On the Spatial Allocation of College Seats: Human Capital Production and the Distribution of Skilled Labor

Yang, Yu (Alan)

Peking University, Guanghua School of Management

1 March 2024

On the Spatial Allocation of College Seats: Human Capital Production and the Distribution of Skilled Labor

Yu (Alan) Yang*

March 1, 2024

Abstract

How to allocate college seats across regions is an important yet largely neglected issue. It may imply a policy tradeoff between efficiency in aggregate human capital production and equality of opportunities for people growing up in different places. Furthermore, the flow of college attendance, resulting from the geography of college seats, also impacts the spatial distribution of skilled workers through post-college migration and regional inequality in future development. This paper studies this tradeoff between efficiency and multidimensional inequality in the spatial allocation of college seats by focusing on the province-based college admission quotas in China, the largest college market in the world. Combining national administrative data and surveys, I estimate a structural model of college and migration choice under quota constraints, together with a measurement model that can recover the nationally comparable distribution of pre-college human capital in each province. There are substantial skill gaps between college applicants across provinces, but this disparity is not well reflected in the allocated admission quotas. A purely merit-based nationwide admission increases aggregate human capital at the cost of worse college opportunities and substantially more severe brain drain in less developed regions, while equalized admission leads to the opposite outcome. Comparing the current quota system against the efficiency-equality frontier suggests that China places a larger policy weight toward a more equalized spatial supply of skilled labor.

Keywords: Place-based college admission, human capital, spatial sorting, regional inequality

JEL Codes: J24, J61, R23, I24

*Peking University, Guanghua School of Management. Email: alanyang@gsm.pku.edu.cn. I am deeply grateful to Jeff Smith, Chris Taber, and Matt Wiswall for their guidance and encouragement. I thank Naoki Aizawa, Peter Arcidiacono, Chao Fu, Long Hong, John Kennan, Jingnan Liu, Corina Mommaerts, Laura Schechter, Xiaoyang Ye, Anson Zhou, and seminar/conference participants at UW-Madison, Peking GSM and CCER, Jinan IESR, CUHK Shenzhen, ShanghaiTech, Tsinghua, AASLE 2020 and 2021, SOLE 2021, and ES 2023 European Winter Meeting for helpful comments and suggestions. I am grateful to Jinghuan Shi, Fei Guo, and Lan Duo for their help during data application. All errors are mine.

1 Introduction

Where people grow up is one of the most salient factors that affect opportunities and success later in life (Chetty et al., 2014), and the difference in college access across areas is found to play an important role (Hillman, 2016; Ishimaru, 2023). Chetty, Friedman, Saez, Turner and Yagan (2020) show that changing how students are allocated to colleges could substantially reduce segregation and increase intergenerational mobility. Re-allocation of college seats has often been carried out through affirmative action policies based on individual or family characteristics (Arcidiacono and Lovenheim, 2016), while the spatial dimension is largely neglected. However, the geographical distribution of the colleges in a country is generally uneven and skewed toward economically more developed regions (Fu et al., 2022; Fabre, 2023). At the same time, college admission in many countries, at least to some extent, is based on residence. For example, public universities in the US charge much higher tuition to out-of-state students and, in some states, guarantee admission to local high-achieving students or impose out-of-state enrollment caps.[1] Similarly, France implements out-of-catchment area caps. And in China, colleges set admission quotas for each province.

This paper studies how best to allocate college seats spatially. This is an important but often difficult policy question for two reasons. First, when there commonly exists regional inequality in economic development and the quality of pre-college education, a stronger student-college sorting may achieve higher efficiency, that is, more human capital in aggregate, if higher ability students benefit more from attending higher quality colleges (Dillon and Smith, 2020), but pure meritocracy in college admission can worsen the opportunities faced by students in poor regions and decrease intergenerational mobility (Brotherhood et al., 2023). Second, the spatial allocation of college seats further connects the geographical distribution of skilled labor through dynamic migration. Many studies find that college location influences subsequent work location choices (Groen, 2004; Kennan, 2021; Huang et al., 2022). Admitting more students from poor areas to colleges located in developed regions can improve individual mobility but exaggerate the brain drain in the origin, as these students tend not to return. This enlarged disparity in human capital stock could further hinder local economic growth and deteriorate future regional inequality (Fleisher et al., 2010). Thus, the spatial allocation of college seats embeds a tradeoff between efficiency in human capital production and multidimensional regional inequality in college opportunities and the supply of

[1] For example, the University of California System has an 18% cap, and the University of Virginia has a one-third cap on out-of-state students. Public college systems in Texas and California offer guaranteed admission to in-state high school students if they are ranked above certain thresholds in their graduating classes.

skilled workers, which has not been systematically studied.

I examine this important tradeoff in the context of China. China has the largest college market in the world, with ten million students taking the National College Entrance Examination (NCEE, or *Gaokao* in Chinese) each year. Importantly, China's college admission is place-based and uses provincial quotas to allocate college seats. The current pre-college education is decentralized at the provincial level, and the education authority in each province administers their own version of the NCEE. Prior to NCEE, each university sets an admission quota (that is, how many students to admit) for every province with the approval of the central government. Each province then independently runs a centralized admission to these allocated college seats based on its NCEE scores. It is worth emphasizing that a quota is not simply a college seat, but an assignment of students from their home province to a college location, potentially in a different province. Therefore, it determines not only the local admission rate, but also the migration flow at the college stage and potentially further influences the post-college location choice.

College access is not equal under the quota system. The number of total four-year college seats received by each province ranged from 0.21 to 0.49 per college applicant during 2006-2011, and this large spatial variation has persisted to the present.[2] On average, economically developed provinces enjoy larger quota allocated to their students, but it is not clear whether and to what extent this reflects the difference in merit because the provincial NCEE is not nationally comparable. Furthermore, how does the spatial quota allocation shape both the college choice and the eventual distribution of millions of college students in China every year? And what are the policy objectives in operating this sophisticated place-based quota system for college admission? Quantifying the tradeoff between efficiency and multidimensional regional inequality could also provide valuable references to other countries who have a policy goal on addressing systematic inequality in higher education.

Answering these questions requires individual-level data that cover the entire country and track decisions from high school graduation to at least entering the labor market, which is not readily available in China. I construct a novel dataset by combining individual-level administrative data on all Chinese college students admitted during 2006-2011 with college student survey data for the same six cohorts, and further supplement it with administrative data on education quality of all undergraduate institutions. This dataset tracks student choices in college admission, during

[2] China's college capacity keeps expanding over time, but the variation in quota allocation is not reduced. In 2022, the total number of four-year college seats received by each province ranged from 0.32 to 0.73 per NCEE exam taker.

college studies, and after graduation, with detailed characteristics of students and colleges. It permits an analysis of admissions to all four-year colleges in all provinces, which has not been done in previous studies but is crucial for analyzing the distributional effect of college access and migration choices at the national level.

To overcome the non-comparability of individual college entrance exams in each province, I build and estimate a measurement model to construct one universal scale of pre-college human capital and recover the distribution for each province that is comparable nationwide. The measurement model is based on a dynamic latent factor framework (Cunha and Heckman, 2008; Cunha et al., 2010) and combines information from both provincial test scores and college GPA. The former provides information about relative rankings of students within a province, while the latter allows comparison of students originating from different provinces within the same college. With students in each province attending various colleges and each college admitting students from multiple provinces, this crossing structure effectively links the distribution in different provinces. This approach is in the spirit of Abowd, Kramarz and Margolis (1999) in identifying both worker and firm heterogeneity in wages using between-firm mobility of individual workers. The measurement model allows for general patterns of measurement errors and heterogeneity across provinces and colleges. I find a substantial variation in average pre-college human capital levels across provinces, which is positively related with local economic development and per student public expenditure before college. Moreover, the observed allocation of admission quota is not fully proportional to merit, meaning that equally achieving students from different provinces can face very different college choice sets. I then examine the important implications of this finding using a structural model.

With the recovered comparable measure of student quality, I estimate a dynamic structural model of college choice under admission quota constraints and post-college migration decision. The model has two distinctive features to facilitate a more comprehensive analysis of the spatial allocation of college seats. First, I allow for a flexible relationship between student and college quality in the technology of human capital production, which fundamentally determines the efficiency of student-college sorting. Second, the dynamic linkage of location choices in the college and work stages is explicitly modeled, which highlights the causal effect of college location, shaped by the quota allocation in the first place, on post-graduate migration. I show how admission cutoffs in colleges can be jointly used to identify preferences for colleges and locations, and to separate the causal effect of college location from a correlation in which students who prefer living in a region choose to go to college there and then ultimately settle down in the same place. The basic intuition is that each admission cutoff forms a fuzzy regression discontinuity locally, and there are many cutoffs available for identification. Having an NCEE score right above a cutoff "exogenously" expands the choice set of feasible schools by one additional college with particular quality and location, which causally changes the choice probability in college and dynamic locations. I find that student and college quality are complements; for one standard deviation increase in pre-college human capital, the marginal return from attending a better college is 38% higher in initial earnings, a level comparable to that estimated in the US (Dillon and Smith, 2020). Controlling for the unobserved location preference that persists over time, staying in the college location for work provides a value equivalent to 11% of annual consumption, and migrating to other places incurs additional losses that increase with distance.

I use the estimated model to evaluate various alternative allocations of college seats, ranging from merit-based nationwide admission without place-based quota to equally distributed quota across provinces. Compared to the current quota system, merit-based admission increases the student-college sorting and, as a result, increases total initial earnings by 1.3% and average student welfare by 3.5%. However, the disparity in the college admission rate between the richest and poorest quartiles of the provinces rises substantially from 9.1 to 14.8 percentage points (or 63%) in full meritocracy. Moreover, once the quota constraint on where students can go for college is removed, conditional on national ranking, more students from disadvantaged regions choose to attend college in developed provinces and tend to stay after graduation. The number of college workers per capita decreases by nearly 20% in the least developed region, exacerbating the brain drain. Compared to full meritocracy, admission that gives fully equal admission rates to all provinces generates the overall opposite outcome. Interestingly, it still achieves slightly higher aggregate efficiency than the current quota system if the only restriction is on the admission rate so that high-achieving students can still compete for the most favored college nationally before the equal cap of admission rate binds in their home province, without being constrained by the allocated quota in the current system. Together, these results point to a policy tradeoff between efficiency as measured by aggregate earnings, equality in opportunities for students born in different places, and equality in supply of skilled labor across regions that can further affect future development and inequality.

Given these estimated tradeoffs, the last part of the paper attempts to infer the policy weights placed by China in designing such sophisticated place-based quota system. I first compare the quota system against the potential frontier of efficiency and equality, which combines the various counterfactual allocations of college seats. The quota system is a clearly inferior policy when we only consider efficiency versus equality in college access. However, it generates the lowest inequality in the spatial supply of skilled labor, and thus the least brain drain, compared to other counterfactual allocation schemes that I have examined. This may be justified by the need to address the substantial regional inequality in development amidst China's spectacular overall economic growth (Kanbur and Zhang, 2005; Xu, 2011). I estimate that a one percentage point decrease in the provincial gap in the share of college-educated workers is approximately equivalent to at least 0.15% of the aggregate output in the government's objective function. In the long run, whether retaining highly skilled workers in underdeveloped areas for future local development is more favorable than alternative policies is left for future research. It is worth noting that the counterfactual analysis is conducted in a partial equilibrium framework so that the re-allocation of students in college admission does not affect college quality and the re-allocation of one cohort of college graduates in the short run does not affect the equilibrium wage level or productivity externality in the local labor market. A general equilibrium analysis that incorporates these considerations is also left for future research.

Related literature

This paper bridges two largely separate and parallel literature on higher education and human capital. The first strand of literature documents the effect of re-allocation of college seats, frequently through affirmative actions, on different groups of students and the debate over the associated quality-fit tradeoff. These policies are primarily based on race (Arcidiacono and Lovenheim, 2016; Bleemer, 2022) or socioeconomic status (Chetty et al., 2020; Mello, 2022; Brotherhood et al., 2023).[3] The second strand examines spatial inequality in the supply of skilled labor, with underlying mechanisms ranging from skill sorting in general (Diamond, 2016) to, more specifically, migration in response to the access, quality and cost of college education (Sjoquist and Winters, 2014; Knight and Schiff, 2019; Kennan, 2021; Anstreicher, 2024). This paper connects and contributes to these two research programs by showing the importance of combining both the way in which college seats are allocated and the dynamic migration decisions from birthplace to college and then labor market in analyzing the tradeoff between efficiency and multidimensional inequal-

[3] One exception is Barrow, Sartain and de la Torre (2020), who study a place-based affirmative action policy in Chicago's high school admission.

ity. I provide place-based evidence and show that consideration of inequality in college access may have implications not only for individual students, but also for regions in terms of the supply of college graduates. Due to geographical sorting and segregation of households (Diamond and Gaubert, 2022), the result of this paper is also valuable for affirmative action policies even if they only target individual or family characteristics.

In contemporaneous work, Fabre (2023) shows that the uneven spatial distribution of colleges, together with migration costs, largely contributes to the spatial inequality in educational attainment in France. Eliminating mobility friction will attract more students from low-opportunity areas migrating to higher education hubs and they tend not to return, thus causing a tradeoff between individual opportunity improvement and hometown brain drain. The key distinction compared to Fabre (2023) is that I emphasize more on where these colleges should distribute admission opportunities and explicitly consider spatial gaps in the quality of college applicants and their complementarity with college quality in human capital production. This permits the quantification of aggregate efficiency, in addition to the multiple aspects of inequality, among various policy alternatives. To this end, this paper is also related to Hendricks, Koreshkova and Leukhina (2022) and Diao, Liu and Zhong (2022), both of whom examine how the level of meritocracy in college admission affects the tradeoff between efficiency and earnings inequality, while they do not consider spatial distribution or migration.[4]

More generally, this paper is related to the broad literature on place-based policies and their implications for efficiency (Glaeser and Gottlieb, 2008; Busso et al., 2013; Kline and Moretti, 2014; Lu et al., 2019; Fajgelbaum and Gaubert, 2020). Recent work, such as Gaubert, Kline and Yagan (2021), has begun to explicitly examine the redistributive goal and efficiency-equity tradeoff in these place-based policies. I contribute to this new direction in the literature by studying the effect of a place-based education policy on the efficiency of human capital production and regional inequalities. Given the nature of college seats as a form of limited resources, this paper also adds to the literature on misallocation (Hsieh and Klenow, 2009; Hsieh et al., 2019), especially in developing countries during economic transition and reform (Tombe and Zhu, 2019; Adamopoulos

[4] There is a small group of earlier studies on China's college quota allocation. Guo, Loyalka and Ye (2018) overcome the non-comparability of the NCEE scores by giving a standardized test to a sample of college students in six provinces, and find that the current quota allocation is not fair by merit standard. Chan, Wang and Zhao (2019) analyze the behavior of quota setting in the Nash equilibrium of the top 100 colleges in China. They find that elite colleges systematically prefer students from richer provinces with better pre-college education. Pu (2020) analyzes college admissions in three Chinese provinces and shows that pooling provincial quotas increases average student welfare if the student quality distribution is the same across provinces.

et al., 2022), by looking at the college market and the productivity of human capital.

The paper proceeds as follows. The next section introduces the institutional background and provides basic data patterns. Sections 3 and 4 present the structural model and the measurement model. Section 5 describes data and estimation strategy. The estimation results are provided in Section 6. Section 7 examines counterfactual allocations of college seats, and Section 8 concludes.

2 Institutional background and basic data patterns

2.1 China's college admission and provincial quota system

There are 2,359 postsecondary education institutions in China as of 2011 (Ministry of Education, 2011). 1,131 of them are four-year colleges, most of which are public schools. The rest are three-year vocational and technical schools. Four-year colleges are classified into three tiers by the Ministry of Education based on school quality, and schools in higher tiers have priority in admitting students (Bo et al., 2019; Jia and Li, 2021). The top tier is generally considered as elite colleges.[5] Most of them belong to the Project-211, which was initiated in 1995 by the Ministry of Education with generous funding and policy support to raise the standards of education and research of high-level universities. A subset of them is further sponsored by the more prestigious Project-985 starting from 1998 as a national strategy to increase international recognition and competitiveness.[6] The second tier includes non-elite public colleges that are administrated by provincial governments, but are still widely recognized and are able to recruit students nationally. The third tier schools mostly admit students from their own province or surrounding provinces, and about 15% of them are private schools. College tuition in China is fixed by the government at an affordable level for most families and does not vary by student's residence location.

All four-year colleges and three-year schools admit students through the National College Entrance Examination (NCEE) based solely on their NCEE scores.[7] The NCEE in most provinces has two tracks based on the high school curriculum: Sciences and Humanities. Students in both tracks take the NCEE tests in Chinese, English, and Mathematics. In addition, students in Sciences

[5] Some colleges are treated as the first tier only in their own province during admission. Throughout the paper, I consider a college to be the first tier only if it is treated as such in all provinces. I classify colleges as second tier if they are treated as second tier or higher in all provinces. The rest of the universities are classified as the third tier.

[6] There are 116 colleges in Project-211 as of 2011. 39 of them are further covered by Project-985.

[7] Some top colleges are also allowed by the Ministry of Education to admit students through self-organized tests and interviews, but the scale is small. Less than 2% of college students were admitted through this channel during the period studied in this paper.

take tests in Physics, Chemistry, and Biology, while students in the Humanities take History, Geography, and Political Science.[8] Students specialized in arts or sports follow a different admission procedure and are not studied in this paper. Since 2001, the provincial education authority has been allowed to design its own high school curriculum. As a result, the NCEE tests for each track and subject are also designed and administered at the provincial level, and are not comparable across province, track, and year. Certain provinces do not have sufficient resources to develop their own NCEE tests for some or all subjects, in which case the Ministry of Education provides a set of general versions for those provinces to choose from. Nevertheless, even in these instances, the exam grading is still organized by each province autonomously.

After the NCEE is held in June each year, every province operates its own autonomous centralized college admission. Most provinces adopt a hybrid of the Boston and Deferred Acceptance mechanism with a restricted application list, typically allowing for 6-8 schools (Chen and Kesten, 2017; Bo et al., 2019). Students also provide a preference list of 3-5 college majors within each school. The admission algorithm operates at the college level. According to the listed ranking, the algorithm sends students to a school as long as there are still available seats in any major. After receiving them, the school assigns students to a major given the NCEE score and the listed preference for majors. If a student's proposed major cannot be fulfilled, the college assigns her to other available majors following some school-specific rules.[9]

China's college admission features a province-based admission quota system. In each year, the Ministry of Education coordinates with the National Development and Reform Commission to determine how many students each university can admit during the next NCEE. Each school submits its plan for the allocation of college seats between provinces and tracks (that is, admission quotas), and this plan must be approved by the provincial and then the central government before the NCEE (Kang, 2005; Pan and Cui, 2017). Almost all colleges allocate quota to more than one province. The quotas received by a province are a set of available college seats (located in various provinces) that are only open to students in that province. It determines not only the number of students that can go to college, but also the distribution of student migration flows into each province for college studies.

[8] A few provinces do not separate tracks in the high school curriculum and their NCEE has only one track.

[9] Students can refuse the college major assignment outside of their listed majors when submitting the application. In this case, they will be returned by the school and will not be considered by other schools in the same admission round. Given this high risk, students are generally advised not to refuse the major adjustment.

2.2 Data patterns of quotas, college choice and migration

There is a large spatial disparity of college opportunities across provinces under China's quota system, defined as the total number of four-year college admission quotas received by each province divided by the number of NCEE exam takers in that province. The left panel of Figure A1 plots the overall allocation of admission quotas across provinces, averaged between 2006 and 2011. In provinces ranked among the top, more than 45% of NCEE exam takers have the opportunity to go to a four-year college. This number reduces substantially to less than 25% in the provinces at the bottom. The large gap in the college admission rate between provinces persists in recent years (Figure A1, right panel).

This spatial disparity in college access is positively correlated with economic development, as shown in the left panel of Figure A2. The right panel of Figure A2 shows that the geographical distribution of the colleges is also skewed towards more economically developed provinces, which is similar to other countries such as the US (Fu et al., 2022) and France (Fabre, 2023). In addition to central planning at the national level, the allocation of college quota can also be influenced by the local government. Almost all second- and third-tier colleges in China are governed and funded by the provincial government.[10] Even top-tier schools, which are governed by the central government, often receive funds and favorable policies (e.g., free land) from the local government of the place where they are located (Pan and Cui, 2017). On average, a large share of the capacity is kept for local students: between 2006 and 2011, first-tier colleges on average allocated 45% of their quotas to the local province, and the second and third-tiers allocated 75% and 81% to the local province, respectively. In this paper, I do not explicitly consider who makes the decision of the quota allocation, but rather take the allocation as given when modeling student choices. In counterfactual analysis, the quota system is changed exogenously from a social planner's perspective.

Importantly, the allocation of quota determines the migration flow of students from each origin at the college stage. The student's choice of the college location is further correlated with the choice of a subsequent work location after graduation. Table 1 shows that between 2006 and 2011, among the 69.1% of students who attended a college in their home province, 71.9% stayed after graduation and only 28.1% migrated to another province. Among the others who attended a college outside their home province, 32.4% stayed at the college location, while 28.2% returned home, and the rest migrated to a third province. This suggests that students not only have home attachment, but

[10] Private colleges in China also receive partial government subsidies and are subject to strict government regulation.

also seem to prefer the place where they live and study for four years during college. The structural model introduced in the next section will carefully model these features and try to understand how students make college and residential choices under admission quota constraints.

