LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Demography

Labor-market status and first-time parenthood: The experience of female minorities in the United States, 1979-2012

Hongwei Han
ede14hha@student.lu.se

Abstract: Background. The United States has been a diverse country since it is establishment. The effects of socioeconomic characteristics including immigration, age, employment status, educational attainment on total fertility rate have been discussed in the previous studies.
Objective. The objectives in this study are to explore whether the difference in the timings of first childbirth among Hispanic women, Black women, and Non-Hispanic, Non-Black women exist in the United States and what factors affect the timings of first childbirth by means of exploiting the National Longitudinal Survey.
Design. The National Longitudinal Survey 1979 is used as the dataset in the study. 5276 female respondents, whose age is from 17 to 22 years old, participated in the survey. The Cox Proportional Hazard Model estimates the risk of first childbirth among various races by 2012. The effects of ethnicity and educational attainment on the timing of first birth are calculated using an interaction regression.
Results. The female employment exerts negative impacts on the timing of first childbirth. The extent to which the female employment prevents parenthood is different among various ethnicities. The risk of first childbirth in female minority with higher education is not different with that in Non-Hispanic/Non-Black women. The risk of first childbirth in single women is the lowest.

Key words: Childbirth, Ethnicity, Employment status, Educational attainment, The United States

EKHM52
Master thesis, second year (15 credits ECTS)
August 2016
Supervisor: Kirk Scott; Annika Elwert
Examiner: Martin Dribe
Word Count: 13000

Table of Contents

1. Introduction

In 1620, the Mayflower arrived in the new world, America, in which there was spacious land. It is not surprising that the 102 passengers were not aware that they would usher in a new era. Immigrants from a variety of countries and even continents had, for different reasons, reached the new world. Independent from the British regime the first immigrants established a culturally diverse country, the United States. In contrast with China, for example, in which diverse races had been, in general, assimilated and different cultures had been integrated, as can be seen in the unique Han culture, the diversity of American ethnic communities have been maintained (Sowell, 1981). It is not surprising that minority cultures and traditions are somewhat separated from the mainstream in the United States. In turn, the United States laid a great foundation for the research on different behaviors of ethnic groups, from the demographic perspective. One of popular indicators in the demographic study is total fertility rate.

However, the difference in the fertility rates among various ethnic groups is blur, prior to the 20th century, due to the shortage of the effect contraception. It is arduous to distinguish whether or not the parenthood was planned. The situation had been changed since supreme court approved that adults could legally use "the pill", which is one of the common forms contraception, in the 1970s. In comparison with the difference of the fertility rates among various ethnic groups, moreover, it is taken for granted that the age-specific marital and total fertility rates in the United States are consecutively stagnant at a low level from the 1960 onwards. According to Goldin and Katz (2002), this phenomenon has been attributed to diminishing restrictions on "the pill". Hotz, Klerman and Willis (1997) state that the low cost of the contraception facilitates the postponement of first childbearing, compared with contraceptive failure. Hotz's, et la (1997) statement consolidates that though raised awareness of postponement appeared prior to the appearance of "the pill," awareness has been fulfilled by "the pill". In this study, the priority is to research on the awareness of the timing of parenthood. The awareness among various races in the United States is able to cause different, and valid contraception is one of important prerequisites.

Although total fertility rate is determined by the timing of first childbirth and interval, this paper only focuses on the timing of first childbirth. Scholars have raised questions which are associated with the timing of first parenthood among various cohorts. When Andersson and Scott (2005)

conducted analyses, in terms of the patterns of childbearing among immigrants, they found that since the 1990s, the trends in childbearing of immigrant women are consistent with that of Swedish-born women. Andersson and Scott (2005) infer that immigrant women had adjusted the Swedish atmosphere regarding childbearing, and suggest that women somehow distributed their time between the non-market activities and labor market activities, based on the New Home Economics. In light of the aforementioned statement, the immigrant women adapted to the general climate of childbearing in their destination country. Notwithstanding, I wonder if the judgement in terms of Swedish demography is applicable in the United States. According to Hotz, Klerman and Willis (1997), though the timing of first parenthood was delayed among whites and non-white women in the United States in the 1980s onwards, the extent to which the timing of first birth is postponed toward the later age is more significant in white women. It seems interesting to research whether the timing of first childbirth is consistent among non-white women and white women in the United States. If U.S. non-white women, as immigrant women in Sweden, had adapted to the climate of childbearing, the timings of first parenthood among various races should be similar, in reference of the U.S. white women. As with what Andersson and Scott have done, I will analyze the relationship between women's employment and the timing of first birth. Based on the National Longitudinal Survey 1979, I hypothesize:

- Due to educational attainment, the timings of first childbirth among female minorities are not significantly different, in comparison to white women in the United States
- There is an adverse effect of female employment on first childbirth in the United States

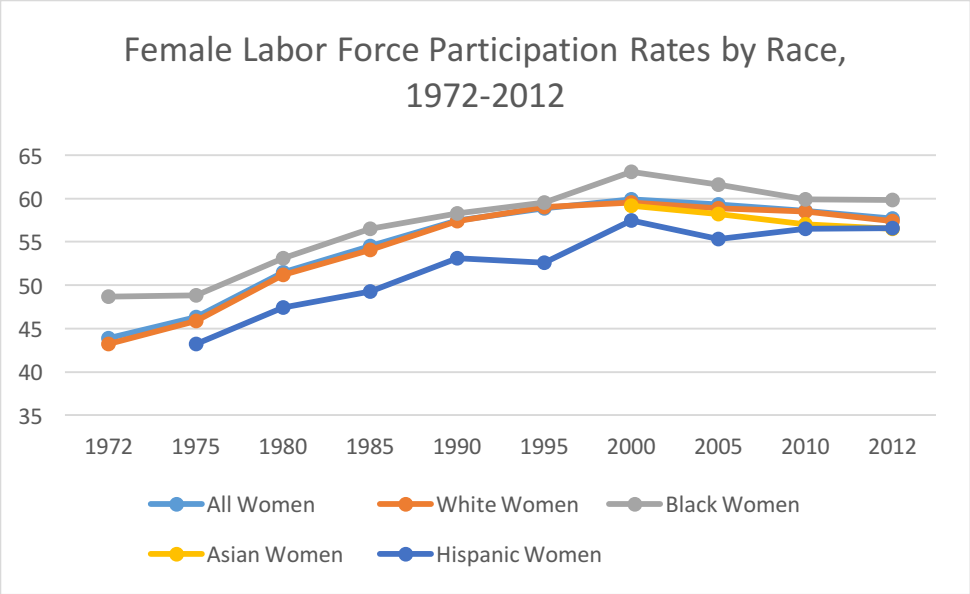


CHART 1: FEMALE LABOR FORCE PARTICIPATION RATES BY RACE, 1972-2012.
SOURCE: CURRENT POPULATION SURVEY, U.S. BUREAU OF LABOR STATISTICS

Before discussing Micro-economic characteristics, a general outline about Marco analyses in the United States is necessary. The macro-data, in terms of the female labor force participation rate, outlines an overall trend with respect to the female labor force in the United States with time. According to the World Bank, labor force participation rate is defined as the ratio between the population within working age and total population in the particular cohort. Based on the Current Population Survey in the United States, female labor force participation rates among various races have risen universally since the 1972. The trends of female labor force participation rates by races, to some extent, are consistent with the decrease in the fertility rate in the United States, though the female labor force participation rate is still relatively low. It is worthwhile to point out that labor force participation rates in white women and Asian women are almost equal in the whole time span. Based on the share of population among various races, it is reasonable that I will treat the white women as the reference, during researching on the labor force participation in the United States. Compared to white women, the female labor force participation rate in Black women generally is higher. One of interpretations is that Black women's spouse's earnings is unable to meet the household's cost, so that the Black women have to join in the labor market. In addition, the single Black women have much more incentive to participate in the labor force. On the contrary, the female labor force participation rate in Hispanic women is lower than white

counterparts. The culture plays an important role of the lower female labor force participation rate in the Hispanic group. When I conduct analysis on the effect of women's employment on the first parenthood, the trend in labor force participation rate by races should be taken into account. According to the Chart 1, the female labor force participation rates are saliently different among various races.

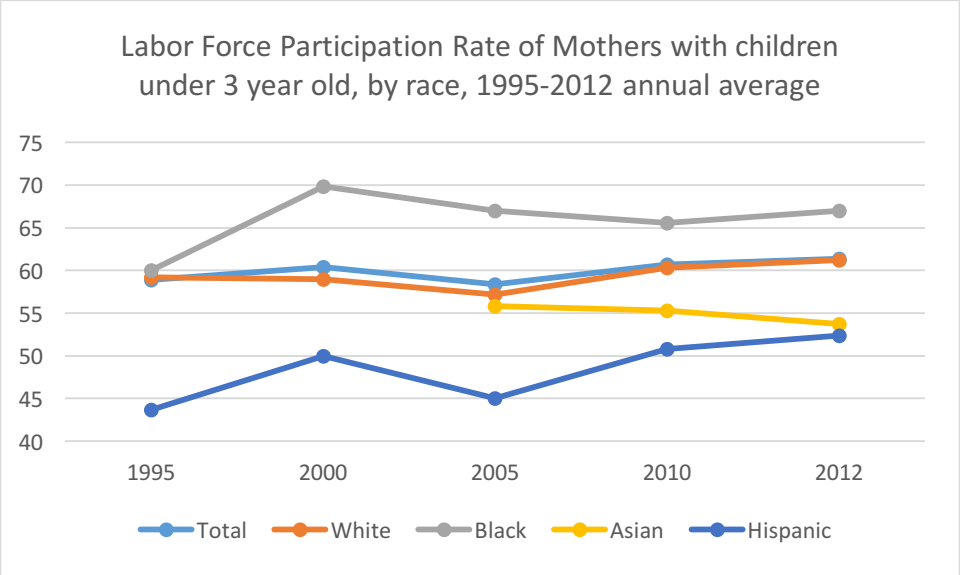


CHART 2: LABOR FORCE PARTICIPATION RATES OF MOTHERS WITH CHILDREN UNDER 3-YEAR-OLD, BY RACE 1995-2012

SOURCE: DATA FROM THE CURRENT POPULATION SURVEY, U.S. BUREAU OF LABOR STATISTICS, 1995-2010

After the interpretation of the female labor participation rate, I proceed to analyze the labor force participation rate of mothers with children under 3-year-old. All trends of the labor force participation rate of mothers with children under 3-year-old are paralleled from 1995 to 2012. All of the trends in the labor force participation rate of mothers with child under 3-year-old by races are flattened in the 2000s, compared with rising prior to 2000. Except from Hispanic mothers in 1995, all labor force participation rate of mothers with children under 3-year-old are higher than 60 percent. It is not so surprising that the labor force participation rate of mothers with children under 3-year-old have drastically increased since 1979. As Chart 1, the trend in white women is still intertwined with the total trend in Chart 2, so that the white women are still properly treated

as the reference category. In Chart 2, the lowest trend is indicated by the Hispanic women with children. Compared with other races, the trend in Hispanic women is interpreted by that Hispanic women relatively lack motivation to participate in the labor force since parenthood. The trends in other races are flat. When I put Chart 1 into account, the trends among mothers with children under 3-year-old are consistent with the trends of labor force participation rate in total women. Except from Black women, the labor force participation rate of mother with children is lower than total female labor force participation. It is worth to point out that in Hispanic women, the labor force participation rates of mothers with Children under 3-year-old is significantly lower than the female labor force participation rate. What I found based on Chart 1 and Chart 2, strengthens Mather's assessment. Compared with Chart 1 and Chart 2, therefore, the childbirth and sequent childrearing prevents women from labor force participation, especially the first childbirth.

