Marriage, Intergenerational Schooling Effect, and Gender Gap in College Attainment

Suqin Ge†  Fang Yang‡
Virginia Tech  SUNY-Albany

First draft: August 2007
This version: April 30, 2008

Abstract

One striking phenomenon in the U.S. labor market is the reversal of the gender gap in college attainment. Females have outnumbered males in college attainment since 1987. We develop a discrete choice model of the college entry decision to study the effects of changes in relative earnings, changes in parental education, and changes in the marriage market on time series observations of college attainment by gender. We find that the increasing relative earnings between college and high school persons and the increases of parental education have important effects on the increase in college attainment for both genders, while the decrease of marriage rates is crucial in explaining the reversal of gender gap in college attainment.

Keywords: gender, education, marriage, women, intergenerational schooling persistence
JEL Classification: J24, J16, I20

†We thank Michele Boldrin, Betty Daniel, Mariacristina De Nardi, Zvi Eckstein, John Jones, and seminar participants at Virginia Tech, 2008 AEA meetings, and University of Virginia for helpful comments and suggestions. We are grateful to the formal discussant, Raquel Fernandez, at 2008 AEA meetings. All remaining errors are our own.

‡Ge: Department of Economics, Virginia Tech, Blacksburg, VA 24061 (email: ges@vt.edu).

‡Yang: Department of Economics, University at Albany, Albany, NY 12222 (email: fyang@albany.edu).
1 Introduction

One striking phenomenon in the U.S. labor market is the reversal of the gender gap in college attainment. 57% of young men aged 25 to 34 in 1980, as compared with 46% of young women, had some college education by age 34. In 1996, however, female college attainment reached 64%, 5 percentage points higher than that of males. In fact females have taken over males in college attainment since 1987.

There is large empirical work on the determinants of education outcomes which emphasizes the role of the earnings premium as a key explanatory variable (see, for example, Becker (1967), Mincer (1974), and Willis and Rosen (1979)). In addition, an extensive literature shows that family background is an important determinant of schooling decision (see, among others, Kane (1994), Cameron and Heckman (1998, 2001), Eckstein and Wolpin (1999), and Ge (2006)). Moreover, many papers argue that, both empirically and theoretically, assortative matching in marriage market is important in schooling decision (see, among others, Ge (2006), Chiappori, Iyigun and Weiss (2006)).

Based on the above literature, we construct a formal economic model including the potential costs and benefits which determine individual college decisions. In our model, individuals with different ability first make college decisions. Then they might form marriages and have children. Parents are altruistic and value children’s ability, which is increasing with parents’ education. Forward-looking individuals take into account the impact of their own schooling on their children’s ability. Other factors that affect an individual’s education choices include the expected direct labor market returns to college, the expected marriage market returns to college, and effort costs for attending college. These costs and benefits can differ by gender.

We estimate the parameters of the model by matching data on aggregate college attainment and college attainment conditional on parents’ education by gender from
1980-1996 Panel Study of Income Dynamics (PSID). We present evidence on the fit of the model to the data. We then use the parameter estimates to simulate counterfactual experiments to decompose the sources of changes in college attainment into the effects of changes in relative earnings, changes in parental education, and changes in the marriage market.

What accounts for the increase of college attainment? We find that the increasing relative earnings between college and high school persons have important effects on the increase in college attainment for both genders. We also emphasize the importance of intergenerational persistence in schooling on the increase in college attainment for both genders. A college-educated parent is substantially more likely to have a college-educated daughter or son than a non-college graduate, even after controlling for the education of the other parent. This link between parents’ and children’s schooling provides an inter-generational propagation mechanism: as the number of college-educated parents increases, their children become more likely to attend college. Thus, the gradual transformation of the parental education itself acts as a propagation mechanism in changes in college attainment.

What accounts for females overtaking males in college attainment? We find that the decreasing marriage rates are crucial in explaining the relative increase in female college attainment. Two factors are relevant here. First, when married, the returns to college education are higher for males than those for females. Second, when single, the return to college education is higher for females than for males. As marriage rates decrease, the returns to college for females being single become high enough to compensate for the low returns to college for females being married, and female college enrollment exceeds that of males.

This paper contributes to an active and growing literature on gender differences in educational attainment. Several papers have studied college enrollment and grad-
uation by gender for one cohort. Averett and Burton (1996) focus on those individual from age 14 to 22 in 1979 and use a human capital model to examine gender differences in college enrollment. Card and Lemieux (2001) argue that large cohort sizes might have accounted for the declining of college enrollment in 1970s. Rios-Rull and Sanchez-Marcos (2002) construct a model to explain why males had a higher college attainment than females in the 1970s. Jacob (2002) finds that higher non-cognitive skills and college premiums among women account for nearly 90 percent of the gender gap in higher education enrollment in 1988. Those work focus on one cohort only and thus cannot examine the trends.

Among works that study the reversal of gender gap in higher education enrollment over time, Anderson (2002) suggests the role of increasing discount rates over time in explaining gender gap in higher education enrollment. Charles and Luoh (2003) emphasize the effect of uncertainty of future wages on relative schooling by gender. Chiappori, Iyigun and Weiss (2006) show that women can acquire more schooling than men if gender wage gap exists but it narrows with the level of education. To our knowledge, the paper is the first that incorporates several factors in a structure model to explain the gender gap in college attainment.