3 The structural model of college and residential choices

This section develops a dynamic discrete choice model of college choice and subsequent choice of residential location after graduating from college. The quality distribution of college applicants in each province is taken as given in the structural model. The next sections will discuss how to recover these distributions using a measurement model and how the structural model developed in this section is integrated with the measurement model in estimation. The structural model has three stages.

Stage 1: High school graduates apply to four-year colleges in the centralized admission in their residence province, subject to provincial admission quota constraints, and attend the one that admits them.

Stage 2: At college graduation, students choose to either go to graduate school or enter the labor market. If choosing graduate school, subsequent choices are not tracked, and the model ends immediately with a terminal value. If choosing to enter the labor market right after college, students move to stage 3.

Stage 3: College graduates choose a province in which to live and work.

3.1 Primitives

Denote the set of all provinces as $\mathcal{N}$. Each province has a cohort of high school graduates with a different size and pre-college human capital distribution, and a set of four-year colleges with different quality and capacity. Denote the set of all four-year colleges in the country as $\mathcal{S}$. To reduce the notational burden, individual subscripts are suppressed in all variables.

3.1.1 Initial condition

Consider one cohort of high school graduates in year t. Each individual enters the model at the end of high school with the following initial conditions:

$$\Omega = \{ \underbrace{h}_{\substack{\text{home}\\\text{province}}}, \underbrace{z}_{\substack{\text{high school}\\\text{track}}}, \underbrace{\theta^0}_{\substack{\text{pre-college}\\\text{human capital}}}, \underbrace{NCEE_{hzt}}_{\substack{\text{NCEE}\\\text{score}}}, \underbrace{\tau}_{\substack{\text{idiosyncratic}\\\text{location preference}}} \}.$$

Each individual takes the NCEE of track $z \in \{\text{sciences}, \text{humanities}\}$ in her home province $h \in \mathcal{N}$ and year t.[11] The test score $NCEE_{hzt}$ is a function of her pre-college human capital θ^0 plus a measurement error. Denote the distribution of θ^0 in each province as $F_h(\theta^0)$. The functional form of $F_h(\theta^0)$ and its relationship with $NCEE_{hzt}$ will be specified in detail in the measurement model in Section 4. Pre-college human capital θ^0 is known to individuals but not to the econometrician.

It is natural that people's location preference is persistent over time (Howard and Shao, 2023). Accounting for this type of unobserved idiosyncratic location preference is potentially important for isolating the causal effect of college location on subsequent residential choices. Denote it as τ, which is also known to individuals but not to the econometrician. Each province has a common value of amenity, but an individual's preference may deviate from these common values, and τ captures this idiosyncratic deviation. The τ is independent of θ^0 and follows a finite mixture distribution of discrete types: $\tau \in \{\emptyset\} \bigcup \mathcal{N}$.[12] To keep the structural model tractable, I impose that the first type, $\tau = \emptyset$, does not have a preference for any province, while each of the remaining types has a preference for the corresponding province in $\mathcal{N}$ but no preference for the other provinces. Students may have such a preference for location if they have family connections in a certain place. They may also establish an attachment to a location either because they have been there previously or because of other reasons, but they may not necessarily have any particular knowledge about elsewhere. Denote the type probability of $\tau = \emptyset$ as $\pi_\emptyset$. Each of the remaining types has an equal type probability $(1 - \pi_\emptyset)/N$, where N is the number of provinces in the set $\mathcal{N}$. The magnitude of this idiosyncratic preference is χ for the college location and $\rho\chi$ for the work location after college, where ρ is a scalar.

3.1.2 Flow payoff of college

The flow payoff of attending college $s \in \mathcal{S}$ located in province $j_s \in \mathcal{N}$ is:

[11] Track z_i was chosen two years before taking the NCEE, and is taken as exogenous in the model.

[12] An alternative approach to model unobserved location preference in the literature is to have an idiosyncratic scalar in front of the amenity level γ^c_{3,j_s} (Diamond, 2016; Huang et al., 2022). It imposes that the idiosyncratic location preference is proportional to the ranking of the common amenity values in each province, which is less flexible than the mixture distribution specification.

$$u_s^c = \underbrace{Z_s'\gamma_1^c}_{\substack{\text{characteristics}\\ \text{of school } s}} + \underbrace{D(j_s,h)'\gamma_2^c}_{\substack{\text{distance function}\\ \text{(college, home)}}} + \underbrace{\gamma_{3,j_s}^c}_{\substack{\text{amenity}\\ \text{level in } j_s}} + \underbrace{\chi\mathbb{1}(j_s{=}\tau) + \xi_s^c}_{\substack{\text{idiosyncratic}\\ \text{preference}}}, \tag{3.1}$$

where college characteristics Z_s includes a quality measure V_s, type of elite programs (Program-211 or 985), college tier, type (comprehensive, STEM-, humanities-, or economics-focused college) and an indicator of being located in the capital city of a province. For tractability, Z_s is fixed in the model, so I do not further account for changes in peer effects but will discuss the potential implications in counterfactual analysis.[13] The distance function takes the following form: $D(j_s,h)'\gamma_2^c = \gamma_{21}^c\mathbb{1}(j_s{=}h) + \gamma_{22}^c d(j_s,h) + \gamma_{23}^c d(j_s,h)^2$, where $d(\cdot,\cdot)$ is the distance (in 1,000 km) between two provinces. The first term is an indicator for attending college in the home province, so γ_{21}^c captures home preference. The quadratic distance terms capture both the migration cost (monetary and psychological) and the preference for adjacent places that share similar climate and culture. The term γ_{3,j_s}^c is a provincial fixed effect and represents the common value of amenities in the college location j_s. Lastly, the idiosyncratic preference has two parts. The first is the individual's persistent location preference for province $j_s = \tau$. The second part, ξ_s^c, is an idiosyncratic preference shock and follows the Type-I extreme value distribution.

I assume that all students prefer four-year college over any outside options, and the value of outside options is normalized to zero.[14] College major is also abstracted away. As discussed in Section 2, it is possible to be admitted by a school even when the quota of the proposed major is already filled, if students allow the school to assign them to other available majors.

3.1.3 *lifetime payoff of graduate school*

After graduating from college s, students can enter the labor market ($g = 0$) or go to graduate school ($g = 1$). If choosing to go to graduate school directly after college ($g = 1$), the lifetime

[13] Keeping track of peer composition in this model would require solving an additional equilibrium when students make college choices, within each round of model estimation, which is not computationally feasible. See Baird, Engberg and Opper (2023) for a formal discussion of the optimal allocation of seats in the presence of peer effects in the context of job training.

[14] Re-taking the NCEE and applying to college again in the next year is not allowed in the model because I cannot distinguish between first-time NCEE takers and re-takers in the administrative data. This rules out the possibility that the counterfactual change in college access affects NCEE re-taking. However, although the share of re-takers (including unknown number of multiple-time re-takers) among all NCEE participants is stable between 20-25% from 1999 (Si, 2021), on average only 4.2% of the *admitted* students (both four-year and vocational colleges) are re-takers (Bai et al., 2021). This suggests that most re-takers are at the bottom part of the distribution, not in the baseline model (which includes only four-year college students), and not likely to be affected by the counterfactual policies considered in this paper.

payoff takes the following parsimonious form:

$$u_s^g = \gamma_0^g + \underbrace{\gamma_1^g \theta^0}_{\substack{\text{pre-college} \\ \text{human capital}}} + \underbrace{\gamma_2^g V_s}_{\text{college quality}} + \gamma_3^g \theta^0 V_s + \underbrace{Z_s' \gamma_4^g}_{\substack{\text{college} \\ \text{characteristics}}} + \underbrace{\xi^g}_{\substack{\text{idiosyncratic} \\ \text{preference}}} . \tag{3.2}$$

The error term ξ^g follows the Type-I extreme value distribution. It will realize and be observed at graduation in stage 2 before students choose between working and graduate school. The capacity and admission of graduate schools are not modeled, but are implicitly captured in Equation (3.2). To fit the data, the estimated Equation (3.2) should yield a lower payoff from graduate studies for students with lower human capital and from lower-quality undergraduate colleges, making them less likely to attend graduate school.

3.1.4 *lifetime payoff of working after college graduation*

If the individual enters the labor market after graduation and chooses to live and work in province $k \in \mathcal{N}$, the lifetime payoff is:

$$u_{sk}^w = \underbrace{\sum_{t=t_g}^{T} \beta^{t\text{-}1} \ln C_t}_{\substack{\text{lifetime payoff} \\ \text{from consumption}}} + \underbrace{D(k, j_s)' \gamma_1^w + D(k, h)' \gamma_2^w}_{\substack{\text{distance function} \\ \text{(work, college) (work, home)}}} + \underbrace{\gamma_{3,k}^w}_{\substack{\text{amenity} \\ \text{level in } k}} + \underbrace{\rho \chi \mathbb{1}(k{=}\tau) + \xi_k^w}_{\substack{\text{idiosyncratic} \\ \text{preference}}} . \tag{3.3}$$

The first term is the lifetime utility from consumption, where the utility in each period equals the log period consumption. Following Arcidiacono (2005) and Kennan and Walker (2011), I assume that college graduates have access to a perfect credit market so that they can borrow or save without restriction at a given interest rate.[15] The consumption utility maximization then implies the same consumption level in all post-college periods:

$$\sum_{t=t_g}^{T} \beta^{t\text{-}1} C_t = \underbrace{E\left[\sum_{t=t_g}^{T} \beta^{t\text{-}1} W_{skt}\right]}_{\substack{\text{Expected lifetime} \\ \text{earnings in } k}}, \tag{3.4}$$

$$C_t = E\left[\sum_{t=t_g}^{T} \beta^{t\text{-}1} W_{skt}\right] \Big/ \sum_{t=t_g}^{T} \beta^{t\text{-}1}, \quad \forall t, \tag{3.5}$$

[15] The financial relationship between young adults and their parents is typically very close in China. For example, many Chinese parents will help their children buy their first house (Wei et al., 2017).

where W_{skt} is the earnings in province k in year t after graduation, and the expectation is taken with respect to the distribution of the future earning shocks.

The distance function $D(k, j_s)$ in Equation (3.3) is intended to capture the causal effect of college location on residential choices. The location of the college may matter for two main reasons. First, searching for jobs near the college location may have some labor market advantage (Huang et al., 2022). Students may acquire more information on the local than a distant labor market after spending four years studying and living there. It is also less costly to take internships and job interviews within a certain area near the college location. The second is the migration cost, which is broadly defined. People may get accustomed and develop attachment to the place where they have spent four years living, and leaving the college location incurs both monetary and psychological migration costs, as well as a one-time cost of throwing away the location-specific capital.

China's current migration control policy (i.e., residence registration, or *hukou* in Chinese) in general gives fresh four-year college graduates large freedom to choose their first residence location and obtain the local *hukou*,[16] except in Beijing and Shanghai, which have a much stricter requirement when offering *hukou* to migrant workers (Zhang et al., 2019; Ge and Wu, 2020).[17] Inside the distance function $D(k, h)$, there is an additional interaction between an indicator function $\mathbb{1}(k{=}h)$ and an indicator for h being Beijing or Shanghai, to capture the special *hukou* restriction in the top-two Chinese cities for migrants and thus the advantage of being a local student or resident. After college graduation, individuals can freely choose their working province $k \in \mathcal{N}$. However, once residence is established after college, further inter-provincial migrations are abstracted away. This is consistent with China's *hukou* and migration restriction on workers other than fresh college graduates (Zhang et al., 2019). An alternative interpretation is that starting residence and career in a province leads to a specific expected distribution of subsequent migration choices. In this distribution, most people are permanent stayers, as observed in the data, but some people, especially those in certain places like Beijing and Shanghai, may be more likely to leave later if they cannot obtain a local *hukou*. Such distributions associated with each initial residence choice are not explicitly modeled; nonetheless, the expected lifetime payoff based on these expected distributions is absorbed by the provincial fixed effect $\gamma^w_{3,k}$.

Similarly to the college stage, the idiosyncratic preference in Equation (3.3) also contains two

[16] A common method to obtain a local *hukou* is through the local worker service center run by the government, especially for fresh college graduates who are employed by private firms and/or rent apartments. When they purchase local housing at a later time, they can change their *hukou* address from the worker service center to their own residence location.

[17] For more details of the *hukou* system, see Song (2014) and Colas and Ge (2019).

parts. The first part captures the dynamic correlation in location preference through the unobserved type τ. Students may like a place and therefore choose it for both college and work. Alternatively, students may choose to study in one place just because they prefer to live there after college, and attending college in that place can help them find a job there. Therefore, it is important to control for such selection to identify the causal effect of college location on work location choices. The second part of the idiosyncratic preference, shock ξ_k^w, follows the Type-I extreme value distribution. The ξ_k^w will be realized and observed in stage 3 before students choose their residential province k. Preference shocks ξ_k^w in stage 3 and ξ^g in stage 2 are assumed to be independent from each other for analytical tractability.

3.1.5 Earnings

Lastly, log earnings in province k in year t after graduation are specified as follows:

$$\ln W_{skt} = \delta_0 + \underbrace{\delta_1 \theta^0}_{\substack{\text{pre-college} \\ \text{human capital}}} + \underbrace{\delta_2 V_s}_{\text{college quality}} + \delta_3 \theta^0 V_s + \underbrace{g_{kt}}_{\substack{\text{earnings} \\ \text{growth in } k}} + \underbrace{\varepsilon_{kt}}_{\substack{\textit{i.i.d.}\text{ shock} \\ \sim N(0,\sigma_\varepsilon^2)}} . \tag{3.6}$$

College graduates are paid based on their human capital in the labor market. Earnings are a function of the student's pre-college human capital, college quality and their interaction. This specification implicitly incorporates human capital production during college. Earnings shock ε_{kt} is assumed to be *i.i.d.* for tractability. Theoretically, the shock term ε_{kt} contains both the human capital production shock during college and the *i.i.d.* earnings shock in the labor market, but they cannot be separately identified. From the student's perspective, the human capital production shock during college cannot be observed when they choose which college to attend, so remains zero in expectation. Thus, from an ex-post point of view, ε_{kt} may follow a distribution of $N(\zeta_s, \sigma_\varepsilon^2)$, in which ζ_s is the realized human capital production shock in college s. However, ex-ante, ζ_s is zero in expectation.

Earnings after college also depend on local productivity; g_{kt} captures the average life-cycle earnings profile of college-educated workers in province k.[18] The earnings should be considered as the average across industries, given that the industry structure in each local labor market is not included in the model.

[18] The earnings equation is in real terms, so the earnings growth term g_{kt} does not contain inflation.

3.2 Individual's problem

The individual's problem is solved backward. After college graduation, if a student chooses to enter the labor market in stage 2 ($g = 0$), she will subsequently choose a province in which to live and work in stage 3. The individual's problem in stage 3 is:

$$\max_{k \in \mathcal{N}} \overline{u}_{sk}^{w} + \xi_k^w, \tag{3.7}$$

where the $\overline{u}$ term represents the deterministic part of the corresponding payoff. The preference shock ξ_k^w is realized and observed at the beginning of stage 3 before choosing k. Given that ξ^w follows the Type-I extreme value distribution, the probability of choosing province k in stage 3 conditional on graduating from college s, p_{sk}^w, follows the convenient logit type:

$$p_{sk}^{w} = \frac{\exp\left(\overline{u}_{sk}^{w}/\sigma^{w}\right)}{\sum\limits_{k' \in \mathcal{N}} \exp\left(\overline{u}_{sk'}^{w}/\sigma^{w}\right)}, \tag{3.8}$$

and the expected payoff of entering the labor market directly after college s, denoted as EV_s^w, has a closed form expression.

In stage 2, college graduates choose between graduate school ($g = 1$) and entering the labor market ($g = 0$):

$$\max_{g \in \{0,1\}} \overline{u}_s^g \mathbb{1}(g = 1) + EV_s^w \mathbb{1}(g = 0) + \xi^g. \tag{3.9}$$

In terms of timing, the shock ξ^g is realized and observed before choosing g, while ξ^w is not observed at this stage. Under the assumption that ξ^g follows the Type-I extreme value distribution and is independent from ξ^w, both the probability of choosing graduate school over working after college s, $p_s^{g=1}$, and the expected lifetime payoff of graduating from college s, EV_s, have closed-form expressions.

In stage 1, each college $s \in \mathcal{S}$ allocates an admission quota Q_{hzt}^s to province h track z in year t. Given the allocated set of quotas $\{Q_{hzt}^s\}_{s \in \mathcal{S}}$, each province runs an independent centralized admission based on $NCEE_{hzt}$. Each individual submits an application list $\mathcal{L}$ to maximize the expected lifetime utility:

$$\max_{\mathcal{L}} \sum_{s \in \mathcal{S}} \left(\overline{u}_s^c + \xi_s^c + \beta^4 EV_s\right) P\left(\text{admitted by } s | \mathcal{L}, NCEE_{hzt}\right). \tag{3.10}$$

When making college application decisions, students know the distribution of current and previous years' NCEE scores and previous years' admission cutoffs of each college in their province. Following the literature, I assume that individuals take expected admission cutoffs as given and therefore imagine that their individual choices will not affect the equilibrium in large markets (Calsamiglia et al., 2020). Given that the main interest of this paper is student's college and location choices rather than the admission mechanism and its associated algorithm, details about how students form the probability of being admitted by each college given the application list are not needed for identification and thus not modeled.

3.3 Identification of college choices

The identification builds on the *asymptotic stability* assumption of the admission outcome (Fack et al., 2019; Agarwal and Somaini, 2020). (Strict) *Stability* means every student is admitted by her favorite *ex-post feasible* school given realized admission cutoffs. It allows for the identification and estimation of individual preferences without the need to model the choice of application list and solve for the equilibrium of college application (Fack et al., 2019).

Under the stable admission outcome, the ex-post feasible college set, denoted as $\mathcal{S}^f$, can be recovered by comparing an individual's NCEE score with ex-post admission cutoffs. Note that all admission quota constraints are already embedded in these equilibrium admission cutoffs. The individual's problem in Equation (3.10) is then reduced to a standard discrete choice problem: to choose college s from her feasible set $\mathcal{S}^f$ to maximize her expected lifetime utility:[19,20]

$$\max_{s \in \mathcal{S}^f} \overline{u}_s^c + \xi_s^c + \beta^4 EV_s. \tag{3.11}$$

Since the preference shock for college ξ_s^c follows the Type-I extreme value distribution,[21] this gives the standard logit-type choice probability:

[19] Infeasible colleges effectively have a flow payoff of negative infinity: $\overline{u}^c(s|\Omega_i) = -\infty$ if $s \notin \mathcal{S}_i^f$.

[20] An alternative way to analyze the application stage is to assume students can apply to all colleges. After receiving the admission outcome (acceptance or rejection) from all colleges, students choose one college to attend from those that accept them. These colleges equivalently form the feasible set $\mathcal{S}^f$.

[21] Conceptually, the shock ξ_s^c captures the idiosyncratic preference for each college s. In reality, it also contains application mistakes that individuals make. There is a small but growing literature showing that students have different levels of sophistication during college application (Luflade, 2017; Calsamiglia et al., 2020), especially among students from different family backgrounds or between urban and rural areas (Ye, 2023). Allowing for correlation between the pre-college human capital θ^0 and the shock ξ_s^c would substantially complicate the analysis, if not make it impossible, given that the distribution of θ^0 is not directly observed and needs to be recovered during estimation.

$$p_s^c = \frac{\exp\big((\overline{u}_s^c + \beta^4 EV_s)/\sigma^c\big)}{\sum\limits_{s' \in \mathcal{S}^f} \exp\big((\overline{u}_{s'}^c + \beta^4 EV_{s'})/\sigma^c\big)}. \tag{3.12}$$

Denote this as the stability-based estimator.

The stability-based estimator, utilizing the ex-post feasible college set, is essentially a fuzzy regression discontinuity (RD) design. Take the identification of the college location's impact on residential location choice as an example. When unobserved heterogeneity in location preferences are controlled for through individual types, the term $D(k, j_s)$ in the payoff of working in province k in Equation (3.3) is identified by comparing students whose NCEE scores are equal to or slightly above the admission cutoff of college s_1 to students whose NCEE scores are slightly below. The former students are eligible for all the colleges for which the latter students are eligible, but the former are also eligible for this additional college s_1. Assuming that s_1 is located in province j_1, all else equal, the former students are (weakly) more likely to attend colleges in j_1 compared to the latter students, because one more college is available to choose in j_1. Of course, the RD here is necessarily fuzzy. Importantly, the example above is just one place where the discontinuity occurs. The discontinuity will occur at every admission cutoff level, and all of them jointly provide identification for the causal effect of the college location on residential location choices, captured by $D(k, j_s)$.

