DISCUSSION PAPER SERIES

IZA DP No. 3245

Fertility Differences between Married and Cohabiting Couples: A Switching Regression Analysis

Junfu Zhang
Xue Song

December 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Fertility Differences between Married and Cohabiting Couples: A Switching Regression Analysis

Junfu Zhang
Clark University
and IZA

Xue Song
Thomson Healthcare

Discussion Paper No. 3245
December 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3245
December 2007

ABSTRACT

Fertility Differences between Married and Cohabiting Couples: A Switching Regression Analysis*

Little is known about why cohabiting couples have fewer children than married couples. We explore the factors that explain the difference in fertility between these two groups using a switching regression analysis, which enables us to quantify the contribution of different factors through a decomposition of the difference. We find that married couples have more children than cohabiting couples primarily because marriage provides stronger incentives for specialization in household production. Unobserved self-selection plays a less important role.

JEL Classification: J12, J13

Keywords: fertility, marriage, cohabitation, switching regression, self-selection, household specialization

Corresponding author:

Junfu Zhang
Department of Economics
Clark University
950 Main Street
Worcester, MA 01610
USA
E-mail: juzhang@clarku.edu

* We would like to thank Robert Moffitt, Bruce Hamilton, Yingyao Hu, David Neumark, Eric Slade, and seminar participants at the Public Policy Institute of California for their comments and suggestions.

1. Introduction

The percentage of men and women in the U.S. who cohabit has risen dramatically. According to Bumpass and Sweet (1989), about 35% of those born in the early 1960s lived with someone of the opposite sex before age 25, compared to less than 8% of those born in the early 1940s. Roughly one-half of first marriages in the 1980s were preceded by cohabitation (Bumpass, Sweet, and Cherlin, 1991). Data from the National Survey of Families and Households (NSFH) show that in the late 1980s (1987-1988), about 7% of those 25-29 years old were cohabiting; by the early 1990s (1992-1994), this figure had risen to 13% (Waite, 1995). The 2000 Census revealed a total of 4.9 million unmarried opposite-sex-partner households, accounting for 8.1% of total coupled households in the nation (Simmons and O'Connell, 2003). Clearly, cohabitation is becoming an increasingly common household union.

Cohabitation is a very different type of union from marriage. According to sociological studies, cohabitation has a higher dissolution rate and a shorter duration. Forty percent of all cohabiting couples either marry or stop living together within a year and only one third still live together after two years (Bumpass and Sweet 1989; Thorton, 1988). Spanier (1983) finds that cohabiting men are less likely to be employed than married men, but cohabiting women are more likely to be employed than married women. Rindfuss and VandenHeuvel (1990) point out that married couples tend to have higher incomes than cohabiting couples.

Other studies have shown that cohabitation is a less integrated union than marriage. Cohabiting couples are less likely to pool their financial resources and more likely to assume that each partner is responsible for supporting himself or herself financially (Blumstein and Schwartz, 1983; Heimdal and Houseknecht, 2003; Winkler, 1997). Cohabitors are more likely to engage in infidelity than married people, even after controlling for personal values regarding extramarital sex

(Treas and Giesen, 2000). While cohabiting men and married man spend similar time on housework, cohabiting women spend significantly less time on housework than married women (South and Spitze, 1994).

Along with the increase in cohabitation rates, childbearing behavior has also changed dramatically in the United States. Before 1960, marriage was a virtual precondition for childbearing in the United States; in 1970, nearly 90% of all births occurred to married couples. Since then, out-of-wedlock births have grown steadily, with 33.5% of births occurring outside of marriage by 2001 (National Center for Health Statistics, 2003). Births to cohabiting couples constitute a large percentage of nonmarital births. Among births to unmarried women under age 40, those occurring to cohabiting parents increased from 29% in the 1980-1984 period to 39% during 1990-1994, accounting for almost all of the increase in nonmarital childbearing between 1980-1984 and 1990-1994 (Bumpass and Lu, 2000).

In response to this trend, researchers across the social sciences have paid increasing attention to childbearing within cohabitation and learned a great deal about fertility behavior of cohabiting couples (see, e.g., Manning, 1993, 1995; Musick, 2002; and Raley, 2001). It is now well established that cohabiting couples have fewer children and are more likely to remain childless than married couples (Bachrach, 1987; Rindfuss and VandenHeuvel, 1990). However, investigations into the cause of this difference are scarce.

Conceptually, fertility discrepancies are attributable to three sources: 1) Distinct behaviors in marriage and cohabitation. Marriage and cohabitation are inherently distinct unions and therefore, married and cohabiting couples could behave differently even if they have identical characteristics. 2) Observable differences in the characteristics of married and cohabiting couples. For example, married couples tend to be older, and older age is usually associated with a higher number of children. 3) Unobservable self-selection. For example, people may choose marriage simply because they want to have children and regard marriage as the proper union for doing so.

Sorting out the relative importance of these factors helps us understand the implications of the rising trend of cohabitation. For example, if the fertility difference stems solely from self-selection, then the trend of cohabitation simply reflects the fact that more people are choosing to have fewer children, suggesting that cohabitation itself is not a reason for the decline of fertility. Conversely, if people with similar characteristics behave differently under marriage and cohabitation, then the rising trend of cohabitation may have a profound impact on population growth through its influence on fertility rates.

This paper considers the differences in fertility behavior between married and cohabiting couples. Using switching regression models, we identify and quantify potential sources that contribute to the fertility difference. We are fully aware that the static switching regression will not capture all the inherent dynamics of marital choice and that a dynamic approach (such as the one taken by Lillard, Brien, and Waite, 1995) has its merit. However, our approach does have the advantage of allowing for the decomposition of the fertility gap into different factors. Quantifying the relative importance of each factor in explaining the fertility difference between married and cohabiting couples helps identify issues for further investigation along this line. Thus we believe our analysis is a good starting point for future work in this area.

The remainder of the paper is organized as follows. Section 2 reviews count data switching regression models and the estimation methods. Section 3 describes the data used for the empirical work. Section 4 presents the empirical results. The final section concludes and considers directions for future research.