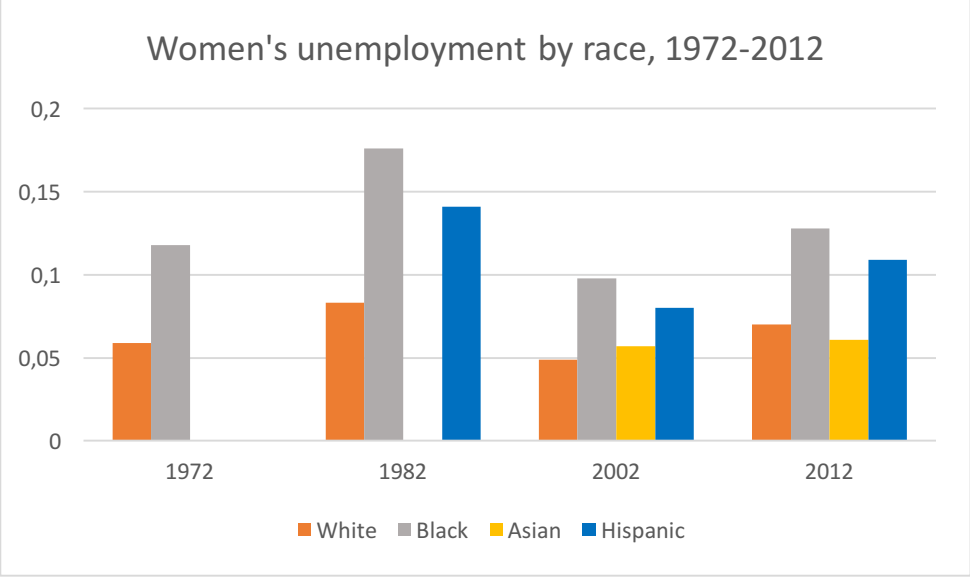


CHART 3: WOMEN'S UNEMPLOYMENT BY RACE, 1972-2012.
SOURCE BUREAU OF LABOR STATISTICS, CURRENT POPULATION SURVEY, 2012 ANNUAL AVERAGES

Finally, U.S. Bureau of Labor Statistics also offers Marco data with respect to women's unemployment by races (Chart 3). The unemployment is defined as the ratio between the number of individuals who are unemployed and the number of individuals who are in the labor force. In Chart 3, the data regarding the Asian women in 1972 and 1982 is omitted. Based on women's unemployment rate among various races, Black women have been most vulnerable to lose their

job since 1972. Compared with white counterparts, though the discrepancy in women's unemployment has shrunk, Black women's unemployment is still higher. Hispanic women also suffer from a relatively high risk of the unemployment, compared with white women in the United States. The Macro data sheds light on the fact that women's unemployment rates among various races are different. Moreover, Mather (2012) demonstrates the macro interpretation and argues that relative high unemployment among Black Americans and Hispanic Americans contributes the decline in fertility. I wonder whether there is causality relation between the different employment status among races and their maternal decisions.

In section 2, I summarize previous literatures and analyze what are crucial determinants to affect the timing of first parenthood. The theoretical model is outlined in the section 3. In this thesis, I primarily apply Kaplan-Meier estimates and Cox Proportional Hazard model. Prior to the empirical analysis, it is necessary to introduce the characteristics of the Cox PH model. The data description is located in the section 4. The National Longitudinal Survey Youth 1979 is the dataset in this thesis. Based on the theoretical model and the dataset, I engage in the empirical analyses and gain couple of results in the section 5. It is undeniable that the research still exists some limitations. I describe them in the section 6, in order to improve the study in the future. The summation and conclusion are located in the section 7.

2. Previous research

Fairfax (2014) states that there is an association between parenthood and marriage, and that we should take marriage into account whilst we conduct the analysis in terms of the first-time parenthood in various ethnicities. Nevertheless, the association between marriage and childbearing has been diminished, because of the prevalence of cohabitation. The cohabitation has been prevailing since 1960s. In some developed countries such as Sweden, people treat cohabitation much the same as marriage. However, it is undeniable that marriage is supposed to lay a strong foundation for parenthood.

In reference to Beckerian economic theory, beside marriage, the components which determine fertility, also consist of child costs, knowledge, uncertainty, and taste (Becker, 1960). According with Malthus's theories, Becker (1960) states that the economic framework enables us to generalize his theory and is better than other scales, when Becker had attempted to analyze an Indianapolis

survey. It is a brilliant assumption that children are considered as a durable consumption good, in order to meet parents' psychic requirements. We can apply the demand and supply model to estimate what parents decide concerning parenthood. Becker (1960) concludes that parents do not stop childbearing until the marginal cost of children outweighs their psychic or economic benefit. Notwithstanding, Becker (1960) also advocates that there is a divergence between the expected number of children and the real number of children, if the ability to produce children is taken into account. It is taken for granted that children are only produced by their parents, in comparison with other durable goods. It is also understandable that parents do not start the first childbirth until they want.

Based on the dynamic Swedish fertility model, the female wage exerts a robust influence on the timing of births (Heckman & Walker, 1990; Butz & Ward, 1979). Heckman and Walker (1990) also imply that the neoclassical model, in which wage is taken into account, has better application. In other words, the higher female wage strongly prompts the first birth, when time trends, age, and marital status have been controlled. Wrigley and Schofield (1981) also argue that the positive effects of real income on fertility are significant, when they research on the positive or negative influence among population, income and fertility.

Based on the Willis' model, Ermisch (1989) points out that compared with mothers' time, fathers rarely contribute their time and spirit to foster child quality and to engage into other related chores. Moreover, a lifetime budget constraint has been introduced in the static model. A lifetime budget constraint indicates that the total amount of time available to the mother is certain. Women tend to engage in much more non-market activities, at the expense of their labor force participation. Butz and Ward (1979) outline that maximizing household utility is associated with market goods, spare time for leisure and child service. Child rearing is more intensive than other chores in non-market activities. In order to pursue the maximum utility, women should consider the time allocation between market activities and child rearing. Nevertheless, I should bear in mind that the husband in the household usually plays a crucial role of financial supporters. Butz and Ward (1979) point out that the increase of male earnings positively affects the fertility rate given female earnings. Additionally, the increase in income induces both the quality and the quantity of children in household and Malthus insisted that the family size is in response to the household's income

(Becker, 1960). Although I primarily research the effects of women's wage on the first parenthood, the male earnings and household income should be taken into account as a control variable.

It is not so surprising that there is an association between the income and occupations. Williams (2012) states that the occupation inequality, to some extent, plays a role of the wage inequality in between-occupations, when fundamental labor market changes including the educational attainment and collective bargaining have been controlled. Based on what Williams (2012) argues, it is plausible that the spouse's occupation can be treated as a proxy in the study. The between-occupation inequality indicates the differential in spouse's income. Moreover, the rising in between-occupation inequality can be associated with the growing between-class inequality (Williams. 2012). The statement enables a consolidation of the assumption that the impact of a women's spouse's occupation has on the timing of her first childbirth, by means of income and class of work.

In terms of the female occupation, Budig and England (2001) point out the outdated knowledge is supposed to be depreciated and the intermittence in the female employment experience prevented women from gaining the benefits from the on-the-job training. Gustafsson (2003) states that the cost of childbearing and sequential child rearing consists of the diminishing direct forgone wages and human capital loss. The occurrence of the postponement of first birth is prevailing, in case of the depreciation of human capital due to non-use and the time spent out of the labor force (Gustafsson, 2003). On the basis of pre-parental human capital, women who are endowed with market specific human capital tend to delay childbearing, in order to prevent a depreciation of income. Furthermore, the life-cycle model of fertility also illustrates the negative relationship between the cumulative utility that mothers enjoy the leisure time and the marginal utility of childrearing. In reference to the life-cycle models, the benefit derived from the postponement of childbearing is involved in increasing leisure time, increasing future wage due to the labor force participation, and reducing the maternal time inputs because of childrearing (Hotz, Klerman & Willis, 1997, Gustafsson, 2003). Ermisch (1989) also states that the adverse effects of childbearing on the women's occupations are obvious, so that the employment status should be considered in the study. Women who join in the labor market, in turn, are prone to delay the parenthood, on the basis of the life-cycle model.

Not only the labor force participation, also education attainment exert influence on the timing of parenthood. Gustafsson (2003) illustrates that the women with higher education were inclined to postpone the first birth compared with lower educated counterparts. Higher education enables a postponement of the timing of first birth by means of women's initial salary and the growth in the salary. Ermisch (1989) claims that the high level of a wife's initial income is associated with the lower completed fertility. The initial income is a pivotal indicator to predict the future income. The higher initial income hints that the cumulative income is higher and income disparity of income will widen with time elapsed. Women with higher initial income insist that the utility of cumulative labor force participation outweighs the utility of childbearing. Moreover, from the perspective of salary growth, Butz and Ward (1979) mention that the increase in women's wage refers to the rise of the cost of child rearing, by virtue of the lifetime budget constraint. The opportunity cost stemmed from parenthood is difficult to diminish. Although the purchased child care is a plausible way to address the rising of cost of parenthood (Ermisch, 1989), the purchased child care is not ubiquitous in the United States. On basis of the price and opportunity cost of parenthood, referred to the depreciation of human capital, women prefer childbearing either instantaneously after marriage or at the limit of the fecund period (Gustafsson, 2003). Due to higher price and opportunity cost of childrearing, it is possible that women with higher education tend to postpone the timing of first childbirth until the end of the fecund period. Mather (2012) agrees with this statement and moves on arguing that the high fertility rate among African American and Hispanic American is associated with less education and lower incomes.

In terms of different races, Sowell (1981) argues that the fertility rate is consistent with the household earnings and the variance of fertility rates among different races is obvious. It makes sense that the ethnic group with lowest household income contributes the highest fertility rates, whereas the marital behaviors in the Blacks and Hispanic households with good education do not significantly differs from white and Asian counterparts. Race still plays an important role of effect on the timing of first childbirth because of different cultures and customs. Hispanic women's attitude toward the fertility is more pronatal, compared with other races (Bogue, 2010). The differential in timing of first childbirth among various races can be attributed as different attitudes

toward Natalism. Other socioeconomic characteristics of different ethnic groups exerts influence on the timing of first time childbearing.

The timing of first childbirth is also tightly associated with characteristics of immigrants and native American in the United States. According to Caponi and Plesca (2014), especially in middle and higher educated women, the total fertility rate among female immigrants is higher. I assume that the timing of first childbirth between immigrants and native Americans is different. In addition, Bouvier and Gardner (1986) assert that the Hispanic immigrants are dominant in the low-income industries. Moreover, in reference to the 1980 census, Hispanic American account for 77 percent of undocumented immigrants (Bouvier & Gardner, 1986). It is possible that the large share of undocumented Hispanic immigrants, who lack well education results in that share of Hispanic immigrants as highest in the low-income industries. The ratios between immigrants and native American are not proportional among various races. The share of Hispanics and Asians is over 80 percent of legal immigrants (Bouvier & Gardner, 1986), even though the share of White and Black Americans are dominant in the U.S. population. It is necessary to control for immigration in the study, in order to get rid of the intervention.

The association between the fertility rate and age should be kept in mind (Sowell, 1981). Age at menarche among various ethnicities in the United States is also to be taken into account. Chumlea, Schubert, Roche, Kulin, Himes and Sun (2003) state that based on the research from 1988 to 1994, 90 percent of girls in the United States by 13.75-year-old and the median age is 12.43 years. Moreover, the age at menarche in 1988 was almost equal to the age at menarche in 1973 (Chumlea et al., 2003). It is applicable that I focus on the women whose age range is from 14 years to 22 years. In light of age at menarche, it is not necessary to research on the girls whose age is under 14-year-old. In addition, Chumlea and his colleagues (2003) also argue that in reference to the Non-Hispanic/Non-Black women, the age at menarche in the Black American is earlier and the age at menarche in the Hispanic women is roughly equal. It sheds light on the fact that from biological perspective, the timing of first childbirth in the Black women is possibly earlier than that in other counterparts, without other interventions. It is also taken into account that the biological explanation cannot strongly support the differential in timing of first childbirth between Hispanic women and Non-Hispanic/Non-Black women in the United States.