Admittedly, the assumption of *strictly* stable college admission outcomes is not realistic. However, as shown in Fack, Grenet and He (2019), college admission using a single priority index (the NCEE score in this paper) in a large market (as in China) is likely to be *asymptotically* stable. As the number of students and the capacity (thus quota) of each college go to infinity proportionally, the fraction of students not matched with their favorite feasible school (due to an imperfect ability to predict admission cutoffs or other randomness) converges to zero. Fack et al. (2019) proves that the stability-based estimator specified in Equation (3.12) is consistent under asymptotic stability.[22]

3.4 Identification of other parameters

Utility parameters for graduate school and residential choices are identified following the standard discrete choice literature (McFadden, 1974; Rust, 1994). The parameter in front of the lifetime utility from consumption in u^w in Equation (3.3) is normalized to 1. Given this normalization, util-

[22] Fack et al. (2019) provide numerical evidence suggesting that typical real-life college markets are sufficiently large for a good asymptotic approximation.

ity parameters determine the relative difference in payoffs between choice alternatives, and thus the probability of choosing each alternative. Because the idiosyncratic preference shocks, ξ's, follow the Type-I extreme value distribution and are independent from each other, we have the logit-type choice probabilities specified in the individual's problem. Matching these choice probabilities with the observed choice distribution identifies these utility parameters.

The unobserved heterogeneity in location preference is identified by the dynamic correlation of location choices, once the effect of college location on work location choices is pinned down by the fuzzy RD discussed above. The amenity terms in u^c and u^w capture the common value of each location, and the distance terms between the college location and the residential location capture the average locational correlation. If we observe that a subgroup of people systematically chooses a particular college location τ with a probability higher than that implied by the common values and, at the same time, disproportionally chooses to work there after college, it suggests that these people have an unobserved preference for this location. The average share of these subgroups identifies the type probability π_τ. The choice probability of college location that deviates from the level implied by the common location values identifies the magnitude of the unobserved preference χ. Lastly, the correlations between home, college and work locations in these subgroups help identify the dynamic scaler parameter ρ.

4 Measuring the distribution of student quality across provinces

This section builds a measurement model to recover the distribution of students' pre-college human capital θ^0 in each province that is comparable nationwide, overcoming the non-comparability of the provincial NCEE scores. Recovering a comparable distribution of college applicants in each province permits a direct evaluation of the admission quota allocation from a merit perspective. Moreover, it is the key prerequisite to evaluate any counterfactual admission policies.

4.1 Measurement model

There are two measures of the human capital level of each student: the NCEE score received in the province h track z in year t, and the cumulative college GPA at graduation in college s major m. To restore the comparability of NCEE scores, the main intuition is to find a measure that is comparable *across* provinces. College GPAs can help compare students who come from different provinces but are enrolled in the same university. Consider two students, A and B. Student A has an

NCEE score of 600 in hzt, and student B has an NCEE score of 550 in $h'z't$. Both attend the same college and major. Both have a college GPA of 3.5. Assume that both the NCEE and the GPA are functions of human capital and, for simplicity, ignore all measurement errors and human capital production shocks in this example. First, A and B should have the same level of human capital at the end of college, implied by the same college GPA. Furthermore, since they receive the same college education, regardless of the technology of human capital production during college, their human capital at college entry should also be the same, as measured by their NCEE scores. Thus, the NCEE score of 600 in hzt should be equivalent to 550 in $h'z't$. Admittedly, both NCEE scores and college GPAs are noisy measures of human capital, and human capital production during college is also stochastic, so matching an NCEE score in hzt to another score in $h'z't$ is not as straightforward as discussed above. The latent factor model specified below considers these factors and formalizes the idea of combining the NCEE scores and college GPAs to recover the comparable pre-college human capital.

I view an individual's latent pre-college human capital θ^0 as a one-dimensional measure that aggregates their innate ability and learned knowledge and skills upon college entry. Consider θ^0 in a linear latent factor model with two measures: the NCEE score received in the province h track z in year t, $NCEE_{hzt}$, and the cumulative college GPA in the college s major m, GPA_{sm}.

$$NCEE_{hzt} = \kappa_{1,hzt} + \lambda_{1,hzt}\theta^0 + \sigma_{1,hzt}\nu_1; \tag{4.1}$$

$$GPA_{sm} = \kappa_{2,sm} + \lambda_{2,sm}\theta^0 + \sigma_{2,sm}\nu_2. \tag{4.2}$$

In Equation (4.1), the raw NCEE score, $NCEE_{hzt}$, is a function of θ^0 plus a measurement error ν_1. The parameters $\kappa_{1,hzt}$, $\lambda_{1,hzt}$ and $\sigma_{1,hzt}$ are subscripted by hzt to allow for the province-track-year-specific testing technology of the NCEE. Unlike Equation (4.1), which is a pure measurement equation, the GPA equation in Equation (4.2) contains information on both the human capital production and the measurement during college.[23] The parameters $\kappa_{2,sm}$, $\lambda_{2,sm}$ and $\sigma_{2,sm}$ are subscripted by sm to allow for school-major-specific college inputs in both human capital production and GPA measurement technology. The error term ν_2 aggregates the production shock and the

[23] To see this, consider human capital production during college in school s major m in the following form:

$$\theta^c = \tau_{0,sm} + \tau_{1,sm}\theta^0 + \varepsilon^0_{sm},$$

where θ^c is the latent human capital at the end of college, and the parameters $\tau_{0,sm}$ and $\tau_{1,sm}$ capture the college-major-specific production technology. Then consider the cumulative college GPA as a noisy measure of human capital measurement error.

Equations (4.1) and (4.2) apply to the population of all NCEE takers in a given year, including both admitted (for which college GPA are observed) and not admitted students (if they had been admitted). In addition, the parameters in the GPA equation are assumed to be stable across years, so there is no t-subscript in Equation (4.2).[24] Following the standard approach in factor model analysis, I normalize $\kappa_{1,(hzt)^*} = 0$ and $\lambda_{1,(hzt)^*} = 1$ for a specific $(hzt)^*$ to pin down the position and scale of the latent human capital θ^0. Both ν_1 and ν_2 have zero mean and unit variance, are $i.i.d.$ across individuals, and are independent from θ^0 and from each other in the population.[25]

4.2 Identification of measurement parameters

Identification of this factor model takes advantage of its structure: the NCEE equation does not depend on sm, so it can measure students from the same province attending different colleges, while the GPA equation does not depend on hzt, so it can measure students enrolled in the same college and major but originally from different provinces. Under the current college admission regime, students from the same province go to multiple colleges in different locations. At the same time, each college admits students from multiple provinces. Intuitively, students from the same province are effectively measured by multiple colleges using GPAs, with each attended college measuring one part of the students in that province. These various college GPA measures are then linked by the same NCEE measurement in that province. At the same time, each college provides comparability of the NCEE scores of students from different provinces using their own GPA measurement, and this NCEE score comparison is also done multiple times by different colleges and for different parts of the NCEE score distribution based on college selectivity. This identification strategy utilizing the "crossing" structure is similar to that in Abowd, Kramarz and Margolis (1999) in the worker and firm setting. The key variation in Abowd et al. (1999) for identifying worker and firm heterogeneity in wages is the between-firm mobility of the individual

upon graduation:

$$
\begin{aligned}
GPA_{sm} &= \varphi_{0,sm} + \varphi_{1,sm}\theta^c + \varepsilon^c_{sm} \\
&= (\varphi_{0,sm} + \varphi_{1,sm}\tau_{0,sm}) + (\varphi_{1,sm}\tau_{1,sm})\theta^0 + (\varepsilon^c_{sm} + \varphi_{1,sm}\varepsilon^0_{sm}) \\
&= \kappa_{2,sm} + \lambda_{2,sm}\theta^0 + \sigma_{2,sm}\nu_2.
\end{aligned}
$$

where the φ parameters allow for college-major-specific measurement technologies of the college GPA. The τ's and φ's would not be separately identified.

[24] I do not find any evidence for college GPA inflation during the sample period.

[25] Given that θ^0 is an one-dimensional aggregate measure of human capital, this categorization of ν_1 and ν_2 implies that all other observables, such as family income or parental education, only affect θ^0, but not ν_1 or ν_2.

workers. Analogically, the students in this paper are workers who are initially employed by a firm called high school hzt and then move to the second firm called college sm.

Beyond the basic intuition discussed above, identification is further complicated by the selection on NCEE scores. This measurement model can only be estimated using a selected sample. It only contains students who are admitted by a college so that I observe their college GPAs, and this college choice is based on both the underlying ability θ^0 and the measurement error shock ν_1 in test scores. As a result, the conditional mean of ν_1 and its conditional correlation with θ^0 will not necessarily be zero, different from the normalization and assumption for unconditional distributions. The solution is to jointly estimate the measurement model and the behavioral model and rely on the latter to simulate the selection terms (e.g. conditional moments of ν_1 and θ^0 in a given college) needed for identifying the measurement parameters, which is discussed in more detail in the next subsection. Appendix B provides the formal proof of identification in the presence of college selection and discusses how the identification that started with certain province-college combinations can be extended to cover all provinces and colleges.

4.3 Empirical specifications to recover the provincial distributions

The main goal is to recover the distribution of θ^0 in each province, $F_h(\theta^0)$. They will be used in the estimation of the structural model. I assume that the pre-college human capital θ^0 of the NCEE exam takers in each province follows a normal distribution, $F_h(\theta^0) \sim N(\mu_h^\theta, (\sigma_h^\theta)^2)$, and this distribution is stable across the six cohorts studied in this paper.[26] The choice of the high school track z is assumed to be independent of $F_h(\theta^0)$, therefore $F_h(\theta^0|z) = F_h(\theta^0)$ for both tracks.[27] I also assume both error terms in the measurement model, the unconditional ν_1 and ν_2, follow the standard normal distribution $N(0, 1)$. Note that each of them will have a scaling parameter σ to account for potentially different variances across provinces or colleges.

Given these normal distribution assumptions, the linear measurement equation implies that $NCEE_{hzt}$ should also be normally distributed. As with most raw test scores, it is difficult to map the scale and distribution of raw NCEE scores to the underlying latent skills (the θ^0 in this paper).

[26] The normal distribution assumption is made for identification. In principle, once the measurement parameters are identified, the distribution of θ^0 can be nonparametrically identified for students admitted from each province by combining the subsamples of all the colleges. However, this procedure is tedious in practice and is restricted by the small sample size of some subsamples. Furthermore, even if the distribution of admitted students is nonparametrically identified, some form of interpolation still needs to be performed to get the distribution of all NCEE exam takers.

[27] Although there is clear gender difference in choosing the tracks (more female students in humanities), there is no clear evidence for or against the similar achievement distribution between two tracks.

Most standardized test scores, such as the SAT in the US, use some form of normalization. I normalize and standardize the raw $NCEE_{hzt}$ as follows. First, I convert the raw NCEE scores to percentile rankings within each hzt. Second, these percentile rankings are mapped to the standard normal distribution and assigned the values from the inverse cumulative density function of the standard normal distribution. After this transformation, the normalized NCEE scores in each hzt follow the standard normal distribution. Note that the raw values of the NCEE scores are not important. The purpose of the measurement model is to recover the underlying distribution of θ^0, which is anchored to an arbitrarily chosen position and scale because the unobserved θ^0 does not have a natural position and scale. Normalizing the values of the NCEE scores will only change the value and interpretation of the measurement parameters $\{\kappa, \lambda, \sigma\}$.

For college GPAs, they should ideally be defined at the level of each college major in each school to account for potentially different measurement across college majors within the same school. The student survey data that I use indeed provide information on college majors, but the sample size in each school-major cell is usually too small. Due to this data limitation, the measurement parameters in the GPA Eq.(4.2) are assumed to be homogeneous within college s for all major m's in the main empirical analysis. In a further robustness check, I relax the restriction by allowing for a college major fixed effect.[28]

With these empirical specifications, the provincial distribution $F_h(\theta^0)$, that is, the mean μ_h^θ and the standard deviation σ_h^θ for each province h, can be recovered from the measurement model, because $F_h(\theta^0)$ is fully determined by $\{\kappa, \lambda, \sigma\}$ and the NCEE score distribution. Given that the normalized $NCEE_{hzt} = (\kappa_{1,hzt} + \lambda_{1,hzt}\theta^0 + \sigma_{1,hzt}\nu_1) \sim N(0,1)$ and the assumption that $\nu_1 \sim N(0,1)$ and is independent from θ^0, we can derive that

$$\mu_h^\theta = -\kappa_{1,hzt}/\lambda_{1,hzt} \quad \text{and} \quad \sigma_h^\theta = \sqrt{1-\sigma_{1,hzt}^2}\Big/\lambda_{1,hzt}, \tag{4.3}$$

which impose over-identifying restrictions for the estimation.

[28] I use the one-digit classification of the college majors. There are in total 12 categories.

5 Data and estimation strategy

5.1 Data source and sample selection

Administrative data of college admission

The first dataset is the administrative records of college admission outcomes from the Ministry of Education of China. It covers the universe of students admitted to four-year colleges through the NCEE each year in China. I obtain a 10% random sample of the micro-level student data for six cohorts that were admitted between 2006 and 2011. The sample consists of 1,645,728 students and records each student's home province, NCEE track, NCEE score and its provincial ranking, as well as the admitted college and major. I drop students admitted by military, police, arts, and sports colleges, and restrict the sample and analysis to 30 out of 31 provinces in China, excluding Tibet (Xizang). College admission in Tibet is both special and complicated due to various pro-Tibetan policies. Many Tibetan students study in high schools located in other provinces under affirmative action policies and follow favored college admission rules, procedures, and cutoffs. Given the small size of the student population in Tibet (on average 0.15% of all NCEE exam takers each year in China) and these complications, Tibet is excluded from the analysis throughout the paper.

The Chinese College Student Survey (CCSS)

The second dataset is the Chinese College Student Survey (CCSS), which is administered by Tsinghua University and covers a nationally representative sample of the same six cohorts of college students in the administrative data. The CCSS first randomly selects schools from all colleges in China, stratified by college tier and location. Within each college, it further randomly selects and surveys students at the time of college graduation. Similar to the administrative data, I restrict the sample to students who were admitted only through the sciences or humanities track in the NCEE and studied in non-military/police schools. The final sample used in the analysis covers 79 four-year colleges and 29,789 students. Each college was surveyed at least once by the CCSS, and many were surveyed in multiple years. On average, 149 students were surveyed each year in each school, which is about 6% of the average cohort size in one college.

One disadvantage of the administrative data of college admission is that it does not track students after entering college, so I do not observe their performance during college or their choices after college. The CCSS supplements the administrative data well. It contains all variables in the administrative records and further collects detailed information on individual characteristics, family background, college performance (including the four-year cumulative GPA), and post-graduation choices. If students enter the labor market after college graduation, the survey collects the first job's information including location, annual earnings, characteristics of the job and the employer. For students who go to graduate school, the name of the institution is recorded. Unfortunately, the CCSS does not track college graduates beyond the first job after college. Recent literature shows that the variation in initial earnings persists largely throughout an individual's career in both the US (Carr and Wiemers, 2022) and China (Jia and Li, 2021). I then supplement the CCSS with the China Family Panel Studies (CFPS) to obtain the earnings growth pattern, which is discussed below. Additional details of data sources and sample construction are described in Appendix C.

Life-cycle earnings profile by location

I use the China Family Panel Studies (CFPS) to obtain the life-cycle profile of earnings after college. The CFPS is a nationally representative longitudinal survey, similar to the Panel Study of Income Dynamics (PSID) in the US. The baseline survey was launched in 2010 and started with a sample of 14,608 households and 33,600 individuals. These households and individuals are subsequently surveyed every two years. I use the six rounds of the survey from 2010 to 2020.

The CFPS is relatively new and lasts for only eleven years, so I adopt a synthetic cohort approach to calculate the life-cycle earnings profile. I restrict the sample to individuals aged between 22 and 62 in each round of the survey, living in urban area, and who are college educated. Constrained by the multi-stage sampling design of the CFPS, the sample size in some provinces is not large enough to get accurate estimates, so I aggregate provinces according to the division of administrative regions: Beijing and Shanghai, other East provinces, Central, West, and Northeast. I run an individual fixed-effect regression of the log annual earnings for each region separately to get the estimated coefficients for age dummies. These coefficients are used as the earnings growth term for each province and working period.[29]

College quality measures

I extract the quality measures for each college from the Higher Education Undergraduate Teaching Quality Report in 2014, published by the Higher Education Evaluation Center of the

[29] One limitation of this approach is that it ignores the possibility of spatial sorting on the individual's unobserved ability, which may affect the earnings growth profile in each location.

Ministry of Education. Each college was required to report a list of consistently defined statistics on students, faculty, infrastructure, funding, and expenditure related to undergraduate teaching. It would be ideal to have these reports covering the time period in the student-level data discussed above, but the year of 2014 is the only data currently available to researchers.

There are two main advantages of using this dataset over commonly seen college rankings in measuring college quality. First, it covers the universe of undergraduate institutions in China, whereas university rankings generally have poor coverage and accuracy at the bottom half of the school distribution. Second, it is specifically designed to collect information on undergraduate teaching for quality control purposes, and thus suits well for the estimation of human capital production during college. In contrast, college rankings generally incorporate other aspects of information as well, such as research, reputation, etc.

I use principal component analysis (PCA) to aggregate college quality measures into a one-dimensional variable. To account for other unobserved aspects of college quality affected by college tier and ownership, I also include indicators for Project-985 schools, Project-211 schools, and private schools in the PCA.[30] The first principle component is further normalized to have a mean of zero and a standard deviation of one to be used as the one-dimensional college quality measure for V_s in the model. Table A1 in the appendix reports the raw measures and the estimated loading factors of the first principle component associated with each measure. Table A2 reports the top 20 schools ranked by the estimated V_s and corresponds well to the general consensus and various university rankings.

5.2 Joint estimation of the behavioral model and the measurement model

The structural model of college and residential choices is estimated via the simulated method of moments (SMM) following the approach in Gourieroux, Monfort and Renault (1993). I target a set of moments in the SMM based on the simulated data to the observed counterparts that can be calculated using the administrative data on college admission and data of college student surveys. These moments cover college choice, choices upon college graduation, residential locations, dynamic correlation of locations, wages, and conditional mean, variance, and covariance of NCEE scores and college GPAs from the measurement model. Appendix D lists the targeted moments used in the SMM.

[30] Project-985 and Project-211 colleges receive comprehensive support from the central government in teaching and research. These supports may not be fully captured by the measures in teaching quality reports. Private colleges in general receive much less government support compared to public schools.

During the SMM estimation, for each province, track and year, a group of college applicants is simulated with randomly assigned initial conditions, including the pre-college human capital θ_0, the NCEE exam shock ν_1, and the idiosyncratic location preference τ. The distribution of pre-college human capital θ_0 in each province, $F_h(\theta^0)$, will be recovered from the measurement model. Once the measurement model and $F_h(\theta^0)$ are identified, the structural model can be estimated. However, this is not straightforward because the estimation of the measurement model requires first calculating several selection terms regarding the correlation between the pre-college human capital θ_0 and the NCEE exam shock ν_1, which need to be derived from the sorting patterns between students and colleges in the behavioral model as discussed in Section 4.2. Given the nature of the two interdependent estimation parts, the structural model and the measurement model are jointly estimated in the SMM.

6 Estimation results

6.1 Recovered pre-college human capital

Provincial distributions. After estimating the mean (μ_h^θ) and standard deviation (σ_h^θ) of NCEE exam takers' pre-college human capital in each province, the distribution of θ^0 in each province is first rescaled to set the national mean at zero and the standard deviation at one for easier interpretation. Overall, there is a large variation in average pre-college human capital levels across provinces. The gap in the mean of θ^0 between the provinces that have the highest and lowest average student quality is approximately 1.3 national standard deviations, while the spread of pre-college human capital within each province is relatively comparable. Figure A3 provides more details. To better understand the potential source of this disparity in student achievement, Figure 1 plots the provincial mean of θ^0 against the log GDP per capita in panel (a) and against the log of the total public expenditure per student averaged over primary, middle and high school in each province in panel (b). Perhaps unsurprisingly, the estimated average pre-college human capital level is positively related with local economic development and per student public expenditure on education before college. In terms of correlation, these two factors can account for roughly 55% of the variation in the average pre-college human capital between provinces.

Student quality vs. college quotas. With the recovered θ^0 distribution, I can formally evaluate the observed allocation of college quotas across provinces from a merit perspective. Figure 2 plots the average number of four-year college quota per NCEE exam taker between 2006 and 2011 against the mean of θ^0 in each province. The allocation of quota is positively correlated with the average quality of college applicants to a limited extent, with the correlation equal to 0.609. Another approach to test to what extent the allocation of quota is driven by student merit is to compare marginal students, that is, the last student admitted from each province, as shown in Figure A4. Following the consumer theory that the marginal utility per dollar spending should be the same for all goods, if the policy goal is to send the best group of students in China for higher education up to college capacity, the quotas should be allocated such that the θ^0 of the last admitted student is equalized across provinces. However, the gap of θ^0 among the bottom group of admitted students in each province is large, up to 1.1 national standard deviations.[31]

Validity check. To verify the validity of the recovered measure of student's pre-college human capital using the measurement model, I compare θ^0 with the result of a series of cognitive tests conducted by the China Family Panel Studies (CFPS). The CFPS survey administered standardized tests on words, math, and number series to all respondents during the 2010 and 2012 rounds. I restrict the sample to students aged 12 to 17, currently in middle or high school. Age-adjusted raw scores from the three cognitive tests are combined in the principal component analysis. Figure 3 plots the provincial mean of θ^0 and the external standardized cognitive test scores for the 25 provinces sampled by the CFPS. The correlation between the two is 0.834. This exercise provides strong support for the validity of the comparable measure of the pre-college human capital in each province.

