Washington University in St. Louis
Washington University Open Scholarship

Spring 5-15-2018

Understanding Child Maltreatment Report Risks as a Function of Age, Socioeconomic Status, Race, and Neighborhood

Hyunil Kim
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Social Work Commons

Recommended Citation

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Economics

Dissertation Examination Committee:
Juan Pantano, Chair
George-Levi Gayle, Co-Chair
Ian Fillmore
Limor Golan
Carl Sanders

Essays on Educational Choices
by
Sie Won Kim

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2018
St. Louis, Missouri

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply grateful to my advisor Juan Pantano for guiding me through my graduate studies. Without his guidance this dissertation would not have been possible. I also would like to thank George-Levi Gayle and Carl Sanders for their continued support from the beginning of my studies. I am also grateful to Limor Golan and Ian Fillmore for their time and helpful comments.

Special thanks to the Department of Economics for the generous academic and financial support I received throughout my graduate studies. I also would like to thank the National Center for Education Statistics for providing me the data set used in this dissertation. I would like to express my thanks to my colleagues, Duksang Cho, Hee Chun Kim, Minho Kim, Jungho Lee, Sunha Myong, and Jaevin Park.

I would like to express my gratitude to my family members including my parents, my sister, and my parents-in-law for their support and prayers. Especially, I would like to thank my baby boy, Louis, who has given me the joy of my life.

Lastly, I owe my deepest gratitude to my wife, Jeongmin, for her unconditional love.

Sie Won Kim

Washington University in St. Louis

May 2018

ABSTRACT OF THE DISSERTATION

Essays on Educational Choices

by

Sie Won Kim

Doctor of Philosophy in Economics

Washington University in St. Louis, 2018

Doctor Juan Pantano, Chair

Professor George-Levi Gayle, Co-Chair

Chapter 1 evaluates two policy counterfactuals designed to promote access to community colleges: tuition subsidies for community college students and easier transition from two- to four-year colleges. I estimate a structural model of employment and college choices, including two-year colleges that provide post-secondary education at a lower cost and an opportunity to transfer to four-year colleges. First, I find that full tuition subsidies at two-year colleges increase the number of students initially enrolled in two-year colleges by 72.1 percent, and mostly attract students who would have not attended college otherwise. This increase in two-year college enrollment translates into an increase of $20,812 in the present value of lifetime income, which is larger than the average cost of $12,550. However, this does not translate into an increase in the number of transfers and four-year college degrees completed. Second, when two-year colleges provide better preparation for students for their transition to four-year institutions, I find that the number of students initially enrolled in two-year colleges increases by 19 percent, the transfer rate increases by 27.5 percent, and the completion rate at four-year colleges increases by 3.1 percent. The average present discounted value of lifetime income also increases by $16,589.

Chapter 2 investigates the effect of the quality of public schools on students' enrollment decisions and their academic achievements in college. I use the National Education Longitudinal Study of 1988 to create a measure of the quality of high schools by taking the average of the standardized test scores of students in each school. I estimate a model of college enrollment and continuation choices in which the quality of high schools could affect students' educational outcomes. I find that the quality of high schools has a positive effect on the probability of receiving a high SAT score and the probability of being admitted to high-quality colleges. The counterfactual results show that centralizing public education results in a higher college enrollment rate of students from low-quality school districts. When a policy imposes a lower bound for high school quality, I find that college enrollment increases by 9.1 percent and college completion increases by 7.8 percent for students attending the lowest quality quartile schools. Lastly, a 10 percent increase in the ability of students from the lowest 25th percentile quality schools increases enrollment and completion by 59.5 percent and 33.3 percent, respectively. The counterfactual results suggest that improvement in the quality of public schools has a modest effect on college enrollment and completion compared to a direct increase in the ability of the students.

Chapter 1

The Effect of Promoting Access to Community Colleges on Educational and Labor Market Outcomes

1.1 Introduction

In 2012, 7.2 million undergraduate students were enrolled in two-year colleges across the U.S., and 10.6 million were enrolled at four-year colleges. Community colleges have an open admissions policy and are typically less expensive than four-year colleges. Post-secondary education at community colleges is an affordable option for higher education and a stepping stone to a bachelor's degree. Approximately 41 percent of high school graduates who pursue higher education initially attended two-year colleges.[1]

Two types of policies promote access to community colleges to improve college attendance and degree completion. The first type affects access to higher education by making two-year colleges more affordable. For example, in 2015 the White House proposed "America's College Promise," which would provide tuition-free community college education to qualified students. This policy assumes that by completing two years at a community college students

---

[1]Digest of Education Statistics 2013, Table 305.10.

would earn half of the credits needed for a bachelor's degree.[2] The second type of policy makes it possible for community colleges to better prepare students for the transition to a four-year college. One reform would require two-year colleges to offer more courses at a four-year college level. Another example is a statewide transfer agreement between community colleges and four-year schools. For example, the California Community Colleges have transfer agreements with both the California State University and the University of California systems making it easier for students to transfer from two- to four-year colleges.

Although both types of policies are expected to increase community college enrollment, it is unclear how increased initial enrollment translates into long-term degree completion and labor market outcomes. Two-year colleges provide post-secondary education to students who would not have continued their education otherwise. Community college graduates benefit from the higher earnings associated with additional years of education. Moreover, a two-year college student could transfer to a four-year college, earn a bachelor's degree, and benefit from even higher expected earnings. At the same time, community colleges could divert students away from four-year colleges. Two-year college attendance could be associated with a lower four-year college completion rate relative to four-year college attendance.

To quantify the net effect, I develop and estimate a structural model of students' educational and employment choices. In the model, students can enroll in two- or four-year colleges, transfer, drop out, or work full-time. The probability of a student completing a four-year college degree is given as a function of years spent at both two- and four-year colleges. I estimate the model with simulated method of moments using the data from the National Longitudinal Survey of Youth 1997.

Given these estimation results, I simulate two policy counterfactuals designed to affect access to community colleges. Outcomes of interest are enrollment choices, degree completion rates, and wages. First, I reduce the cost of two-year colleges. Second, I improve the

---

[2]Executive Office of the President, "America's College Promise: A Progress Report on Free Community College." 2015.

preparation for transition to a four-year college by setting a year at a community college equivalent to a year at a four-year school for a bachelor's degree completion probabilities. I track which group of students attends community colleges and then compare the changes in their educational choices to baseline outcomes.

The findings show that two-year college subsidies increase the percentage of students initially enrolled in two-year college by 33.7 percent and 72.1 percent with a subsidy rate of 0.5 and 1, respectively. At the same time, the subsidies decrease the percentage of students who work initially after high school by 6.9 percent and 13.5 percent with a subsidy rate of 0.5 and 1, respectively. Lower community college tuition mostly attracts students who would not have attended college otherwise. This increase in two-year college enrollment translates into an increase in students' earnings of $10,854 or $20,812 in the average present discounted value of lifetime income depending on the subsidy rate. Benefits of both subsidies exceed the costs; however, it does not translate into an increase in the number of transfers and students' bachelor's degree attainment. In the second policy counterfactual, when two-year colleges better prepare students for the transition to a four-year college, the percentage of students initially enrolled in two-year college increases by 19 percent, the transfer rate increases by 27.5 percent, and the bachelor's degree completion rate at age 26 increases by 3.1 percent. As a result, the average present discounted value of lifetime income increases by $16,589. The counterfactual results suggest that lower costs increase enrollment, but not the number of students earning four-year degrees. In contrast, when community colleges better prepare students for the transition to four-year colleges, the increase in enrollment is larger than in the half-subsidy case, helping more students attend community colleges to eventually earn bachelor's degrees.

The rest of the paper is structured as follows. Section 1.2 reviews related literature. Section 2.2 describes the data and descriptive statistics. Section 2.3 develops the structural model. Section 1.5 discusses estimation strategy. Section 1.6 presents the estimation results

and the model fit. Section 1.7 discusses the policy counterfactuals and welfare analysis. Section 2.7 concludes.

1.2 Literature Review

Researchers often separate the effects of community colleges into the democratization effect and the diversion effect. The former affects two-year college students who would not have attended college otherwise, whereas the latter affects two-year college students who would have attended a four-year college otherwise. Rouse (1995) uses the distance to two-year and four-year college and the average state two- and four-year college tuition as instrumental variables. By comparing the coefficients on two- and four-year attendance, she finds that the magnitude of the democratization effect is larger than that of the diversion effect. Long and Kurlaender (2009) use the propensity score matching method controlling for students' educational expectations. Recent papers explore case studies in specific states to estimate the effect of community colleges on educational outcomes. Goodman et al (2015) exploit SAT score thresholds in Georgia and find that the diversion effect on bachelor's degree attainment exists for those who were drawn to two-year colleges. Denning (2014) uses the expansion of community college taxing district in Texas as an instrumental variable and reports that a $1,000 decrease in tuition results in 7.1 percentage points increase in community college enrollment. Local average treatment effects are difficult to compare since the treatment groups are different. To compare the effects of policies promoting access to community colleges on various outcomes, I estimate a structural model of employment and schooling.

The model used in this paper is closely related to the discrete choice models in which individuals choose between schooling and employment options. Keane and Wolpin (1997) use a dynamic structural model to study the schooling and occupational choices of young men. Arcidiacono (2005) estimates a structural model including application, enrollment and

college major choices. Epple, Romano and Sieg (2006) and Fu (2014) estimate structural equilibrium models to understand tuition, applications, admissions, and enrollment decisions jointly. Other papers use structural models including community colleges as one of the schooling options. Johnson (2013) finds that tuition subsidies have larger effect on educational attainment than an equal amount of the loan limit increase. A tuition subsidy of $1,500 to two- and four-year colleges increases overall college enrollment by 6.4 percentage points and bachelor's degree completion rate by 5.3 percentage points. Trachter (2015) reports that a 15 percent decrease in two-year college tuition results in an increase in two-year college enrollment from 15.7 percent to 26.6 percent. At the same time four-year college enrollment drops from 27.4 percent to 17.3 percent.

To understand the educational and employment patterns observed in the data, I examine the returns to post-secondary education. Researchers have documented that college education is associated with higher earnings. In particular, researchers exploited exogenous variation to estimate the returns to schooling including two-year colleges. Such variations include distances to two- and four-year colleges (Kane and Rouse (1995)) and an admissions threshold (Zimmerman (2014)). Kane and Rouse find that an associate's degree is associated with 7 percent higher annual earnings than high school graduates, and a bachelor's degree is associated with 28 percent higher annual earnings than high school graduates. Zimmerman uses a regression discontinuity to report that if a student is just above the admissions threshold at a public university in Florida, earnings increase by 22 percent between 8 and 14 years after high school graduation.

Community colleges serve as a stepping stone to a bachelor's degree through an opportunity to transfer to a four-year institution. A statewide articulation policy facilitates this transfer process for community college students. Researchers have investigated the effect of this policy on students' educational outcomes. Anderson et al (2006) find that a statewide articulation policy has little effect on students' transfer rates, but they acknowledge that

they focus on small number of states and these policies may need time to take effect on the transfer rates. Roksa and Keith (2008) argue that statewide articulation policies should be evaluated by outcomes that occur after students transfer to four-year institutions such as number of credits transferred, time to a bachelor's degree, and completion of a bachelor's degree. Authors report that statewide articulation policies have no effect on relevant outcomes, but they explain that this outcome could be due to the data limitations and variation in articulation policies among states.

This paper contributes to the existing literature by comparing the effects of two policy counterfactuals, which promote access to community colleges. I quantify the effects of subsidizing tuition at community colleges and providing smoother transitions to four-year institutions in one model. This paper reports the effects on enrollment at two- and four-year colleges, a bachelor's degree attainment, and labor market outcomes and the group of students affected by each type of policy.

1.3   Data

I use the panel data from the National Longitudinal Survey of Youth 1997 (NLSY97). The NLSY97 is a longitudinal study of a nationally representative sample of youth ranging in age from 12 to 18 at the end of 1996. The NLSY97 surveyed 9,021 youth born between 1980 and 1984. I use the data from round 1 (1997-1998) through round 15 (2011-2012). In round 15, respondents ranged in age from 26 to 32. The NLSY97 has variables on schooling including a type of college, an enrollment status, and a four-year college degree status. I obtain information on demographic characteristics such as race, gender, and age. In addition, I have information on region, family income, parents' education, and the Armed Services Vocational Aptitude Battery (ASVAB) math-verbal percentile score. Table 1.4 summarizes the variables and the data sources. I use the school enrollment status on October each year

to determine whether an individual is enrolled at a two- or four-year college. The highest degree received prior to the start of each academic school year is used to determine whether or not an individual has a four-year college degree. The ASVAB math-verbal percentile score is used as a measure of ability. I obtain gross household income in thousands of dollars when each respondent was 18 years old. Total income from wages and salary each year is selected to construct the wage profile.

I restrict the sample to male high school graduates between 18 and 26 years old. I focus on males because of their strong labor market participation as the model considers only their employment choices before 26. I choose age 26 as the end of the sample, because few people complete a four-year college degree after 26. The completion rate at four-year colleges increases by only 0.01 percentage points between ages 25 and 26. I exclude high school dropouts and observations with missing data on the ASVAB math-verbal percentile score. The final sample size is 751 with 6,008 individual-year observations.

To complement the NLSY97 wage data, I use the National Longitudinal Survey of Youth 1979 (NLSY79). Since the NLSY97 sample is relatively young, it has limited wage data after college graduation. The most recent round 16 (2013-2014) interviewed youth of ages from 28 to 34. It is difficult to estimate the expected lifetime income and to project out into an individual's retirement age of 65. I estimate the expected lifetime income using the data from the NLSY79 and combine it with the NLSY97.

For college-level data, I use the Integrated Postsecondary Education Data System (IPEDS). The IPEDS contains surveys conducted by the U.S. Department of Education's National Center for Education Statistics. It reports information on enrollment, program completion, graduation rates, faculty and staff, finances, institutional prices, and student financial aid. I use the tuition data to compute state-average tuition of two- and four-year colleges. I use the state of residence in high school to assign the tuition cost to a student.[3] In addition, I

---

[3] Due to the data availability, I aggregate the states into four census regions.

compute state-average instructional expenditure for two- and four-year colleges that I use in the cost-benefit analysis.

Figure 1.1 shows the proportion of school and work choices over time for the sample. Initially, 15.45 percent of high school graduates attend two-year colleges and 35.55 percent attend four-year colleges. The remaining 49 percent of high school graduates join the labor market. Students who initially attend two- and four-year colleges enter the labor market after a few periods. Labor market participation increases gradually and reaches 86.42 percent at age 26. To explain the steady increase in work and a decrease in schooling over time, one year of schooling would be associated with an increase in future earnings which compensates for the opportunity cost of schooling.

Table 2.1 presents the descriptive statistics of key variables in the model. I group the full sample into three groups – a work, a two-year, and a four-year – based on students' initial choices. As the level of schooling increases, from the first column to the third column, the ASVAB math-verbal percentile score, the log wage at age 26, and the family income are higher. The proportion of non-Hispanic and non-black students is higher at four-year schools than at two-year schools and for high school graduates. Initial enrollment choice is correlated with the number of years spent at two- or four-year colleges and a bachelor's degree completion. Those who enrolled at a two-year college spend 2.76 years at a community college and 0.84 years at a four-year college. Those who enrolled at a four-year college spend 4.21 years at a four-year institution and 0.25 years at a two-year institution. Those who choose not to go to college spend less than a year at either type of college. As a result, the fraction of a bachelor's degree holder for initial two-year college attendees is 17 percent compared to 60 percent for initial four-year college attendees. These differences show that student characteristics are different at two- and four-year colleges, and two types of schools would prepare students differently for a four-year college degree completion.

Table 1.3 shows the Mincer regression estimates of the NLSY79 sample. I use the estimates to construct the present discounted value of lifetime income and use it as the terminal value as a function of number of years spent at college, a four-year college degree, work experiences, and race.

1.4 Model

1.4.1 Overview

Individual i chooses from options $j \in \{2yr, 4yr, work\}$ at $t \leq T$. Individuals can either enroll at two-year or four-year colleges, work full-time, drop out, or transfer between schooling options. At $t = T + 1$, everyone enters the labor market. I assume that everyone works after $t = T + 1$, and receives a present discounted value (PDV) of lifetime income based on the total number of years spent at college, a bachelor's degree status (BA), and student's ability. The model allows 3^8 choice combinations. Only choice restriction of the model is that students no longer make decisions after getting a four-year degree. Four-year college completion probabilities depend on the number of years spent at two- and four-year college and a student's ability. Students have correct beliefs about the probabilities of completing a four-year college.

1.4.2 Preferences

Utility of a schooling option $j \in \{2yr, 4yr\}$ at time t is

$$U_{jt} = \bar{U}_{jt}\left(X_{it}, Z_j; \beta_j\right) + \epsilon_{jt}$$

$$U_{jt} = \beta_{0j} + \beta_{1j}A_i + \beta_{2j}tuition_j + \beta_{3j}\mathbf{1}\left[D_{jt-1} = 0\right] + \beta_{4j}\mathbf{1}\left[type_i = 2\right] + \epsilon_{jt} \qquad (1.1)$$

where A_i is the ability measure, and $tuition_j$ is the average tuition level at school j in i's census region. $\mathbf{1}\left[D_{jt-1} = 0\right]$ is an indicator which equals 1 if an individual i attended different type of school at $t-1$. β_{3j} captures the transfer costs. Unobserved type $type_i = \{1, 2\}$ affects the intercept of the utility. I assume that ϵ_{jt} follows a Type 1 Extreme Value distribution. This utility function allows for different parameters for two- or four-year colleges.

Utility of working full-time ($j = work$) at time t is

$$U_{work,t} = \rho \ln\left(W_t\left(X_{it}; \alpha\right)\right) + \epsilon_{work,t}$$

$W_t\left(X_{it}; \alpha\right)$ represents the annual wage at t as a function of individual characteristics. I assume that $\epsilon_{work,t}$ follows a Type 1 Extreme Value distribution. Wage function is

$$W_{it} = \exp\left(\alpha_0 + \alpha_1 A_i + \alpha_2 s_{2yr,t} + \alpha_3 s_{4yr,t} + \alpha_4 BA_{it} + \mu_t\right) \qquad (1.2)$$

where A_i is the ability measure and s_{jt} is the years of schooling spent at college type j until t including choice at t $\left(s_{jt} = \sum_{t'=1}^{t} D_{jt'}\right)$. BA_{it} is a dummy variable which equals 1 if i has a bachelor's degree at the beginning of t.

1.4.3 Unobserved Heterogeneity

It is likely that there exist information only observed by a student when making the decisions in the model. Example of such information is the levels of aspiration or motivation. To account for such information I allow two unobserved types, $type_i \in \{1, 2\}$. Type-specific intercepts affect the utility of schooling option j. I assume that the unobserved type affects the wage equation and probability of BA completion indirectly through schooling choices.

Unobserved types could explain variation in choices of students who have similar observed characteristics.

1.4.4 Probability of Four-year College Degree Completion

Students attending four-year colleges face four-year college degree completion probabilities. The probability of receiving a BA degree depend on ability, and years of schooling at two- and four-year colleges. The probability of receiving a BA degree at time $t+1$ is

$$Pr\left(BA_{t+1} = 1|D_{4yr,t} = 1\right)$$

$$= \Lambda\left(\lambda_0 + \lambda_1 A_i + \lambda_2 s_{2yr,t} + \lambda_3 s_{4yr,t}\right) \cdot \mathbf{1}\left[s_{4yr,t} \geq min_{4yr}\right] \tag{1.3}$$

where $\Lambda\left(\cdot\right)$ is the logistic function. As in the wage equation, A_i is the ability measure and s_{jt} is the years of schooling spent at college j until t including choice at t. Students need at least min_{4yr} years at $4yr$ to graduate four-year college. I assume that students need to attend at least two years at four-year colleges to complete, $min_{4yr} = 2$. Coefficients on years of schooling, λ_2 and λ_3, determine how two types of colleges contribute to completing a BA degree. Estimates would be similar if there is no difference in two- and four-year college's academic preparation for a BA degree attainment.

1.4.5 Decision Problem

The value of each option j at t, the alternative-specific value function is

$$V_{jt}\left(\Omega_t\right) = U_{jt} + \beta E\left[V\left(\Omega_{t+1}\right)|\Omega_t, D_{jt} = 1\right]$$

where

$$V\left(\Omega_t\right) = \max_j \left\{V_{jt}\left(\Omega_t\right)\right\}$$

For $t \leq T$, individual i chooses among options $j \in \{2yr, 4yr, work\}$. $D_{jt} = 1$ if i chooses option j at t. If i receives a four-year degree $(BA_{it} = 1)$ then i starts working in $t + 1$. For $t > T$, working is an absorbing state, and I assume that everyone works until T_{retire}. A terminal value function $PDV_{T+1}\left(\Omega_{T+1}\right)$ determines the present discounted value of lifetime income associated with working from $T + 1$ to T_{retire}. To complement the wage data in the NLSY97, I use the NLSY79 to compute the wage profiles of individuals over 26.

A student's choice depends on various model parameters. If the coefficient on tuition is high it is less desirable to attend high-tuition schooling option, a four-year college. If the coefficient on previous choice indicator variable increases, I would observe less transfers between schooling options and between school and work choices and more persistence in two- and four-year colleges. For the probabilities of BA completion, the coefficients on the number of years spent at two- and four-year colleges determine the effect of years spent at two- and four-year colleges on a BA completion probability. If the coefficient on two-year is large enough relative to the coefficient on four-year then enrollment at a community college would be higher since two-year attendance does not hurt the chance of BA completion.

The state space is

$$\Omega_t = \left\{\{s_{jt-1}\}_j, BA_{it}, D_{jt-1}, type_i, X_i\right\}$$

Evolution of the number of years spent at college j, s_{jt} between $t-1$ and t is deterministic depending on D_{jt}. The number of years spent at college j evolves as

$$s_{jt} = s_{jt-1} + D_{jt}$$

for $j \in \{2yr, 4yr\}$. Probabilities shown in (1.3) determine the BA degree outcome BA_{it}. D_{jt-1} is the previous period's choice. Unobserved type $type_i$ and observed characteristics X_i including the ability measure remain the same over time. In summary, individual i computes alternative-specific value functions at t and chooses option j with the highest value given $\Omega_t = \left\{ \{s_{jt-1}\}_j, BA_{it}, D_{jt-1}, type_i, X_i \right\}$. Individual i repeats this decision process until $t = T$.

1.4.6   Dynamic Problem Solution

I solve the dynamic model backwards starting from the last period T. At $t = T$, the value of option j is

$$V_{jT} \left(\Omega_T \right) = U_{jT} + \beta E \left[PDV_{T+1} \left(\Omega_{T+1} \right) | \Omega_T, D_{jT} = 1 \right]$$

where $PDV_{T+1} \left(\Omega_{T+1} \right)$ is the present discounted value of lifetime income working from $T+1$ to T_{retire}. Expectation is taken over the future degree outcome $BA_{i,T+1}$.

At $t = T - 1$, the value of option j is

$$V_{jT-1} \left(\Omega_{T-1} \right) = \bar{V}_{jT-1} \left(\Omega_{T-1} \right) + \epsilon_{jT-1}$$

$$= U_{jT-1} + \beta E \left[V \left(\Omega_T \right) | \Omega_{T-1}, D_{jT-1} = 1 \right]$$

where $V \left(\Omega_T \right) = \max_j \left[V_{jT} \left(\Omega_T \right) \right]$.

