Marriage and Earnings

by Christopher Cornwell
and Peter Rupert

Christopher Cornwell is an associate professor of economics at the University of Georgia, Athens, and Peter Rupert is an economic advisor at the Federal Reserve Bank of Cleveland.

Introduction

Uncovering the determinants of earnings is an important and well-researched area in labor economics. In studies of race or sex discrimination, it is imperative to use statistical methods that control for various factors so that the researcher can obtain an unbiased measure of discrimination. Another area that has attracted interest is the interaction of wages and union membership. Again, controlling for certain factors enhances our understanding of how unions affect wages. Of course, knowing exactly *what* to control for is at the heart of the problem. In this paper, we examine one particular control variable that is often used in earnings regressions—an indivdual's marital status. In analyzing and interpreting the results obtained from earnings regressions, we hope to develop a better measure of how any policy affects (or does not affect) individual behavior.

The wage premium attributable to marriage has been well documented in the literature and is typically as large as that associated with union status. The source of this premium, however, remains debatable. Two common explanations of why married men earn more than unmarried men are 1) the division of labor in a married household allocates more of the man's time to the market,[1] and 2) married men have a lower cost of human capital acquisition, since a spouse may be working to help finance the additional human capital.[2] Both of these stories imply that marriage enhances productivity, and therefore wages, as a result of an increase in human capital. We suggest an alternative explanation for the marriage premium, derived from the job-matching literature, in which marriage signals certain unobservable individual characteristics that are valued by employers—including ability, honesty, loyalty, dependability, and

1 As evidence that married men are relatively more market intensive, it is often noted that they work more hours. For example, in the sample we draw from the 1971 wave of the National Longitudinal Survey of Young Men (NLSYM), mean hours per week for married men is 44.17, versus 41.28 for unmarried men. However, after age 22, the mean hours differential drops to less than one hour per week.

2 Some evidence that marriage facilitates human capital acquisition is provided by Kenny (1983), but the potential endogeneity of an individual's marital status is ignored. Furthermore, the argument that marriage makes it cheaper to accumulate human capital is difficult to reconcile with the fact that men who acquire more formal education tend to marry later than those who acquire less (Bergstrom and Schoeni [1992]).

determination.[3] Failure to control for the correlation of the fixed effects with marriage will lead to a bias in the marital status coefficient. That is, some of the returns attributable to marriage will actually be returns to some unobserved qualities correlated with marital status.

Under either of the above scenarios, marital status should not be treated as an exogenous determinant of the wage rate. However, it is assumed to be exogenous in most wage regressions that control for marital status, resulting in an estimated marriage premium that is biased upward.

In this paper, we reexamine the empirical relationship between marriage and wages. Our investigation proceeds along two lines. First, to the extent we can model the process that determines marriage, our cross-section procedures attempt to capture the kind of incentives that the human capital stories imply for the marriage premium. Second, we employ panel data estimation techniques that allow us to control for unobservable individual-specific effects that may be correlated with marital status. If we interpret these individual effects as ability— or as any of the qualities listed above that may lead to better job matches — then panel data estimation addresses a different source of endogeneity than that arising from the human capital arguments.

The clearest picture of the effects of marriage on wages emerges in our panel estimates. In every case where we condition on unobserved individual effects, the estimated marital status coefficient is essentially zero. We argue that these results support the view of marriage as a "signal'' of some underlying characteristics. Furthermore, specification tests confirm the importance of unobserved heterogeneity and reject the exogeneity of marital status.

I. Marriage, Wages, and Individual Effects

Virtually all cross-section wage regressions that control for marital status report a large, statistically significant wage premium for married men. Some of the more prominent examples are discussed by Reed and Harford (1989) and Korenman and Neumark (1991). Our view is that the marriage premium commonly reported in cross-section wage regressions is largely a statistical artifact, at least for young men. The wage premium can be explained in terms of unobservable individual characteristics that are positively correlated with marriage and wages. The characteristics that lead to "good'' (long and stable) marriages are the same characteristics that produce "good'' (long and stable) jobs and higher wages.

This view has some additional support from another strand of the literature relating to the returns to job tenure. Abraham and Farber (1987) propose that workers in long-tenure jobs earn more *in every year* on the job, and that most of the cross-section return to tenure is due to unobserved individual and job-match effects. They test their proposition by estimating wage equations conditional on predicted job duration. Interestingly, the results of their job duration model indicate that marriage has a large and positive statistically significant effect.

Further evidence exists in the quit behavior of married men. Consistent with the positive relationship between marriage and job duration is the depressing effect marriage has on quits. Shaw (1987) reports that the quit rate for married men aged 25–54 is less than half that of unmarried men. He also finds that marriage has its strongest deterrent effect on the quit behavior of younger men.