2. Switching Regression Models

We will measure fertility by the number of children, and estimate count data models that deal with discrete non-negative dependent variables. While both the basic count data models (such as the Poisson model and the negative binomial model) and the linear endogenous switching models have become standard statistical tools, count data models with endogenous switching represent a fairly recent development of estimation techniques. Therefore, in this section, we shall specify our econometric model, and discuss our estimation and identification strategies in detail.

2.1 Model specification

We consider count data models in a switching regression framework as follows:

$$E[y_i \mid x_i, M_i = 1] = \exp(x_i\beta_m + \varepsilon_{1i}) \tag{1}$$

$$E[y_i \mid x_i, M_i = 0] = \exp(x_i\beta_c + \varepsilon_{0i}) \tag{2}$$

$$M_i^* = z_i\gamma + \mu_i \tag{3}$$

$$M_i = \begin{cases} 1 \text{ if } M_i^* > 0 \\ 0 \text{ if } M_i^* \le 0 \end{cases} \tag{4}$$

$$(\varepsilon_{1i}, \varepsilon_{0i}, \mu_i)' \sim N_3[0, \Sigma], \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{10} & \sigma_{1\mu} \\ \sigma_{10} & \sigma_0^2 & \sigma_{0\mu} \\ \sigma_{1\mu} & \sigma_{0\mu} & 1 \end{pmatrix}.$$

Equations (1) and (2) describe the fertility behavior of married and cohabiting couples, respectively. The dependent variable *y* is the number of children a couple has. *x* is a vector of variables that determine the expected number of children. Notice that unlike standard count data models, our model here does not explicitly define the underlying distribution of variable *y*. Thus equations (1) and (2) are consistent with either the Poisson or the negative binomial model. Equations (3) and (4) describe a couple's choice between marriage and cohabitation, with *M=1* representing marriage. The vector z refers to variables that affect the couple's marital choice. ε_0, ε_1, and μ are random terms.

A general framework like (1)-(4) constitutes a switching regression model. A classic example of a switching regression model is the study of wage differences between union and non-union workers (Lee, 1978). In many cases where only the behavior of one group is observable, the model reduces to the standard Heckman-type self-selection model (Heckman, 1979).

Substantial evidence suggests that the negative relationship between fertility and cohabitation could be a result of self-selection. For example, Barber and Axinn (1998) find that young women who want more children are more likely to choose marriage. A switching model explicitly takes into account self-selection by adding an equation of selection between marriage and cohabitation. If the sample selection equation leaves important unobserved variables unaccounted for, we will have $Cov(\varepsilon_1, \mu) \neq 0$ and/or $Cov(\varepsilon_0, \mu) \neq 0$.

A key advantage of using the switching model is that it allows the choice between marriage and cohabitation to affect fertility behavior in other ways. For example, the fertility of married and cohabiting couples may be influenced by their characteristics in different ways, which will be reflected in different parameters of equations (1) and (2). In addition, married and cohabiting couples may have very different observable characteristics, i.e., different values of explanatory variables in equations (1) and (2).

The choice of explanatory variables draws from those that are emphasized in standard fertility theories (e.g., Becker, 1960, 1981; Easterlin, 1968). According to Becker, an increase in the wage rate would raise the opportunity cost of bearing and raising children for both married and cohabiting women, with the consequence of reducing their demand for births. Family income affects a couple's fertility decisions in two ways. If children are normal goods, higher income increases the demand for children; on the other hand, a rise in income could reduce fertility if wealthier couples care more about the quality of children and the quality is substitutable with the quantity of children.

Current marital status and current fertility decision, according to Becker (1981), are affected by a woman's expected economic opportunities, such as her expected wage rate in the labor market. Household specialization theory implies that the more different the man and woman are in market and household productivity, the higher gains they obtain from the formation of a marriage. We shall use the predicted male-female wage difference to measure their comparative advantages. We expect that a couple with a higher wage difference is more likely to choose marriage, thereby facilitating household specialization.

The Pennsylvania school of fertility pays more attention to factors that shape parents' tastes and preferences toward children (Easterlin 1968). One of the longest-standing and best-documented findings in the study of fertility is the positive and strong correlation between parents' fertility behavior and children's fertility preferences and behavior (see, for example, Axinn, Clarkberg, and Thornton, 1994). To capture this effect, we include woman's and man's number of siblings in vector x. We expect that individuals growing up in larger families tend to have more children themselves.

In addition, x also includes some other variables of interest, such as race, tenure of the relationship, living in a metropolitan area, and region of residence. z includes variables that are likely to affect a couple's marital choice, such as religious preference, region of residence, race, and county unemployment rate.

2.2 Model estimation

In principle, count data models with endogenous switching could be estimated using the maximum likelihood method. However, the likelihood function of our model involves high-order integrals, which makes the maximum-likelihood estimation computationally burdensome. To avoid this problem, we develop an estimation strategy that is similar to the one introduced by Lee (1978).

Lee (1978) considered an endogenous switching model with two linear models and a probit switching equation. He proceeded by applying Heckman's two-stage method twice. At the first stage, he estimated the probit model, and used the estimated parameters to calculate the so-called Inverse Mill's Ratio for every observation. In the second stage, he ran two OLS regressions separately to estimate the parameters for the two linear models, with the Inverse Mill's Ratio included as a control variable.

One could follow the same strategy to deal with count data models with endogenous switching if a similar Heckman-type two-stage method is available for the estimation of count data models with self-selection. Some recent work by Greene (1994, 1995, and 2001) and Terza (1998) has made it possible.

A two-stage method for count data models with self-selection was first proposed by Greene (1994). He considered a Poisson model as follows:

$$\Pr(y_i \mid x_i) = e^{-\lambda_i}\lambda_i^{y_i} / y_i!, \tag{5}$$

$$\lambda_i = \exp(x_i\beta + \varepsilon_i), \tag{6}$$

$$d_i^* = z_i\gamma + u_i, \quad d_i = sign(d_i^*), \tag{7}$$

$$(\varepsilon_i, u_i) \sim N_2[(0,0), (\sigma^2, \rho\sigma, 1)],$$
$$(y_i, x_i) \text{ observed iff } d_i = 1.$$

Greene's estimation method is a direct analog to Heckman's two-step procedure. The first step is to obtain a maximum-likelihood estimate of the probit equation, and then compute

$$\hat{C}_i = \frac{\phi(z_i\gamma)}{\Phi(z_i\gamma)} \tag{8}$$

for all observations with $d_i = 1$. In the second step, estimate the parameters (β, θ) (where $\theta = \rho\sigma$) of the Poisson model in the form

$$E(y_i) = \lambda_i = \exp(x_i\beta + \theta C_i) \tag{9}$$

by maximum likelihood, substituting $\hat{C}_i$ for the unobserved C_i.