6.2 Structural model parameters

The estimated value and standard error of the structural model parameters are reported in Table A3. Several aspects of the model estimation results deserve further discussion. They have important implications for evaluating counterfactual policies on college seat allocation.

Technology of human capital production. The estimated earnings equation shows that both academic preparation prior to college and college quality are productive in developing the human capital of college graduates and increasing the return on the labor market. Importantly, the parameter of the interaction between pre-college human capital θ^0 and college quality V_s is positive

[31] Figure A5 further illustrates the bias of using raw NCEE scores, instead of the nationally comparable θ^0 distribution, to compare students across provinces. Therefore, recovering a comparable measure of students is essential before examining any counterfactual allocations of college seats.

and statistically significant in the earnings equation, demonstrating the complementarity between student and college quality. Starting from a nationally average college applicant ($\theta^0 = 0$), if her pre-college human capital increases by one standard deviation ($\theta^0 = 1$), the marginal return from attending a better college is 38% higher in initial earnings. This complementarity between student and college quality in higher education is consistent with studies of other countries like the US (Dillon and Smith, 2020). Direct evidence on the interaction effect between student and college quality in human capital production is scarce in the literature. To my best knowledge, this is the first study to provide an estimate for the level of complementarity in China.[32] Interestingly, the estimated magnitude of the complementarity in earnings in China is similar to that found in Dillon and Smith (2020) for college graduates in the US.[33]

Preference for location. To better interpret the estimated location preferences, I convert the parameters of the location and distance terms into monetary values. Since the parameter of the discounted lifetime payoff from consumption is normalized to 1 (Eq. 3.3), and the smoothed period consumption equals the discounted lifetime earnings divided by a sum of the discount factor (Eq. 3.5), each additional term γ in the utility (Eq. 3.3) is equivalent to a change Δ in consumption or earnings as follows:

$$\begin{aligned}
&\sum_{t=1}^{T} \beta^{t\text{-}1} \ln C_t + \gamma = \sum_{t=1}^{T} \beta^{t\text{-}1} \ln(C_t + \Delta);\\
&\Delta = \left[\exp\left(\frac{\gamma}{\sum_{t=1}^{T} \beta^{t\text{-}1}}\right) - 1\right] C_t\\
&\quad = \left[\exp\left(\frac{\gamma}{\sum_{t=1}^{T} \beta^{t\text{-}1}}\right) - 1\right] \frac{E\left[\sum_{t=1}^{T} \beta^{t\text{-}1} W_{skt}\right]}{\sum_{t=1}^{T} \beta^{t\text{-}1}}.
\end{aligned}$$

For each payoff parameter γ, I calculate the value of Δ for each individual and then take the average. The parameter is then converted to an annual monetary value in Chinese Yuan (CNY) that is equivalently added to the annual consumption or earnings over the life cycle. The average starting annual salary for college graduates is 27,644 CNY in 2012 value, and the average annual

[32] Diao et al. (2022) also find complementarity in earnings between the student's within-province percentile of the NCEE score and college ranking, but they did not provide a quantitative estimate.

[33] Dillon and Smith (2020) find that the marginal return from college quality will increase by 18-37% if student quality moves from the 50th to the 75th percentile. Mapping the 50th and 75th percentile back to the recovered national distribution of θ_0, I estimate that the marginal return from college quality will increase by 27% in earnings.

consumption (smoothed over the life cycle) among individuals is 115,113 CNY.[34]

The location preference values are summarized in Table 2. Students show a high level of home attachment during both the college (for 4 years) and the work (assumed to be 40 years) stage. For example, the value of living and working in the home province after college is equivalent to 11.17% of annual consumption (12,861 CNY) per year. If leaving the home province, the additional utility loss is a concave function of the distance between the college or work location and the home province. The bottom panel of Table 2 shows that the college location plays an important role in the subsequent choice of work location, which is consistent with recent evidence both in China (Huang et al., 2022) and the US (Kennan, 2021). Controlling for the unobserved location preference that persists over time, staying in the college location for work provides a value equivalent to 10.88% of consumption (12,525 CNY) per year, combining the broad benefits in the job market, local connections, migration, etc., and migrating to other provinces incurs additional losses that increase with distance. This causal effect of college location has important policy implications: how to allocate college seats across regions and where these college seats come from can potentially change the spatial distribution of college-educated workers. The estimated idiosyncratic location preference is non-negligible, accounting for 27.3% of the total college location correlation (4.09/(4.09+10.88)); therefore, it is necessary to control for the dynamic correlation in unobserved location preferences. Lastly, the estimated value of amenity is unsurprisingly the highest for the most developed quartile of provinces (Table A4).

6.3 Model fit

With the estimated parameters, I examine how well the model can fit the observed student choices at each stage. I start by plotting the average NCEE scores (all rescaled to be between 0 and 1) of students admitted by each tier of colleges from each province, track and year in model simulation against the data (Figure A6). In college admission that is solely based on test scores, those scores can be seen as student's purchasing power for higher education. As a result, the average test score of admitted students reflects the overall attractiveness of a college and the student's preference. Figure A6 shows that the model can reasonably fit the admission results by college quality and location; the largest and average deviation of the simulated average NCEE scores from the actual ones are 6.19% and 0.06%, respectively.

[34] The age-earnings profile of recent cohorts in China is much steeper than that in the US (Fang and Qiu, 2023). In 2012, 1 USD $\approx$ 6 CNY, and GDP per capita in China and the US is 6,316 and 51,603 USD, respectively.

Following their undergraduate studies, Table A5 displays the average percentage of students who proceed directly to graduate school, averaged across six cohorts. Because the administrative data do not track students after college admission, this share is calculated based on the CCSS-surveyed schools, and the simulated share is calculated by collecting all students who graduate from these schools in model simulation. Given the estimated complementarity between student and college quality in the return of graduate studies, the share of students from top-tier colleges continuing with graduate studies is nearly twice that in the middle-tier and four times that in the bottom-tier colleges. The model can fit the share of graduate studies in all three tiers of colleges almost exactly.

For students entering the labor market after college, Figure A7 first plots the distribution of location choices in the data versus the simulation, also based on the colleges covered by the CCSS survey. The model can well fit the magnitude of spatial share in residential location choices. What is more crucial for counterfactual analysis is the ability to generate the dynamic linkage between location choices from college to work. In Figure A8, the left panel shows the share of students choosing to work in their home province by each home province h, and the right panel plots the share of students who work in the place where their *alma mater* is located by each college province j. Overall, the model performs well in fitting the observed spatial dispersion in these two dimensions.

7 Counterfactual college seat allocations: policy tradeoffs

Utilizing the estimated structural model and the recovered distribution of pre-college human capital in each province, in this section, I examine several counterfactual schemes of spatial college seat allocation. The primary objectives are twofold. The first is to compare the provincial quota system currently used in China's college admission to two natural alternative college seat allocations that do not use spatial quotas, a nationwide merit-based admission and an admission with fully equalized admission rate across provinces, and study their implications on both efficiency and equality. The second goal is to explore the efficiency-equality frontier and position the current quota system within this frontier space to consider the implied policy weights and tradeoffs between the two dimensions.

7.1 Setup of alternative college admissions

I start the counterfactual analysis by considering two natural alternative college admission schemes that do not use spatial quota. The first is a nationwide merit-based admission, denoted as [Merit], and the second scheme gives the same admission rate to each province for each tier of colleges, denoted as [Equal]. Both [Merit] and [Equal] eliminate province-based NCEE testing and quotas. Instead, students in all provinces are given a standardized college entrance exam to ensure that all scores can be ranked nationally.

Under [Merit], a nationwide admission is conducted to admit students up to the existing college capacity using the Serial Dictatorship algorithm, in which students choose the college sequentially based on their national ranking of test scores, regardless of their home province. Therefore, it is solely merit-based and does not have any equality considerations. On the other extreme, the [Equal] admission restricts the admission rate to each tier of colleges to be the same for all provinces. This equals the total number of seats in all colleges in each tier divided by the total number of college applicants nationwide. At first, [Equal] conducts admission in the same way as [Merit] and students choose the college sequentially based on their national ranking. However, once the share of students from the same residence province admitted to a certain college tier reaches the common admission rate cap, subsequent students from that particular province become ineligible for the available seats in that particular college tier. They can choose a college in other tiers if there are still available seats and the corresponding admission rate is not yet binding; otherwise, they will not be admitted by any four-year colleges and take the default outside option. Unlike the quota system in the baseline, [Equal] only caps the overall admission rate without limiting the flow of individual students.[35]

The year 2008 is taken as the baseline in the counterfactual analysis for comparison, which is the middle of the period analyzed in this paper. I simulate the same number of NCEE exam takers in 2008, in total 9,996,601. Each student's pre-college human capital θ^0 is randomly drawn from the recovered provincial distribution $F_h(\theta^0)$. Throughout the counterfactual analysis, the distribution of student size and quality (θ^0), as well as college quality (V_s) and geographical distribution, are fixed to the level of 2008. This setup implies several important underlying assumptions. The policy-invariant distribution of θ^0 rules out the possible response of the testing effort to changes in college opportunities. Note that the recovered θ^0 should be productive in both NCEE and the

[35] I do not consider a *completely* equal quota allocation in which *each university* is forced to allocate college seats equally across provinces, which would be too unrealistic.

college study. Since the preparation effort oriented to the test is generally not productive after the test, it would not be part of θ^0. In addition, the relationship between winning probability and equilibrium effort is usually nonlinear in a contest (Clark and Riis, 1998), so it is difficult to know whether and how testing effort will change in the counterfactual.[36] The college quality V_s is also held constant in the counterfactual, which means that the peer quality change during the re-sorting between students and colleges is abstracted away, similarly in Dillon and Smith (2020). I discuss the implication of this assumption when presenting the counterfactual results.

To make the comparison between baseline and counterfactual easier, provinces are grouped by the quartile of GDP per capita in 2008. Table A6 summarizes the average college resources and student quality in each region. More colleges, especially high-quality colleges, are disproportionally located in more economically developed regions. In the baseline of 2008, provinces in the most developed quartile on average host a large college capacity of 0.373 seats per local college applicant, and those colleges are of the highest average quality. College applicants in the top-quartile provinces also have the highest average pre-college human capital. In contrast, in the least developed quartile, the college capacity is 40% lower, the college quality is lower by more than half the national standard deviation, and the preparedness of college applicants is also lower by 0.36 the national standard deviation. This skewed spatial distribution in terms of college and student resources will have a fundamental impact on [Merit] versus [Equal] admission outcomes in human capital production, equality of college access, and the spatial distribution of skilled labor.

7.2 Policy tradeoffs in efficiency and equality

College opportunities. I first compare college access in different regions of the baseline quota system with counterfactual admission schemes [Merit] and [Equal] in Table 3. In the baseline, college admission rate is determined by the total number of quotas received by the student's residence province. 33.6% of NCEE exam takers can be admitted to a four-year university in the most developed provinces, while the admission rate is 23.7-24.5% in the bottom half (column 1). This allocation of college seats in the baseline, which is found earlier in Figure 2 not fully consistent with the provincial difference in student quality (θ^0), in fact reduces the inequality in college

[36] The policy-invariant distribution of θ^0 also rules out the possibility that the counterfactual change in college access may affect NCEE re-taking and college re-application. In addition, selection into or out of taking the NCEE is abstracted away but would not be a big concern for the purposes of this counterfactual. Because such selection mostly happens at the lower end of the student distribution, counterfactual admission schemes for four-year colleges are not likely to affect these students, and this paper does not study vocational college admission.

opportunities between places to some extent. When moving to [Merit], the allocation of college seats based entirely on merit under nationwide competition, the inequality in admission rate increases to 35.8% versus 21.0% in the richest and poorest regions (column 2). This is driven by the large inequality in the distribution of the human capital of college applicants between provinces, which combines the student's learning ability and the quality of education they have received up to high school graduation.[37] At the other extreme, [Equal] admission, by design, fully equalizes the college admission rate between provinces (column 3).

Combined with changes in admission rate, the average quality of admitted students increases under [Merit] by 0.045 the national standard deviation (Table 3, columns 4 vs. 5), because all college seats are allocated to the highest portion of college applicants nationally without distortion introduced by provincial quotas. However, if the policy goal is to provide equal level of college access for students growing up in different places, the cost would be a lower quality of incoming students (column 6). Importantly, many high-performing students in the top quartile of provinces lose the opportunity to have a four-year college education.

Student-college sorting. Eliminating the distortion of provincial quotas could enhance the sorting between students and colleges, leading to better efficiency of human capital production in higher education due to the estimated complementarity between student and college quality. Figure 4 compares the student-school match between the baseline and two counterfactuals following the approach in Dillon and Smith (2017). Students and colleges are divided into four quartiles based on pre-college human capital θ^0 and the college quality measure V_s. Then, I calculate the gap between the two to measure the undermatch and overmatch of the students. In general, students concentrate in the zero-gap group, indicating assortative matching, but there are apparent large mismatches (≤ -2 and ≥ 2) in the baseline. The [Merit] admission substantially increases the sorting between students and colleges and reduces large mismatches when students are free to choose the college given their national ranking. This indicates that the existing allocation of quota hinders certain high-performing students from obtaining a college education of sufficient quality endorsed by merit., and they are mainly from the developed region according to the previous Table 3. Interestingly, the [Equal] scheme also slightly increases sorting despite admitting a body of students with a lower average quality. The key distinction between [Equal] and the baseline lies in

[37] Separating the two is beyond the ability of the measurement framework in this paper. In the counterfactual, the quality of pre-college education and the distribution of student's pre-college human capital by students in each province are taken as exogenous.

their methods of imposing restrictions. [Equal] only limits the overall admission rate without affecting the flow of students, whereas a spatial quota fixes both the origin (student's home province) and destination (college location).

It is worth noting that assortative matching is not perfect even when students are free to choose a college according to their national ranking under [Merit]. As found in the structural model estimates, students value not only college quality, but also other characteristics, especially location, which matters for work choices. Another possible reason is that peer quality is not included in the college quality measure V_s. When students and colleges are more assortatively matched, peer quality in better schools also increases, which should further increase the sorting. As a result, the positive effect of eliminating admission quotas on student-college sorting found in this analysis should be considered as a lower bound. Improved student-college sorting will increase the aggregate output of human capital, which will be analyzed next.

Spatial distribution of skilled labor. The re-distributed college seats across provinces further affect students' post-college locational choices. Table 4 reports the change in the spatial distribution of college graduates who choose to enter the labor market and their initial earnings.[38] It is important to note that local labor demand is treated as exogenous in the counterfactual and does not respond to changes in the supply of college graduates, so the wage level for college graduates in each location is fixed (Kennan and Walker, 2011). The result should be considered as capturing the short-run effect if one or two cohorts of college students are affected by counterfactual admission, which is a small change in labor supply compared to the entire existing workforce. The assumption of exogenous skill prices is also consistent with recent evidence that the return to college education does not decrease with the local skill supply due to the endogenous adoption of skill-biased technology by firms (e.g. Feng and Xia, 2022).

I first calculate the number of college graduates working in each region, regardless of home location, divided by the number of local college applicants in the same cohort in columns (1)-(3) of Table 4. This statistic is more directly relevant to local economic development than the college admission rate, as students who grow up in a place may migrate to another region after college. Admission based on merit exacerbates the already significant inequality in the geographical distribution of skilled labor, leading to a higher concentration of skilled workers in developed areas. The fact that the number of college workers per capita will decrease by nearly 20% in the least devel-

[38] The capacity of graduate school is flexible in simulation. The change in the share of college graduates choosing graduate study is less than 1% in both the [Merit] and the [Equal] counterfactuals.

oped region under [Merit] is particularly alarming. This is the result of student's dynamic location choices (Figure A9 and A10) after eliminating the migration arrangement from home to the college location imposed by admission quotas. Under [Merit], a larger number of students from less developed regions pursue their higher education in the top quartile provinces, and many of them stay there after college as these provinces are more attractive and graduating from a college there could help them secure a career in the same region. In contrast, majority of students originating from the developed region tend to return to their hometown after college in all admission schemes. Consequently, brain drain is more pronounced in less developed provinces.

Despite the distributional change, worker quality generally increases in all regions, benefiting from more efficient human capital production during college under improved sorting. Columns (4)-(6) in Table 4 compute the sum of the first year's earnings of all new college graduates working in each region, which is used as a proxy for total output and efficiency. Combining the higher return to human capital in the developed region with a more concentrated supply of high quality labor, at the national level, [Merit] admission increases the total earnings of the entire cohort by 1.3%, which is a substantial gain in efficiency considering that China's current annual GDP growth rate is around 5-6%. At the same time, the average welfare of the student increases by 3.5%, resulting from a combination of increased lifetime earnings and higher average amenities as more college graduates live and work in more attractive locations. Lastly, it is interesting that the [Equal] scheme also slightly increases brain drain and total earnings compared to the baseline, despite equal access to colleges. Again, this comes from the fact that [Equal] does not limit the flow of students, whereas a spatial quota fixes both the origin (student's home province) and destination (college location). This somewhat surprising finding prompts an exploration of the government's objective behind implementing the quota system, which is discussed in the next subsection.

It is worth noting that the counterfactual analysis here is conducted in a partial equilibrium framework and the notion of efficiency is mainly measured by total initial earnings. How the change in college seat allocation and the re-distribution of college graduates affect equilibrium wage levels (Heckman et al., 1998) in the presence of productivity externalities (Moretti, 2004) is left for future research in a general equilibrium framework.

Together, these results point to a policy tradeoff between efficiency in aggregate output, equality in college opportunities for students born in different places, and equality in supply of skilled labor across regions, which may further affect future development and inequality.

7.3 Efficiency-equality frontier

To summarize the efficiency-equality tradeoff analyzed above, I plot the baseline, the two counterfactuals, as well as ten equally-spaced college seat allocations between [Merit] and [Equal] in the space of efficiency and equality measures in Figure 5.[39] Efficiency on the vertical axis is measured by the total initial earnings. I get very similar results when measuring efficiency with student welfare. Two equality measures on the horizontal axis are the provincial gap of college admission rate in the left panel and the provincial gap of skilled worker per capita in the right.

In Figure 5 panel (a), if we only look at the tradeoff between output and equality in college access, the current quota system seems to be a clear inferior policy. I use the term "frontier" in a loose or conservative way, as there are many more feasible allocations of college seats, and some of them could potentially be located upper right to the curve connecting [Merit], [Equal], and the points in between. Nevertheless, the baseline is obviously off the potential policy frontier to a significant extent. However, panel (b) provides a different story if the policy goal in terms of equality is to reduce the gap in supply of skilled labor in the local economy and thus the brain drain. The admission quota system used by China appears to trade output for a more spatially equalized supply of college graduates, which may be justified by the need to address the substantial regional inequality in development amidst the remarkable overall economic growth (Kanbur and Zhang, 2005; Xu, 2011). This objective is accomplished by precisely assigning college seats located in each area to particular student home locations as specified. Figure A11 illustrates that the quota in the baseline would be ineffective without the influence of the college location on post-graduation migrations.

Motivated by the potential policy preference suggested by the above evidence, the last exercise attempts to infer the policy weights placed by China in designing the college quota system. Taking the output value as the benchmark, I convert the value of reducing the gap of the provincial admission rate by one percentage point to the percentage equivalence of output growth, and similarly for the value of reducing the gap of the share of skilled labor by one percentage point. Figure 6 highlights the conditions in which the baseline quota system is the preferred policy. A one percentage point decrease in the disparity of the share of college graduates working locally is approximately equivalent to at least 0.15% of the aggregate output when the government's empha-

[39] To generate these ten in-between allocations, for each province, I calculate ten middle points between the admission rate realized under [Merit] and the equal admission rate under [Equal], then impose them as the cap of local admission rate for the corresponding province, one at a time in totally ten counterfactuals.

sis on equal access to college is modest. In the long run, whether retaining highly skilled workers in underdeveloped areas for future local development is more favorable than alternative policies, such as supplementing merit-based college seat allocation with interregional transfer payments, is beyond the scope of this paper but worth future research.

8 Concluding remarks

How to allocate college seats across regions is an important yet largely neglected issue. To tackle this question, this paper builds and estimates a structural model that combines college choice with post-college residential location choice. With the estimated complementarity between student and college quality and the causal effect of college location on subsequent migration, I highlight a policy tradeoff in the spatial allocation of college seats between efficiency in aggregate human capital output, equality of opportunities for people growing up in different places, and long-run regional development benefiting from the supply of skilled labor. The analysis in this paper is based on Chinese data and setting, which itself has the largest college market and the second largest economy, but the question is not unique to a particular country. The general environment associated with this research, such as uneven spatial distribution of colleges, unequal quality of the local education system prior to college, the technology of human capital production in higher education, as well as the student preference in location, are common.

The findings of this paper suggest that the current policy objective in China's higher education is a mix of efficiency and equality. The government cares about inequality of college access of different groups across regions, and places greater policy weight toward a more equalized spatial supply of skilled labor. Not going full meritocracy in college seat allocation is seen in policy practices in other places. In the US, public universities in California have recently moved to test-blind admissions (Hendricks et al., 2022). This paper provides a new and more comprehensive perspective on efficiency versus multidimensional regional inequality for these policy practices that aim to address systematic inequality. Future research should examine the long-run effect of the distributional change of college students due to college admission changes and the spatial sorting in the general equilibrium. This places the reform of the college admission system in a larger policy space and is potentially beneficial for a more coordinated design of higher education and spatial policies (Fajgelbaum and Gaubert, 2020; Blouri and Ehrlich, 2020).