Only Nakosteen and Zimmer (1987) argue explicitly that marriage does not significantly affect wages. However, the empirical support for their argument is weak. Using a cross-section of 576 employed men between the ages of 18 and 24, taken from the 1977 wave of the Panel Study of Income Dynamics (PSID), they estimate an earnings equation in which marriage is modeled as a treatment effect, thereby making marital status endogenous.[4] When they restrict marital status to be exogenous and apply ordinary least squares (OLS) to their wage equation, Nakosteen and Zimmer obtain a statistically significant marital status coefficient estimate of 0.370. Relaxing the exogeneity restriction actually causes this coefficient estimate to *rise*, although it is no longer statistically significant. Furthermore, specification tests of exogeneity are inconclusive. Nevertheless, Nakosteen and Zimmer find that the true marriage premium is not significantly different from zero.

3 Reed and Harford (1989) provide another alternative—that the marriage premium represents a compensating differential required to induce married men to accept "undesirable" working conditions. In their view, marriage is related to the purchase of costly "family goods" such as children.

4 See Barnow, Cain, and Goldberger (1980) for an explanation of treatment effect models.

Korenman and Neumark, whose results are based on the 1976, 1978, and 1980 waves of the NLSYM, take the opposite stance. They claim that the gains to marriage are large for young (white) men, even after controlling for unobserved heterogeneity. In their preferred specification, years married and its square are included in the wage equation, along with the marital status dummy. Marriage premiums are derived from both cross-section and fixed effects ("within") estimates. The marriage premium at mean years married, calculated from the fixed effects estimates, is 15 percent, which is about 20 percent smaller than the premium yielded by their cross-section estimates.

However, the large gains to marriage reported by Korenman and Neumark may be misleading. First, only the estimated years-married coefficients are statistically significant. This is true in both their cross-section and fixed effects regressions. Thus, more than one-fifth of these premiums are due to shift-parameter estimates that are not significantly different from zero.

Second, when we include years married and its square along with tenure and its square in our sample, the former never enters statistically significantly, while the latter always does. Because married men hold longer jobs (experience less turnover), years married may be playing a role similar to tenure in the Korenman/Neumark regressions.

Third, Korenman and Neumark's results suggest that each additional year of marriage translates into a 1 to 2 percent wage gain. As indicated by Bergstrom and Schoeni (1992), this implies that men who married at age 17 should earn 10 to 20 percent more at age 40 than men who married at age 27, all else equal. However, their results show that, controlling for current age, men who married at 17 make 25 percent *less* on average than men who married at 27.[5] Finally, Cornwell and Rupert (1996) demonstrate that adding another year (1971) to the Korenman/Neumark sample changes the results substantially. This change can be attributed to the fact that most of the marital status changes in their sample represent individuals who are either leaving or entering marriage for the second time.[6]

II. Econometric Framework

Our econometric methodology addresses the possibility of marital status endogeneity in a variety of ways. Each of the techniques has shortcomings as well as merits. Therefore, we cover a wide range of procedures and attempt to derive a consensus regarding the contribution of marital status to earnings.

We begin with a general model of wage determination for married (M) and single (S) men, where the wage-generating process is initially assumed to be different for each type:[7]

$$(1) \quad y_{it}^{M} = \delta^{M} + X_{it}^{M\prime} \beta^{M} + u_{it}^{M}$$

$$(2) \quad y_{it}^{S} = \delta^{S} + X_{it}^{S\prime} \beta^{S} + u_{it}^{S},$$

where i indexes individuals ($i = 1, \ldots , N$) and t indexes time periods ($t = 1, \ldots T$), y_{it} is the natural logarithm of the real wage, X_{it} is a vector of explanatory variables, and the u_{it}'s are disturbances with time-invariant and time-varying components, expressed as

$$(3) \quad u_{it}^{M} = \alpha_{i}^{M} + \epsilon_{it}^{M}$$

$$(4) \quad u_{it}^{S} = \alpha_{i}^{S} + \epsilon_{it}^{S}.$$

The α_i's, which vary over individuals but not over time, capture unobserved, individual-specific attributes that may be valued in both the labor and marriage markets. The ϵ_{it}'s, which vary over individuals and time, reflect aspects of the wage-determining process that can be represented as statistical noise.

In our sample, we cannot reject the null hypothesis that $\beta^M = \beta^S$.[8] Thus, we express the model given in equations (1)–(4) in terms of a single wage equation,

$$(5) \quad y_{it} = \delta^{S} + X_{it}' \beta + (\delta^{M} - \delta^{S}) M_{it} + \eta_{it},$$

where

$$(6) \quad \eta_{it} = u_{it}^{S} + (u_{it}^{M} - u_{it}^{S}) M_{it}$$

and M_{it} is a dummy variable indicating marital status.

5 Bergstrom and Schoeni point out that to reconcile these results, one would have to argue that if men who married at 27 had married at 17, they would have earned 35 to 40 percent more per year than men who actually did marry at 17.

6 By appending the additional (earlier) year to the Korenman/Neumark data set, we can look at the earnings of younger men or, more specifically, at the earnings of men who have never been married. The evidence suggests that to the extent there is a gain to marital status, it is purely an intercept shift, with no additional effect attributed to the number of years married.