The paper is organized as follows. In Section 2, we present some empirical results from the PSID documenting college attainment rates in 1980-1996. In Section 3, we present our model. Section 4 provides parameters estimated from the data that are used in the model. Section 5 presents the quantitative results of the benchmark model and investigates the quantitative importance of changes in relative earnings, changes in parental education, and changes in the marriage market. Brief concluding remarks are provided in Section 6.
2 Data on College Attainment

We use the PSID to calculate college attainment rates. The PSID is a longitudinal survey of U.S. families and the individuals who make up those families. Approximately 4,800 U.S families are sampled in 1968 and these families have been re-interviewed annually until 1997. From 1997 onwards PSID was changed to biennial data collection and two major changes were made to the PSID: a reduction of the core sample, and the addition of a new sample of post 1968 immigrant families and their adult children.

We select individuals in the core sample whose age is between 25-34 in each year and with valid information on parents’ education.\(^1\) We use completed schooling among mature adults as a measure of the schooling.\(^2\) Individual who has more than 12 years of education completed by age 34 is defined as having college education. College attainment rate is calculated as the fraction of individuals that have college education among each specific group.

\[\text{College attainment rate} = \frac{\text{Number of individuals with college education}}{\text{Total number of individuals}}\]

Figure 1: College attainment rates by age 34 among those aged 25-34. Source: Authors’ calculations from the PSID data files.

\(^1\)We thus use the average college attainment of 10 birth cohorts. The sample size in the PSID is too small for us to analyze each birth cohort.

\(^2\)We want to focus on schooling attainment among the population of Americans whose schooling activity can reasonably be expected to have ended. See Charles and Luoh (2003) for a discussion of the advantage of using schooling attainment over enrollment.
Figure 1 illustrates the changes in relative college attainment by males and females over the sample period considered here, 1980 to 1996.\textsuperscript{3} 57\% of young men aged 25 to 34 in 1980, as compared with 46\% of young women, had some college education. In 1996, however, female college attainment reached 64\%, 5 percentage points higher than that of males. In fact females have taken over males in college attainment since 1987.\textsuperscript{4} \textsuperscript{5}

We also calculate college attainment rates conditional on parent’s education. We first find parents’ education for the selected sample by linking parents and children from Individual Files (1968-2005). The PSID facilitates the intergenerational linkage by providing parent’s ID in the Individual Files. If a linkage can not be found in Individual Files, we use 2003 Parent Identification File to link an individual with his/her parents. If the above procedure fails to provide parent’s education information, we

\textsuperscript{3}We choose this period due to the availability of data. The latest year of data available to us is 2005 the PSID. Since we use education completed by age 34, individuals at the age of 34 in 2005 was 25 in 1996. After 1997, education by age 34 will not be available for individuals at age 25 in each year.

\textsuperscript{4}Other studies (see, for example, Charles and Luoh (2003), Goldin, Katz and Kuziemko (2006)), which use different measures of education or different data sets, find similar patterns.

\textsuperscript{5}The sample size in PSID is too small if we divide sample by race/ethnicity. The convergence and ultimate taken-over in completed schooling among successive generations of men and women is also evident when we divide sample by race/ethnicity using the CPS.
find parents’ education by using parents’ and parents-in-laws’ education reported by the head in Family Files. In 1974 questions were asked about how much education had been completed by the household head’s parents and by the spouse’s parents. In the later waves, these parental education questions were asked for new heads and spouses. By merging Individual Files with Family Files, we are able to find parents’ education for those who were heads or spouses or siblings of the heads.

Figure 2 shows female college attainment conditional on parental education. We observe that, a college-educated parent is substantially more likely to have a college-educated daughter than a non-college-educated parent, even after controlling for the education of the other parent. For example, among those who were at age 25-34 in 1980, 84% of females whose parents had college education had attended some college, 5 percentage points higher than those whose father had college education and mother did not, 20 percentage points higher than those whose mother had college education and father did not. Therefore, the marginal effect of father’s education on children’s education is bigger than that of mother’s.

3 The Model

The economy is populated by overlapping generations that live for 2 periods. We assume that going to college entails an idiosyncratic non-pecuniary effort cost \( D \in [0, \infty) \). In each period, the adult population is characterized by a distribution of effort costs. At the beginning of the first period, individuals with different costs make schooling decisions. In the second period, they might form marriages and have children. Parents are altruistic and value their children’s ability. We assume that the higher is a parent’s education, the higher is his/her children’s ability. Forward-looking

---

6 Similar patterns hold for sons and the results are available from the authors upon request.
7 We can interpret the effort cost as net of psychic benefit of attending college.
individuals take into account the impact of their own schooling on their children’s ability. Other factors that affect an individual’s education choices include the direct labor market returns to college and marriage market returns to college. These costs and benefits can differ by gender. We now describe the model in more detail.

3.1 Marriage and the Labor Market (Second Period)

In the second period, individuals of schooling type $s_f$ and $s_m$ marry at an exogenously given rate, and they work. Denote the education of male $s_m = \{0, h, c\}$, and the education of female $s_f = \{0, h, c\}$, where 0 stands for being single. Let $Y_{g,s_m,s_f}$ denote the earnings of an individual of gender, $g = \{f, m\}$, male’s education $s_m$, and female’s education $s_f$.

Each individual values his/her own consumption and children’s learning ability, if he/she has children. We abstract from out-of-wedlock birth and assume that a single individual does not have children. If a person does not marry, he/she enjoys his/her own consumption. The lifetime utility function for a single female and a single male of schooling $s$ is, respectively,

\begin{align}
U^f (s_m = 0, s_f = s) &= \log(Y_{f,0,s}), \\
U^m (s_m = s, s_f = 0) &= \log(Y_{m,s,0}).
\end{align}

A married couple has children. We assume that fertility is exogenous and the total number of children a couple has is independent of each spouse’s education.