Fertility is also positively associated with the degree of altruism in the household (Becker & Barro, 1986). It is a common hypothesis that if parents are altruistic toward their children they are prone to descend and disperse their gene. The altruism theory is a complement of the life-cycle model. In contrast with Easterlin's (1973) theory that fertility is negatively associated with the wealth of parents, Becker and Barro (1986) state that fertility insofar is positively relevant to the growth of wealth across generations, if we take the altruism rather than a lifetime budget constraint into consideration. However, the accurate degree of altruism among various races is subject to the viable measure because it is intangible. In this study, I place the altruistic behavior as one of cultures into races. For example, Asian culture advocates the accumulation of fortune and transmission to the offspring. It is consistent with Beck and Barro's (1986) theory that Asian culture also encourages the higher fertility rate and early parenthood. It is possible that different extent of altruism among various races causes the timings of first childbirth.

Compared with micro-economic traits, macroeconomic factors carry influence on the individuals' behaviors in demography. The geographic information is one of the important macroeconomic factors. Olopha and Aladeniyic (2015) argue that an individual's address of residence is tightly correlates with the contribution of the demographic variables such as the fertility intentions of women. Differences in age at birth of first child are strongly associated with the rural/urban fertility differentials (Olopha & Aladeniyic, 2015). The relationship between the age at birth of first child and the rural/urban fertility differentials is determined by largely the opportunity and accessibility of high education. Women who dwell in the urban areas have greater opportunities to attain education and reduce their proclivity to childbirth. Tarver (1969) designed a model to measure the influence of rural-urban differences among women. The increase in fertility of women is positively correlated to the distance from a metropolitan downtown (Tarver, 1969). Not only is there an educational factor, but employment influences the fertility of women. The geographic factors by means of education and employment affect the timing of first childbirth in the United States.

In contrast with the Nordic countries, the U.S. government did not implement the same level of parenthood-friendly policies. When we research on the parenthood behaviors of an ethnic minority, our results are supposed to match predictions of Beckerian economic theory, in contrast with what

Andersson and Scott had noted in the Swedish context. According to Andersson and Scott (2005), compared to Germany, the Swedish government have enacted several policies and encouraged women's labor force participation. Nevertheless, the institution in the United States tends to cater to the affluent (Hummer and Hamilton, 2010). It makes sense that individual awareness results in the difference of the timing of first-time parenthood among different races in the United States, regardless of the parenthood-friendly policy.

3. Methodology

3.1. Fertility measures

According to Andersson and Scott (2005), the relative risk of the first childbirth varies for different ethnicities of childless women in the United States. Based on variables supplied from the NLSY79 dataset, the relative risk of the first childbirth among different ethnicities of childless women with various characteristics should be taken into account. The characteristics consist of age of the woman, labor market activity, spousal characteristics and macro-economic characteristics. Indirect standardization plays important roles for the estimation (Andersson & Scott, 2005). The proportional hazard model, by which time dependent event sequences have been analyzed, is applied. By means of the proportional hazard model, standardizing an array of variables enables one to estimate the relative risk of first birth among childless women since menarche.

I calculate descriptive statistics, by ethnicity, educational attainment, employment status and women's income. After the test for statistical significance, all of the descriptive statistics are significant, due to the large sample size.

The hazard of first parenthood was estimated by year. Stata, version 14 is the statistical tool used in order to analyze the survey. The survival analysis and the Cox PH model are also applied in this study. The Cox regression is selected mainly due to the characteristics of the Cox regressions to estimate continuous-time event history data and to cover time-varying covariates (Marital Status and Employment Status).

3.2. Survival Analysis

Prior to thoroughly discussing the survival model, several outlines in the survival model should be taken into account. Kaplan-Meier estimator is equipped with the survival function. Dynamic

selection is supposed to be considered in this case. Due to the fact that women with higher educational attainment are prone to postpone their parenthood, for example, women who take a longer time to in education have less opportunity to have early childbirth.

3.2.1. Kaplan-Meier estimator

According to Assburg (2012), the Kaplan-Meier estimator, which is the non-parametric models, is a graphical form in order to depict the time-to-event data after censoring. The estimator for S(t) has been evaluated by means of the maximum likelihood estimation for h_i. After logarithm, the Kaplan-Meier estimator is defined as:

$$\hat{S}(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i})$$

$$n_i = n_{i+1} + c_i + d_i$$

Where n_i is the number in risk set at t_i, $d_i$denoted the number of events at t_i. In addition, $c_i$stands for the number of observation censored from t_i to t_{i+1}. *The trend of Kaplan-Meier estimator usually is depicted by graphic method.*

3.2.2. Left-truncated and Right Censored Observation

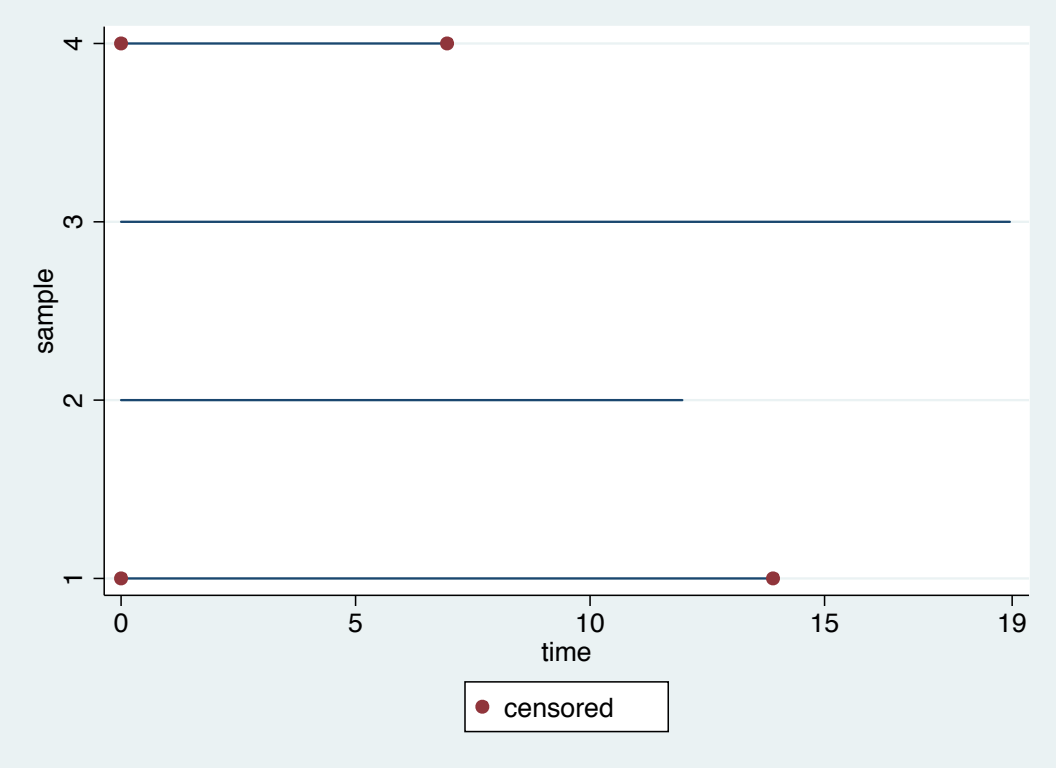


CHART 4: SAMPLES TO EXPLAIN THE LEFT-TRUNCATED AND RIGHT CENSORING IN KAPLAN-MEIER ESTIMATOR

After the analysis of Kaplan-Meier estimators, it is worthwhile to point out that there are four types of possible censoring (Klein & Moeschberger, 2003). Right-censoring is the most important censoring, which should be considered in the survival function. Right-censoring literally means that the subject leaves the survey prior to occurrences. In other word, the event does not occur, when the survey is accomplished, and the true occurrences cannot be observed given the censoring time. However, it is inappropriate to get rid of the observations in the right-censoring directly. Directly ignoring right censoring observations is associated with bias towards short spells. In the chart 4, Sample 2 and Sample 3 respectively represent right censoring. In Sample 2, the respondent had exited the survey. On the contrary, in Sample 3 the particular event does not occur until the survey finished. In comparison with Sample 2 and Sample 3, Both Sample 1 and Sample 4 indicate that the events respectively are occurring in 14 units and 7 units since they entered the survey.

The left truncation, moreover, should be considered in this research. It means that the subject had experienced a particular event prior to the survey. It is meaningless to take the observations into account, whereas left-truncation causes the underestimation in the research. For example, several female respondents had experienced parenthood prior to the survey. I have to remove the female respondents who had experienced childbirth prior to the survey.

3.2.3. Survivor function:

In reference to Cameron and Trivedi (2015), the time span in which a subject survives in the initial status is associated with duration, and then the cumulative distribution function of the duration, T, is illustrated as:

$$F(t) = P(T \leq t)$$

Given the cumulative distribution function (CDF), the survivor function indicates the probability of surviving with respect to duration is listed as

$$S(t) = p\ (T > t) = 1 - F(t)$$

In the empirical analyses, the survivor function should calculate the fraction of workers who have maintained the situation after t periods:

$$\hat{S}(t) = \frac{1}{N}\sum_{i=1}^{N} I(t_i > t)$$

Where $t_i$is the duration of individual I =1, …, N and I(.) is the indicator function.

3.3. Cox Proportional Hazard model

Compared with the Kaplan-Meier estimator, the Cox PH model is a Semi-parametric model. According to Cameron and Trivedi (2015), the characteristics of semi-parametric models consist of a certain distribution function, and no parametric form of the survivor function. It is worthwhile to mention that incorrect results are possibly gained due to the wrong assumption, when the parametric model has been selected.

Based on the survivor function, Hazard function is defined as:

$$h(t) = \lim_{\delta t \to 0}\left(\frac{\Pr(t \le T < t + \delta t | T \ge t)}{\delta t}\right) = \frac{F(t+\delta t) - F(t)}{\delta t} * \frac{1}{1 - F(t)} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

Where S(t) denotes first derivative of the survivor function, and δt stands for the differential of the density function of T. Prior to research on the Cox PH model, it is necessary to illustrate that the hazard function enables to stand for the survivor function. The hazard model can be transformed as:

$$h(t) = -\frac{lnS(t)}{dt}$$

The value of the cumulative distribution function at t=0 is zero, and then the cumulative distribution function can be rephrased as:

$$F(t) = 1 - \exp\left[-\int_0^t h(s)ds\right], t \ge 0$$

Finally, given the CDF of T, in which the hazard function, h(t), has been considered, the density function of T, f(t) as:

$$f(t) = h(t) * \exp\left[-\int_0^t h(s)ds\right]$$

It is applicable that the hazard function is able to account for the probability in the survivor function.

Cox (1972) stated the proportional hazard model defined below:

$$\ln h(t) = h_0(t) + b_1 X_1 + b_2 X_2(t)$$

where h(t) stands for hazard rate, $h_0(t)$ denotes the baseline hazard function, $X_1$are time-constant variables and $X_2(t)$ are time-varying covariates.

In the Cox PH model, the distribution form (baseline hazard) with respect to the duration is not specified. In order words, the estimation concerning $h_0(t)$ is not necessary. Due to features of the Cox PH model, the variation in the hazard function is hypothetically associated with the covariates multiplicatively shift. Cox (1992) states that partial maximum likelihood estimator for the coefficients is not associated with the baseline hazard estimation and hazard model assumption. As the intercept in the ordinary regression, the baseline hazard function is an analogy of the reference group in the Cox PH model. The interpretation of the baseline hazard function is the probability of an event occurred controlling other explanatory variables. As previously mentioned, the baseline hazard makes no assumption of shape over time because it contains no parameterization. I should bear in mind that the assumption is that the characteristics of all individuals are similar.