7 Our approach is similar to that employed by Robinson (1989) in his analysis of union wage effects.

8 The value of the *F*-statistic from a comparison of regressions of married and single men over the two periods covered in our data set is only 0.75, and the 95 percent critical value is about 2.8.

If the correlation between marriage and the error term is zero (that is, $E[M_{it}\eta_{it}] = 0$), then treating (5) as a standard cross-section wage regression and estimating by OLS produces a consistent estimate of the marriage premium, $(\delta^M - \delta^S)$. However, for reasons outlined in the previous section, this is *not* likely to be the case. Thus, in general, the marriage premium cannot be identified by OLS.

Since M_{it} is unlikely to be orthogonal to η_{it}, identification of the marriage premium is problematic. The difficulty is that the wage differential comprises two terms: $(\delta^M - \delta^S)$ and $(u^M_{it} - u^S_{it})$. The first term is fixed and the same for all men. If men are randomly distributed across married and single states, then this term represents the true wage premium. Even if assignment across married and single states is random, however, standard cross-section estimation may still be inappropriate. If the source of endogeneity of M_{it} is unobserved individual attributes that are valued by employers as well as potential marriage partners, then consistent estimation would imply conditioning on the α_i's. Panel data are important in this regard, a point that we elaborate on below.

If assignment to the married and single states is not random, then identification of the true wage premium is complicated by the second term, $(u^M_{it} - u^S_{it})$. In this case, the expected increase in the wage rate as a result of marriage is $(\delta^M - \delta^S) + E[(u^M_{it} - u^S_{it}) \mid M_{it} = 1]$. Separate identification of $(\delta^M - \delta^S)$ requires additional restrictions on our model so that the process generating marital status can be parameterized. The restrictions may involve orthogonality conditions that define a set of exogenous variables for marital status, or distributional assumptions like bivariate normality of y_{it} and M_{it}. The validity of such restrictions is an empirical question. Imposing them when they are empirically invalid is a misspecification, the statistical consequences of which may be less acceptable than the failure to control for nonrandom assignment.

Alternative Estimators and Specification Tests

We consider three different approaches to estimating the effect of marriage on wages: instrumental variables (IV), the inverse Mills ratio (IMR) method, and methods that exploit the availability of panel data. Each approach imposes a different set of restrictions on the model. The choice of which approach to use is determined largely by whether the restrictions can be justified.

The appeal of both the IV and IMR methods hinges on the ability to specify the process governing marital status.[9] Let the reduced form for this process be expressed as

$$(7) \qquad M^*_{it} = Z'_{it}\gamma + v_{it},$$

where M^*_{it} is a latent index representing the net gain to marriage, Z_{it} is a vector of explanatory variables that includes X_{it}, and v_{it} is a zero mean disturbance with variance σ^2_v. We observe only a discrete realization of M^*_{it}, which we define as the dummy variable M_{it}. Note that $M_{it} = 1$ if $M^*_{it} > 0$, and $M_{it} = 0$ if $M^*_{it} \leq 0$.

Instrumental Variables

The IV procedure exploits the orthogonality condition

$$(8) \qquad E[g(Z_{it})\eta_{it}] = 0,$$

where Z_{it} contains at least one regressor not in X_{it}, and g is some known transformation of Z_{it}. The set of restrictions in (8) are used in (7) to construct an instrumental variable for M_{it} that has been purged of correlation with η_{it}. No distributional assumption about the v_{it}'s is necessary, although one may be imposed. For example, if the v_{it}'s are independently and identically distributed (i.i.d.) standard normal and are independent of Z_{it}, then (7) can be estimated by probit maximum likelihood. The resulting instrument for M_{it} would be $\hat{M}_{it} = \Phi(-Z'_{it}\hat{\gamma})$, where Φ is the standard normal cumulative distribution function. In any case, the instrumental variable is inserted in a least squares regression of y_{it} on $(X_{it}, \hat{M}_{it})$, and the estimated coefficient of $\hat{M}_{it}$ is taken to be the measure of the effect of marriage on wages. However, if the condition in (8) is violated, the IV estimate of the marriage premium will not be consistent. Assuming (8) holds, the IV estimator provides a natural contrast to OLS, thereby providing a Wu–Hausman-type test of the exogeneity of marital status (see Wu [1973] and Hausman [1978]).

9 A detailed discussion of these methods is provided by Heckman and Robb (1985).

Inverse Mills Ratio

The IMR method addresses the endogeneity of marital status in the context of the nonlinear regression function

$$(9)\quad E(y_{it} \mid Z_{it}, M_{it}) = \delta^S + X_{it}'\beta + (\delta^M - \delta^S)M_{it} + \xi h(Z_{it}, M_{it}; \gamma),$$

where

$$(10)\quad h(Z_{it}, M_{it}; \gamma) = M_{it}E(\eta_{it}|Z_{it}, M_{it} = 1) + (1 - M_{it})E(\eta_{it}|Z_{it}, M_{it} = 0),$$

$\xi = \sigma_{\eta v}/\sigma_v^2$, and $\sigma_{\eta v} = \mathrm{cov}(\eta, v)$. Note that h comprises the relevant inverse Mills ratios. If the v_{it}'s are assumed to be i.i.d. standard normal, then the inverse Mills ratios are, respectively, φ/Φ and $-\varphi/(1 - \Phi)$, where φ is the standard normal probability density function, and both φ and Φ are evaluated at $-Z_{it}'\gamma$.