---

8We abstract from tuition costs which should have similar marginal effects on the education of each gender and are unlikely to explain the gender difference in college attainment.

9For simplicity, we do not model marriage as a match outcome. Fernandez, Guner and Knowles (2005) study the interactions between household matching, inequality, and per capita income.

10We plan to extend our analysis in future research.

11Fernandez and Rogerson (2001), and Rios-Rull and Sanchez-Marcos (2002) show that the fertility declines with income and education. Adopting the assumption that the fertility declines with education should only change the results marginally since in our model individual’s decision is not
We abstract from the financial cost of raising children. The cost of having children as the opportunity cost of time will be incorporated in the estimation of earnings process.

The learning ability of the couple’s children, $a'$, is a function of the couple’s human capital, $s_m$ and $s_f$. The production function of children’s ability is Cobb-Douglas

$$a' = s_m^{1-\theta} s_f^\theta.$$  

This functional form captures the fact that when parents are more educated, their children tend to have high learning ability. This could occur because more educated parents provide a better environment for children to flourish, or because parental ability is passed on genetically (Plug and Vijverberg (2003)). Children of different genders from the same family have the same learning ability.

Each spouse gets a share of the total family income, with the weight of each spouse depending on his/her individual relative earnings through a parameter $\lambda \in [0, 1]$. The share of the wife is $(1 - \lambda)0.5 + \lambda Y_{f,sm,sf}/(Y_{m,sm,sf} + Y_{f,sm,sf})$ and the share of the husband is $(1 - \lambda)0.5 + \lambda Y_{m,sm,sf}/(Y_{m,sm,sf} + Y_{f,sm,sf})$. Notice that $\lambda = 0$ is the case of full income-pooling, while $\lambda \in (0, 1]$ implies that each spouse’s weight is increasing in his/her share of household earnings.

The utilities of men and women at a marriage type $(s_m, s_f)$ are given, respectively, by

(4) \quad & U_f(s_m, s_f) = \log[0.5( (1 + \lambda)Y_{f,sm,sf} + (1 - \lambda)Y_{m,sm,sf} )] + \lambda_a \log[s_m^{1-\theta} s_f^\theta], \\
(5) \quad & U_m(s_m, s_f) = \log[0.5( (1 + \lambda)Y_{m,sm,sf} + (1 - \lambda)Y_{f,sm,sf} )] + \lambda_a \log[s_m^{1-\theta} s_f^\theta],

where $\lambda_a$ measures the weight on the utility from children’s ability.

affected by the number of children.
3.2 The College Decision (First Period)

At the beginning of the first period, an individual decides whether to go to college or not. The decision to go to college depends on the cost and the expected returns to college. A female individual chooses whether to attend college, \( s_f = 1 \) (high school) and \( s_f = 2 \) (college), given her conditional marriage probabilities \( P_f(s_m|s_f) \) and her individual cost of schooling \( D \), by solving

\[
\max_{s_f} \left\{ \sum_{s_m=0}^{2} U_f(s_m, s_f = 1) P_f(s_m|s_f = 1), \sum_{s_m=0}^{2} U_f(s_m, s_f = 2) P_f(s_m|s_f = 2) - D \right\}.
\]

Note that \( P_f(0|s_f) \) is the probability of being single. A male’s problem is defined analogously.

An individual is indifferent between going to college or not if the expected utility gain from going to college is equal to the effort cost \( D \). We define the threshold levels as

\[
\begin{align*}
D_f^* & \equiv \sum_{s_m=0}^{2} U_f(s_f = 2, s_m) P_f(s_m|s_f = 2) - \sum_{s_m=0}^{2} U_f(s_f = 1, s_m) P_f(s_m|s_f = 1), \\
D_m^* & \equiv \sum_{s_f=0}^{2} U_m(s_m, s_f = 2) P_m(s_f|s_m = 2) - \sum_{s_f=0}^{2} U_m(s_m, s_f = 1) P_m(s_f|s_m = 1).
\end{align*}
\]

Therefore a female with an idiosyncratic effort cost \( D \) chooses to go to college, \( s_f = 2 \), if and only if \( D < D_f^* \) and a male chooses \( s_m = 2 \) if and only if \( D < D_m^* \).

3.3 Distribution

Each individual receives a draw of effort cost, \( D \), in the first period. We assume that the individual’s learning ability, \( a \), affects the distribution of effort cost from which he/she draws. More specifically, we assume that the effort cost \( D \) is log-
normally distributed with mean $\mu(a)$ and variance $\sigma^2$, where $\mu(a)$ is decreasing in the ability level $a$. Recall from equation (3) that $a$ is determined by parent’s type, $a_{s^{-1}_m,s^{-1}_f}$, where $s^{-1}_j$ is parent $j$’s schooling. In each period, there are 4 different values of $a$. Let $\psi_g^c(s^{-1}_m = i, s^{-1}_f = j)$ denote the college attainment rates of individuals of gender $g$, conditional on parents’ education, which are calculated using cumulative distribution function of $D$ at $D^*_g$

$$\psi_g^c(s^{-1}_m = i, s^{-1}_f = j) = F[D^*_g|a_{i,j}].$$

Notice that although an individual born to any family type would make decisions based only on his/her gender and the value of effort cost $D$, the fraction of individuals born to each family type that go to college will depend on the parents’ type, because the parents’ type determines the average effort cost these individuals bear.