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Adamopoulos, Tasso, Loren Brandt, Jessica Leight, and Diego Restuccia, "Misallocation, Selection, and Productivity: A Quantitative Analysis With Panel Data From China," *Econometrica*, 2022, *90* (3), 1261–1282.

Agarwal, Nikhil and Paulo Somaini, "Revealed Preference Analysis of School Choice Models," *Annual Review of Economics*, 2020, *12*, 471–501.

Anstreicher, Garrett, "Spatial Influences in Upward Mobility," *Journal of Political Economy*, 2024, *132* (1), 149–199.

Arcidiacono, Peter, "Affirmative Action in Higher Education: How do Admission and Financial Aid Rules Affect Future Earnings?," *Econometrica*, 2005, *73* (5), 1477–1524.

_ **and Michael Lovenheim**, "Affirmative Action and the Quality-Fit Trade-off," *Journal of Economic Literature*, 2016, *54* (1), 3–51.

Bai, Chong-En, Ruixue Jia, Hongbin Li, and Xin Wang, "Entrepreneurial Reluctance: Talent and Firm Creation in China," *NBER Working Paper Series*, 2021, *28865*, 1–49.

Baird, Matthew D., John Engberg, and Isaac M. Opper, "Optimal Allocation of Seats in the Presence of Peer Effects: Evidence from a Job Training Program," *Journal of Labor Economics*, 2023, *41* (2), 479–509.

Barrow, Lisa, Lauren Sartain, and Marisa de la Torre, "Increasing Access to Selective High Schools through Place-Based Affirmative Action: Unintended Consequences," *American Economic Journal: Applied Economics*, 2020, *12* (4), 135–163.

Bleemer, Zachary, "Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209," *The Quarterly Journal of Economics*, 2022, *137* (1), 115–160.

Blouri, Yashar and Maximilian V. Ehrlich, "On the optimal design of place-based policies: A structural evaluation of EU regional transfers," *Journal of International Economics*, 2020, *125*, 103319.

Bo, Shiyu, Jing Liu, Ji-Liang Shiu, Yan Song, and Sen Zhou, "Admission Mechanisms and the Mismatch between Colleges and Students: Evidence from a Large Administrative Dataset from China," *Economics of Education Review*, 2019, *68*, 27–37.

Brotherhood, Luiz, Bernard Herskovic, and João Ramos, "Income-Based Affirmative Action in College Admissions," *The Economic Journal*, 2023, *133* (653), 1810–1845.

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the Incidence and Efficiency of a Prominent Place Based Policy," *American Economic Review*, 2013, *103* (2), 897–947.

Calsamiglia, Caterina, Chao Fu, and Maia Guell, "Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives," *Journal of Political Economy*, 2020, *128* (2), 642–680.

Carr, Michael D. and Emily E. Wiemers, "The decline in long-term earnings mobility in the U.S.: Evidence from survey-linked administrative data," *Labour Economics*, 2022, *78*, 102170.

Chan, Jimmy, Zhi Wang, and Shunyi Zhao, "The Institution of Merit: A Chinese Case Study," *Working paper*, 2019.

Chen, Yan and Onur Kesten, "Chinese College Admissions and School Choice Reforms: A Theoretical

Analysis," *Journal of Political Economy*, 2017, *125* (1), 99–139.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Income Segregation and Intergenerational Mobility Across Colleges in the United States," *The Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

__ , Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Clark, Derek J. and Christian Riis, "Influence and the Discretionary Allocation of Several Prizes," *European Journal of Political Economy*, 1998, *14* (4), 605–625.

Colas, Mark and Suqin Ge, "Transformations in China's Internal Labor Migration and Hukou System," *Journal of Labor Research*, 2019, *40* (3), 296–331.

Cunha, Flavio and James J. Heckman, "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 2008, *43* (4), 738–782.

__ , __ , and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

__ and Cecile Gaubert, "Spatial Sorting and Inequality," *Annual Review of Economics*, 2022, *14* (1), 795–819.

Diao, Chenglei, Chenyuan Liu, and Xiaohan Zhong, "Improving Equality in China's College Admissions: Estimation and Policy Interventions," *Working paper*, 2022.

Dillon, Eleanor W. and Jeffrey A. Smith, "Determinants of the Match between Student Ability and College Quality," *Journal of Labor Economics*, 2017, *35* (1), 45–66.

__ and __ , "The Consequences of Academic Match between Students and Colleges," *Journal of Human Resources*, 2020, *55* (3), 767–808.

Fabre, Anaıs, "The Geography of Higher Education and Spatial Inequalities," *Working paper*, 2023.

Fack, Gabrielle, Julien Grenet, and Yinghua He, "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, 2019, *109* (4), 1486–1529.

Fajgelbaum, Pablo D. and Cecile Gaubert, "Optimal Spatial Policies, Geography, and Sorting," *The Quarterly Journal of Economics*, 2020, *135* (2), 959–1036.

Fang, Hanming and Xincheng Qiu, "“Golden Ages”: A Tale of the Labor Markets in China and the United States," *Journal of Political Economy: Macroeconomics*, 2023, *1* (4), 665–706.

Feng, Shuaizhang and Xiaoyu Xia, "Heterogeneous firm responses to increases in high-skilled workers: Evidence from China's college enrollment expansion," *China Economic Review*, 2022, *73*, 101791.

Fleisher, Belton M., Haizheng Li, and Min Qiang Zhao, "Human Capital, Economic Growth, and Regional Inequality in China," *Journal of Development Economics*, 2010, *92* (2), 215–231.

Fu, Chao, Junjie Guo, Adam J. Smith, and Alan Sorensen, "Students' heterogeneous preferences and the uneven spatial distribution of colleges," *Journal of Monetary Economics*, 2022, *129*, 49–64.

Gaubert, Cecile, Patrick Kline, and Danny Yagan, "Place-Based Redistribution," *NBER Working Paper*

Series, 2021, *w28337.*

Ge, Teng and Tao Wu, "Search, Migration, and Social Connections: Solving the Migration Puzzle to Beijing," *China Economic Review*, 2020, *62*, 1–11.

Glaeser, Edward L. and Joshua Gottlieb, "The Economics of Place-Making Policies," *Brookings Papers on Economic Activity*, 2008, *39*, 155–253.

Gourieroux, C., A. Monfort, and E. Renault, "Indirect inference," *Journal of Applied Econometrics*, 1993, *8* (S1), S85–S118.

Groen, Jeffrey A., "The Effect of College Location on Migration of College-Educated Labor," *Journal of Econometrics*, 2004, *121*, 125–142.

Guo, Lina, Prashant Loyalka, and Xiaoyang Ye, "Are Regional Quotas Fair? Simulating Merit-Based College Admissions Using Unique Student-Level Data from China," *Working paper*, 2018.

Heckman, James J., Lance Lochner, and Christopher Taber, "General-Equilibrium Treatment Effects: A Study of Tuition Policy," *American Economic Review*, 1998, *88* (2), 381–386.

Hendricks, Lutz, Tatyana Koreshkova, and Oksana Leukhina, "Causes and Consequences of Student-College Mismatch," *Federal Reserve Bank of St. Louis Working Paper*, 2022, *026.*

Hillman, Nicholas W., "Geography of College Opportunity: The Case of Education Deserts," *American Educational Research Journal*, 2016, *53* (4), 987–1021.

Howard, Greg and Hansen Shao, "The Dynamics of Internal Migration: A New Fact and its Implications," *Working paper*, 2023.

Hsieh, Chang-Tai and Peter J Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

__, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Huang, Mian, Chunbing Xing, and Xiaoyong Cui, "Does College Location Affect the Location Choice of New College Graduates? Evidence from China," *China & World Economy*, 2022, *30* (3), 135–160.

Ishimaru, Shoya, "Geographic Mobility of Youth and Spatial Gaps in Local College and Labor Market Opportunities," *Working paper*, 2023.

Jia, Ruixue and Hongbin Li, "Just Above the Exam Cutoff Score: Elite College Admission and Wages in China," *Journal of Public Economics*, 2021, *196*, 104371.

Kanbur, Ravi and Xiaobo Zhang, "Fifty Years of Regional Inequality in China: A Journey through Central Planning, Reform, and Openness," *Review of Development Economics*, 2005, *9* (1), 87–106.

Kang, Ning, *Institutional Innovation in Higher Education Resource Distribution in a Transitional Economy*, Educational Science Publishing House, 2005.

Kennan, John, "Spatial Variation in Higher Education Financing and the Supply of College Graduates," *Working paper*, 2021.

__ and James R. Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kline, Patrick and Enrico Moretti, "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," *Quarterly Journal of Economics*, 2014, *129* (1), 275–331.

Knight, Brian and Nathan Schiff, "The Out-of-State Tuition Distortion," *American Economic Journal: Economic Policy*, 2019, *11*, 317–350.

Lu, Yi, Jin Wang, and Lianming Zhu, "Place-Based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program," *American Economic Journal: Economic Policy*, 2019, *11* (3), 325–360.

Luflade, Margaux, "The Value of Information in Centralized School Choice Systems," *Working paper*, 2017.

McFadden, Daniel D, "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka, ed., *Frontiers in Econometrics*, Academic Press, 1974, pp. 105–142.

Mello, Ursula, "Centralized Admissions, Affirmative Action, and Access of Low-Income Students to Higher Education," *American Economic Journal: Economic Policy*, 2022, *14* (3), 166–197.

Ministry of Education, "Regular Institutions of Higher Education," *http://www.moe.gov.cn/srcsite/A03/moe_634/ 201104/t20110427_119480.html*, 2011.

Moretti, Enrico, "Human Capital Externalities in Cities," *Handbook of Regional and Urban Economics*, 2004, *4*, 2243–2291.

Pan, Kunfeng and Sheng Cui, "Local Government Competition on Human Capital and the Enrollment Strategy of Provincial Universities," *China Economics of Education*, 2017, *2* (6), 31–50.

Pu, Yun, "College Admission in Three Chinese Provinces: Province-Specific versus Pooling Quotas," *China Economic Review*, 2020, *60*, 1–9.

Rust, John P., "Structural Estimation of Markov Decision Processes," in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, Amsterdam, New York, Oxford: Elsevier Science, North-Holland, 1994, pp. 3081–3143.

Si, Ruichao, "Unexpected pressure? The effect of a curriculum reform on household educational investments and student outcomes," *Journal of Economic Behavior & Organization*, 2021, *190*, 730–746.

Sjoquist, David L. and John V. Winters, "Merit Aid and Post-College Retention in the State," *Journal of Urban Economics*, 2014, *80*, 39–40.

Song, Yang, "What should economists know about the current Chinese hukou system?," *China Economic Review*, 2014, *29*, 200–212.

Tombe, By Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, 2019, *109* (5), 1843–1872.

Wei, Shang Jin, Xiaobo Zhang, and Yin Liu, "Home Ownership as Status Competition: Some Theory and Evidence," *Journal of Development Economics*, 2017, *127*, 169–186.

Xu, Chenggang, "The Fundamental Institutions of China's Reforms and Development," *Journal of Economic Literature*, 2011, *49* (4), 1076–1151.

Ye, Xiaoyang, "Improving College Choice for the Poorest Students: Results from Randomized Experiments in Centralized Admissions," *Education Finance and Policy*, 2023, pp. 1–33.

Zhang, Jipeng, Ru Wang, and Chong Lu, "A Quantitative Analysis of Hukou Reform in Chinese Cities: 2000-2016," *Growth and Change*, 2019, *50* (1), 201–221.

Figure and tables

(a) θ^0 vs. GDP per capita

(b) θ^0 vs. pre-college education expenditure

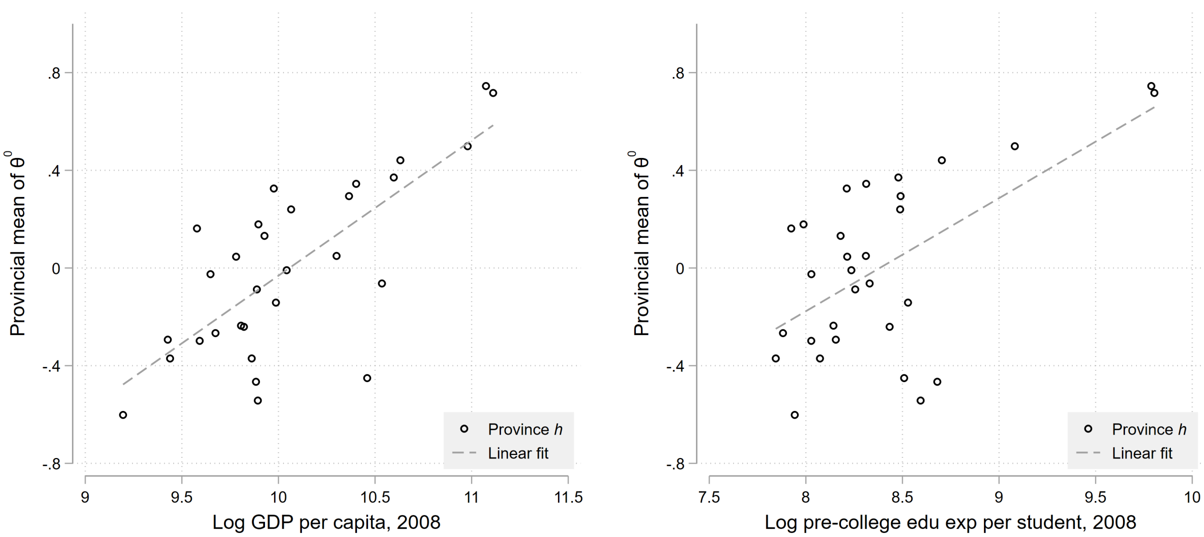


Figure 1: Pre-college human capital θ^0 vs. GDP and pre-college education expenditure

Notes: Provincial GDP per capita in 2008 are from the National Bureau of Statistics of China. Per student pre-college education expenditure data in 2008 are from the Ministry of Education of China and are defined as the per student total expenditure averaged over primary, middle and high school in each province.

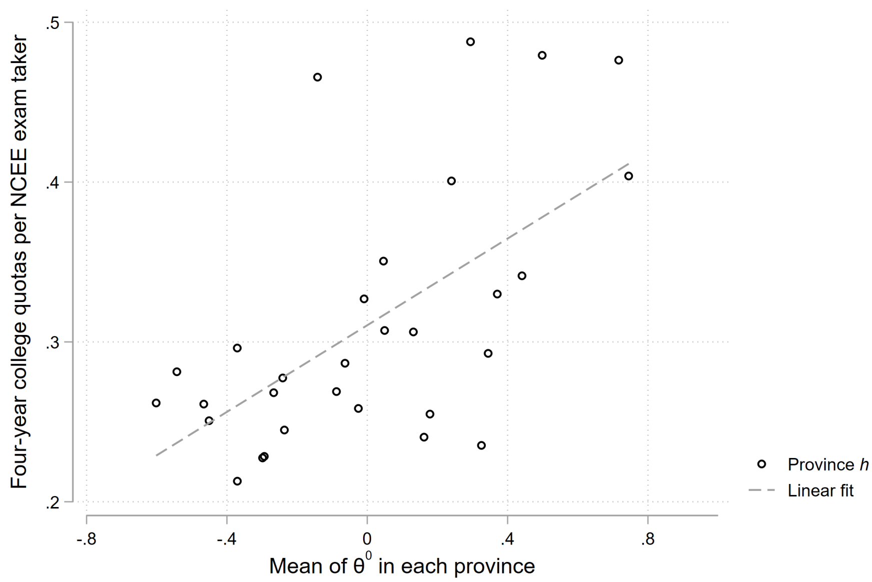


Figure 2: Mean of θ^0 vs. the average number of four-year college quotas per NCEE exam taker, 2006-2011

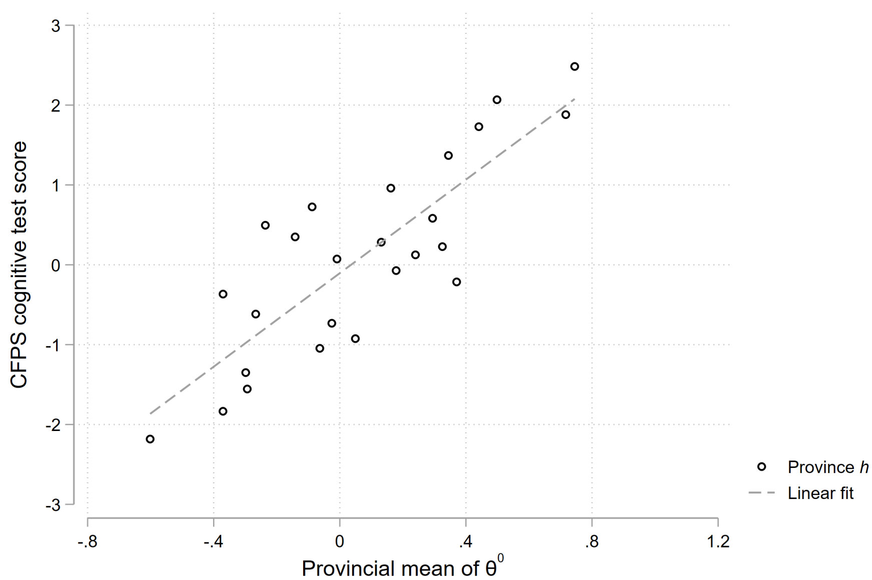


Figure 3: Validity check: $E(\theta^0|h)$ vs. cognitive test scores in the CFPS

Notes: The cognitive test scores in the CFPS data are constructed by combining the standardized cognitive tests on words, math, and number series conducted during the 2010 and 2012 rounds for the 25 surveyed provinces. I use the first principal component of the age-adjusted scores of the three cognitive tests.

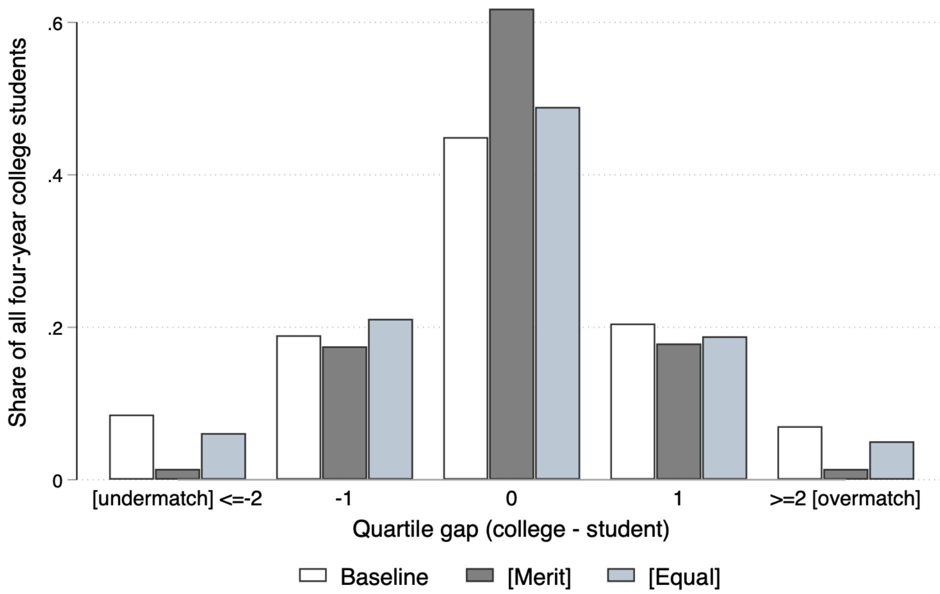


Figure 4: Sorting and mismatch between students and colleges by quality

Notes: Students and colleges are divided into four quartiles based on pre-college human capital θ^0 and the college quality measure V_s. "Baseline" refers to the existing quota allocations observed in 2008. [Merit] is the counterfactual of nationwide purely merit-based admission without provincial quotas. [Equal] is the counterfactual of fully equalized college seat allocation such that for each tier of colleges, the admission rate is the same for all provinces.

(a) Total output vs. equality of college opportunities

(b) Total output vs. distribution of skilled labor

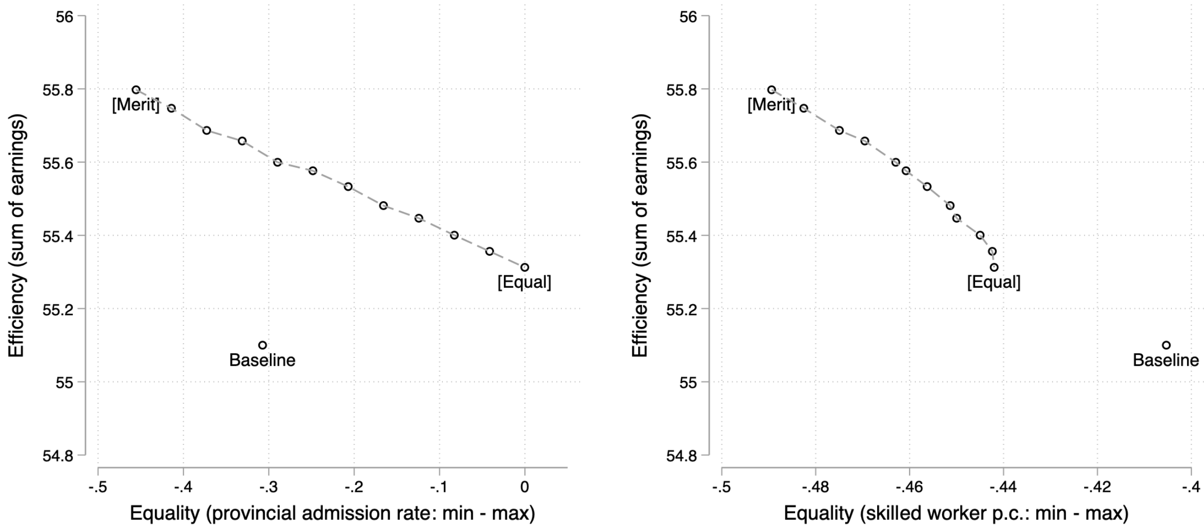


Figure 5: Efficiency-equality frontier

Notes: The baseline, the two counterfactuals, as well as ten equally-spaced college seat allocations between [Merit] and [Equal] are plotted in the space of efficiency and equality measures according to the values in each corresponding simulation.