This method is designed to correct for the self-selection of individuals into marriage; that is, it accounts for whether there is something different about individuals who marry versus those who do not. Estimation of (9) can be accomplished through a simple two-step procedure. First, probit maximum likelihood is applied to (7) to obtain an estimate of γ, which is used to construct $\hat{h}(Z_{it}, M_{it}; \hat{\gamma})$. Then, $\hat{h}$ is substituted for h in (9), and the resulting regression is estimated by least squares.[10] The statistical significance of $\hat{h}$ in this regression provides another test of the exogeneity of marital status.

Consistency of the IMR method also depends on whether (8) is satisfied, as well as on knowledge of the functional form of h. In a given empirical application, the IMR method may not be robust to departure from the functional-form assumption (say, normality). Another problem is that the inverse Mills ratios may simply be proxying for omitted nonlinearities, in which case interpretation as a correction for self-selection becomes difficult.

Panel Data

With panel data, alternative estimators to the cross-section approaches of IV and IMR exist. For convenience, assume that $u_{it}^M = u_{it}^S \;\forall\; i$. This will have no adverse statistical consequences if 1) men have only the same knowledge as the econometrician regarding their ε_{it}'s, but have exclusive knowledge of their α_i's, and 2) men's unobservable characteristics have the same impact on wages regardless of whether an individual is married or single (Robinson [1989]). Then, (6) becomes

$$(11)\quad y_{it} = \delta^S + X_{it}'\beta + (\delta^M - \delta^S)M_{it} + \alpha_i + \varepsilon_{it}.$$

Finally, assume that the α_i's and ε_{it}'s are i.i.d. random variables, uncorrelated with each other, with zero means and constant variances σ_α^2 and σ_ε^2.

In a standard human capital wage equation, it is likely that $E(X_{it}'\alpha_i) \neq 0$, since X_{it} typically contains measures of education, work experience, and job tenure that are correlated with unobserved ability reflected in the α_i's. Furthermore, because men with large amounts of human capital are more attractive as potential spouses, it is likely that X_{it} and M_{it} are correlated. Hence, attempts to estimate the effect of marriage on wages should go beyond correction for self-selection per se.

The "Within" Estimator

The simplest procedure for consistently estimating (11) is the so-called within estimator, which is calculated by applying least squares to data that have been transformed into deviations from individual means. Since the α_i's are treated as fixed parameters, the within estimator is consistent regardless of the relationship between (X_{it}', M_{it}) and α_i. Alternatively, the within estimator can be viewed as an instrumental variables estimator, with the instruments being the deviations from the means, which are orthogonal to the α_i's by construction.

Of course, if the α_i's are uncorrelated with (X_{it}', M_{it}), more efficient estimation of (11) is possible via generalized least squares (GLS). Like the within estimator, GLS can also be computed from a least squares regression on transformed data. In this case, the transformation is to "whiten" the errors.[11] Another advantage of panel data is that this proposition can be tested. The efficiency of GLS under the null hypothesis

10 Other estimation strategies include direct nonlinear least squares estimation of (9) and maximum likelihood. To estimate by maximum likelihood, one must be willing to assume that (η, ν) is distributed bivariate normal. See Barnow, Cain, and Goldberger (1980) and Heckman and Robb (1985) for details.

11 This transformation follows from Fuller and Battese (1973) and amounts to a "$(1 - \theta)$ differencing" of the data, where $\theta = [\sigma_\epsilon^2/(\sigma_\epsilon^2 + T\sigma_\alpha^2)]^{1/2}$.

TABLE 1

Means and Standard Deviations by Year, and Percent of Sample Changing Marital Status

Variable	1971 Cross Section	1971 Panel	1976 Panel
LOG REAL WAGE	1.122 (0.438)	1.179 (0.437)	1.399 (0.414)
MARITAL STATUS	0.718 (0.450)	0.716 (0.451)	0.895 (0.306)
AGE	23.96 (3.18)	24.00 (3.18)	29.00 (3.18)
TENURE	3.209 (2.532)	3.321 (2.551)	6.197 (4.032)
UNION	0.314 (0.464)	0.326 (0.469)	0.342 (0.475)
SOUTH	0.360 (0.480)	0.367 (0.482)	0.379 (0.485)
SMSA	0.679 (0.467)	0.687 (0.464)	0.688 (0.463)
EDUCATION	12.738 (2.519)	12.802 (2.467)	13.321 (2.620)
RACE	0.141 (0.348)	0.149 (0.356)	0.149 (0.356)
No. of observations	1,073	860	860

Marital Status Changers	Never Married to Married	Married to Not Married
Frequency	197	43
Percent of observations	22.9	5.0

NOTE: Standard errors are in parentheses.
SOURCE: National Longitudinal Survey of Young Men.

that $E[(X'_{it}, M_{it})\alpha_i] = 0$, along with the robustness of the within estimator to departures from the null, form the basis of a Wu–Hausman test of the difference between the GLS and within estimates of (11).