Let the total fraction of individuals of gender $g$ attending college be $\Phi^c_g$. Let $p^{-1}(s^{-1}_m = i, s^{-1}_f = j)$ be the fraction of fathers and mothers with education level $i$ and $j$, respectively. Thus the aggregate college attainment, $\Phi^c_g$, is the average of the conditional attainment rates weighted by parents education distribution.

$$\Phi^c_g = \sum_{i,j=1}^2 \psi_g^c(s^{-1}_m = i, s^{-1}_f = j) * p^{-1}(s^{-1}_m = i, s^{-1}_f = j).$$

4 Parameters

We calculate parents’ education distributions, marriage distributions, and earnings as inputs of the model. We compute the distribution of parent’s education from the PSID. Since the Current Population Surveys (CPS) cover longer periods and have a larger sample than the PSID, we use the CPS to estimate earnings and marriage
This section describes the estimation procedure and results of those parameters in detail.

### 4.1 Parents’ Background

We calculate the schooling distribution of our PSID sample’s parents. The sample description is presented in Section 2. We show its change over time in Figure 3. We observe that the number of college-educated parents increases over time. In 1980, 12% of individuals at age 25-34 had parents that were both college-educated, and 69% had parents that were both high school graduates. In 1996, the fractions change to 23% and 50% respectively.

![Figure 3: Parents’ education. Source: Authors’ calculations from the PSID data files.](image)

### 4.2 Marriage Rates

We estimate the probability of being married that each individual faces from the March supplement of the CPS 1964-2007. We define an individual as having college education if he/she completes more than 12 years of schooling. To be consistent with the PSID and the CPS provide similar pattern in college attainment.

To be consistent

---

12 The PSID and the CPS provide similar pattern in college attainment.

13 The CPS changed schooling classification in 1992. Prior to 1991, we have information on the number of grades attended and completed up to 18 years. After 1992, information on an individual’s
with our model, we exclude individuals with marital status as widowed, divorced or separated. Individuals in our sample are either never married or currently married. We define an individual as single if he/she has never married.

For each birth cohort, we first construct a pseudo-panel between age 18 and 65. In each pseudo-panel we construct, we calculate the life-cycle profile of marriage rate which is defined as the fraction of individuals that are married at each age. Usually not the entire life-cycle profile is observed.\textsuperscript{14} We use the 4th-order polynomial in age to estimate the life-cycle profile. Then we pick the probability of being married at age 35 (38) as a proxy of the average lifetime marriage probability for a typical female (male).\textsuperscript{15} Between 1980 and 1996, we compute the average marriage rate of those cohorts who were at age 25-34 in each year.

In the marriage market, there has been a decline of marriage rates for both genders. As Figure 4 shows, from 1980 to 1996 the marriage rate has decreased by 18 percentage points for high school males, 13 percentage points for college males, 15 percentage points for high school females, and 16 percentage points for college females.

\textsuperscript{14}For example, for a cohort born in 1970, the available CPS data only provide us marriage rate profile from age 18 to 37.

\textsuperscript{15}We thus capture the pattern that males usually marry late than females.
points for high school females, and 10 percentage points for college females.\textsuperscript{16}

We also calculate the probability of marrying each type of spouse conditional on being married. We use household and spousal identification information to match couples. Between 1980 and 1996, we compute the marriage distribution of those married individual who were at age 25-34 in each year.

![Figure 5: Probability of marrying a college spouse conditional on own gender and education](image)

Figure 5 confirms the well-known phenomenon that people do not marry randomly and there exists assortative matching (Becker (1973), Mare (1991), and Pencavel (1998)).\textsuperscript{17} A college educated person is more likely to marry a college educated spouse and benefits from the spouse’s earnings. Meanwhile, Figure 5 shows that as female college attainment rate increases, the probability of marrying a college female for a male, with or without college education, increases over time and the opposite happens for a female.

\textsuperscript{16}The marriage rate for males is lower than that for females for each cohort. This does not indicate that the marriage market is not clear at each point in time. Since males marry 3 year later than females, at each point in time, males marry females 3 cohort younger. Shifting males profiles by 3 years to the right reduces most of the differences in marriage rates by gender.

\textsuperscript{17}Benham (1974), Boulier and Rosenzweig (1984), Behrman, Rosenzweig and Taubman (1994), and Weiss (1997), point out that one’s own schooling can improve spousal schooling acquired in the marriage market, but it is difficult to conclude whether this effect is due to human capital accumulation within the household or assortative mating.
4.3 Earnings

We need to estimate expected lifetime earnings at each marriage status for an individual at the beginning of the life cycle. We do not observe wage for those who do not work. If labor force participation is determined stochastically by a process whose random unobservable component is correlated with unobservable in the wage function, a simple OLS regression is biased. To control for the selection-bias, we estimate the wage by a two-stage procedure: First we estimate equations of observed labor market participating status as functions of explanatory variables along with random disturbance terms representing unobservable factors. Then we specify and estimate equations of the logarithm of wage, controlling for participation selection.

4.3.1 Estimation Procedure

We estimate a regression function for each sub-sample of working individuals by gender as

\[ \log w_i = X_i \beta + \lambda_i \alpha + \eta_i, \]

where \( \log w_i \) is the logarithm of real hourly wage, and \( X \) is a vector of characteristics such as schooling, work experience, etc. \( \lambda \), the inverse Mills ratio, represents the selections terms on participation. The detailed explanation of the regression function is provided in Appendix 7.1.

