Education, Marriage and Career: How to Account for It All?

Anna Nesterenko

September 27, 2005

1 Introduction

Issues of labor market participation and educational attainment of women along with their marriage and fertility decisions have been given a lot of attention in economics. In terms of labor force participation, considerable amount of work has been done with the emphasis on (married) female labor supply since recent trends suggest that more women choose to participate at the labor market. The evidence suggests, as well, that more and more females enroll and graduate from colleges with higher and/or more versatile degrees as compared to several decades ago. These changes in the formal education and new opportunities in the labor market are inevitably linked with the marriage and fertility decisions. Changing perceptions about the roles of the spouses, lower fertility rates and higher divorce rates are the most distinct ones, and seem to be associated with seemingly improved labor market outcomes and increased investment in human capital among women. Becker (1985) provides some brief explanation for these changes during the 1960-90-s, while Goldin (2004) discusses dynamics and patterns of female labor market participation for most of the 20^{th} century. The role of women in the society has changed dramatically over the past century, and so had their goals and priorities in terms of family and career outcomes. Choosing between family and career, the most recent cohort of women seems to be pursuing *both* at the same time, which makes it very distinct from the previous generations of women who prioritized one *or* the other. In this paper, I will try to analyze in the same framework the choices women make and the outcomes in terms of their education, family and career. The paper is organized as follows. Section 2 gives a brief review of the existing literature in the field. Section 3 presents the theoretical model of female labor market participation focusing on women's earnings and accounting for their marital status and endogenous choice of number of years of education. The data used in the research are described in Section 4. Section 5 turns to an econometric analysis of the model and presents the main findings. Section **??** concludes, suggesting some policy implications arising from the analysis and possible directions for further research.

2 Brief review of the literature

Before formulating a formal model to study female labor supply and investment in human capital decisions and testing it using the recent data, I will briefly discuss relevant models already developed in the field. The interest in the topic developed in the 1960-s when female labor participation rates started increasing, and more and more women were entering the labor force. Not only their participation, but also its intensity as measured by the amount of working hours supplied to the market, have become important research issues. At the same time, it has been shown that working women represent a rather selected sample: those who have decided to work for pay must consider themselves more productive in the market activities as compared to non-market ones (leisure, household work, child rearing, etc.), though they are not necessarily *the most* productive women among those with otherwise similar characteristics. Female labor market participants must have some comparative advantage in the labor market, and thus expect higher returns to their market time as opposed to non-market. Higher market returns, i.e. higher wages, require bigger investments in human capital, which makes the choice of education endogenous to the labor market participation decisions. While this setup seems to be rather involved already, marriage, fertility, child bearing and child rearing decisions add to its complexity even more.

In his "Treatise on the Family" Becker talks about sexual differences in the division of labor that exist due to both, gains from specialized investment and intrinsic differences between men and women. Investment differences, such as amount of schooling and degrees pursued, reinforce biological differences among the sexes, such as women's biological ability to give birth and their commitment to the care of children, which makes it more difficult to disentangle biological comparative advantage from specialized investments. To a great extent, all these differences reveal themselves through market wages. There is some considerable evidence that wage rates for women are lower than those for men in similar professions and overall in the economy. Becker explains it at least partially by the fact that women invest less in market human capital, and more in household capital since they are more productive in household activities. However, since recent trends are changing, there relationships require much more careful analysis on the recent data.

While female wage rate and their labor market earnings appear to be lower than those of men, there is some empirical evidence that the rate of return on human capital investment, particularly, rate of return to schooling, appears to be higher for women. In his work, Dougherty (2003) summarizes the results of some studies performed on different data sets to conclude that majority of them find higher schooling coefficient for females, and then discusses some of the possible explanations of this effect. Sample selection bias can be responsible for part of the difference in the coefficients. To account for sample selection, a two-stage model with participation and wage equations is used, following an econometric framework developed by Gronau (1974) and Heckman (1979). In this model, labor market earnings are observed only for those females who have decided to participate once their market wage is higher than their reservation wages, and based on the perception of their higher comparative productivity in the labor market and some other unobserved characteristics which are also conjectured to be positively correlated with the factors affecting wages and earnings. Theoretically, if the error terms in two equations are positively correlated suggesting a possible upward bias in the wage equation, then allowing for sample selection by introducing this two-stage model vs. a single wage equation should reduce the rate of return coefficient. There is an empirical evidence, however, that allowing for sample selection reduces, increases or has no effect on schooling coefficients (Dougherty (2003)).