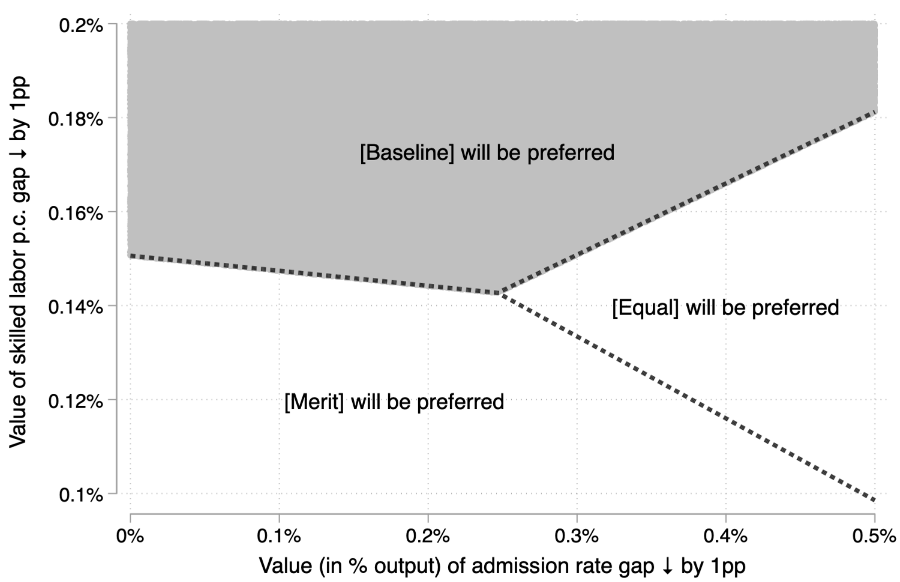


Figure 6: Policy weights for output, college opportunity, and skilled labor distribution

Notes: This figure assumes a policy objective function of a combination of aggregate output (sum of initial earnings), the gap of provincial admission rate, and the gap of skilled labor per local college applicant. The policy weight for output is normalized to one, and the weights for the other two components are placed on the horizontal and vertical axes.

Table 1: Data pattern of migration at college and work stages

<table>
<tr><th>Attend college in</th><th></th><th colspan="3">Work in</th></tr>
<tr><th></th><th></th><th>Home province</th><th>College province</th><th>Other province</th></tr>
<tr><td>Home province</td><td>69.1%</td><td colspan="2">71.9%</td><td>28.1%</td></tr>
<tr><td>Non-home province</td><td>30.9%</td><td>28.2%</td><td>32.4%</td><td>39.4%</td></tr>
</table>

Notes: Administrative data on college admission between 2006 and 2011 and the Chinese College Student Survey (CCSS) covering the same cohorts. See Subsection 5.1 for more data details.

Table 2: Location preferences in monetary value

Parameter	Annual value (% of consumption)		
College at home province		1.89%	
Distance[home, college]	-0.05%	-0.09%	-0.12%
	(500 km)	(1,000 km)	(1,500km)
Work at home province		11.17%	
Distance[work, home]	-3.85%	-6.76%	-8.71%
	(500 km)	(1,000 km)	(1,500km)
Work at college location		10.88%	
Distance[work, college]	-2.73%	-5.21%	-7.44%
	(500 km)	(1,000 km)	(1,500km)
Idiosyncratic location preference		4.09%	

Notes: The estimated payoff parameters are converted to monetary values added to the smoothed annual consumption. The average annual consumption (smoothed over the life cycle) among individuals is 115,113 CNY.

Table 3: Changes in admission rate and quality of admitted students

	College admission rate			Average θ^0 of admitted		
Home province quartile (by GDP p.c.)	Baseline (1)	[Merit] (2)	[Equal] (3)	Baseline (4)	[Merit] (5)	[Equal] (6)
Q1 (top)	0.336	0.358	0.277	1.370	1.338	1.456
Q2	0.301	0.330	0.277	1.265	1.272	1.361
Q3	0.237	0.221	0.277	1.140	1.218	1.044
Q4 (bottom)	0.245	0.210	0.277	1.038	1.167	0.970
Total	0.277	0.277	0.277	1.213	1.258	1.200

Notes: The college admission rate is defined as the number of students admitted per NCEE exam taker under each admission policy. "Baseline" refers to the existing quota allocations observed in 2008. [Merit] is the counterfactual of nationwide purely merit-based admission without provincial quotas. [Equal] is the counterfactual of fully equalized college seat allocation such that for each tier of colleges, the admission rate is the same for all provinces.

Table 4: Distribution of college graduates and labor market outputs

	# skilled workers per capita$^{(a)}$			Total initial earnings (B¥)$^{(b)}$		
Work province quartile (by GDP p.c.)	Baseline (1)	[Merit] (2)	[Equal] (3)	Baseline (4)	[Merit] (5)	[Equal] (6)
Q1 (top)	0.534	0.593	0.561	29.39	32.82	31.07
Q2	0.170	0.162	0.149	10.95	10.65	9.89
Q3	0.130	0.109	0.127	7.56	6.42	7.53
Q4 (bottom)	0.128	0.103	0.119	7.20	5.90	6.82
Total				55.10	55.80 (↑ 1.3%)	55.31 (↑ 0.4%)

Notes: (a) The number of college graduates per capita is defined as the number of college graduates working in a region divided by the number of NCEE exam takers in the same cohort. (b) Earnings are measured in 2012 CNY value. 2012 is the year when the cohort admitted in 2008 graduate and enter the labor market.

Appendices

Appendix A Additional figures and tables

A.1 Data and measurement

(a) 2006-2011 (period analyzed in this paper) (b) 2006-2011 vs. 2022

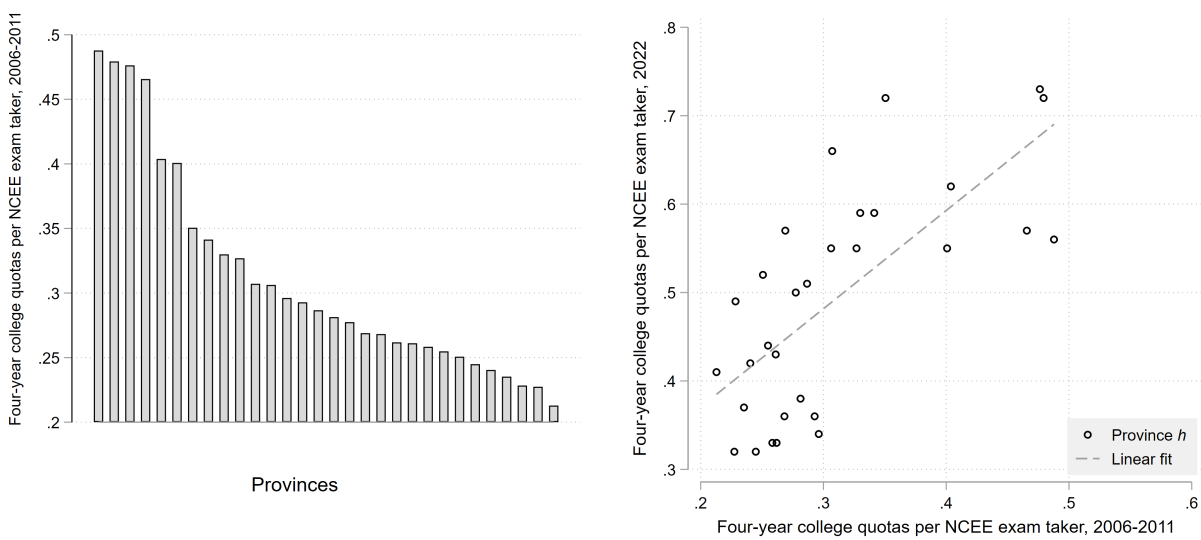
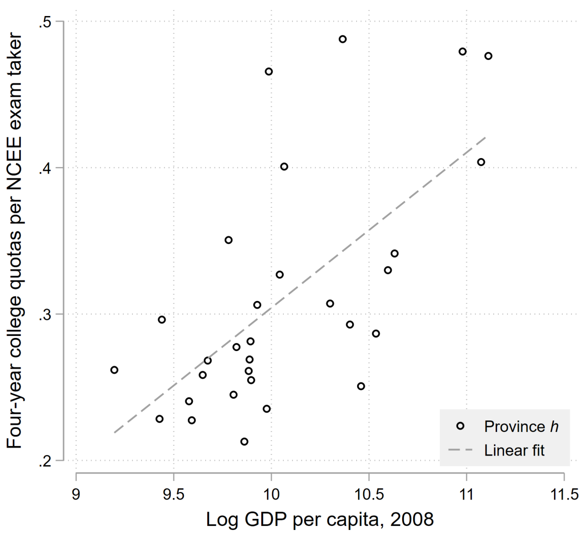


Figure A1: Four-year college admission quotas per NCEE exam taker by province

Notes: Each bar represents the total number of four-year college admission quotas received by each province divided by the total number of NCEE exam takers in that province. The Panel (a) is averaged across years between 2006 and 2011, and Panel (b) compares the period of 2006-2011 and the recent quota allocations in 2022.

(a) Four-year college admission quotas vs. GDP per capita



(b) Four-year college capacity vs. GDP per capita

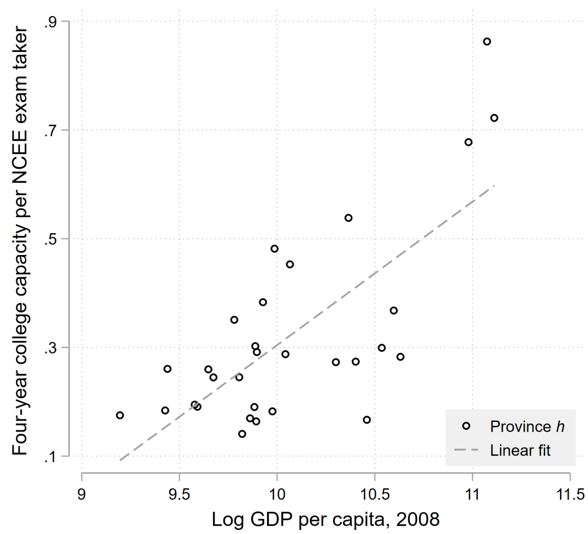


Figure A2: Four-year college admission quotas and capacity vs. GDP per capita

Notes: Provincial GDP per capita series in 2008 are from the National Bureau of Statistics of China. College capacity per capita is defined as the total number of new college seats in all four-year colleges located in a province divided by the total number of NCEE exam taker in that province.

Table A1: Principle component analysis of college quality

Raw measure	Loading factor
Student-faculty ratio	-0.147
Value of teaching/research equipment, per student	0.417
Number of library books, per student	0.083
Public grant for undergraduate teaching, per student	0.275
Teaching expenditure - regular, per student	0.267
Teaching expenditure - experiment, per student	0.290
Teaching expenditure - internship, per student	0.259
Share of courses taught by full professor, per student	0.050
Employment rate	0.149
College tier (1=top, 2=middle, 3=bottom)	-0.411
Project-985 college	0.321
Project-211 college	0.375
Private college	-0.249

Table A2: Top-20 colleges ranked by the estimated college quality V_s

Ranking by V_s	College name	V_s
1	Tsinghua University	3.090
2	Peking University	2.911
3	Shanghai Jiaotong University	2.794
4	Fudan University	2.706
5	Sun Yat-sen University	2.575
6	Zhejiang University	2.478
7	Renmin University of China	2.437
8	Wuhan University	2.399
9	Nanjing University	2.365
10	Beihang University	2.333
11	Xiamen University	2.304
12	Beijing Normal University	2.276
13	Tianjin University	2.250
14	Tongji University	2.225
15	Huazhong University of Science And Technology	2.202
16	Xi'an Jiaotong University	2.180
17	Harbin Institute of Technology	2.159
18	Shandong University	2.139
19	University of Science and Technology of China	2.119
20	East China Normal University	2.101

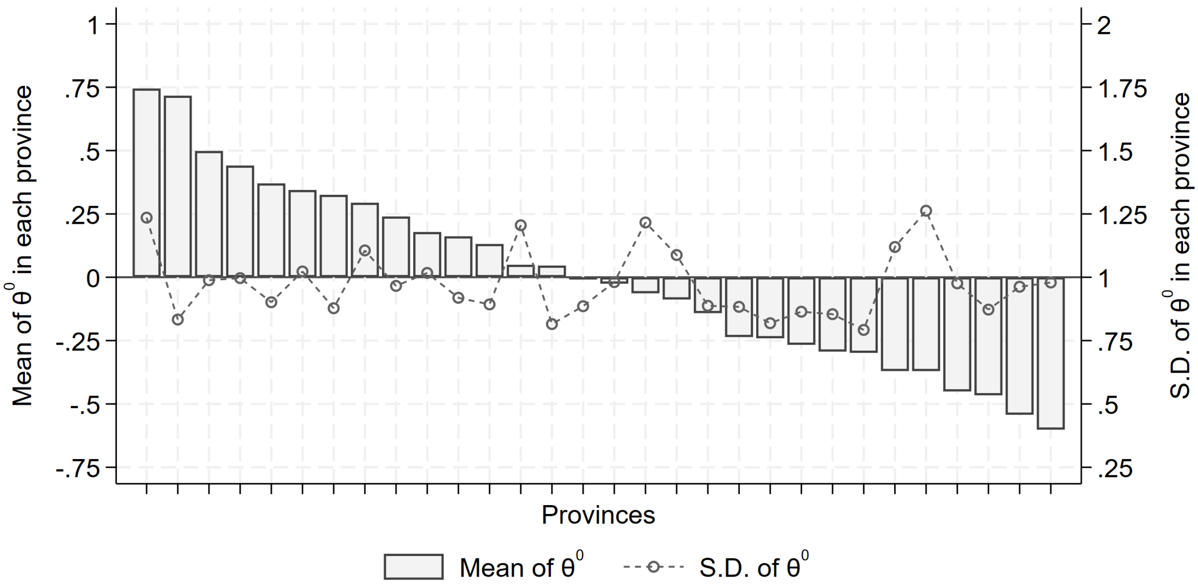


Figure A3: Mean and standard deviation of θ^0 of NCEE exam takers by province

Notes: The pre-college human capital θ^0 of NCEE exam takers in each province, $F_h(\theta^0)$, is assumed to follow a normal distribution. Mean and standard deviation of θ^0 in each province are estimated using the measurement model and standardized to set the national mean to zero and national standard deviation to 1.

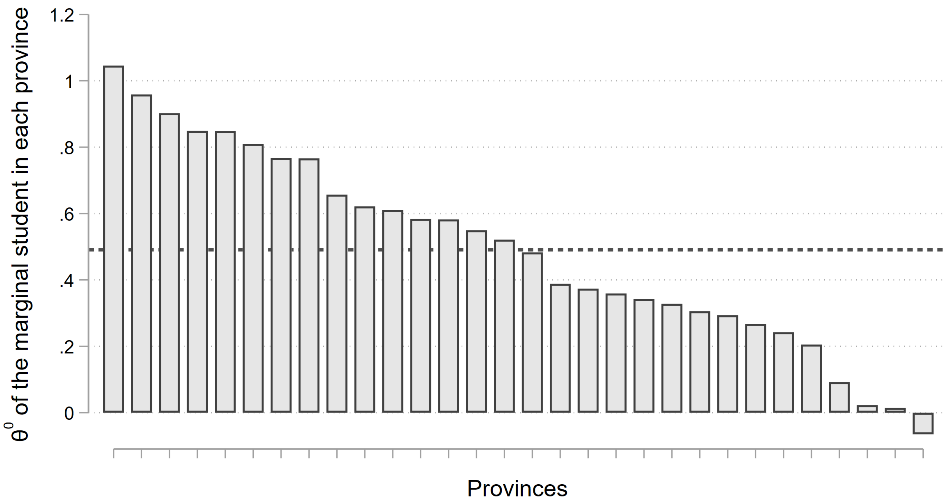


Figure A4: θ^0 of the marginally admitted students under college admission quotas in each province

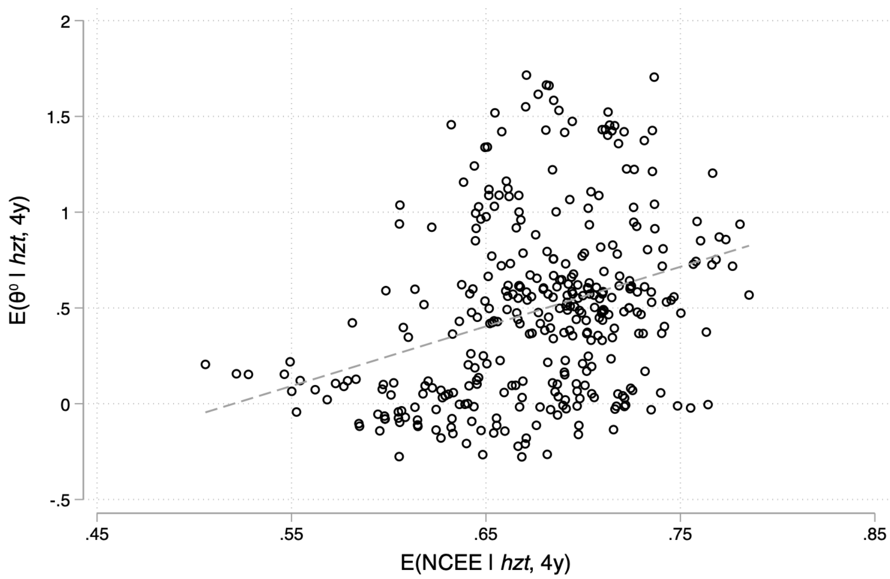


Figure A5: Raw NCEE scores vs. the comparable human capital θ^0 for admitted students

Notes: Each dot represents the students in hzt admitted by any four-year college $s \in \mathcal{S}$. Raw NCEE scores are rescaled to be between 0 and 1 by dividing the corresponding full score. The expected θ^0 for students in hzt, conditional on being admitted by a four-year college, is calculated by integrating over the distribution of the NCEE measurement error ν_1 and incorporating the admission cutoff.

A.2 Model estimation results

Table A3: Parameter estimates

Parameter		Estimate	Standard Error
College	College quality V_s	0.851	0.323
	Elite college: Project-985	0.859	0.043
	Elite college: Project-211	0.196	0.242
	Tier-1 college	3.984	0.218
	Tier-2 college	2.017	0.130
	STEM college	-0.224	0.129
	Non-STEM college	-0.804	0.144
	Econ college	0.155	0.022
	College in capital city of a province	0.273	0.113
	College at home province	4.644	0.036
	Distance[home, college] (in 1,000 km)	-0.242	0.047
	Distance[home, college] squared	0.027	0.103
Working	Work at home province	1.846	0.019
	Work at home province*home is Beijing/Shanghai	9.222	0.001
	Distance[work, home]	-1.579	0.011
	Distance[work, home] squared	0.329	0.019
	Work at college location	1.800	0.025
	Distance[work, college]	-1.028	0.026
	Distance[work, college] squared	0.089	0.063
Earnings	Constant	6.346	0.128
	Pre-college human capital θ^0	0.034	0.010
	College quality V_s	0.090	0.009
	Interaction $\theta^0 * V_s$	0.035	0.006
	Log provincial average income	0.346	0.012
Unobserved heterogeneity	Share of indifferent type ($\tau = \varnothing$)	0.581	0.000
	Size of location preference during college χ	0.988	0.004
	Size scalar for working periods ρ	1.642	0.000
Grad school	Pre-college human capital θ^0	-0.520	0.032
	College quality V_s	-0.950	0.073
	Interaction $\theta^0 * V_s$	1.468	0.166
	Elite college: Project-985	2.827	0.017
	Elite college: Project-211	4.143	0.055
	Tier-1 college	4.608	0.052
	Tier-2 college	7.084	0.032
	STEM college	0.586	0.028
	Non-STEM college	3.586	0.044
	Econ college	1.411	0.007
	College in capital city of a province	-1.432	0.042

Notes: Standard errors are calculated based on the variance-covariance matrix and the numerical derivatives following Gourieroux et al. (1993).