Rust (1987) showed that this expectation expression has a closed form solution. I can rewrite the expectation term as

$$E \left[V \left(\Omega_T \right) | \Omega_{T-1}, D_{jT-1} = 1 \right]$$

$$= E\left[\gamma + \ln\left(\sum_j \exp\left(\bar{V}_{jT}\left(\Omega_T\right)\right)\right)|\Omega_{T-1}, D_{jT-1} = 1\right]$$

$$= \sum_{BA_{i,T}}\left[\gamma + \ln\left(\sum_j \exp\left(\bar{V}_{jT}\left(\Omega_T\right)\right)\right)\right]Pr\left(BA_{i,T}\right)$$

where γ is the Euler constant. I repeat the process until $t = 1$. Once I have computed the expectation term, I can write choice probabilities as

$$Pr\left(D_{jt} = 1|\Omega_t\right) = \frac{\exp\left(\bar{V}_{jt}\left(\Omega_t\right)\right)}{\sum_j \exp\left(\bar{V}_{jt}\left(\Omega_t\right)\right)}$$

I use these choice probabilities to simulate data in the estimation.

1.5 Estimation

1.5.1 Simulated Method of Moments

I estimate the model with simulated method of moments allowing for unobserved types of individuals. Estimation is also feasible using maximum likelihood, and the unobserved types can be integrated out using the EM algorithm. Due to computational convenience, I choose simulated method of moments. I estimate three sets of parameters in the model: utility parameters including unobserved type parameters, wage parameters, and probability of completion parameters. Table 1.5 lists the parameters of the model. I jointly estimate the unobserved type probability parameter π_1 and the type-specific intercepts in utility functions, $\{\beta_{4,2yr}, \beta_{4,4yr}\}$ with other utility parameters.

Simulated method of moments estimator is

$$\hat{\theta}_{SMM} = \arg\min_{\theta} \left[\mathbf{m}_{sim}(\theta) - \mathbf{m}_{data} \right]' \mathbf{W} \left[\mathbf{m}_{sim}(\theta) - \mathbf{m}_{data} \right]$$

where $\mathbf{m}_{data}$ is the M-dimensional vector of data moments and $\mathbf{m}_{sim}(\theta)$ is the M-dimensional vector of simulated moments given the vector of parameters θ. $\mathbf{W}$ is a $M \times M$ weighting matrix. I choose $\mathbf{W}$ to be the inverse of the variance-covariance matrix of the empirical moments. Before the estimation process, I first compute the data moments, $\mathbf{m}_{data}$ and compute the weighting matrix $\mathbf{W}$ by taking the inverse of the variance matrix of the moments. Using the solution from subsection 1.4.6, I generate simulated data given parameter values θ. Simulated method of moments estimator is consistent and asymptotically normal with variance

$$\left(1 + \frac{1}{S} \right) \left[\left(\frac{\partial \mathbf{m}_{sim}(\theta)}{\partial \theta} \right)' \mathbf{W} \left(\frac{\partial \mathbf{m}_{sim}(\theta)}{\partial \theta} \right) \right]^{-1}$$

The following is the list of moments used in estimation (number of moments in parentheses). There are 39 moments and 22 parameters.

- proportion making choice j in each period $t = [1, 8]$ (16)
- average number of years spent at two-year colleges for each period $t = [2, 9]$ (8)
- average number of years spent at four-year colleges for each period $t = [2, 9]$ (8)
- a four-year college degree status for $t = [3, T + 1]$ (7)

1.5.2 Identification

I identify the utility function parameters by exploiting variation in school and work choices each period including switching between two- and four-year schools, dropping out of college,

15

and returning to college from work. To identify the parameter on tuition level, I assume that tuition level depends only on student's state of residence in high school which is independent of unobserved type. The parameter on previous schooling choice is identified by persistence of schooling choices. If the magnitude of the parameter is relatively small, I would observe more switching between two- and four-year colleges.

Unlike the utility, I observe a four-year college degree status and labor market outcomes in the data. However, students select themselves into different type of school, and the selection problem affects the degree status as well. It is difficult to assume that unobserved heterogeneity affecting schooling outcomes are independent of labor market outcomes. The model allows for an unobserved heterogeneity to account for the selection problem. I assume that an unobserved type affects a four-year college degree outcome and wages indirectly through schooling choices. Individuals observe their types, but the schools and employers do not observe them. This assumption is stronger than allowing a four-year college degree and wages to directly depend on the types. Under this assumption, I can identify a four-year college degree probability and wage function parameters separately from the utility parameters. Unobserved type parameters are identified by using variation in choices of students who have similar observed characteristics.

1.6   Results

In this section, I present the estimates of the model and evaluate the model fit. Table 1.6 displays the utility parameter estimates results. The magnitude of a parameter on ability is higher for four-year colleges: 0.167 for two-year and 0.510 for four-year colleges. The parameter on tuition is -0.146 for two-year and -2.669 for four-year colleges. Subsidizing tuition would have a larger effect on four-year than on two-year college enrollment. An individual of type 1 is more likely to attend college than a type 2 individual, holding other individual

characteristics constant. Moreover, the negative utility associated with attending a two-year college is 10 times larger than that of a four-year college: -2.363 for two-year and -0.231 for four-year colleges. Switching cost is positive for two-year colleges. It captures the fact that community colleges have an open admissions policy. The parameter on switching to four-year colleges is also positive, but it is not statistically significant. I use the estimates from the wage regression (1.2) separately outside the model using the NLSY97 as shown in Table 1.2. Similarly I estimate the probability of BA completion equation (1.3) separately using a logit model. Table 1.7 presents the BA completion functions estimates. The parameters on years spent at two- and four-year college are 0.6102 and 0.9795, respectively. For a student who has an ASVAB percentile score of 60, two years at a community college, and two years at a four-year college would have a 28 percent chance of obtaining a BA degree. The same student with 0 years at a community college and four years at a four-year college would have a 45 percent probability to receive a BA degree. The number of years spent at two- and four-year colleges affect the BA completion probabilities differently.

Figures 1.2, 1.3, and 1.4 display the students' schooling and work choices over time for the actual data and simulated data from the model. The model fits the overall choice patterns of the NLSY97 data. Figure 1.5 reports the BA degree attainment. The model overpredicts BA completion rate after $t = 5$. This is partly due to the fact that the probability of 4-year college degree has the logistic functional form. Figure 1.6 shows the average number of years spent at two- and four-year colleges. Both years spent at two- and four-year colleges matches well.

1.7   Counterfactuals

In this section, I discuss two policy counterfactuals designed to influence access to community colleges. First, I reduce the cost to attend two-year colleges. Second, I improve the

preparation for transition to a four-year college by setting a year spent at community colleges equivalent to a year spent at four-year colleges for the probability of a four-year college degree completion. Given the parameter estimates and randomly drawn shocks, I forward simulate schooling and work choices, degree attainment, and labor market outcomes.

1.7.1 Two-year College Subsidy

I simulate data from the model in which I reduce the tuition of community colleges. Table 1.8 summarizes the counterfactual results. The percentage of students initially enrolled in two-year college increases by 33.7 percent when the subsidy rate is 0.5. Figures 1.7 and 1.9 show that the half-tuition subsidy increases two-year college enrollment and decreases the proportion of students who would have worked otherwise.

When I increase the subsidy rate to 1, i.e., free community college, the percentage of students initially enrolled in two-year college increases by 72.1 percent. Full subsidy draws more students to two-year colleges; hence, the average number of years spent at four-year colleges drops more than the half-tuition case.

This policy mostly affects the students who are on the margin of working and attending two-year colleges and does not have a large effect on four-year college students. Percentage of students who transfer from two-year colleges to four-year colleges decreases from 38 percent to 33 percent. Transfer rate drops because most of the students induced to community colleges are the students who would have not attended college at all. Hence, the increase in enrollment does not translate into an increase in four-year college degree completion. As a result of such enrollment change, the four-year college degree completion rate slightly decreases by 0.7 percent and 3.5 percent depending on the subsidy rate. Lastly, these changes in the enrollment and the completion affect the labor market outcomes measured by the present discounted value of lifetime income. Table 1.8 shows that the average PDV

18

of lifetime income increases by \$10,854 (subsidy rate = 0.5) and \$20,812 (subsidy rate = 1) due to the increase in overall years in college.

Proportions in choices are not sufficient to see how students change their schooling choices compared to the baseline. I present the transitions in years of schooling in Table 1.9. Diagonal elements represent the percentages of students who attended the same years at colleges under the baseline and counterfactuals. For example, the second row of Table 1.9 (a) shows that 50.5 percent of students who spent a year at a two-year college make the same choice and 36.8 percent attend two years at community colleges under the first counterfactual when the subsidy rate is 0.5. Students, who initially attended community colleges, stay longer in community colleges under this counterfactual. In Table 1.9 (b), among the four-year college students, who spent four years, 12.3 percent switched from three years at four-year institutes. When grouping the students based on the years spent at four-year colleges, those who spent one year at four-year colleges (i.e., four-year drop outs) tend to switch to two-year colleges more often than the other group of students who spent more than two years at four-year colleges.

Elements in the right upper side of the diagonal elements represent the students who spend more years at community colleges than the baseline. In particular, the first row shows that 19.5 percent of students who did not attend any college in the baseline now attend more than a year at community colleges.

Transition of years spent at colleges for a full-tuition subsidy is shown in Table 1.10. Overall pattern of the transition is the same, but the size of the transition increases. Difference between the half- and full-tuition rate shows that when a subsidy rate is 0.5 those who spent four years at four-year colleges remained the same for 84.9 percent, whereas under the full-tuition subsidy, the same group of students remained the same for 79.5 percent.

1.7.2 Improvement in Transition to Four-year Colleges

Statewide transfer agreements between community colleges and four-year colleges allow students to transfer their two-year school credits to four-year colleges without depreciation. These types of policies would increase two-year college enrollment and transfers to four-year colleges. In particular, I simulate a model in which I set the parameter on years spent at a two-year college to be the same as the parameter on years spent at a four-year college in the probability of a four-year college completion equation (1.3). I change the parameter on years spent at two-year colleges to 0.9795 (λ_3) from the baseline estimate of 0.6102 (λ_2). All other parameters remain the same as the baseline model estimates. Then the probability of receiving a four-year college degree at time $t+1$ becomes

$$Pr\left(BA_{t+1} = 1 | D_{4yr,t} = 1\right)$$

$$= \Lambda\left(\lambda_0 + \lambda_1 A_i + \lambda_3\left(s_{2yr,t} + s_{4yr,t}\right)\right) \cdot \mathbf{1}\left[s_{4yr,t} \geq min_{4yr}\right] \tag{1.4}$$

Figure 1.15 displays that when a year spent at two-year college is equivalent to a year spent at four-year college for BA completion, students tend to attend two-year colleges more, and the relative size of the change is larger than the changes of a half-tuition subsidy and smaller than the changes of a full-tuition subsidy. The proportion in two-year colleges remains relatively high until $t = 6$ then drops to the baseline outcome as seen in Figure 1.15. Figure 1.17 shows that the proportion of students who work decreases compared to the baseline outcome. Table 1.8 reports that the percentage of students initially enrolled in two-year college increases by 19 percent. Fraction of students transferring from community colleges to four-year colleges increases from 38 percent to 42 percent, and the a four-year college degree completion rate at age 26 increases by 3.1 percent. This counterfactual produces more four-year college degree holders because attending two-year colleges does not hurt students'

chances of getting a four-year college degree.

Table 1.11 presents the transition in college enrollment. This counterfactual increases the average number of years spent at two-year colleges, but the group of students induced to community colleges are different from the tuition subsidy counterfactual. For example, 19.5 percent of those who spent zero years at community colleges switch to two-year colleges more than a year in the first counterfactual. But in the second counterfactual, only 6.7 percent of the same group switch to community colleges.

1.7.3   Welfare Analysis

In this section I perform a welfare analysis by comparing the costs and the benefits of both counterfactuals.

For the first counterfactual, the total cost is

$$\sum_i \sum_t \beta^{t-1} \left\{ tuition_{2yr,i} \cdot subsidy\_rate_{2yr} \cdot \mathbf{1}\left[D_{2yr,it} = 1\right] \right\}$$

where I set $subsidy\_rate_{2yr}$ equal to either 0.5 or 1. I calculate the total benefit by computing the change in the present discounted value of lifetime income.

For the second counterfactual, it is not obvious how to compute the cost of improving preparation for transition to a four-year institution. I use the instructional expenditure per student as a measure of the cost outside of the model. Suppose I spend the same amount of instructional expenditure to a student at a community college as a student at a four-year college to make a year at a two-year school equivalent to a year at a four-year school. The total cost is

$$\sum_i \sum_t \beta^{t-1} \left\{ \Delta expenditure\_per\_student_i \cdot \mathbf{1}\left[D_{2yr,it} = 1\right] \right\}$$

where $\Delta expenditure\_per\_student_i$ is the difference in the instructional expenditure per student at two- and four-year institutions. The total benefit is the change in PDV of lifetime income of students. The instructional expenditure per student obtained from the IPEDS for two- and four-year schools are \$1,350 and \$4,023. The difference of \$2,673 is the $\Delta expenditure\_per\_student_i$.

The last two rows of Table 1.8 present the costs and benefits of the counterfactuals. A half-tuition subsidy increases the percentage of students initially enrolled in two-year college by 33.7 percent, but decreases the four-year college degree attainment slightly. The average PDV of lifetime income increases by \$10,854, which is larger than the average cost of \$4,721. Since the PDV of lifetime income depends on the number of years spent at a college and a BA degree, this increase is due to an increase in years spent at two-year schools. The increase in the benefit due to an increase in years in a college is larger than the decrease in benefit due to a decrease in BA degree attainment.

When a subsidy rate increases to 1, the percentage of students initially enrolled in two-year college increases by 72.1 percent. Both changes in the enrollment and the degree attainment result in higher average PDV of lifetime income by \$20,812. When the subsidy rate is higher, enrollment at a two-year school increases and the average cost goes up to \$12,549, but this cost is still lower than the benefit.

For the second counterfactual, the percentage of students initially enrolled in two-year college increases by 19 percent, and a BA degree attainment increases by 3.1 percent. Both increases in the enrollment and the degree completion result in an increase in the average PDV of lifetime income by \$16,589, which is five times larger than the average cost. As discussed earlier, the instructional expenditure per student is not included in the model. In other words, changing the expenditure does not change the model simulation outcomes. Whether it is the increase in the expenditure or some other programs to improve the community college education, when the average cost is \$14,814, the cost would equal the benefit.

This amount is larger than the half-tuition subsidy cost, $10,754, and smaller than the full-tuition subsidy cost, $20,812. Any amount of expenditure below $14,814 would make the benefit greater than the cost.

1.8   Conclusion

This paper investigates students' college and employment choices and evaluates how two policies promoting access to community colleges affect students' educational and labor market outcomes. I develop and estimate a structural model of work and college choices, including community colleges, which offer post-secondary education at a lower cost. Students can enroll at two- or four-year colleges, transfer, drop out, or work full-time. The probability of a student completing a four-year college degree is given as a function of years spent at both two- and four-year colleges. I estimate the model with simulated method of moments using the data from the NLSY97.

The NLSY97 data shows a steady increase in labor market participation and a decrease in college enrollment from for individuals age 18 to 26. The model developed in this study can reproduce these patterns including the number of years spent at two- and four-year colleges. Given the estimation results, I evaluate two policy counterfactuals designed to promote access to community colleges: tuition subsidies for community colleges and easier transition from two- to four-year colleges.

The findings show that two-year college subsidies increase the percentage of students initially enrolled in two-year college by 33.7 percent and 72.1 percent with a subsidy rate of 0.5 and 1, respectively. At the same time, these subsidies decrease the percentage of students who work initially after high school by 6.9 percent and 13.5 percent with a subsidy rate of 0.5 and 1, respectively. Lower community college tuition attracts students who would not have attended college otherwise. This increase in enrollment at two-year colleges translates

into an increase in student earnings of \$10,854 and \$20,812 in the average PDV of lifetime income; however, it does not translate into increases in the number of transfers and four-year college degrees. When two-year colleges better prepare students for the transition to four-year colleges, the percentage of students initially enrolled in two-year college increases by 19 percent, the transfer rate increases by 27.5 percent, and the four-year college degree completion rate at age 26 increases by 3.1 percent. As a result, the average PDV of lifetime income increases by \$16,589.

These counterfactual results suggest that lower costs increase enrollment but decrease the number of students transferring and earning four-year college degrees. In comparison, when community colleges better prepare students for the transition to four-year colleges, the increase in community college enrollment is larger than in the half-subsidy case, helping more students attend community colleges to eventually earn a four-year college degree.

Welfare analysis results show that subsidizing tuition at community colleges and providing easier transition to four-year colleges increase the present discounted value of lifetime income and in all cases the benefits exceed the costs. The second counterfactual result should be carefully interpreted as it is difficult to quantify the cost of better preparation for transition to a four-year college. I choose instructional expenditure per student to measure the cost of such policy. At \$14,814, the cost would equal the benefit; this value is larger than the half-tuition subsidy cost and smaller than the full-tuition subsidy cost. These two counterfactuals affect different groups of students. Tuition subsidies attract those who are on the margin between two-year college entry and high school graduation; hence the resulting increase in enrollment does not increase the number of students who transfer and earn four-year college degrees. In contrast, when a year at a community college is equivalent to a year at a four-year college for degree completion, the number of students transferring and completing a four-year college degree increase slightly.

Tables and Figures

Table 1.1: Sample Means

	Initial Choice			
	Work	2-year	4-year	Total
ASVAB percentile score	48.32	55.67	74.42	58.74
	[25.44]	[25.02]	[21.92]	[26.93]
Total years at 2-year college at age 26	0.66	2.76	0.25	0.84
	[1.10]	[1.61]	[0.76]	[1.38]
Total years at 4-year college at age 26	0.59	0.84	4.21	1.91
	[1.35]	[1.38]	[1.65]	[2.25]
4-year college degree status at age 26	0.07	0.17	0.60	0.28
	[0.26]	[0.38]	[0.49]	[0.45]
Log wage at age 26	9.67	9.74	9.87	9.76
	[0.78]	[0.76]	[0.76]	[0.78]
Family income (in 2000$)	58872.5	72182.05	81623.75	69016.96
	[54098.68]	[46191.18]	[70613.62]	[60302.27]
Race: non-Hispanic and non-black	0.58	0.65	0.73	0.64
Observations	368	116	267	751

Notes: NLSY97 sample of male high school graduates. Columns represent individuals' initial choices. Standard deviations are shown in brackets.

Table 1.2: Wage Regression: NLSY97 Sample

	Estimates	Standard Error
Ability	0.0016	0.0002
Years at 2 year	0.0470	0.0063
Years at 4 year	-0.0445	0.0052
4-year college degree	0.2505	0.0232
Constant	7.8778	0.0175
Observations	38,421	
R^2	0.2002	

Notes: Regression result for the NLSY97 sample between age 18 and 26. This is different from the wage regression using the earlier survey (NLSY79) sample between age 26 to 65. Dependent variable is a log wage. Ability is measured by the ASVAB math-verbal percentile score. Independent variables include age dummy variables.

Table 1.3: Wage Regression: NLSY79 Sample

	Estimates	Standard Error
Work experience	0.049	0.002
Work experience squared	-0.001	0.000
Race: black	-0.285	0.013
Race: non-Hispanic and non-black	0.119	0.012
4-year college degree	0.445	0.020
Constant	9.596	0.014
Observations	40,302	
R^2	0.1656	

Notes: NLSY79 sample. Dependent variable is log wage. Independent variables includes dummies for number of years at college (1 to 8). Coefficients on number of years at college range from 0.115 to 0.288, and all estimates are statistically significant at 99 percent level.

Table 1.4: List of Variables and Data Sources

Variable	Source
School and work choices $(D_{j1}, \ldots, D_{jT})$	NLSY97
Degree status $(BA_1, \ldots, BA_{T+1})$	NLSY97
Family income	NLSY97
Parents' education	NLSY97
ASVAB math-verbal percentile score	NLSY97
Race, gender, age, census region	NLSY97
Tuition (2- and 4-year state average)	IPEDS
Instructional expenditure per student (2- and 4-year state average)	IPEDS
Wage between 18 and 26	NLSY97
PDV lifetime income (wage profile from 26 to 65)	NLSY79

Notes: NLSY79: National Longitudinal Survey of Youth 1979; NLSY97: National Longitudinal Survey of Youth 1997; IPEDS: Integrated Postsecondary Education Data System.

Table 1.5: List of Parameters

	Variable	Parameter
Utility $(j \in \{2yr, 4yr\})$	constant	β_{0j}
	ability	β_{1j}
	tuition	β_{2j}
	transfer cost	β_{3j}
	type 2	β_{4j}
	wage	ρ
$\Pr(BA = 1)$	constant	λ_0
	ability	λ_1
	years at 2-year	λ_2
	years at 4-year	λ_3
Wage	constant	α_0
	ability	α_1
	years at 2-year	α_2
	years at 4-year	α_3
	4-year degree	α_4
$\Pr(type = 1)$	probability of type 1 $(\pi_1 = 1 - \pi_2)$	π_1

Notes: Total number of parameters is 22.

Table 1.6: Estimation Result: Utility

		Estimates	Standard Error
	constant	-0.267	0.040
	ability	0.167	0.003
Utility 2-year	tuition	-0.146	0.068
	switch	0.903	0.334
	type 2	-2.335	0.630
	constant	2.774	0.472
	ability	0.510	0.018
Utility 4-year	tuition	-2.669	0.092
	switch	0.203	0.801
	type 2	-0.238	0.061

Notes: NLSY97 sample. Tuition in thousands of dollars. A dummy variable, type 2 equals to 1 if an unobserved type is 2 and 0 otherwise. Ability is measured by the ASVAB percentile score. Estimates obtained using simulated method of moments.

Table 1.7: Estimation Result: 4-year College Degree Completion

	Estimates	Standard Error
Ability	0.0074	0.0008
Years at 2-year	0.6102	0.0311
Years at 4-year	0.9795	0.0201
Constant	-4.5392	0.0909

Notes: NLSY97 sample who attended at least 2 years at four-year colleges. Ability is measured by the ASVAB percentile score.

Table 1.8: Counterfactual Results

	Baseline simulation	CF1: 2-year college Tuition subsidy		CF2: Easier transition to 4-year college
		Half	Full	
	(1)	(2)	(3)	(4)
% 2-year at age 18-20	0.142	0.190	0.245	0.169
% 4-year at age 18-20	0.345	0.333	0.312	0.335
% work at age 18-20	0.513	0.477	0.443	0.496
% transfer	0.385	0.356	0.331	0.422
Average number of years at 2-year at 26	0.784	1.061	1.417	1.120
Average number of years at 4-year at 26	1.828	1.780	1.723	1.724
4-year college degree completion at 26	0.378	0.376	0.365	0.390
Average PDV of lifetime income (000$)	605.091	615.945	625.903	621.680
ΔAverage PDV of lifetime income (000$)	-	10.854	20.812	16.589
Average cost (Total cost/Nobs, 000$)	-	4.722	12.550	2.993

Notes: Column (1) shows the baseline simulation. Column (2) and (3) show counterfactual 1 result with a subsidy rate = 0.5 and a subsidy rate = 1 (i.e., free community college). Column (4) displays counterfactual 2 result with Δexpenditure= $2,673. Transfer defined as individuals who attend 2-year college at either $t = 1$ or $t = 2$ and attend at least a year at 4-year college afterward.