Along with sample selection bias in the returns to education model, one has to allow for endogeneity of education: being determined to pursue labor market career instead of focusing solely on traditional household activities, women are likely to choose more education, which is then likely to bring higher returns. Assuming rational, forward-looking utility-maximizing behavior, career-oriented women will continue their education beyond high school or four-year college degree, thus making a bigger investment in human capital with expectations of higher future returns. To control for endogeneity, women's educational choices should be modeled in a separate equation. Furthermore, as we are still dealing with a rather selected sample of career-oriented women, this schooling equation should be included along with participation and wage equations to account for all these effects simultaneously.

3 Economic model

The model presented below follows Heckman (1979) framework that accounts for sample selection among working women, and is extended to include endogenous educational choices.

My primary interest at this point is in studying the effects of various factors on female labor income or amount of working hours supplied to the labor market. Focusing on the labor income as a dependent variable, I can specify the *main* equation as follows:

$$Y = \beta X + \varepsilon \tag{1}$$

The vector of independent variables X that determine market earnings includes individual labor market characteristics (education and experience, occupation and industry, etc.), and demographic factors. However, labor market income is observed only for working women, and they are assumed to participate if the market wage they can be offered exceeds their reservation wage, i.e. if $W_m - W_r > 0$, which defines *participation* equation in this case. Market wage and reservation wage equations can be formulated as:

$$W_m = \gamma_m X_m + u_m \tag{2}$$

$$W_r = \gamma_r X_r + u_r \tag{3}$$

respectively. Here, $E(u_i) = 0, i = m, r$, and $E(u_m u_r) = \sigma$. Then participation, or sample selection, criteria would be:

$$\begin{aligned} W_m - W_r &> 0 \\ \gamma_m X_m + u_m - (\gamma_r X_r + u_r) &> 0 \\ u_m - u_r = u_p &> -(\gamma_m X_m - \gamma_r X_r) \end{aligned} \tag{4}$$

and among the factors influencing labor market participation decision of women, their family characteristics would be very important along with their individual ones.

With this sample selection rule in mind, the main regression function for this working subsample can be formulated as:

$$\begin{aligned} E(Y|X, sample\ selection\ rule) &= \beta X + E(\varepsilon|sample\ selection\ rule) \\ E(Y|X, W_m - W_r > 0) &= \beta X + E(\varepsilon|u_p > -(\gamma_m X_m - \gamma_r X_r)) \\ E(Y|X, participation) &= \beta X + \frac{\sigma_{\varepsilon p}}{\sigma_p}\lambda(\alpha_p) \end{aligned} \tag{5}$$

where $\sigma_{\varepsilon p}$ is correlation between the error terms in the *main* and *participation* equations, and σ_p is the standard deviation of the participation equation error term, u_p; $\lambda(\alpha_p) = \frac{\phi(\alpha_p)}{1-\Phi(\alpha_p)}$ is the inverse Mill's ratio or Heckman's λ, with $\alpha_p = -\frac{\gamma_m X_m - \gamma_r X_r}{\sigma_p}$.

When main equation is formulated with labor market earnings as a dependent variable, it is sensible to include, along with individual characteristics, education and experience as regressors, and the coefficient on schooling will produce an estimate for the rate of return on education.

However, completed years of education can not be considered exogenous, since women are likely to choose how many years to study, especially in terms of college education. Hence, endogenous education choice S will be modeled separately in the equation (6) based on the individual characteristics Z_s, family background B, parents' education P, etc.:

$$S = \alpha_s + \gamma_s Z_s + \rho B + \pi P + u_s \tag{6}$$

The set of equations estimated together should capture at least two important factors acting at the same time and affecting labor market outcomes for women: self-selection into the labor force and endogenous choice of education with the perspective of future employment and given particular individual and family characteristics. To the best of my knowledge, there is no single procedure to account for both sample selection and endogeneity of the regressors. However, the estimation can be still performed in several stages using well-known results at each particular stage. From a separate regression for education, I can obtain estimated values for years of schooling. From the participation probit equation, I can obtain Heckman's λ. Finally, I can use these (instrumented) estimates for education, as well as estimated Heckman's λ in the main equation to obtain more accurate estimates. There should be noted, however, that in such joint estimation some correction of standard errors should be performed. However, in the present work I will be relying only on the standard error correction procedure embedded into the estimation of Heckman's λ.