Terza (1998) argued that Greene's method is flawed. He showed that the conditional mean of the dependent variable in this model is:

$$E(y_i \mid d_i = 1) = \exp(x_i\beta^*) * \frac{\Phi(z_i\gamma + \theta)}{\Phi(z_i\gamma)}, \tag{10}$$

where β^* is the same as β apart from the intercept term, which is shifted by $\frac{\sigma^2}{2}$, and $\theta = \rho\sigma$. If $\rho = 0$ or $\sigma = 0$, $\frac{\Phi(z_i\gamma + \theta)}{\Phi(z_i\gamma)}$ becomes 1, indicating zero sample selection bias. Thus, the effect of sample selection on the conditional mean of the count equation is multiplicative, not additive as in the linear case. This suggests that an ad hoc adjustment based on adding the Inverse Mill's Ratio to the conditional mean by analogy with the linear case is inappropriate. Terza also pointed out that under the model specification, the distribution of the observed data is not Poisson. However, this unknown distribution is of minor consequence, since with the known form of the conditional mean function, one can easily apply the nonlinear least squares (NLS) method to obtain a consistent estimator. The two-step procedure would thus involve a maximum likelihood estimation of the probit model, followed by a nonlinear least squares regression of y_i on the conditional mean function presented above. We will call this estimator the "Terza estimator."

A Taylor expansion of $\ln\left(\frac{\Phi(z_i\gamma+\theta)}{\Phi(z_i\gamma)}\right)$ around the point $\theta = 0$ shows that

$$\ln\left(\frac{\Phi(z_i\gamma+\theta)}{\Phi(z_i\gamma)}\right) \approx \theta C_i, \tag{11}$$

where C_i is defined as before. Therefore, Greene's (1994) formulation can be understood as an approximation to Terza's estimator. Subsequently, Greene (1995, 2001) reconciled the two formulations and presented an alternative two-step approach, also based on nonlinear least squares. This procedure also involves two steps: First, estimate the probit by maximum likelihood and compute $\hat{C}_i$ (analogous to Heckman's Inverse Mills Ratio); second, instead of maximum likelihood estimation, use nonlinear least squares, where now the conditional mean function is given by

$$E(y_i \mid x_i) = \exp(x_i\beta^* + \theta C_i), \tag{12}$$

where $C_i = \frac{\phi(z_i\gamma)}{\Phi(z_i\gamma)}$. We will refer to this estimator as the "Greene estimator."

Both Terza's and Greene's approaches can be extended to the cases in which (y_i, x_i) is observable in both regimes (d_i = 1 or 0), as in our endogenous switching model. It can be shown that

$$E(y_i \mid d_i = 0) = \exp(x_i\beta^*) * \frac{1-\Phi(z_i\gamma+\theta)}{1-\Phi(z_i\gamma)}. \tag{13}$$

Again, applying Taylor expansion around $\theta = 0$, we obtain that

$$\ln\left(\frac{1-\Phi(z_i\gamma+\theta)}{1-\Phi(z_i\gamma)}\right) \approx \theta \frac{-\phi(z_i\gamma)}{1-\Phi(z_i\gamma)}. \tag{14}$$

So in the case of $d_i = 0$, the Greene estimator in the second step will be a nonlinear least square estimation of

$$E(y_i \mid d_i = 0) = \exp\left(x_i\beta^* + \theta \frac{-\phi(z_i\gamma)}{1-\Phi(z_i\gamma)} \right). \tag{15}$$

Therefore, we will use Terza's and Greene's two-stage methods to estimate our model. The procedure is summarized as follows:

1) Estimate the parameters of the probit selection equation $\hat{\gamma}$ by maximum likelihood.

2) Compute the Terza estimator by minimizing nonlinear least squares:

$$(\hat{\beta}_m, \hat{\theta}_m) = \arg\min \sum_{i|M_i=1} \left[y_i - \exp(x_i\beta_m) * \frac{\Phi(z_i\hat{\gamma} + \theta_m)}{\Phi(z_i\hat{\gamma})} \right]^2, \tag{16}$$

$$(\hat{\beta}_c, \hat{\theta}_c) = \arg\min \sum_{i|M_i=0} \left[y_i - \exp(x_i\beta_c) * \frac{1-\Phi(z_i\hat{\gamma} + \theta_c)}{1-\Phi(z_i\hat{\gamma})} \right]^2. \tag{17}$$

Similarly, compute the Greene estimator by minimizing nonlinear least squares:

$$(\hat{\beta}_m, \hat{\theta}_m) = \arg\min \sum_{i|M_i=1} \left[y_i - \exp\left(x_i\beta_m + \theta_m \frac{\phi(z_i\hat{\gamma})}{\Phi(z_i\hat{\gamma})} \right) \right]^2, \tag{18}$$

$$(\hat{\beta}_c, \hat{\theta}_c) = \arg\min \sum_{i|M_i=0} \left[y_i - \exp\left(x_i\beta_c + \theta_c \frac{-\phi(z_i\hat{\gamma})}{1-\Phi(z_i\hat{\gamma})} \right) \right]^2. \tag{19}$$

2.3 Model identification

In theory, our count data model with endogenous switching is identified under both estimation methods. Notice that in equations (16)-(19), $x\beta$ and $z\gamma$ always go through a non-linear transformation. Thus even if every variable in vector *z* (choice equation) also appears in vector *x* (fertility equation), all parameters are identified and can be estimated. However, this identification is solely derived from our belief of the function forms: a normal distribution of the error term in the choice equation and an exponential distribution of the fertility data. In practice, such a belief is not trusted and it is advisable to have the exclusion restriction satisfied. In our case, we need at least one variable in the marriage choice equation that is not in the fertility equation.