Table A4: Average amenity in monetary values by college and work region

Region (quartile by GDP p.c.)	Amenity as college location		Amenity as work location	
	Mean	Std. dev.	Mean	Std. dev.
Q1 (top)	-482	304	2,635	7,336
Q2	-476	215	-10,685	8,152
Q3	-1,037	594	-6,743	7,522
Q4 (bottom)	-778	250	-8,857	6,810

Notes: All monetary values are in 2012 CNY. The average amenity in each region is calculated by averaging the provincial fixed effects of college location and work location, and converted to annual monetary values added to smoothed consumption. The value of Beijing is normalized to 0. In 2012, the average annual smoothed consumption is 115,113 CNY.

A.3 Model fit

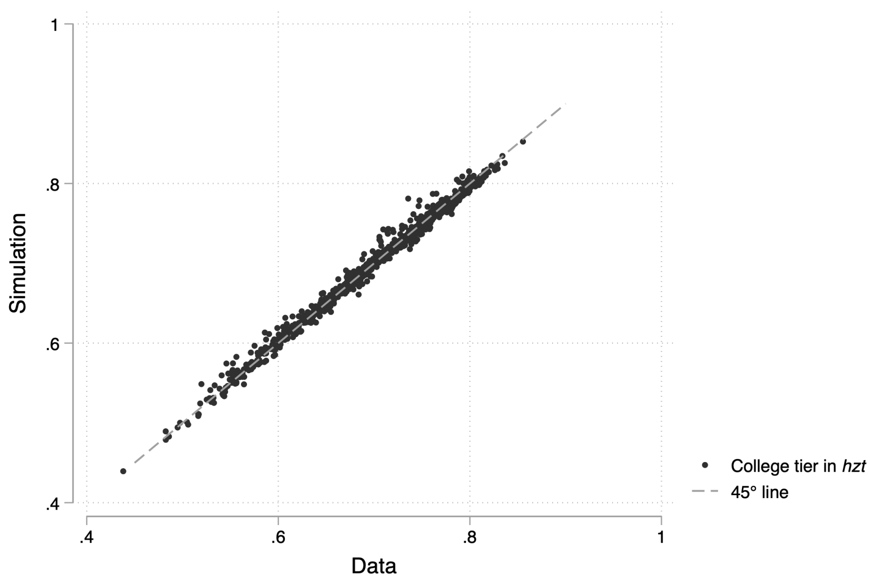


Figure A6: Model fit: average NCEE scores of admitted students

Notes: Each dot represents the average NCEE scores of students admitted by each tier of colleges from each province (h), track (z) and year (t) in administrative records and simulation. All raw NCEE scores are rescaled to be between 0 and 1.

Table A5: Model fit: share of college graduates going to graduate school

College tier	Data	Simulation
Tier-1 (top)	0.322	0.324
Tier-2	0.182	0.182
Tier-3 (bottom)	0.085	0.084

Notes: The share is calculated based on the CCSS-surveyed schools only, because the administrative data do not track students after college admission. The simulated share is calculated using all students who graduate from the CCSS-surveyed schools in the simulation. All numbers are averaged across cohorts (t).

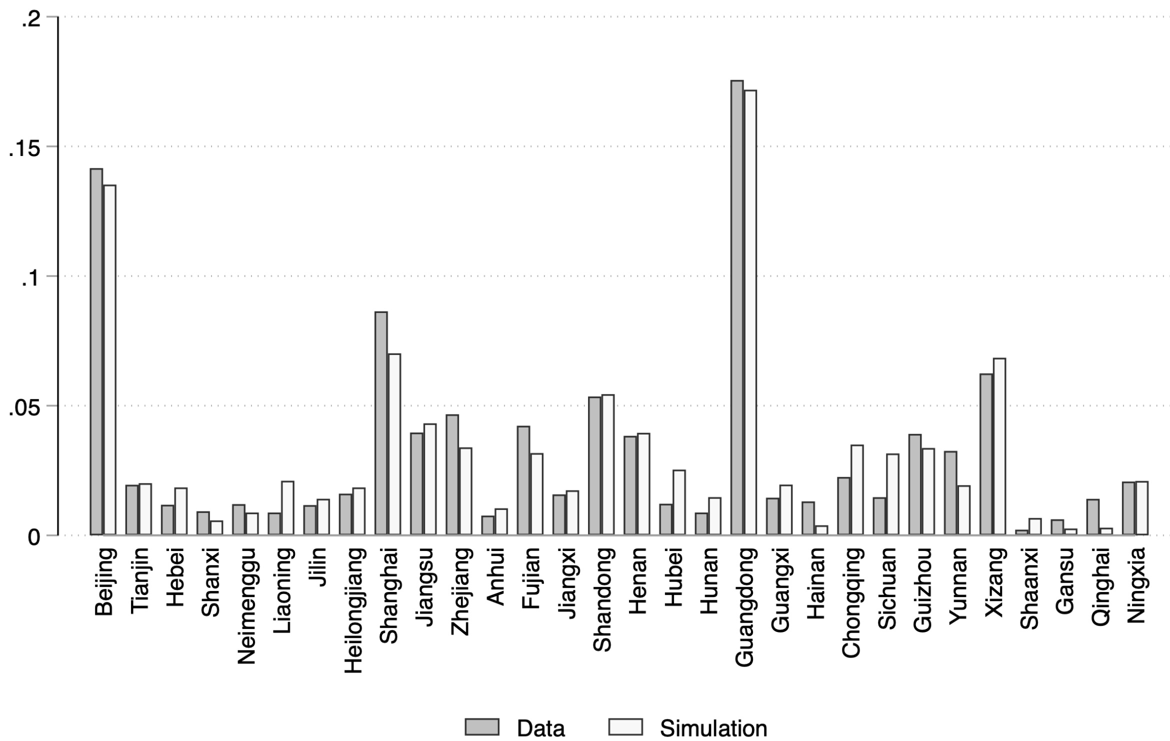


Figure A7: Model fit: provincial distribution of college graduates

Notes: The distribution is calculated based on the CCSS-surveyed schools only, because the administrative data do not track students after college admission. The simulated distribution is calculated using all students who graduate from the CCSS-surveyed schools in simulation. The distribution is calculated combining all cohorts (t).

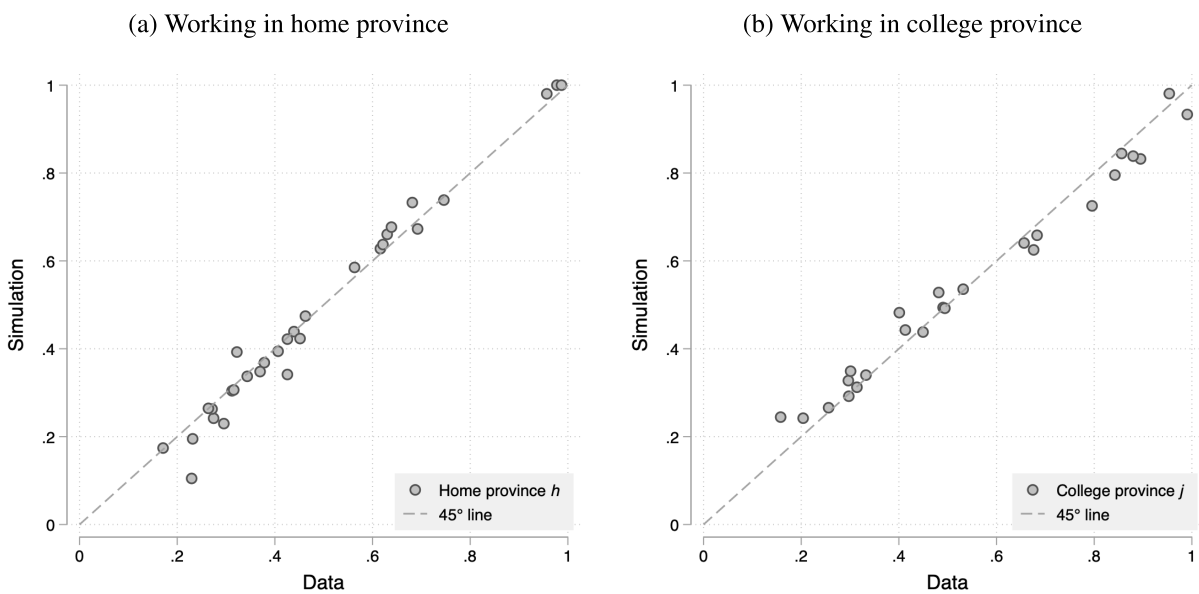


Figure A8: Model fit: share of college graduates working in home and college province

Notes: Same as Figure A7, the share in each sub-figure is calculated based on the CCSS-surveyed schools only. All numbers are averaged across cohorts (t).

A.4 Counterfactual

Table A6: College resources and student quality by region

Province quartile (by GDP p.c.)	Capacity per capita of local colleges (1)	Average quality (V_s) of local college (2)	Average quality (θ^0) of local NCEE exam takers (3)
Q1 (top)	0.373	0.425	0.201
Q2	0.311	-0.023	0.177
Q3	0.226	-0.148	-0.188
Q4 (bottom)	0.214	-0.151	-0.161

Notes: Provinces are grouped into quartiles according to GDP per capita in 2008. The capacity per capita of local colleges is defined as the number of admission seats of colleges located in the region per NCEE exam taker in the region. Both the national distribution of college quality V_s and student pre-college human capital θ^0 have mean zero and standard deviation one.

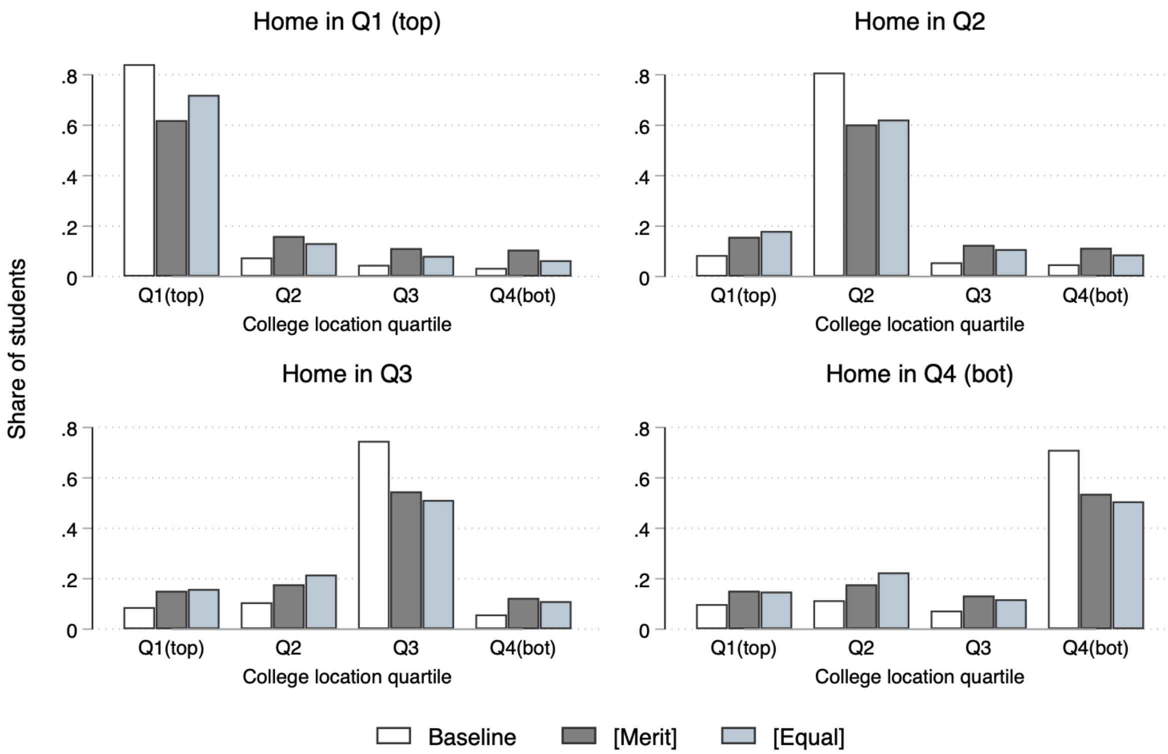


Figure A9: Counterfactual: College location choice by home region

Notes: Provinces are grouped into four quartiles based on GDP per capita in 2008. "Baseline" refers to the existing quota allocations observed in 2008. [Merit] is the counterfactual of nationwide purely merit-based admission without provincial quotas. [Equal] is the counterfactual of fully equalized college seat allocation such that for each tier of colleges, the admission rate is the same for all provinces.

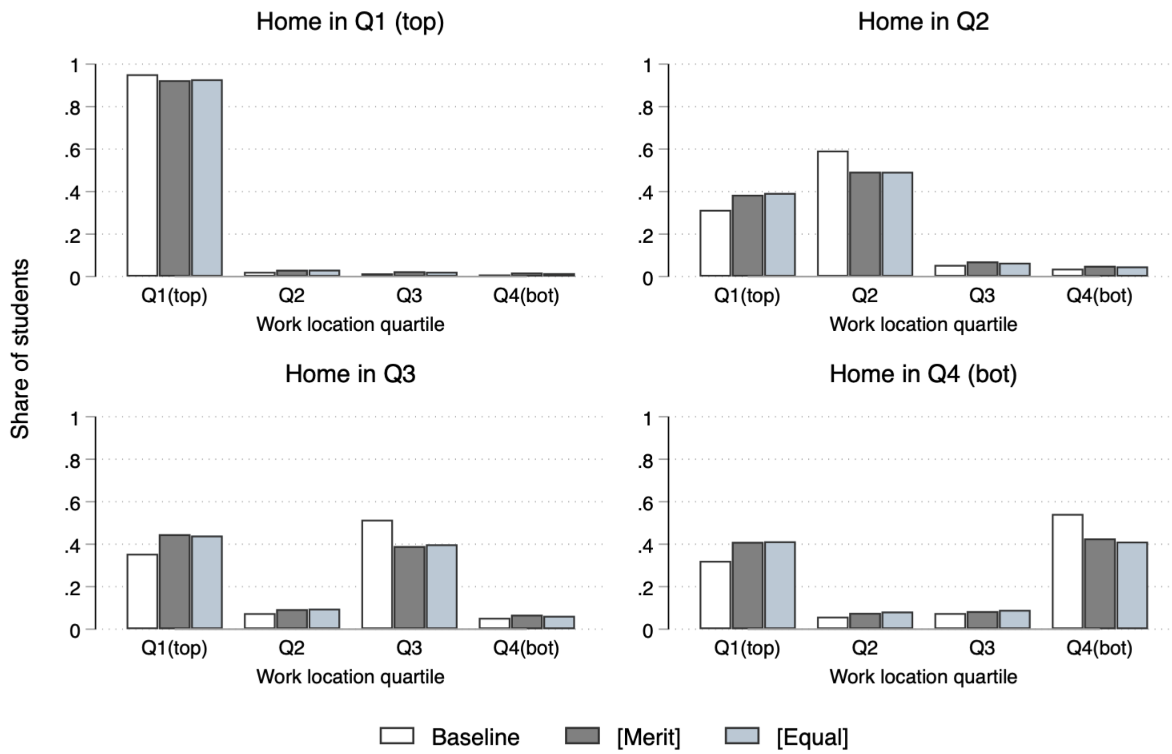


Figure A10: Counterfactual: Work location choice by home region

Notes: Provinces are grouped into four quartiles based on GDP per capita in 2008. "Baseline" refers to the existing quota allocations observed in 2008. [Merit] is the counterfactual of nationwide purely merit-based admission without provincial quotas. [Equal] is the counterfactual of fully equalized college seat allocation such that for each tier of colleges, the admission rate is the same for all provinces.

(a) With the college location effect (Figure 5 panel b)

(b) Removing the college location effect on work location choices

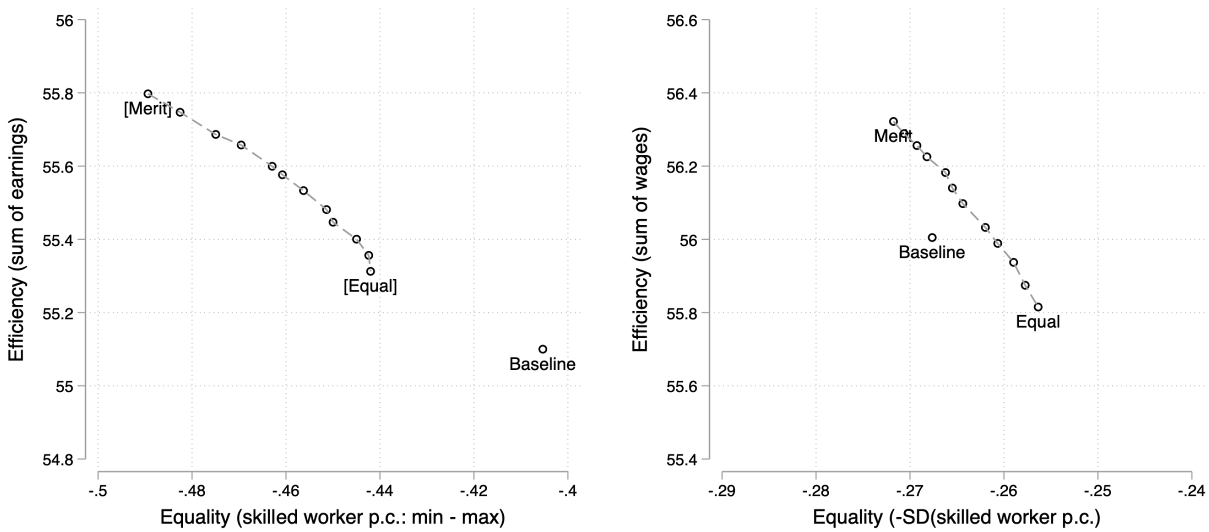


Figure A11: Counterfactual: Total output vs. distribution of skilled labor

Notes: The left panel is for the standard model in the paper. Each simulation plotted in the right panel is generated by restricting the effect of college location on post-graduation migration to zero.

Appendix B Details on identification of the measurement model

B.1 Illustration of the identification strategy

Consider the latent factor model of θ^0 in the main text:

$$\begin{aligned} NCEE_{hzt} &= \kappa_{1,hzt} + \lambda_{1,hzt}\theta^0 + \sigma_{1,hzt}\nu_1; \\ GPA_{sm} &= \kappa_{2,sm} + \lambda_{2,sm}\theta^0 + \sigma_{2,sm}\nu_2. \end{aligned}$$

This measurement model can only be estimated using a selected sample. It only contains students who are admitted by a college so that I observe their college GPAs, and the college sm is also an endogenous choice made by the students. I assume that students know θ^0 and select colleges based on $NCEE_{hzt}$ (i.e. selection by ν_1). However, students do not observe ν_2 nor anything correlated with ν_2 in any college before attending it, so there is no selection by ν_2.[A1] Selection by ν_1 implies that, in general, the conditional distribution of ν_1 does not equal the unconditional distribution:

$$\begin{aligned} &E(\nu_1|hzt, sm, \theta^0) \neq E(\nu_1) = 0; \\ &var(\nu_1|hzt, sm, \theta^0) \neq var(\nu_1) = 1. \end{aligned}$$

While no selection by ν_2 implies

$$\begin{aligned} &E(\nu_2|hzt, sm, \theta^0) = E(\nu_2) = 0, \ \ \forall hzt, sm, \theta^0; \\ &var(\nu_2|hzt, sm, \theta^0) = var(\nu_2) = 1, \ \ \forall hzt, sm, \theta^0. \end{aligned}$$

Under these assumptions, the measurement model parameters $\{\kappa, \lambda, \sigma\}$ are identified using the conditional mean, variance and covariance of the NCEE scores and college GPAs for each

[A1] The error term ν_2 may further include behavioral responses during college conditional on entering with a pre-college human capital θ_0. In one robustness check, I further control for a measure of study effort during college as a proxy for behavioral responses.

combination of the province-track-year hzt and the attended college-major pair sm:

$$
\begin{aligned}
&E(NCEE_{hzt}|sm) = \kappa_{1,hzt} + \lambda_{1,hzt}E(\theta^0|hzt, sm) + \sigma_{1,hzt}E(\nu_1|hzt, sm);\\
&E(GPA_{sm}|hzt) = \kappa_{2,sm} + \lambda_{2,sm}E(\theta^0|hzt, sm);\\
&var(NCEE_{hzt}|sm) = \lambda^2_{1,hzt}var(\theta^0|hzt, sm) + \sigma^2_{1,hzt}var(\nu_1|hzt, sm)\\
&\qquad\qquad + 2\lambda_{1,hzt}\sigma_{1,hzt}cov(\theta^0, \nu_1|hzt, sm);\\
&var(GPA_{sm}|hzt) = \lambda^2_{2,sm}var(\theta^0|hzt, sm) + \sigma^2_{2,sm};\\
&cov(NCEE_{hzt}, GPA_{sm}) = \lambda_{1,hzt}\lambda_{2,sm}var(\theta^0|hzt, sm) + \lambda_{2,sm}\sigma_{1,hzt}cov(\theta^0, \nu_1|hzt, sm)
\end{aligned}
$$

Cancelling out $E(\theta^0|hzt, sm)$ and $var(\theta^0|hzt, sm)$ gives three moment conditions for each combination of hzt and sm, denoted as (M1) to (M3):

$$
\begin{aligned}
&\text{(M1)}\quad E(NCEE_{hzt}|sm) = \kappa_{1,hzt} + \frac{\lambda_{1,hzt}}{\lambda_{2,sm}}\left(E(GPA_{sm}|hzt) - \kappa_{2,sm}\right) + \sigma_{1,hzt}E(\nu_1|hzt, sm);\\
&\text{(M2)}\quad var(NCEE_{hzt}|sm) = \frac{\lambda_{1,hzt}}{\lambda_{2,sm}}\left(cov(NCEE_{hzt}, GPA_{sm}) - \lambda_{2,sm}\sigma_{1,hzt}cov(\theta^0, \nu_1|hzt, sm)\right)\\
&\qquad\qquad + \sigma^2_{1,hzt}var(\nu_1|hzt, sm) + 2\lambda_{1,hzt}\sigma_{1,hzt}cov(\theta^0, \nu_1|hzt, sm);\\
&\text{(M3)}\quad var(GPA_{sm}|hzt) = \frac{\lambda_{2,sm}}{\lambda_{1,hzt}}\left(cov(NCEE_{hzt}, GPA_{sm}) - \lambda_{2,sm}\sigma_{1,hzt}cov(\theta^0, \nu_1|hzt, sm)\right) + \sigma^2_{2,sm}.
\end{aligned}
$$

To illustrate the identification idea, first assume $E(\nu_1|hzt, sm)$, $var(\nu_1|hzt, sm)$ and $cov(\theta^0, \nu_1|hzt, sm)$ are known. The identification will require: 1) data from at least two different (hzt)'s and two different college-major pairs sm's, 2) students from each hzt attend both sm. These data will generate 4 combinations of hzt-sm and the conditional moments in (M1)-(M3) within each cell, resulting in total 4*3=12 moment conditions. Under the normalization in which $\kappa_{1,(hzt)^*} = 0$ and $\lambda_{1,(hzt)^*} = 1$ for a specific $(hzt)^*$, there are in total 10 measurement parameters, so they are (over-)identified. Lastly, going back to $E(\nu_1|hzt, sm)$, $var(\nu_1|hzt, sm)$ and $cov(\theta^0, \nu_1|hzt, sm)$, they will be endogenously determined, given the selection by ν_1 in college choices, and thus need to be derived from the structural model. As a result, the measurement model will be estimated jointly with the behavioral model, which is discussed in the section on the estimation strategy.