Table 1.9: Counterfactual 1: Transition of Total Years at College

(a) Years at 2-year Colleges at 26

	Years	Counterfactual 1				
		0	1	2	3	3+
Baseline	0	91.0	7.0	2.0		
	1	1.1	50.5	36.8	10.5	1.1
	2			46.6	43.1	10.3
	3		3.3		46.7	50.0

(b) Years at 4-year Colleges at 26

	Years	Counterfactual 1					
		0	1	2	3	4	5
Baseline	0	97.2	2.1	0.5	0.2		
	1	68.4	21.1	5.3	5.3		
	2	3.4	13.8	62.1	17.2	3.4	
	3		4.0	20.0	76.0		
	4			2.7	12.3	84.9	
	5				1.2	7.3	91.5

Notes: Tables show transitions in years of schooling from the baseline to the counterfactual. Rows represent years at college in the baseline simulation, and columns represent years at college in counterfactual simulations. Units in percentage (%). Subsidy rate is 0.5.

Table 1.10: Counterfactual 1: Transition of Total Years at College

(a) Years at 2-year Colleges at 26

| | Years | Counterfactual 1 | | | | |
		0	1	2	3	3+
Baseline	0	80.0	13.3	5.1	1.2	0.4
	1	1.1	34.7	21.1	26.3	16.8
	2		1.7	20.7	34.5	43.1
	3			3.3	13.3	83.3

(b) Years at 4-year Colleges at 26

| | Years | Counterfactual 1 | | | | | |
		0	1	2	3	4	5
Baseline	0	97.2	1.9	0.5	0.2	0.2	
	1	73.7	15.8	5.3		5.3	
	2	27.6	3.4	48.3	13.8	3.4	3.4
	3	2.0	6.0	14.0	72.0	4.0	2.0
	4		4.1	2.7	12.3	79.5	1.4
	5			1.2	4.9	13.4	78.0

Notes: Tables show transitions in years of schooling from the baseline to the counterfactual. Rows represent years at college in the baseline simulation, and columns represent years at college in counterfactual simulations. Units in percentage (%). Subsidy rate is 1, i.e., free community college.

Table 1.11: Counterfactual 2: Transition of Total Years at College

(a) Years at 2-year Colleges at 26

	Years	0	1	2	3	3+
			Counterfactual 2			
Baseline	0	93.3	4.3	1.2	0.6	0.6
	1		71.6	14.7	10.5	3.2
	2			55.2	19.0	25.9
	3		3.3	3.3	30.0	63.3

(b) Years at 4-year Colleges at 26

	Years	0	1	2	3	4	5
			Counterfactual 2				
Baseline	0	97.0	2.1	0.7	0.2		
	1	42.1	26.3	15.8	15.8		
	2	13.8	3.4	72.4	6.9	3.4	
	3	2.0	4.0	14.0	72.0	6.0	2.0
	4		5.5	4.1	13.7	76.7	
	5				7.3	12.2	80.5

Notes: Tables show transitions in years of schooling from the baseline to the counterfactual. Rows represent years at college in the baseline simulation, and columns represent years at college in counterfactual simulations. Units in percentage (%).

Figure 1.1: NLSY97 Sample Choices
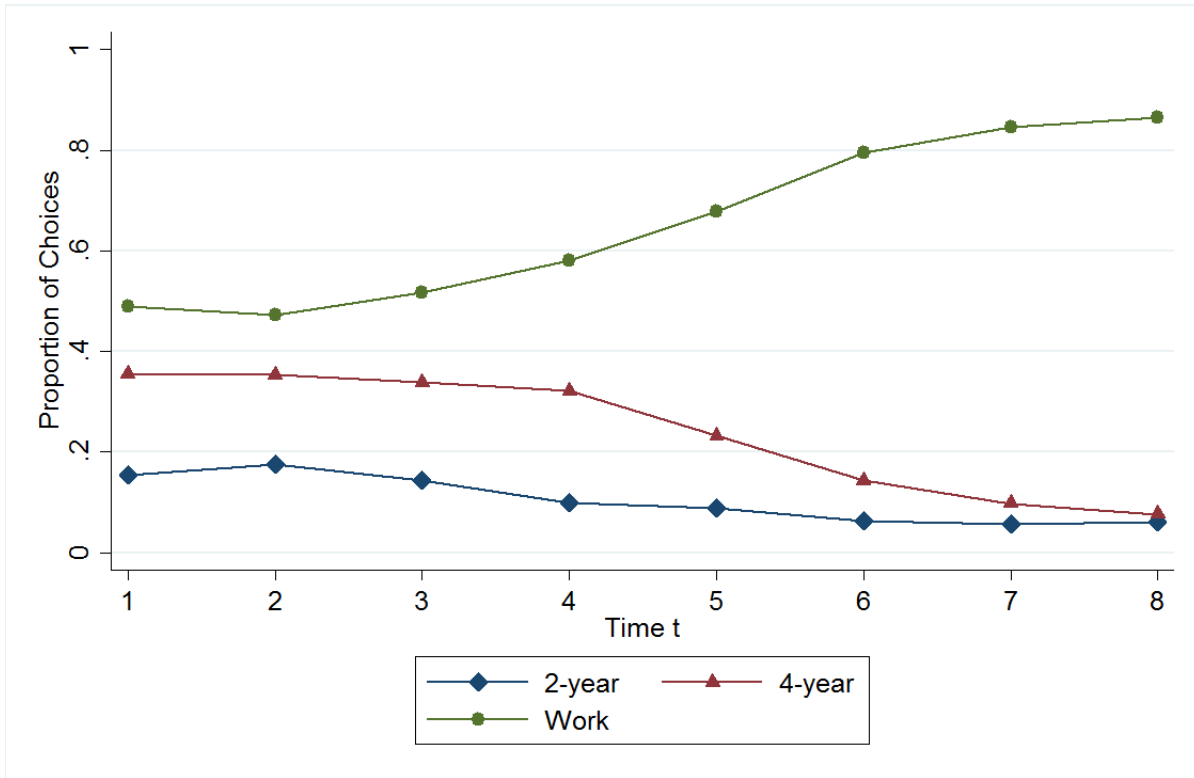
Notes: Data: NLSY97 sample; Observations = 751.
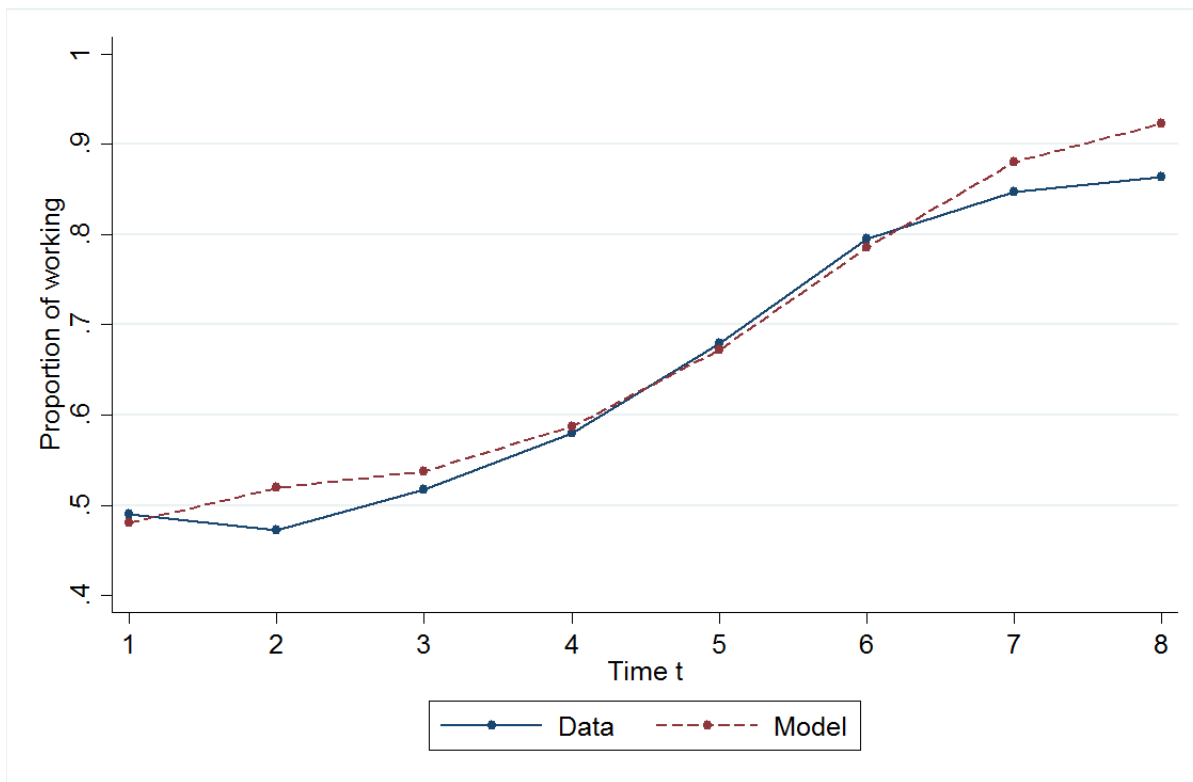
Figure 1.2: Model Fit: 2-year College
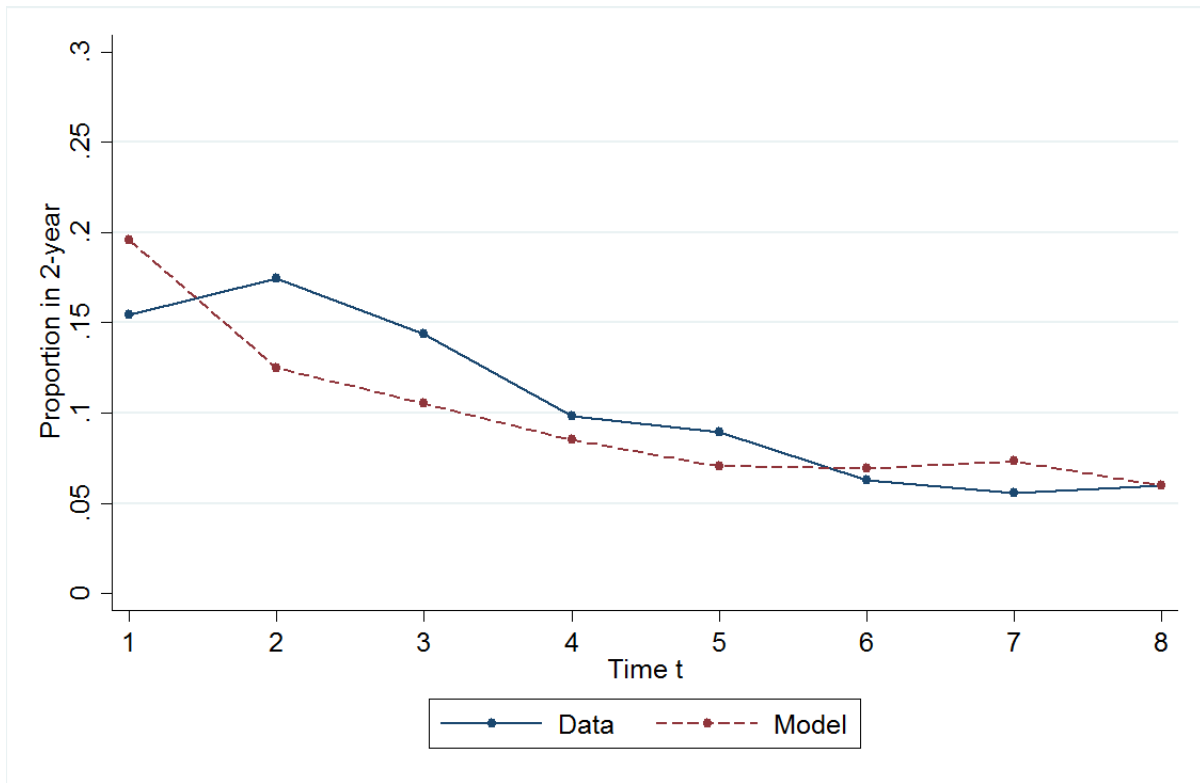
Notes: Data: NLSY97 sample; Model: simulated data.
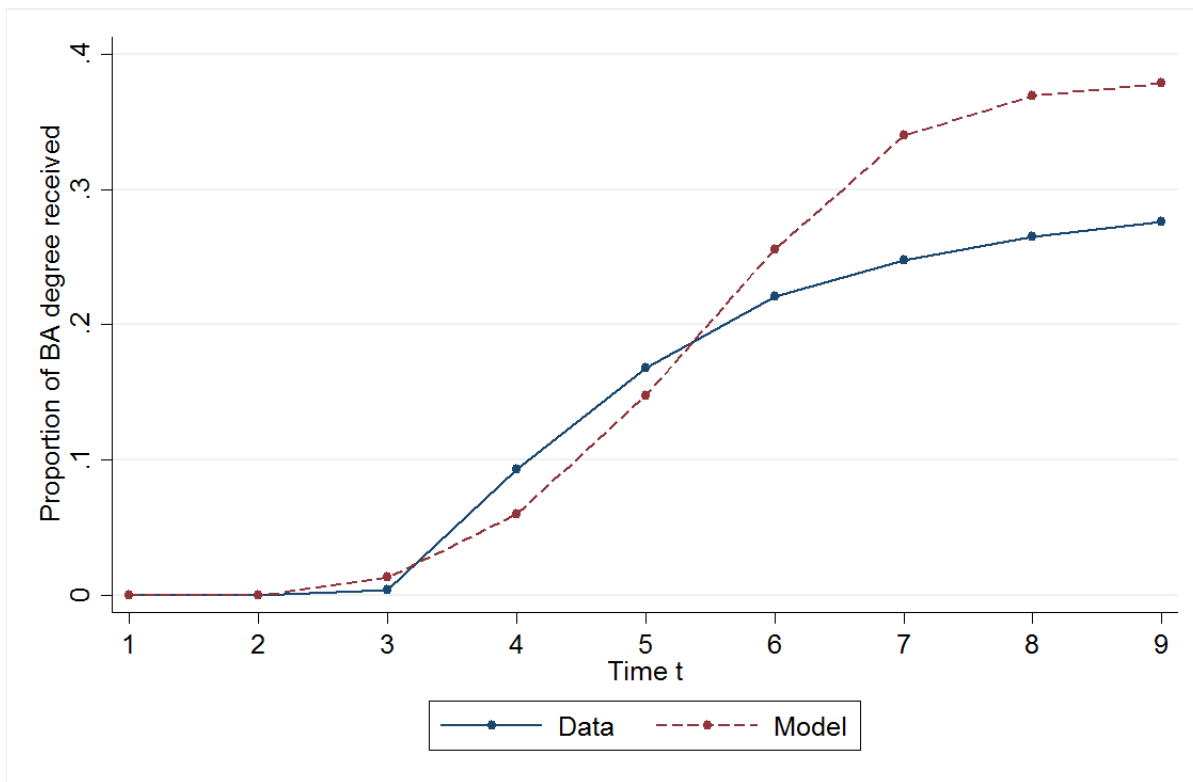
Figure 1.3: Model Fit: 4-year College
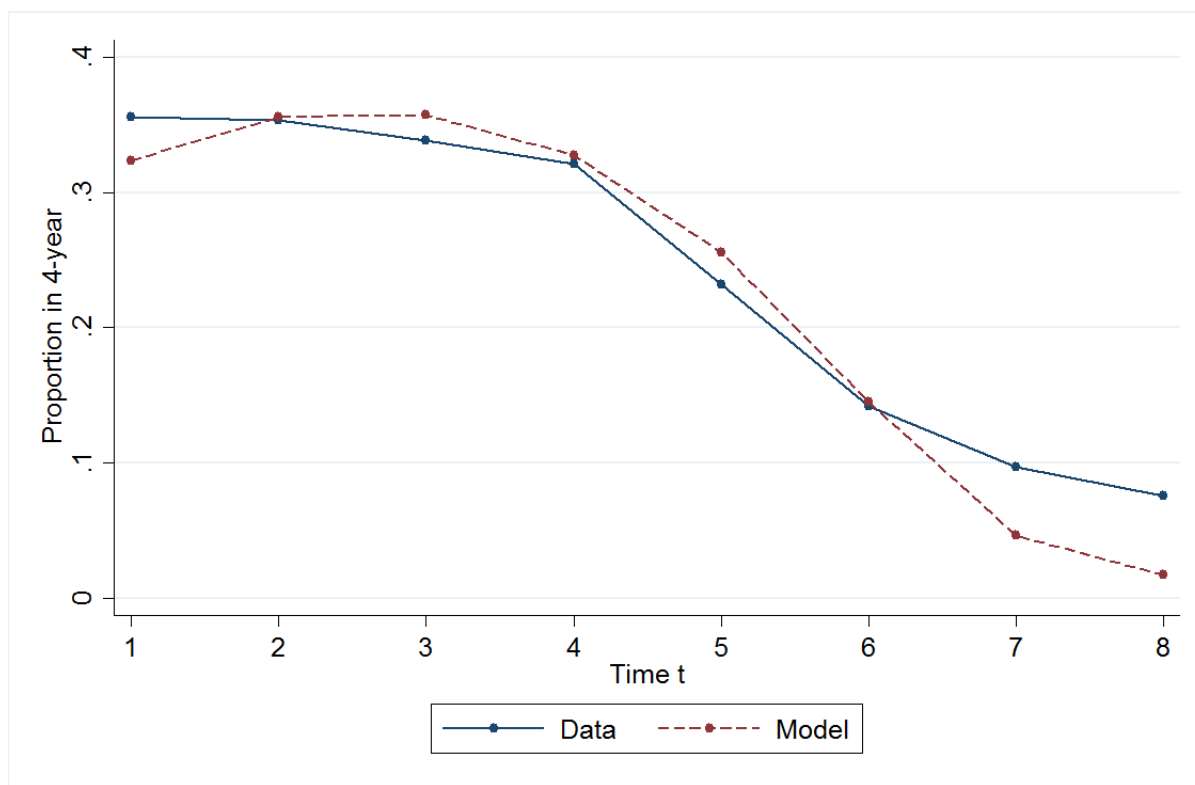
Notes: Data: NLSY97 sample; Model: simulated data.
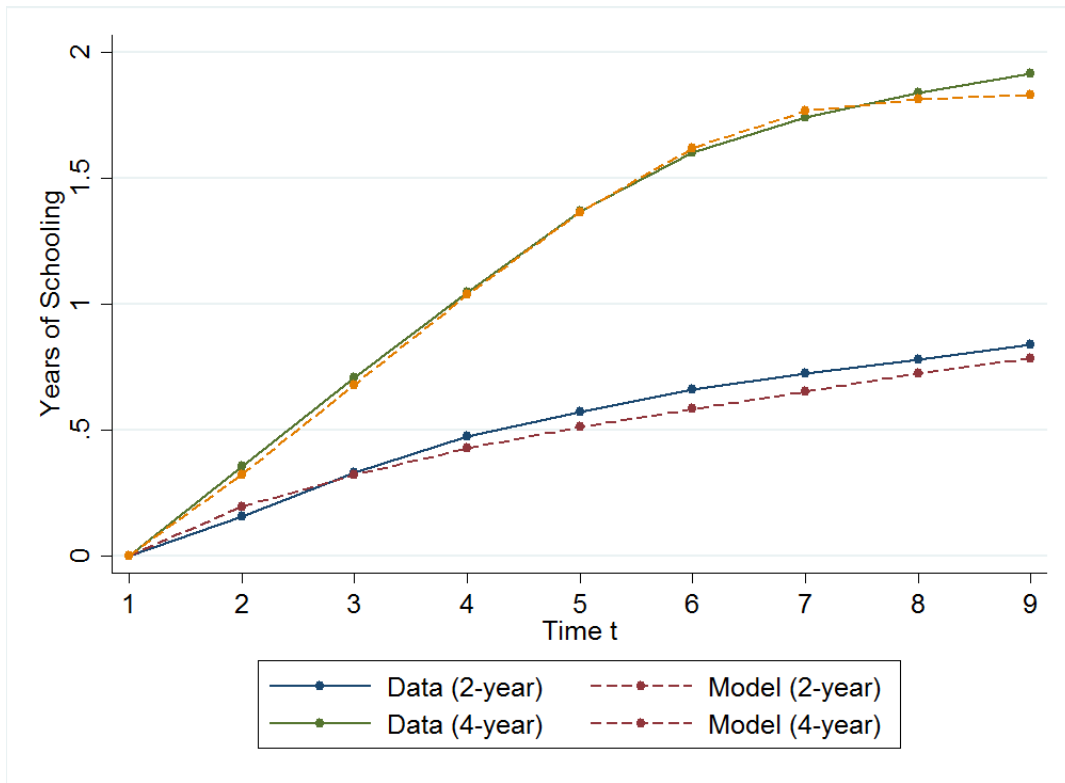
Figure 1.4: Model Fit: Work

Notes: Data: NLSY97 sample; Model: simulated data.
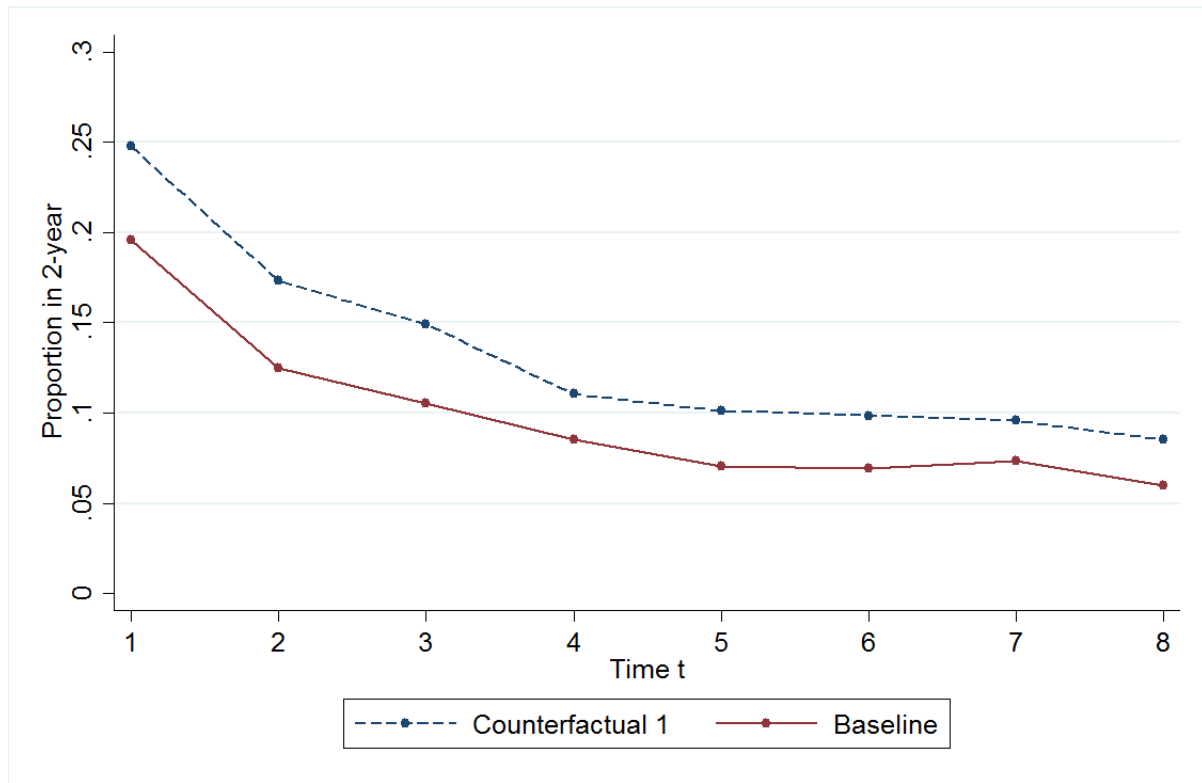
Figure 1.5: Model Fit: 4-year College Degree

Notes: Data: NLSY97 sample; Model: simulated data.

