Rosen, Sherwin, and Paul Taubman. 1982. "Changes in Life-Cycle Earnings: What Do Social Security Data Show?" *Journal of Human Resources* 17(3):321–37.

Tomes, Nigel. 1984. "The Effects of Religion and Denomination on Earnings and the Returns to Human Capital." *Journal of Human Resources* 19(4):472–88.

Weiss, Andrew. 1984. "Determinants of Quit Behavior." *Journal of Labor Economics* 2(3):371–87.