Marital Status and Unobserved Heterogeneity
- Do twins suggest a genetic inheritance?

by William Nilsson[1]

Department of Economics, Umeå University, Sweden and

Centre for Research on Welfare Economics (CREB), Barcelona, Spain

Abstract

The purpose of this study is to reveal the importance of unobserved heterogeneity on marital status. The results for monozygotic (identical) and dizygotic twins indicate an important genetic component in marital status. Simply controlling for marital status could be misleading.

JEL: J12

Keywords: marriage, divorce, heterogeneity, selection, twins

Umeå Economic Studies 685, 2006

1. Introduction

An increasing amount of empirical literature studies the effects of marital status, such as marriage, divorce or living with a partner, on earnings, health status and labor supply. Marital status is, however, a choice variable and this should be considered in the econometrical specification. Otherwise, the results could be biased due to endogeneity. In addition to this, important unobserved heterogeneity could be left out which could lead to further bias, due to omitted variables.

The purpose of this study is to use twin data to reveal the importance of (usually) omitted background variables, including genes, on marital status. Bivariate probit models are estimated separately for male and female monozygotic (identical) and dizygotic (fraternal) twin siblings' marital status. The models include variables that typically can be found in data sets used in the literature. The correlation of the error terms on the marital status equations for the twin siblings gives an omnibus measure of the importance of common unobserved heterogeneity. In addition, measures for the aggregate background effect on marital status are also estimated from the models.

2. Model

The latent propensity to be married, M_{sit}^{*}, is assumed to be determined by

$$M_{1it}^{*} = \boldsymbol{\beta}_{1}^{'}\mathbf{x}_{1it} + \eta_{1i} + \xi_{1it} \tag{1}$$
$$M_{2it}^{*} = \boldsymbol{\beta}_{2}^{'}\mathbf{x}_{2it} + \eta_{2i} + \xi_{2it} \tag{2}$$

$$M_{sit} = I(M_{sit}^{*} > 0)$$
$$s = 1, 2, \ i = 1,...,N, \ \text{and} \ t = 1,...,T$$

Subindex $s = 1$ indicates the first group of twins, and $s = 2$ indicates the second group of twins consisting of the sibling twins of $s = 1$. The model, accordingly, consists of an equation for each twin sibling. Explanatory variables are included in $\mathbf{x}_{sit}$ and $\boldsymbol{\beta}_{s}$ are vectors of parameters to be estimated. Subindex i refers to the pair of twins $i = 1,…,N$ and

$t = 1,\ldots,T$ refers to different periods. Individuals are either observed to be married ($M_{sit} = 1$), or not ($M_{sit} = 0$). $I(M_{sit}^* > 0)$ is an indicator function which takes the value 1 if the inequality is satisfied, and zero otherwise. The error term, v_{sit}, includes an individual specific effect, η_{si}, and an orthogonal white noise, ξ_{sit}. The error term is assumed $v_s \sim N(0,1)$. In addition, $Cov(v_1, v_2) = \rho$. The correlation between the error terms is interesting since it measure to what extent unobservables are correlated for twin siblings. Any factor that is shared by the twins and influences their marital status, apart from what is captured by the included variables, will contribute to the coefficient. In addition to this measure, it is also possible to estimate the Aggregate Background Effect,

$$ABE_1 = \left(\frac{\sum\limits_{i \in (M_{2it}=1)} \Pr(M_{1it}=1|M_{2it}=1)}{\sum\limits_i M_{2it}} \right) - \left(\frac{\sum\limits_{i \in (M_{2it}=0)} \Pr(M_{1it}=1|M_{2it}=0)}{\sum\limits_i (1-M_{2it})} \right)$$

where the probabilities are calculated from the estimated bivariate probit model as indicated below;

$$\Pr(M_{1it} = 1 \mid M_{2it} = 1) = \frac{\Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_{1it}, \boldsymbol{\beta}_2'\mathbf{x}_{2it}; \rho)}{\Phi(\boldsymbol{\beta}_2'\mathbf{x}_{2it})} \tag{3}$$

$$\Pr(M_{1it} = 1 \mid M_{2it} = 0) = \frac{\Phi_2(\boldsymbol{\beta}_1'\mathbf{x}_{1it}, -\boldsymbol{\beta}_2'\mathbf{x}_{2it}; -\rho)}{\Phi(-\boldsymbol{\beta}_2'\mathbf{x}_{2it})} \tag{4}$$

$\Phi(.)$ and $\Phi_2(.)$ are the cumulative density functions of univariate and bivariate standard normal distributions. The Aggregate Background Effect compares the average probability to be married for two different groups, i.e. depending on the marital status of the twin sibling. (Note that ABE_2 can be estimated by conditioning on the *first* twin). If the background is important for the marital status, observing one twin as married (M_{sit}), divorced (D_{sit}) or having a spouse (S_{sit}), will increase the probability of witnessing his twin sibling in the same marital status. If the background hardly influences the marital

status at all, the average probability will be similar for the different groups and the Aggregate Background Effect will be close to zero.