The Hausman–Taylor Estimator

A consistent and potentially more efficient alternative to the within estimator is Hausman and Taylor's (1981) efficient instrumental variables procedure (HT-IV).[12] HT-IV exploits knowledge of the uncorrelatedness of certain columns of X'_{it} with α_i to increase the instrument set beyond that of the within estimator.[13] Computation involves an IV regression on the same data transformation required for GLS. Thus, the efficiency gains to HT-IV come from an expanded instrument set and whitened errors.

More efficient estimation of the parameters of (11) is one motivation for considering HT-IV. Another is that HT-IV permits a direct test of the uncorrelatedness of M_{it} with α_i. The GLS/within contrast does not indicate which explanatory variables are correlated with the effects if the null is rejected. However, using a Wu–Hausman test of the difference between HT-IV estimates that maintain a legitimate set of over-identifying restrictions with HT-IV estimates that take M_{it} to be exogenous, we can determine whether marital status is in fact exogenous.

Note that both the within and HT-IV estimators are distinguished from the cross-section procedures, which depend on over-identifying restrictions like those defined in (8) and/or on special distributional assumptions. In addition, within and HT-IV allow consideration of another source of nonidentifiability of the marriage premium that is not addressed by the cross-section approaches, namely, correlation between X_{it} and the effects. Consistent estimation of $(\delta^M - \delta^S)$ is not possible if $E(X'_{it}\alpha_i) \neq 0$, unless X_{it} is orthogonal to M_{it}, *even if* $E(M_{it}\alpha_i) = 0$.

III. Estimation of the Marriage Premium

Data

Our data are drawn from the 1971 and 1976 waves of the NLSYM. The primary advantage of the NLSYM is that it allows us to follow individuals moving from the single (and not previously married) to the married state. Between 1971 and 1976, roughly 23 percent of the young men in our sample went from never married to married, while only about 5 percent were divorced or separated. The 1971 cross-section data set consists of 1,073 young men who were between the ages of 19 and 29 in 1971 and who worked more than 40 weeks during the year. Our panel is constructed from the individuals in the cross-section data set who are also observed in 1976. Attrition reduces the number of observations in both periods to 860, of which about 15 percent are black.

In addition to marital status, we observe (in both years) each man's wage, age (AGE), years

12 Extensions of this procedure to broader classes of instrument sets can be found in Amemiya and MaCurdy (1986) and Breusch, Mizon, and Schmidt (1989).

13 HT-IV uses the individual means and deviations from means of each time-varying exogenous variable in X_{it} as separate instruments (see Breusch, Mizon, and Schmidt [1989]).

TABLE 2

Cross-Section Marital Status Premiums from the 1971 NLSYM

	OLS	IV1	IV2	IMR
CONSTANT	–2.156	–5.381	–4.858	–3.835
	(0.697)	(2.721)	(2.278)	(1.330)
MARITAL STATUS	0.054	–0.691	–0.570	–0.334
	(0.025)	(0.554)	(0.488)	(0.253)
AGE	0.188	0.471	0.425	0.335
	(0.059)	(0.238)	(0.199)	(0.115)
AGE^2	–0.003	–0.008	–0.007	–0.006
	(0.001)	(0.004)	(0.004)	(0.002)
UNION	0.214	0.223	0.221	0.219
	(0.023)	(0.032)	(0.030)	(0.026)
SOUTH	–0.122	–0.095	–0.100	–0.108
	(0.024)	(0.038)	(0.035)	(0.028)
SMSA	0.157	0.156	0.156	0.156
	(0.023)	(0.030)	(0.028)	(0.025)
EDUCATION	0.032	0.027	0.028	0.029
	(0.004)	(0.007)	(0.007)	(0.005)
RACE	–0.169	–0.287	–0.268	–0.231
	(0.033)	(0.103)	(0.087)	(0.053)
h				0.196
				(0.149)

NOTE: Standard errors are in parentheses.
SOURCE: National Longitudinal Survey of Young Men.

of tenure with his employer (TENURE), years of education (EDUCATION), union status (UNION = 1 if a union member), race (RACE = 1 if black), and residence (SOUTH = 1 if resident of the South; SMSA = 1 if resident of a Standard Metropolitan Statistical Area).[14] The variable MARITAL STATUS has a value of 1 if the individual is married and living with his spouse, and zero otherwise. Table 1 provides the means and standard deviations for the variables in our samples, and documents the percentage of men who change their marital status.

Results

First, we estimate our model using purely cross-sectional methods. In each case, X_{it} contains the explanatory variables AGE, AGE^2, UNION, SOUTH, SMSA, EDUCATION, and RACE. The dependent variable in every regression is the natural logarithm of the real wage. The results of cross-section estimation are presented in table 2. The OLS estimates, given in the first column, serve as a baseline for comparing estimates obtained from procedures that allow for marital status endogeneity.

