Modeling life Transitions on Fertility in Young Adults

Maciej Misztal
University of North Carolina at Chapel Hill

October 22, 2010
1 Introduction

Cohabitation went from being fairly uncommon in the 1960's to the predominant form of union in America at the turn of the century Schoen et al. (2007). Out of wedlock births rose to a third of all births over the same period. These shifts in the establishment of families are seen through every ethnic and socioeconomic group Kennedy and Bumpass (2008). Although there has been an overall delay in age of first birth since the 1950's, it has occurred almost completely among educated women. This increase in cohabitation and related decrease in marriage naturally increased the amount of children born out of wedlock. Due to the greater instability of cohabiting couples, more of these children will grow up in less stable households. Cohabitation also has a significantly different impact on the workforce. While single women work considerably more than married women, their labor force participation is in some data sets statistically indistinguishable to that of cohabiting women Gemici and Laufer (2010). Further more, there are changes in trends within cohabiting couples. The portion of women likely to be married to the partner they are cohabiting with after five years is falling Schoen et al. (2007).

Economists have been slower than other social scientists to focus on cohabitation and changing transitions in youth. A large part of this reason is sparse, incomplete, or unreliable data. Moreover, studying transitions into young adulthood provides problems to fundamental assumptions of labor models. While it is reasonable to assume that women over the course of their entire lives will be forward-looking and rational with respect to their labor, spending, and fertility decisions, it is less certain that all most pregnancies among teenagers are planned.

This paper intends model the schooling, labor, marriage and cohabitation decision on fertility. The focus will be on young women between the ages of 14 to 26 in the NLSY97 cohort. The fertility study will be restricted to children who remain with their mother. While pregnancy may not be rational among these young women, the decision to keep the child within the household rather than give it up for adoption or older caretakers is more justifiable. The outcomes we model are likely correlated through unobserved time-varying and permanent characteristics. We will estimate using the discrete random effects method that will help provide unbiased estimates of fertility decisions on future wages. This paper intends to add to the literature in
three ways. First, while the NLSY97 dataset has been used widely in the study of adolescents, its rich data concerning further education and full time labor decisions is only beginning to be released. Comparing estimates to the widely used NLSY79 may be useful in addressing which aspects of the life cycle decisions have changed over the last generation. Second, it is able to better measure cohabitation in addition to marriage when addressing the fertility decision. Not only is cohabitation more prevalent, but the data is better designed to track it. Finally, in contrast to studies using the National Longitudinal Study of Adolescent Health (Add Health), this model is able to track detailed year by year information, not only about wages and work, but other relevant behaviors. This study aims to take advantage of these aspects of the data to gauge the impact of changing fertility and living pattern on wages.

2 Literature Review

The paper grazes topics that have been examined in depth in the literature. For models of labor supply, consider Blundell and Macurdy (1999). Although our model only seeks to predict outcomes, not preferences, much work has been done in the dynamic structural framework to estimate sequential joint decision of young women. van der Klaauw (1996) creates a female life cycle model that is able to structurally estimate women’s labor contributions along with their (potential) marriage and divorce decisions with respect to education, race and potential earning of each spouse. He finds the utility of marriage is increasing in the husband’s wage but decreasing in the women’s wage. This suggests the theory of specialization to explain the marriage premium. The marriage may be less useful and binding for both parties as their contributions to the household become substitutes.

Keane and Wolpin (2006) provide perhaps the most complete structural model of women’s life cycle decisions. They use a model that incorporates schooling, marriage, welfare acceptance, labor, and fertility. They use the NLSY79 cohort through 1990. They confirm unobserved heterogeneity is crucial to the understanding of younger women’s decisions in particular. They find that causes of discrepancies in the outcomes of minority and white women are due to a number of complex interrelated factors, unobserved to the econometrician. No one decision is
responsible for less productive outcomes in the others. They also conclude that the distribution
of unobserved “types” of women is imperative to understand the differences in outcome within
and between minorities as well as whites. This was similar to the result of their research
concerning youths dropping out of school Eckstein and Wolpin (1999), which also confirmed that
certain permanent unobservable types in every group where simply more predisposed towards
dropping out or staying in.