In the estimation of the IV corrected sample selection model one more issue has to be addressed. I conjecture that years of schooling are endogenous in the main equation, and thus have to be instrumented, so that the predicted values for education could be used in the estimation of the main equation. However, since education can be observed for all the women in the sample, the predicted education values would be obtained without differentiating between working and non-working women. This distinction, however, appears to be important for the next estimation step, where educational choices are conjectured to be associated with the labor market status, and thus different for working and non-working women. In other words, estimated coefficients for education are obtained based on the sample of both working and non-working women, and later they are used in the estimation for the sub-sample of working women only, since in this step sample selection is explicitly accounted for. One of the consequences of such estimation procedure is incorrectly estimated standard errors, which casts some doubt on the efficiency of the obtained estimates in the main equation.

One way to circumvent this problem might be to perform instrumental variables estimation that can produce standard errors which are consistent in the presence of some violations of the assumption of the i.i.d. errors. For this purpose, the following model will be used:

$$
\begin{aligned}
Y &= \beta_1 X + \gamma S + \varepsilon \\
S &= \beta_2 X + \phi Z + \xi
\end{aligned}
\tag{7}
$$

where Y is a dependent variable (yearly income or total working hours), X is a vector of individual characteristics, S is education variable (years of schooling), and Z is a vector of excluded instruments. Following the assumptions of the IV model, the error terms ε and ξ should be uncorrelated, and excluded instruments Z should be exogenous to the main equation. As a result, IV model will produce consistent estimates of the β_1 and γ coefficients. However, this two-step estimation procedure will not account for possible sample selection among women, since "main" equation will be estimated on the pooled sample of women, thus assuming that individual characteristics have the same effect on the dependent variable among working and non-working women. However, IV estimation would allow performing the tests of the validity of the instruments used for the IV corrected sample selection model.

4 Data

The data used in this paper comes from the 2001 wave of the Panel Study of Income Dynamics (PSID) — a longitudinal study of a representative sample of U.S. individuals and the family units in which they reside. While some information is collected about all individuals in the family unit, the greatest level of detail is ascertained for the primary adults heading the family unit. According to the PSID guidelines[1], for the time of interview each family unit has one and only one current Head. Originally, if the family contained a husband-wife pair, the husband was arbitrarily designated the Head to conform with Census Bureau definitions in effect at the time the study began. The Head of the household must be at least 16 years old and the person with the most financial responsibility for the family. If this person is female and she has a husband (or boyfriend with whom she has been living for at least one year) in the household, then he is designated as Head. However, if the husband or boyfriend is incapacitated and unable to fulfill the functions of Head, then the family unit will have a female Head. Merging individual and family data, then focusing only on the working age[2] females who are either

[1] Available from http://psidonline.isr.umich.edu/Guide/Overview.html

[2] The age of the women in this sample ranges from 24 to 61, with the mean age of 41.

Table 1: Marital status and family role

Marital status	Relation to head			
	Head	Legal wife	Female cohabitor	Total
Married	23	2,926	2	2,951
Never married	580	0	116	696
Widowed	87	0	4	91
Divorced	519	0	63	582
Separated	181	0	7	188
Total	1,390	2,926	192	4,508

heads or wives in the families, and having deleted invalid observations[3], I obtained 4,508 observations. Table 1 gives cross-tabulated results in terms of marital status and family role of the women in the sample. Less than one third of women were reported as the heads of the households, and about two thirds of all the women in this sample were married at the time of interview.