In general, the results of our OLS regressions are typical of those found elsewhere in the literature. Of interest here is the coefficient of marital status, which is estimated to be 0.054 with a standard error of 0.025 (statistically significant at the 5 percent level). We have argued that the marriage premium estimated by OLS is biased upward due to the endogeneity of marital status. A Wu–Hausman test of the exogeneity of marital status can be performed by comparing OLS and IV estimates of our model.

To construct an instrument for M_{it}, we need an empirical specification of the decision rule that determines marriage (equation [7]); that is, a regressor (or set of regressors) that is not included in X_{it}. From the NLSYM, we obtain two family-background variables for this purpose: number of siblings (NSIB) and years of education of the father (FED). The Wu–Hausman test amounts to a test of significance of $\hat{M}_{it}$ in a regression of y_{it} on $(X'_{it}, M_{it}, \hat{M}_{it})$. We conduct the test using a $\hat{M}_{it}$ calculated from both OLS and probit maximum likelihood estimates of γ. The p-values of the test statistics for the null hypothesis that the coefficient of $\hat{M}_{it}$ equals zero are 0.089 and 0.108, respectively. Taken together, these tests provide weak evidence against the exogeneity of marital status.

Instrumental Variables

Assuming that marital status is endogenous, we turn to the IV estimates of our model, which are presented in the second and third columns of table 2. The results in the IV1 column are based on a $\hat{M}_{it}$ obtained by OLS, while those in the IV2 column are based on a $\hat{M}_{it}$ derived from probit maximum likelihood. In both cases, we fail to reject, even at a 10 percent level of significance, the null hypothesis that the estimated marital status coefficient is zero. Moreover, both IV marital status coefficient estimates are less than zero and are large in absolute value.

Consistency of these IV estimates depends on the validity of the over-identifying restrictions exploited in the construction of $\hat{M}_{it}$. A test of the over-identifying restrictions examines whether (NSIB, FED) jointly adds to the

14 The wage-rate variable provided by the NLSYM is the hourly wage for hourly workers and salary/usual hours for salaried workers. Our wage-rate variable is deflated by the Consumer Price Index and is expressed in 1970 dollars.

predictive power of our model. The value of the test statistic, which is distributed as $F_{3,992}$, is 2.337. Hence, the restrictions cannot be rejected at the 5 percent level of significance. On the other hand, the p-value of the F-statistic is only 0.949, and only NSIB (aside from certain variables in X_{it}) enters (7) statistically significantly. This may be part of the reason that, using our IV procedures, we are unable to arrive at any stronger conclusions about the impact of marital status endogeneity on the estimated marriage premium.[15]

Inverse Mills Ratio

As an alternative to IV, we estimate our model using the two-step IMR procedure. First, we estimate (7) by probit maximum likelihood to obtain an estimate of γ. This estimate is then used to form the estimates of the relevant IMR terms that comprise h in (10). Second, we apply OLS to (9), where $\hat{h}$ is substituted for h. The results of this two-step procedure are given in the fourth column of table 2. The marital status coefficient estimate is –0.334 and, like the IV estimates, is not statistically significant. Furthermore, the "selection term," $\hat{h}$, does not enter the regression statistically significantly, thereby providing no (additional) evidence of the endogeneity of marital status.[16]

In sum, cross-sectional approaches to the identification of the marriage premium yield no definitive conclusions. The evidence against the exogeneity of marital status is relatively weak. When marital status is treated as endogenous, the resulting coefficient estimates are large and negative. Moreover, when the estimated marital status coefficient is large in absolute value, the fact that it is not statistically different from zero may have little meaning.

Next, we consider estimation techniques that exploit the availability of panel data.

Panel Data

We proceed under the assumptions of (11) and focus on the role of unobservable individual-specific characteristics in estimating the return to marriage. The advantage of panel data is the ability to control for such characteristics, which may be correlated with (X_{it}', M_{it}). Consequently, we can address another likely source of non-identifiability of the marriage premium, namely, the potential correlation of X_{it} with the effects. Furthermore, our panel data procedures do not require the data to satisfy orthogonality conditions such as those in (8), or to meet any special distributional assumptions. This is important, since the cross-section estimates appear to be sensitive to these kinds of restrictions.

Under the null hypothesis that the explanatory variables are uncorrelated with the effects, OLS applied to the individual means of the variables (the so-called between estimator) yields consistent estimates of all the coefficients in our model. Conditional on X_{it} as defined above, we estimate our model with our panel data set using the between estimator as a basis for comparison with procedures that are consistent under the alternative hypothesis.

The between estimates, which exploit only the cross-section dimension of the panel, are presented in the first two columns of table 3. Like the OLS estimates in table 2, they are typical of cross-section estimates derived from human capital wage equations.