They take into account that the trends regarding cohabitation support the theory of specialization. Gemici and Laufer (2010) use NLSY97 to form a dynamic model of household creations and dissolution, focusing on the matching aspect. Their model suggests cohabitation as a third choice between marriage and remaining single. They then predict policies that would change the cost of divorce. Their model suggests that marriage is optimal when the specialization within the household is greater. Musick (2007) uses data from the NSFG to study intended and non-intended pregnancies for married and cohabiting partners. She uses Applied Maximum likelihood to jointly model intended pregnancies, unintended pregnancies, and the entrances into marriage and cohabiting relationships, in an effort to control for endogeneity. After introducing controls for unmeasured heterogeneity, she finds a significant reduction in the effect of cohabitation on birth, suggesting that many of the women who give birth in cohabiting relationships are of the marrying “type” but have a hard time finding a willing marriage partner. This paper is related to a series of papers using the methods established by Lillard (1993). It creates a set of simultaneous hazard functions that are interdependent. In the first paper they are conception and marriage. The models further allow multiple instances and failures of an event, in this case marriage or conception. The models are estimated using MLE and are dependent on specific parametric assumptions. They are identified by the fact that the data has events overlapping in different patterns. The consistent result of models using this technique is the endogeneity bias involved in simple predictions of relationships on fertility decisions among young women.

There is evidence in the literature that the poor schooling and wage outcomes associated
with teenage mothers are driven by omitted variable bias and unobserved heterogeneity. Geronimus and Korenman (1992) uses PSID, NLSYW, and NLSY79 data to examine teenage mothers as compared to their sisters. They find evidence that the negative consequences of childbearing
are the result of family characteristics as well as unobserved individual characteristics. They suggests that the prevalent cross-sectional literature unfairly attributes substandard income to youth pregnancies. Hoffman et al. (1993) considers evidence of the cost of teenage pregnancy using PSID information. The authors use the sisters of teen mothers to control for background. They estimate the cost of teenage pregnancy is not much less significant than is usually assumed and it is the disadvantaged background of most teen mothers that drives their poor outcomes.

Upchurch et al. (2002) suggest that motherhood outside of marriage may be relatively desirable for economically disadvantaged women who do not expect potential husbands to contribute much, and other groups for whom marriage-appropriate men are scarce. They also want to incorporate multiple births into their model, including for those women who have been previously married. The paper concerns itself only to nonmarital conception leading to live births. They intend to quantify the effects of education marriage and marital dissolution and prior fertility on nonmarital childbearing and compare models that control for heterogeneity but not endogeneity.

The results complement Brien et al. (1999). Non-black women who drop out of school before they conceive are more likely to get pregnant than their enrolled counterparts, although there is no long term effect. For all women, having been married decreases the probability of nonmarital pregnancy after divorce. Those that are predisposed towards marriage are also predisposed towards nonmarital pregnancy. Prior childbearing experience both within and outside of marriage will also significantly affect further fertility decisions. Finally, race and having come from a disadvantaged background have significant effects on the joint and sequential decisions of women.

This paper will add to these strands of literature by combining both labor/schooling decisions with the union/fertility decisions. We control for the endogeneity by jointly estimating the equations. We also control for time varying and permanent heterogeneity. The latter will control for “types”.
3 Data

The NLSY97 data set follows a nationally representative sample of individuals across the country born between 1980 and 1984. Data was first collected in 1997, when the respondents were between the ages of 12 and 14. During the first year of the interview, extensive data was collected on the household in which the respondents grew up, including an interview with the caretakers. Data is now available for 12 rounds, into their mid-twenties. The data is designed primarily with the labor decisions in mind, and keeps yearly updates of all the work the respondents have done. There is also information about schooling, including information about the high schools and colleges as well as training programs.

Crucial to the purposes of this study is the information about relationship and fertility decisions. Data is collected starting at the age of 16 and goes into detail about frequency and length of relationships, as well as intentions of pregnancy and use of contraception. Special attention is paid to tracking the length and number of cohabiting relationships throughout the first 5 rounds. To help ensure truthful answers for topics such as sexual activity or drug use, certain parts of the interview are self-administered via laptop and headphones.