We use county unemployment rate as an exclusion variable. We assume that county unemployment rate affects a woman's choice on marriage but not her childbearing decision. An economy with lower unemployment rate particularly encourages women to participate in the labor force. As Ressler and Waters (1995) argued, a higher labor force participation of women implies a higher demand for the flexibility offered by cohabitation relative to marriage. Thus lower unemployment rates tend to be associated with higher cohabitation rates. While one may argue that a lower unemployment rate also raises the opportunity cost of having children, we believe that such effects will be at most second order because a mother's expected earnings in the labor market is jointly decided by both her wage rate and the employment prospect. Indeed, county unemployment rate has been tested in the fertility equation with various specifications, and found to be statistically insignificant in all cases.

In addition, we also included in the marriage choice equation two dummy variables indicating whether the women's parents completed high school. Both variables are assumed to be uncorrelated with fertility behavior and excluded from the fertility equation. Again, we tried alternative specifications and concluded that these are innocuous assumptions because our results are not sensitive to them.

3. Data

We use the data from the Panel Study of Income Dynamics (PSID) for our empirical analysis. The PSID, starting in 1968 with approximately 4800 families, is a longitudinal study of a representative sample of U.S. family units and individuals that reside in them. These families have been interviewed annually from 1968 to 1997 and biannually after 1997. PSID has collected a wide range of socioeconomic information on these families and their members.

Before 1983, female cohabitors and wives had the same code, which made it impossible to distinguish them. From then on, female cohabitors have been separately coded from legally married wives. Thus cohabitors are now distinguishable from legally married couples at the family level. However, for a long time, the data after 1993 were at an early release stage, which did not have several variables important to our study.[1] Therefore, our empirical analysis presented here is based on the PSID 1983-1993 waves.

Our sample only covers male-headed families in which the male head is married with a wife or is cohabiting with a female. Due to a lack of information on some variables, we drop female-headed families and the cohabiting couples that formed only very recently.[2] Observations with missing values on independent variables and observations with mis-coded values are excluded from the sample. Following the common practice in the fertility literature, we also drop women older than 45 years from the sample.

We use the number of children in a family to measure fertility. This variable does not include the husband's (family head's) stepchildren because PSID coded them separately. However, if the husband has had children with his ex-wife or ex-partner and they live in the household, there is no way to exclude them. In addition, PSID does not distinguish legally adopted children from biological children, although this is unlikely to create a serious bias.

[1] The 1994-1996 family level data was not released until March 2004, when we had already completed the bulk of our empirical analysis. (See http://psidonline.isr.umich.edu/data/zipCore.aspx, accessed on June 27, 2005).

[2] PSID treats the partner of a female family head and the first-year cohabitor of any family head as “other family members,” and many questions asked to a male family head and his wife or long-term cohabitor are not asked to those individuals. Thus we cannot construct all the independent variables for those individuals. Fortunately, female-headed families are extremely rare cases, constituting less than 0.35% of the total families in any year in our sample. Thus dropping those observations is unlikely to create a non-negligible bias. However, the exclusion of first-year cohabitors does bias the average relationship duration and average number of children for cohabiting couples, and thus underestimate the fertility difference between married and cohabiting couples.

Since the PSID is a panel data set, we can observe a couple for several periods. A simple pooled cross-section estimation would bias the standard errors downward because of multiple observations on the same individuals. Instead, we use the last observation on a couple in the sample.

Some marriages in the dataset (390 observations) started as cohabitation.[3] It is clearly inappropriate to include these observations in both marriage and cohabitation samples because that would cause correlations between the error terms. We choose to include these observations in the cohabitation sample but not in the marriage sample, primarily because the cohabitation sample is relatively small. For these couples, the information in their last cohabiting years is used in the regression analysis. Essentially, we are assuming that a couple's child-bearing decision during the cohabitation period was made without knowing for sure that they would get married. In other words, although these couples eventually became married, in terms of fertility, they behaved just like other cohabitors when they were still in cohabitation.

The consequence of including this group in the cohabiting sample is ambiguous. On the one hand, it is possible that these couples decided to get married precisely because they had children while cohabiting and considered marriage as the more appropriate union for raising children. In that case, our estimation of the fertility differences between cohabiting and married couples is biased downward. On the other hand, it is also possible that these cohabiting couples are anticipating future marriage and postpone child-bearing to the future, and thus avoid having children in cohabitation. As a result, our estimation of the fertility difference could be biased upward.

Some descriptive statistics are presented in Table 1. The cohabiting sample consists of 655 couples, and the married sample consists of 3701 couples. On average, a married couple has 1.63

[3] This sub-sample of 390 couples who started with cohabitation and subsequently got married suggests a study of fertility in cohabitation and marriage in a dynamic context. We leave this possibility open for future research.

children, considerably more than the average of 0.42 among cohabiting couples. About 61% of cohabiting couples are childless, while only 22% of married couples have no children.

Male and female wage rates are relevant variables in both marriage and fertility decisions. However, we cannot use the actual wage rates, partly because the actual wage rate is unknown for non-workers. In fact, even if the actual wage rate were available for everybody, it is endogenous (especially for women) because a woman's fertility behavior affects her labor market experience and thus her wage rate.

Following standard practice among labor economists, we use estimated wage rates. Thus both male and female wage rates in Table 1a are estimated potential earning abilities based on education, potential experience (age - education - 6), and geographic location (see Table 2). Female wage estimation additionally takes into account potential selection biases from labor force participation.[4] Married individuals have a slightly higher estimated potential wage rate. Married couples have higher male-female wage differences, which indicate a better match of comparative advantages.

On average, married individuals are more than four years older, and are slightly better educated. Cohabiting couples are more likely to live in big cities, less likely to reside in the South, and more likely to be black. The average tenure for marriage is 11.7 years, compared to 3 years for cohabitation. For both married and cohabiting women, less-educated women and black women tend to have more children.

4. Empirical Results

[4] Whether there is a child under age 6 and non-wife family income (family income minus wife's income) are assumed to affect women's labor force participation but not their earning ability. And these variables are used to identify the wage equation for women.

This section presents and discusses regression results. All regressions use the sample weight to correct for non-random sampling in the PSID.

4.1 Choice between marriage and cohabitation

The maximum likelihood estimation of the marital choice equation is presented in Table 3. The coefficient of the male-female wage difference is positive and significant, indicating that individuals are more likely to choose marriage over cohabitation when the difference is larger. Since the difference indicates the male's and female's comparative advantages, a larger difference implies higher marital gains through household specialization. Thus, this result is consistent with the household specialization theory.