Focusing on the returns to marriage, the first column reports a statistically significant marriage premium of 7 percent. However, the regression associated with the first column does not condition on tenure. When TENURE and TENURE^2 are included, the estimated MARITAL STATUS coefficient declines by one-third and is no longer statistically significant. At the same time, the estimated returns to tenure are sizable. This pattern is repeated in Korenman and Neumark when they introduce years married to their regressions. But if Abraham and Farber are correct, tenure and years married simply capture individual characteristics that lead to longer-lasting and higher-wage jobs.

This might be expected given the evidence cited in Abraham and Farber (1987) on the relationship between marital status and job duration (completed tenure). Table 4 presents further evidence on this relationship from our cross-section data set. Here, we report the results obtained from estimating an exponential hazard of job duration (correcting for right-censoring) conditional on MARITAL STATUS, AGE, AGE^2, UNION, BLUE COLLAR (= 1 if the individual is employed in a blue-collar occupation), EDUCATION, and RACE. Other than the effect of union membership, marriage has the largest positive impact on job duration.

15 The instrument set (NSIB, FED) represents our best attempt, using our 1971 NLSYM cross-section sample, at a specification of (7).

16 We also estimated (9) by maximum likelihood, which assumes (η_{it}, ν_{it}) to be bivariate normal. The maximum likelihood estimate of $(\delta^M - \delta^S)$ is 0.032 with a standard error of 0.033, so it is not statistically significant. In addition, the correlation coefficient $\rho_{\eta\nu}$ is only –0.117 and is also not statistically significant.

TABLE 3

Estimated Marital Status Premiums from the 1971 and 1976 NLSYM

	Between	Between	Within	HT–IV
CONSTANT	–1.908	–1.851		–0.968
	(0.809)	(0.794)		(0.161)
MARITAL STATUS	0.070	0.0461	–0.012	–0.017
	(0.037)	(0.036)	(0.025)	(0.024)
AGE	0.150	0.141	0.118	0.124
	(0.062)	(0.061)	(0.022)	(0.021)
AGE^2	0.002	–0.002	–0.002	–0.002
	(0.001)	(0.001)	(0.000)	(0.000)
TENURE		0.055	0.015	0.013
		(0.012)	(0.007)	(0.007)
TENURE^2		–0.003	–0.001	–0.001
		(0.001)	(0.000)	(0.000)
UNION	0.162	0.142	0.190	0.181
	(0.025)	(0.025)	(0.027)	(0.020)
SOUTH	–0.093	–0.084	–0.058	–0.082
	(0.023)	(0.023)	(0.056)	(0.022)
SMSA	0.198	0.197	0.070	0.158
	(0.023)	(0.023)	(0.041)	(0.021)
EDUCATION	0.040	0.041	0.045	0.054
	(0.004)	(0.004)	(0.014)	(0.012)
RACE	–0.203	–0.202		–0.193
	(0.031)	(0.030)		(0.035)
χ^2			34.770	8.030

NOTE: Standard errors are in parentheses.
SOURCE: National Longitudinal Survey of Young Men.

TABLE 4

Exponential Hazard of Job Duration

	Coefficient Estimate	Standard Error
CONSTANT	1.526	2.412
MARITAL STATUS	0.284	0.086
AGE	–0.076	0.205
AGE^2	0.004	0.004
UNION	0.512	0.083
BLUE COLLAR	–0.012	0.087
EDUCATION	–0.008	0.018
RACE	0.125	0.107
Log-likelihood	–1,597	
No. of observations	1,073	

SOURCE: National Longitudinal Survey of Young Men.

Within Estimates

The between estimates are consistent only if (X'_{it}, M_{it}) is orthogonal to the effects. The within estimates, given in the third column of table 3, are consistent regardless of whether the effects are correlated with the explanatory variables. Our within regression yields an estimated marriage premium that is essentially zero. The MARITAL STATUS coefficient estimate is –0.012 with a standard error of 0.025. Conditioning on unobserved heterogeneity has a similar effect on the return to tenure, as the estimated TENURE coefficient declines from 0.055 to 0.015. One interpretation of these results, consistent with Abraham and Farber, is that the cross-section gains to marriage and tenure largely reflect unobserved individual characteristics that are valued by the firm.

The outcomes of two specification tests support this interpretation. First, the assumption of the between regressions that (X'_{it}, M_{it}) is orthogonal to the effects is soundly rejected. A Hausman test of the difference in the between and within estimates yields a test statistic that is asymptotically distributed as χ^2_9 and that has a value of 34.770. The interpretation of the test statistic is as follows: The within estimates are consistent under the null or alternative hypothesis, while the between estimates are consistent only if there is no correlation between (X'_{it}, M_{it}) and α_i, implying that the estimates should be somewhat close if no correlation exists. However, the statistic reported above shows that the estimates are quite far apart, leading us to reject the assumption of no correlation.

Hausman–Taylor Estimates

While the Hausman test demonstrates the general importance of unobserved heterogeneity in our model, it does not specifically address the uncorrelatedness of M_{it} with α_i. This is accomplished by appealing to the HT-IV estimator of Hausman and Taylor (1981). Efficiency gains over the within estimator are obtained by exploiting information about the uncorrelatedness of certain columns of X'_{it} with α_i and accounting for the variance components. The null hypothesis that M_{it} is orthogonal to α_i can be tested by comparing HT-IV estimates derived from a legitimate set of over-identifying restrictions with HT-IV estimates computed assuming M_{it} is exogenous.