The next section gives the complete proof of identification and how the identification started from the initial four hzt-sm cells can be extended to cover all provinces, tracks, years and colleges.

B.2 Identification proof

To reduce notational burdens, ignore NCEE track z and year t. The identification will be the same when adding these dimensions back. Consider province h and h', and college s and s':

$$NCEE_h = \kappa_{1,h} + \lambda_{1,h}\theta^0 + \sigma_{1,h}\nu_1 \qquad NCEE_{h'} = \kappa_{1,h'} + \lambda_{1,h'}\theta^0 + \sigma_{1,h'}\nu_1$$
$$GPA_s = \kappa_{2,s} + \lambda_{2,s}\theta^0 + \sigma_{2,s}\nu_2 \qquad GPA_{s'} = \kappa_{2,s'} + \lambda_{2,s'}\theta^0 + \sigma_{2,s'}\nu_2$$

Normalize $\kappa_{1,h} = 0$ and $\lambda_{1,h} = 1$. There are four groups of students depending on $h * s$ combination.

To begin, calculate the variances and covariances of NCEE scores and college GPAs for the four groups:

$$var(NCEE_h|s) = var(\theta^0|h,s) + \sigma_{1,h}^2 var(\nu_1|h,s) + 2\sigma_{1,h}cov(\theta^0,\nu_1|h,s)$$
$$var(NCEE_h|s') = var(\theta^0|h,s') + \sigma_{1,h}^2 var(\nu_1|h,s') + 2\sigma_{1,h}cov(\theta^0,\nu_1|h,s')$$
$$var(NCEE_{h'}|s) = \lambda_{1,h'}^2 var(\theta^0|h',s) + \sigma_{1,h'}^2 var(\nu_1|h',s) + 2\lambda_{1,h'}\sigma_{1,h'}cov(\theta^0,\nu_1|h',s)$$
$$var(NCEE_{h'}|s') = \lambda_{1,h'}^2 var(\theta^0|h',s') + \sigma_{1,h'}^2 var(\nu_1|h',s') + 2\lambda_{1,h'}\sigma_{1,h'}cov(\theta^0,\nu_1|h',s')$$
$$var(GPA_s|h) = \lambda_{2,s}^2 var(\theta^0|h,s) + \sigma_{2,s}^2$$
$$var(GPA_s|h') = \lambda_{2,s}^2 var(\theta^0|h',s) + \sigma_{2,s}^2$$
$$var(GPA_{s'}|h) = \lambda_{2,s'}^2 var(\theta^0|h,s') + \sigma_{2,s'}^2$$
$$var(GPA_{s'}|h') = \lambda_{2,s'}^2 var(\theta^0|h',s') + \sigma_{2,s'}^2$$
$$cov(NCEE_h, GPA_s) = \lambda_{2,s} var(\theta^0|h,s) + \lambda_{2,s}\sigma_{1,h}cov(\theta^0,\nu_1|h,s)$$
$$cov(NCEE_h, GPA_{s'}) = \lambda_{2,s'} var(\theta^0|h,s') + \lambda_{2,s'}\sigma_{1,h}cov(\theta^0,\nu_1|h,s')$$
$$cov(NCEE_{h'}, GPA_s) = \lambda_{1,h'}\lambda_{2,s} var(\theta^0|h',s) + \lambda_{2,s}\sigma_{1,h'}cov(\theta^0,\nu_1|h',s)$$
$$cov(NCEE_{h'}, GPA_{s'}) = \lambda_{1,h'}\lambda_{2,s'} var(\theta^0|h',s') + \lambda_{2,s'}\sigma_{1,h'}cov(\theta^0,\nu_1|h',s')$$

Canceling out each $var(\theta_i^0|\cdot)$ in the variance equations using the corresponding covariance equation gives:

$$var(NCEE_h|s) = \frac{1}{\lambda_{2,s}}cov(NCEE_h, GPA_s) + \sigma_{1,h}^2 var(\nu_1|h,s) + \sigma_{1,h}cov(\theta^0, \nu_1|h,s)$$
$$var(NCEE_h|s') = \frac{1}{\lambda_{2,s'}}cov(NCEE_h, GPA_{s'}) + \sigma_{1,h}^2 var(\nu_1|h,s') + \sigma_{1,h}cov(\theta^0, \nu_1|h,s')$$
$$var(NCEE_{h'}|s) = \frac{\lambda_{1,h'}}{\lambda_{2,s}}cov(NCEE_{h'}, GPA_s) + \sigma_{1,h'}^2 var(\nu_1|h',s) + \lambda_{1,h'}\sigma_{1,h'}cov(\theta^0, \nu_1|h',s)$$
$$var(NCEE_{h'}|s') = \frac{\lambda_{1,h'}}{\lambda_{2,s'}}cov(NCEE_{h'}, GPA_{s'}) + \sigma_{1,h'}^2 var(\nu_1|h',s') + \lambda_{1,h'}\sigma_{1,h'}cov(\theta^0, \nu_1|h',s')$$
$$var(GPA_s|h) = \lambda_{2,s}cov(NCEE_h, GPA_s) - \lambda_{2,s}^2\sigma_{1,h}cov(\theta^0, \nu_1|h,s) + \sigma_{2,s}^2$$
$$var(GPA_{s'}|h) = \lambda_{2,s'}cov(NCEE_h, GPA_{s'}) - \lambda_{2,s'}^2\sigma_{1,h}cov(\theta^0, \nu_1|h,s') + \sigma_{2,s'}^2$$
$$var(GPA_s|h') = \frac{\lambda_{2,s}}{\lambda_{1,h'}}cov(NCEE_{h'}, GPA_s) - \frac{\lambda_{2,s}^2}{\lambda_{1,h'}}\sigma_{1,h'}cov(\theta^0, \nu_1|h',s) + \sigma_{2,s}^2$$
$$var(GPA_{s'}|h') = \frac{\lambda_{2,s'}}{\lambda_{1,h'}}cov(NCEE_{h'}, GPA_{s'}) - \frac{\lambda_{2,s'}^2}{\lambda_{1,h'}}\sigma_{1,h'}cov(\theta^0, \nu_1|h',s') + \sigma_{2,s'}^2$$

Due to selection on ν_1 in college choices, $var(\nu_1|\cdot)$ and $cov(\theta^0, \nu_1|\cdot)$ will be derived from the structural model. After they are known, the 8 equations above have 7 unknown parameters: $\lambda_{1,h'}$, $\lambda_{2,s}$, $\lambda_{2,s'}$, $\sigma_{1,h}$, $\sigma_{1,h'}$, $\sigma_{2,s}$ and $\sigma_{2,s'}$, and they are over-identified.

Next, calculate the expected NCEE scores and college GPAs:

$$E(NCEE_h|s) = E(\theta^0|h,s) + \sigma_{1,h}E(\nu_1|h,s)$$
$$E(NCEE_h|s') = E(\theta^0|h,s') + \sigma_{1,h}E(\nu_1|h,s')$$
$$E(NCEE_{h'}|s) = \kappa_{1,h'} + \lambda_{1,h'}E(\theta^0|h',s) + \sigma_{1,h'}E(\nu_1|h',s)$$
$$E(NCEE_{h'}|s') = \kappa_{1,h'} + \lambda_{1,h'}E(\theta^0|h',s') + \sigma_{1,h'}E(\nu_1|h',s')$$
$$E(GPA_s|h) = \kappa_{2,s} + \lambda_{2,s}E(\theta^0|h,s)$$
$$E(GPA_{s'}|h) = \kappa_{2,s'} + \lambda_{2,s'}E(\theta^0|h,s')$$
$$E(GPA_s|h') = \kappa_{2,s} + \lambda_{2,s}E(\theta^0|h',s)$$
$$E(GPA_{s'}|h') = \kappa_{2,s'} + \lambda_{2,s'}E(\theta^0|h',s')$$

Canceling out each $E(\theta_i^0|\cdot)$ in above equations gives:

$$
\begin{aligned}
E(GPA_s|h) &= \kappa_{2,s} + \lambda_{2,s}\Big(E(NCEE_h|s) - \sigma_{1,h}E(\nu_1|h,s)\Big)\\
E(GPA_{s'}|h) &= \kappa_{2,s'} + \lambda_{2,s'}\Big(E(NCEE_h|s') - \sigma_{1,h}E(\nu_1|h,s')\Big)\\
E(GPA_s|h') &= \kappa_{2,s} + \frac{\lambda_{2,s}}{\lambda_{1,h'}}\Big(E(NCEE_{h'}|s) - \kappa_{1,h'} - \sigma_{1,h'}E(\nu_1|h',s)\Big)\\
E(GPA_{s'}|h') &= \kappa_{2,s'} + \frac{\lambda_{2,s'}}{\lambda_{1,h'}}\Big(E(NCEE_{h'}|s') - \kappa_{1,h'} - \sigma_{1,h'}E(\nu_1|h',s')\Big)
\end{aligned}
$$

Since $\lambda_{1,h'}$, $\lambda_{2,s}$, $\lambda_{2,s'}$, $\sigma_{1,h}$ and $\sigma_{1,h'}$ are already identified, $\kappa_{1,h'}$, $\kappa_{2,s}$, $\kappa_{2,s'}$ can be identified using the four equations above. Now all parameters in the two provinces and two colleges are identified.

To identify parameters for a third province h'', we need some of its students to attend an identified college. Assume they attend college s. Calculate the conditional mean, variance and covariance of the NCEE score and college GPA:

$$
\begin{aligned}
&E(NCEE_{h''}|s) = \kappa_{1,h''} + \lambda_{1,h''}E(\theta^0|h'',s) + \sigma_{1,h''}E(\nu_1|h'',s)\\
&E(GPA_s|h'') = \kappa_{2,s} + \lambda_{2,s}E(\theta^0|h'',s)\\
&var(NCEE_{h''}|s) = \lambda_{1,h''}^2 var(\theta^0|h'',s) + \sigma_{1,h''}^2 var(\nu_1|h'',s) + 2\lambda_{1,h''}\sigma_{1,h''}cov(\theta^0,\nu_1|h'',s)\\
&var(GPA_s|h'') = \lambda_{2,s}^2 var(\theta^0|h'',s) + \sigma_{2,s}^2\\
&cov(NCEE_{h''},GPA_s) = \lambda_{1,h''}\lambda_{2,s}var(\theta^0|h'',s) + \lambda_{2,s}\sigma_{1,h''}cov(\theta^0,\nu_1|h'',s)
\end{aligned}
$$

Cancelling out $E(\theta^0|h'',s)$ and $var(\theta^0|h'',s)$ gives the following three equations:

$$
\begin{aligned}
E(NCEE_{h''}|s) &= \kappa_{1,h''} + \frac{\lambda_{1,h''}}{\lambda_{2,s}}\left(E(GPA_s|h'') - \kappa_{2,s}\right) + \sigma_{1,h''}E(\nu_1|h'',s);\\
var(NCEE_{h''}|s) &= \frac{\lambda_{1,h''}}{\lambda_{2,s}}\big(cov(NCEE_{h''},GPA_s) - \lambda_{2,s}\sigma_{1,h''}cov(\theta^0,\nu_1|h'',s)\big)\\
&\qquad + \sigma_{1,h''}^2 var(\nu_1|h'',s) + 2\lambda_{1,h''}\sigma_{1,h''}cov(\theta^0,\nu_1|h'',s);\\
var(GPA_s|h'') &= \frac{\lambda_{2,s}}{\lambda_{1,h''}}\big(cov(NCEE_{h''},GPA_s) - \lambda_{2,s}\sigma_{1,h''}cov(\theta^0,\nu_1|h'',s)\big) + \sigma_{2,s}^2.
\end{aligned}
$$

Since $\kappa_{2,s}$, $\lambda_{2,s}$ and $\sigma_{2,s}$ are known, $\kappa_{1,h''}$, $\lambda_{1,h''}$ and $\sigma_{1,h''}$ are identified once $E(\nu_1|h'',s)$, $var(\nu_1|h'',s)$ and $cov(\theta^0,\nu_1|h'',s)$ are calculated from the structural model. By the same fashion, parameters for a third college s'' can be identified as long as some of its students come from an identified province.

In China, all colleges admit students from more than one province, and many admit from all provinces. Since the GPA measurement in each college is assumed to be stable over time, the college GPA can link different cohorts of students within the same college. Given these facts and assumptions, the identification chain can be extended to all province-track-years and colleges.

Appendix C Details on data and sample selection

C.1 Administrative data on college admission

C.1.1 Data availability

There are a few exceptions in the NCEE track setting and data availability:

- Guangdong: No track difference in 2006 (coded as sciences in the data and analysis); two tracks since 2007.
- Jiangsu: No track difference in 2006 and 2007 (coded as sciences in the data and analysis); two tracks from 2008. Due to regulatory restrictions, admission data between 2009 and 2011 are not available.
- Zhejiang: Due to regulatory restrictions, admission data in 2010 and 2011 are not available.

Note that having missing data for a province in some (but not all) years do not affect the identification of the student quality distribution $F_h(\theta^0)$ in that province. Based on the measurement model, $F_h(\theta^0)$ can be identified using only one year's data, although data from multiple years provide over-identification and improve estimation efficiency.

C.1.2 NCEE score outliers of admitted students

Some non-sports and non-arts colleges still offer sports- or arts-related majors, and most of their students are admitted through specialty tracks of the NCEE other than the standard sciences and humanities tracks. Many of these students have much lower NCEE scores. To prevent using their scores when forming the admission cutoff of each college, which will otherwise produce extremely low cutoffs, I drop all students in sports- and arts-related majors, which consist of 1.76% of the sample.

C.2 The Chinese College Student Survey (CCSS)

C.2.1 Sampling issues

I use CCSS graduating class surveys from 2010 to 2015. After 2015, the sampling design was changed. In 2010, the CCSS randomly selected 100 colleges stratified by tier and region (Municipalities, East other than Municipalities, Central, West, and Northeast), weighted by college capacity. The survey was rolled out among selected colleges during 2010-2015. Unfortunately, due to an unexpected funding cut in 2014 and 2015, 10 colleges that were scheduled for the survey in those years didn't participate. As a result, in total 90 colleges are surveyed, among which 8 are vocational colleges and 82 are four-year colleges. Among the 82 four-year colleges, one is a police school and is dropped from my sample because their students enter the police system after graduation and have restricted work location options. Another two colleges are also dropped because each of them has less than 10 students recorded in the sample due to unexpected issues in survey administration. Therefore, in total 79 four-year colleges are kept in the sample. Each college participated at least one year in the CCSS. Many colleges participated in multiple years.

Since some colleges that were originally selected in the sampling design are dropped from the final sample, I re-construct the school and individual weights to restore the national representativeness of the final sample using the school capacity information from the administrative data.

C.2.2 Sample selection

To begin, the 2010-2015 four-year college sample contains 32,879 students. 1,054 (3.21% of all students) are dropped from the sample because they are admitted through self-admission by each individual college, not through the regular NCEE. Another 2,076 (6.31% of all students) are dropped because they are in the specialty tracks of the NCEE, mostly arts or sports. As with the administrative sample, students from Tibet are also excluded from the CCSS sample. As a result, 29,789 students are left for the final sample and all of them are admitted through the regular NCEE in the sciences or humanities tracks. On average, 149 students are surveyed each year in each school, which is about 6% of the average cohort size in one college.

C.2.3 Post-graduation choice and initial job

The CCSS collected each student's post-graduation choices during the survey, which is administered about two months before college graduation. For post-graduation choices (all numbers are weighted using adjusted sampling weights), 74.5% choose to enter the labor market after graduation, 19.7% go to graduate school or prepare for the graduate school entrance exam, 2.2% go overseas for study or work, and 3.7% do not yet have a plan. Among those who choose to work after graduation, 88.8% have searched for jobs and 78.9% have at least one offer at the time of the survey. Among those who have offers, detailed information about the best job offer is collected. 70.5% accepted this offer, 7.8% were still in consideration, and 21.7% chose not to accept and continued the job search. Perhaps not surprisingly, the initial earnings in offers that are not accepted are 14. 3% lower, after controlling for schools, major, NCEE, college GPA, individual and family background. However, the distribution of the job offer's location is similar when comparing offers that are accepted and rejected. Initial earnings are trimmed at the top and bottom 1%.

I use the accepted job offers to estimate the earnings equation and the work location choices in the model. I interpret the wage rate in accepted job offers as an unbiased (but certainly noisy) measure of post-college human capital. I assume that all students will eventually find and accept such a job offer and treat those who have not at the time of the survey (reject current offers, no offer yet, or not start searching yet) as being in a frictional job search process. In terms of location choices, given that the distribution of the job location is similar when comparing offers that are accepted and rejected, estimating location preferences using only the sample of accepted job offers is less of a concern. Effectively, I assume there is no unobserved heterogeneity affecting the timing of receiving acceptable job offers and that the timing does not systematically vary across regions.

Appendix D Details on model estimation

D.1 List of targeted moments

I target a set of moments in the SMM based on the simulated data to the observed counterparts that can be calculated using the administrative data on college admission and data from college student surveys.

The first set of moments is directly related to college choices, including

- The share of eligible college applicants choosing each college $s \in \mathcal{S}$,
  - at each admission cutoff point c^*
  - in each NCEE province h, track z, and year t,

  where students consist of those with NCEE score $\geq c^*$, and the choice set $\mathcal{S}$ includes colleges with admission cutoffs $\leq c^*$.

The second set of moments is related to choices upon graduation, including the following:

- Work location choice:
  - The share of college graduates choosing to work in each province k;
- Dynamic correlation of locations:
  - The share of college graduates working in their home province;
  - The share of college graduates working in their college province;
- Initial earnings:
  - The mean of log earnings by pre-college human capital θ^0, college quality V_s, and work location k;
  - The covariance between log earnings and θ^0, V_s, the interaction $\theta^0 * V_s$, and average wage level in k;
- Graduate study:
  - The share of college graduates choosing graduate studies in each college tier;
  - The share of college graduates choosing graduate studies in each CCSS college s.

The third consists of a set of moment conditions derived from the measurement model:

- The relationship between the mean of NCEE scores and the mean of college GPAs:

  $E(NCEE_{hzt}|s) = \kappa_{1,hzt} + \frac{\lambda_{1,hzt}}{\lambda_{2,s}} \left(E(GPA_s|hzt) - \kappa_{2,s}\right) + \sigma_{1,hzt} E(\nu_1|hzt, s);$

- The relationship between the variance of NCEE scores and the covariance between NCEE scores and college GPAs:

$$\begin{aligned} var(NCEE_{hzt}|s) =& \frac{\lambda_{1,hzt}}{\lambda_{2,s}} \big(cov(NCEE_{hzt}, GPA_s) - \lambda_{2,s}\sigma_{1,hzt} cov(\theta^0, \nu_1|hzt, s) \big) \\ &+ \sigma^2_{1,hzt} var(\nu_1|hzt, s) + 2\lambda_{1,hzt}\sigma_{1,hzt} cov(\theta^0, \nu_1|hzt, s); \end{aligned};$$

- The relationship between the variance of college GPAs and the covariance between NCEE scores and college GPAs:

$$var(GPA_s|hzt) = \tfrac{\lambda_{2,s}}{\lambda_{1,hzt}} \left(cov(NCEE_{hzt}, GPA_s) - \lambda_{2,s}\sigma_{1,hzt} cov(\theta^0, \nu_1|hzt, s) \right) + \sigma^2_{2,s}.$$