3. Data

This study uses information from the Swedish Twin Registry, (STR), and Statistics Sweden, (SCB). 1001 female and 851 male pairs of monozygotic twins that were born between 1949 and 1958 are available. In addition, 1361 female and 1421 male pairs of dizygotic twins are available. Monozygotic twins have identical genes, while dizygotic twins are no more genetically alike than ordinary siblings. It is therefore interesting to compare estimates for monozygotic and dizygotic twins. In addition to information about the twins, biological parents and social siblings are connected to the twins. Social siblings are observed in the Population and Housing Census for the years 1965, 1970 and 1975. To all individuals in the population information from the Income and Wealth Register is included for the years 1968, 1971, 1974, 1977, 1980 and 1983. Variables from the longitudinal database LOUISE are included for the years 1994-1999. During this period, i.e. when the twins were between 36 and 50, marital status is observed. In this study three different measures are used. Apart from being married or being divorced, the twins can also be identified as living with a spouse. A spouse is defined as a legal marriage or a twin living together with a partner, with whom he/she has at least one common child. For the female (male) monozygotic sample, about 62 (59) percent were married, 14 (10) percent divorced and 67 (66) percent were identified as living with a spouse.

4. Results

Results from the bivariate probit models for the female and male monozygotic samples are included in Table 1. Whether the biological father or mother was observed as divorced or not when the twins were between 20 and 22 years old seems to matter for the marital status. Note, however, that a significant coefficient does not reveal if it is the event, or a biologically genetic motive for a correlation. No marginal effects are calculated since the main interest is ρ and the measure of Aggregate Background Effect. These results are included in Table 2, together with the same measures for the dizygotic samples.

For the female monozygotic sample, ρ is large and significantly different from zero in all cases. The corresponding coefficients for the dizygotic sample are estimated to be smaller, and in fact, only significantly different from zero when the probability of being married is estimated. The results indicate an important role for genetics in determining marital status in adulthood for the female sample. The measures for Aggregate Background Effects are also substantially larger for the monozygotic sample.

ρ is found to be significantly different from zero for both male monozygotic and dizygotic samples and when both marriage and spousal status are studied. ρ is also found to be positive when the bivariate probit models for divorce are estimated, but in these cases it is not significantly different from zero. The Aggregate Background Effects are again found to be quite large and this suggests that the background is important for studies of marital status.

ρ in all cases is estimated to be larger for the monozygotic twins. It is, of course, difficult to ascribe all of the differences between the samples to the genes, since it cannot be ruled out that monozygotic twins, in general, are treated more alike. The conclusion that genetics play an important role seems, however, to be robust. The reason is that dizygotic twins also share genes, although to a lesser extent, and this also contributes to the correlation for the dizygotic sample.

5. Concluding remarks

The results in this study underline that controlling for unobserved heterogeneity in studies that estimate the causal effects of marital status is indeed, very important. The results indicate an important role for genes. The social background, including the genes, is likely to affect health, labor supply and labor income in adulthood and the risk for omitted variables is evident. While the social background and the genes are fixed over the life-time, it is important to note that a fixed-effect model that uses variation over time, can still fail to control for unobserved heterogeneity. Bergstrom & Bagnoli (1993) suggest in a theoretical model that "males who regard their prospects as unusually good choose to wait until their economic success is revealed before choosing a bride". Comparing the income before and after marriage would, accordingly, be misleading as an estimate for the causal effect of

marriage. It is possible that a fixed-effect method would rather capture unobserved ability that has different returns on the labor market and the marriage market at different ages. Further, a fixed-effect model cannot handle reverse causality, i.e. a good situation on the labor market can increase the chance of becoming married.

Instrumental variables (IV) methods that uses exogenous variation in marital status can also fail to measure the causal effect of marriage. First, it is in general, difficult to find valid instruments. It is also possible that the exogenous information only affects a small share of the population. In that case the interpretation of the causal effect could be limited to that specific group. If the effect of marital status varies among individuals and individuals are aware of this when deciding to marry or not, IV could give misleading results (Heckman, 1997). It is not unreasonable that this is the case.

The results in this study suggest that empirical studies of the effects of marital status should deal with unobserved heterogeneity. The degree of unobserved heterogeneity found by comparing twins is of course specific to the sample. It would, accordingly, be informative to see results for different ages and different countries.

References

Bergstrom, T. C. and M. Bagnoli, 1993, Courtship as a Waiting Game, *Journal of Political Economy* 101(1), 185-202.

Heckman J. J., 1997, Instrumental Variables: A Study of Implicit Behavioral Assumptions in One Widely Used Estimator, *Journal of Human Resources* 32(3), 441-462.

Table 1 – Estimates from bivariate probit models for monozygotic twins.