Another way to look at the sample at hand is to compare the levels of educational attainment across different marital categories. As it can be seen from the Table 2, about one third of the women in this sample have completed high school only, about 27% have done some college work, almost 15% received college degree, and only 8.5% were enrolled in some graduate and post-graduate studies. With some slight discrepancies, about 30 percent of women in each marital status category have high school degrees, though somewhat higher percentage of currently married and never married women have received their college degrees, as opposed to those widowed, divorced or separated. Quite noticeably, there is the highest percentage of graduate and post-graduate students among the married women.

To get some preliminary idea about labor force participation, I have constructed the following variables based on the amount of working hours per week in 2000. I have defined women to be out of labor force if, on average, they worked zero hours on their main job(s), working less than part-time if they reported some positive number of hours but less than 20, working part-time if they worked between 20 but less than 40 hours a week, and working full time if they, on average, worked 40 or more hours per week. With such a notation (Table 3), about 17% of the sample appear to be out of labor force, and almost 57% have reported

[3]Dealing with the survey data, I deleted observations with "refused to answer" and "don't know" replies to the questions about the years of education, as well as kept only those respondents who reside in the US and live with their respective family units.

Table 2: Distribution of educational attainments and marital status[a]

	Marital status					
Education level	Married	Never married	Widowed	Divorced	Separated	Total
Less than high school	406	183	33	115	58	795
	(13.76)	(26.29)	(36.26)	(19.76)	(30.85)	(17.64)
High school only	943	208	32	191	53	1,427
	(31.96)	(29.89)	(35.16)	(32.82)	(28.19)	(31.65)
Some college	804	164	21	184	56	1,229
	(27.25)	(23.56)	(23.08)	(31.62)	(29.79)	(27.26)
College degree	511	96	3	55	9	674
	(17.32)	(13.79)	(3.30)	(9.45)	(4.79)	(14.95)
Some post-grad work	287	45	2	37	12	383
	(9.73)	(6.47)	(2.20)	(6.36)	(6.38)	(8.50)
Total	2,951	696	91	582	188	4,508
	(100)	(100)	(100)	(100)	(100)	(100)

[a] Total number of observations in each category and (column percentage in parentheses).

working full-time. As compared among different education levels, there are higher rates of full-time labor market participation among more educated women, while the shares of those out of labor force drop in each higher educational category. My distinction between "less than part-time" and "part-time" work was somewhat arbitrary, and there is no clear pattern between different educational categories.

5 Estimation and discussion of the results

In Section 3 I presented two alternative estimation approaches: instrumental variables model and IV corrected sample selection model. As a starting point, I will discuss the results of the IV model first, since some of them might be useful for the subsequent estimations.

Instrumental variables model is estimated following equation (7). First-stage regression of years of education is performed by OLS, where dependent variable is the number of years of school completed, and excluded instruments are the dummy variables to report whether each of the women's parents have completed high school or college. The instrumental variables regression of log of labor income uses predicted values for the years of education from the

Table 3: Distribution of educational attainments and labor force participation[a]

	Labor force status (by total yearly hours reported)					
Education level	Out of LF	0 – 300	300 – 1600	1600 – 2000	2000 and more	Total
Less than high school	236	33	199	140	187	795
	(29.69)	(4.15)	(25.03)	(17.61)	(23.52)	(100)
High school only	243	31	289	350	514	1,427
	(17.03)	(2.17)	(20.25)	(24.53)	(36.02)	(100)
Some college	158	26	293	332	420	1,229
	(12.86)	(2.12)	(23.84)	(27.01)	(34.17)	(100)
College degree	81	16	149	182	246	674
	(12.02)	(2.37)	(22.11)	(27.00)	(36.50)	(100)
Some post-grad work	36	12	95	108	132	383
	(9.40)	(3.13)	(24.80)	(28.20)	(34.46)	(100)
Total	754	118	1,025	1,112	1,499	4,508
	(16.73)	(2.62)	(22.74)	(24.67)	(33.25)	(100)

[a] Total number of observations in each category and (row percentage in parentheses).

first step along with the rest of the regressors. Estimation results for both stages are presented in the Table 4.

