Safety First: Perceived Risk of Street Harassment and Educational Choices of Women*

Girija Borker†

JOB MARKET PAPER

15 January 2018

Latest Version

Abstract

This paper examines the impact of perceived risk of street harassment on women's human capital attainment. I assemble a unique dataset that combines information on 4,000 students at the University of Delhi from a survey I designed and conducted, a mapping of potential travel routes to all colleges in the students' choice set using an algorithm I developed in Google Maps, and crowd-sourced mobile application safety data. Using a random utility framework, I estimate that women are willing to choose a college in the bottom half of the quality distribution over a college in the top quintile in order to travel by a route that is perceived to be one standard deviation (SD) safer. Furthermore, women are willing to spend INR 18,800 (USD 290) per year more than men for a route that is one SD safer – an amount equal to double the average annual college tuition. These findings have implications for other economic decisions made by women. For example, it could help explain the low female labor force participation in India.

*Thanks to Andrew Foster for his invaluable guidance and continuous support, and to Kaivan Munshi, Emily Oster, and Matthew Turner for their generous advice and insights. I would like to thank Jesse Shapiro, Dan Bjorkegren, Bryce Steinberg, Kenneth Chay, Anja Sautmann, Rebecca Thornton, John Friedman, Nancy Luke, Margarita Gafaro Gonzalez, and Akhil Lohia for their helpful conversations and comments, and the participants at the Applied Microeconomics Seminars, the Population Studies and Training Center (PSTC) gender workshop at Brown University, the PAA Annual Meeting 2017, and the Population Health Science Research Workshop 2017. I am grateful to Inderjeet S. Bakshi, Harjender Singh Chaudhary, Shubham Chaudhary, Rajiv Chopra, Sanjeev Grewal, Rabi Narayan Kar, Kawar Jit Kaur, Pravin Kumar, Amna Mirza, Shashi Nijhawan, Pragya, Rajendra Prasad, Savita Roy, Alka Sharma, Neetu Sharma, Malathi Subramaniam, Dinesh Yadav, Vimlendu and the team of dedicated survey enumerators for their support during data collection at University of Delhi. I would like to thank Ashish Basu and Kalpana Vishwanath for providing access to the SafetiPin mobile app data. This project was made possible by the expertise of Michael Davlantes and the capable research assistance provided by Peeyush Kumar. I gratefully acknowledge financial support from the UN Foundation, the PSTC, Global Health Initiative, Brown India Initiative, Pembroke Center for Teaching and Research on Women, and the Department of Economics at Brown University. All remaining errors are my own.

†Ph.D. Candidate, Department of Economics, Brown University. Email: *girija_borker@brown.edu*.

1 Introduction

Gender-specific constraints may help explain a significant portion of the economic mobility differentials between men and women in developing countries. These constraints include laws that limit women's access to and ownership of productive assets (Rakodi 2014), poor access to credit (Asiedu et al. 2013), and women bearing disproportionate responsibility for household work (Ferrant, Pesando, and Nowacka 2014). In this paper, I examine an additional constraint that restricts women's economic mobility – safety in public spaces.

Street harassment, or sexual harassment faced in public spaces, is a serious problem around the world. In Delhi, where this study is based, 95 percent of women aged 16-49 report feeling unsafe in public spaces (UN Women and International Center for Research on Women 2013). Women incur significant psychological costs from sexual harassment (Langton and Truman 2014) and actively take precautions to avoid such confrontations (Pain 1997). For example, 84 percent of women aged 40 years or younger in India said that they avoid an area in their city because of harassment (Livingston 2015).

In this paper, I document that women in Delhi choose to attend worse ranked colleges than men, both in absolute terms and conditional on their choice set. There can be several explanations for this observation. In a country where over 85 percent of marriages are arranged (Borker et al. 2017), it may be that women do not care about college quality if an undergraduate degree only has signaling value in the marriage market. It could also be that women prefer to attend a local, lower quality institution because of family obligations. Another potential explanation could be that women do not like competitive environments, hence they choose not to attend high quality colleges with high quality peers (Niederle and Vesterlund 2007, Niederle and Vesterlund 2011, Buser, Niederle, and Oosterbeek 2014). This paper posits another explanation: in a context where the majority of students live at home and travel to college every day, women choose to attend worse-ranked colleges in order to avoid street harassment. Unrelated to individual or family preferences that may result in optimal choices, street harassment imposes an external constraint on women's behavior that could potentially lead to suboptimal choices.

Choosing a worse ranked college is likely to have long-term consequences since college quality affects a student's academic training (Zhang 2005), network of peers (Winston and Zimmerman 2004), access to labor opportunities (Pascarella and Terenzini 2005), and lifetime earnings (Brewer, Eide, and Ehrenberg 1999; Eide, Brewer, and Ehrenberg 1998). In fact, such misallocation of students to colleges, where high achieving females sort to low quality colleges, may not only affect women's long-term outcomes but could also have important aggregate productivity effects (Hsieh et al. 2016).