The data I am currently using is only geocoded for region and presence in a Metropolitan Statistical Area. I hope to gain access to zip code and county data in the future. This will allow me to at minimum control for state policies and county policies, as well as schools.

I also have information on the amount of time and source of childcare, down to who drives the children to school. Women without children in the home are regularly asked if they have relatives or close friends who would be willing to take care of their hypothetical child.

In addition to the Census data regarding the makeup of states and counties, there are several policy variables that can be used for identification. During the formative years of the participants, the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) aimed to change the incentives related to teenage pregnancy. Around the same time, many states with below average rates of university attendance implemented large scale merit-based scholarships. Also, in 1997 legislation passed that encouraged states to extend Medicaid coverage
of low-income youth. In all these cases the policy shift is exogenous and impacts the NLSY cohort. Because the implementation was done by state and varied over time, it will tease out variation of the data in the estimation.

4 Theoretical Model

(Notice: this section is not crucial at this time) The following theory motivates the empirical model but will not be fully implemented. The empirical model is not structural, does not predict preferences, and simply models outcome. we start with a general life cycle interpretation of the neoclassical model incorporating fertility decisions, as set forth in Becker et al. (1960). Women will be followed from age 14. The woman will have a budget constraint and a time constraint. She will strive to maximize current utility subject to them. The woman derives utility from consumption goods $C_{it}$ and leisure time $L_{it}$.

The woman may also derive utility from her marital/cohabitation status $m_{it}$ if she is married/cohabitating at time $t$. This utility may depend on the partner’s compatibility and characteristics, and transitioning into and out of either state is presumed to be costly. The mother derives utility from her children $n_{it}$ and has a one time additional positive shock when the child is a newborn $nb_{it}$. There is also utility from moving away from the parental household $q_{it}$. The Utility each period may also be influenced by time varying preference parameters over work, schooling, and marriage, as well as time invariant preference parameter. Finally, it will be influenced by age and demographic characteristics $x_{it}$.

$$U_{it} = U(C_{it}, L_{it}, m_{it}, n_{it}, q_{it}, \theta_{it}, \mu_{it}, x_{it})$$ (1)

Time available to the woman is fixed in each period and can be distributed towards hours worked $h_{it}$, and schooling $s_{it}$ if one attends school(or training) in this period $T^*$. What is left is attributed to leisure and non market activities. If the model incorporates childcare, then $n_{it}$ will interact with $m_{it}$, $q_{it}$ and $r_{it}$, availability of near by relatives, within the utility function. We
expect cohabiting partners to not provide as much childcare as husbands. Time for childcare is implicitly incorporated into $L_{it}$.

$$T = h_{it} + T^s * s_{it} + L_{it}(N_{it}, m_{it}, r_{it}, q_{it}) \tag{2}$$

$C_{it}$ is determined within the budget constraint. For simplicity I assume the budget is binding, although I do have data for assets and depts in the relevant periods. The budget is made up of hourly wages multiplied by hours worked $h_{it}$. Wages $w(S_{it}, \tau_{it})$ are influenced by $S_{it}$, the stock of Education, and $\tau_{it}$, the stock variable of indicating tenure at a particular job. $q_{it}^s$ represents cost of living away from home. Unearned income can be contributed by an individual’s partner $I_{it}^m$, one’s parents ($I_{it}^p$). There is also a cost to schooling per period, $t_{it}$ and children child care costs $c(N_{it}, n_{it})$. I begin by assuming that income from partner and parents is treated as exogenous. $g_{it}$ represents government assistance. The welfare eligibility rules would be strict and depend on state and other factors reflected by actual welfare requirements. The equation for eligibility is not yet modeled. Total income uncertain in every period, regardless of labor choices. The error term may manifest itself as a one time unexpected cost or some significant gift. It is represented by $\varepsilon_{it}$. It should be noted that the female discovers the value of the random error after the decision to work in that period has been made. Consider living with

$$C_{it} = h_{it} * w(S_{it}, \tau_{it}) - q_{it}^s + g_{it} + I_{it}^m + (I_{it}^p) - p_{it} * T^s * s_{it} + c(N_{it}, n_{it}) + \varepsilon_{it} \tag{3}$$

We allow cohabilitation $m_{it} = 1$ and marriage $m_{it} = 2$ status to be a probabilistic event which is dependent on the following:

$$Pr(m_{it} = 0, 1, 2) = (N_{it}, S_{it}, w_{it}, \theta_{it}, q_{it}^s, x_{it}) \tag{4}$$
We will incorporate a stock variable $M_{it}$ that increases with tenure of the partner. This stock variable is presumed to be larger in the case of marriage. The factors that affect the probability of marriage will also be relevant in case of divorce or separation. We would expect the cost of a divorce with a husband to be larger than the cost of a separation from a cohabitor.