Several results from the IV estimation are of particular interest. First of all, dummy variables for parents' education seem to perform well as excluded instruments. They are statistically significant in the first-stage OLS regression, and have expected signs. To further justify the validity of these instruments, below reported are the results of the relevant tests[4]. Anderson canonical correlations likelihood-ratio statistic is used to test whether the equation is identified, i.e., the excluded instruments are relevant. Under the null hypothesis, the model is underidentified, meaning that the matrix of reduced form coefficients has $rank = K - 1$, with K – number of of regressors (here, 12). Relevant test statistic is distributed as χ^2 with $(L - K + 1)$ degrees of freedom, with L – number of included and excluded instruments (here, 15). For this model specification, calculated Anderson LR statistic is 459.99, which far exceeds critical value of $\chi^2_{(4)}$= 9.49 at 5% significance level. Hence, the null hypothesis is rejected, meaning that the model is identified and the instruments used are relevant to the model. However, this result should be taken with some caution, since weak instruments prob-

[4]Tests discussed below are performed within *ivreg*2 estimation command in Stata.

Table 4: Instrumental variables estimation results

Variable	Coefficient	(Std. Err.)
First-stage regression of years of education		
Working weeks	0.0189**	(0.0018)
Age	0.0509†	(0.0308)
Age^2	-0.0007*	(0.0003)
Black	-0.3692**	(0.0857)
All other race	-1.8867**	(0.1298)
Married	0.5302**	(0.1732)
Not married	-0.1138	(0.1223)
Respondent flag	-0.2182**	(0.0831)
Kids	-0.1851**	(0.0315)
Female head	0.2967†	(0.1728)
Father: high school degree	0.4264**	(0.0855)
Father: college degree	1.7802**	(0.1164)
Mother: high school degree	0.3313**	(0.0809)
Mother: college degree	1.0411**	(0.1308)
Intercept	11.2681**	(0.6092)
IV regression		
Education	0.2129**	(0.0369)
Working weeks	0.1725**	(0.0017)
Age	0.0055	(0.0251)
Age^2	-0.0001	(0.0003)
Black	0.1545*	(0.0738)
All other race	0.3586**	(0.1357)
Married	-0.2515†	(0.1427)
Not married	-0.1425	(0.0994)
Respondent flag	0.1674*	(0.0680)
Kids	0.0092	(0.0266)
Female head	-0.1349	(0.1411)
Intercept	-1.0287	(0.6666)
N	4508	
R^2	0.7704	
$F_{(11,4496)}$	1378.2356	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

lem might still be present. Another important test is the one of overidentifying restrictions (Hansen–Sargan test). Under the null hypothesis, the instruments are uncorrelated with the error term, and excluded instruments are truly exogenous to the main equation. Calculated Sargan test statistic is 3.54, and under the null it is distributed as χ^2 in the number of overidentifying restriction (here, 3). Since calculated test statistic does not exceed the critical value of $\chi^2_{(3)}$= 7.81 at 5% significance level, the null hypothesis can not be rejected, meaning that the excluded instruments are valid and correctly excluded from the main estimated equation.

Secondly, estimated parameters from the IV regression can be commented on since they might serve as a reference point for subsequent estimations. Estimated coefficient on years of education is statistically significant, and it appears that every additional year of schooling is likely to bring 21% increase in the total yearly labor income.

However, IV estimation on the whole sample of women assumes that each additional year of education is likely to bring the same increase in yearly labor income for working women, no matter how many hours they have been putting in, as well as for non-working women. Such assumption ignores sample selection among women, and thus causes biased parameter estimates. To avoid this problem, sample selection should be accounted for. Keeping the issue of endogeneity in mind, years of education should be instrumented with the parents' education variables that proved to be reasonable instruments in the above estimation.

The first step in the IV corrected sample selection model would be to obtain estimated coefficients for education from the OLS regression following equation (6). The main regressors of interest in this auxiliary regression (see Table 5) are parents' education, which turned out to be statistically significant and with expected signs. Thus, once one or both parents have completed high school or college, it has a positive effect on the number of years of schooling their daughters have undertaken. Including parental education, women's age, marital status and race accounts for family background and individual characteristics that are likely to affect the choice of education. In this way, predicted values for the education variable can be used in further estimation thus somewhat circumventing the endogeneity problem.

