Signals and Human Capital in Admission Tournament

Jane (Jingnan) Liu

Labor Field Paper

Abstract

This paper develops a structural model of pre-college educational investment in college admission tournament. Students are heterogeneous in ability, family wealth, and preferences for colleges and can purchase tutoring service to improve their human capital and test scores. They also face borrowing constraints. The score distribution, admission thresholds, and college assignment are joint equilibrium outcomes. The model is estimated with Korean ELS: 2005 data and can be used to study Korea's tutoring market with a wide range of policy candidates, including taxing private tutoring, reducing noise in admission. A tax lowers the overall spending on tutoring. It is the students from middle-income families that are most responsive to the price change. Reduced signal noise incentivizes the tutoring expenditure of high-ability students, and improves their chances of attending prestigious colleges.

1 Introduction

Scarcity leads to competition. It applies to educational opportunities as well. To get into a top university, being smart is not enough. In the US, wealthy parents can send their kids to expensive private schools or live in a better school district. The latter choice is also costly in the form of higher rents or property taxes. In many other countries, such as Turkey and South Korea, parents compete through private after-school tutoring, which prepares students for college entrance exams. In 2010, Korean families spent 10.7% of their income on such informal education for each student (OECD 2012). The industry has grown exceedingly large. The expenditure on private tutoring amounts to 1.8% of Korean national GDP and 54.6% of the annual budget for public education in 2009 (Statistics Korea, 2010). The public has been complaining about the skyrocketing tutoring expenditure. However, many rounds of reforms proves unsatisfactory.

The Korean case is not unique. In many countries, enrollments of elite universities are inelastic, and students take tutoring to catch up, keep up, or get ahead of their peers in admission. What complicates the design of policies is that individual spending on tutoring, or, more generally, on education, not only depends on observable and unobserved individual characteristics (e.g., preference), but also responds to and influences other students' spending. Another difficulty is due to the dual function of educational investment: it generates genuine human capital and signals in the admission tournament. Take a tax on private tutoring for example. It is likely to decrease tutoring spending. However, it also depresses human capital formation before college and affects the ranking orders and, thus, the college and labor market outcomes of students. The optimal policy need to take these quantitative implications into account.

This paper studies policy designs in a structural model. The model features an admission tournament, in which households purchase tutoring service to compete for the fixed capacities of selective colleges. I also allow tutoring service to influence human capital formation and signals production at the same time. A structural approach is necessary for counterfactual policy evaluations mainly for three reasons. First, the determinants of individual tutoring spending, such as the distribution of tutoring spending and the admission cutoffs, are equilibrium objects that are not invariant to policy changes. Second, it helps quantify the two roles of educational investment, which is needed in the quantitative analysis. Last, the structural model can be used for a wide range of policy experiments, including the ones that have never been tried.

The model captures several salient features of the high-stake college admission process. In the model, selective colleges have a fixed supply of seats, and the admissions is a rank-order tournament depending on the relative signals (Lazear and Rosen, 1981). Students are heterogeneous in ability, family wealth, and preferences for colleges and choose the level of private tutoring investment, which raises one's signals and human capital. The admission probability is determined in equilibrium such that given admission cutoff, the number of seats in selective colleges is equal to the number of attendees. The human capital distribution, admission cutoff, and college assignment are all equilibrium outcomes.

The model quantifies the dual function of tutoring investment. On the one hand, it produces genuine human capital which can be useful in labor market, as emphasized by Becker (1962) and Mincer (1974). On the other hand, it leads to better signals and, hence, admission to a higher ranked college. As the admission depends on the rank of signals, human capital is generally over invested in response to competitive pressure. The two channels can have quite different policy implications. If tutoring works only through the human capital channel, then the existence of a tutoring market is generally good, and the necessary policy intervention is to subsidize tutoring market or to provide students with credit. If private tutoring works only through the signaling channel, then we have two consequences. First, the positional externality of the admission tournament implies overinvestment of private tutoring. Subsidy or cheaper credit may only exacerbate the wasteful investment. Second, with the presence of borrowing constraints, the very existence of the tutoring market can propagate the advantage of wealthy families, and subsidy or cheaper credit may offset some of that distortion in the student-college assignment. It is not true that we can do better by simply cracking down the tutoring market. Any reform about it must balance the two functions and pay attention to distributional consequences.

I estimate the model with a nationally representative sample in South Korea, the Korean Educational Longitudinal Study of 2005. The information on tutoring expenditure, academic performance and post-college outcome allows me to separate the effects of tutoring on human capital and on signals. The unobserved preference for colleges can be revealed from one's tutoring choice conditional on her initial academic performance. I find that both functions of tutoring are economically and quantitatively important.

As a policy experiment, I explore the implications of taxing and subsidizing the tutoring market. The experiment helps understand how peer competition shapes tutoring expenditure, and to what extent tutoring magnifies achievement gap in an admission tournament. Subsidizing the tutoring services increases the overall human capital. Households with medium income are most responsive to the price reduction under subsidy, by increasing their tutoring expenditure. Tax has the opposite effects: tutoring expenditure and human capital get lower. I further evaluate the impact of information. I find that reduced signal noise incentivizes the tutoring expenditure of high-ability students, whereas discourages the low-ability students due to the more "rigid" ranking order. High-ability students are benefited in college assignment.

This paper is related to a small but growing literature on endogenous pre-college human capital formation in an admission tournament. These studies focus on student effort under various admission policies. For example, Bodoh-Creed and Hickman (2018) and Grau (2018) study various forms of affirmative action rules, using B&B data and Chilean administrative data respectively. Arslan (2018) emphasizes the role of preference over colleges using Turkey college admissions data. These studies do not consider borrowing constraint – or to say, they interpret the cost as disutility. Myong (2018) investigates the effects of different scholarship on student effort, and the borrowing constraint is assumed on attending private high schools and colleges. Domina (2007) finds that more access to scholarship in universities leads to increased attendance of advanced courses in high school. As mentioned above, the tutoring services can be bought from the market, so that they are quite different from utilitarian effort costs – the borrowing constraint can interact with wealth levels of students and play an important role in the admission tournament. The first contribution of this paper is to quantitatively analyze variable educational expenditure (i.e., tutoring) with borrowing constraint in an admission tournament. Note that studying tutoring service has another advantage: It is observable in data.

The second contribution of this paper is to separate and quantify the two outcomes of educational investment (including effort): genuine human capital and signals in the admission tournament. Existing studies use academic achievement (such as GPA, test scores) as a proxy for both pre-college human capital and signals. This is fine under two alternative assumptions: First, the two are perfectly correlated. Second, pre-college human capital plays no role in subsequent studies and work. If either assumption is true, or to say, if we do not take for granted that one of these two assumptions is satisfied, then we should allow genuine human capital to be different from signals. That is what this paper plans to do. I model the production of signals and pre-college human capital. The two production functions can be separately identified with the information of post-college outcomes.

The paper also contributes to the literature on private tutoring. One strand of the studies have investigated the effect of tutoring on academic outcome and obtained mixed results (Dang, 2007, for Vietnam; Gurun and Millimet, 2008, for Turkey; Ono, 2007, for Japan; Ryu and Kang, 2013, for Korea; Zhang, 2013, for China). Another strand of literature examines the policy impacts on tutoring expenditure in Korea, and find little to no effect (Choi and Choi, 2016; Kim and Lee, 2010). One of the limitations of these studies is as follows: even if we understand what policies can reduce the overall tutoring spending, we are still not sure whether they improve welfare. A structural model can help quantitatively evaluate a wide range of policy candidates, including taxing private tutoring, expansion of selective universities, and adjusting admission policies. These experiments should be not only interesting for Korea but also a broad set of countries with high-stake exams and active private tutoring market.

This paper is organized as follows. Section 2 lays out the model and estimation strategy. Section 3 describes the institutional background and data. Section 4 presents the estimation results of the baseline model, including parameter estimates and model fit. A few counterfactual policy experiments are displayed in Section 5. Section 6 concludes.