Variable	Female sample Divorced $D_{st}=1$ $(\beta_1=\beta_2)^a$	Female sample Married $M_{st}=1$ $(\beta_1=\beta_2)^a$	Male sample Divorced $D_{st}=1$ $(\beta_1=\beta_2)^a$	Male sample Married $M_{st}=1$ $(\beta_1=\beta_2)^a$
Biological father or mother was divorced, when twin observed between age 20 and 22.	0.1556 (1.35)	-0.3154*** (-3.06)	0.2744** (2.09)	-0.1567 (-1.36)
Biological father not identified or dead before twin turned 25	0.2151 (1.59)	-0.0761 (-0.54)	-0.1046 (-0.49)	0.1656 (0.94)
Biological mother not identified or dead before twin turned 25	-0.0096 (-0.04)	0.0697 (0.30)	0.2416 (0.92)	0.1665 (0.66)
Relative averaged income for biological father when twin was 17-25 [c]	-0.1481*** (-2.85)	0.0587 (1.37)	0.0204 (0.45)	0.1469*** (3.16)
Twin's oldest sibling was an older social sister, when observed between 12 and 16 years old.	-0.1003 (-1.02)	0.0862 (1.02)	-0.0339 (-0.33)	0.1335 (1.54)
Twin 41-45 years old [d]	0.2590*** (4.38)	0.0879* (1.82)	0.0322 (0.47)	0.1780*** (3.50)
Twin 46-50 years old [d]	0.4091*** (4.85)	0.1359* (1.84)	0.1818* (1.99)	0.2889*** (3.60)
Education, B (1=upper secondary school) [e]	-0.0494 (-0.50)	-0.0198 (-0.22)	0.0863 (0.82)	0.0792 (0.95)
Education, C (1=post secondary school and post graduate education) [e]	-0.1825 (-1.64)	0.1115 (1.16)	-0.0749 (-0.59)	0.4053*** (3.94)
Relative averaged income for twin when twin was 17-25 years old [c]	-0.6670*** (-4.33)	0.2838** (2.26)	-0.1024 (-0.70)	0.3141** (2.35)
Constant	-0.7551*** (-4.97)	-0.0029 (-0.02)	-1.3998*** (-8.53)	-0.5163*** (-3.49)
ρ	0.3155*** (21.5337) [b]	.4433*** (79.1846) [b]	0.1273 (2.1217) [b]	0.4325 *** (67.5171) [b]
Number of observations	5445	5445	4686	4686
Log pseudo-likelihood	-4095.77	-6922.09	-2839.13	-5863.30

Notes: Estimated coefficients are found in the first line for each variable and second row includes *t*-ratios. The standard errors are corrected for repeated observations from the same twins over the years. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. a) Parameters for variables are constrained to be the same for both sets of twins, i.e. $\beta_1 = \beta_2$. b) Numbers in parenthesis are χ^2(d.f.=1). c) The incomes are relative to the average income among 30-60 years old in municipality. The average refers to up to three observations during the period. d) The reference case for the twins' age is 36-40 years old. e) The reference case for education is compulsory school.

Table 2 - Measures for ρ and Aggregate Background Effects on marital status

| Measure | Female samples | | | | Male samples | | | |
| | Monozygotic | | Dizygotic | | Monozygotic | | Dizygotic | |
	Twin 1	Twin 2	Twin 1	Twin 2	Twin 1	Twin 2	Twin 1	Twin 2
Divorce								
Aggregate Background Effect	0.1792	0.1769	0.0563	0.0490	0.0535	0.0473	0.0325	0.0356
Marginal probability, $P(D_s = 1)$	0.1344	0.1355	0.1289	0.1278	0.0917	0.0915	0.1020	0.1018
ρ	0.3155*** (0.0634)		0.0852 (0.0628)		0.1273 (0.0865)		0.0746 (0.0672)	
Marriage								
Aggregate Background Effect	0.3001	0.2940	0.1187	0.1174	0.3013	0.3015	0.1663	0.1762
Marginal probability, $P(M_s = 1)$	0.6206	0.6194	0.5979	0.5987	0.5892	0.5916	0.5625	0.5627
ρ	0.4433*** (0.0430)		0.1760*** (0.0421)		0.4325*** (0.0458)		0.2457*** (0.0397)	
Spouse								
Aggregate Background Effect	0.2750	0.2708	0.0403	0.0377	0.2361	0.2377	0.1231	0.1294
Marginal probability, $P(S_s = 1)$	0.6859	0.6852	0.6866	0.6877	0.6694	0.6715	0.6485	0.6487
ρ	0.4176*** (0.0443)		0.0563 (0.0447)		0.3525*** (0.0481)		0.1953*** (0.0405)	

Note: Regression results for the dizygotic samples, and when spouse is used as dependent variable, are available from the author on request. The measures for ρ are marked * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$. Standard errors are included in parenthesis.