The next step is estimating Heckman selection model according to the following basic framework. In the main equation (1), the dependent variable is a natural logarithm of 2000 yearly "labor income", which, according to the PSID survey structure, includes not only wages and salaries, but also bonuses, overtime, tips, commissions, and miscellaneous labor income. This labor income variable can be observed only for the working women, and hence the selection equation (4) is specified with dummy dependent variable equal to 1 for those women who reported having worked some particular amount of hours per year. To ensure some robustness of the model to the choice of different participation criteria, three specifications were adopted.

Table 5: OLS estimation results

Variable	Coefficient	(Std. Err.)
Dependent variable: Years of education		
Age	0.0315	(0.0307)
Age squared	-0.0004	(0.0004)
Married[a]	0.2541*	(0.1070)
Not married[a]	-0.1128	(0.1238)
Black[b]	-0.4776**	(0.0852)
All other race[b]	-2.0826**	(0.1309)
Father: high school diploma	0.4611**	(0.0870)
Father: college degree	1.8066**	(0.1185)
Mother: high school diploma	0.3845**	(0.0823)
Mother: college degree	1.1410**	(0.1330)
Intercept	12.0846**	(0.6064)
N	4508	
R^2	0.1987	
$F_{(10,4497)}$	111.4978	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

[a] The omitted base category is ”Single, never married”, while ”Not married” includes widowed, divorced and separated women.

[b] The omitted base category is ”White”, while ”All other race” includes Hispanics, American Indians, Asian and other races not mentioned above.

Specification 1 differentiates between the women who were out of labor force or worked very few hours, and those who worked more than 300 hours (about 15 weeks or more of part-time work). Specification 2 uses 1600 hours and above per year as a participation criteria. Finally, Specification 3 splits the sample into full-time working women and the rest, including part-time workers and those out of labor force. These different participation criteria are likely to involve different participation incentives, as well as exert different effects on the changes in labor income. In all of the specifications, women's decision to participate at the labor market is conjectured to depend upon their age, race, marital status, number of kids, and whether or not they are considered to be heads of the households. Independent variables in the main equation include number of weeks worked in that year, age and age squared, number of years of schooling estimated in the previous step, as well as respondent flag.

Estimation results for Heckman selection model are presented in the Tables 6–8, where estimated coefficients from three specifications are listed respectively.

Among the three model specifications, those that consider part-time workers as labor market participants exhibit a lot of similarities, as opposed to the one that focuses mainly on the full-time working women. Two variables that enter participation equation and do not belong into the main income equation are the number of children under the age or 18 and dummy variable for the women who are the heads of the households. I assume that these two factors would affect women's decision to enter the labor force, but they should not have any effect on the yearly labor income. Throughout all model specifications, having more children under the age of 18 make it less likely for the women to start working, since their household time is more valuable. Female head dummy variable is statistically significant in the model specifications for the part-time workers. Being a head of a household increases the probability to work at least part-time, since these women can be conjectured to have more financial responsibilities and they have to provide for the rest of their family members. However, being a head of the household is not statistically significant in the model for full-time working women, and estimated coefficient even has a negative sign. One of the possible explanations might be that the amount of responsibilities these women have does not allow them to work full-time since there would be nobody to take care of the family needs when they are away at work.

Turning on to the main equation for (log of) labor income, some distinctions should be made according to different model specifications. Interesting enough, the number of weeks worked during the year have a different impact on the changes on labor income according to women's labor market status. The sub-sample of women who work above 300 hours a year includes a lot of part-time workers, and for them the more weeks they work, the higher yearly income they get. However, being a full-time worker does not suggest that more weeks would

Table 6: Estimation results : IV corrected sample selection

Variable	Coefficient	(Std. Err.)
(Log of) Income Equation		
Work weeks	0.0341**	(0.0024)
Education	0.1670**	(0.0271)
Age	0.0591**	(0.0205)
Age squared	-0.0005*	(0.0002)
Black	0.1563*	(0.0660)
All other race	0.3752**	(0.1295)
Married	0.0909	(0.0825)
Not married	-0.1682†	(0.0916)
Respondent flag	-0.0693	(0.0514)
Intercept	5.1154**	(0.5545)
Participation equation (300+ hours)		
Age	-0.0209**	(0.0026)
Black	0.0178	(0.0538)
All other race	-0.2663**	(0.0741)
Married	0.0940	(0.1108)
Not married	0.1128	(0.0827)
Kids	-0.1674**	(0.0187)
Female head	0.2591*	(0.1086)
Intercept	1.8039**	(0.1482)
Mill's λ	-1.6529**	(0.3226)
N	4508	
Censored	872	
Uncensored	3636	
$\chi^2_{(14)}$	386.886	