2 Background and Data

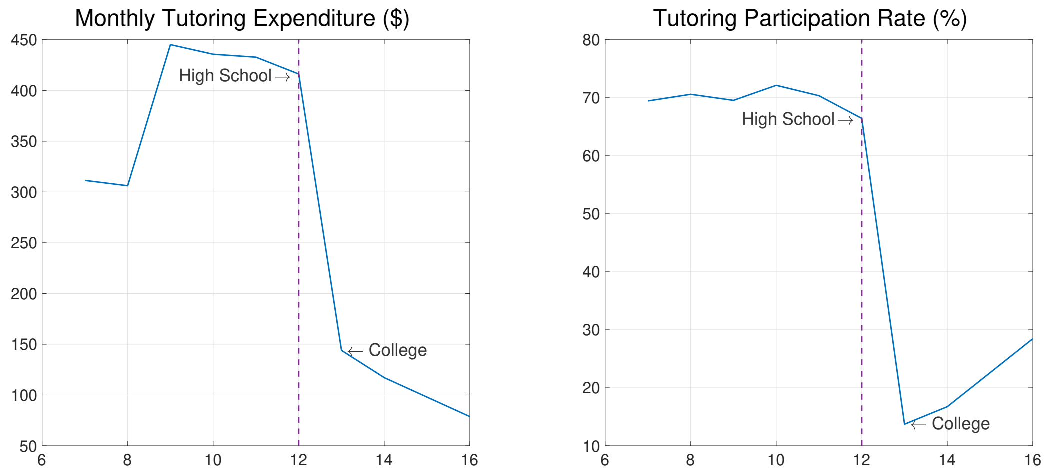
2.1 Institutional Background

The academic rat race among Korean high-school students for college admissions is an annual competition for seats at a diverse set of tertiary institutions from most prestigious universities to two-year colleges. College rankings are fairly well-agreed upon and stable over time. Graduating from prestigious universities brings substantial economic and non-economic premiums. For example, the top three Korean colleges, accommodating 1% of college graduates, account for 74% of the CEOs (Lee, 2007), 63.7% of senior officials and 58.1% of congressmen in South Korea (Chae, Hong, and Lee 2005).

The College Scholastic Ability Test (CSAT) scores play a key role in college admission. All high school students who intend to attend colleges must pass the annual national CSAT. Near-perfect CSAT scores are required at top three colleges. As of 2010, this national test consisted of 5 sections: Korean language arts, mathematics, English, social studies/science, and the second foreign languages. Students are informed of the scores and percentile rankings of all subjects before application. Admission quotas are pre-specified and determined by Korean Ministry of Education. Colleges have an explicit formula, including weight, to calculate the final score in admission. There are two rounds of admission each year: early decision and regular admission. The early decision is based on a combination of high school records, CSAT scores, extracurricular activities, recommendation letters and interview, while regular admission relies exclusively on the ranking of CSAT scores.

Throughout high school, students exert time and money to prepare for standardized tests. The use of tutoring is prevalent and primarily for academic purposes. In secondary education, 90% of tutoring expenditure is for academic purpose, among which 92% is spent on the commonly administered subjects in CSAT: Korean language arts, mathematics and English (2010 Survey of Private Education Expenditure). Since poor parents are

Figure 1: Motivating Fact



not able to foot the bills for private tutoring, the heavy reliance on private tutoring in Korea creates an inequitable distribution of education resources.

Tutoring expenditure drops dramatically after college attendance. Figure 1 shows the tutoring participation rate and the average monthly expenditure among the participants. Around 70% of students in secondary school take tutoring. The participation rate drops 50% after attending college. The average monthly tutoring expenditure falls to one-third. The dramatic decline cannot be easily rationalized by the human capital motive alone, suggesting that the incentives to compete for good colleges can play an important role in tutoring decision.

2.2 Data

The Korean Educational Longitudinal Study of 2005 is used in analysis. The KELS 2005 is a longitudinal survey that began in 2005 with a nationally representative sample of 6,908 Korean seventh graders (first year middle school). The survey follows the cohort annually before 2012 and biennially afterwards. The data includes information on the students' academic performance as measured by GPA and standardized test scores, tutoring expenditure, high school characteristics, family background, college attendance, and students' perceived labor market earnings. Information on initial academic performance and family background allows me to model the incentives facing households when they make tutoring decision. Pre-college test scores and post-college outcomes help disentangle the two effects of tutoring: producing genuine human capital and generating signals in the admission tournament.

The national College Scholastic Ability Test (CSAT) is is a standardized test that examines individuals' abilities for entering a university. It is held once per year and is made up of five sections. Korean language arts, mathematics and English are mandatory subjects that account for more than 60% of total points in CSAT. Science and Second Foreign Language sections are elective. Students choose one subject of each elective section depending on the majors they plan to apply for. I focus on the three mandatory subjects in measuring the initial academic performance, CSAT score and tutoring expenditure because CSAT scores of mandatory subjects are comparable across individuals. Besides, test scores of elective subjects are highly correlated with that of mandatory subjects with correlation coefficient 0.8.

In my analysis, colleges are grouped into three tiers with Tier 1 representing the top 15% of college seats. KELS 2005 dataset does not contain direct measure of college quality. I use the lowest CSAT score of students admitted through regular admission as a proxy for college quality. As the regular admission process is solely based on CSAT scores, the lowest score is a meaningful reflection of admission cutoff and hence college quality. I further assume that signals are commonly observed and evaluated in the same way by all colleges. The lack of college names limits one's ability to link a university identifier to its actual admission policy.

I focus on students who attend colleges right after graduating from high school, which represent 76% of the full sample. This number is higher than but still comparable to the national level of college enrollment rate 70% in the same year (Korea Educational Development Institute, 2017). High school graduates entering into labor market (6.64% of full sample) or retaking the national exam for college admissions (12.97% of full sample) are excluded from the analysis. I further restrict my analysis to a sample of 1300 students who have complete information on standardized test scores (grade 7, 12), tutoring expenditure (grade 7-12), household income (grade 7-12), high school and college characteristics.

Table 1 provides descriptive statistics for the estimation sample and the full sample. Initial test score, CSAT and GPA are normalized to have mean zero and unit standard deviation among the full sample. The mean test scores in the estimation sample is positive and has a standard deviation below one. This primarily reflects that the low-performing students are less likely to report their CSAT scores and are more likely to work right after high school graduation. Students in restricted sample attend better high schools and take slightly more tutoring, largely because I require students to attend colleges right after high school. Household income is measured by average monthly income after excluding the education spending on siblings. The distributions of household income and expected wage in estimation sample resemble that in full sample. Although the estimation sample includes more high-performing students than would a nationally representative sample, it is those students who actively take tutoring to compete for elite colleges, and are more responsive to policy changes regarding tutoring market and college admission. The estimation sample still covers a broad range of key players in the admission tournament.

Table 1: Descriptive Statistics

	Estimation		Full Sample		
	Mean	Std. Dev.	Mean	Std. Dev.	Obs.
	Panel A. Households Characteristics				
Male (%)	35.85		52.37		6908
Initial Test Score (Grade 9)	0.4152	0.9188	-7.0e-8	1	6622
CSAT Score (Grade 12)	0.1727	0.9286	-1.4e-7	1	3857
High School GPA	0.2052	0.8801	-1.1e-7	1	4827
HH Income ($100/Month)	41.667	38.899	40.639	40.231	5100
	Panel B. School Characteristics				
High School Quality	0.6169	0.1855	0.5326	0.2280	5354
Tier 1 College (%)	15.00		12.55		3514
Tier 2 College (%)	56.23		52.65		
Two-Year College (%)	28.77		34.80		3514
Expected Wage ($100/Month)	17.428	5.6258	17.244	5.8078	2923
	Panel C. Choice Variables				
Tutor Expense ($100/Month)	2.8950	2.7458	2.7087	3.1289	5100

Unit: $ in 2010.