Based on the initial Cox PH model, the exponential function, which called as the hazard proportional function, is defined as (Jiang, Li & Sanchez-Barricarte, 2015):

$$h_i(t) = h_0(t)\exp\left(\beta_1 x_{i1} + \beta_2 x_{i2} \ldots + \beta_k x_{ik}\right)$$

where $h_i(t)$ is defined as the hazard rate, $h_0(t)$ denotes the baseline hazard function. x_{ik} are covariates influencing event occurrence. β are regression coefficients to be determined in the model.

After dividing by $h_0(t)$ in both sides and the logarithm of the results, the final Cox regression model was obtained:

$$log\frac{h_i(t)}{h_0(t)} = \beta_1 x_{i1} + \beta_2 x_{i2} \ldots + \beta_k x_{ik}$$

In this study, I primarily research on the impact of ethnicity (Sowell, 1981), educational attainment (Gustafsson, 2003 and employment status (Butz and Ward, 1979; Gustafsson, 2003) on the timing of women's first parenthood. Meanwhile, several of women's characteristics such as marital status (Fairfax, 2014), spousal occupation (Williams, 2012; Butz & Ward, 1979), geographical information (Olopha & Aladeniyic, 2015), Women's income (Ermisch, 1989; Wrigley & Schofield, 1981; Heckman & Walker, 1990; Butz & Ward, 1979), and citizenship (Bouvier & Gardner, 1986; Caponi & Plesca, 2014) should be taken into account.

The hazard ratio is equal to the exponent of the coefficients in the hazard function. The hazard ratio is indicated as an odds ratio. For example: The dummy variable, educational attainment, is equal to zero when the subject does not attend in the college while the hazard ratio for the variable, educational attainment, is the relative odds of respondents attended the college in comparison with respondents without the college. If the value in the hazard ratio for this variable is 1.2, then it indicates that the odds of the sample with college is 20% higher compared with the sample without college. To summarize, the interpretation of hazard ratios is with respect to the reference category.

3.3.1. The proportional hazards assumption

Prior to the interpretation of the Cox PH model, the proportional hazards assumption should be empathized in this section. There is an assumption with respect to the Cox PH model that the proportional hazards exist in a scenario where there is a stable relationship between the dependent variable and the explanatory variables. Due to the assumption that the constant relationship

between the dependent variable and the explanatory variables, any individuals at any point in time are proportional. This means that it is necessary to test the assumption prior to the interpretation.

Commonly, the tests consist of graphical evidence and Schoenfeld residuals in order to test the assumption. If the results fail to meet the assumption, there is a heterogeneity in the model. In other word, the characteristics among various sub-groups are different over time. Of course, it is also soundly within reason that the variables vary with respect to time. The possible solutions include stratification or separate models. However, in this study, neither of aforementioned solutions is able to sort the issue out.

4. Data Description in the NLSY79

The NLSY79 consisted of young men and women who were 14-22 years old, since the first survey in 1979. The NLSY79 was supported by the U.S. Department of Health and Human Services, the National Institute on Aging, the National Institute on Alcohol Abuse and Alcoholism, the U.S. Department of Defense and the Armed Services and the National Institute of Education. NLSY79 imitated studies of both young men and young women cohorts in the 1960s, in order to evaluate the change of employment and training programs since the 1977 amendments to the Comprehensive Employment and Training Act (Zagorsky, 1999). It is impressive that the survey had lasted over 2 decades, so that respondents experienced the transition from the school to the labor force market. The surveys have collected information in terms of labor market experiences of various groups in different time points. We should bear in mind that the primary purpose of the NLSY79 is to research on the respondent's labor force attachment and investments in education and training. Prior to 1994, the data gathered is updated yearly. Due to the chronological survey, researchers are able to thoroughly analyze to the American men and women who were born between the 1950s or 1960s.

The Sample of the NLSY79 is both young women and men living in the United States and born from January,1st, 1957 to December, 31st, 1964. The overall sample of the NLSY79 consists of the cross-sectional sample (6111), the supplemental sample (5295) and the sample sourced to the military (1280) (Zagorsky, 1999). The sources of the sample are listed below: The cross-sectional sample is designed to include the sample whose age is from 14 to 21 in December 31, 1978. In the other word, the cross-sectional sample includes respondents who represent the population living

in the United States since the first round of the survey (Zagorsky, 1999). The supplemental sample is involved in the Hispanic, Black, and economically disadvantaged non-Black/non-Hispanic youths living in the United States during 1979. On the basis of the design in the NLSY79, the dataset is a flow sample, which includes individuals who start their spell in a particular time interval (1979-2012). The purpose of supplemental sample is to supply a wider Hispanic and Black sample living in the United States. Of course, respondents' age is also between 14 and 21. Respondents are sourced from the military to represent the population born from January 1st, 1957 to December 31st, 1961 and enrolled in the military during the survey. We should bear in mind that most of respondents in the cross-sectional sample had gone through the survey in NLSY79. Nevertheless, 1079 respondents in the military sub-sample were no longer able to participate the survey starting in 1985, by virtue of cost limitations (Zagorsky, 1999). As compensation, there are 201 respondents who had been randomly selected and continued the survey in the NLSY79. Furthermore, administrators in the survey had made a decision in order to remove the 1634 respondents who were economically disadvantaged, Non-Black/Non-Hispanic subsamples. In summation, the NLSY79 dataset consists of the cross-sectional sample following 1979, in which the number of female respondents is 3289, and the supplement sample, in which the number of female respondents is 1096. Prior to moving on next section, I also should point out that the NLSY79 survey consists of two phases, when researchers conducted on the eligible sample selection. In the first phase, based on the several characteristics such as ethnicity, income, region and others, researchers selected the eligible sample by means of stratified multi-stage. In the second phase, the stratified sample was randomly selected. Overall, there are 8984 respondents in the sample, when researchers removed 38 out of 9022 respondents, which were identified as duplicated data.

The original purpose for the NLSY79 was to interview both civilian and militant youths during 1979-2012. Notwithstanding, the sample size in the NLSY79 has been reduced to 11607 due to the fact that the interview in the full military sample was ceased since 1985. In the beginning of 1991, researchers stopped gathering the 1643 respondents who are the economically disadvantaged, Non-Black/Non-Hispanic supplemental sample. Until 1989, paper-and-pencil interviewing was predominant in the NLSY79 survey. Computer assisted personal interview (CAPI) have been implemented in order to eradicate problems stemming from manual mistakes.

According to Zagorsky (1999), the retention rates throughout the first 16 rounds of the survey maintain approximate 90 percent. The deceased respondents during survey are included in the survey respondents, whereas the respondents who were permanently dropped out are excluded from the NLSY79 dataset. Common reasons for non-interview consist of an interview being refused by the respondent or by the respondent's parent, the family address not being successfully pinpointed or the respondent being reported as deceased. In addition, the National Opinion Research Center (referred to NORC) is reluctant to review particular cases in which they determined the high difficulty to interview and the funding had been slashed. Finally, the NLSY79 was continued in 2012. Though the NLSY79 dataset consists of hundreds of thousands of variables, this study primarily focuses on demographic variable included gender, age, the timing of first childbirth, ethnicity, and socioeconomic characteristics such as educational attainment, employment status, marital status, Income, spousal occupation, SMSA, and citizenship. Specific introductions concerning different variables are listed in following section.

Gender

In this study, I primarily concern female respondents, rather than male counterparts, for the sake of the research on first parenthood in the United States. Nevertheless, the observations concerning the gender are subject to a small degree of error, when they were recorded during the annual survey. Though NORC had already been aware of this error and verified the sample size of error is not significant, I still note that there are 42 errors occurred in variables named as "Sex of R" during entire survey.

Age Group

In order to understand the effects of age, I select the variable named as "age of R in the 1979" in this thesis. The respondents' age in 1979 is from 14-yeasr-old to 22-years-old. It is kept in mind that a number of factors contributed to inconsistencies in the NLSY79, due to the multiple survey points. In terms of the multiple survey points, age had been inquired at different time points. In other words, it is possible that different respondents would had different age recordings, even though their birthdays are identical. The other latent issue associated with Age Group is that the determination of age is subject to the interviewers' judgment. In the survey, there are not systemic criteria for reckoning the age, so that a deviation among various respondents who were born in

same day can possibly occur. When I checked Age, there were 5276 valid respondents in the sample The age distribution meets the normal distribution, in which the numbers of both youngest and oldest respondents are few in comparison with the number around the median age.

Ethnicity

In the NLSY79, the variable, ethnicity, includes "Hispanic", "Black", and "Non-black/non-Hispanic". I should bear in mind that the ethnicity is subject to the self-awareness. The self-awareness is subjective perception, in contrast with the objective traits such as the skin color. Based on the prerequisite, "Hispanic" is assigned to respondents who self-identified as Hispanic. "Black" included respondents whose race is coded as black. In addition, "Non-black/non-Hispanic" covers those whose ethnicity was illustrated by "white" or "other". 3618 out of 5276 respondents have experienced childbirth in the research interval. *The share of Hispanic women is 18 percent of the total parenthood in the survey. 59 percent of childbirths occurred in the Non-Hispanic, Non Black Women, in comparison with 23 percent of childbirths in Black women in the research period.*

Educational attainment

In terms of the education attainment, I select the variable named as "highest grade completed since last interview" (HGC). The variable is available since the 1979 survey, so that I am able to trace back the respondents' educational attainment. I should be aware of inconsistencies, in terms of the educational attainment. The most obvious inconsistency includes the highest grade completed reducing over time and inconsistency between the high school diplomat and highest grade completed (Zagorsky, 1999). Even after a clean-up in 1994, the inconsistencies had not been removed completely. I set grade 12 as a threshold, when I redesign the variable, educational attainment, based on HGC. Where respondents declared that they had accomplished the grade 13 onwards, the variable, education attainment is recorded as college, in comparison with below college. In the end of the survey, 88 percent of respondents completed high school.

Marital Status

I applied the variable named marital status in this case. Although remarried and reunited are included in the variable, respondents who are remarried or reunited are supposed to be assigned as

married in the NLSY79 (Zagorsky, 1999). In terms of cohabitation, information was only available from 1990. It includes the partners regardless of gender. The collection of information on partnership has varied over time in the NLSY79. Sometimes, the questionnaires were illustrated as fertility or childcare in the section compared with "marriage." In the end of the NLSY79 survey, the share of married women was 64 percent, in comparison with never married women, which was 32 percent of those who had children. The separated women and divorced women respectively compose 2 percent of first childbirths in the NLSY79 survey.

Employment status

In the reference to the employment status recode (1979-1998), I am able to conduct this on research associated with the labor force participation. The variable named as the employment status recode was consistent with Current Population Survey (CPS). Moreover, it had recalculated the week in activity and hours worked. This variable also accounts for whether working for pay, whether looking for work, what the job seeker is doing, and whether and why respondents are voluntarily absent from jobs. The sub-group named in the active forces implies the respondents currently serve in the military. Due to the CPS Changes in 1994, the revise in the NLSY79 also had been undertaken, in order to be consistent with the CPS. First, U.S. labor force classifications had been changed. Second, the way to ask questions has been changed. Third, Advanced computer technology had been introduced. Nevertheless, the NLSY79 does not include the CPS information, when respondents were on active military duty. If the respondents were under 15-year-old in early survey, the CPS questions were not inquired. Based on the data in NLSY79, the share of the employed group initially increases and then diminishes with time. It matches with the change in the share of the sub-group named "out of employment" with time. I assume that women were voluntarily taken out of labor force participation until 1998. 53 percent of the childbirths in this survey are to employed women. The share of the unemployed women is 8 percent of childbirths. It should to be pointed that the share of women out of labor force is 38 percent of total childbirths in the survey.

Geographic residence

I apply the information concerning whether the current residence is in the Standard Metropolitan Statistical Area (MSA), in order to analyze geographic residence. On the basis of zip code, state,

and county, the location of the respondent is supposed to be assigned as within or outside of a MSA (Zagorsky, 1999). Based on the definition, the observations have been assigned into different subgroups, such as “not in SMSA”, “SMSA, not central city”, “SMSA, central city not know”, and “SMSA in central city”. The variable in this case is simplified as Unknown, SMSA, and Non in SMSA. The distribution of the observations in SMSA are consistent with the distribution of the observations in urban or rural areas. In sum, 26 percent of births in the survey occurred in rural areas, compared with 43 percent of childbirths occurring in the SMSA. The share of Unknown geographic information is 30 percent of birth in the survey.