This paper measures the extent to which perceived risk of street harassment can help explain women's college choices in Delhi. For this, I evaluate the gender differential in trade-offs between travel safety, college quality, travel costs, and travel time in a model of college choice. The difference in trade-offs captures the cost of street harassment for women, since men in Delhi do not face such harassment. I provide what to my knowledge is the first evidence of the effect daily harassment has on a durable human capital investment such as higher education. I also estimate the first revealed preference estimates of street harassment in terms of travel costs and travel time, augmenting existing (Aguilar, Gutierrez, and Soto 2016) and mostly ongoing research on harassment.

I face three key identification challenges. First, I must define the *de facto* set of colleges that a student can actually choose from. Second, I need to define the set of routes a student could take to each of the colleges in their choice set. Finally, I need to determine the perceived safety of each travel route, which needs to be a measure that is likely to capture the safety perceptions of college students.

To address the main identification challenges, I assemble a unique dataset and exploit the set-up of University of Delhi (DU). DU is an umbrella entity that is composed of several colleges that are spread across Delhi. Each college has its own campus, classes, staff, and placements and operates essentially like an independent university. Admissions in DU are strictly based on students' high school exam scores. I infer students' comprehensive choice set of colleges, using detailed information on 4,000 students from a survey that I conducted in DU. Using the mapping capabilities of Google Maps and an algorithm that I developed, I map students' travel routes by travel mode, including both the reported travel route and the potential routes available to students for every college in their choice set. Finally, I combine the information on travel routes with crowd-sourced safety data from two mobile applications. The first mobile application, SafetiPin, provides perceived spatial safety data in the form of safety audits conducted at various locations across Delhi. The second mobile application, Safecity, provides analytical data on harassment rates by travel mode. The route and safety data together allow me to assign a safety score to each travel route.

To assess whether women face different trade-offs between travel safety and college quality, travel costs, and travel time compared to men, I use two approaches. First, I take advantage of the DU's admissions procedure to approximate a random allocation of college choice sets. At DU, students apply to all colleges in the University based on their high school exam scores. Each college has a "cutoff score" or the minimum score required to gain admission to a major in the college. I compare the choices of each student with other students of the same gender who live in the same neighborhood, study the same major, and have the same admission year. Given the discrete cutoffs, a change in students' relative exam scores also changes their choice sets, relative to their neighbors. I find that while women's choices seem to take into account both route safety and college quality, men's choices only depend on quality and are consistent with a model in which preferences are lexicographic in college quality.

Second, I use a combination of structural econometric methods to estimate the magnitude of the students' willingness to pay for travel safety. I estimate students' indirect utility function using a nested logit model and heterogeneity in these preferences using a mixed logit model. In my benchmark specification, students' indirect utility depends on college quality, route safety, travel costs, and travel time, and is allowed to fully vary by gender. The analysis uses spatial variation in students' location, destination colleges, route choices, mode choice and area safety. Identification is based on the assumption that the difference between men and women's unobserved preferences for a route to a college is uncorrelated with observed college quality, perceived travel safety, travel costs, and travel time.

I find that women are willing to choose a college that is in the bottom half of the quality distribution over a college in the top 20 percent for a route that is perceived to be one standard deviation (SD) safer. Men on the other hand are only willing to go for from a top 20 percent college to a top 25 percent college for an additional SD of perceived travel safety. Translating perceived safety to actual safety, an additional SD of perceived safety while walking is equivalent to a 3.1 percent decrease in the rapes reported annually. Using the travel cost method, I am able to value harassment and I find that women are willing to incur an additional expense of INR 20,000 (USD 310) per year to travel by a route that is one SD safer. This is a significant sum of money, double the average annual tuition in DU and seven percent of the average annual per capita income in Delhi. Women's willingness to pay (WTP) in terms of annual travel costs is much higher than men's WTP of INR 1,200 (USD 19) for an additional SD of perceived route safety. I find similar estimates in the trade-off between travel safety and college quality using the mixed logit model that allows for heterogeneity of preferences.

This paper relates to several streams of literature: the psychology and criminology literature that provides qualitative evidence on the effects of street harassment, the broader economics literature that studies the distortive effects of fear, the value of statistical life literature that estimates individual's willingness to pay for small reductions in mortality risks, and the school choice literature in economics that assesses the factors influencing school choice. It has been shown that fear of imagined dangers affects individual behavior (Becker and Rubinstein 2011). Specifically, there is evidence that harassment by strangers strongly effects women's perceptions of safety across social contexts (Macmillan, Nierobisz, and Welsh 2000) and that women change their behavior in response (Keane 1998). One common response is changing their mobility patterns (Hsing-Ping Psu 2011). While there have been several studies that estimate value of a statistical life using implicit trade-offs between different risks and money (Viscusi and Aldy 2003), there has been no attempt, to my knowledge, to measure the misallocation effects associated with sexual harassment. The school choice literature examines the institutional attributes that families value. Families have been found to value high academic attainment, proximity, and certain composition of students in terms of race and socio-economic status. (Gallego and Hernando 2009, Burgess et al. 2015, Hastings, Kane and Staiger 2009, Carneiro, Das, and Reis 2013, Muralidharan and Prakash 2017). This is the first paper to consider travel safety in a model of college choice, a factor that is likely to be especially relevant for educational choices of women in rapidly urbanizing developing countries.