The timing of the model is such. In the beginning of the period the woman observes a wage offer and a potential partner. In addition she knows the exogenous variables in the model including non-earned income, preference parameters as well as individual and area characteristics. The woman also observes the current state of stock variables including children, tenure, and schooling as well as the resulting wage. She is also aware for the value of the her choice variables in the last period. So her information at the beginning of the period is

$$Z_{it} = (M_{it}, I_{M, i,t-1}, r_{i,t-1}, I_{P, i,t-1}, G_{i,t-1}, \theta_{i}, \mu_{it}, X_{it}, B_{it}, S_{it}, \tau_{it}, N_{it}, w_{it}, s_{i,t-1}, H_{i,t-1}, b_{it}) \quad (5)$$

She then decides how to allocate her time into work $H_{it}$, schooling $s_{it}$, and leisure $L_{it}$ and whether to give add a child in the next period. Afterwards, the present value of lifetime utility associated with to income $\varepsilon_{it}$ is realized, the woman decides whether or not to add another child to the family $N_{it+1} = N_{it} + 1$.

Presumably, the woman would then maximize her present lifetime utility function.

5 Empirical Model

During the young adult period, there is great amount of endogeneity surrounding any decision. If we are to focus on the period between the mid-teens and the mid-twenties, we may not be able to account for the complex relationships between the decision to continue education, move from the parents’ residence to that of a partner, the decision to start a family or take care of one’s child. The defining characteristic of this part of life is transition from a fairly narrow set
of choices to freedom to optimize most facets that will lay the groundwork for the rest of their lives. The decision to stay in school between freshman and sophomore year is different than that of staying through another year of college education. The investment in human capital over this time will have a larger impact than those of any other time period. The decision to have child will put a large time and financial strain on anyone, and would force the alterations of the timing of all other decisions.

Allow $H_{ijt}$ to summarize the entire history of hours worked for person $i$ in county $j$ in up till period $t$. In the same way regard $M_{ijt-1}$, $Q_{it-1}$, $S_{ijt-1}$, and $N_{ijt-1}$ contain the history of marriage/cohabitation, living with parents, Schooling/training, and number of children,respectively. They can be considered stock variables. $X_{ijt}$ includes all exogenous characteristics, including spousal income, money and childcare from relatives, race, gender, government transfers, and so on. Consider $P_{ijt}$ to be the vector of community characteristics for each of the above decisions, including prices, tuition, male/female ratio, average rent, and so on. The model follows the discrete factor random effects methodology in Mroz (1999). Therefore errors can be decomposed into a time varying individual component $\kappa_{it}$, a permanent individual component $\omega_{t}$ and an idiosyncratic moment $\theta_{it}$. The timing of the model:

1. The female observes the history of all the decision variables from the last period as well as her personal characteristics and all community-level conditions.
2. The female observes a wage draw $w_{t}$. If I had modeled a draw for a different husband/cohabitor, it would also be observed at this point.
3. The female makes a simultaneous decision over employment, living at a parent’s residence, marriage/cohabitation, and schooling.
4. The woman then decides whether to accumulate another child for the next period.