We extend the Heckman (1979) and Lee (1978) two-stage estimation methods to this model to obtain consistent estimates. First, we estimate equations of observed labor market participating status as functions of explanatory variables along with random disturbance terms representing unobservable factors. Second, we use these estimates to construct the inverse Mills ratios. Then, we run OLS of log wage
equations on $X$, using the estimated inverse Mills ratios as additional regressors, as is specified in (11). Finally, we predict hourly wage for each individual using the following fitted equations

\[
\log \hat{w}_i = X_i \hat{\beta},
\]

(12)

where $\hat{\beta}$ is the consistent estimation of $\beta$.

4.3.2 Estimation Results

The model is estimated on March CPS from 1964 to 2007. We keep individuals at age 18 to 65 who are not in the armed forces, and not self-employed. To be consistent with the decision model, we restrict our attention to individuals who are either married or single (never married). Hourly wage is deflated to 2006 dollar using the CPI. Definitions of variables are given in Appendix 7.2. We run separate probit wage selection and log wage regression for each gender in each year. The reduced-form probit selection results in 2007 are provided in Appendix 7.3.

Estimated coefficients and asymptotic $t$-statistics of the wage equations in 2007 corrected for selections are found in Table 1. Estimated coefficients on education, experience, occupation dummies, race, and region dummies are similar to estimates from the typical wage equations found in the literature. Collage education attainments are generally more important for women’s wage than for men. Experience has more positive impact on men’s wage than on women’s wage.

Selectivity biases are particularly interesting. The coefficients of $\lambda$ (defined in equation (17) in Appendix 7.1) are positive and statistically significant for both men and women. Therefore, observed wage patterns of men and women are higher than the population mean pattern would have been. One should expect that individuals with higher wage potential should be more likely to participate in the labor force.
The estimation results confirm that individuals who expect to earn more are more likely to participate in the labor force.

4.3.3 Lifetime Earnings

The earning concept that is consistent with our model is the expected life-time earnings. To calculate the life-time earnings, first we use the following procedure to estimate the average life-cycle profiles of earnings from the CPS. For a typical individual who is in birth cohort $t$, with gender $g = \{f, m\}$, education of the husband $s_m = \{0, h, c\}$, and education of the wife $s_f = \{0, h, c\}$, we denote earnings at age $\textit{age}$ as $y_{g,s_m,s_f}^t(\textit{age})$. For each birth cohort $t$, we first construct a pseudo-panel between age 18 and 65. Then $y_{g,s_m,s_f}^t(\textit{age})$ is calculated as the product of mean predicted hourly wage (as in equation (12)) and mean annual hours worked of the particular gender, education, marital status, age, and cohort. In the pseudo-panels we construct,
usually not the entire life-cycle earnings profiles are observed.\footnote{For example, for a cohort born in 1960, the CPS 1964-2007 data only gives us estimates of earnings profile from age 18 to 47.} We use the following polynomial in age, $age$, to estimate the life-cycle earnings profile for each birth cohort by gender, education of both spouses

\begin{equation}
y_{g,s_m,s_f}^t(age) = \beta_{0,g,s_m,s_f} + \beta_{1,g,s_m,s_f} \cdot age + \beta_{2,g,s_m,s_f} \cdot age^2 + I(cohort = t) + \varepsilon_{g,s_m,s_f}^t(age),
\end{equation}

where $I(cohort = t)$ is a dummy for birth cohort $t$.

We calculate the life-cycle earnings by gender, marital status, and cohort using the estimated life-cycle earnings profiles. For a male who is in birth cohort $t$, with gender $g = \{f, m\}$, education of the husband $s_m = \{0, h, c\}$, and education of the wife $s_f = \{0, h, c\}$, we calculate total discounted life-cycle earnings at the beginning of his adult life, $Y_{m,s_m,s_f}^t$, as

\begin{equation}
Y_{m,s_m,s_f}^t = \sum_{age=18}^{65} \left( \frac{1}{1 + r} \right)^{age-18} y_{m,s_m,s_f}^t(age), \text{ if } s_m = h
\end{equation}

\begin{equation}
Y_{m,s_m,s_f}^t = \sum_{age=22}^{65} \left( \frac{1}{1 + r} \right)^{age-18} y_{m,s_m,s_f}^t(age), \text{ if } s_m = c
\end{equation}

where $r$ is the annual real interest rate and $y_{m,s_m,s_f}^t(age)$ is the annual real earnings at age $age = \{18, 19, ..., 65\}$.\footnote{We assume college students can not work, thus do not have earnings during age 18-21.} A female’s lifetime earnings are calculated analogously. In the calculation, we use predicted age-specific earnings from equation (13). Furthermore, an interest rate of $r = 4\%$ is used.

Between 1980 and 1996, we compute the average discounted life-cycle earnings of those cohorts who aged 25-34 in each year, by gender and marriage type. Figure 6 shows lifetime earnings for males at each marriage status. We observe that high school graduates on average have much less earnings than college graduates, regardless
of marriage status. We also notice that, married males on average earn more than never-married do. The marriage premium is the highest for those whose spouses have college degrees. Over time, lifetime earnings by cohort are decreasing slowly for males at all marital status, especially for married males.\footnote{Our finding is consistent with Kambourov and Manovskii (2005) who show that the life-cycle profiles of males’ earnings for younger cohorts are lower than those for older cohorts.}

Figure 7 shows lifetime earnings for females at each marriage status. We observe that high school graduates on average have much less earnings than college graduates, regardless of marriage status. However, unlike males, we do not notice that married females earn more than singles. The marriage premium for females is negligible among high school graduates and is in fact negative among college graduates. Over time, lifetime earnings for females are increasing gradually for married females, partially due to the increasing of female labor supply, partially due to the increasing of wages. On the contrary, lifetime earnings for single females are decreasing gradually, since the increasing of wages is offset by the decreasing of labor supply.\footnote{McGratten and Rogerson (2004) use Census and find decline of hours worked by single female.} Compared with lifetime earnings for males shown in Figure 6, females earn much less than males at the same marital status. The differences on earnings between married males and

---

Figure 6: Male’s lifetime earnings. Source: Authors’ estimation from the CPS data files.
married females are decreasing gradually over time.