2 Institutional Setting: College Choice and Harassment

2.1 Structure of University of Delhi

DU is one of the top non-technical universities in India (BRICS University Rankings 2015). DU is composed of 77 colleges of which 58 offer general undergraduate majors in Humanities, Commerce and Science.[1] There are over 180,000 undergraduate students at DU (University of Delhi Annual Report 2013-14), which represents around 8 percent of all students who passed the Class 12 qualifying exam in India.[2] DU is also the main public central university in Delhi that offers a liberal arts education. Other public universities offering general undergraduate majors are either significantly smaller in comparison or have limited overlap with the majors offered by DU.[3] Another option for students in Delhi are the private universities. However, these private institutions not only offer limited courses but are also considerably more expensive than DU. For example, one of the biggest private universities in Delhi charges on average 9 to 18 times DU's average annual tuition.

Colleges in DU are spread across Delhi. Figure 1 shows the spread of colleges in Delhi. Colleges vary in the size of their student population; on average, a college has about 2,800 students. Undergraduate studies at DU are for three years,[4] and each college offers multiple majors. On

[1] These exclude colleges that offer professional degrees like law and medicine. Of the 58 general education undergraduate colleges, 22 colleges are women only and eight colleges are evening colleges where classes take place after 2 pm.

[2] This represents around 6.6 percent of all students who appeared for the Class 12 qualifying exam in India.

[3] These general public universities in Delhi are Jamia Millia Islamia University, Jawahar Lal Nehru University, and Guru Gobind Singh Indraprastha University. Jamia Milia has less than 14 percent of DU's annual undergraduate intake of which up to 50 percent is reserved for Muslim students. Jawahar Lal Nehru University offers undergraduate programs only in foreign languages. Guru Gobind Singh Indraprastha University only offers one course that is offered by DU's general undergraduate colleges.

[4] In 2013, DU attempted to move to a four-year undergraduate program (FYUP). However, this decision was met

average a college offers about 12 majors with most colleges having a large overlap in the majors they offer.[5] Each college has its own campus, staff, classes, and placements.[6] Students within a major across colleges take a common university-wide exam at the end of each academic year. This exam on average accounts for 75 percent of the students' final grades. The remaining 25 percent of the final grade is based on internal college evaluation of the student.

A distinguishing feature of DU is that admission to colleges is strictly based on students' high school exam scores. Each college specifies a cutoff score or the minimum percentage required to gain admission to a major. Every student who scores above this cutoff score is guaranteed admission to the college. In line with previous studies (Black and Smith 2004), I use selectivity in admissions as an indicator of college quality, measured by a colleges' cutoff score. Based on these cutoff scores I am able to rank each college in absolute terms and within a student's choice set, where a higher rank indicates a lower cutoff score and hence worse quality. The absolute rank rates colleges within a major and admission year using the first cutoff list. Rank within a student's choice set ranks the colleges to which the student was admitted by their cutoff score for the students' major and admission year. Figures 2a and 2b show the cumulative distribution function (CDF) of the absolute rank of a college and the CDF of rank within a student's choice set, respectively. We can see that women's CDF lies below the CDF for men, indicating that women choose worse ranked colleges than men across the distribution. Women choose worse ranked colleges than men in absolute terms for the most part of distribution, and they choose worse quality colleges from the ones that they were eligible to attend for the entire distribution. This is despite women scoring higher than men on exams at the end of high school, as shown in Figure 3.[7]

Another feature of DU is that majority of the students (72 percent) live with their parents and

with widespread protests and was embroiled in controversy since its implementation. The FYUP was rolled back in 2014 and the University returned to its three-year undergraduate program.

[5]Only few colleges offer additional specialized courses such as Bachelors in Journalism and Bachelors in Elementary Education.

[6]While there is a Central Placement Cell that is open to all students enrolled in the University, the majority of the placements take place in the individual colleges. A Right to Information appeal revealed that the Central Placement Cell has placed only 5,800 students in past five years, equal to 13 percent of the total number of students who registered with the Cell.