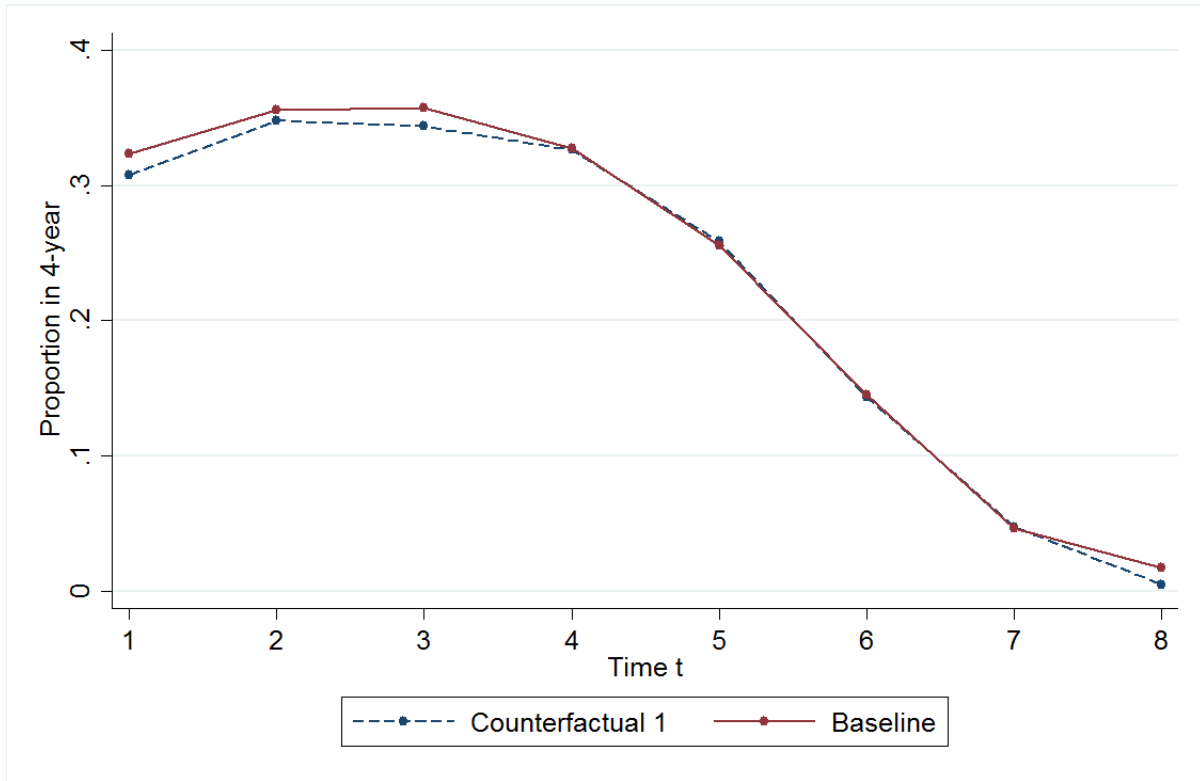Figure 1.6: Model Fit: Average Years of Schooling

Notes: Data: NLSY97 sample; Model: simulated data.

Figure 1.7: Counterfactual 1. Half Tuition Subsidy: 2-year College
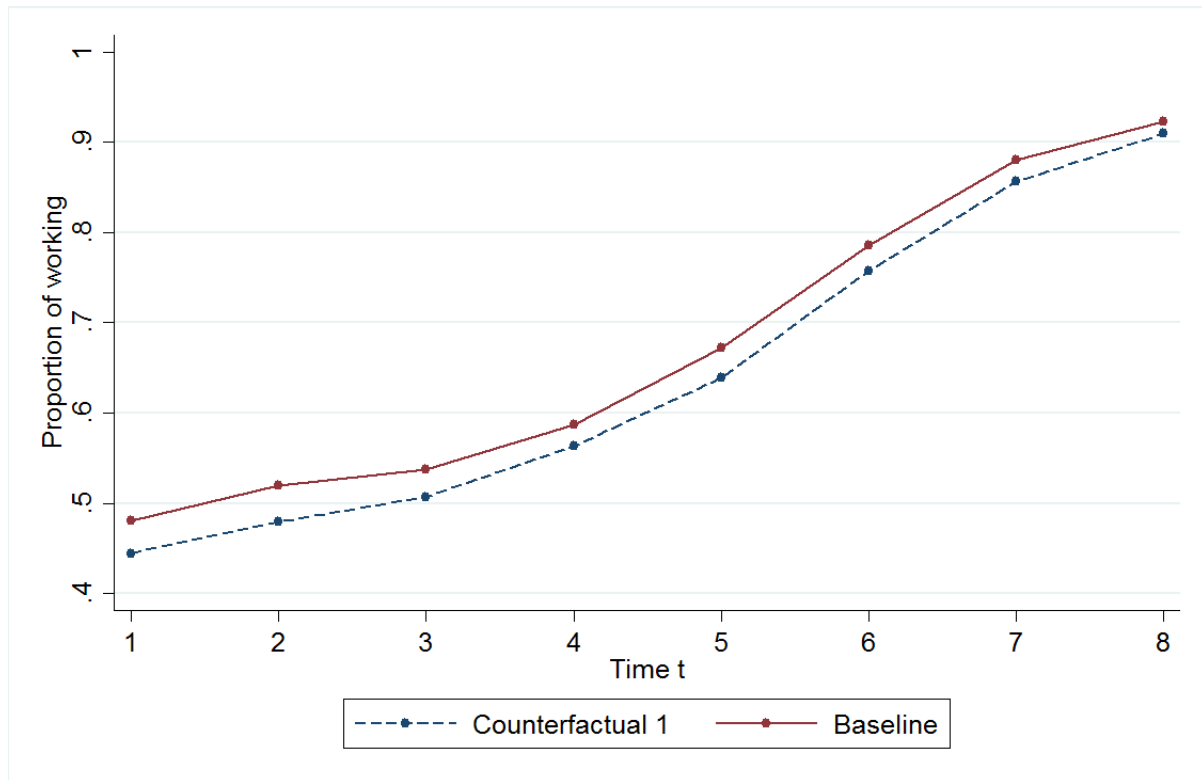

Notes: Simulated data from counterfactual 1. Subsidy rate = 0.5

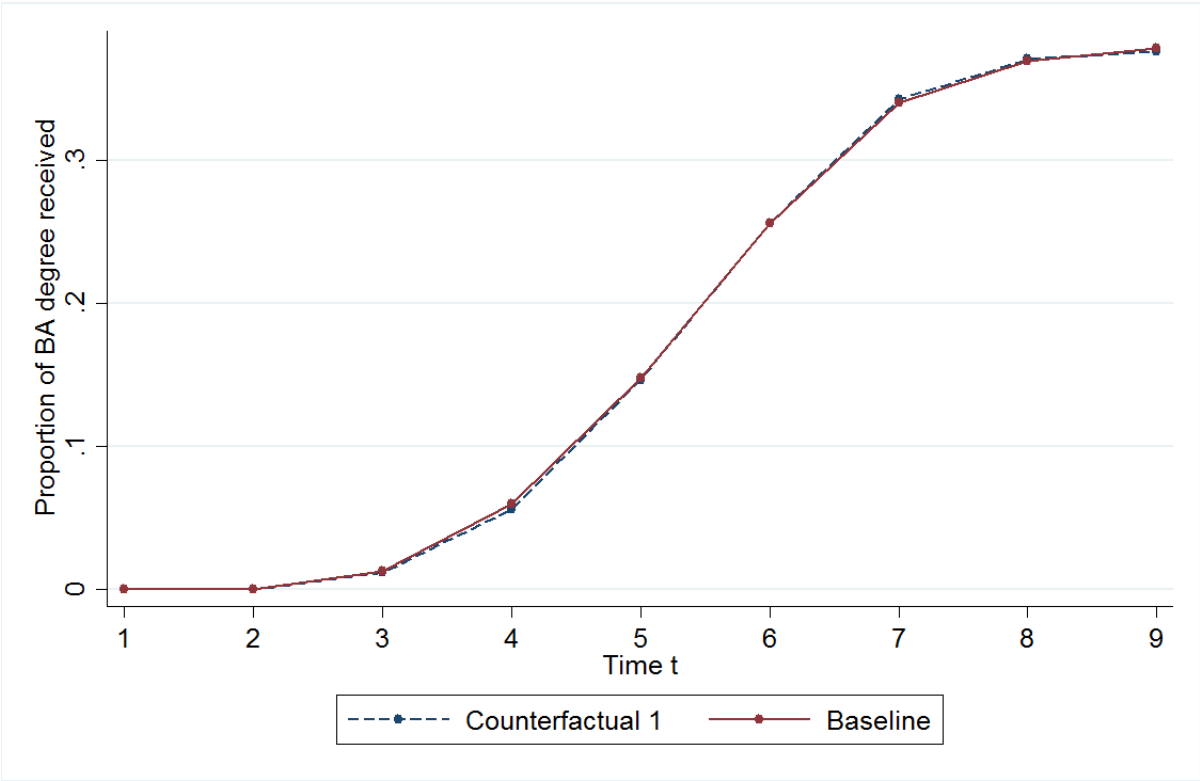Figure 1.8: Counterfactual 1. Half Tuition Subsidy: 4-year College

Notes: Simulated data from counterfactual 1. Subsidy rate = 0.5

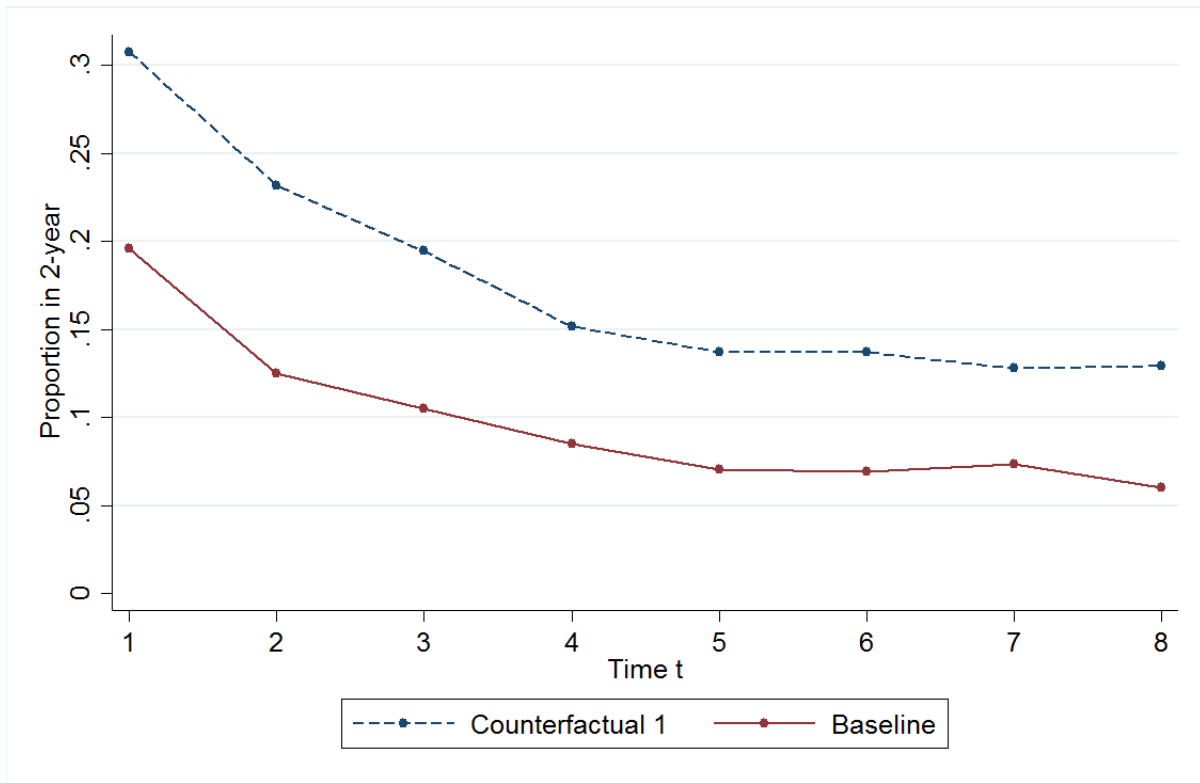Figure 1.9: Counterfactual 1. Half Tuition Subsidy: Work

Notes: Simulated data from counterfactual 1. Subsidy rate = 0.5

Figure 1.10: Counterfactual 1. Half Tuition Subsidy: 4-year College Degree
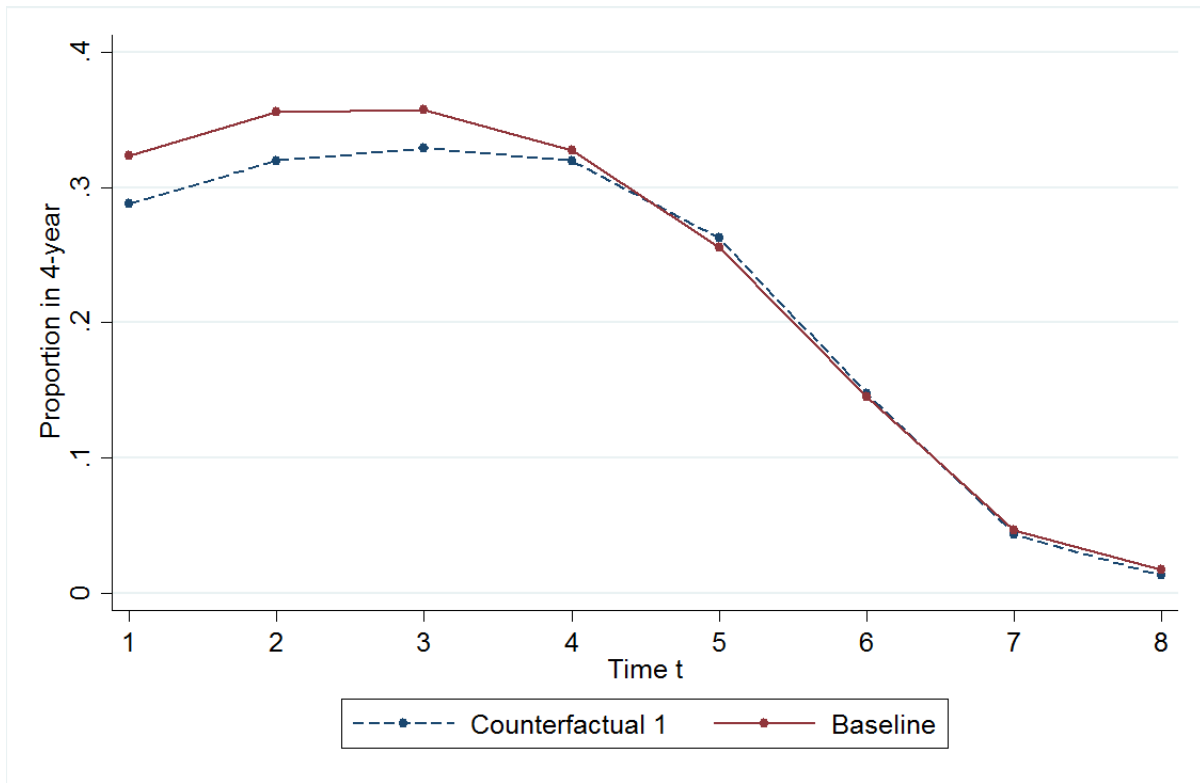

Notes: Simulated data from counterfactual 1. Subsidy rate = 0.5

Figure 1.11: Counterfactual 1. Full Tuition Subsidy: 2-year College

Notes: Simulated data from counterfactual 1. Subsidy rate = 1

Figure 1.12: Counterfactual 1. Full Tuition Subsidy: 4-year College

Notes: Simulated data from counterfactual 1. Subsidy rate = 1

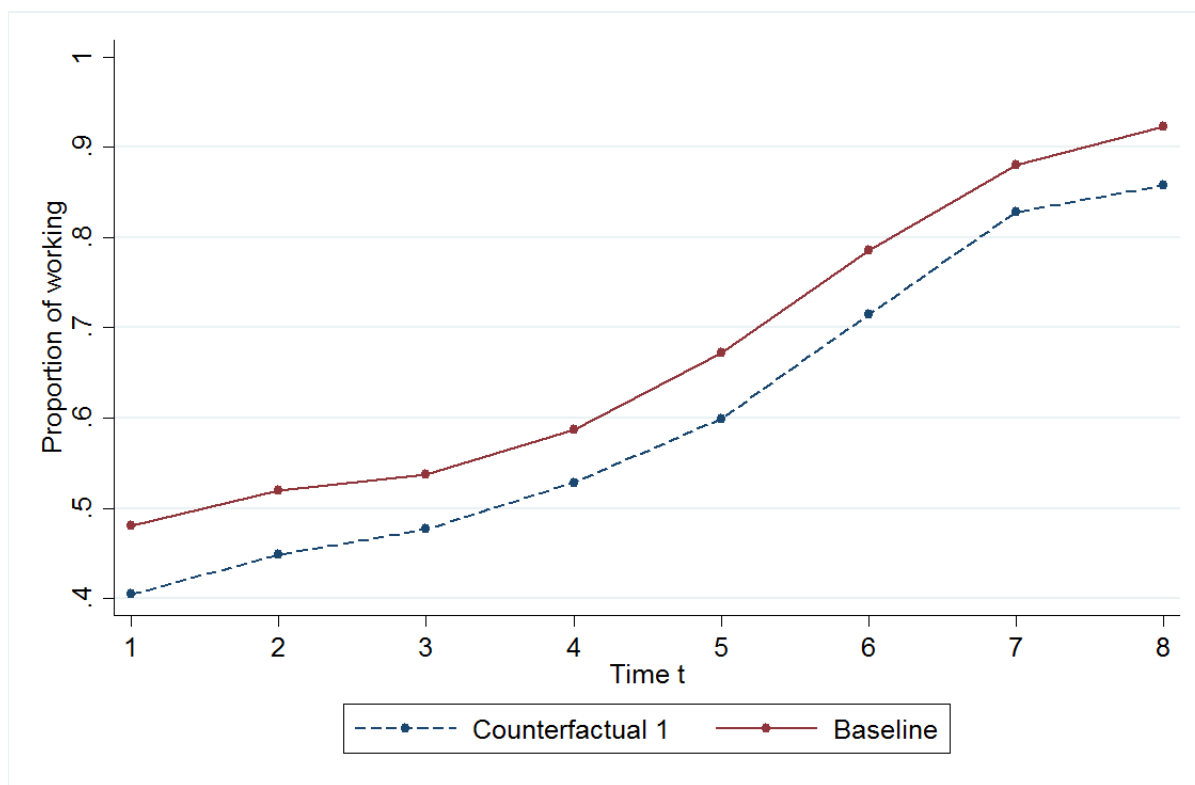Figure 1.13: Counterfactual 1. Full Tuition Subsidy: Work

Notes: Simulated data from counterfactual 1. Subsidy rate = 1

Figure 1.14: Counterfactual 1. Full Tuition Subsidy: 4-year College Degree

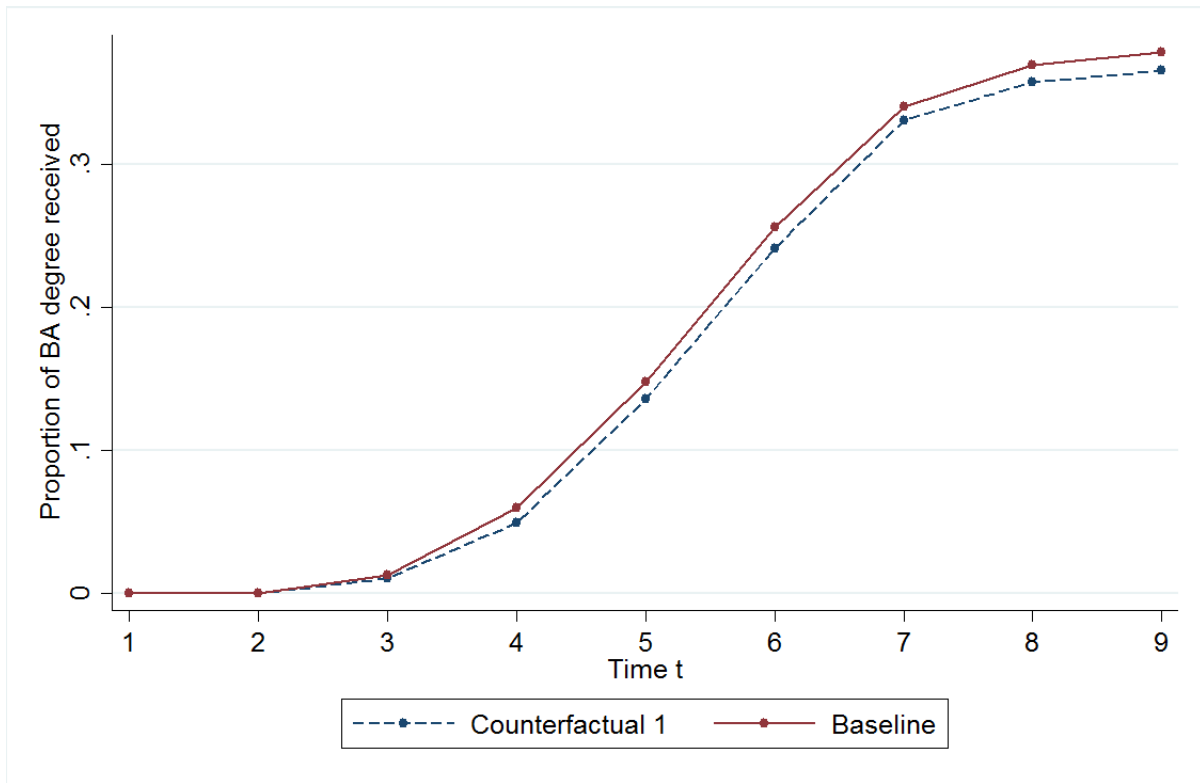Notes: Simulated data from counterfactual 1. Subsidy rate = 1

Figure 1.15: Counterfactual 2. Easier Transition to 4-year Colleges: 2-year College
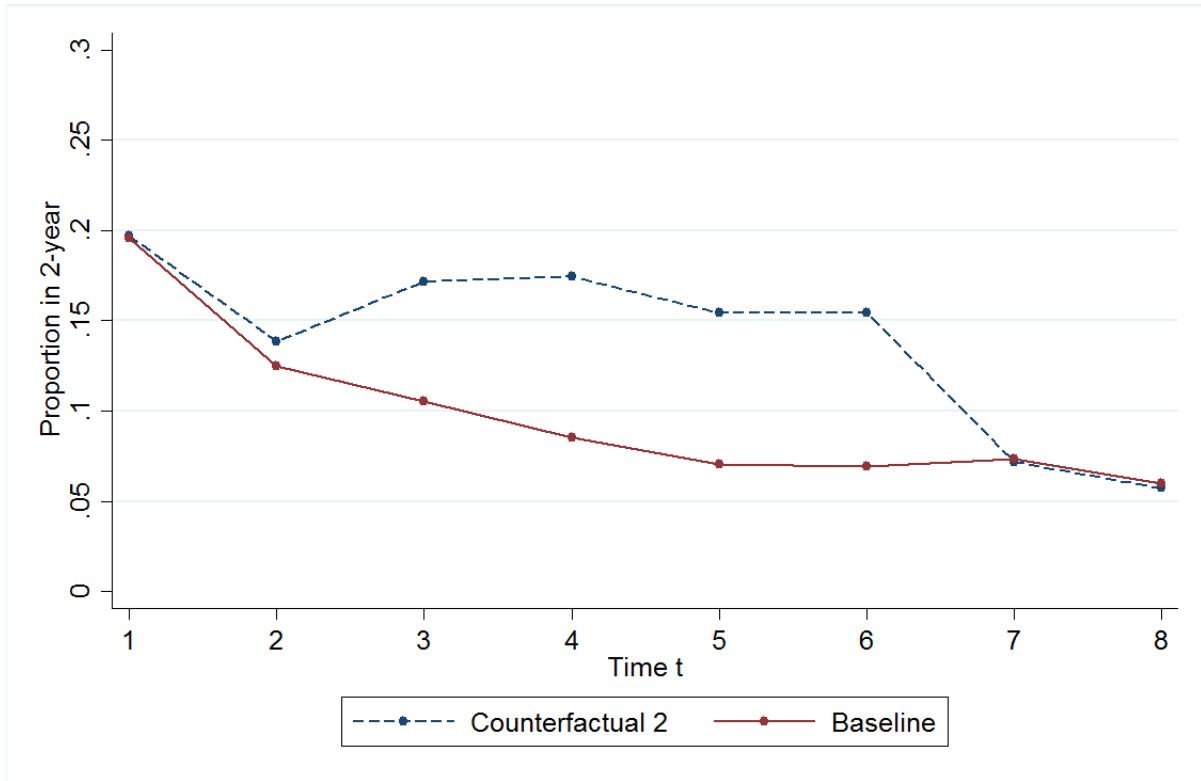
Notes: Simulated data when a year at a 2-year college is equivalent to a year at a 4-year college.