If we assume extreme value distributed idiosyncratic terms, we can estimate simultaneously the long odds of a woman being employed as:

$$\ln \left[ \frac{p(h_{ijt} > 0)}{p(h_{ijt} = 0)} \right] = \alpha_{0} + \alpha_{1} H_{ijt-1} + \alpha_{2} M_{ijt-1} + \alpha_{3} Q_{it-1} + \alpha_{4} S_{ijt-1} + \alpha_{5} N_{it} + \alpha_{6} P_{ijt} + \alpha_{7} X_{ijt} + \alpha_{8} \kappa_{it} + \alpha_{9} \omega_{i} $$

(6)
The binary logit for an additional year of schooling/training vs not having an additional year:

\[
\ln \left[ \frac{p(s_{ijt} = 1)}{p(s_{ijt} = 0)} \right] = \beta_0 + \beta_1 H_{ijt-1} + \beta_2 M_{ijt-1} + \beta_3 Q_{it-1} + \beta_4 S_{ijt-1} + \beta_5 N_{it} + \beta_6 P_{ijt} + \beta_7 x_{ijt} + \beta_8 \kappa_{it} + \beta_9 \omega_i
\]

(7)

And likewise for \( Q_{ijt-1} \).

\[
\ln \left[ \frac{p(q_{ijt} = 1)}{p(q_{ijt} = 0)} \right] = \gamma_0 + \gamma_1 H_{ijt-1} + \gamma_2 M_{ijt-1} + \gamma_3 Q_{it-1} + \gamma_4 S_{ijt-1} + \gamma_5 N_{it} + \gamma_6 P_{ijt} + \gamma_7 x_{ijt} + \gamma_8 \kappa_{it} + \gamma_9 \omega_i
\]

(8)

The marriage equation is just a multinomial logit of being married \( (m_{ijt} = 2) \) or cohabiting \( (m_{ijt} = 1) \) relative to being single \( (m_{ijt} = 2) \) :

\[
\ln \left[ \frac{p(m_{ijt} = d)}{p(q_{ijt} = 0)} \right] = \delta_0 + \delta_1 H_{ijt-1} + \delta_2 M_{ijt-1} + \delta_3 Q_{it-1} + \delta_4 S_{ijt-1} + \delta_5 N_{it} + \delta_6 P_{ijt} + \delta_7 x_{ijt} + \delta_8 \kappa_{it} + \delta_9 \omega_i d = 1, 2
\]

(9)

with work supply intensity, conditional on working given by

\[
h_{ijt} = \sigma_0 + \sigma_1 H_{ijt-1} + \sigma_2 M_{ijt-1} + \sigma_3 Q_{it-1} + \sigma_4 S_{ijt-1} + \sigma_5 N_{it} + \sigma_6 P_{ijt} + \sigma_7 x_{ijt} + \sigma_8 \kappa_{it} + \sigma_9 \omega_i
\]

(10)

According to mincer’s equations, we can observe log wages at time \( t \), conditional on being employed as

\[
\ln(w_{it} | h \neq 0) = \rho_0 + \rho_1 \tau + \rho_2 \tau^2 + \rho_3 H_{ijt-1} + \rho_4 M_{ijt-1} + \rho_5 Q_{it-1} + \rho_6 S_{ijt-1} + \rho_7 N_{ijt-1} + \rho_8 P_{ijt}^w + \rho_9 x_{ijt} + \rho_{10} \kappa_{it} + \rho_{11} \omega_i + \theta_{it}^w
\]

(11)
where \( \tau \) is work tenure and \( P^w \) refers only to the community vectors dealing with wage.

Finally the log odds ratio of acquiring a child relative to not acquiring one is

\[
\ln \left[ \frac{p(nu_{ijt+1} = 1)}{p(nu_{ijt} = 0)} \right] = \rho_0 + \rho_1 H_{ijt} + \rho_2 M_{ijt} + \rho_3 Q^r_{it} + \rho_4 S_{ijt} + \rho_5 N_{ijt} + \rho_6 P^n_{ijt} + \rho_7 x_{ijt+1} + \rho_8 \kappa_{it} + \rho_9 \omega_i + \theta^n_{it} \tag{12}
\]

### 6 Identification and further work

I plan to use the availability of informal childcare to aid in identification, but am happy to take additional suggestions.

I am also uncertain about the framework and having birth as the dependent variable and the justification for the timing.

At one point I considered adding welfare as yet another jointly estimated equation, but believe this to be intractable.

The same with adding some sort of matching framework to cohabitation and marriage.

Also, given the specification so far, there is nothing in my model that does not allow me to estimate it for men in my data set.

I also have to deal with attrition. Any suggestions are welcome, including concerning the title.
References