Spouse’s Occupation

To carry out the effects of women’s spouses first parenthood, spouses’ occupation is one of applicable variables. The occupation of women’s spouse mainly describes the working situation in last year. The occupation has 417 separate categories in total. In the NLSY79 dataset the occupation has been reassigned into 12 major groups. All occupations were in respect of the 1970 codes prior to 1982. Since the 1982 interview, the occupation of women’s spouse was defined by the 1980 codes. Based on the characteristics of occupations, I have reassigned the occupations into four different groups: white collar, physical work, military service, and out of employment. In turn, 28 percent of childbirths occurred while the women’s spouse engage in white collar jobs, compared with 27 percent in physical work. The share of total childbirth in the army and out of labor force are respectively equal to 4 percent and 1 percent. In 40 percent of birth occurrences, the survey cannot indicate what is women’s spouse’s occupation.

Income

The income questions in NLSY79 refer to the previous calendar year. From 1979 to 1982, if respondents met one of criteria such as attaining 18 years of age, having a child, enrolling in college, marrying or living outside their parents’ home, they should answer some simple questions in terms of income. It should be pointed that women’s income in NLSY79 is not constant and suffers from the inflation. In order of the deflation with economic series, Consumer Price Index, CPI, as deflator has been introduced. CPI is offered by U.S. Bureau of Labor statistics. The base period is from 1982 to 1984, so that the average CPI in 1982-84 is supposed to be equal to 100. Index includes consumer expenditure such as food and beverages, housing, apparel, transportation,

medical care, recreation, education and communications, and other goods and services. However, CPI includes the expenditure by urban wage earners in 87 urban areas across the United States, so that the estimated CPI is deviant from real consumer income. In turn, 45 percent of first childbirths occur in the income group, in which women's salary is from 0 to 100 based on 1982-1984 CPI. The share of the first childbirth in the income group in which the women's salary is from 101 to 200 and from 201 to 300 is separately 21 percent and 6 percent in comparison with approximate 2 percent of birth in women's salary above 300 in CPI. The share of births among women whose salary is none, is 14 percent. Finally, 12 percent of births occurred when women's salary was unknown.

Citizenship

For foreign-born respondents, NLSY79 database covers information in terms of citizenship in 1984. Citizenship stands for time-invariant variable. As supplement of the sample race, the variable Citizenship consists of Yes, No, and Unknown. In comparison with non-citizen and Unknown, citizen is predominant in the NLSY79 database. 86 percent of childbirths occurred to American citizens, compared with 6 percent of childbirths to immigrants. 8 percent of births were among women whose identity cannot be verified.

5. Empirical analysis

Failure is defined when respondents attain the particular threshold in the Stata. In this case, the failure stands for respondents' first childbirth. Based on the NLSY79 database (Appendix 4), there are 5276 subjects who meet prerequisites such as the female and no experience of parenthood prior to the study. The number of respondents who experienced childbirth during the survey is 3618 out of 5276. Since 1979, the total number of the observations in the study is 81989. It is taken into account that the observations after the first parenthood are invalid. The number of valid observations is equal to 65146. When the right-censored observations are taken into account, the observations that respondents who do not experience the failure in the end of the research, should also be considered. In reference to Appendix 4, the time span at risk is 34 years. The respondents who had left the survey prior to the end due to the decease or other reasons, are still considered. So the total observations depict the characteristics of respondents who experienced the first childbirth and who were without a childbirth at the end of survey. The total amount of the observations in this research is 65146.

In reference to Appendix 4, moreover, the number of subjects and records respectively has been listed. The distribution of the time of exit also is also provided. The average time of birth is 15 years. In other word, respondents have experienced their first childbirth after 15 years since they participated the survey. The median in the time of birth is 10 years. The comparison between the average time of birth and the median in the time of birth indicates that the distribution of the first childbirth has positive skewness. The distribution in the timing of birth in the beginning of survey is denser. The time at risk also indicates that the total number of the observations is 82363. The value in the time at risk is consistent with the value in exit time. In Appendix 4, the total failure indicates that there are 3618 respondents who had experienced first childbirth during the survey.

Appendix 5 primarily indicates that time-constant variables and time-varied variables change over time. In the Cox PH model, the variables such as ethnicity, educational attainment, and employment status are drawn attention, when the marital status, geographical information, spouse's occupation women's salary, and citizenship have been controlled. In the Stata, the attribution of all variables concerning the NLSY79 is listed in Appendix 5. Due to the fact that the socioeconomic characteristics of subjects have been varied over survey, the value in particular variable has changed. It should be pointed out that race and citizenship in 1984 and age group in 1979 are constant. In the variable, Educational attainment, 3793 out of 5276 are constant whereas 1483 out of total subjects varied. In the variable, employment status, there is 1411 constant, compared with 112 always missing observation and 1394 sometimes missing observations. The variable, marital status, consists of 2180 constant subjects, 133 subjects who are always missing in the survey and 1398 subjects who are sometimes missing. In the variable, spouse's occupation, 2157 subjects are constant, in contrast with 3119 varied subjects. In terms of the variable, spouse's occupation, the omitted observations have been assigned into an artificial group named as Unknown. Regarding the geographical information, the number of subjects which is constant is 2036, in comparison with 3240 subjects varied observations. The variable, women's salary, consists of 959 constant subjects and 4317 varied subjects. I also reassign the omitted observations into a new subgroup named Unknown. Based on the NLSY79 dataset, the life table has been established in Appendix 6. In Appendix 6, I am able to trace back the number of failed subjects in

every interval and calculate the survival function. On the basis of life table, Kaplan-Meier estimator enables to be drawn.

5.1. Result and Discussion

After the descriptive statistic, univariate analysis enables this paper to supply much more information. Using Stata, the log-rank test of equality is applied, in order to exam the categorical variable (Bruin, 2006). Bruin (2006) recommends that the p-value of 0.2-0.25 or less indicates the significant predictor. In other word, if the p-value of the predictor is higher than 0.25 in the univariate analysis, it is possible that the predictor is intervened by other sub-groups. According to Bruin (2006), the log-rank test enables an examination of the equality of the survivor function among the different groups.

I carry out the univariate analyses by Kaplan-Meier survival estimates in order to test the categorical variables which are the interested variables including races and education. In Appendix 7, the Kaplan-Meier estimates indicate the races and education respectively. The Kaplan-Meier estimated curves comparing Non-Hispanic, Non-Black women, with female minority group (Appendix 7, A) implies that the risk of first childbirth in Non-Hispanic, Non-Black women is constantly lower than that in female minority group over time. On the contrary, the Kaplan-Meier estimated curves for the risk of first childbirth comparing Hispanic women and Black women are overlapped in the fourth year and intertwined (Appendix 7, B). Based on the Kaplan-Meier estimates, it should be born in mind that the difference in timing of first childbirth among U.S. minorities is not significant. I am prone to ignore this issue and move on, however, when I take the research aim into account, which primarily is to distinguish the risk of first childbirth between Non-Black, Non-Hispanic women and female minority. The Kaplan-Meier estimated curves for the risk of first childbirth comparing women who attained college, versus women without college degree (Appendix 7, C), implies that the risk of first childbirth in women who attained college is lower than that in women without college degree over time. In addition, the robustness of the measure is checked. The test of proportional hazard assumption also has been applied in the Cox Proportional Hazard model, in order to deeply analyze the internal relationship among different subgroups. The result from the proportional hazard model is illustrated in Appendix 8.

I proceed to the interpret the hazard ratio of the Cox Proportional Hazards Model (the Cox PH Model). The tables in Appendix 8 consists of Model 1 and Model 2. In comparison with Model 1, the interaction between ethnicity and educational attainment has been applied in Model 2. Overall, the estimated hazard ratios from the Cox PH Test, to some extent, are consistent with what I expect in Model 1 and Model 2 (Appendix 8). However, the coefficients in several variables such as SMS Areas are insignificant in the regression. The trend of risk of birth is an exponentially growth with women's age in 1979. It implies that the older women are prone to childbirth compared with younger counterparts. Whereas the test of proportional hazard assumption is passed, the coefficients in both age and age square are not significant in the 90% confidence interval. So even though indicators in age and age square are reasonable, I cannot trust the estimated hazard ratio.

When I focus on the variable, ethnicity in Model 1, the hazard ratios, which respectively are equal to 1.286 in Hispanic women and 1.612 in Black women, indicating that risk of the first childbirth in Hispanic women and Black women is separately 28.6 percent higher and 61.2 percent higher. Although the coefficients in Ethnicity are significant in 99% confidence interval, the hazard of childbirth among different groups is not proportional over time, on the basis of the result in the test of proportional hazards assumption. The test result rejects the null hypothesis that the Cox PH assumption is valid. It is possible that the hazards of childbirth between certain subgroups and the reference group are overlapping. Due to the overlapping occurring, I cannot assert that the risk of childbirth in Hispanic women or Black women is any higher over time. The hazard ratio in the variable, ethnicity, is unable to solidly verify the effects of ethnicity on the timing of first childbirth in Model 1.

Sowell (1981) judges that the socioeconomic characteristics among different races also play important roles for the timing of childbirth. When I move on the variable, education attainment, the risk of first childbirth in women with higher education is 28.9% less, which is inferred by the hazard ratio in the variable, educational attainment (0.711). Moreover, the hazard ratio in educational attainment is significant in 99% confidence interval. Therefore, the educational attainment serves as a determinant to postpone the timing of first childbirth in the United States (Gustafsson, 2003). However, due to the insignificant result in the test of proportional hazard

assumption, I cannot draw a solid conclusion, in terms of the relationship between educational attainment and the timing of first birth based on Model 1.

When I continue to research on the effects of ethnicity and educational attainment on the timing of first childbirth, the interaction regression between education attainment and ethnicity has been exploited in Model 2. Except from the subgroups Non-Hispanic/Non-Black women with college degree and Black women without college degree, others are valid in the Cox PH test. The risk of first childbirth in Hispanic women with college and Black women with college is respectively 8.4% lower and 8.3% higher as can be inferred to the hazard ratio of 0.916 and 1.083. However, the hazard ratios of subgroups, both Hispanic women with college and Black women with college are insignificant in 90% confidence interval. Therefore, it is reasonable to postulate that the timing of first birth in female minority with high education is contingent to the timing in Non-Hispanic/Non-Black women. In addition, the risk of timing in Hispanic women without college degree is 31.8% higher in comparison with Non-Hispanic/Non-Black women without college degree. In reference to the Non-Hispanic/Non-Black women, the timings of first childbirth in Hispanic women possessed different degree are divergent. The timing of Hispanic women without college degree is still significantly earlier than white counterparts, whereas the fertility behavior of Hispanic women with college degree is close to that in the Non-Hispanic/Non-Black women. The timing of first childbirth in Black women with a high educational degree is close to that in Non-Hispanic/Non-Black women. The educational attendance diminishes the differentials in timings of first childbirth between female minorities and Non-Hispanic/Non-Black women. As Solow's (1981) statement, the fertility behavior in Black women with higher education is approaching to the Non-Hispanic/Non-Black counterparts.

Another important socioeconomic characteristics is the labor force participation. Controlling other variables, the possibility of first childbirth in unemployed women is 44.7% higher in the United State, and the risk of first birth in the subgroup, women out of labor force, is 1.8 fold. Notwithstanding, Andersson and Scott (2005) judge that labor force participation boosts the proclivity of childbirth in Sweden. The different parenthood policy in the United States should be considered. In Sweden, the affirmative action programs implemented by the government enables to protect the maternal leave with payment. Until now, on the contrary, the United States do not

implement any similar policies. Andersson and Scott (2005) state that the conservative welfare system is unable to encourage the fertility in the employed women. Therefore, the working women in the United States suffer from many more disadvantages from parenthood, compared with counterparts in Sweden. Butz and Ward (1979) claim that the employment of young women causes the countercyclical fertility movements. Moreover, the cultural difference is another reason. In England, in which culture is much more close with the United States, female unemployment induces the increase in birth (Aksoy, 2016). In sum, female employment or out of labor force determinants significantly increase the risk of first birth in the United States. The empirical results strongly support this statement.