Figure 1.16: Counterfactual 2. Easier Transition to 4-year Colleges: 4-year College
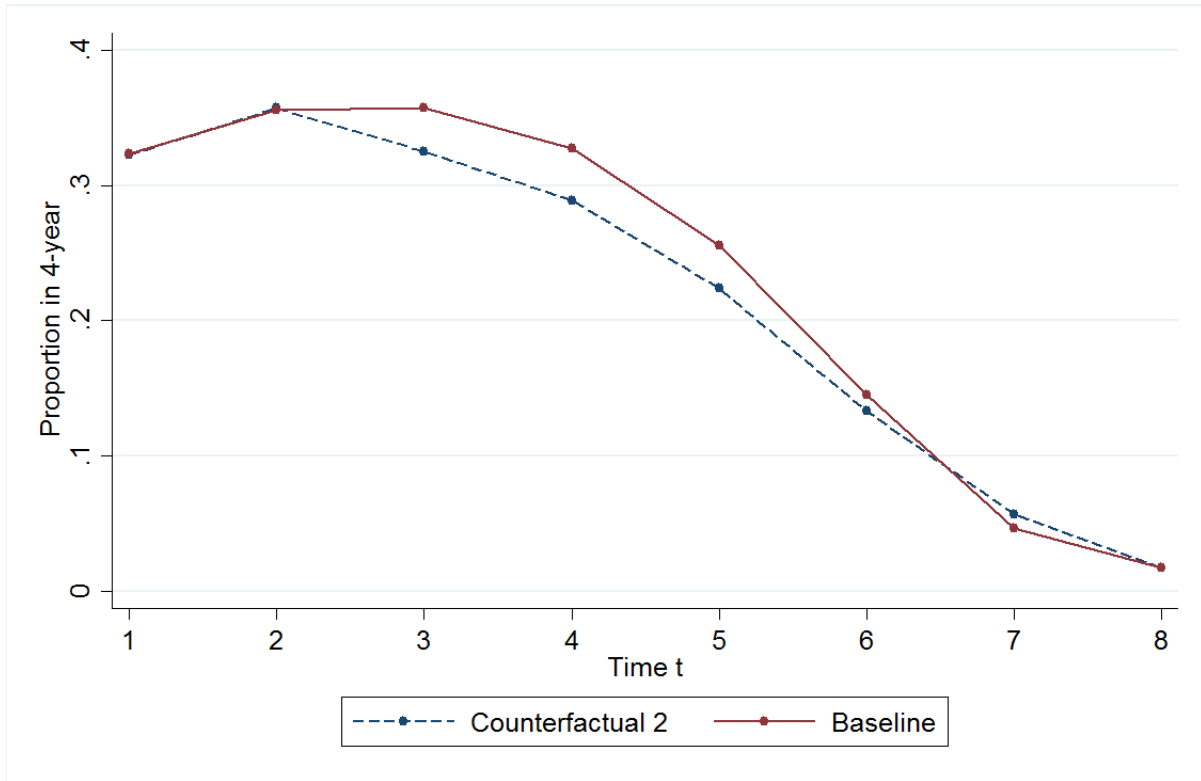
Notes: Simulated data when a year at a 2-year college is equivalent to a year at a 4-year college.

Figure 1.17: Counterfactual 2. Easier Transition to 4-year Colleges: Work

Notes: Simulated data when a year at a 2-year college is equivalent to a year at a 4-year college.

Figure 1.18: Counterfactual 2. Easier Transition to 4-year Colleges: 4-year College Degree

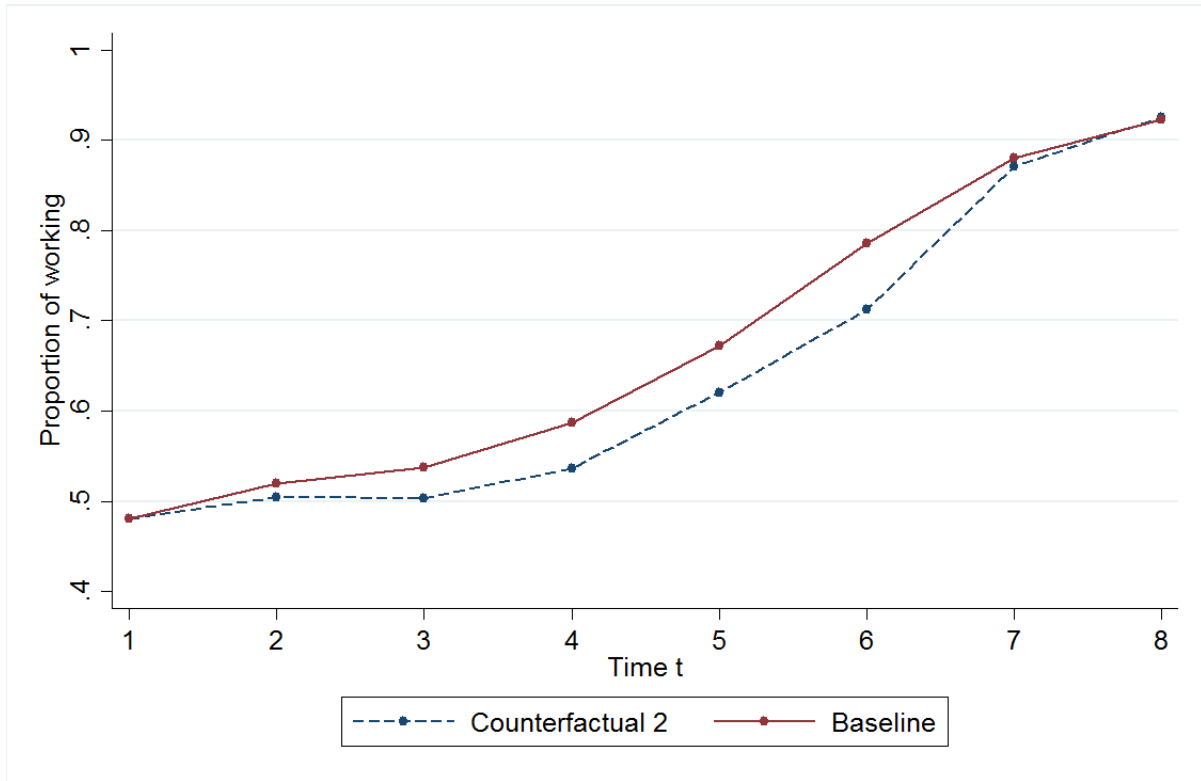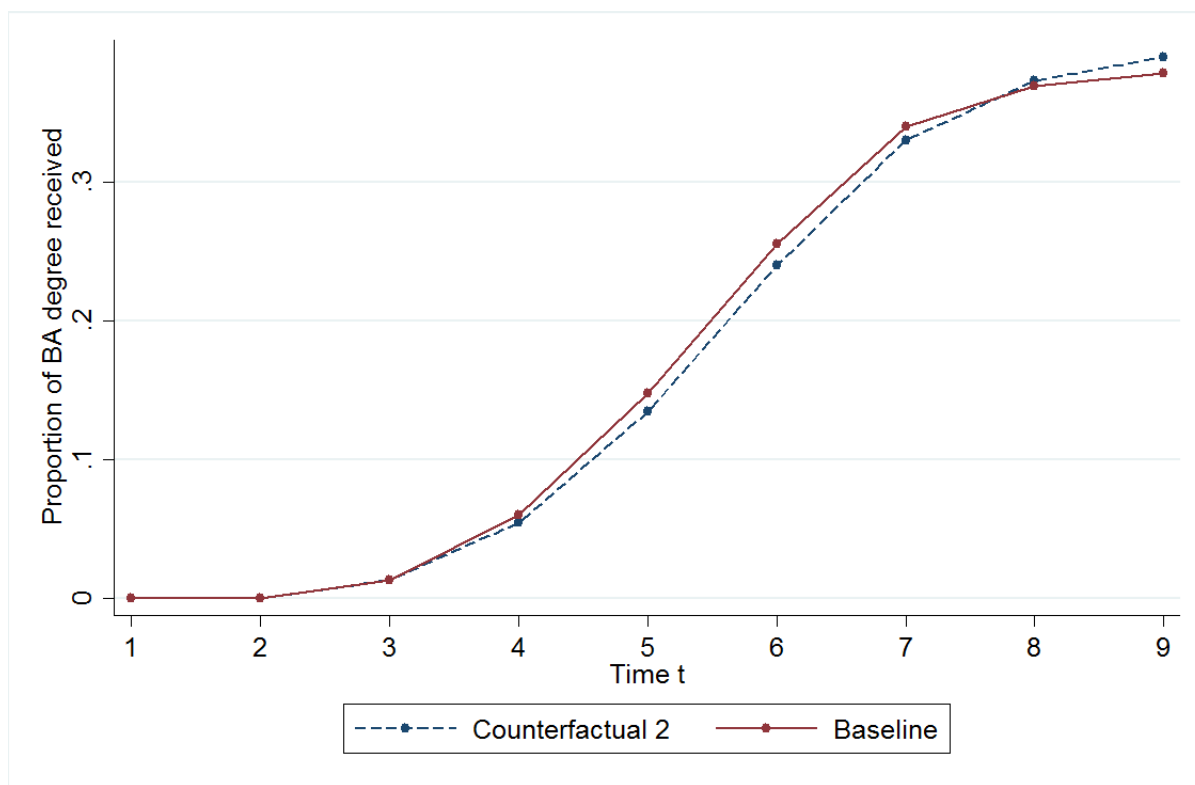Notes: Simulated data when a year at a 2-year college is equivalent to a year at a 4-year college.

Chapter 2

Quality of Public High Schools and Educational Choices

2.1 Introduction

Public high schools in the U.S. are largely funded by local property taxes, and the school funding formula relies on property taxes. As a result, high-income school districts have more educational resources than low-income school districts. Suppose different levels of funding cause the quality of high schools through different factors, and local school districts or the federal government were to achieve the goal of improving school quality. What would happen to students' educational outcomes given this improvement? For example, Illinois plans to make school funding more equitable by allocating more funding to districts with low property values. High school quality seems to make difference for student performance and educational choices, but it is difficult to identify the causal relationship between quality and outcome. The main challenge is that it is difficult to find an exogenous variation in schooling inputs to estimate a causal effect.

This paper investigates the effect of public school quality on students' enrollment decisions and their academic achievements in college. I investigate how school quality affects observable educational choices by estimating a structural model. I use a structural model to

quantify the effect of high school quality on students' educational choices for an outcome-based analysis. This paper focuses on school performance instead of schooling inputs that may affect students' performance. I investigate college enrollment and post-secondary education completion rates when the school quality were to change across schools.

I use the National Education Longitudinal Study of 1988 to create a measure of high school quality by taking the average of the standardized test scores for students at each school. In my model of college enrollment and continuation choices, the quality of high schools directly affects students' SAT scores, the probabilities of being admitted to high-quality colleges, and college grades.

The first counterfactual results show that centralizing public education results in a higher college enrollment rate of students from a low-quality school districts. Because of the strong correlation between quality of high school and family income level, those who benefit most from this counterfactual policy are the students enrolled in a low-quality high school located in a low-income district. I find that college enrollment increases by 19.8 percent and college completion increases by 17.6 percent for the students from the lowest quartile quality schools. Secondly, when a policy imposes a lower bound for the high school quality, I find that college enrollment increases by 9.1 percent and college completion increases by 7.8 percent for the students attending the lowest quality quartile schools. Since the students attending the lowest-quality schools are mostly from the lowest family income background, I find a small increase for the low-income families. Lastly, a 10 percent increase in the ability of students from the lowest 25th percentile quality schools increases enrollment and completion by 59.5 percent and 33.3 percent, respectively. In addition, this counterfactual policy increases enrollment by 27.6 percent and completion by 16.1 percent for a group of students from a family income below \$35,000. Compared to a direct increase in the ability of the students, an increase in the quality of public schools has a modest effect on college enrollment and completion.

Many previous studies focus on how different factors affect the school quality and academic achievement, but the results are mixed (See Hanushek (1986) and Hedges et al (1994)). Examples of specific schooling inputs that have been widely discussed include: reducing the class size (Krueger (2003)), adopting the school accountability systems (Kane and Staiger (2002) and Jacob (2005)), increasing the school choice options (Hastings et al (2005)), reforming the school finance system (Card and Payne (2002) and Guryan (2003)), and improving teacher quality (Rivkin et al (2005)). This paper differs from the existing literature by focusing on the student performance as a measure of school quality. I measure a high school quality by taking the average of the standardized test scores for students at each school.

In terms of the model choice, this paper is related to structural models used in Arcidiacono (2005) and Stange (2012) in the context of higher education. Arcidiacono estimates a structural model where students apply to college, choose where to attend, and choose what subject to major in, while the schools select students for admission and financial aid offers. To understand how the quality of schools affects the educational choices, this paper adds quality of high school to potentially affect the SAT score, college quality, and college grades. Quality of high schools indirectly affects educational choices through a SAT score, college quality, and college grades.

The rest of the paper is structured as follows. Section 2.2 introduces the data and descriptive analysis. Section 2.3 describes the model, and Section 2.4 explains the estimation strategy. Section 2.5 presents the estimation results and the model fit . Section 2.6 discusses the policy counterfactuals targeting high school qualities. Finally, Section 2.7 concludes.

2.2 Data and Descriptive Statistics

2.2.1 Data Construction

I use the restricted-use version of the National Education Longitudinal Study of 1988 (NELS88). In the spring of 1988, the NELS88 surveyed 24,599 8th grade students attending 1,052 high schools in the U.S. The follow-up surveys were conducted in 1990, 1992, 1994, and 2000. The third follow-up survey (1994) contains data for participants who would be sophomores if they decided to attend college after graduating from a high school. The fourth follow-up survey (2000) provides data for participants who would have either graduated or dropped out if they went to college. The NELS88 surveys from 1992, 1994, and 2000 fit well with the timeline of the model in which students make decisions to enroll in college and to stay or drop out.

I restrict the sample to male high school seniors who graduated as the class of 1992 and participated in follow-up surveys in 1994 and 2000. I use the transcript data to determine whether an individual dropped out at certain year or completed college. Transfers are not considered in the sample. Total number of the sample is 1914.

I obtain both student and school level data from the NELS88 data set. Student level data includes college grades, enrollment and drop-out status, standardized mathematics test scores, SAT scores, and family income. I compute the cumulative grades using transcript data, and I use standardized mathematics scores tested in the base-year survey as a measure of ability. School level data includes the key variable, high school quality, college quality, and tuition costs. To have a standardized measure of the quality throughout public high schools, I take the average of the standardized mathematics test scores for enrolled students if a high school have more than 16 students in the sample. The number of high schools in the sample is 600. I use 13,825 observations to compute qualities of 600 high schools in this sample. To measure college quality, I use the percentile SAT scores of incoming students for a given

college in 1992 obtained by linking the NELS88 data set to the Integrated Postsecondary Education Data System (IPEDS).

Since the NELS88 data set has limited income data for a short period of time after college, it is difficult to estimate the expected income and to project this value to an individual's retirement age of 65. To supplement the NELS88, I estimate the expected lifetime income using the data from the National Longitudinal Survey of Youth 1979 (NLSY79). I assume that individuals in my sample from the NELS88 look at the income from NLSY79 to form expectations about their future earnings.

2.2.2   Descriptive Statistics

Table 2.1 shows the descriptive statistics for the main variables in the model. I divide the full sample of students into two groups based on quality level of the high school they attended. The first two columns display the sample means of variables in the full sample. The next two columns represent students who enrolled in college, and the last two columns show the sample means of the students who completed college. A pattern showing a high percentage of enrollment, completion, math scores, high-income, and non-black and non-Hispanic ethnicity remain the same regardless of educational status. Probabilities for enrollment and completion are higher for the students from better quality high schools. Moreover, the correlation between the family income and high school quality illustrates the fact that the U.S. public education system depends on local property taxes.

Figures 2.2-2.4 illustrate qualities of high schools in the data. Figure 2.2 shows the density of a high school quality in the sample. The mean is 51.27, and the standard deviation is 5.2. Figure 2.3 illustrates the fraction of family income levels within each school quality quartile. For example, the fraction of the highest family income category is 43.97 percent if a high school quality is the highest, whereas the same fraction drops to 5.2 percent if a high school

quality is the lowest. Except for the lowest and the highest family income categories, there is enough variation in family income for each school quality quartile.

Figure 2.4 plots the probability of college enrollment and completion for each high school quality quartile. I find a positive correlation between enrollment and completion probabilities and high school quality.

2.3   Model

The structural model consists of two stages, and individuals make at most four choices: an enrollment decision $E_{t=1}$ and three continuation decisions D_t when they are enrolled in college $(2 \leq t \leq 4)$. Figure 2.1 displays the timeline of the model. The model allows individuals to drop out and start working at the beginning of sophomore, junior, and senior year. If an individual decides not to attend college or drop out of college, the individual starts working and makes no further schooling choices.

High school quality affects the enrollment decision through two channels. First, it affects the quality of college admitted, SAT score, and college grades. It is likely that high quality schools better prepare students academically for college-level studies. Second, high schools affect their students by informing about the college option, such as college counseling. This model partly captures this effect by allowing q_{hs} to affect the utility of attending a college.

In the first stage, an individual decides to attend college or to work based on the quality of high school q_{hs} and the ability A_i as measured by the a standardized mathematics test score. SAT score is determined by one's ability A_i, high school quality q_{hs}, and family income y_{fam}. Given a SAT score and individual- and school-level characteristics, the quality of college q_c is determined. If an individual decides to attend college, the first-year grade G_{i1} is determined conditional on individual characteristics A_i, q_{hs}, SAT_i, y_{fam}, race, and quality of college. I do not model the application and admission process explicitly.

In the second stage, conditional on the enrollment an individual decides whether to stay or drop out of college depending on the individual and school-level characteristics and the grade G_{it}. I assume that students cannot come back to college after dropping out.

2.3.1 Preference

Utility of enrolling in college c of quality q_c is the following

$$u_i\left(q_c\right) = \bar{u}_i\left(X_i, Z_c; \alpha_x, \alpha_w\right) + \epsilon_{i,c}$$

X_i is a vector of individual characteristics including q_{hs}, ability, SAT score, GPA, family income, race, and type r. Z_c is a vector of college-level characteristics such as q_c and the net cost. I assume that $\epsilon_{i,c}$ follows a Type 1 Extreme Value distribution.

It is likely that there exist information only observed by a student when making decisions in the model. Let α_i be the individual specific preference for a college option. If there exists only one type of individuals then α_i is the same for all i. When I allow for unobserved heterogeneity and add multiple types of individuals, α_i depends on type r. I assume that the probability of being a certain type depends on the family income. Probability of being a type r is

$$\Pr\left(type_i = r\right) = \Lambda\left(\gamma_0 + \gamma_1 low\_inc_i\right)$$

where $\Lambda\left(\cdot\right)$ is the logistic function.

2.3.2 Labor Market

At $t = 5$, individuals who finish college enter the labor market. Value of starting to work at time $t = 5$ is

$$V_{s=2,t=5}\left(q_c, G_i, \epsilon_{2,5}\right) = \overline{V}_{2,5}^0\left(q_c, G_i\right) + \epsilon_{2,5}^0$$

$$= \lambda \sum_{\tau=t+17}^{65} \beta^{\tau-t-17} y_{i\tau} + \epsilon_{2,5}^0$$

where $\overline{V}_{2,t}^0$ is the expected present value of lifetime income net of unobservable preference when i starts working at t. Throughout the model, I assume that both $\epsilon_{s,t}^0$ and $\epsilon_{s,t}^1$ follow a Type 1 Extreme Value distribution. $y_{i\tau}$ is the expected income of τ-year old individual i at time t ($\tau = t + 17$). This income depends on the schooling choices. For individuals who decided to attend college and graduate ($E_{t=1} = 1$ and $D_t = 1, \forall t \geq 2$) (i.e., college graduates), income at age τ is

$$y_{i\tau} = y_{i\tau}\left(q_c, G_{it}, t, s_{it}, X_i\right), t \geq 5$$

where s_{it} is the total number of years spent at college, and X_i includes gender, race, and age.

2.3.3 Stage 2: College Continuation

I solve the model backwards starting from the last period ($t = 4$) in the second stage. In this stage, if there exist a period t such that present value of lifetime utility of staying in school is lower than that of working, an individual drops out of college at t and goes to work. An individual makes at most three choices in this stage.