Female age is a significant factor. Compared with younger women, older women tend to end up in marriage rather than in cohabitation, which is expected. However, it is surprising that male age has a negative coefficient, although small in magnitude and only significant at the 10% level.

Female education has a significant and positive coefficient of 0.270. However, since education also enters an individual's wage equation, its total effect must be adjusted accordingly. In the female wage equation (Table 2), the coefficient of education is 0.128; the male-female wage difference in the marital choice equation has a coefficient of 2.138. Thus the indirect effect of female education through the wage difference variable is -0.274 (= -2.138*0.128). And therefore, female education's total effect on marital choice is negative but very small (-0.004=0.270-0.274). That is to say, although higher education alone makes women more likely to choose marriage, they will have difficulties finding partners that match their comparative advantages. And these two effects almost cancel each other. A similar calculation suggests that male education has a positive total effect that is not negligible (2.138*0.112-0.171 = 0.068). That is, a well-educated man has a higher chance of being married rather than cohabiting.

Black women have a significantly lower probability of being married. Also, living in the south or the west is associated with a higher probability of getting married. When the county unemployment rate is high, the probability of choosing marriage increases significantly, probably because marriage provides a type of unemployment insurance. If the female's mother has completed high-school education, she is more likely to be married. On the other hand, her father's education and her religious preference do not significantly affect the marital choice.

4.2 Number of children

Two estimators, Terza's and Greene's, are used to estimate the count data models specified in equations (1)-(4). Tables 4 and 5 present the results.

The two estimators give fairly similar results. The coefficient of wage difference is positive and statistically significant for both married and cohabiting couples. That is, when the man's potential wage rate is much higher than the woman's, the couple tends to have more children. This result is also consistent with the economic theory of fertility that emphasizes household specialization.

As confirmed by many other studies, family income is negatively and significantly correlated with the number of children for married couples. However, the income effect is positive and insignificant for cohabiting couples. Perhaps the lower expectation about the future of the cohabiting relationship coupled with low earnings makes child-bearing a particular risky “investment” for low-income cohabitors, and thus they avoid having children and make the negative income effect disappear.

Older women tend to have had more children, which is true in both types of relationship. The direct effect of female education is positive, which means better educated women would like to have more children. Yet again, such individuals have higher earning abilities and hence higher opportunity costs. In Table 4, for married couples, the coefficient of female education is 0.653; the coefficient of female wage is -5.348; and the effect of female education on female wage is 0.128 in Table 2. Thus, the net effect of female education is negative: -0.031=0.653-5.348*0.128. Similar calculations show that the net effect of female education is also negative for cohabiting couples. However, male education has a positive net effect for both married and cohabiting couples.

Both married and cohabiting couples living in large cities tend to have more children. Being black increases the number of children for married couples but not for cohabiting couples. Compared to living in the north, living in the South or the West increases the number of children for married couples, but has insignificant effects on cohabiting couples. Living in north central region of the country, however, reduces cohabiting couples' fertility significantly but has an insignificant effect on married couples. In both marriage and cohabitation, longer relationship duration increases the number of children, although the positive effect decreases as the squared duration has negative parameters.

The value of coefficient θ is positive and statistically significant at the 5% level for the marriage sample, which suggests that people wanting more children are more likely to choose marriage. For the cohabitation sample, the value of θ is also positive but smaller in magnitude and not statistically significant.[5]

4.3 Decomposition of fertility difference

[5] For cohabiting couples, although θ is positive, the selection effect is actually negative. See the negative sign built in for θ in equation (19).

We have shown that married couples tend to have more children than cohabiting couples. This section further examines the major factors that contribute to this difference.

Letting $\hat{y}_{1i}$ be the log of predicted $E(y_i \mid M_i = 1)$ and $\hat{y}_{0i}$ the log of predicted $E(y_i \mid M_i = 0)$, from equations (10) and (13), we derive that

$$\hat{y}_{1i} = x_i \hat{\beta}_m + \ln \frac{\Phi(z_i \hat{\gamma} + \theta_m)}{\Phi(z_i \hat{\gamma})}, \text{ and} \tag{20}$$

$$\hat{y}_{0i} = x_i \hat{\beta}_c + \ln \frac{1 - \Phi(z_i \hat{\gamma} + \theta_c)}{1 - \Phi(z_i \hat{\gamma})}. \tag{21}$$

Taking the average over the marriage and cohabitation samples respectively, we obtain

$$\bar{\hat{y}}_1 - \bar{\hat{y}}_0 = \bar{x}_m (\hat{\beta}_m - \hat{\beta}_c) + (\bar{x}_m - \bar{x}_c) \hat{\beta}_c + (\bar{T}_m - \bar{T}_c), \tag{22}$$

where $\bar{x}_m = \frac{1}{N_m} \sum_i x_i M_i$, $\bar{x}_c = \frac{1}{N_c} \sum_i x_i (1 - M_i)$, $\bar{T}_m = \frac{1}{N_m} \sum_i \ln \frac{\Phi(z_i \hat{\gamma} + \theta_m)}{\Phi(z_i \hat{\gamma})} M_i$,

$\bar{T}_c = \frac{1}{N_c} \sum_i \ln \frac{\Phi(z_i \hat{\gamma} + \theta_c)}{\Phi(z_i \hat{\gamma})} (1 - M_i)$, and N_m and N_c are the size of the marriage and cohabitation samples, respectively. The first term in equation (22) represents behavioral differences between married and cohabiting couples, the second term corresponds to the differences attributable to explanatory variables, and the third term can be thought of as the difference due to unobserved self-selection.

Table 6 presents the decomposition based on the Terza estimator. All numbers in the three columns are normalized by the total difference (1.577), so that they add up to 1. The Greene estimator gives similar decomposition results.

The last column of Table 6 shows that the contribution of selection biases is positive, meaning that people who want more children tend to choose marriage. The magnitude of the number implies that unobserved self-selection accounts for about 40% of the fertility differences between married and cohabiting couples.