Figure 7: Female’s lifetime earnings. Source: Authors’ estimation from the CPS data files.

5 Findings

Can the model replicate the change of college attainment that occurred between 1980 and 1996? To do this, we use data reported in Section 4 and estimate other parameters by matching aggregate and conditional moments on college attainment by gender in the data. We run counterfactual simulations to study the effects of different mechanisms on college attainment by comparing college attainments from each simulation with those in the benchmark.

5.1 Benchmark

We use calculated earnings, marriage distributions, and parent education distributions as inputs of the model and estimate the rest of the parameters to match observed aggregate and conditional attainment rates by gender. The estimated parameters are presented in Table 2. We find that $\lambda = 0$, indicating that couples are pooling income completely. $\theta_s$ is less than 0.5, indicating that fathers’ education affects children’s
ability more than mothers’ education, consistent with the observations from Figure 2 that the marginal effect of father’s education on children’s education is larger than that of mother’s.\textsuperscript{22} Effort cost distribution parameter $\mu_{1,1}$ corresponds to $\mu(a_{1,1})$, etc. The fact that $\theta_s$ is less than 0.5, implies that $a_{1,1} < a_{1,2} < a_{2,1} < a_{2,2}$. Since $\mu(a)$ is decreasing in $a$, the rank order of $\mu$’s is consistent with that of $a$’s.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference</td>
<td>$\lambda_a = 2.976, \lambda = 0$</td>
</tr>
<tr>
<td>Ability production</td>
<td>$\theta_s = 0.483$</td>
</tr>
<tr>
<td>Effort cost distributions</td>
<td></td>
</tr>
<tr>
<td>Means</td>
<td>$\mu_{1,1} = 0.592, \mu_{2,1} = 0.387$</td>
</tr>
<tr>
<td></td>
<td>$\mu_{1,2} = 0.457, \mu_{2,2} = 0.292$</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>$\sigma = 0.220$</td>
</tr>
</tbody>
</table>

Table 2: Parameters used in the benchmark model

Figure 8 compares college attainment rates from the model with those in the data. The model is able to generate the pattern that college attainments for females were lower in 1980 and higher in 1996 than those for males, as is observed in the data. In the model, female college attainment began to exceed that of males in 1988, one year later than observed in the data.

\textsuperscript{22}This is consistent with empirical study by Behrman and Rosenzweig (2002).
Figure 9 compares females’ college attainment rates conditional on parent’s education from the model with those in the data. The model is able to generate the pattern that a college-educated parent is substantially more likely to have a college-educated daughter than a non-college graduate, even after controlling for the education of the other parent. In our model, parents’ type determines the average effort cost these individuals bear. Thus the rank order of μ’s, \( \mu_{1,1} > \mu_{1,2} > \mu_{2,1} > \mu_{2,2} \), is critical to generate the rank order of school attainment by parent’s type.

![Figure 9: Female’s college attainment rates](image)

5.2 Counterfactual Simulations

We run counterfactual simulations to study the effects of different mechanisms on college attainment by comparing college attainments from each simulation with those in the benchmark. For each simulation, we keep the value of the variables that we want to focus fixed in the 1980 level and keep value of other variables as in the benchmark. Therefore the comparison between each simulation and the benchmark results will quantify the effect of those variables.

\(^{23}\)Similar patterns hold for males and the results are available from the authors upon request.
5.2.1 Parents’ background

In the first simulation, we investigate the intergenerational schooling effects. The results are shown in Figure 10. When parents’ schooling distribution is fixed at 1980 level, college attainment would drop by 7.6 and 7.0 percentage points in 1996 for males and females, respectively. In the benchmark model, the gradual transformation of family schooling composition acts as a propagation mechanism in changing college attainment: as the number of college-educated parents increases, so does the proportion of children with high ability, which then helps to increase the attainment rate of the children generation. This propagation mechanism seems to affect female and male in the similar magnitude, thus, similar as in the benchmark, we notice the reversal of college attainment at the same year. In summary, parental education is an important source of the increase in college attainment, but is not about to account for the reversal of the gender gap.

![Figure 10: No change in parents distribution since 1980](image)

5.2.2 Earnings

In the benchmark model, $\lambda = 0$ implies full income pooling between spouses. Therefore for married couple, the relevant concept of earnings is household lifetime
earnings. To be consistent with our model, for singles, we compute the ratio of life-cycle earnings between college and high school males and females. For married females, we compare earnings of households in which the wife has but the husband does not have college education with earnings of households where both spouses are high school graduates. For married males, we compare earnings of households in which the husband has but the wife does not have college education with earnings of households where both spouses are high school graduates.\textsuperscript{24}

Figure 11 presents the earnings return to college by gender and marital status. Several patterns are observed from Figure 11 over the period 1980-1996. First, the earnings return to college is increasing for both genders and for all marital status.\textsuperscript{25} Second, the earnings return to college is higher for single females than for single males. In 1980, single college-educated males had 67% more earnings than single high school males, while single college-educated females had 102% more earnings than single high school females. This pattern remains unchanged over time. Third, the earnings return to college is similar for married females as for married males. For a typical household in 1980 in which neither spouse attained college, having the man going to college gives the household 24% more earnings, while having the female going to college gives 25% more earnings.