Fairfax (2014) states that the distribution of the fertility rate is associated with marriage. Initially, the Cox PH test indicates that all subgroups in the variable marital status are valid. In reference to the never married women, the possibility of first childbirth among married women is approximate 5 folds as can be inferred by 4.927 of the hazard ratio in Model 1. The risk of first birth in separated women and divorced women respectively is 1.6 folds and 45.6% percent higher. The hazard ratios in subgroups, separating women and divorced women, are significant in 99% confidence interval. In the United States, the marriage is still an imperative prerequisite of childbearing, in contrast with the cohabitation, which is prevailing nowadays. The small sample size in the subgroup, widowed, can explain the insignificant hazard ratio. The hazard ratio in the variable, marital status, in Model 2 is consistent with that in Model 1.

The variable, spouse's occupation, indicates that spouse's socioeconomic characteristics influence the timing of first childbirth. William's (2012) theory implies that spouse's occupation with respect to the income is associated with the timing of first birth. The spouse's income in white collar job is far more than that in the physical job, Army service, and out of labor force. Although the hazard ratios in spouse's occupation between Model 1 and Model 2 are different, the disparity is slight. In Model 1, women whose spouse engages in the white collar job are prone to experience the first childbirth. What the hazard ratios are interpreted as in Model 1 is consistent with that Butz and Ward's (1979) theory that the households, in which the husband earns higher salary, tend to have children. Military service, in which the spouse participates, induces the postponement in the first childbirth, based on the hazard ratio. The result matches with what Teachman, Tedrow, and

Anderson's (2015) judgement. It is not surprising that the decreased risk of first birth is attributed by the spouse being out of labor force.

Beside the women's employment status, the women's income also plays a crucial role for the fertility rate (Heckman & Walker, 1990; Andersson & Scott, 2005). Andersson and Scott (2005) state that the increasing risks of the parenthood is associated with the rising women's income. And relied on a longitudinal dataset, Heckman and Walker (1990) reinforce the conclusion. In the research, CPI has been used as the deflator. Based on the hazard ratio in women's salary, women's salary referred to 1982-84 CPI has the positive effect on the timing of first childbirth. The risks of timing in first childbirth are arising with the women's relative salary based on the categorical variable. The hazard ratios in subgroups, 101-200, 201-300, and Above 300 are significant, so I can trust these results.

In terms of geographic information, although the risk of first birth of women in subgroup, out of SMSA, is 6.2 percent lower in Model 1, the hazard ratio is not significant. Tarver (1969) states that due to the difference in education and job opportunity, the increase of fertility rate is in response of the distance from SMSA. I assume that what Tarver (1969) judges is too outdated to meet the socioeconomic context in the 1979 onwards. For example, Yang (1977) states that there is no significant effect of rural-urban difference on fertility given occupation class. The SMSA nowadays does not have significantly effect on the timing of first birth.

Caponi and Plesca (2014) point out that the total fertility rate among immigrant women is higher, in comparison with that in citizens. It shows that the risk of first childbirth in female immigrants is 4.6 percent higher in Model 1. The hazard ratio of the variable, citizenship, exhibits that immigrants have more proclivity to have child compared with native Americans. Notwithstanding, the hazard ratio of female immigrants is not significant in 90% confidence interval, so that the result, to some extent is not trustable. The insignificant hazard ratio in Model 2, in which the value is 0.926, accentuates the doubt.

Sowell (1981) judges that the behavioral deviances are significant in the same race. Bouvier and Gardner (1986) also reinforce the theory when they consider immigration, especially illegal

immigration. After general analyses, Model 3, Model 4 and Model 5 in Appendix 9, respectively exhibit the Cox PH model among the different races. Relied on Model 3, Model 4 and Model 5, I attempt to conduct the research about the effects of education attainment and employment status on the timing of first birth in intra-ethnicity. Based on the rank-log test, all of covariates in different races are valid. The robustness of measures in all models are also significant. Concerning subjects in various ethnicities, there are 3179 subjects in the Non-Hispanic/Non-Black women and 2052 out of 3179 subjects have experienced the childbirth in the survey. The number of failure in the Hispanic women and Black women respectively is 599 and 810. To assess covariant in Model 3, Mode 4, and Model 5 enables to analyze the difference in socioeconomic characteristics, when I get rid of the impacts of races.

Initially, the hazard ratios in education attainment indicate that to attend education universally impedes the first childbirth in all races. The adverse effects of educational attainment on the timing of first childbirth are consistent among various races. Especially, the differential of the risk of first birth in Black women is lowest in reference to other races. It is supported by Gustafsson's (2003) statement that to attend higher education prompts the postponement of childbirth in the United States. In addition, the hazard ratios in education attainment are significant in 99% confidence interval, so I can trust the result. The results circumstantially support that the educational attainment attenuates the differential in the timing of first childbearing among races. The fertility behaviors in female minority with higher education approaches that in Non-Hispanic/Non-Black women in the United States.

Concerning employment status, the differential in timing of first childbirth in Black women is not as obvious as that in other ethnicities. The risk of first birth in unemployed Black women is comparatively 17.8% higher as can be insignificantly inferred by the hazard ratio (1.178). The differential between employment and unemployment in the Black women is narrower than that in other women. It is able to be interpreted that the Black women mainly suffer from the involuntary unemployment. The chart concerning the women's unemployment by race (Chart 3) depicts that the unemployment rate in Black women is strikingly higher than that in other races. I may speculate that most unemployed Black women are prone to work, even though they are unemployed. Willingness for the employment prevents Black women from childbirth. On the contrary, Hispanic

women are prone to leave the labor force since first childbirth, in reference to the labor force participation by races (Chart 2). The lower motivation to work in the Hispanic women is the highest hazard ration differential between unemployed women and employed women in Model 3.

In terms of marital status, the relative differential in the risk of first childbirth in Black women is lower than other races. Melcher-Post (2011) assesses that the population with high level of unintended pregnancy rates possess traits including low income, less education, and ethnic minorities. Those traits are compatible with the socioeconomic characteristics of Black women. Moreover, the US government restrictively forbids abortion. The relatively lower differential in the risk of first childbirth in Black women enables to be interpreted as that the single mother in Black cohort is prevailing. In reference to Non-Hispanic/Non-Black women, the lower differential in the risk of first birth in Hispanic is attributed by the lower awareness of contraceptive methods (Craig, Dehlendorf, Borrero, Harper and Rocca, 2014). The lower awareness of contraceptive methods, to a great extent, accounts for the lower differential in risk of first childbirth in Hispanic women.

Black women are the most reluctant to become parents when their spouse participates in physical work compared with white collar job. The differential in risk of first birth infers the maternal decision-making in Black women is tightly associated with the women's spouse's occupation. In this study, the spouse's occupation serves as a proxy of spouse's income and social hierarchy. Therefore, Black women are more sensitive with respect to the spouse's occupation, income and hierarchy, when they make a decision in terms of the timing of parenthood.

Concerning the salary group and SMS Areas, the hazard ratios are insignificant and consistent with the timing of first childbirth in integrated cohorts, so I prefer to proceed to the variables of citizenship in Model 3, Model 4, and Model 5. In Model 3, the risk of first birth in Hispanic women without citizenship is higher. The result suggests that Hispanic immigrants have lower knowledge associated with contraceptive use compared to U.S.- born Hispanics (Craig, et la, 2014). In Model 4, nevertheless, the risk of first birth in Black women without citizenship is 22.7% lower as can be inferred by hazard ration (0.773). The possibility of first birth in Black women is insignificant. In

addition, most immigrants in the United States since 1960s come from South America and Asia (Bouvier & Gardner, 1986). The sample size of Black immigrants is small.

6. Research Limitation

6.1. In-balanced research

In the theoretical hypothesis, the respondents are supposed remain throughout the life of the survey. However, there are observations which have been missed, for several reasons. The dataset established is in-balanced. In other word, there is a deviance between what I estimate and the objective fact due to the omitted observations. When I conducted on the research in terms of hypotheses, the accuracy in the Cox PH model has been decreased due to omitted observations in the variables marital status and employment status. Because one of hypotheses is associated with the effects of employment status, I have to put the variable employment status into model, at the expense of the fewer observations. Though the interpolation or extrapolation, to some extent, addresses the issues owing to omitted observations, it is not applicable in this case. For example, though an array of observations in particular respondents are omitted regarding educational attainment, based on the last known observation I am able to speculate what the value should be. Nevertheless, it is difficult to conjure values in variables such as marital status, due to the frequent variance in those variables.

6.2. Information concerning respondents' spouse

In the NLSY79, only marriage is considered as a prerequisite of respondent's spouse. The information regarding cohabitants is also omitted in the NLSY79. I cannot research the hazard ratios of women's first parenthood when they are cohabitating with others. In addition, the insufficient history concerning spouse's parenthood is one of the other limitations in the study. Compared with women's first parenthood, the history of spouses' parenthood is difficult to acknowledge. It is taken for granted that respondents' spouses have the possibility to experience the marriage and childbirth prior to the study. Based on this assumption, observations in the NLSY79 cannot completely stand as the objective fact. However, the NLSY79 dataset cannot supply adequate information concerning respondent's spouses, in order to figure this issue out.

6.3. The timing of women's first childbirth VS. women's parenthood in the lifespan

Only women's first parenthood has been focused on in this thesis. Nevertheless, most of other researchers were interested in the women's parenthood across the whole lifespan. I cannot ignore the difference due to the different research questions. It is still necessary to consider the interval of the parenthood among women of different ethnicities. In comparison with women possessing lower educational attainment, women with higher degrees tend to postpone their first parenthood. However, as compensation, the women with higher degrees are inclined to shrink the interval between their first parenthood and sequent parenthoods. The differential in the total fertility rate among various races is not as salient as the differential in the timings of first childbirth, when the difference in the awareness of contraceptive method among various races are taken into account.

7. Conclusion

Since the new continent was discovered, a variety of people have gradually congregated towards America. Nevertheless, people of different races still maintain their intrinsic characteristics rather than the integration in the mainstream society (Sowell, 1981). I attempt to research on the effects of socioeconomic characteristics among different ethnicities regarding fertility. Hotz, Klerman and Willis (1997) outlines that though women in the United states generally have tended to delay their first childbirth since the 1980s, this extent is the highest in white women. According to Andersson and Scott (2005), the timings of first childbirth among different ethnicities in Sweden are delayed and convergent with time. Moreover, what Andersson and Scott (2005) expound enlightened me to research whether female employment affects the timing of their first childbirth in the United States.

The U.S. department of Health and Human service has implemented the National Longitudinal Survey, based on the 1960 youth survey. The study lasts over 20 years and highlights the transition from school to the labor market. The subjects in the study consist of women who were born between the 1950s and 1960s. In order to research the first childbirth, the samples who had experienced childbirth have been removed from the study. The ethnicity in the study consists of Hispanic, Black, and Non-Hispanic/Non-Black. Due to economic reasons, the sample in the army has been removed from study since 1984. I have deleted the sample, in which the observations are derived from the army service. After the preliminary cleanup, the sample size is 5276. The failure in the sample is 3618 out of 5276. The socioeconomic characteristics in the study consist of

subject's age, ethnicity. educational attainment, employment status, marital status spouse's occupation, SMSA areas, women's salary and citizenship.