The other two terms of the decomposition explain about 60% of the fertility differences. It is interesting to note that they have opposite signs: $\bar{x}_m(\hat{\beta}_m - \hat{\beta}_c)$ is positive while $(\bar{x}_m - \bar{x}_c)\hat{\beta}_c$ is negative. This has two implications: First, marriage changes a couple's behavior so dramatically that if cohabitors and married couples had similar socioeconomic characteristics, and if cohabitation lasted equally as long as marriage (i.e., $\bar{x}_m = \bar{x}_c$), the fertility difference would be even greater than currently observed. Second, married couples have more children not because of their observed characteristics; in fact, if married couples behaved the same way as cohabiting couples (i.e., $\hat{\beta}_m = \hat{\beta}_c$), they would have had fewer children.

The numbers in the first two columns indicate the relative importance of each variable in explaining the overall difference. The wage difference variable has positive contributions in both columns. The positive value in the first column implies that married couples respond more to the same wage differences, probably because marriage is relatively stable and thus the same wage difference provides more incentives for household specialization in marriage. The positive value in the second column implies that a couple with a higher wage difference is more likely to choose marriage, which in turn leads to more children in marriage because fertility rises with increasing wage differences.

Like the wage difference, female education and female age have positive contributions in both columns. Again, this suggests that the fertility of married couples responds more to education and age and that married couples are on average better educated and older. On the other hand, male education and male age have different effects in the two columns. Again, to understand the overall effects of these variables, it is important to recognize that they also enter the equation through the wage difference variable.

The duration variable is also very important, which is natural since couples are likely to have more children the longer they stay together. However, if we add up the four values in the two columns that are associated with “tenure” and “tenure squared,” we find a fairly small total effect. That is, although the duration of marriage is much longer, the duration itself has a much smaller effect on married couples. The two effects almost cancel out and thus account for very little of the difference in fertility between married and cohabiting couples.

Overall, our decomposition shows that while self-selection explains a large proportion of the fertility difference between married and cohabiting couples, it is not the most important factor. The difference is mostly explained by the fact that married couples and cohabiting couples behave very differently under the two residential unions. It is likely that the behavioral difference stems from the distinct natures of the two unions. Because marriage is a closer bond that is more stable, it allows married couples to specialize more according to their comparative advantages: The man concentrates on market production and becomes a specialized wage earner, whereas the woman specializes more in household production by raising more children.

Given that cohabitation is becoming a more common residential union, and more and more marriages are preceded by cohabitation, household specialization will become increasingly uncommon and only occur among higher-age couples. An implication of this is that fertility rate will continue to drop along with the rising trend of cohabitation in the United States.

5. Conclusion

Although previous studies have consistently shown that cohabiting couples tend to have fewer children than married couples, such studies have done little to explain why this difference in fertility occurs. Identifying and quantifying the cause of the fertility difference between these two groups is necessary for fully understanding the implications of the rising trend of cohabitation. In this paper, we develop a switching regression model that allows us to measure the contribution of each factor through a decomposition of the fertility difference.

We model fertility behavior by estimating count data models within an endogenous switching framework. For both married and cohabiting couples, our results show evidence of household specialization. We find that the number of children of married and cohabiting couples is positively correlated with the male-female wage difference in a household. We interpret this relationship as demonstrating a link between fertility and intra-household comparative advantage, which is driven by the fact that a couple with a higher wage difference has more to gain for the man to specialize in earning wages and the woman to specialize in household production, according to their comparative advantages.

Based on the regression results from the count data models, we decompose the difference in fertility between married and cohabiting couples. We show that although married and cohabiting couples have different characteristics, these observed selection variables contributed little to their fertility difference. Selection biases (i.e., unobserved self-selection) account for about 40% of the fertility difference. The most important factor for explaining the fertility difference between these two groups is the behavioral difference associated with marriage and cohabitation. We find that cohabiting couples respond to many independent variables very differently than married couples. In fact, if married and cohabiting couples had the same characteristics, the fertility difference between these two groups would be even larger than currently observed.

These results from the decomposition imply that married couples have more children primarily because they are in marriage, rather than because of the characteristics that influenced their decision to enter wedlock. In other words, the marriage itself, rather than any other individual characteristics, influenced their fertility behavior. Why does marriage make a couple behave so differently from a cohabiting couple? Most likely, it is because the closeness and stability of marriage enables married couples to specialize according to their comparative advantages, where a woman will spend more time raising children with little concern of the resulting decline of earnings capacity in the labor market. In contrast, the shorter expected duration of cohabitation makes it less feasible for such couples to specialize according to comparative advantage and invest in “durable goods” like children. Given the differences in fertility behavior associated with cohabitation, its rising trend is likely to have a substantial effect on population growth in the United States.

Our static analysis calls for further investigation along this line. It is known that many cohabiting couples become married later on. Hence future research may directly investigate whether this change from cohabitation to marriage affects the couple's fertility behavior. The results from our findings suggest that this change should have a great effect on fertility, but our static approach prevents us from drawing such a strong conclusion. Thus a dynamic study that takes full advantage of the longitudinal data from the PSID will likely shed more light on this topic.

References

Axinn, William G., Marin E. Clarkberg, and Arland Thornton, 1994, “Family Influences on Family Size Preferences,” *Demography* 31, 65-79.

Bachrach, Christine A., 1987, “Cohabitation and Reproductive Behavior in the U.S.,” *Demography* 24, 623-37.

Barber, Jennifer S. and William G. Axinn, 1998, “The Impact of Parental Pressure for Grandchildren on Young People's Entry into Cohabitation and Marriage,” *Population Studies* 52, 129-44.

Becker, Gary S, 1960, *An Economic Analysis of Fertility: Demographic and Economic Changes in developed Countries*, Princeton University Press and the National Bureau of Economic Research.

Becker, Gary S, 1981, *A Treatise on the Family*, Cambridge: Harvard University Press.

Becker, Gary and H. G. Lewis, 1973, “On the Interaction between the Quantity and Quality of Children,” *Journal of Political Economy* 81, S279-88.

Blumstein, P. and P. Schwartz, 1983, *American Couples*, New York: Morrow.

Bumpass, Larry, and Hsien-hen Lu, 2000, “Trends in Cohabitation and Implications for Children's Family Contexts in the United States,” *Population Studies* 54, 29-41.

Bumpass, Larry L. and James A. Sweet, 1989, “National Estimates of Cohabitation,” *Demography* 26, 615-25.