In survey, students are asked wage expectation in the year they graduate from college. There are six options of expected pre-tax annual income categories to choose from. Students are also asked actual monthly pre-tax income after they get employed. But because the most recent survey data made available is collected in 2014, which is the senior year for four-year college students, the actual earnings are not observed for the majority of the sample. To investigate how accurate the wage perceptions are, I compare the wage expectation with the actual wage, focusing on a sample of two-year college graduates. Table 2 presents the distribution of expected monthly wage. The median expected wage falls into category $(\$1105, \$1473]$. This is consistent with the median actual wage $1238, and mean actual wage \$1245 of the same sample. While these data are based on small sample sizes and only two-year college group, they are still informative and suggesting that expectations data are predictive of actual realizations.

Table 2. Expected Monthly Wage of 2Yr College Graduates

Exp. Wage (\$)	≤ 737	$(737, 1105]$	$(1105, 1473]$	$(1473, 1842]$	$(1842, 2210]$	> 2210
Percent (%)	2.53	30.38	39.24	22.15	4.43	1.27
Observations	4	48	62	35	7	2

Unit: \$ in 2010.

3 The Model

3.1 Environment

There is a continuum of households of unit mass. Each household has a child in high school, and is endowed with $X_i = (A_i, q_i, y_i, \nu_i)$. Ability A_i represents the child's stock of skills at the beginning of high school, and is perfectly measured by initial test score. q_i is the high school quality, y_i is household income. ν_i represents household i's preference for Tier 1 colleges. Individual preference ν_i is private information, while the population distribution is common knowledge.

Colleges are categorized into three tiers: high quality four-year (Tier 1), low quality four-year (Tier 2), and non-selective two-year colleges (Tier 3), with total mass one. Four-year colleges have higher quality than two-year colleges, which is measured by wage return. Anyone can attend two-year colleges, whereas the admission process for four-year colleges can be competitive because of capacity constraints. Colleges in the same tier are identical for a household. All households agree on the ranking of colleges. Colleges wish to admit the best students possible but genuine human capital is private information. They rank students based on commonly observed set of signals including test scores.

Once a student enters college, her belonging household makes no other decisions: the college is an absorbing state.

The model starts from 1st year high school. Household i with $X_i = (A_i, q_i, y_i, \nu_i)$ chooses how much to spend on private tutoring, while taking into account how much they value colleges, and how tutoring decisions will affect their admission chances. At the end of high school, human capital is produced, signals crucial for college admission are generated. Students are assigned to colleges based on the rank-order of signals.

3.2 Admission

The admission policy is a combination of CSAT score, high school GPA and quality, other factors that are unobserved to econometrician. The admission criteria is pre-specified as

$$s_i = \delta_1 CSAT_i + \delta_2 GPA_i + \delta_3 q_i + \delta_4 A_i + \delta_5 A_i^2 + \delta_6 y_i + \delta_7 y_i^2 + \xi_i. \tag{1}$$

Here $CSAT_i$ represents scores in College Scholastic Ability Test, a national test held once per year. GPA_i is high school GPA, q_i is high school quality measured by the school's advancement rate into college. $\delta_4 A_i + \delta_5 A_i^2$ captures the student ability that is not reflected in end-of-high school test scores but is correlated with initial performance (A_i). This ability component can be observed by colleges through recommendation letter and during interview in the early admission. $\delta_6 y_i + \delta_7 y_i^2$ captures unmeasured admission signals such as extracurricular activities which matter for early admission decision. Households from wealthy background can afford and often spend significant amount of money building application packages including extracurriculars. ξ is a random matching shock commonly observable to colleges during admission process, but not to the student while making tutoring decision. ξ is assumed to be normally distributed for tractability, and is normalized as $\xi \sim N(0, 1)$ as s is scale-free.

3.3 Signals Generation

Test scores are generated from

$$\begin{aligned} CSAT_i &= \gamma_{10} + \gamma_{11} e_i + \gamma_{12} e_i^2 + \gamma_{13} A_i + \gamma_{14} A_i^2 + \gamma_{15} q_i + \epsilon_{1i} \\ GPA_i &= \gamma_{20} + \gamma_{21} e_i + \gamma_{22} e_i^2 + \gamma_{23} A_i + \gamma_{24} A_i^2 + \gamma_{25} q_i + \epsilon_{2i} \end{aligned}, \tag{2}$$

where A_i is the initial ability measured by test score at the beginning of high school, e_i is the monthly tutoring expenditure during high school. Parameters $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$ describe the signaling channel, in which tutoring improves test scores and thus admission chances. γ_{25} captures the "small-pond-big-fish" effect because GPA is not comparable across high schools. ϵ_1, ϵ_2 are independent shocks in scores generating process and follow normal distribution.

3.4 Preference

Households value current consumption, wages in labor market, non-pecuniary benefits from attending Tier 1 colleges. For tractability, the three components are assumed to be additively separable:

$$u_i = \ln(y_i - e_i) + \nu_i E_{\epsilon_1, \epsilon_2, \xi} I_{1i}(e_i) + \beta E_{\epsilon_1, \epsilon_2, \xi, \varepsilon_c, \varepsilon_w} \ln(w_i). \tag{3}$$

Here y_i is the household income available for household consumption and the student's education (excluding the education spending on siblings). Parameter β captures the importance of labor market payoff. ν_i represents one's discounted non-pecuniary utility value from attending Tier 1 colleges. The non-pecuniary benefits may include social status, alumni network etc. ν_i follows is assumed log-normally distributed, under which each individual strictly prefers Tier 1 colleges over other college tiers.

$I_{1i} \in \{0, 1\}$ indicates whether student i gets accepted into Tier 1 colleges. At the time of choosing tutoring, household i formulates probability $E_{\epsilon_1, \epsilon_2, \xi} I_{1i}(e_i)$ of attending Tier 1 colleges. It is the probability that the signal surpasses admission cutoff c_1:

$$
\begin{aligned}
E_{\epsilon_1,\epsilon_2,\xi} I_{1i}(e_i) = Pr \quad & \{\delta_1\left(\gamma_{10} + \gamma_{11}e_i + \gamma_{12}e_i^2 + \gamma_{13}A_i + \gamma_{14}A_i^2 + \gamma_{15}q_i + \epsilon_{1i}\right) + \delta_3 q_i + \delta_4 A_i \\
+\delta_5 A_i^2 + \delta_6 y_i + \delta_7 y_i^2 \quad & +\delta_2\left(\gamma_{20} + \gamma_{21}e_i + \gamma_{22}e_i^2 + \gamma_{23}A_i + \gamma_{24}A_i^2 + \gamma_{25}q_i + \epsilon_{2i}\right) + \xi_i \geq c_1\} \quad .
\end{aligned} \tag{4}
$$

3.5 Expected Wage

Logarithm of labor market entry wage is given by

$$
\ln(w_i) = \rho_1 e_i + \rho_2 e_i^2 + \rho_3 A_i + \rho_4 A_i^2 + \rho_5 q_i + \sum_{j=1}^{3} r_j I_{ji} + \varepsilon_{ci} + \varepsilon_{wi}, \tag{5}
$$

where r_j denotes the monetary payoff from attending colleges of Tier $j \in \{1,2,3\}$. ε_{ci} refers to the human capital shock realized during college, ε_{wi} is the wage shock realized in labor market. $\varepsilon_{ci}, \varepsilon_{wi}$ are assumed independent from all information one has prior to college entrance and with zero mean. Parameters ρ_1, ρ_2 capture the marginal productivity of tutoring expenditure in producing genuine human capital. After the realization of human capital shock in college ε_{ci}, a student forms expectation on her labor market outcome:

$$
E_{\varepsilon_w} \ln(w_i) = \rho_1 e_i + \rho_2 e_i^2 + \rho_3 A_i + \rho_4 A_i^2 + \rho_5 q_i + \sum_{j=1}^{2} r_j I_{ji} + \varepsilon_{ci}. \tag{6}
$$

Expected wage is surveyed at the end of college. This subjective expectation reflects a student's perceived monetary return to college. As the tutoring choice is jointly determined by perceived monetary return and non-pecuniary preference, observing expectations allows making more accurate inference on individual preference. For the purpose of estimating preference parameters, it would be ideal to have information on the expectations agents hold at the time of making tutoring choice. But due to data limitation, the end-of-college expectations are the closest approximations.