[7]We can see from Figure 2b that 62 percent of men and 39 percent of women choose one of the top three colleges in their choice set.

travel to college daily. Of the students who are residents of Delhi, 99.1 percent live with their parents and travel to college every day. This is primarily because of the social norm of living with parents and because of lack of facilities at the University. About 18 percent of colleges have on-campus residence facilities that can accommodate about 5 percent of students enrolled in the University (NDTV 2015). The students travel to college by either public or private transport. In my sample, 83 percent of students use some form of public transport to travel to college every day. By focusing on Delhi residents who live with their parents and travel to college every day, I have a sample of students that does not sort on the basis of college location, since it is unlikely that parents' choice of residence is influenced by their children's future choice of college.

2.2 Street Harassment

Gender-based street harassment is defined as "unwanted comments, gestures, and actions forced on a stranger in a public place and is directed at them because of their actual or perceived sex" (Stop Street Harassment 2015). According to a nationally representative survey in the US, 65 percent of women have experienced street harassment (Stop Street Harassment 2014). Similarly, 86 percent of women living in cities in Thailand and 86 percent in Brazil have been subjected to harassment in public (Action Aid 2016). Delhi, infamously known as the "rape capital" of India, is notorious for both verbal and physical harassment on public transportation (YouGov 2014). In my sample, 89 percent of female college students have faced some form of harassment while traveling in Delhi. In particular, 63 percent of female students have experienced unwanted staring, 50 percent have received inappropriate comments, 40 percent have been touched, groped or grabbed, and 26 percent have been followed. Many women take precautions to avoid harassment, for example in my sample 72 percent of female students report avoiding an unsafe area, 67 percent avoid going out after dark, 31 percent move away from the harasser, and only 4 percent of women report taking no action to avoid harassment while traveling to college in Delhi.

This paper focuses on women enrolled in college as they are vulnerable to sexual attacks due to their age (17-21 years) and lack of experience in dealing with harassment. A survey of women 18 years and older in Chennai, another major city in India, found that 75 percent of women had their first encounter with sexual harassment between 14 to 21 years of age (Mitra-Sarkar and Partheeban 2011). For a majority of children in Delhi, both girls and boys, the main mode of transport to and from school is the official school bus. Once they finish high school, they are expected to take responsibility for their travel, as colleges have neither an official provision for transport nor standardized times for classes. Next, I present a simple stylized model of college choice to characterize how women may face trade-offs between travel safety and college quality.

3 Stylized Model of College Choice

In this section I show a simple stylized model of college choice. This model explicitly captures how women might have to choose worse quality colleges in order to avoid travel by unsafe routes. In the 2×2 matrix in Figure 4a, the high-scoring students are in the first row (high school exam score = H) and low-scoring students are in the second row (high school exam score = L). In the columns, there is a low quality "Not-so-good college" (Quality$=n$) and a high quality "Good college" (Quality $=g$) with $g>n$. In between these two colleges there is a "danger" area that is unsafe, a travel route becomes unsafe if it passes through this unsafe area. There is an equal number of high and low-scoring males and females located in each college's neighborhood. A high scoring student is eligible to attend both the good and the not-so-good colleges given that their high school exam score is above the cutoff for both colleges $(H>g>n)$. A low-scoring student, on the other hand, is only eligible to attend the not-so-good college given that their high school exam score is below the cutoff for the good college $(g>L>n)$. In this model, I assume that women have two options when choosing their travel routes: they can either avoid unsafe areas or travel by a safer but more expensive mode of transport and women prefer the former.

Figure 4b and 4c show the choices made by high-scoring and low-scoring males respectively.

Both high-scoring males attend the good college and both low-scoring males attend the not-so-good college. Given the set-up, this means that $\frac{1}{2}$ of the males travel by unsafe routes, denoted by the arrows, and a male student on average attends a college with quality = $\frac{n+g}{2}$. Figure 4d shows the choices of women who do not face a safety-quality trade-off. The high-scoring female chooses the good college and the low scoring female chooses the not-so-good college. In Figure 4e, we can see the choice of a high scoring female who would have to take an unsafe route i.e. cross the unsafe area if she were to choose the good college. By assumption she avoids the unsafe area and chooses the lower quality not-so-good college. Finally, Figure 4f shows the decision of the low scoring woman who would have to cross the unsafe area to attend the only college she is eligible for. She chooses a safe but more expensive route to travel to the Not so good college, denoted by the dashed green arrow. In such a case this woman could have also chosen to not attend college at all, denoted by the thick arrow. Given that this study examines choices of students currently enrolled in DU, I am unable to evaluate the effects of safety on the decision to attend college. However, if selection into college is similar to the selection into high and low quality colleges, then my estimates provide a lower bound of the effects of travel safety because there might be a host of women who choose to not attend college at all in order to avoid harassment. Based on this stylized example, for the students who decide to attend college, we can see that the embedded quality-safety trade-off manifests itself in all women traveling by safer routes compared to half of the men, women attending lower quality colleges relative to men, and women incurring higher travel costs than men.