Bumpass, Larry L., James A. Sweet, and Andrew Cherlin, 1991, “The Role of Cohabitation in Declining Rates of Marriage,” *Journal of Marriage and the Family* 53, 913-27.

Cameron, A. Colin and Pravin K. Trivedi, 1998, *Regression Analysis of Count Data*, Cambridge University Press.

Easterlin, Richard A., 1968, *Population, Labor Force, and Long Swings in Economic Growth: The American Experience*, New York, NY, National Bureau of Economic Research.

Greene, William H., 1994, “Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Model,” working paper No. EC-94-10, Department of Economics, Stern School of Business, New York University.

Greene, William H., 1995, “Sample Selection in the Poisson Regression Model,” working paper No. EC-95-6, Department of Economics, Stern School of Business, New York University.

Greene, William H., 2001, “FIML Estimation of Sample Selection Models for Count Data,” in *Economic Theory, Dynamics and Markets, Essays in Honor of Ryuzo Sato*, edited by T. Negishi, R. Ramachandran, K. Mino. Boston, MA: Kluwer.

Heckman, James J., 1979, “Sample Selection Bias as a Specification Error,” *Econometrica* 47, 153-61.

Heimdal, Kristen R. and Sharon K. Houseknecht, 2003, “Cohabiting and Married Couples’ Income Organization Approaches in Sweden and the United States,” *Journal of Marriage and Family* 65, 539-549.

Lee, Lung-Fei, 1978, “Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables,” *International Economic Review*, 19, 415-433.

Lillard, Lee A., Michael J. Brien, and Linda J. Waite, 1995, “Premarital Cohabitation and Subsequent Marital Dissolution: A Matter of Self-Selection?” *Demography* 32, 437-57.

Manning, Wendy D., 1993, “Marriage and Cohabitation Following Premarital Conceptions,” *Journal of Marriage and the Family* 55, 839-50.

Manning, Wendy D., 1995, “Marriage, Cohabitation, and Entry into Motherhood,” *Journal of Marriage and the Family* 57, 191-200.

Musick, Kelly, 2002, “Planned and Unplanned Childbearing among Unmarried Women,” *Journal of Marriage and Family* 64, 915-29.

National Center for Health Statistics, 2003, *Health, United States, 2003: With Chartbook on Trends in the Health of Americans*. Hyattsville, Maryland.

Raley, R. Kelly, 2001, “Increasing Fertility in Cohabiting Unions: Evidence for the Second Demographic Transition in the United States?” *Demography* 38, 59-66.

Ressler, Rand W. and Melissa S. Waters, 1995, “The Economics of Cohabitation.” *Kyklos* 48, 577-92.

Rindfuss, Ronald R. and Audrey VandenHeuvel, 1990, “Cohabitation: A Precursor to Marriage or An Alternative to Being Single?” *Population and Development Review* 16, 703-26.

Rosenzweig, Mark R., 1999, “Welfare, Marital Prospects, and Nonmarital Childbearing,” *Journal of Political Economy* 107(6), S3-S32.

Simmons, Tavia and Martin O'Connell, 2003, “Married-Couple and Unmarried-Partner Households: 2000,” Census 2000 Special Reports, U. S. Census Bureau, February 2003. Online at: http://www.census.gov/prod/2003pubs/censr-5.pdf.

South, Scott J. and Glenna Spitze, 1994. “Housework in Marital and Nonmarital Households.” *American Sociological Review* 59, 327-47.

Spanier, Graham B., 1983, “Married and Unmarried Cohabitation in the United States: 1980,” *Journal of Marriage and Family* 45, 277-88.

Terza, Joseph V., 1998, “Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects,” *Journal of Econometrics* 84, 129-54.

Thornton, Arland, 1988, “Cohabitation and Marriage in the 1980s,” *Demography* 25, 497-508.

Treas, Judith and Deirdre Giesen, 2000, “Sexual Infidelity Among Married and Cohabiting Americans,” *Journal of Marriage and Family* 62, 48-60.

Waite, Linda J., 1995, “Does Marriage Matter?” *Demography* 32, 483-507.

Winkler, Anne E., 1997, “Economic Decision-Making by Cohabitors: Findings Regarding Income Pooling,” *Applied Economics* 29, 1079-1090.

Table 1a: Descriptive Statistics

	Married Couples		Cohabiting Couples	
	Mean	Std. Dev.	Mean	Std. Dev.
number of children	1.63	1.20	0.42	0.79
estimated female wage	2.14	0.35	2.04	0.36
estimated male wage	2.59	0.33	2.44	0.32
male wage - female wage	0.46	0.25	0.40	0.29
log family income	10.62	0.79	10.29	0.89
female age	34.05	6.72	29.76	6.82
female education	13.09	2.24	12.65	2.19
male age	36.51	7.67	32.37	8.23
male education	13.26	2.41	12.65	2.46
female is Baptist	0.21	0.41	0.22	0.41
female is Jew	0.024	0.15	0.015	0.12
female other religious preference	0.51	0.50	0.50	0.50
tenure of the relation	11.66	7.43	2.94	2.05
female's number of siblings	3.42	2.48	3.64	3.14
male's number of siblings	3.43	2.55	3.55	2.65
female is black	0.090	0.29	0.15	0.36
live in SMSA	0.51	0.50	0.62	0.49
live in north central	0.26	0.44	0.24	0.43
live in south	0.36	0.48	0.28	0.45
live in west	0.19	0.39	0.22	0.42
county unemployment rate	6.82	2.43	6.43	2.44
sample size	3701		655	

Table 1b: Birth frequency

	Married Couples		Cohabiting Couples	
Number of Children	Frequency	Percent	Frequency	Percent
0	813	21.97	402	61.37
1	861	23.26	138	21.07
2	1,229	33.21	82	12.52
3	539	14.56	22	3.36
4	191	5.16	9	1.37
6	54	1.46	1	0.15
7	10	0.27	1	0.15
8	1	0.03	0	0
9	2	0.05	0	0
10	1	0.03	0	0
Total	3,701	100	655	100

Table 1c: Birth by Marital Status, Education, and Race

		Children	Std. Dev.	N
Married	less than high school	1.99	1.38	577
	high school	1.68	1.14	1608
	more than high school	1.49	1.19	1516
Cohabiting	less than high school	0.77	0.92	159
	high school	0.48	0.86	297
	more than high school	0.17	0.52	199
Married	black	1.75	1.34	947
	non-black	1.62	1.19	2754
Cohabiting	black	0.74	1.01	234
	non-black	0.36	0.74	421