The validity of using end-of-college expectations hinges on two assumptions. Firstly, students report their expectation truthfully. This assumption is implicitly made when using any survey data and is not specific to expectations data. Second, students do not systematically change their beliefs on college premium (r_j) during college. Since the idiosyncratic match quality realized during college is embedded in the error term ε_{ci}, the changing expected college premium is mainly driven by college-specific information shocks and dropout decisions. Given the great emphasis on education and low dropout rate in Korea, it is not a strong assumption that households are well informed of the college premium when making tutoring decision.

The expected wage is also used to identify the human capital and wage formation. To get unbiased estimates, the expectations are required to be predictive of actual realizations. The wage comparison made in Table 2 suggests that the wage expectations are informative of the actual realizations.

3.6 Equilibrium

Households compete for the fixed and pre-determined amount of slots in selective colleges. The colleges capacity, production technology, admission criteria, and the joint distribution of household endowments are common knowledge prior to households' choices of tutoring. In a large contest with a continuum of households and under rational expectation, households can anticipate the correct admission cutoffs without uncertainty. Given admission cutoffs and with borrowing constraint, each household chooses the optimal tutoring expenditure:

$$\max_{e_i \geq 0} \ln\left(y_i - e_i\right) + \nu_i E_{\epsilon_1,\epsilon_2,\xi} I_{1i}\left(e_i\right) + \beta E_{\epsilon_1,\epsilon_2,\xi,\varepsilon_c,\varepsilon_w} \ln\left(w_i\right). \tag{7}$$

Consistent with the lack of financial loans designed for pre-college education, there is no borrowing possible to finance the tutoring cost. And consistent with the generous provision of financial aids in college, households are assumed to have access to perfect financial market during and after college. Therefore, the utility from workforce monetary payoff can be expressed as a function of expected present value of the lifetime earnings, that is, $\beta E_{\epsilon_1,\epsilon_2,\xi,\varepsilon_c,\varepsilon_w} \sum_{t=1}^{T} \frac{\ln(w_{it})}{(1+r)^t}$. Since the wage measures is observed only once and in early career, assumptions have to be made on how wages evolve over the life cycle and across college types. The current version of wage equation (5) is time-invariant, which at least, implicitly assumes that the growth rate of wage is the same for all colleges.

The first-order condition gives:

$$\beta(\rho_1 + 2\rho_2 e_i) + (\beta r_1 + \nu_i) \frac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{1i}(e_i)}{\partial e_i} + \beta r_2 \frac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{2i}(e_i)}{\partial e_i} \leq \frac{1}{y_i - e_i}. \tag{8}$$

At the margin, households are trading off the tutoring cost with the future benefits of improving admission chances and obtaining human capital. The admission cutoff c_j of Tier $j \in \{1,2\}$ colleges is determined by market clearing condition:

$$\int P\left(c_{j+1} \leq s \leq c_j\right) d\mathcal{F}(A, q, y, \nu) = \kappa_j, \tag{9}$$

where κ_j is the capacity of Tier j colleges. The number of admitted students is equal to the number of college seats, conditional on households' optimal tutoring choices.

3.7 Estimation Strategy

Under the assumptions that random terms $\{\epsilon_{1i}, \epsilon_{2i}, \xi_i\}$ are independent and normally distributed, I estimate the signal generation equations (2) by ordinary least squares, and the admission equation (1) with ordered probit. The expected wage equation (6) is estimated by maximum likelihood because expected wages are categorical in data. Let π_{ik} be the probability of expected wage being in category $(\varpi_{k-1}, \varpi_k]$, then

$$\begin{aligned}\pi_{ik} &= \Pr\left(\ln\left(\varpi_{k-1}\right) \leq E_{\varepsilon_w} \ln\left(w_i\right) \leq \ln\left(\varpi_k\right)\right) \\ &= \Phi\left(\ln\left(\varpi_k\right) - \rho_1 e_i - \rho_2 e_i^2 - \rho_3 A_i - \rho_4 A_i^2 - \rho_5 q_i - \textstyle\sum_{j=1}^2 r_j I_{ji}\right) \\ &\quad -\Phi\left(\ln\left(\varpi_{k-1}\right) - \rho_1 e_i - \rho_2 e_i^2 - \rho_3 A_i - \rho_4 A_i^2 - \rho_5 q_i - \textstyle\sum_{j=1}^2 r_j I_{ji}\right)\end{aligned}$$

Here $\Phi(\cdot)$ represents the cumulative normal distribution function with standard deviation σ_c. The log likelihood of observed wage expectations for a dataset with N observations and 6 wage categories can be written as

$$\mathcal{L}_1\left(\overrightarrow{\rho}, \overrightarrow{r}, \sigma_c\right) = \sum_{i=1}^{N} \ln\left(\Pr\left(E_{\varepsilon_w} \ln\left(w_i\right) \mid e_i, A_i, q_i, \vec{I}_i\right)\right) = \sum_{i=1}^{N}\sum_{k=1}^{6} \mathbb{1}\left(k\left(i\right) = k\right) \cdot \ln\left(\pi_{ik}\right).$$

The parameters of the utility function (3) are also estimated by maximum likelihood. Let $\lambda_i(e_i)$ be the likelihood of the household i choosing the observed tutoring investment. Conditional on initial endowment $\{A_i, q_i, y_i\}$, tutoring investment e_i is determined by non-pecuniary preference ν_i.

$$\lambda_i\left(e_i\right) = \Pr\left(e_i \mid A_i, q_i, y_i\right) = \Pr\left(\nu_i\left(e\right) \mid A_i, q_i, y_i, e_i\right).$$

The mapping $e \mapsto \nu_i(e)$ can be derived from the first order condition (8):

$$\nu_i\left(e\right) \leq \frac{\dfrac{1}{y_i - e_i} - \beta\left(\rho_1 + 2\rho_2 e_i\right) - \beta r_2 \dfrac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{2i}\left(e_i\right)}{\partial e_i}}{\dfrac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{1i}\left(e_i\right)}{\partial e_i}} - \beta r_1,$$

where "$\leq$" holds at corner solution $e = 0$. Therefore, $\lambda_i(e_i)$ can be further written as

$$\lambda_i(e_i) = \begin{cases} \Psi\left(\dfrac{\dfrac{1}{y_i - e_i} - \beta(\rho_1 + 2\rho_2 e_i) - \beta r_2 \dfrac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{2i}(e_i)}{\partial e_i}}{\dfrac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{1i}(e_i)}{\partial e_i}} - \beta r_1\right), & \text{if } e_i = 0 \\ \varphi\left(\dfrac{\dfrac{1}{y_i - e_i} - \beta(\rho_1 + 2\rho_2 e_i) - \beta r_2 \dfrac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{2i}(e_i)}{\partial e_i}}{\dfrac{\partial E_{\epsilon_1,\epsilon_2,\xi} I_{1i}(e_i)}{\partial e_i}} - \beta r_1\right), & \text{if } e_i > 0 \end{cases}$$