At $t = \{2, 3, 4\}$, an individual enrolled in college compares the present value of lifetime utilities of staying ($D_{it} = 1$) and the value of outside option to work ($D_{it} = 0$):

$$V_{s=2,t}\left(q_c, G_{it}, t, \epsilon_{2,t}\right) = \max\left\{\overline{V}_{2,t}^1\left(q_c, G_{it}\right) + \epsilon_{2,t}^1; \overline{V}_{2,t}^0\left(q_c, G_{it}\right) + \epsilon_{2,t}^0\right\}$$

Value of staying at college at time t is

$$\overline{V}_{2,t}^1\left(q_c, G_{it}\right) + \epsilon_{2,t}^1 = u\left(q_c\right) + \beta E\left[V_{2,t+1}\left(q_c, G_{i(t+1)}, t+1, \epsilon_{2,t+1}\right) \mid D_{it} = 1, q_c, G_{it}\right] + \epsilon_{2,t}^1$$

where the expectation is taken over the future cumulative grade $G_{i(t+1)}$ and $\epsilon_{2,t+1}$'s. I can rewrite the expectation term as

$$E\left[\max\left\{\overline{V}_{2,t+1}^1\left(q_c, G_{i(t+1)}\right) + \epsilon_{2,t+1}^1; \overline{V}_{2,t+1}^0\left(q_c, G_{i(t+1)}\right) + \epsilon_{2,t+1}^0\right\} \mid D_{it} = 1, q_c, G_{it}\right]$$

$$= \sum_{G_{i(t+1)}} E_\epsilon\left[\max\left\{\overline{V}_{2,t+1}^1\left(q_c, G_{i(t+1)}\right) + \epsilon_{2,t+1}^1; \overline{V}_{2,t+1}^0\left(q_c, G_{i(t+1)}\right) + \epsilon_{2,t+1}^0\right\} \mid D_{it} = 1, q_c\right]$$

$$\cdot Pr\left(G_{i(t+1)} \mid G_{it}, q_{hs}, q_c, A_i, SAT_i\right)$$

Rust (1987) showed that this expectation expression has a following closed form solution:

$$E_\epsilon\left[\max\left\{\overline{V}_{2,t+1}^1\left(q_c, G_{i(t+1)}\right) + \epsilon_{2,t+1}^1; \overline{V}_{2,t+1}^0\left(q_c, G_{i(t+1)}\right) + \epsilon_{2,t+1}^0\right\} \mid D_{it} = 1, q_c\right]$$

$$= \gamma + \ln\left[\exp\left(\overline{V}_{2,t+1}^1\left(q_c, G_{i(t+1)}\right)\right) + \exp\left(\overline{V}_{2,t+1}^0\left(q_c, G_{i(t+1)}\right)\right)\right]$$

where γ is the Euler constant. Combining above expressions together I can rewrite $\overline{V}_{2,t}^1\left(q_c, G_{it}\right)$ as

$$\overline{V}_{2,t}^1\left(q_c, G_{it}\right) = u\left(q_c\right) + \beta \sum_{G_{i(t+1)}}\left\{\gamma + \ln\left[\exp\left(\overline{V}_{2,t+1}^1\left(q_c, G_{i(t+1)}\right)\right) + \exp\left(\overline{V}_{2,t+1}^0\left(q_c, G_{i(t+1)}\right)\right)\right]\right\}$$

$$\cdot Pr\left(G_{i(t+1)} \mid G_{it}, q_{hs}, q_c, A_i, SAT_i\right)$$

Value of dropping out and getting a job at time t is

$$\overline{V}_{2,t}^0 (q_c, G_{it}) + \epsilon_{2,t}^0 = \lambda \sum_{\tau=t+17}^{65} \beta^{\tau-t-17} y_{i\tau} (q_c, G_{it}, A_i, t, X_i) + \epsilon_{2,t}^0$$

where $\overline{V}_{2,t}^0$ is the present value of lifetime income net of unobservable preference when i starts working at t. An individual chooses to stay in college if $\overline{V}_{2,t}^1 (q_c, G_{it}) + \epsilon_{2,t}^1 \geq \overline{V}_{2,t}^0 (q_c, G_{it}) + \epsilon_{2,t}^0$. The probability of staying in college is

$$Pr\,(D_{it} = 1) = \frac{\exp\left(\overline{V}_{2,t}^1\right)}{\exp\left(\overline{V}_{2,t}^1\right) + \exp\left(\overline{V}_{2,t}^0\right)}, 2 \leq t \leq 4$$

2.3.4 SAT Score, College Quality, and College Grades

When an individual decides to attend college, a SAT score is drawn from a probability distribution based on $X_{hs} = \{q_{hs}, A_i, y_{fam}, sex, race\}$. I discretize SAT scores into J categories and assume that SAT_i follows the logit probabilities.

$$Pr\,(SAT_i = j \mid E_i = 1, q_{hs}, A_i, y_{fam}, female) = \frac{\exp\left(X_{hs}\rho_j\right)}{\sum_{l=1}^{J} \exp\left(X_{hs}\rho_l\right)}$$

I characterize colleges by their qualities and discretize into L categories. The lowest category, $L = 1$, belongs to the two-year colleges since most of them have an open admissions policy. I assume that the school's optimal admission rule results in logit probabilities. The probability of being accepted to a $L = l$ quality college is

$$Pr\,(q_c = l \mid E_i = 1, q_{hs}, A_i, SAT_i, y_{fam}, female) = \frac{\exp\left(X_c\delta_l\right)}{\sum_{j=1}^{L} \exp\left(X_c\delta_j\right)}$$

where X_c is a vector of school and individual characteristics including $q_{hs}, A_i, y_{fam}, female,$

and SAT_i.

Conditional on the enrollment, the model determines the first-year college grade. I discretize cumulative grades G_{it} into K values. Let $G_{it}^* = X_g'\eta + \nu_i$ be a latent variable for G_{it}. Define $G_{it} = g^{(k)}$ if $\mu_{k-1} < G_{it}^* < \mu_k$ where $\mu_0 = -\infty$, $\mu_K = \infty$, and $g^{(k)}$ being the kth possible grade. I assume that ν_i follows a standard normal distribution. Using the result from the ordered probit model, the probability of $G_{it} = g^{(k)}$ is the following for $2 \leq t \leq 4$:

$$Pr\left(G_{it} = g^{(k)} \mid X_g = \left(G_{i(t-1)}, t \cdot G_{i(t-1)}, q_{hs}, q_c, A_i, SAT_i, female\right)\right)$$

$$= \Phi\left(\mu_k - X_g\eta\right) - \Phi\left(\mu_{k-1} - X_g\eta\right), 1 \leq k \leq K$$

where μ_k is a threshold parameter strictly increasing in k, and Φ is a cumulative distribution function of a standard normal distribution.

In the case of the first-year grade, X_g does not include previous cumulative grade and its interaction with time t, $G_{i(t-1)}$ and $t \cdot G_{i(t-1)}$. I estimate two sets of threshold parameters μ_k's separately: one for freshman year and another common set of thresholds among sophomores, juniors, and seniors.

2.3.5 Stage 1: College Enrollment

In the first stage $(t = 1)$, an individual chooses to attend college or not by comparing the present value of lifetime utilities from attending college $(E_i = 1)$ and working $(E_i = 0)$:

$$V_{s=1,t=1}\left(q_{hs}, A_i, y_{fam}\right) = \max\left\{\overline{V}_{1,t}^1\left(q_{hs}, A_i, y_{fam}\right) + \epsilon_{1,t}^1; \overline{V}_{1,t}^0\left(q_{hs}, A_i, y_{fam}\right) + \epsilon_{1,t}^0\right\}$$

Value of enrolling in college is

$$\overline{V}^1_{s=1,t=1}(q_{hs}, A_i) + \epsilon^1_{1,t} = \sum_{SAT_i} \sum_{q_c} \sum_{G_{i1}} \{u(q_c) + \beta E\left[V_{s=2,t+1}(q_c, G_{it+1}, t+1, \epsilon_{2,t+1}) \mid q_{hs}, A_i\right]\}$$

$$\times Pr\left(G_{i1} \mid q_{hs}, q_c, A_i, SAT_i\right) \times Pr\left(q_c \mid enroll\_college = 1, q_{hs}, A_i, SAT_i, y_{fam}\right)$$

$$\times Pr\left(SAT_i \mid enroll\_college = 1, q_{hs}, A_i, y_{fam}\right) + \epsilon^1_{1,t}$$

Similar to the stage 2, I replace the conditional expectation term with the following closed form expression and rewrite $\overline{V}^1_{s=1,t=1}(q_{hs}, A_i)$ as

$$\overline{V}^1_{s=1,t=1}(q_{hs}, A_i) = \sum_{SAT_i} \sum_{q_c} \sum_{G_{it}} \left\{u(q_c) + \gamma + \ln\left[\exp\left(\overline{V}^1_{2,t+1}\left(q_c, G_{i(t+1)}\right)\right) + \exp\left(\overline{V}^0_{2,t+1}\left(q_c, G_{i(t+1)}\right)\right)\right]\right\}$$

$$\times Pr\left(G_{i(t=1)} \mid q_{hs}, q_c, A_i, SAT_i\right) \times Pr\left(q_c \mid enroll\_college = 1, q_{hs}, A_i, SAT_i, y_{fam}\right)$$

$$\times Pr\left(SAT_i \mid enroll\_college = 1, q_{hs}, A_i, y_{fam}\right)$$

Value of starting to work is

$$\overline{V}^0_{1,t}(q_{hs}, A_i) + \epsilon^0_{1,t} = \lambda \sum_{\tau=18}^{65} \beta^{\tau-18} y_{i\tau}(A_i, X_i) + \epsilon^0_{1,t}$$

An individual decides to enroll if $\overline{V}^1_{1,t}(q_{hs}, A_i, y_{fam}) + \epsilon^1_{1,t} \geq \overline{V}^0_{1,t}(q_{hs}, A_i, y_{fam}) + \epsilon^0_{1,t}$. The probability of enrolling in college is

$$Pr\left(E_i = 1\right) = \frac{\exp\left(\overline{V}_{1,t}^1\right)}{\exp\left(\overline{V}_{1,t}^1\right) + \exp\left(\overline{V}_{1,t}^0\right)}, t = 1$$

2.4   Estimation

2.4.1   Maximum Likelihood Estimation

The model is estimated with maximum likelihood. The outcomes of the model are E_i, SAT_i, q_c, D_{it}, and G_{it}, college enrollment, SAT score, college quality, continuation decision, and college grades, respectively. I categorize the parameters θ into two groups: utility parameters θ_u and transition parameters θ_f.

Table 2.2 lists the structural parameters of the model. Utility parameters include α_E, α_D, and α_c. Note that utility parameters include coefficient on unobserved types. Transition parameters include coefficients on X_{hs} and on X_c, grade threshold parameters μ's and coefficients on X_g, η's, and the parameters of the probability of being type r.

Individual likelihood contribution is

$$L_i\left(\theta\right) = Pr\left(E_i \mid q_{hs_i}, A_i, y_{fam}\right) \times f\left(SAT_i \mid E_i = 1, q_{hs_i}, A_i, y_{fam}\right)$$

$$\times \sum_{r=1}^{R} \{f\left(q_{c_i} \mid E_i = 1, q_{hs_i}, A_i, SAT_i, y_{fam}\right) \times f\left(G_{i1} \mid q_{c_i}, E_i = 1, q_{hs_i}, A_i, SAT_i, y_{fam}\right)$$

$$\times \prod_{t=2}^{4} [Pr\left(D_{it} \mid G_{it}, q_{c_i}, E_i = 1, q_{hs_i}, A_i, SAT_i, y_{fam}, type_i\right)$$

$$\times f\left(G_{it} \mid G_{i(t-1)}, q_{c_i}, E_i = 1, q_{hs_i}, A_i, SAT_i, y_{fam}\right)\Big]\Big\} Pr\left(type_i = r\right)$$

The maximum likelihood estimator $\hat{\theta}_{MLE}$ maximizes the log-likelihood function $\mathcal{L}$,

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \log\left[\prod_i L_i\left(\theta\right)\right]$$

Decisions in each stage additively contribute to the total likelihood function. Hence, I can break down the log-likelihood function $\mathcal{L} = \log\left[\prod_i L_i\left(\theta\right)\right]$ into separate likelihood contributions.

$$\mathcal{L}\left(\theta\right) = \mathcal{L}_{SAT}\left(\rho\right) + \mathcal{L}_{q_c}\left(\delta\right) + \mathcal{L}_{G_1}\left(\mu_{G_1,1}, \ldots, \mu_{G_1,K-1}, \eta_{G_1}\right)$$

$$+\mathcal{L}_G\left(\mu_1, \ldots, \mu_{K-1}, \eta\right) + \mathcal{L}_D\left(\alpha_D, \alpha_c, \lambda, \mu_1, \ldots, \mu_{K-1}, \eta\right)$$

$$+\mathcal{L}_E\left(\alpha_E, \alpha_D, \alpha_c, \lambda, \rho, \delta, \mu_{G_1,1}, \ldots, \mu_{G_1,K-1}, \eta_{G_1}, \mu_1, \ldots, \mu_{K-1}, \eta\right)$$

I first obtain the estimates of transition parameters $\left(\hat{\rho}, \hat{\delta}, \hat{\mu}'s, \hat{\eta}\right)$ by maximizing $\mathcal{L}_{SAT}$, $\mathcal{L}_{q_c}$, $\mathcal{L}_{G_1}$, and $\mathcal{L}_G$ then use the estimates to estimate utility parameters $(\alpha_E, \alpha_D, \alpha_c, \lambda)$.

2.4.2   Identification

Utility and type parameters are identified by the variation in choices of students including dropout and continuation decisions. I use the average tuition level of students' residence in high school which is assumed to be independent of unobserved type to identify the parameter

64

on the cost. I control for various individual characteristics including an ability measure, test score, family background, and demographics. If two individuals have similar characteristics, the model would predict that they should make the same choices over time. If the model prediction does not match the data, I would attribute the difference to the unobserved heterogeneity which is only observed to each individual when making choices.

Transition parameters including SAT scores, college grades, and college quality are estimated outside the model. I assume that unobserved types affect these outcomes indirectly through educational choices. This is a stronger assumption then allowing unobserved heterogeneity to affect these outcomes directly. Under this assumption, I use the observed data from the NELS88 to estimate the transition parameters.

2.5 Results

The model results match the overall data pattern quite well. Figure 2.5 shows the educational choices over time of the full sample. Individuals gradually work more and enroll in college less. Figure 2.6 displays the schooling choice transitions. The model matches the educational outcomes in the data by high school quality and by different educational groups as shown in Figures 2.7 and 2.8. The model results for the SAT scores and GPA show patterns similar to the actual data. Given that a student is enrolled in college, the math score is higher, the fraction of low family income background is lower, and the fraction of non-black and non-Hispanic population is higher in the simulated data set as in the actual data set from the NELS88. This outcome is similar for the group of students who completed college as shown in Figure 2.8.

As discussed in the estimation section, I first estimate the transition parameters and then I take the estimates as given to estimate the utility parameters. Tables 2.3-2.6 present the estimates of the transition parameters in the order of SAT score, college quality, and

first-year grade. Then I present the utility and type parameter estimates in Tables 2.7 and 2.8, respectively.

2.5.1 Transition Parameters

SAT score   An individual's SAT score is determined conditional on the enrollment decision and individual- and school-level characteristics. I discretize the scores into three categories with $J = 3$ being the highest score. Cutoff points are $(0, 400, 600, 800)$. Table 2.3 presents the logit SAT estimates. Both the individual ability and the high school quality have a positive effect on the probability of achieving the highest SAT score ($J = 3$). Family income and race do not have a statistically significant effect on the probabilities of receiving the highest SAT score ($J = 3$). Figure 2.9 shows the probabilities of receiving a SAT score between 400 and 600. A non-Hispanic and non-black student with a mean ability from a low-income background has 49 percent and 85 percent chances of receiving a SAT score between 400 and 600 when attending a high school quality of 40 and 60, respectively. The same student from a high-income background has 54 percent and 84 percent chances of achieving the same SAT score when attending a school quality of 40 and 60, respectively.

College quality   Conditional on the enrollment decision and a SAT score, a college quality is determined by logit probabilities. I allow the college quality to be low or high. Table 2.4 shows the logit college quality estimates. An individual's SAT score and ability have a positive and statistically significant effect on the probability of acceptance to a high-quality college. Coefficients on race and family income are not statistically significant. Figure 2.10 displays the adjusted prediction of the logit result.

College grades   Conditional on the enrollment, a SAT score, and college quality, the first-year grade is determined by the ordered probit probabilities. I discretize the grades into four values with $K = 4$ being the highest grade. Cutoff points are $(0, 1, 2, 3, 4)$. Table 2.5 presents the first-year grade estimates. The individual's ability and a SAT score have a positive effect on the probability of earning a higher grade. Coefficient on a high school quality is no longer statistically significant, but the quality does have an indirect effect on the grade through a SAT score and quality of college. Coefficients on race are not statistically significant. Upper class grade estimates are shown in Table 2.6. After controlling for previous grades in college, the quality of high school no longer has a statistically significant effect.

2.5.2   Utility and Type Parameters

I obtain utility parameter estimates after taking the estimates of the transition parameters as given. Table 2.7 shows the parameter estimates and standard errors.[1] Quality of high school has a positive effect on the utility of college enrollment. Attending a high-quality college yields a large positive utility benefit. Parameter on a type 1 has a positive effect on attending college implying that a large portion of educational choice is not explainable by observed characteristics. A type 1 individual is more likely to attend college than a type 2 individual, holding other individual characteristics constant. A Hispanic student has a more negative utility than a non-Hispanic and non-black student. A black student has a slightly positive utility compared to a non-Hispanic and non-black student.

Table 2.8 shows type parameter estimates and standard errors. The probabilities of being type 1 for a student from low- or high-income are 42.4 percent and 59.7 percent, respectively.

---

[1]Standard errors are calculated through bootstrapping procedure with 1000 replications.

2.6 Policy Counterfactuals

After estimating the model, I investigate the outcomes generated from the model simulations. In the model, public school quality directly affects the probability of receiving a high SAT score, the probability of entering a high-quality college, and the probability of getting a higher first-year grade. Thus, public school quality affects the educational outcomes by affecting students' college preparedness. This section discusses three policies regarding public high school quality to understand the effect of school quality on students' educational outcomes.

2.6.1 Uniform Quality

The first counterfactual investigates a hypothetical policy which restricts the quality of public schools to ensure uniformity regardless of the school districts. More centralized public education system and more equal distribution of school funding would have less variation in quality among schools. I restrict the quality of public schools at the median level and simulate and track students' educational choices. Then I compare the enrollment and dropout rates with the baseline result.

Centralizing public education results in a higher college enrollment rate of students from low-quality school districts as shown in Figure 2.11 (a). If quality of a high school is far below from the median, increases in enrollment and completion is higher. I find that college enrollment increases by 19.8 percent and college completion increases by 17.6 percent for the students from the lowest quartile quality schools. Enrollment and completion slightly decrease for the students from the top 50th percentile quality schools. Figure 2.11 (b) plots the counterfactual result by family income level. Because of the high correlation between quality of high school and family income level, those who benefit most from this counterfactual policy are the students enrolled in a low-quality high school located in a low-income district.

2.6.2 Lower Bound

The second counterfactual analyzes the effect of a quality threshold for high schools. This policy is more feasible than the first counterfactual that the government could impose this cutoff by providing incentives, direct subsidies, or penalties such as in the No Child Left Behind Act. As described in the introduction, Illinois plans to allocate more funding to districts with low property values, which would potentially have a positive effect on the students in low-quality schools.

I fix the qualities of the lowest quartile high schools to the 25 percentile level ($q_{hs} = 48.69$). Figure 2.12 (a) displays the changes in choices for students by high school quality quartile. Because this model is a partial equilibrium model in which students do not compete for seats in college, students attending high schools with quality above the 25th percentile do not alter their choices. For the bottom 25th percentile schools, I find that college enrollment increases by 9.1 percent and college completion increases by 7.8 percent. Figure 2.12 (b) shows the counterfactual result by family income level. Since the students attending the lowest-quality schools are mostly from the lowest family income background, I find a small increase for low-income families.

2.6.3 Increase in Ability

The third counterfactual investigates a 10 percent increase in the ability of students from the lowest 25th percentile quality schools instead of changing quality of high schools. To understand the effect of the ability, I assume that a high school quality does not change due to the increase in individuals' abilities. Figure 2.13 displays the changes in choices by high school quality and by family income level. Figure 2.13 (a) illustrates that a student from a lowest-quality quartile experiences an increase of 59.5 percent and 33.3 percent for enrollment and completion, respectively. For a student from the second lowest-quality quartile, enrollment

and completion rate increase by 29.5 percent and 31.6 percent. Figure 2.13 (b) shows that directly affecting the ability measure benefits the high-income families, as well. In particular, this policy increases enrollment by 27.6 percent and completion by 16.1 percent for a group of students from a family income below $35,000. This amount is the largest increase among three counterfactuals for this particular family income category. Similar to the second policy counterfactual, since a college market is not modeled, this relatively large effect does not consider the fact that the seats at colleges are limited especially for high-quality colleges.

2.7 Conclusion

This paper investigates how public school quality affects post-secondary school enrollment and completion decisions. In the first stage of the model, an individual chooses to attend college or to work. Conditional on the enrollment, an individual makes continuation decisions while staying in college. I estimate the model with maximum likelihood, and given the estimates, I perform three policy simulations regarding the changes in the quality of high school.

The counterfactual results show that centralizing public education results in a higher college enrollment rate of students from low-quality school districts. Because of the strong correlation between quality of public schools and family income level, those who benefit most from this counterfactual policy are the students enrolled in a low-quality high school located in a low-income district. When a policy imposes a lower bound for the high school quality, I find that college enrollment increases by 9.1 percent and the college completion increases by 7.8 percent for the students attending the lowest quality quartile schools. Lastly, a 10 percent increase in the ability of students from the lowest 25th percentile quality schools increases enrollment and completion by 59.5 percent and 33.3 percent, respectively. The counterfactual results suggest that improvement in the quality of public schools has a modest

effect on college enrollment and completion compared to a direct increase in the ability of the students.

Tables and Figures

Table 2.1: Sample Means

High school quality	Full Sample		Enrolled		Completed	
	Low	High	Low	High	Low	High
Prob. of enrolling	0.58	0.79				
Prob. of completing	0.38	0.59	0.65	0.74		
Standardized math test	52.13	58.50	55.01	59.72	56.57	60.47
	[9.52]	[9.64]	[9.42]	[9.22]	[9.55]	[8.98]
Low family income	0.45	0.24	0.36	0.20	0.33	0.67
Non-black & non-Hispanic	0.75	0.91	0.80	0.92	0.81	0.93
Observations	959	955	561	762	366	564

Notes: Full sample is the high school class of 1992 who participated in 3rd and 4th follow-up surveys. Enrolled sample is the individuals enrolled in full time public or private colleges among the full sample. Completed sample is the individuals who received a bachelor's degree or higher as of 2000. High school quality grouped into high if q_{hs} is greater than the median ($q_{hs} = 52.33$). Low family income equals 1 if a household income is less than \$35,000 in 1991 dollars. Standard deviations are shown in brackets.