Table 2: Estimating Earning Potential

(Dependent variable: log(wage))

	Women	Men
education	0.128 (0.008)***	0.112 (0.006)***
experience [a]	0.036 (0.010)***	0.054 (0.006)***
experience squared	-0.0005 (0.0003)	-0.001 (0.0002)***
city	0.128 (0.028)***	0.058 (0.024)**
north central	-0.121 (0.042)***	-0.112 (0.039)***
south	-0.144 (0.041)***	-0.179 (0.036)***
west	-0.010 (0.042)	-0.083 (0.036)**
lambda [b]	-0.280 (0.097)***	--- ---
constant	0.182 (0.121)	0.573 (0.090)***
number of observations	4356	4356

Note: Standard errors are in parentheses. *-significant at the 10% level, **-significant at the 5% level, ***-significant at the 1% level.
a: experience = age – years of education – 6.
b: selection variable.

Table 3: Choice between Marriage and Cohabitation: Probit

(Dependent variable: 1 if married and 0 if cohabiting)

	Coefficients	Standard Errors
constant	-3.112***	0.454
male wage - female wage	2.138***	0.525
female age	0.072***	0.013
female education	0.270***	0.067
male age	-0.022*	0.013
male education	-0.171***	0.052
live in SMSA	-0.095	0.073
female black	-0.262***	0.100
female Baptist	-0.042	0.105
female Jewish	0.009	0.301
other religious preference	-0.068	0.077
live in north central	0.226**	0.098
live in south	0.513***	0.102
live in west	0.287***	0.103
female's mother completed high school	0.155**	0.077
female's father completed high school	0.003	0.074
unemployment rate	0.040***	0.015
log likelihood	-1463.53	
number of observations	4333	

*-significant at the 10% level, **-significant at the 5% level, ***-significant at the 1% level.

Table 4: Fertility Behaviors in Marriage and Cohabitation: Terza's Estimator

(Dependent variable: number of children)

	MARRIED COUPLES		COHABITING COUPLES	
	Coefficient	Std. Err.	Coefficient	Std. Err.
constant	-3.814***	0.259	-3.387***	1.049
male wage - female wage	5.348***	0.184	4.831***	1.086
log family income	-0.069***	0.009	0.035	0.030
female age	0.087***	0.005	0.077**	0.035
female education	0.653***	0.023	0.601***	0.144
male age	-0.099***	0.004	-0.124***	0.021
male education	-0.507***	0.018	-0.543***	0.087
live in SMSA	0.236***	0.018	0.253**	0.112
female black	0.110***	0.032	-0.045	0.204
female Baptist	-0.057**	0.029	0.404***	0.158
female Jewish	0.049	0.059	1.121**	0.506
other religious preference	0.004	0.020	0.137	0.141
female's siblings	0.008**	0.003	0.026	0.015
male's siblings	0.008**	0.003	0.024	0.015
tenure	0.099***	0.006	0.574***	0.077
tenure squared	-0.0016***	0.0002	-0.041***	0.007
live in north central	0.013	0.025	-0.475***	0.155
live in south	0.147***	0.031	-0.412	0.277
live in west	0.408***	0.030	-0.148	0.186
theta	0.968**	0.494	0.335	0.520
number of observations	3679		654	

*-significant at the 10% level, **-significant at the 5% level, ***-significant at the 1% level.

Table 5: Fertility Behaviors in Marriage and Cohabitation: Greene's Estimator

(Dependent variable: number of children)

	MARRIED COUPLES		COHABITING COUPLES	
	Coefficient	Std. Err.	Coefficient	Std. Err.
constant	-3.957***	0.344	-3.557***	1.130
male wage - female wage	5.414***	0.211	4.982***	1.087
log family income	-0.069***	0.009	0.034	0.030
female age	0.089***	0.006	0.082**	0.035
female education	0.662***	0.027	0.620***	0.144
male age	-0.100***	0.004	-0.126***	0.021
male education	-0.513***	0.020	-0.554***	0.087
live in SMSA	0.234***	0.019	0.248**	0.111
female black	0.105***	0.033	-0.067	0.204
female Baptist	-0.058**	0.029	0.400**	0.157
female Jewish	0.049	0.059	1.121**	0.507
other religious preference	0.002	0.020	0.130	0.141
female's siblings	0.008**	0.003	0.026	0.015
male's siblings	0.008**	0.003	0.024	0.015
tenure	0.098***	0.006	0.576***	0.077
tenure squared	-0.0015***	0.0002	-0.041***	0.007
live in north central	0.018	0.026	-0.459***	0.154
live in south	0.159***	0.037	-0.377	0.276
live in west	0.417***	0.033	-0.129	0.186
theta	0.667***	0.208	0.428	0.541
number of observations	3679		654	

*-significant at the 10% level, **-significant at the 5% level, ***-significant at the 1% level.

Table 6: Decomposition of Fertility Difference between Married and Cohabiting Couples

	FERTILITY DIFFERENCE		
	$\bar{x}_m(\hat{\beta}_m - \hat{\beta}_c)$	$(\bar{x}_m - \bar{x}_c)\hat{\beta}_c$	Selection Bias
constant	-0.271	0	
male wage - female wage	0.149	0.172	
log family income	-0.695	0.007	
female age	0.219	0.209	
female education	0.437	0.165	
male age	0.587	-0.326	
male education	0.299	-0.209	
live in SMSA	-0.005	-0.018	
female black	0.009	0.002	
female Baptist	-0.061	-0.001	
female Jewish	-0.016	0.006	
other religious preference	-0.042	0.000	
female's siblings	-0.038	-0.004	
male's siblings	-0.037	-0.002	
tenure	-3.515	3.175	
tenure squared	4.791	-4.645	
live in north central	0.079	-0.005	
live in south	0.127	-0.020	
live in west	0.068	0.003	
Sum	2.085	-1.490	0.405
logE_m( )-logE_c( )	0.349 – (-1.228) = 1.577		

The contribution of each variable is normalized by the total difference (1.577), thus they sum up to 1.