Hausman tests contrasting HT-IV and within estimates reveal that a legitimate instrument can be constructed by assuming that AGE, AGE^2, UNION, SOUTH, SMSA, and RACE are uncorrelated with the α_i's. This result comes from the fact that the within estimator is consistent regardless of the correlation of the α_i's with the above variables, while the HT-IV's are not. Therefore, if the results are close in some sense (as defined by a Hausman test, for example), that would lead to the acceptance of the variables being uncorrelated with the α_i's. The test statistic, which is asymptotically distributed as χ^2_5, has a value of 8.03, which is well within the 95 percent statistic of 12.8. The HT-IV estimates based on this instrument set are presented in the last column of table 3. In general, they corroborate the results from within estimation. The estimated marital status coefficient is again very small and statistically insignificant. The HT-IV tenure coefficient estimate is 0.013, quite close to the within estimate.

A Hausman test of the difference between the HT-IV estimates in the last column of table 3 and those obtained from HT-IV estimation with marital status added to the instrument set rejects the exogeneity of marital status. The test statistic equals 3.89 and is asymptotically distributed as χ^2_1 with a p-value of 0.048.

IV. Conclusion

We use cross-section and panel data estimation procedures to determine the effect of marriage on the wages of a sample of young men drawn from the NLSYM. Whenever we control for unobservable individual effects, the estimated returns to marriage are virtually zero. In addition, specification tests reject the hypothesis that marital status is uncorrelated with the effects. We conclude that the usual cross-section marriage premium is essentially a statistical artifact, at least for young men. Within the job-matching literature, this conclusion has a reasonable interpretation: As an explanatory variable in human capital wage equations, marital status appears to fulfill a role similar to that of tenure, namely, proxying for unobservable individual-specific characteristics that are valued by the firm.

References

Abraham, K.G., and H.S. Farber. "Job Duration, Seniority, and Earnings," *American Economic Review*, vol. 77, no. 3 (June 1987), pp. 278–97.

Amemiya, T., and T.E. MaCurdy. "Instrumental-Variable Estimation of an Error Components Model," *Econometrica*, vol. 54, no. 4 (July 1986), pp. 869–80.

Barnow, B.S., G.G. Cain, and A.S. Goldberger. "Issues in the Analysis of Selectivity Bias," *Evaluation Studies Review Annual*, vol. 5 (1980), pp. 43–59.

Bergstrom, T., and R. Schoeni. "Income Prospects and Age at Marriage," University of Michigan, Center for Research and Economic Social Theory, Working Paper 92-10, March 1992.

Breusch, T.S., G.E. Mizon, and P. Schmidt. "Efficient Estimation Using Panel Data," *Econometrica*, vol. 57, no. 3 (May 1989), pp. 695–700.

Cornwell, C., and P. Rupert. "Unobservable Individual Effects, Marriage, and the Earnings of Young Men," *Economic Inquiry*, 1996, forthcoming.

Fuller, W., and G. Battese. "Transformations for Estimation of Linear Models with Nested-Error Structure," *Journal of the American Statistical Association*, vol. 68, no. 343 (September 1973), pp. 626–32.

Hausman, J.A. "Specification Tests in Econometrics," *Econometrica*, vol. 46, no. 6 (November 1978), pp. 1251–72.

_________, and W. Taylor. "Panel Data and Unobservable Individual Effects," *Econometrica*, vol. 49, no. 6 (November 1981), pp. 1377–98.

Heckman, J., and R. Robb. "Alternative Methods for Evaluating the Impact of Interventions: An Overview," *Journal of Econometrics*, vol. 30, nos. 1 and 2 (October/November 1985), pp. 239–67.

Kenny, L.W. "The Accumulation of Human Capital during Marriage by Males," *Economic Inquiry*, vol. 21, no. 2 (April 1983), pp. 223–31.

Korenman, S., and D. Neumark. "Does Marriage Really Make Men More Productive?" *Journal of Human Resources*, vol. 26, no. 2 (Spring 1991), pp. 282–307.

Nakosteen, R.A., and M.A. Zimmer. "Marital Status and Earnings of Young Men: A Model with Endogenous Selection," *Journal of Human Resources*, vol. 22, no. 2 (Spring 1987), pp. 248–68.

Reed, W., and K. Harford. "The Marriage Premium and Compensating Wage Differentials," *Journal of Population Economics*, vol. 2, no. 4 (1989), pp. 237–65.

Robinson, C. "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models," *Journal of Political Economy*, vol. 97, no. 3 (June 1989), pp. 639–67.

Shaw, K. "The Quit Propensity of Married Men," *Journal of Labor Economics*, vol. 5, no. 4, part 1 (October 1987), pp. 533–60.

Wu, D. "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica*, vol. 41, no. 4 (July 1973), pp. 733–50.