There are three main challenges in estimating these trade-offs in practice, outside a 2×2 set-up. There are many colleges that a student can choose from, many routes that a student can take to each of the colleges in their choice set, and each route can have a different level of safety. I address each of these challenges in the following data section.

4 Data

I have three main types of data – student information from DU, travel routes from Google Maps, and mobile application safety data. This data enables me to address the aforementioned challenges. Using students' exam scores and DU's admissions information, I create students' complete choice set of colleges. Using Google Maps, I map students' reported and potential travel routes to each college in their choice set. Finally, I combine the mapped routes with mobile app safety data to compute the perceived safety of each travel route. Section 4.1 describes the student data, Section 4.4 describes the route creation using Google Maps, and Section 4.5 outlines the mobile app safety data.

4.1 Student Data

I have student information from three main sources: a sample of students from eight colleges in DU where a detailed survey was conducted, confidential administrative data on the entire student population of these eight colleges, and a sample of students from 32 other colleges in DU where a shorter survey was conducted.

Full Survey Data: In Spring 2016, I conducted a detailed survey in eight colleges in DU. As part of the survey, I collected data on 4,000 male and female students. This paper survey was conducted in class at a time that was previously scheduled with the professors. On average, students took about 25 minutes to complete the survey. From the full survey, I have information on students' current and permanent residential locations, exact daily travel route as a sequence of landmarks and modes of travel, high school exam scores by subject, parental and household characteristics, and measures of exposure to harassment for female students.

The eight colleges were purposefully chosen based on their geographic location and variation in quality. We can see from Figure 1 that the colleges are spread out across the city. Two colleges in sample are women only and one college is an evening college. Figure 5 shows the students in the full survey sample. From the figure, we can see that students travel to college from most parts of the Delhi National Capital Region. Based on the full survey data, I have a sample of 2,695 students, who live in Delhi with their families and travel to college every day, which comprise 71 percent of all students surveyed and 99.1 percent of Delhi residents who were surveyed.

Administrative Data: For the eight colleges in the full survey sample, I have confidential administrative data on all students enrolled in the colleges. I have information on students' genders, current and permanent residential locations, courses of study and social categories. For one of these colleges, I also have students' aggregate high school scores and parental occupations.

Short Survey Data: In addition to the detailed survey in eight colleges, I also conducted a short survey across 32 other colleges in DU. Data on 887 male and female students was collected through a combination of online (34 percent) and intercept (66 percent) surveys. From the short survey, I have information on students' current and permanent residential locations and high school exam scores by subject.

For the online survey, the staff and/or students in the 32 colleges were contacted. For the intercept survey, the students were approached outside their college campuses by enumerators and requested to fill the survey form. From the short survey data, I have a sample of over 669 students, who live in Delhi with their family and travel to college every day, which comprise 75 percent of all students surveyed and 99.5 per cent of the Delhi residents who were surveyed.

Representability of Full Survey Sample: The colleges in the full survey sample are fairly evenly distributed across the quality distribution, as shown in Figure 6 where each colored bar represents a college in the full survey sample.

Additionally, students in the full survey sample are also representative of the wider student body in the eight colleges and the University. Table 1 compares the characteristics in the full survey sample, the short survey sample and the administrative data. Test statistic for two sample t-tests comparing the sample means of the full survey data with the short survey data and administrative data are also reported. Based on the t-tests, I am unable to reject the null hypothesis of equality of sample means between the short survey sample and the full survey sample in terms of most admission categories of students and their high school exam scores, for both men and women. [8]

[8] The mean fraction of students in each admission category is similar between the full survey data and the administrative data except for male students belonging to the general category students and to Other Backward Castes (OBC), and Schedule Castes (SC) female students. These social categories are officially designated groups of historically disadvantaged people in India. SC were formerly referred to as "untouchables" and OBC is the collective term used by the Government to classify castes which are socially and educationally disadvantaged but not SC. Even though equality of means is statistically rejected, the difference of a maximum of five percentage points is economically not significant.

The mean distance to college and distance to city center are similar across samples except that women tend to live closer to the city center in the short survey sample compared to the full survey sample.

4.2 Admissions in DU

To gain admission in DU, students have to complete the Common Pre-admission Form. This is a single form that is used for admission to all colleges in the university. A student has to specify the major(s) they wish to apply for. Following the submission of the form, each college releases the first list of cutoff scores. The cutoff score is the minimum average percentage score a student needs in high school to gain admission into a college.[9] The high school scores are based on the national Senior School Certificate Examinations.[10,11] The college cutoff score is calculated separately for each major on the basis of the seats available in a college, the high school scores of applicants and the cutoff score in previous years according to the Delhi University Standing Committee on Admissions 2015. The cutoffs vary by social category,[12] subjects studied in Class 12 and in some cases by gender of the student.[13] Following the release of the cutoff list, students have about three days to register in a college of their choice. Students are required to submit their original degree certificate and pay the first year's annual fees at the time of admission. The colleges are obligated

[9]The average for each student is calculated on a "best of four" basis, where most often the students can exclude their lowest scoring subject while calculating the average. Most colleges require students to include at least one language in this average.