The preference parameters can be estimated from the log likelihood of observed tutoring investment:

$$\mathcal{L}_2(\beta, \mu, \sigma_\nu) = \sum_{i=1}^{N} \ln(\lambda_i(e_i)).$$

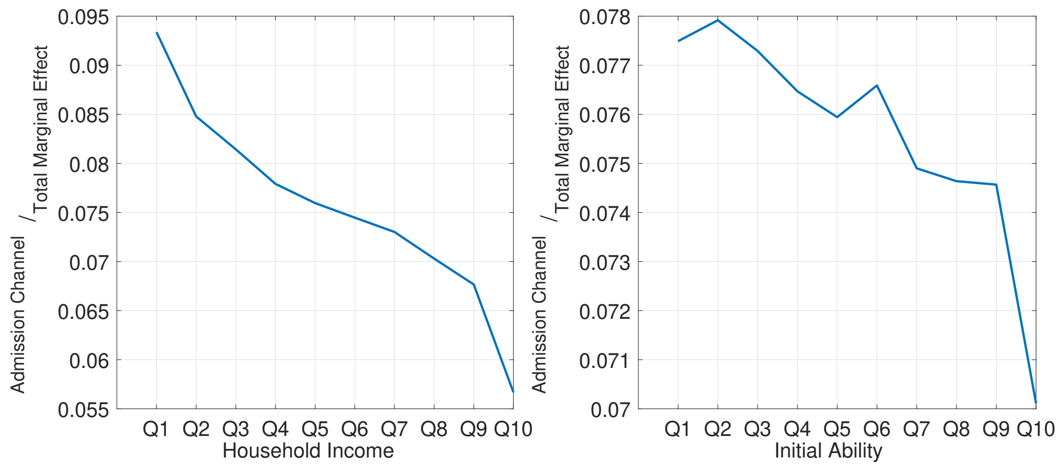
4 Estimation Results

4.1 Signals and Wage

Table 3 reports the estimates for the signals and wage equations. Tutoring expenditure exhibits diminishing marginal return. A \$100 change of monthly tutoring expenditure e from median level leads to 0.08 standard deviation change of CSAT score, and 0.02 standard deviation change in GPA. The marginal effect of tutoring on CSAT is 4 times as its effect on GPA. Tutoring plays a more important role on CSAT scores.

Tutoring produces genuine human capital. Conditional on ex-post college assignment, spending an extra \$100 on tutoring from median values every month can raise wage by 2.09%, equivalent to \$34 monthly wage gain for a mean wage earner. Given the wage gain over the life cycle, the labor market return to tutoring is sizable. Note that the human capital return could be overestimated driven by the omitted individual college quality. Conditional on college tiers, the omitted college quality is likely to be positively correlated with tutoring expenditure. This omitted variable bias can be resolved by adding college

Figure 2: Admission and Human Capital Channels in Wage Determination



fixed effects. However, adding college tiers implies that the quality rankings are less likely to be agreed upon by all households, because households may have idiosyncratic preference for college characteristics such as location and amenity. This will complicate the rank-order tournament setup and create computational burden in solving equilibrium outcomes.

There are two channels that tutoring can affect wage: producing genuine human capital, improving admission probability. Figure 2 shows the relative magnitude of the two channels. The horizontal axis represents the quantile rank of household income and ability. The vertical axis describes the ratio of the marginal effect through improving admission probability to the total marginal effect. As the ratio is below 0.5 for all, tutoring impacts wage mainly through the production of human capital.

Ability takes a greater role in generating CSAT and GPA than in affecting wages. One standard deviation change from median initial test score can boost up CSAT by 0.55 standard deviation, but only raise wage by 2.76%. The estimates are consistent with a common view that previous academic performance helps one enter a better college, but post-college wage is mainly driven by the human capital in college, which is accrued primarily through college quality.

Table 3. Signals and Wage Parameters

	CSAT		GPA		ln(Wage)	
Variables	Coef.	St. Dev.	Coef.	St. Dev.	Coef.	St. Dev.
Tutor e	0.1119	0.0168	0.0356	0.0194	0.0250	1.89e-7
Tutor e^2	-0.0049	0.0015	-0.0027	0.0018	-0.0007	1.64e-8
Ability A	0.3848	0.0253	0.2358	0.0293	0.0092	2.22e-8
Ability A^2	0.1740	0.0221	0.1501	0.0255	0.0190	1.37e-7
HS Quality q	0.9055	0.1096	-0.8232	0.1268	0.0444	1.69e-7
Tier 1 I_1					0.2071	4.06e–7
Tier 2 I_2					0.0925	4.12e-7
Constant	-0.9685	0.0736	0.4028	0.0852	2.6119	1.37e-7
Std. Error	0.7060		0.8167		0.3013	1.15e-8

4.2 Admission

Estimates for admission criteria are displayed in Table 4. The marginal impact of one standard deviation change in CSAT is larger than the marginal effect of one standard deviation change in GPA. Conditional on CSAT score, household income and initial performance explain a substantial proportion of variation in admission outcomes. This is consistent with the fact that about 50% of students in sample enter colleges through early admission, where other criteria, such as essays and letters of recommendation, extracurricular activities, aptitude examinations or interviews.

Incorporating initial performance to admission equation weakens the marginal impacts of CSAT and GPA, and thus, the estimated effect of tutoring expenditure on admission chances. But as many tutoring centers help students prepare for application packages and college interviews in early admission, the model may underestimate the impact of tutoring on admission probability. The agents may be more responsive to college competition incentives than the model would suggest.

Table 4. Admission Preference Parameters

Parameter	Description	Value	St. Dev.
δ_1	admission weight on $CSAT$	0.4483	0.0597
δ_2	admission weight on GPA	0.3404	0.0456
δ_3	weight on high school quality	0.4616	0.2068
δ_4	weight on initial performance g	0.0935	0.2038
δ_5	weight on initial performance g^2	0.2033	0.0393
δ_6	weight on household income y	0.0062	0.0022
δ_7	weight on household income y^2	-1.4e-5	6.6e-6
c_1	admission cutoff of Tier 1 colleges	2.2474	0.1683
c_2	admission cutoff of Tier 2 colleges	0.1422	0.1523

4.3 Preference

Table 5 describes the preference parameters. The estimates of the structural parameters indicate that while expected wage is a statistically significant determinant of the tutoring expenditure, they play a rather small role in the choice. Figure 3 compares the relative magnitudes of the marginal utility benefit of tutoring through wage $\frac{\partial E_{\epsilon_1,\epsilon_2,\xi,\varepsilon_c,\varepsilon_w}\beta\ln(w)}{\partial e}$ and the marginal benefit through college admission $\frac{\partial \nu_i E_{\epsilon_1,\epsilon_2,\xi} I_1(e)}{\partial e}$. The vertical axis displays the ratio of college competition incentive to the total marginal benefit of tutoring expenditure. The high ratio implies that competition for Tier 1 colleges is the driving force for tutoring investment. The competition incentive is stronger for high ability students. As a counterfactual exercise, I shut down the competition channel, so that tutoring investment cannot impact college assignment. Households significantly lower their tutoring expenditure, with the majority of households not even purchasing tutoring. This exercise suggests over-production of human capital in competing for prestigious colleges.