The second simulation analyzes the case when there is no change in earnings since 1980. The results are shown in Figure 12. Both male and female attainment rates would drop by 9.9 and 11.1 percentage points in 1996 for males and females, respectively. This indicating that the increasing returns to college in the labor

\textsuperscript{24}We also experienced comparing earnings of households where both spouses are college graduates with earnings of households in which the wife has but the husband does not have college education, the returns are only slightly higher and the overtime trends are almost identical.

\textsuperscript{25}Using cross-sectional earnings or wages, many authors have documented recent increases in the earnings return to college (See, for example, Juhn, Murphy and Pierce (1993), Katz and Murphy (1992), Card and DiNardo (2002), and Eckstein and Nagypál (2004)). Our measure using lifetime earnings gives similar results.
market for those cohorts, as is shown in Figure 11, have important impact on college attainment for those cohorts. The change of earnings has larger effect on college attainment for female than that for male. This is due to the fact that, over time, the earnings return to college for single female has been increased more substantially than that for male. Similar as in the benchmark, we notice the reversal of college attainment at the same year. Thus the change of earnings over time is not about to account for the reversal of the gender gap in college attainment.

Figure 12: No change in earnings since 1980
5.2.3 Marriage market

The next two simulations try to isolate the effects of changes in the marriage market. Without change in the rates of marriage, as shown in Figure 13, both males and females would reach higher college attainment in 1996 and female's college attainment is always lower than that of males. As is shown in Figure 11, the earnings returns to college in the labor market is higher for singles than for married couples. However, married individuals receive additional benefit from college through increasing their children’s learning ability. Under our parameters, the return from children for married couples dominates their lower return in the labor market, thus the returns to college increase with marriage rate. As marriage rate declines, returns to college decrease and so does college attainment. Moreover, as marriage rates decline female college attainment decreases less than that of males. This is because single females receive larger return to college in the labor market than single males. Moreover, under our parametrization, $\theta_s$ is less than 0.5, therefore married females receive smaller return to college from children than married males. The declining of marriage rate decreases college attainment for females less than that for males. This shows that changes in marriage rates are crucial in accounting for the reversal in gender gap in college attainment.

In the fourth simulation we fix the conditional marriage probabilities as in 1980 and keep the marriage rates in the data. The results are shown in Figure 14. The college attainment in 1996 would be 5.0 and 3.8 percentage points lower for males and females. Therefore the change of conditional marriage probabilities plays a role in accounting for the increasing in college attainment for both genders. The change of marriage rate has larger effect on college attainment for male than that for female. This is in part due to the fact that, over time, the probability of marrying a college spouse for male has been increased quite substantially while that for female has
barely changed. In the benchmark, over time males benefit more from marring college spouses, thus college attainment increases more, compared with the experiments when marriage probabilities are fixes as in 1980.

![Figure 13: No change in single rates since 1980](image)

![Figure 14: No change in conditional marriage probability since 1980](image)

### 6 Conclusions

We develop a dynamic model that incorporates marriage, education and income to study the effects of changes in relative earnings, changes in parental education, and changes in the marriage market on changes in college attainment by gender. We find
that the increases of parental education and relative earnings between college and high school persons have important effects on the increase in college attainment for both genders, while the decrease of marriage rates is crucial in explaining the reversal of gender gap in college attainment.

There are several directions in which this work can be extended. First, we abstract from divorce and out-of-wedlock birth and assume a single individual does not have children. The divorce rate has been stabilized since the early 1980s but out-of-wedlock birth has been increasing. Those changes in family structure might affect female and male’s college attainment decision differently. Secondly, we assume earnings are exogenous. An extension that we wish to explore is the relationship among college attainment, marriage and labor supply for both genders. Even though labor earnings are sacrificed, a parent who stays at home and takes care of children would contribute to the household by increasing the ability of children. We plan to study these issues in future work.

References


7 Appendix

7.1 Estimation procedure

Consider the following wage function on a sample of working men and women:

$$\log w_i = X_i \beta + \mu_i$$

where $\log w_i$ is the logarithm of hourly wage, and $X$ is a vector of characteristics such as schooling, work experience, etc. It is argued, however, that the sample of employed workers is not a random sample, and that this selectivity might bias the coefficients. Formally, we can write down a participation equation

$$E_i = 1 \text{ if } Z_i \gamma + \epsilon_i \geq 0,$$
$$E_i = 0 \text{ if } Z_i \gamma + \epsilon_i < 0,$$

where $Z$ includes variables that predict whether or not a person works. Therefore the probability of an individual working is

$$\Pr(E_i = 1) = \Pr(\epsilon_i \geq -Z_i \gamma) = \Phi\left(\frac{Z_i \gamma}{\sigma}\right),$$

where $\sigma^2$ is the variance of $\epsilon_i$, and $\Phi(\cdot)$ is cumulative distribution function of the standard normal.