[10]The Senior School Certificate Examinations are evaluated in a double blind manner.

[11]The majority of schools in India come under the purview of the Central Board for Secondary Education (CBSE), a board of education that conducts the Senior School Certificate Examination. The only other national board is the Indian Certificate of Secondary Education. Other boards of education are at the state level. In our sample over 96 percent of students' board of examination was the CBSE.

[12]Social categories refers to the officially designated groups of historically disadvantaged people in India. These categories are used for the purposes of affirmative action. These are General category (Gen), Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Castes (OBC). Another category used for admissions is Physically Handicapped (PH). Gen is the unreserved category, SC were formerly referred to as "untouchables", ST are the indigenous people, and OBC is the collective term used by the Government to classify castes which are socially and educationally disadvantaged but not SC.

[13]In minority colleges, cutoffs are lower for students belonging to the minority religion. A few colleges also take into account the subjects studied in Class 10, mostly for language courses. A sample cutoff list is shown in Figure A1. In this cutoff list, the cutoff score are listed by college major (rows) and students' social categories (columns). We can see that the minimum score required by a general category male student to gain admission in Economics is 95 percent, for female students the cutoff score is 1 percentage point lower at 94 percent.

to admit every student who approaches the college with a score above the released cutoff score.[14] After three days if there are seats available in a college then the college revises its cutoffs downward and releases the second cutoff list. The same process is repeated until all seats in every college are filled. In 2015, DU released 12 cutoff lists. Based on these objective cutoffs it is possible to construct the choice set of colleges for each student conditional on choice of major.[15]

4.3 Choice Set Creation

I construct student's choice set conditional on major choice using students' high school scores by subject and each college's publicly available cutoff lists. For every student in the sample, I compute an aggregate score following guidelines specified by each college in DU. If the student's aggregate score percentage is greater than the cutoff specified by a college, then that college is in the student's choice set. I repeat this procedure for all cutoff lists released by every college, which is equivalent to using the lowest cutoff score.[16] On average, a student has 22 colleges in their choice set. As expected, the number of colleges in a student's choice set is positively correlated with their high school exam score and the cutoff score of their chosen college, as shown in Figure A2.

Accurate choice sets are crucial for my analysis. Most importantly, there should not be any systematic errors in choice sets by gender. Since the choice sets are created based on students' reported high school exam scores, I test if there is any systematic misreporting of exam scores by gender. For this, I match students from the full survey sample to the college administrative data at the one college for which I have students' high school exam scores. The students are matched on the basis of their residential locations, genders, social categories and parental occupations.[17] I find

[14] There are some instances where colleges have claimed to run out of registration forms to prevent students from registering once the college had reached its sanctioned limit (Hindustan Times 2013).

[15] In principle, only a student with scores above the cutoff can be granted admission. However, in my data I find about 10 percent of the students enrolled in a college where the cutoff score is above their high school exam score. This could be because of misreporting of the high school exam score or patronage or if the student was admitted under a different category than stated. For example, some seats in every college are reserved for students who have excelled in sports and extra-curricular activities, and the cutoffs for these students are not made public by all colleges.

[16] Two colleges are excluded from the analysis because they followed a different procedure for admissions.

[17] I was able to match 78 percent of the Delhi residents in my full survey sample to the administrative data for the

that on average students report 0.75 to 1 percentage point higher scores in the survey data, but there is no gender differential in this misreporting.

4.4 Route Mapping using Google Maps

Students' reported and potential travel routes are mapped using an algorithm I develop in Google Maps. I map students' reported travel routes as a sequence of landmarks and travel modes, taking into account the departure times. The travel information collected as part of the full survey and its mapping in Google Maps fills a major data gap in India, since there are no detailed travel surveys in the country. The existing data on daily travel from the Census of India is aggregated at the district level making it impossible to study travel choices by individual attributes.[18] To create students' potential routes to the chosen college and the colleges in their choice set, up to four routes are extracted per Google Maps based travel option, i.e., driving only, walking only and public transit, giving a total of up to 12 travel routes for every student to each college in their choice set. The public transit routes are then broken into separate legs based on travel modes. I drop students who are outliers in terms of travel time and I also drop potential routes that have a travel time greater than the 99th percentile of reported travel time. Allowing for variation in departure times, the reported travel route is one of the options suggested by Google Maps between the origin and destination for over 90 percent of the students in sample.[19] Ultimately, for every student I have their reported travel route and potential travel routes to the college they chose and the potential travel routes they could have taken to each college in their choice set.[20]

An example of route mapping is given in Figure A3. Figure A3a shows a student and the college

one college, without any conflicts.