Figure 3: Wage and Admission Incentives

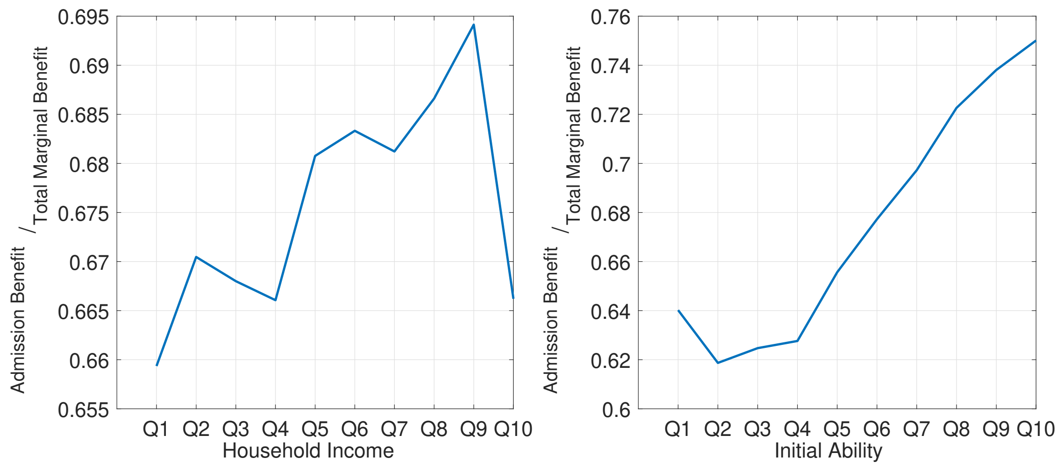


Table 5. Admission Preference Parameters

Parameter	Description	Value	St. Dev.
β	preference for log(wage)	0.0882	1.9e-5
μ	preference for Tier 1, $\ln(\nu) \sim N\left(\mu, \sigma_\nu^2\right)$	0.8073	2.2051
σ_ν	preference for Tier 1, $\ln(\nu) \sim N\left(\mu, \sigma_\nu^2\right)$	3.5169	4.9611

Non-pecuniary preference for Tier 1 colleges play a major role in tutoring choice. This is consistent with the substantial non-economic premiums of graduating from an elite college. But admittedly, the non-pecuniary preference ν may also captures pecuniary benefits, such as wage growth, that are associated with Tier 1 colleges. Note that w_i measures one's wage expectation at the beginning of career. If graduates from Tier 1 colleges enjoy lower unemployment rate and higher wage growth, by construction, those benefits will be contained in preference term ν.

4.4 Model Fit

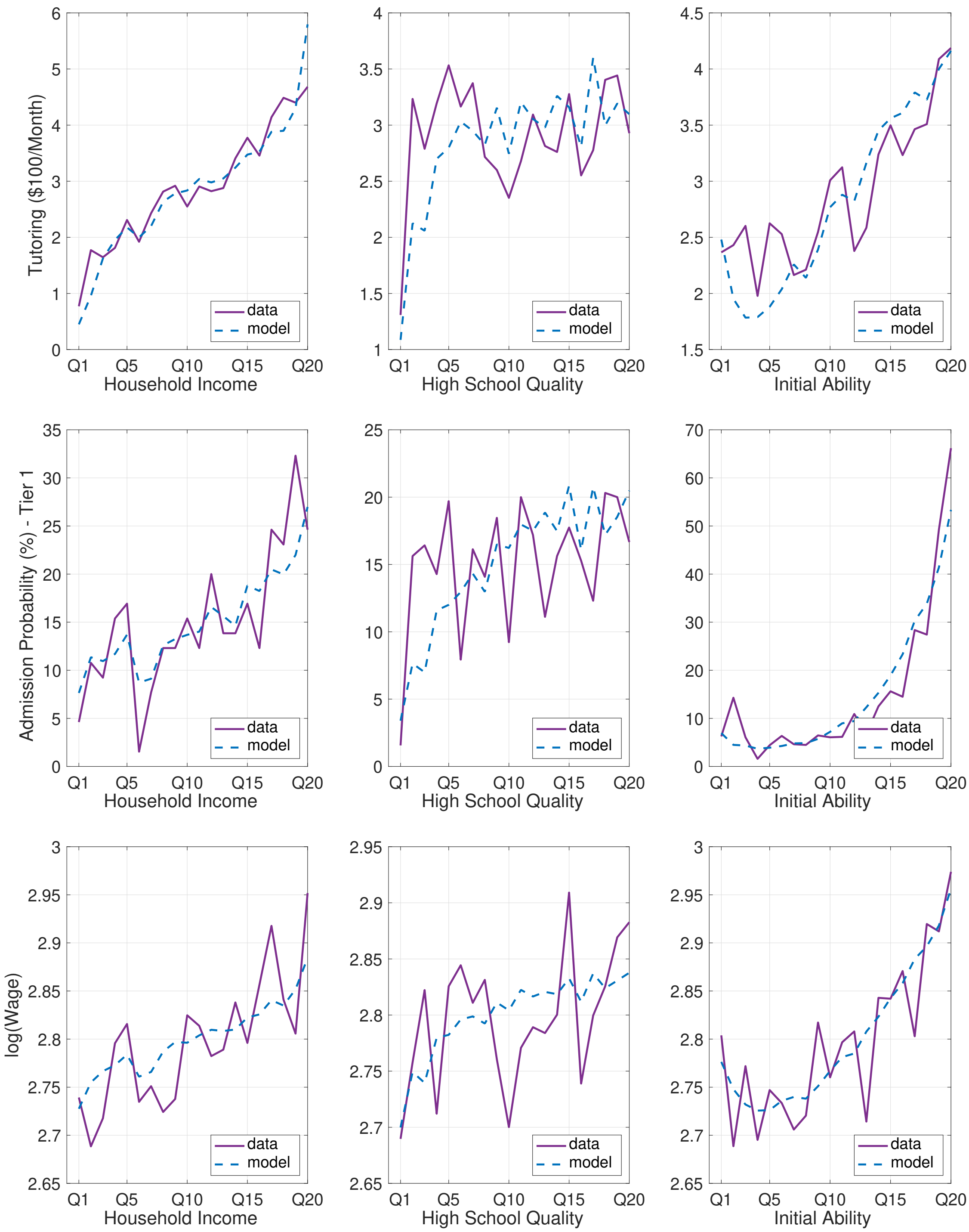
Figure 4 depicts the model predictions by household income, high school quality and initial academic performance. The tutoring expenditure of households from the top 5% income group is over-predicted. There are two possible explanations for this over-prediction.

First, the model does not allow borrowing and lending before college. In the model, the opportunity cost of purchasing tutoring is the lost consumption. However in reality, households are trading off current consumption, human capital return, with financial return. It is the wealthy households who hold more financial asset and receive higher financial return. The heterogeneous asset return may explain the declining marginal tutoring spending with income. Second, the model does not consider the time constraint. Faced with binding time constraint, although wealthy households can afford and are willing to purchase more tutoring, students may not have extra time to learn.

There is a potential tension between model fit and the usefulness of the model for counterfactual analysis. To help improve model fit, mostly the convex relationship between the admission rate and initial ability in the data, I include initial ability and its quadratic term in the admission criteria (1). It captures the idea that initial ability may be observed by colleges through recommendation letters and interviews in the early admission process. But adding initial ability into the admission criteria also mechanically reduces the importance of tutoring in admission.[1] To make model fit better, more weight is given to initial ability, a predetermined variable, which effectively makes admission probability insensitive to policy changes. In addition, the current model still underpredicts the Tier 1 admission probability for students with top-decile ability (the third panel of the second row in Figure 4). Alternatively, one could replace initial ability and its quadratic form with a quadratic term of CSAT to help fit the convexity without putting too much weight on predetermined variables. But that comes at a cost: the admission probability would lose closed form solutions, due to the error term in CSAT in the quadratic term, so that estimation is computationally demanding and is therefore not pursued in the current version.

[1]Another issue is that initial ability should be useful for only a fraction of students who have a chance of early admission.