The selectivity problem is apparent by taking expectations of the wage function over the sample of employed workers

$$E(\log w_i|E_i = 1, X_i) = X_i \beta + E(\mu_i|\epsilon_i \geq -Z_i \gamma).$$

Suppose $\mu_i$ and $\epsilon_i$ are jointly normally distributed, let $\sigma_{\mu,\epsilon}$ be the covariance between $\mu_i$ and $\epsilon_i$. We can now write

$$E(\mu_i|\epsilon_i \geq -Z_i \gamma) = \frac{\sigma_{\mu,\epsilon} \phi(Z_i \gamma/\sigma)}{\sigma} \Phi(Z_i \gamma/\sigma),$$

where $\phi(\cdot)$ is the standard normal density. When $\sigma_{\mu,\epsilon}$ is not zero, selectivity bias occurs. To estimate the potential wage consistently, we need to add the selection term (the inverse Mills ratio)

$$\phi(Z_i \gamma/\sigma) \quad \Phi(Z_i \gamma/\sigma) \equiv \lambda_i$$

in the OLS regression as in

$$\log w_i = X_i \beta + \lambda_i \alpha + \eta_i.$$
7.2 Definitions of Variables for Tables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Respondent’s age.</td>
</tr>
<tr>
<td>Age²</td>
<td>Square of Age.</td>
</tr>
<tr>
<td>HI</td>
<td>Dummy variable: 1 if respondent is a high school dropout.</td>
</tr>
<tr>
<td>HG</td>
<td>Dummy variable: 1 if respondent is a high school graduate.</td>
</tr>
<tr>
<td>SC</td>
<td>Dummy variable: 1 if respondent has some college education.</td>
</tr>
<tr>
<td>CG</td>
<td>Dummy variable: 1 if respondent is a college graduate.</td>
</tr>
<tr>
<td>Exp</td>
<td>Respondent’s years of work experience.</td>
</tr>
<tr>
<td>Exp²</td>
<td>Square of Exp.</td>
</tr>
<tr>
<td>Black</td>
<td>Dummy variable: 1 if respondent is black.</td>
</tr>
<tr>
<td>Marry</td>
<td>Dummy variable: 1 if respondent is married.</td>
</tr>
<tr>
<td>Nchild</td>
<td>Number of own children in household.</td>
</tr>
<tr>
<td>Nchlt5</td>
<td>Number of own children under age 5 in household.</td>
</tr>
<tr>
<td>Northeast</td>
<td>Dummy variable: 1 if household is located in northeast region.</td>
</tr>
<tr>
<td>Midwest</td>
<td>Dummy variable: 1 if household is located in midwest region.</td>
</tr>
<tr>
<td>South</td>
<td>Dummy variable: 1 if household is located in south region.</td>
</tr>
<tr>
<td>West</td>
<td>Dummy variable: 1 if household is located in west region.</td>
</tr>
<tr>
<td>Metro</td>
<td>Dummy variable: 1 if household is located in a metropolitan</td>
</tr>
<tr>
<td>Manager</td>
<td>Dummy variable: 1 if respondent is a manager or professional.</td>
</tr>
<tr>
<td>Whitecollar</td>
<td>Dummy variable: 1 if respondent has white-collar occupation other than those in management.</td>
</tr>
<tr>
<td>Bluecollar</td>
<td>Dummy variable: 1 if respondent has blue-collar occupation.</td>
</tr>
<tr>
<td>λ</td>
<td>See equation (17), based on estimates in Table 3.</td>
</tr>
</tbody>
</table>

7.3 Estimation results: probit selection

The reduced-form probit selection rules in equations (16) is estimated in each year for men and women, respectively. We estimate these probits year by year because there are some evidences that how individuals select themselves into workforce have shifted over time (Mulligan and Rubinstein (2007)). Table 3 presents estimated coefficients and asymptotic t-statistics of the reduced form participation probit for 2007.\(^{26}\) Our findings are generally in accord with previous research. Specifically, we find that education attainment has a positive and statistically significant impact on the probability of participation for both men and women. The probability of working increases in age at a decreasing rate for both men and women. Black men are less likely to participate compared with non-blacks. Men who are married or have children are more likely to participate, even though the effect of number of children is not statistically significant. Married women and women with children are less likely to participate.

\(^{26}\)Estimates for other years are available from the authors.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Males Coefficient</th>
<th>Males t</th>
<th>Females Coefficient</th>
<th>Females t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.5929</td>
<td>-41.75</td>
<td>-2.6571</td>
<td>-43.10</td>
</tr>
<tr>
<td>HG</td>
<td>0.3134</td>
<td>15.67</td>
<td>0.5029</td>
<td>24.77</td>
</tr>
<tr>
<td>SC</td>
<td>0.3882</td>
<td>18.68</td>
<td>0.6520</td>
<td>32.05</td>
</tr>
<tr>
<td>CG</td>
<td>0.7044</td>
<td>31.09</td>
<td>0.8212</td>
<td>38.86</td>
</tr>
<tr>
<td>Age</td>
<td>0.1627</td>
<td>46.82</td>
<td>0.1448</td>
<td>41.89</td>
</tr>
<tr>
<td>Age²</td>
<td>-0.0022</td>
<td>-52.07</td>
<td>-0.0018</td>
<td>-44.29</td>
</tr>
<tr>
<td>Black</td>
<td>-0.3328</td>
<td>-16.14</td>
<td>-0.0018</td>
<td>-0.10</td>
</tr>
<tr>
<td>Marry</td>
<td>0.4641</td>
<td>22.41</td>
<td>-0.0499</td>
<td>-2.91</td>
</tr>
<tr>
<td>Nchild</td>
<td>0.0396</td>
<td>4.82</td>
<td>-0.0800</td>
<td>-12.86</td>
</tr>
<tr>
<td>Nchlt5</td>
<td>0.0315</td>
<td>1.79</td>
<td>-0.2708</td>
<td>-22.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of obs.</td>
<td>48,145</td>
<td>51,315</td>
</tr>
<tr>
<td>-2 ln(likelihood ratio)</td>
<td>8285.05</td>
<td>5252.96</td>
</tr>
<tr>
<td>$\chi^2$ degree of freedom</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3: Participation Selection Rules: Probit Analysis (CPS 2007)