[18]One exception to this is the travel survey conducted by Bansal et al. (2016) in three major cities in India. However, the focus of their travel survey is vehicle ownership with a few questions on average travel patterns, as opposed to details of daily travel routes by mode, which I collected.

[19]These checks were conducted on a 15 percent random sample of the data, stratified by travel mode.

[20] To my knowledge, Google Maps does not factor in travel safety in their route suggestion algorithm. Given that the observed routes and hypothetical routes highly overlap, under the null of zero safety effect, the routes created using Google Maps seem to perform well as choice set routes for the students.

he chose to attend. Figure A3b shows the actual route he travels by every day where he steps out of his house and takes a rickshaw to the closest metro station, he then takes a bus to a bus stop near his college from where he walks to college. Figure A3c shows potential route options to the chosen college and Figure A3d shows the potential route options to each of the 32 college in this student's choice set.

4.5 Safety Data

The final piece of data I use is safety data from two popular mobile applications in Delhi – area safety data from the SafetiPin mobile app and safety by travel mode from the Safecity mobile app.

Safetipin Mobile Application Data: SafetiPin is a mobile app that allows its contributors to conduct "safety audits" of a location. These safety audits allow the user to characterize the safety of a location based on nine parameters. The nine parameters are openness of spaces, visibility or "eyes on the street", presence of security personnel, the condition of the walking path, presence of people specifically women and children on the street, access to public transport, extent of lighting, and the overall feeling of safety. The contributors can rate a location by assigning a score from 0 (low safety) to 3 (high safety) on each of the nine parameters. Details of each parameter and a description of the audit rubric are given in Table A1. For my benchmark specification, I use a composite area safety index of the nine parameters computed using principal component analysis. I check for robustness by excluding one safety parameter from the safety index each time.

SafetiPin was launched in November 2013 in Delhi, and the app is now available in 28 cities across 10 countries. The SafetiPin data is partially crowd-sourced and partially collected by trained auditors. The latter enables SafetiPin to have a wider and more representative coverage of the city (Vishwanath and Basu 2015).

I have data on over 26,500 audits from November 2013 to January 2016, as shown in Figure 7a. In this sample, 98 percent of the contributors are 39 years or younger and 70 percent of the users are female.[21] I interpolate these audits to create a safety surface using Inverse Distance Weighting, this base level of area safety is shown in Figure 7b. Each pixel is 300 meters×300 meters.

Safety Data by Mode of Travel: SafetiPin audits do not capture the safety of a travel mode. Hence, I use data on safety of a travel mode from analytical data based on another safety mobile app called Safecity. Safecity allows its users to record personal stories of harassment and abuse in public spaces. In these stories, the users mention the mode of transport they were using when they experienced harassment. The data I use is based on 5,500 crowd sourced reports of harassment. This information is used to weight area safety by the travel mode, while computing the safety of a travel route. Table 2 provides information on mode usage by gender in the full survey data and proportion of harassment reports by mode from Safecity's analytical data. Students use a variety of modes to travel to college, with 38 percent of students traveling by a public or private bus for some portion of their daily route. Men are more likely to travel by bus than women. The metro is the most popular mode of transport for all college students and is more popular among women by a significant margin. Of the women who travel by the metro, 86 percent reported exclusively traveling in the ladies-only compartment. A large proportion of both men and women are likely to walk some part of their travel route, with men being more likely than women to have a walking part. From the last column of Table 2, we can see that, in line with anecdotal evidence, buses are the most unsafe mode of transport with about 40 percent of the harassment incidents involving a bus or the people in it. This is followed by the metro which covers about 16 percent of the incidents.

4.6 Calculating Route Safety

I assign a safety score to the reported and potential routes by computing a weighted average of the area safety data, where the weights are the proportion of a route and harassment by travel mode (m) in each safety pixel (p). Specifically, the safety score or the route shown in Figure 8 is calculated

[21] Contributor characteristics are available for 80 percent of the data.

as:

$$Safety\ of\ travel\ route = \Sigma_m \Sigma_p \left[\frac{Area\ safety_p \times Route\ length_{mp}}{Total\ route\ length} \times (1 - Harassment_{mp}) \right]$$

Here the base level of safety is from the SafetiPin data; route length divided by the total route length gives the proportion of route in pixel p; and the final term is to take into account harassment based on mode m used in pixel p. I use $(1 - Harassment)$ since Safecity data is about harassment while the SafetiPin area safety data is about the feeling of safety such that a higher value indicates higher perceived safety. For example, $Harassment_{m=walk} = 0$ while $Harassment_{m=bus} = 0.4$, using the above formula this means that in the same area and with equal length routes, route safety in a bus is 40 percent lower than the route safety while walking. This is the route safety measure I use in the benchmark specification. I check for robustness by using alternative safety measures.