Figure 4: Model Fit



5 Counterfactual Experiments

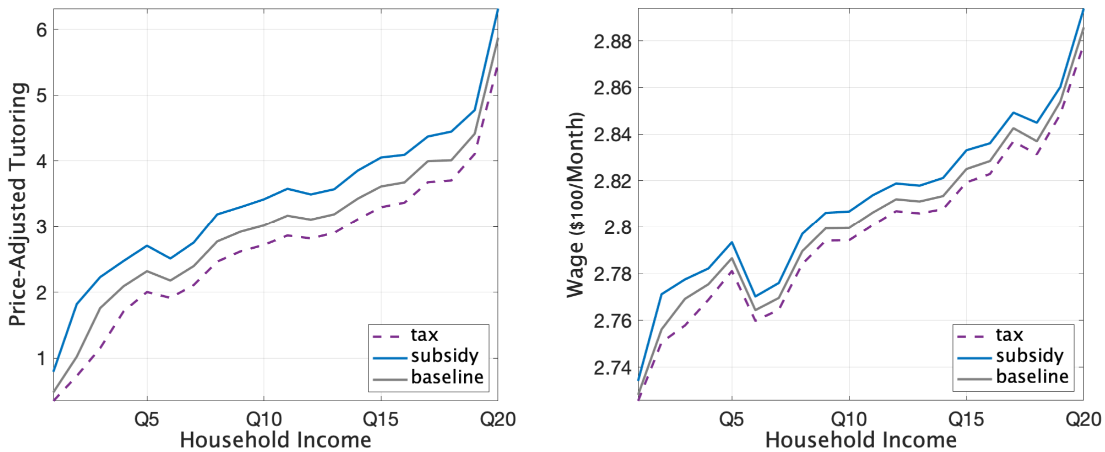
5.1 Tax and Subsidy

I now explore how tax and subsidy in tutoring market affect the tutoring incentives and equilibrium outcomes. It is theoretically ambiguous how households adjust their tutoring choices in response to price change. When price is high, human capital return to tutoring spending declines, but admission chances may improve because small increases in signal result in the student surpassing a larger fraction of competitors. Below I present counterfactuals where the price of tutoring increases (decreases) by 30% due to tax (subsidy).

Figures 5 presents the heterogeneous effects on tutoring expenditure and expected wage. The tutoring expenditure has been adjusted by price so that it measures the units of tutoring service purchased. One might think that because the top-income households purchase the most tutoring service, their expenditure should decrease the most with a proportional tax. However, it is the middle-income households' expenditure that respond the most in the experiment. Note that the middle-income households are more likely to be constrained, so that the tax is more "expensive" because tutoring service implies a greater loss in marginal utility. In addition, the unconstrained top-income households face less competition, so their marginal return of tutoring is higher.

The human capital accumulated gets lower, while admission probability as a function of income is almost unchanged, the latter perhaps due to the lack of responsiveness to competition incentives. Last, note that the distributional effects are across every ability level, so on average the admission probability as a function of ability is not changed. A subsidy would have the opposite effect.

Figure 5: Counterfactual Tax and Subsidy



5.2 Reduce Admission Noise

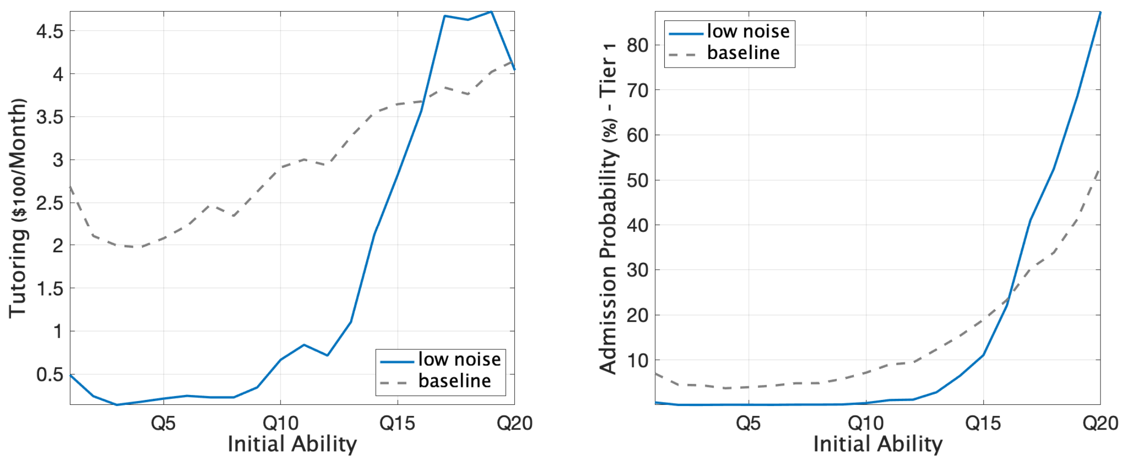
In this experiment, I evaluate the importance of signal noise by assuming away the random matching shock ($\xi = 0$). The admission is determined by

$$s_i = \delta_1 CSAT_i + \delta_2 GPA_i + \delta_3 q_i + \delta_4 g_i + \delta_5 g_i^2 + \delta_6 y_i + \delta_7 y_i^2. \tag{10}$$

The declining noise-to-signal ratio provides households more certainty when making tutoring decision. Now the distributional effect is mostly across ability. Reduced noise lowers the probability that a low-ranking student is perceived as a high-ranking student. Therefore, the high-ability students can more effectively to purchase tutoring service to defend their positions in the ranking order. On the other hand, the low-ability students have less incentives of doing so due to the more "rigid" ranking order. These effects are reflected on Figure 6.

Across the income distribution, households decrease spending while the middle- and bottom-income decrease the most, the latter of which is perhaps due to the discouraged competition incentives as high-ability students can more easily stand out. As a result, it is the students of high ability and from high income family benefit in admission.

Figure 6: Counterfactual - No Admission Noise



6 Conclusion

This paper develops a structural model to study pre-college educational investment (i.e., tutoring) in a college admission tournament. Methodologically, there are mainly two contributions. First, it allows educational investment to separately affect human capital accumulation and signals production. Second, it quantitatively studies the educational spending decision in a admission tournament. I find that tutoring produces genuine human capital, but also results in over-production driven by competition for prestigious colleges. The response of tutoring spending with respect to college admission is economically and quantitatively important.

As a result, the model provides policy implications. For example, conventional wisdom says a tax on educational investment should universally reduce investment and human capital by everyone. In this model, however, I find that the reduction of expenditure is most prominent among students from middle-income families. This is because the middle-income households are more likely to be constrained, so that the tax is more "expensive" because tutoring service implies a greater loss in marginal utility. Furthermore, I explore the impact of signal noise. Reduced noise incentivizes the tutoring expenditure of high-ability students, whereas discourages the low-ability students due to the more “rigid” ranking order. The admission chances of high-ability students get improved.

Future work can include further counterfactual analysis, for example, expansion of the selective universities or restriction on the quantity of tutoring service (e.g., limiting the hours of tutoring schools). Another possibility is to decompose the two channels of tutoring, especially their roles in explaining the counterfactural experiment results. Therefore, we can gain a better understanding of the value of incorporating the two channels in a unified study.

References

[1] Peter Arcidiacono, Patrick Bayer, and Aurel Hizmo. Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, 2(4):76–104, 2010.

[2] Hayri Alper Arslan. An empirical analysis of college admissions with endogenous entrance exam scores. 2018.

[3] Gary S Becker. Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49, 1962.

[4] Aaron L Bodoh-Creed and Brent R Hickman. College assignment as a large contest. *Journal of Economic Theory*, 175:88–126, 2018.

[5] Hai-Anh Dang. The determinants and impact of private tutoring classes in vietnam. *Economics of Education Review*, 26(6):683–698, 2007.

[6] Raquel Fernandez and Jordi Gali. To each according to...? markets, tournaments, and the matching problem with borrowing constraints. *The Review of Economic Studies*, 66(4):799–824, 1999.

[7] Nicolás Grau. The impact of college admissions policies on the academic effort of high school students. *Economics of Education Review*, 2018.

[8] Brent R Hickman. Pre-college human capital investment and affirmative action: a structural policy analysis of us college admissions. *University of Chicago. Unpublished*, 2013.

[9] Ed Hopkins. Job market signaling of relative position, or becker married to spence. *Journal of the European Economic Association*, 10(2):290–322, 2012.

[10] Edward P Lazear and Sherwin Rosen. Rank-order tournaments as optimum labor contracts. *Journal of political Economy*, 89(5):841–864, 1981.

[11] Sanghoon Lee. The timing of signaling: To study in high school or in college? *International Economic Review*, 48(3):785–807, 2007.