Significance levels : † : 10% $*$: 5% $**$: 1%

Table 7: Estimation results : IV corrected sample selection

Variable	**Coefficient**	**(Std. Err.)**
(Log of) Income Equation		
Work weeks	0.0006	(0.0061)
Education	0.1413^{**}	(0.0275)
Age	0.0577^{**}	(0.0208)
Age squared	-0.0006^{*}	(0.0003)
Black	-0.0053	(0.0576)
All other race	0.0922	(0.1173)
Married	0.0086	(0.0682)
Not married	-0.1633^{*}	(0.0769)
Respondent flag	-0.0484	(0.0523)
Intercept	7.3270^{**}	(0.6353)
Participation equation (1600+ hours)		
Age	-0.0167^{**}	(0.0023)
Black	0.2026^{**}	(0.0467)
All other race	-0.1778^{*}	(0.0698)
Married	0.1265	(0.0974)
Not married	0.2152^{**}	(0.0695)
Kids	-0.2322^{**}	(0.0173)
Female head	$0.1699^{\dagger}$	(0.0950)
Intercept	0.9377^{**}	(0.1266)
Mill's λ	-0.5064^{**}	(0.1578)
N	4508	
Censored	1897	
Uncensored	2611	
$\chi^2_{(14)}$	160.071	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Table 8: Estimation results : IV corrected sample selection

Variable	Coefficient	(Std. Err.)
(Log of) Income Equation		
Work weeks	-0.0224*	(0.0112)
Education	0.1461**	(0.0394)
Age	0.0844**	(0.0309)
Age squared	-0.0009*	(0.0004)
Black	0.0501	(0.0815)
All other race	0.0075	(0.1616)
Married	-0.0129	(0.0975)
Not married	-0.3317**	(0.1143)
Respondent flag	-0.0385	(0.0752)
Intercept	8.1510**	(0.9992)
Participation equation (2000+ hours)		
Age	-0.0172**	(0.0023)
Black	0.1261**	(0.0468)
All other race	-0.0424	(0.0734)
Married	-0.0820	(0.0975)
Not married	0.2919**	(0.0691)
Kids	-0.1809**	(0.0181)
Female head	-0.1096	(0.0949)
Intercept	0.4633**	(0.1258)
Mill's λ	-0.5057*	(0.2411)
N	4508	
Censored	3009	
Uncensored	1499	
$\chi^2_{(14)}$	135.5788	

Significance levels : † : 10% $*$: 5% $**$: 1%

translate into higher yearly income. This finding might be consistent with an explanation that women who have joined the labor market are not necessarily the most productive women in that particular occupation, hence their longer working hours would not add to the amount they are paid.

Another variable of interest from the main equation is education. In all of the specifications for the IV corrected sample selection model, the estimates of the rate of return to schooling range from 14 to 16 percent. These estimated coefficients are smaller than the one from the instrumental variables models, a finding consistent with my previous conjecture that IV estimates might be upward biased if participation decisions (or sample selection into the labor market) are not accounted for.

These findings seem to be consistent with the earlier results in the field, and yet they show another way to tie together educational choices, family status and labor market participation decisions of women based on a rather recent data set.

References

Gary S. Becker. *A Treatise on the Family*. Harvard University Press, Cambridge, Massachusetts; London, England, enlarged edition.

Gary S. Becker. Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 3(1):S33–S58, January 1985.

Christopher Dougherty. Why is the rate of return to schooling higher for women that for men? Center for Economic Performance, London School of Economics and Political Science, August 2003.

Claudia Goldin. The lond road to the fast track: Career and family. NBER working paper 10331, available at http://www.nber.org/papers/w10331, March 2004.

Reuben Gronau. Wage comparisons – A selectivity bias. *The Journal of Political Economy*, 82(6):1119–1143, 1974.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–162, January 1979.