Andersson and Scott (2005) used the Cox proportional hazard model, in order to research the risk of the first childbirth between native Swedish and immigrants. The Kaplan-Meier estimators and the Cox PH model have been applied in this study. During the empirical analysis, the effects of left-truncated and right censored observation should be born in mind. Removed left truncated observations are possibly to contributing the underestimated bias. Nevertheless, the Kaplan-Meier estimated curve for the risk of first birth (Appendix 7, B) implies the difference of timing of first birth between Hispanic women and Black women is not significant. In the Cox PH model, the log-rank test in which the equal survivor function for different groups is the null hypothesis, has been exhibited. The test of proportional hazard assumption has been applied, for the sake of the sophisticated scrutiny among various subgroups. Because all subjects with different ages had joined in the survey in the 1979 onwards, I research on a flow sample, rather than the stock sample.

Based on the Cox PH model and NLSY79 dataset, I had engaged in the empirical analyses. Initially, parents' characteristics have been discussed. As Beckerian economic theory suggests, parents' characteristics determine fertility. Though cohabitation and parenthood in cohabitation nowadays are prevailing, the relationship between marriage and parenthood remains strong (Fairfax, 2014). The result stemming from the empirical analysis is still consistent with theory. The timing of first childbirth in single women is significantly later. In the neoclassical model, the effects of the women's income on the timing of first childbirth is positively significant (Heckman & Walker, 1990; Butz & Ward, 1979; Wrigley & Schofield, 1981). When the women's salary is adjusted by CPI, the possibility of first childbirth in higher salary groups is earlier in women with higher income.

In contrast with female income, the adverse effect of female employment on the timing of childbirth is significant (Hotz, Klerman and Willis, 1997). It is observed that except from Black women, the timings of first birth in employed women in both Hispanic and Non-Hispanic/Non-Black are significantly later, compared with unemployed counterparts. Although the timing of first birth in Black women is insignificant, the hazard ratio is consistent with others. So I can assume

that if women have a job with higher initial wage, they are prone to delay the timing of first birth, due to an intermittent career path and the huge maternal time inputs in parenthood. Gustafsson (2003) strengthens this point from the human capital perspective.

Beckerian economic theory also illustrates that the household earnings influence on the total fertility rate. Beside the wife's income, the husband's income is involved in the household earning. The husband's income plays a role of the fertility rate, given female earnings (Butz & Ward, 1979). Nevertheless, the husband's income is altered by the spouse's occupation in the research due to data limitations on husband's income. Williams (2012) points out that the positive relationship between the inter-occupation and wage inequality is significant. It still makes sense that the spouse's occupation takes over the husband's income in the study. In reference to the white collar job, physical work, army or out of labor force, in which women's spouse engages, this induces the postponement of parenthood.

Sowell (1981) discusses the effects of race on the fertility rate by means of the educational attainment. The educational attainment facilitates the postponement of parenthood (Gustafsson, 2003). The *pre parental human capital* plays a vital role in the timing of parenthood. The interaction results indicate that the possibility of childbirth is diminished in minority group with high education. In reference to the Non-Hispanic/Non-Black women without a college degree, the hazard ratio in minorities with college degree are insignificantly different. The differentials in timing of first childbirth between female minority with college and Non-Hispanic/Non-Black women is tiny. I postulate that the effects of education attainment on the first childbirth attenuates the impact of races. According to Caponi and Plesca (2014), citizenship in the United States also is taken into account. Although the hazard ratio is insignificant in the empirical analysis, the differential in the risk of first childbirth between citizens and immigrants is compatible with Bouvier and Gardner's (1986) statement.

In sum, the socioeconomic characteristics among different ethnicities as determinants exert far-reaching influence on the maternal decisions. Although the study is carried out based on in-balanced research and alterative variables, reasonable conclusions have been laid out in terms of

the effects of ethnicity, employment and education attainment on the timing of first childbirth in the United States.

8. Reference list

Aksoy, C. G. (2016). The Effects of Unemployment on Fertility: Evidence from England. The BE Journal of Economic Analysis & Policy, 16(2), 1123-1146.

Andersson, G., & Scott, K. (2005). Labour-market status and first-time parenthood: the experience of immigrant women in Sweden, 1981–97. Population studies, 59(1), 21-38.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis.The journal of political economy, 9-49.

Becker, G., & Barro, R. J. (1986). A reformulation of the economic theory of fertility.

Bogue, D. J. (2010). Contraception, attitude-practice, and fertility differentials among US Hispanic, African–American and white women. Journal of Population Research, 27(4), 275-292.

Bouvier, L. F., & Gradner, R. W. (1986). Immigration to the US: the unfinished story. Population Bulletin, 41(4), 3-50.

Bruin, J. 2006. newtest: command to compute new test. UCLA:
Statistical Consulting Group. http://www.ats.ucla.edu/stat/stata/ado/analysis/.

Budig, Michelle & Paula England (2001), 'The wage penalty for motherhood.' American Sociological Review, 66: 204-25.

Butz, W. P., & Ward, M. P. (1979). The emergence of countercyclical US fertility. *The American Economic Review*, *69*(3), 318-328.

Burnette, Joyce & Maria Stanfors (2012), 'Was there a family gap in late nineteenth century manufacturing? Evidence from Sweden.' History of the Family, 17: 31-50.

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge university press.

Caponi, V., & Plesca, M. (2014). Empirical characteristics of legal and illegal immigrants in the USA. Journal of Population Economics, 27(4), 923-960.

Chumlea, W. C., Schubert, C. M., Roche, A. F., Kulin, H. E., Lee, P. A., Himes, J. H., & Sun, S. S. (2003). Age at menarche and racial comparisons in US girls.Pediatrics, 111(1), 110-113.

Cleves, M., 2008. *An introduction to survival analysis using Stata*. Stata Press.

Cox, D.R., 1992. Regression models and life-tables. In *Breakthroughs in statistics* (pp. 527-541). Springer New York.

Craig, A. D., Dehlendorf, C., Borrero, S., Harper, C. C., & Rocca, C. H. (2014). Exploring young adults' contraceptive knowledge and attitudes: disparities by race/ethnicity and age. Women's Health Issues, 24(3), e281-e289.

Easterlin, R. A. (1973). Relative economic status and the American fertility swing.

Ermisch, J. F. (1989). Purchased child care, optimal family size and mother's employment Theory and econometric analysis. *Journal of Population Economics*, *2*(2), 79-102.

Fairfax, C. N. (2014). Social Work, Marriage, and Ethnicity: Policy and Practice. Journal of Human Behavior in the Social Environment, 24(2), 83-91.

Goldin, Claudia & Lawrence Katz (2002), 'The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions.' Journal of Political Economy, 110: 730-70.

Gustafsson, S. (2003). Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe. In *Family, household and work* (pp. 345-367). Springer Berlin Heidelberg.

Heckman, J. J., & Walker, J. R. (1990). The relationship between wages and income and the timing and spacing of births: evidence from Swedish longitudinal data. *Econometrica: journal of the Econometric Society*, 1411-1441.

Hotz, V. J., Klerman, J. A., & Willis, R. J. (1997). The economics of fertility in developed countries. *Handbook of population and family economics*, *1*(Part A), 275-347.

Hummer, R. A., & Hamilton, E. R. (2010). Race and ethnicity in fragile families. The Future of Children, 20(2), 113-131.

Jiang, Q., Li, Y. and Sánchez-Barricarte, J.J., 2016. Fertility Intention, Son Preference, and Second Childbirth: Survey Findings from Shaanxi Province of China. *Social Indicators Research*, *125*(3), pp.935-953.

Klein, J. P., & Moeschberger, M. L. (2005). Survival analysis: techniques for censored and truncated data. Springer Science & Business Media.

Loughran, D. S. & Zissimopoulos, J. M. (2009), 'Why wait? The effect of marriage and childbearing on the wages of men and women.' Journal of Human Resources, 44: 326-49.

Mather, M. (2012). Fact sheet: The decline in US fertility. Population Reference Bureau, World Population Data Sheet.

Melcher-Post, J. R. (2011). Contraception Use Among Women in an Urban Clinic (Doctoral dissertation, The University of Arizona).

Olopha, P., & Aladeniyi, O (2015), Demographic Analysis of the Effect of Some Determinants

of Fertility on Fertility Intentions – The Rural and Urban Factor. Journal of Natural Sciences Research Vol.5, No.18, 2015, pp.35-47.

Sowell, T. (1981). Ethnic America: a history. Basic Books.

Tarver, J. D. (1969). Gradients of urban influence on the educational, employment, and fertility patterns of women. Rural Sociology, 34(3), 356.

Teachman, J., Tedrow, L., & Anderson, C. (2015). The Relationship between Military Service and Childbearing for Men and Women. Sociological Perspectives, 58(4), 595-608.

Williams, M. (2012). Occupations and British wage inequality, 1970s–2000s. European Sociological Review, jcs063.

Wrigley, E. A. and Schofield, R. S. (1981) The Population History of England 1541-1871. A reconstruction. Chapter 7 and Chapter 11. London: Edward Arnold.

Yang, S. O. W. (1977). Rural-Urban Differences in Factors Affecting Fertility: A Case Study of Ohio.

Zagorsky, J. L., & White, L. (1999). The National Longitudinal Surveys: NLSY79 user's guide. Columbus, OH: Ohio State University. Center for Human Resource Research.

9. Appendix

Reference	Code
Non-interview	-5
Valid Skip	-4
Invalid Skip	-3
Don't Know	-2
Refusal	-1

Appendix 1: The meaning of negative numbers in the NLSY79 dataset.

Source: The National Longitudinal Surveys: NLSY79 user's guide

Variable	No. of Obs	Mean	Stan Dev	Max	Min
Age Group	3,618				
14		0.09	0.29	1	0
15		0.15	0.36	1	0
16		0.16	0.36	1	0
17		0.14	0.35	1	0
18		0.13	0.34	1	0
19		0.12	0.33	1	0
20		0.10	0.30	1	0
21		0.09	0.28	1	0
22		0.02	0.14	1	0
Sample Race	3,618				
Hispanic		0.18	0.38	1	0
Black		0.23	0.42	1	0
Non-Hispanic Non-black		0.59	0.49	1	0
Educational attainment	3,618				
College		0.37	0.48	1	0
Below college		0.63	0.48	1	0
Employment status	3,461				
Employed		0.53	0.50	1	0

Unemployed		0.08	0.28	1	0
Out of labor force		0.38	0.49	1	0
Marital status	3,461				
Never married		0.32	0.47	1	0
Married		0.64	0.48	1	0
Separated		0.02	0.13	1	0
Divorced		0.02	0.15	1	0
Widowed		0.00	0.03	1	0
Spouse's occupation	3,618				
Unknown		0.40	0.49	1	0
White collar		0.28	0.45	1	0
Physical work		0.27	0.44	1	0
Army		0.04	0.20	1	0
Out of Labor		0.01	0.10	1	0
SMSA areas	3,618				
Unknown		0.30	0.46	1	0
Not in SMSA		0.26	0.44	1	0
Central city		0.27	0.45	1	0
Central city		0.16	0.37	1	0
Women's salary (CPI)	3,618				

None		0.14	0.35	1	0
1-100		0.45	0.50	1	0
101-200		0.21	0.41	1	0
201-300		0.06	0.23	1	0
301-400		0.01	0.10	1	0
401-500		0.00	0.04	1	0
Above 500		0.01	0.07	1	0
Unknown		0.12	0.32	1	0
Citizenship	3,618				
No		0.06	0.23	1	0
Yes		0.86	0.35	1	0
Unknown		0.08	0.27	1	0

Appendix 2: Descriptive Statistics for Micro Level Variables in NLSY79

Year	**Annual CPI**
1979	72.6
1980	82.4
1981	90.9
1982	96.5
1983	99.6
1984	103.9
1985	107.6
1986	109.6
1987	113.6
1988	118.3
1989	124.0
1990	130.7
1991	136.2
1992	140.3
1993	144.5
1994	148.2
1995	152.4
1996	156.9
1997	160.5
1998	163.0
1999	166.6
2000	172.2
2001	177.1
2002	179.9