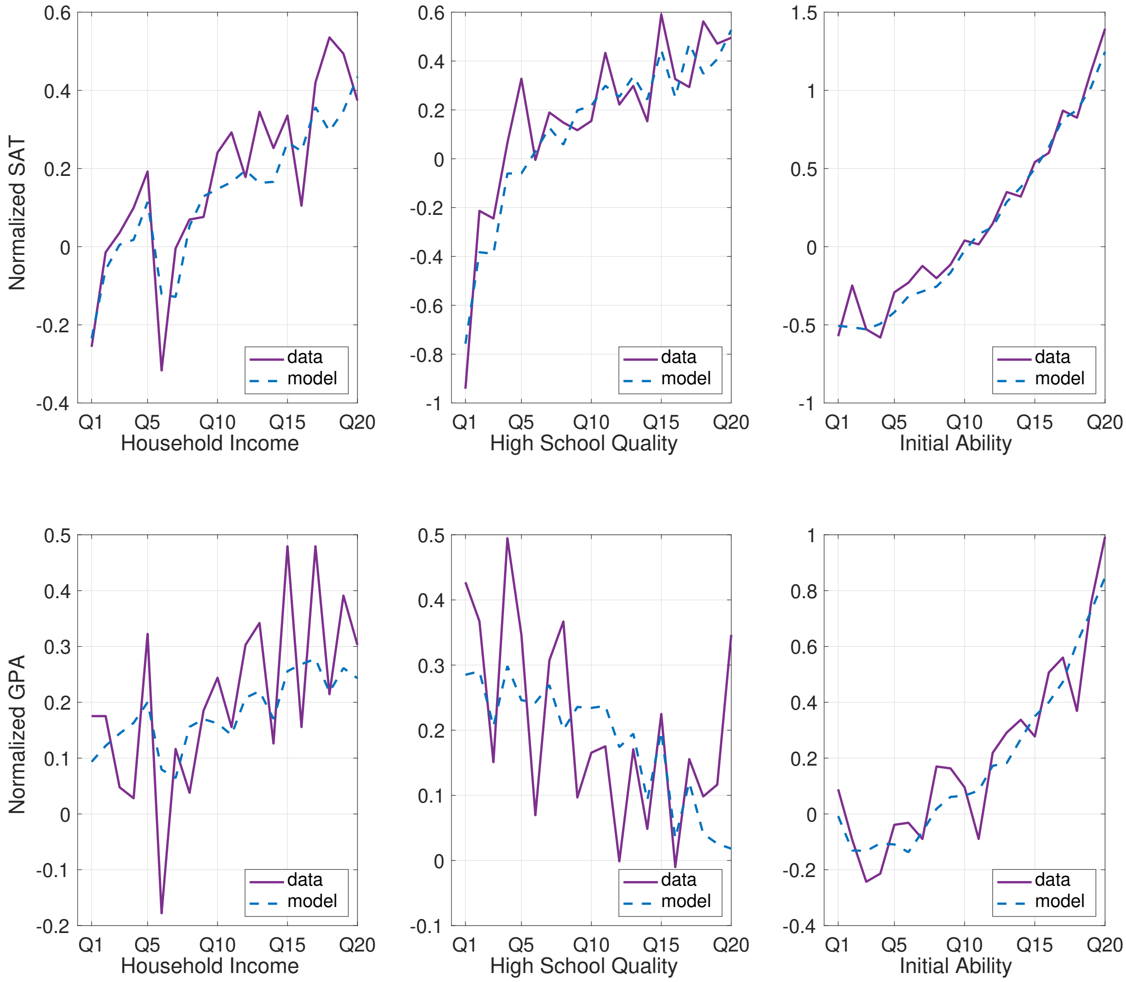
[12] Jacob Mincer. Schooling, experience, and earnings. human behavior & social institutions no. 2. 1974.

[13] Sunha Myong. Need based aid from selective universities and the achievement gap between rich and poor. 2016.

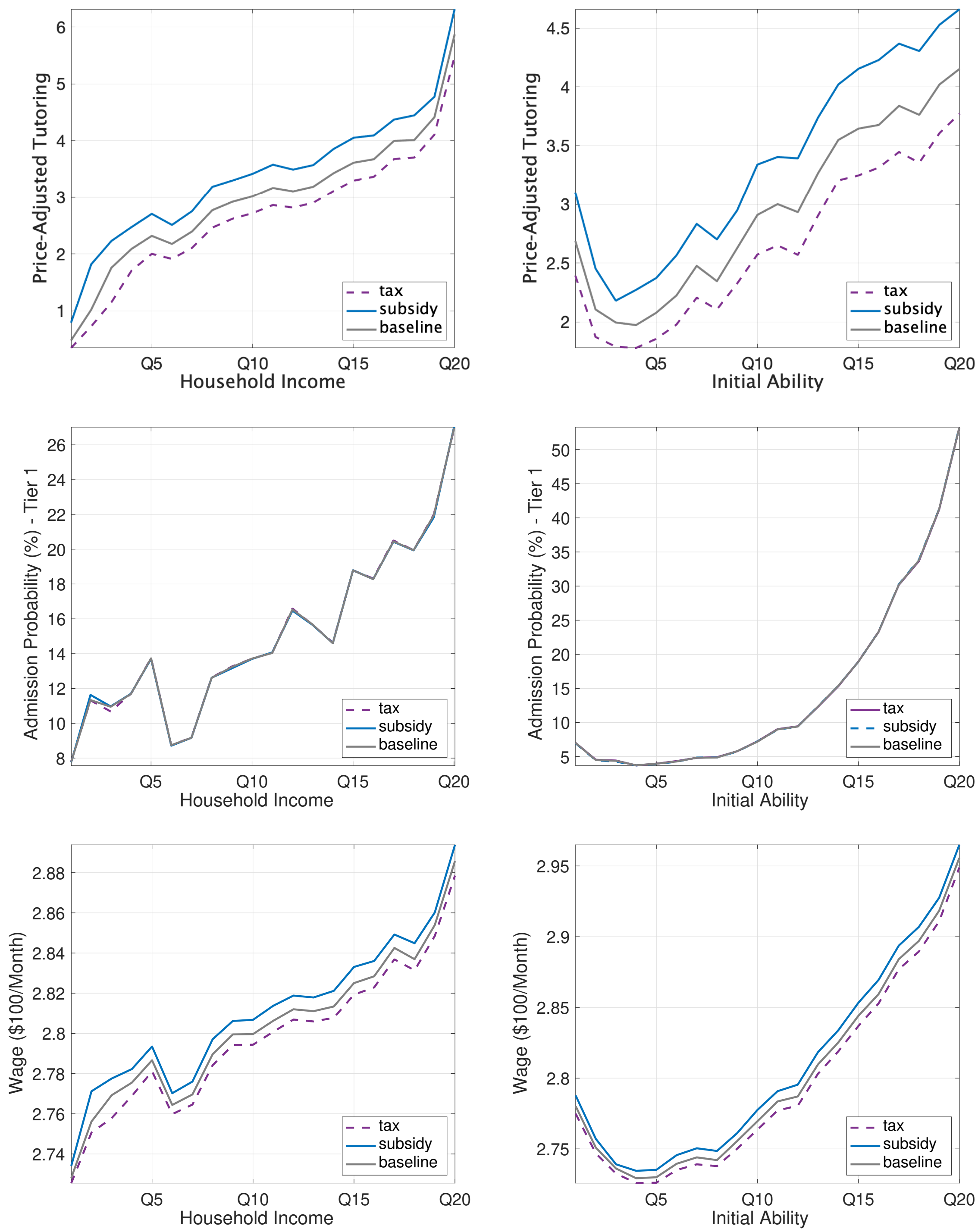
[14] Deockhyun Ryu and Changhui Kang. Do private tutoring expenditures raise academic performance? evidence from middle school students in south korea. *Asian Economic Journal*, 27(1):59–83, 2013.

[15] Michael Spence. Job market signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

Appendix 1. Model Fit



Appendix 2. Counterfactual: Tax and Subsidy



Appendix 3. Counterfactual: Admission Noise