Table 2.2: Summary of Structural Parameters

Utility Parameters θ_u			Transition Parameters θ_f		
	Variable	Parameter		Variable	Parameter
Enrollment	q_{hs}	α_1	SAT (J sets)	q_{hs}	ρ_1
choice	A_i	α_2		A_i	ρ_2
	q_c	α_3		y_{fam} (5)	ρ_3
	SAT_i	α_4		$race$ (3)	ρ_4
	G_{it}	α_5		constant	ρ_0
	y_{fam}	α_6	College quality	SAT_i	δ_1
	$race$	α_7		q_{hs}	δ_2
	$cost$	α_8		A_i	δ_3
	type 1	α_9		y_{fam} (5)	δ_4
	constant	α_0		$race$ (3)	δ_5
				constant	δ_0
			First-year grade	q_c	$\eta_{G_1,1}$
				SAT_i ($J=3$)	$\eta_{G_1,2}$
				q_{hs}	$\eta_{G_1,3}$
				A_i	$\eta_{G_1,4}$
				$race$ (3)	$\eta_{G_1,5}$
				thresholds	μ_{G_1}
			Cumulative grade (2nd to 4th-year)	q_c	η_1
				SAT_i ($J=3$)	η_2
				q_{hs}	η_3
				A_i	η_4
				$race$ (3)	η_5
				$G_{i(t-1)}$ ($K=4$)	η_7
				thresholds	μ
			Type probabilities	type 1	γ_1
				constant	γ_0

Table 2.3: Logit SAT Estimates

	J = 1 (Lowest)		J = 3 (Highest)	
	Estimates	Std. Error	Estimates	Std. Error
High school quality q_{hs}	-0.077	0.020	0.050	0.022
Ability A_i	-0.198	0.014	0.254	0.020
Family income category 2	0.281	0.315	0.167	0.493
Family income category 3	0.255	0.305	0.081	0.487
Family income category 4	-0.009	0.303	0.609	0.466
Family income category 5 (Highest)	-0.505	0.350	0.694	0.460
Race: black	0.591	0.350	-1.673	1.205
Race: non-black & non-Hispanic	-0.393	0.430	-0.124	0.521
Constant	13.820	0.285	-21.033	1.818

Notes: $J = 2$ is the base outcome. Hispanic is the baseline race. Number of observation is 1323.

Table 2.4: Logit College Quality Estimates

	J = 2 (High)	
	Estimates	Standard Error
SAT score category 2	0.908	0.173
SAT score category 3 (Highest)	2.614	0.478
High school quality q_{hs}	0.022	0.016
Ability A_i	0.032	0.010
Family income category 2	-0.380	0.254
Family income category 3	-0.173	0.247
Family income category 4	-0.107	0.247
Family income category 5 (Highest)	0.380	0.277
Race: black	0.271	0.364
Race: non-black & non-Hispanic	-0.099	0.247
Constant	-2.548	0.868

Notes: $J = 1$ is the base outcome. Hispanic is the baseline race. Number of observation is 1323.

Table 2.5: Ordered Probit First-year Grade Estimates

	Estimates	Standard Error
High-quality college	0.004	0.072
SAT score category 2	0.333	0.088
SAT score category 3 (Highest)	0.649	0.143
High school quality q_{hs}	0.002	0.006
Ability A_i	0.018	0.004
Race: black	0.001	0.174
Race: non-black & non-Hispanic	0.158	0.115
Threshold $\mu_{G_1,1}$	-0.024	0.370
Threshold $\mu_{G_1,2}$	1.056	0.369
Threshold $\mu_{G_1,3}$	2.375	0.373

Notes: Number of observation is 1323. Hispanic is the baseline race.

Table 2.6: Ordered Probit Year 2-4 Grade Estimates

	Estimates	Standard Error
High-quality college	0.069	0.052
SAT score category2	0.265	0.062
SAT score category 3 (Highest)	0.478	0.098
High school quality q_{hs}	-0.005	0.004
Ability A_i	0.009	0.003
Race: black	-0.125	0.118
Race: non-black & non-Hispanic	0.176	0.080
Previous grade category 2	0.272	0.134
Previous grade category 3	0.971	0.131
Previous grade category 4 (Highest)	2.132	0.137
Threshold $\mu_{G_1,1}$	-0.337	0.279
Threshold $\mu_{G_1,2}$	0.856	0.280
Threshold $\mu_{G_1,3}$	2.428	0.282

Notes: Number of observation is 3164. Hispanic is the baseline race. Ordered probit estimates obtained from pooled data of colleges grades from year 2 to 4.

Table 2.7: Utility Parameter Estimates

Variable	Estimates	Standard Error
High school quality q_{hs}	0.043	0.0017
Ability A_i	0.007	0.0034
High-quality college	0.835	0.0360
SAT score category 2	0.020	0.0414
SAT score category 3 (Highest)	0.480	44.2960
Current GPA	1.117	0.0295
Low family income	-0.519	0.1046
Race: Hispanic	0.148	0.0735
Race: black	0.432	0.1375
Cost	-3.090	0.0524
Constant	-1.474	0.2720
PDV of $y_{i\tau}$	5.593	0.3372
Type 1	1.420	0.6376

Notes: Total number of sample is 1914. Non-black and non-Hispanic is the baseline race. Standard errors are calculated through bootstrapping with 1000 replications.

Table 2.8: Probability of Type Parameter Estimates

Variable	Estimates	Standard Error
Constant	0.393	0.1380
Low income	-0.697	0.2798

Notes: Total number of sample is 1914. Standard errors are calculated through bootstrapping with 1000 replications.

Figure 2.1: Model Timeline
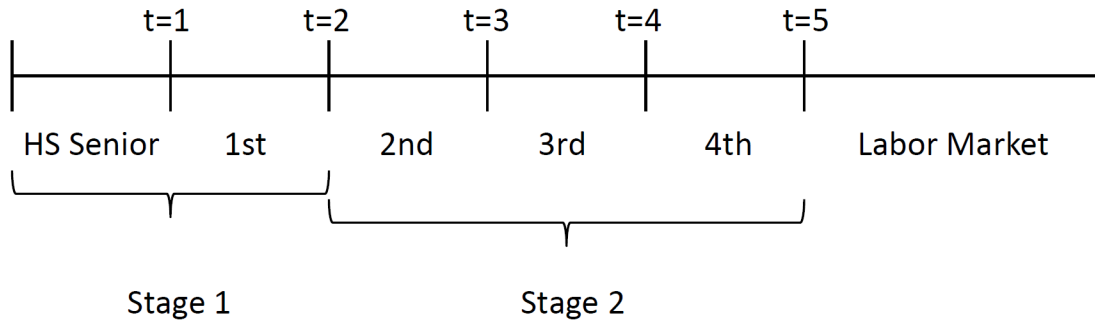
Notes: Students make at most 4 decisions. If a students drops out before $t = 4$, then one starts to work.

Figure 2.2: High School Quality Density
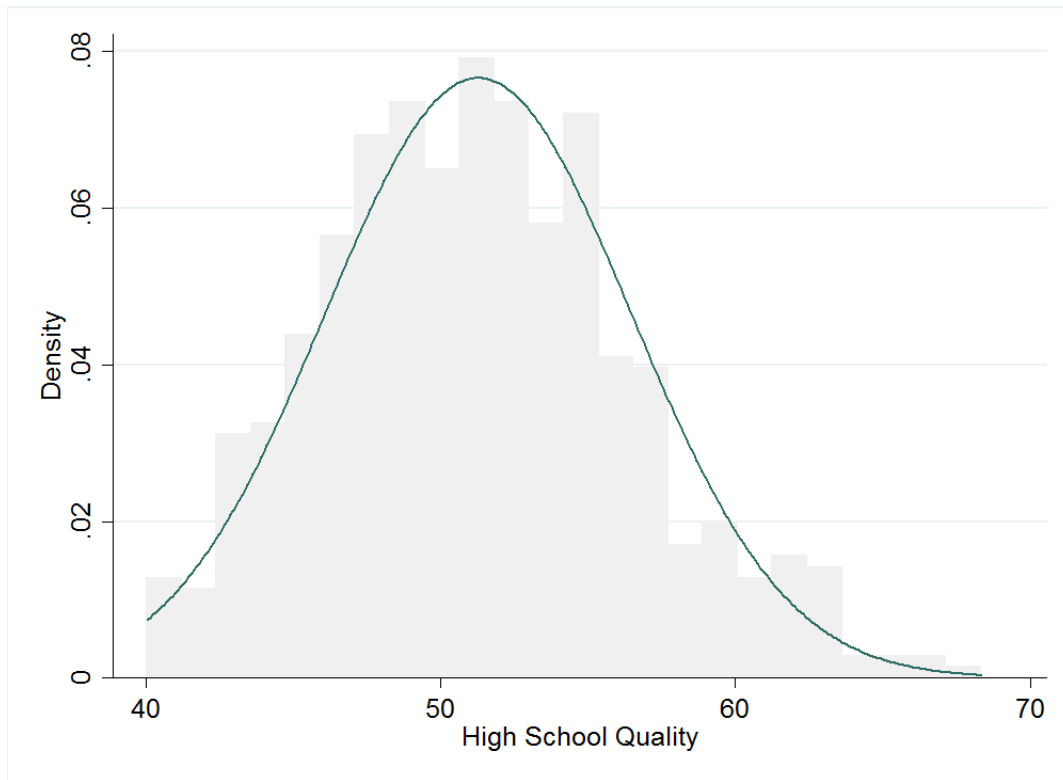
Notes: Total number of high schools in the sample is 600. I use 13,825 observations to compute high school qualities in this sample.

Figure 2.3: Family Income by High School Quality
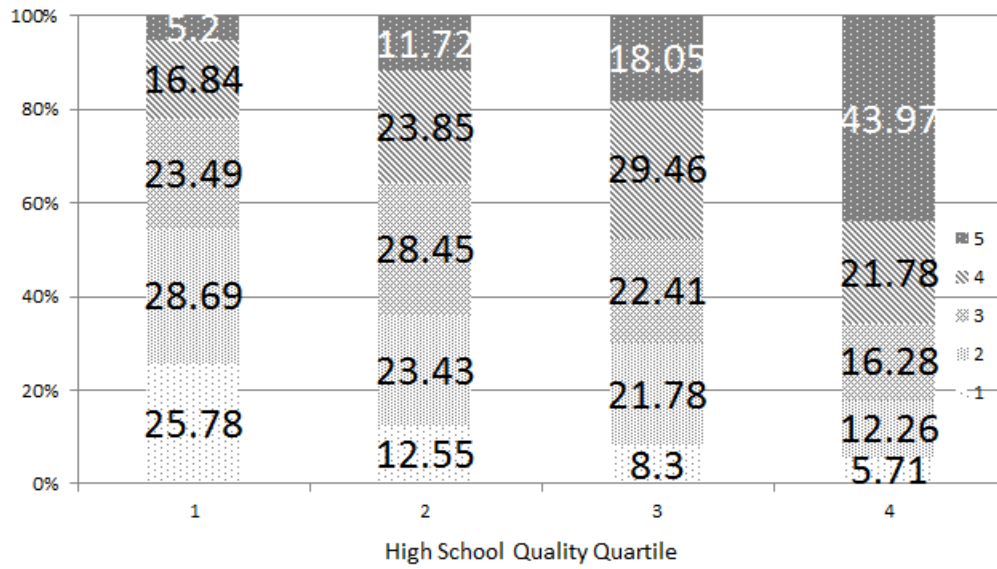
Notes: Family income category 1: below $19,999; 2: [$20,000, $34,999]; 3: [$35,000, $49,999]; 4: [$50,000, $74,999]; 5: above $75,000.

Figure 2.4: Probability of Enrollment and Completion
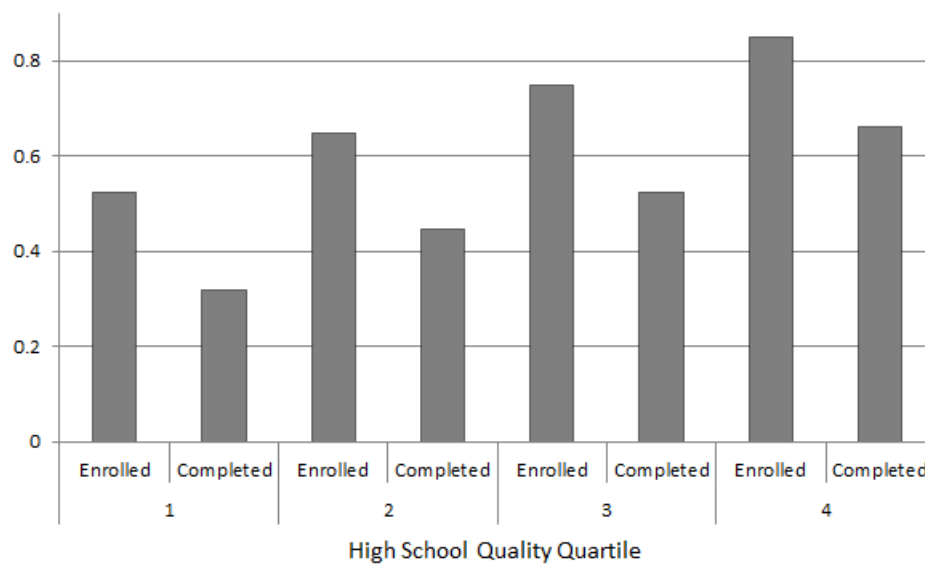
Notes: NELS88 sample.

Figure 2.5: Model Fit: Choices Over Time
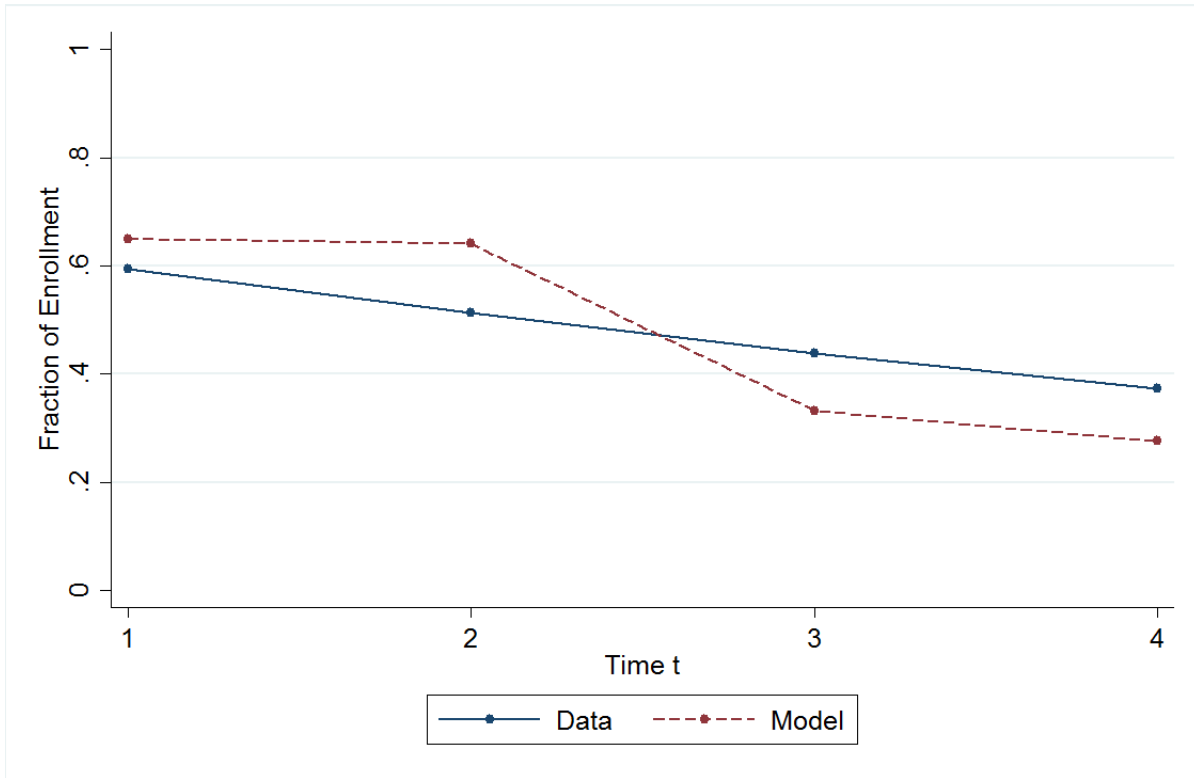
Notes: Fraction of enrollment in college over the full sample.

Figure 2.6: Model Fit: Transition Over Time
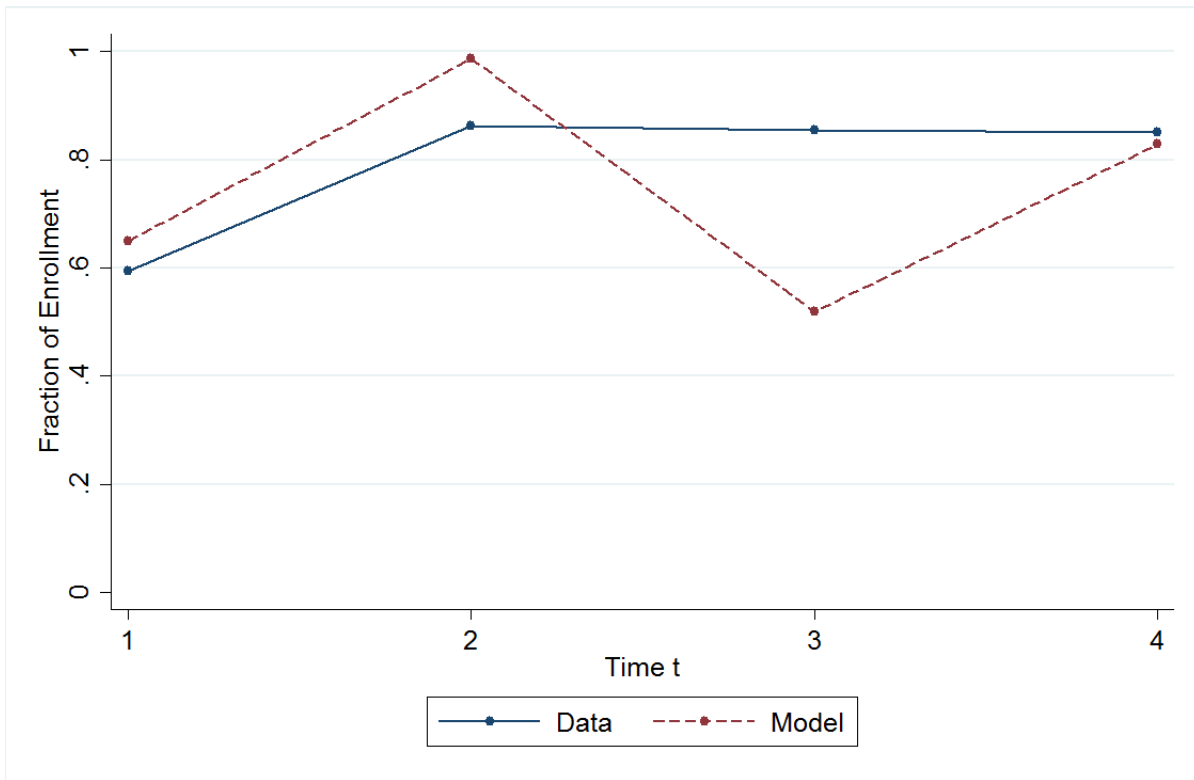
Notes: Fraction of enrollment in college conditional on previous choice.

Figure 2.7: Model Fit: Enrolled Sample by High School Quality
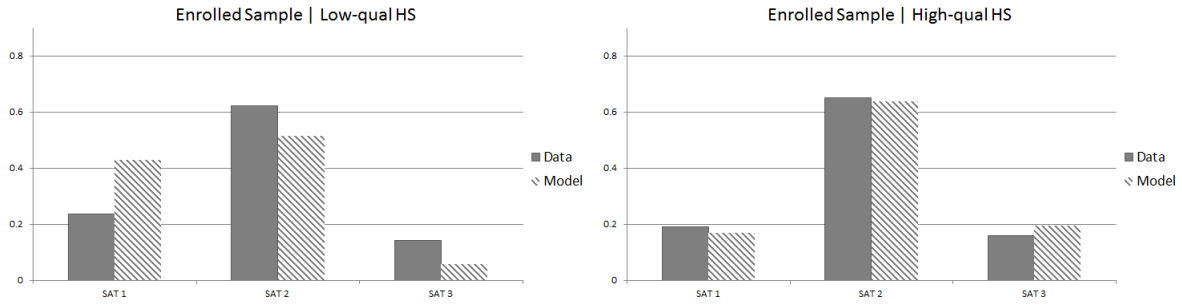
(a) SAT scores
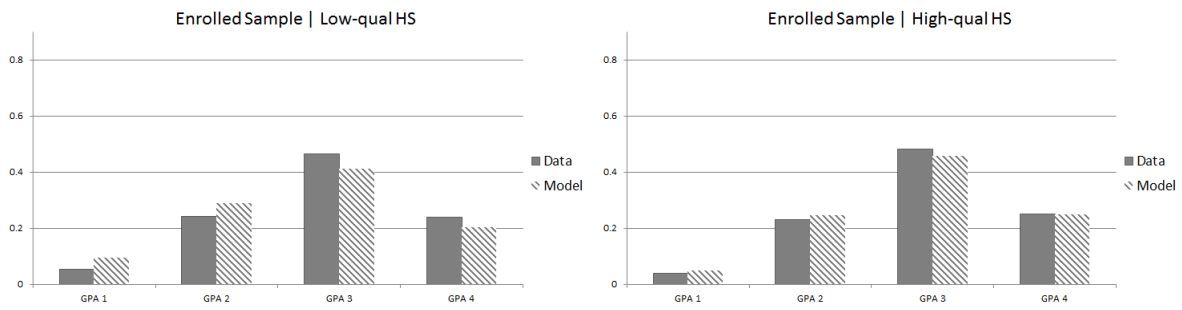
(b) First-year grade
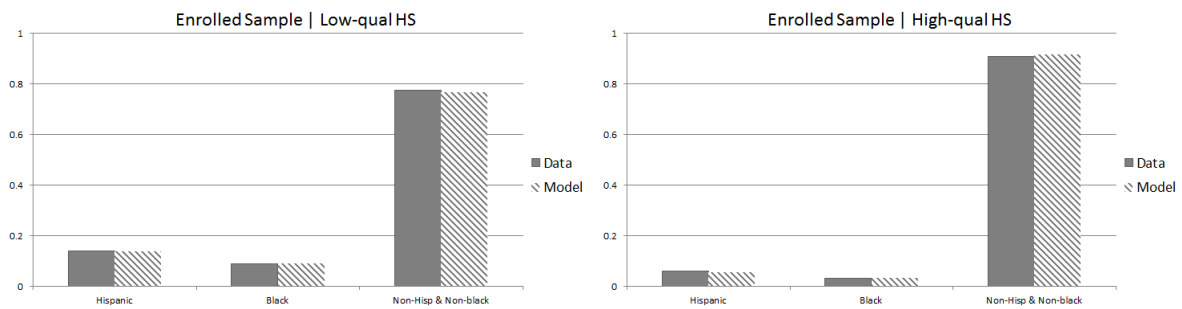
(c) Race

Notes: Sample enrolled in college.