2003	184.0
2004	188.9
2005	195.3
2006	201.6
2007	207.342
2008	215.303
2009	214.537
2010	218.056
2011	224.939
2012	229.594

Appendix 3: Consumer Price Index-All urban consumers. Base period: 1982-84
Source: U.S. Bureau of Labor Statistics

Category	total	mean	min	median	max
no. of subjects	5276				
no. of records	65146	12.34761	1	10	25
(first) entry time		0	0	0	0
(final) exit time		15.53999	1	10	34
subjects with gap	0				
time on gap if gap	0	.	.	.	.
time at risk	81989	15.53999	1	10	34
failures	3618	0.6857468	0	1	1

Appendix 4: The Description for the Survival Function

Variable	Constant	Varying	Never missing	Always missing	Sometimes missing
Age Group	5276	0	5276	0	0
Ethnicity	5276	0	5276	0	0
Educational attainment	3793	1483	5276	0	0
Employment status	1411	3753	3757	112	1407
Marital status	2180	2984	3770	112	1394
Spouse's occupation	2157	3119	5276	0	0
SMSA areas	2036	3240	5276	0	0
Women's salary (CPI)	959	4317	5276	0	0
Citizenship	5276	0	5276	0	0

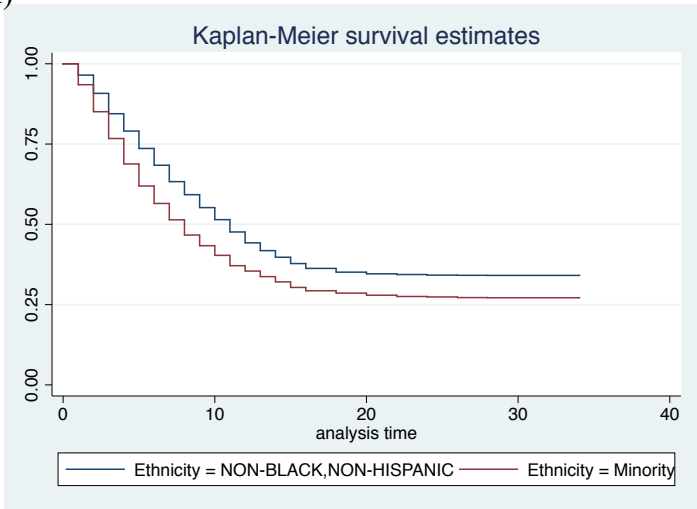
Appendix 5: The Description for Time-constant Variables and Time-varied Variables

Time	Beg Total	Fail	Net Lost	Survival Function	Std Error	[95% Conf. Int.]
1	5276	247	0	0.9532	0.0029	0.9471 0.9586
2	5029	355	0	0.8859	0.0044	0.8770 0.8942
3	4674	375	0	0.8148	0.0053	0.8041 0.8250
4	4299	336	0	0.7511	0.006	0.7392 0.7626
5	3963	315	0	0.6914	0.0064	0.6788 0.7037
6	3648	280	0	0.6384	0.0066	0.6252 0.6512
7	3368	269	0	0.5874	0.0068	0.5740 0.6005
8	3099	228	0	0.5442	0.0069	0.5306 0.5575
9	2871	198	0	0.5066	0.0069	0.4931 0.5200
10	2673	183	0	0.4719	0.0069	0.4584 0.4854
11	2490	190	0	0.4359	0.0068	0.4225 0.4493
12	2300	144	0	0.4086	0.0068	0.3954 0.4219
13	2156	114	0	0.387	0.0067	0.3739 0.4002
14	2042	100	0	0.3681	0.0066	0.3551 0.3811
15	1942	99	0	0.3493	0.0066	0.3365 0.3622
16	1843	70	0	0.3361	0.0065	0.3233 0.3488
18	1773	53	0	0.326	0.0065	0.3134 0.3387
20	1720	30	0	0.3203	0.0064	0.3078 0.3329
22	1690	15	0	0.3175	0.0064	0.3049 0.3301
24	1675	9	0	0.3158	0.0064	0.3033 0.3283

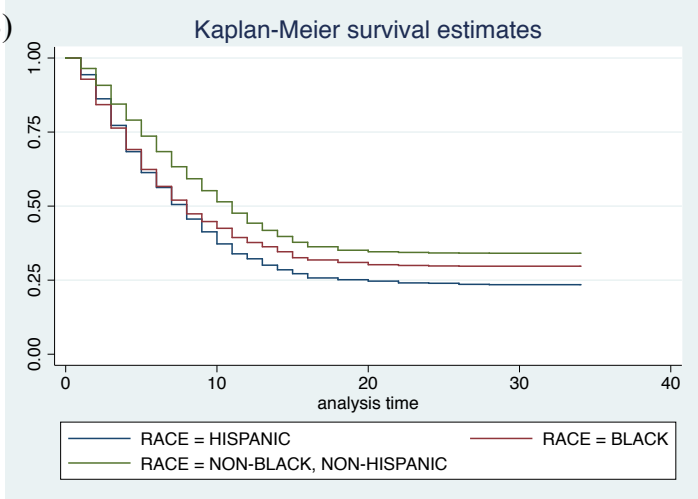
26	1666	6	0	0.3146	0.0064	0.3021 0.3272
28	1660	2	0	0.3143	0.0064	0.3018 0.3268
34	1658	0	1658	0.3143	0.0064	0.3018 0.3268

Appendix 6: Life Table constructed based on the NLSY79 Database.

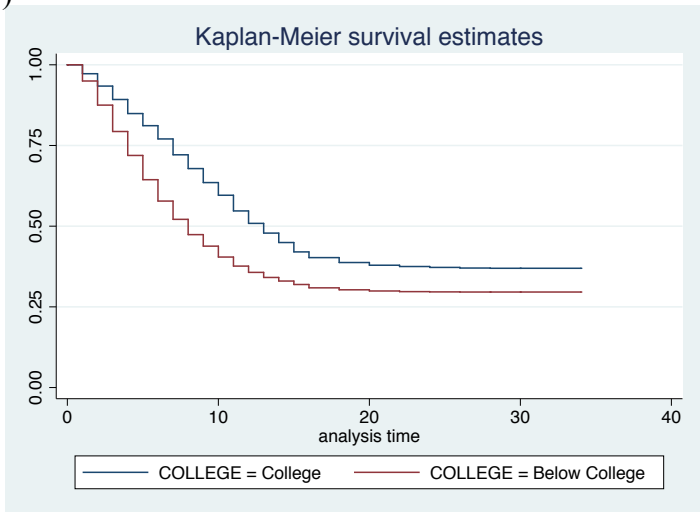
(A)



(B)



(C)



Appendix 7: Kaplan-Meier survival estimates. The number at risk are available every year from zero to 16th year and are available every two years from 16th year to 34th year. (A) Kaplan-Meier estimated survival curves for the risk of first childbirth comparing Non-Hispanic, Non-Black women, versus female minority. (B) Kaplan-Meier estimated survival curves for the risk of first childbirth comparing Non-Hispanic, Non-Black women, versus Hispanic women, versus, Black women. (C) Kaplan-Meier estimated survival curves for the risk of first childbirth comparing women attained college, versus women without college degree.

	Model 1		**Model 2**	
	Haz. Ratio	Test of Proportional hazard assumption	Haz. Ratio	Test of Proportional hazard assumption
Age_Group2	0.995	0.0637	0.997	0.1005
Age_Group	1.110	0.0504	1.061	0.0699
SAMPLE_RACE				
Non-Hispanic Non Black (Reference)	-	-	-	-
HISPANIC	1.286***	0.0031	-	-
BLACK	1.612***	0.0000	-	-
Educational attainment				

College	0.711***	0.0000	-	-
Below College (Reference)	-	-	-	-
EDUCATION_RACE				
Non-Hispanic Non-Black *College	-	-	0.734***	0.0000
Non-Hispanic Non-Black * Below College (Reference)	-	-		.
Hispanic*College	-	-	0.916	0.6241
Hispanic*Below College	-	-	1.318***	0.5697
Black*College	-	-	1.083	0.7533
Black*Below College	-	-	1.698***	0.0003
Employment status				
Employed (reference)	-	-	-	-
Unemployed	1.447***	0.0478	1.451***	0.0481
Out of labor force	2.877***	0.1753	2.861***	0.1496
Marital status				
0:0 never married (reference)	-	-	-	-
1:1 married	4.927***	0.1236	5.064***	0.0603
2: 2 separated	2.556***	0.0208	2.612***	0.0106
3: 3 divorced	1.456***	0.5136	1.494***	0.4264
6: 6 widowed	1.614	0.2327	1.579	0.2694
Spouse's occupation				
Unknown	0.664***	0.0580	0.669***	0.0315
White collar (Reference)	-	-	-	-
Physical Work	0.878***	0.1254	0.872***	0.2427
Army	0.514***	0.1153	0.582***	0.4822
Out of Labor	0.463***	0.9815	0.449***	0.9247
SALARY_GROUP				
None	0.692***	0.0139	0.719***	0.0013
1-100 (Reference)	-	-	-	-
101-200	1.089*	0.7681	1.084*	0.6265
201-300	1.254***	0.0974	1.253***	0.0959
Above 300	1.334**	0.4450	1.334**	0.4326
Unknown	0.557***	0.0031	0.550***	0.0081

SMS Areas				
Unknown	0.929*	0.5704	0.915**	0.8039
Not in SMSA	0.938	0.0516	0.936	0.0395
SMSA (Reference)	-	-	-	-
Citizenship				
No	1.046	0.0903	0.926	0.4823
Yes (Reference)	-	-	-	-
Unknown	0.822*	0.1909	0.560***	0.0143
Number of obs	50174		50174	
No. of subjects	5164		5164	
No. of failures	3461		3461	
Time at risk	58402		58402	
LR chi2(26)	3289.72		3364.24	
Log likelihood	-25920.938		-25883.677	

Appendix 8: Hazard Ratio from the Cox Proportional Hazard Model with interactive variables between ethnicity and citizenship in 1990

*** p<0.01, **p<0.05 and *p<0.1

Source: NLSY79

	Model 3	**Model 4**	**Model 5**
	Hispanic	Black	Non-Hispanic/Non-Black
Age_Group2	1.007	0.981**	0.997
Age Group	0.723	1.778**	1.064
Educational attainment			
College	0.710***	0.748***	0.720***
Below College (Reference)	-	-	-
Employment status			
Employed (reference)	-	-	-
Unemployed	1.611***	1.178	1.479***
Out of labor force	2.581***	2.539***	2.972***
Marital status			
0:0 never married (reference)	-	-	-
1:1 married	6.045***	2.517***	7.197***

2: 2 separated	3.611***	1.836**	3.379***
3: 3 divorced	1.845**	0.765	2.073***
6: 6 widowed	4.466	0.636	2.704
Spouse's occupation			
Unknown	0.848	0.589**	0.668***
White collar (Reference)	-	-	-
Physical Work	1.010	0.579***	0.925
Army	0.901	0.427***	0.567***
Out of Labor	0.590	0.345***	0.567***
Women's salary			
None	0.765**	0.744***	0.626***
1-100 (Reference)	-	-	-
101-200	1.140	1.039	1.074
201-300	1.325	1.075	1.190*
Above 300	1.289	1.161	1.246
Unknown	0.412***	0.460***	0.759**
SMS Areas			
Unknown	0.871	0.938	0.950
Not in SMSA	0.977	0.872	0.957
SMSA, not central city (Reference)	-	-	-
Citizenship			
No	1.097	0.773	1.194
Yes (Reference)	-	-	-
Unknown	1.075	1.004	0.756*
Number of obs	7582	11364	31228
No. of subjects	830	1164	3170
No. of failures	599	810	2052
Time at risk	8795	13704	35903
LR chi2(22)	648.53	328.39	2433.85
Log likelihood	-3336.266	-5123.9605	-14083.801

Appendix 9: Hazard Ratio from the Cox Proportional Hazard Model among different races

*** p<0.01, **p<0.05 and *p<0.1

Source: NLSY79