Figure 2.8: Model Fit: Completed Sample by High School Quality
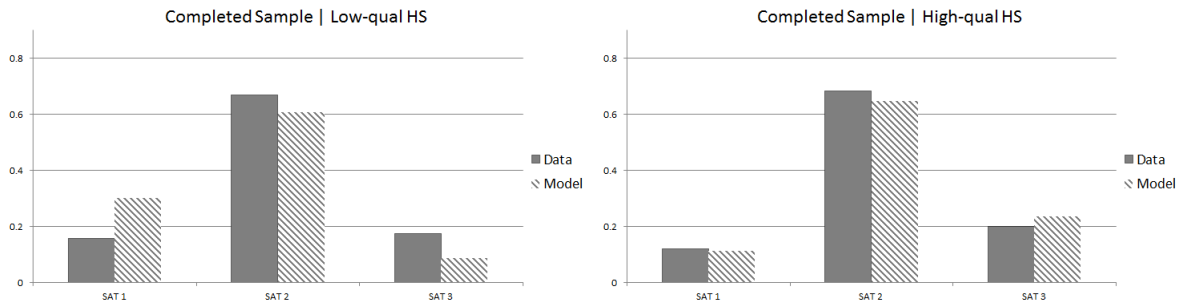
(a) SAT scores

(b) First-year grade
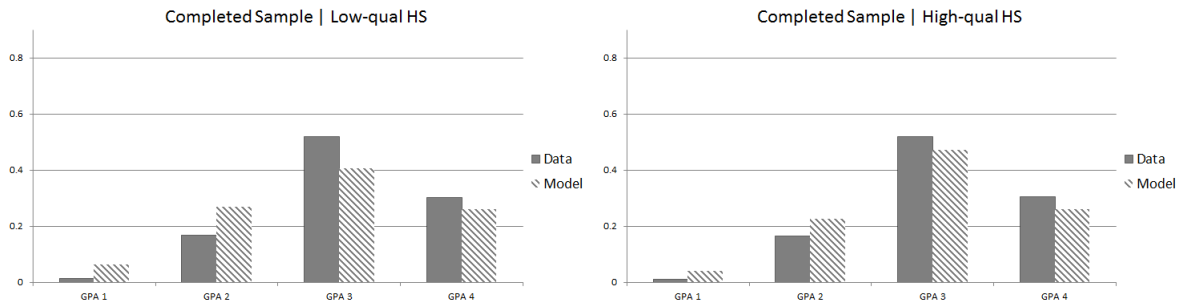
(c) Race
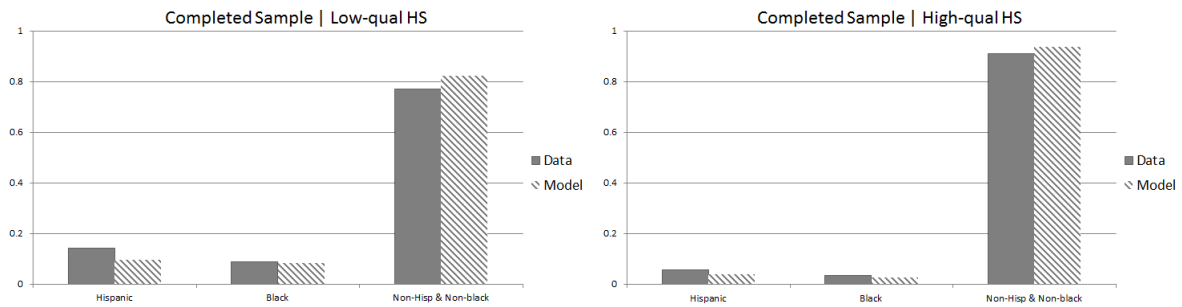
Notes: Sample completed college.

Figure 2.9: Predicted Probability of SAT Score

(a) Low family income

(b) High family income
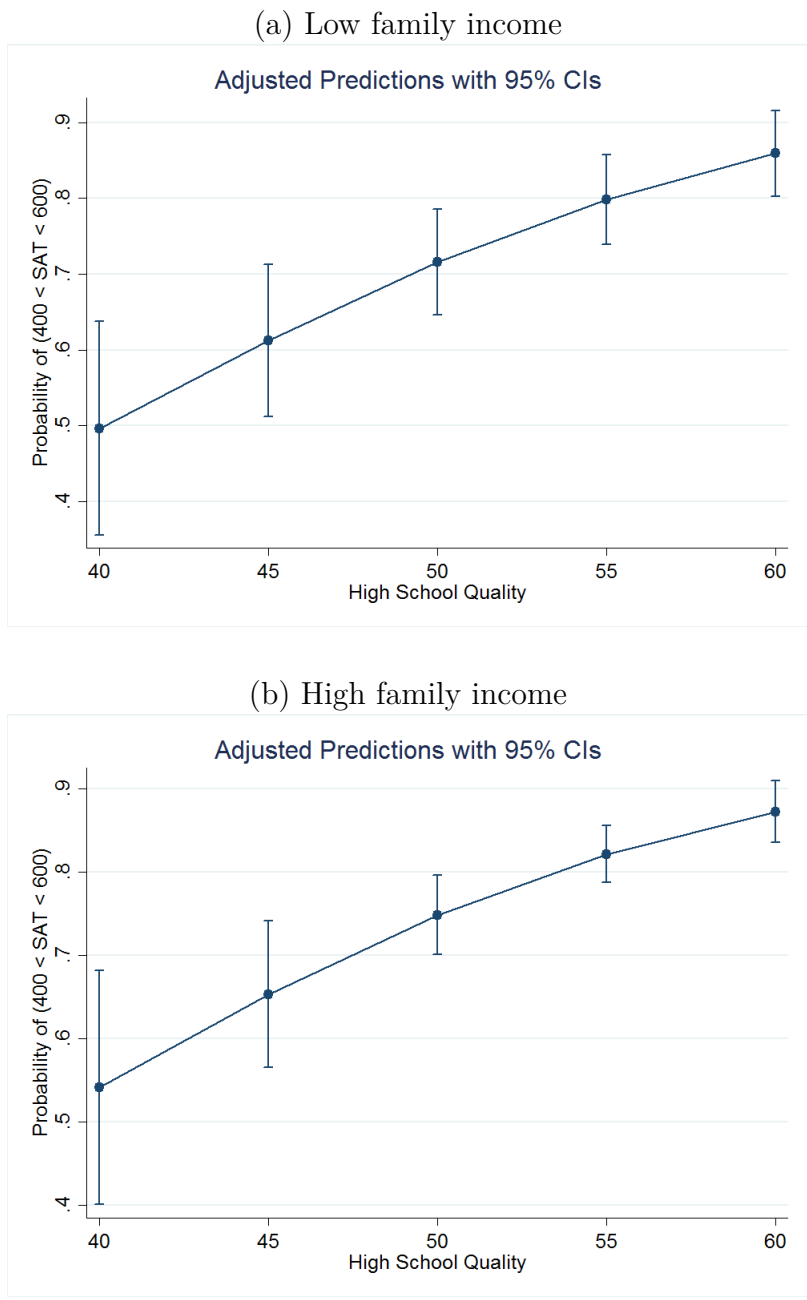
Notes: $Prob\left(400 \leq SAT_i < 600\right)$ (a) below \$35,000; (b) above \$35,000. All other covariates are fixed at mean values.
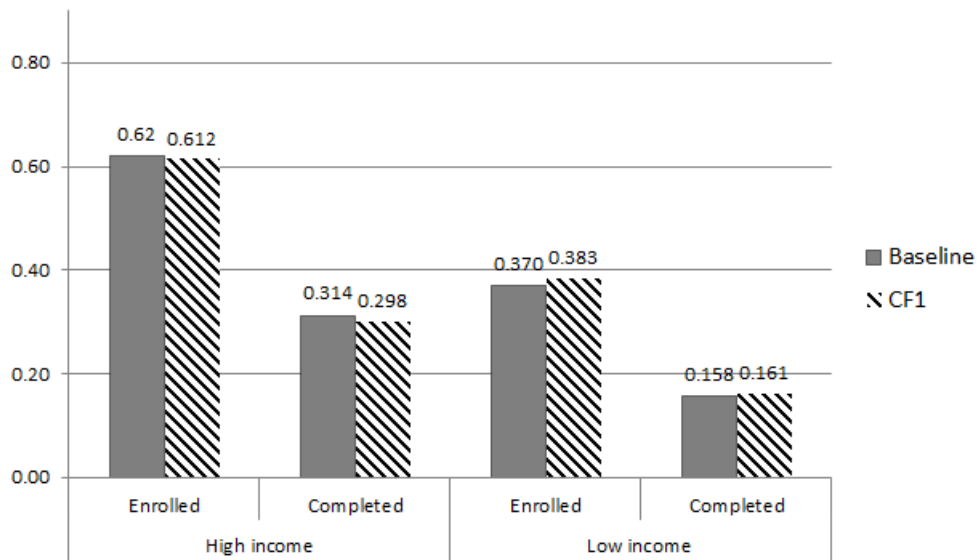
Figure 2.10: Predicted Probability of High-quality College

(a) Low family income

(b) High family income
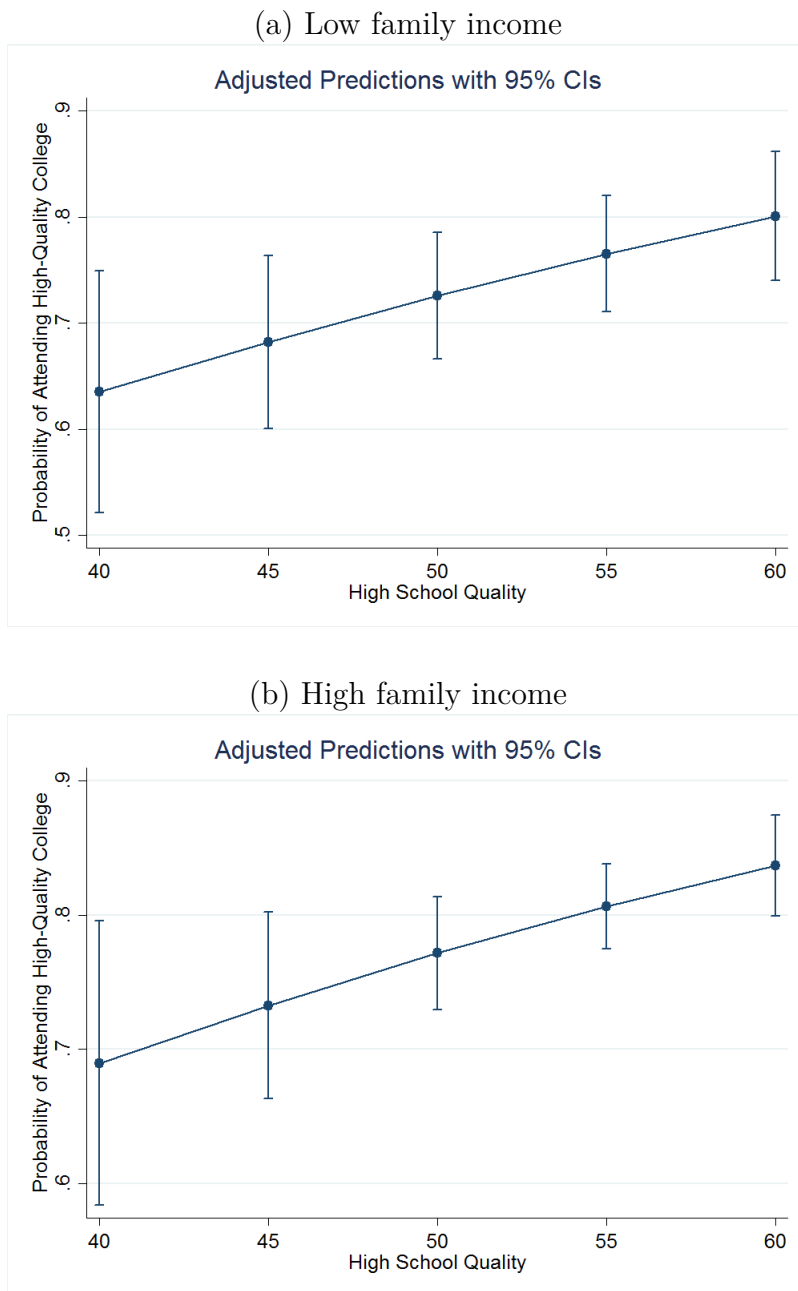
Notes: $Prob\left(550 \leq q_c < 800\right)$ (a) below \$35,000; (b) above \$35,000. All other covariates are fixed at mean values.

Figure 2.11: Counterfactual 1: Choices

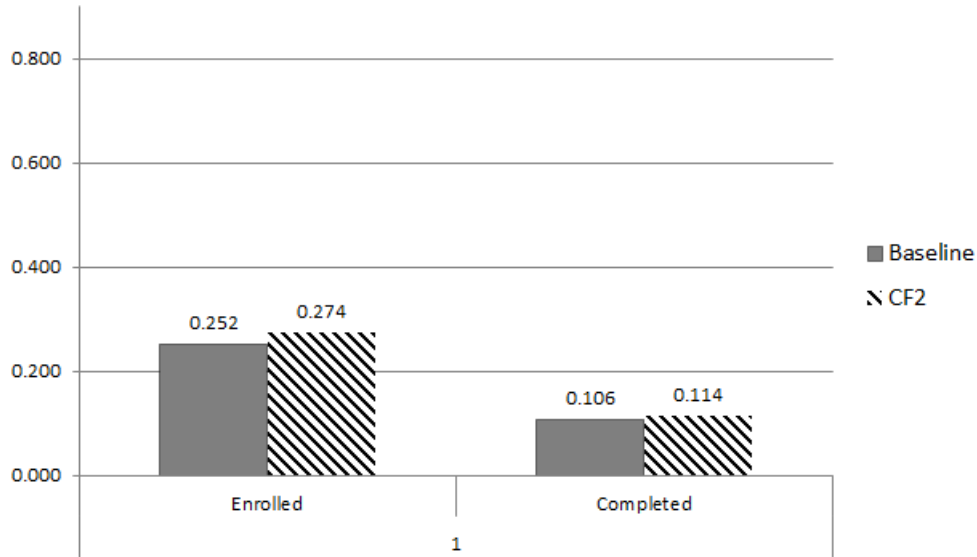(a) Choices by high school quality quartile
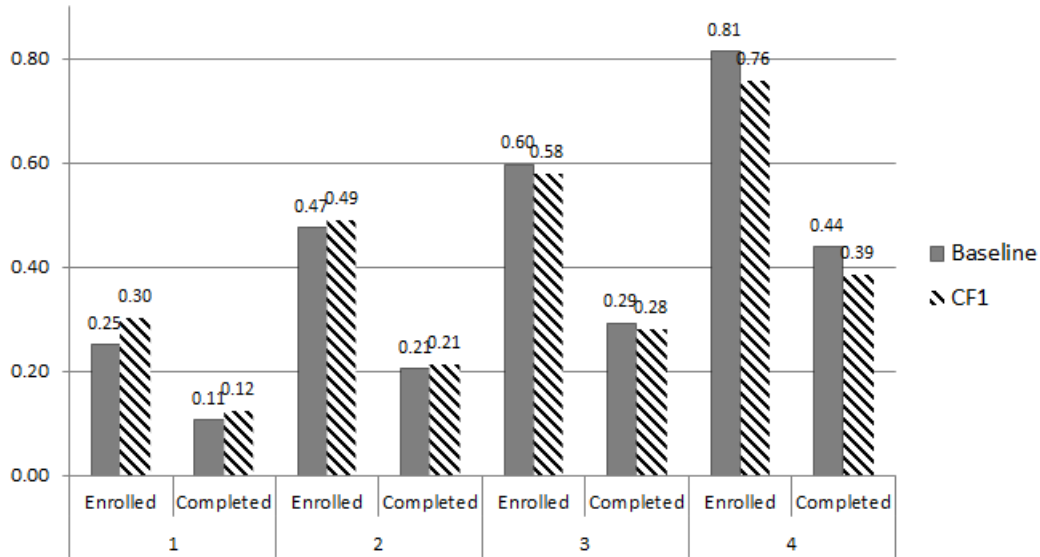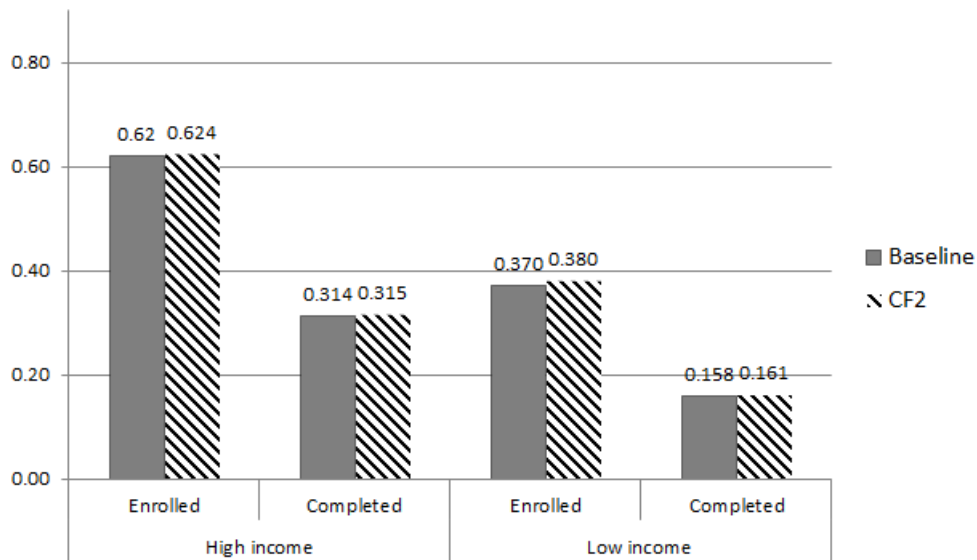

(b) Choices by family income level

Notes: (a) High school quality quartile. 25th percentile: $q_{hs} = 48.69$; 50th percentile: $q_{hs} = 52.30$; 75th percentile $q_{hs} = 55.65$. (b) Low income category if a household income is less than \$35,000.

Figure 2.12: Counterfactual 2: Choices

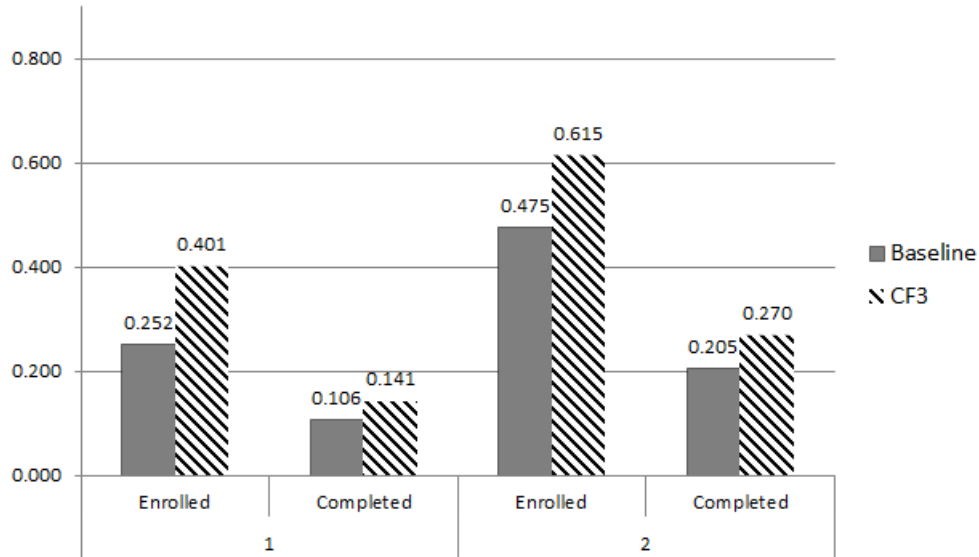(a) Choices by high school quality quartile
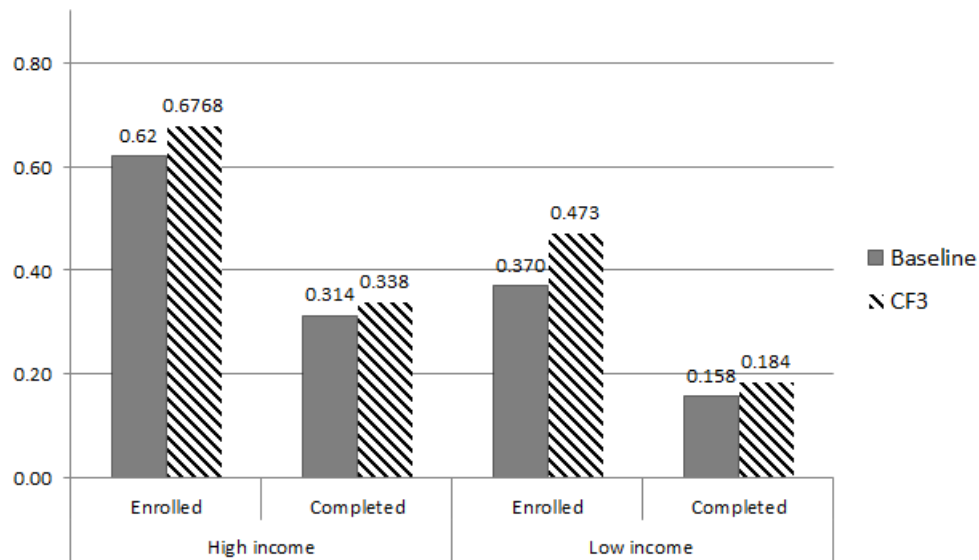

(b) Choices by family income level

Notes: (a) High school quality quartile. 25th percentile: $q_{hs} = 48.69$; 50th percentile: $q_{hs} = 52.30$; 75th percentile $q_{hs} = 55.65$. (b) Low income category if a household income is less than \$35,000.

Figure 2.13: Counterfactual 3: Choices

(a) Choices by high school quality quartile

(b) Choices by family income level

Notes: (a) High school quality quartile. 25th percentile: $q_{hs} = 48.69$; 50th percentile: $q_{hs} = 52.30$; 75th percentile $q_{hs} = 55.65$. (b) Low income category if a household income is less than $35,000.

Bibliography

Anderson, G., Sun, J. C., & Alfonso, M. (2006). Effectiveness of statewide articulation agreements on the probability of transfer: A preliminary policy analysis. *The Review of Higher Education*, *29*(3), 261–291.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, *73*(5), 1477–1524.

Card, D., & Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, *83*(1), 49–82.

Denning, J. T. (2014). College on the cheap: Costs and benefits of community college.

Epple, D., Romano, R., & Sieg, H. (2006). Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, *74*(4), 885–928.

Fu, C. (2014). Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy*, *122*(2), 225–281.

Goodman, J., Hurwitz, M., & Smith, J. (2015). College access, initial college choice and degree completion. Tech. rep., National Bureau of Economic Research.

Guryan, J. (2001). Does money matter? regression-discontinuity estimates from education finance reform in massachusetts. Tech. rep., National Bureau of Economic Research.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1177.

Hastings, J. S., Kane, T. J., & Staiger, D. O. (2005). Parental preferences and school competition: Evidence from a public school choice program. Tech. rep., National Bureau of Economic Research.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? a meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, *23*(3), 5–14.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of Public Economics*, *89*(5), 761–796.

Johnson, M. T. (2013). Borrowing constraints, college enrollment, and delayed entry. *Journal of Labor Economics*, *31*(4), 669–725.

Kane, T. J., & Rouse, C. E. (1995). Labor-market returns to two-and four-year college. *The American Economic Review*, *85*(3), 600–614.

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives*, *16*(4), 91–114.

Keane, M. P., & Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, *105*(3), 473–522.

Krueger, A. B. (2003). Economic considerations and class size. *The Economic Journal*, *113*(485), F34–F63.

Long, B. T., & Kurlaender, M. (2009). Do community colleges provide a viable pathway to a baccalaureate degree? *Educational Evaluation and Policy Analysis*, *31*(1), 30–53.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Roksa, J., & Keith, B. (2008). Credits, time, and attainment: Articulation policies and success after transfer. *Educational Evaluation and Policy Analysis*, *30*(3), 236–254.

Rouse, C. E. (1995). Democratization or diversion? the effect of community colleges on educational attainment. *Journal of Business & Economic Statistics*, *13*(2), 217–224.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, *55*(5), 999–1033.

Stange, K. (2012). An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, *4*(1), 49–84.

Trachter, N. (2015). Stepping stone and option value in a model of postsecondary education. *Quantitative Economics*, *6*(1), 223–256.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, *32*(4), 711–754.