Table 3 reports summary statistics on the variables we use for subsequent analysis. As mentioned previously, the relevant sample for this study is Delhi residents who live with their family. In this sample, 65 percent of the students are female. Relative to men, women on average come from households with a higher socio-economic status.[22] In terms of college choice, women choose colleges that have more than a one percentage point lower cutoff score than men's chosen colleges and attend colleges that are on average ranked 5th within their choice set, compared to men who attend their 3rd or 4th ranked college. The chosen college is equally far for both men and women. Women seem to choose colleges that have a larger student population, offer more majors, and are more likely to have boarding facilities. In this sample, 44 percent of women attend women only colleges. In terms of route choice, relative to men women choose routes that are safer, more expensive, and have a shorter travel time. The descriptive statistics are in line with the outcomes from the stylized example in Section 3.

[22] Students' socio-economic status is measured by an index variable created using principal component analysis. The index is based on whether a student lives in rented or owned house, students has own laptop, computer, or both, the number of cars, scooters and motorcycles owned by household, price of most expensive car owned by household, "pocket money" or money spent per month excluding travel expense, indicator for whether student attended private school, and mother's and father's years of education.

5 Descriptive Evidence: Response to Changes in Choice Set

In this section, I evaluate students' responses to changes in their choice sets to better understand their underlying preferences. The ideal experiment for this exercise would require random allocation of college choice sets to students. Then evaluating students' responses in terms of choice attributes to the variation in their choice sets would help me better understand their underlying preferences. Since I do not have full control over students choice sets, I exploit DU's admissions process to approximate the ideal experimental design. I use the fact that students' high school exam scores combined with colleges' cutoff scores completely determine each student's choice set.

I compare students' choices in terms of travel safety, college quality, travel time and costs, with those of a "similar" student as their relative exam scores change. Given the discrete cutoffs, a change in the student's relative exam score also changes their relative choice set. I define a "similar" student as a student who lives in a 1.5 km radius neighborhood[23] around the index student, is of the same gender, studying the same major and has the same admission year. A student with a greater high school exam score faces a superior choice set in terms of college quality and a larger, though not necessarily superior, choice set in terms of route attributes compared to a neighbor with a lower cutoff score. I have 2,951 unique pairs that use information on 69 percent of the students in my sample.

To better understand how analyzing students' choices relative to a close neighbor can help me understand their underlying preferences, consider two extreme cases. First, if students have lexicographic preferences in terms of quality, then we would observe that the relative quality of the index student's chosen college would increase with an increase in the index student's score relative to her neighbors, while the relative route safety could move in any direction. Relative travel time and travel cost could also change in any direction with an increase in the index student's score gap. In the other extreme case, if students have lexicographic preferences in terms of safety then we would observe no change or an increase in the safety difference between the index student's chosen route relative to her neighbor's chosen route with an increase in the index student's relative exam

[23]The 1.5 km radius is the minimum distance for which 90 percent of the students have at least one neighbor.

score. The safety difference would remain constant with an increase in the score difference in the special case when the college with the safest travel route has a lower cutoff score than all students' high school exam scores in every neighborhood. The relative college quality of the chosen college, the relative travel time, and cost could move in any direction with a increase in the index student's relative exam score.

Figure 9 plots the binned scatter plots of difference in safety, quality, time, and cost between the index students and their neighbors' choice against the difference between the index student's high school exam score and their neighbor's, distinguishing between males and females. The score bins are of a two-point absolute score difference. In the student-neighbor pair, the index student is the student who has a greater high school exam score. In these figures, a greater score difference implies that the index student faces a bigger choice set in terms of both colleges and travel routes. I find that women choose higher quality colleges that lie on safer travel routes that are longer and marginally more expensive with an expansion in their choice set. Men also choose higher quality colleges and routes that are marginally more expensive but they do not respond in terms of safety or time. From Figure 9a, we can see that there is a positive relation between safety difference and the score difference for females while there is a no such systematic relation for males. This means that while females choose safer routes relative to their neighbors as their college choice set and hence their route choice set expands, this is not the case for males whose choice of relative route safety is almost flat across the score differences. From Figure 9b, the positive relation between quality difference and the score difference for both males and females signifies that an increase in the index student's score relative to their neighbor's is associated with an increase in relative college quality for both men and women. The quality gradient is significantly lower for females compared to males. Figure 9c shows that women choose relatively longer routes with an increase in their relative scores, compared to men. There is only a marginal difference between men and women's relative travel costs with a change in their relative scores in Figure 9d. The equivalent linear regression results are reported in Table A2.

It is important to note that the binned scatter plots show the total effects, as opposed to partial effects, associated with the expansion of a student's choice set. Based on these total effects, I find that women value safety differently compared to men. And while women's choices seem to take into account both route safety and college quality, men's choices only depend on quality and are in fact fairly consistent with the hypothesized preferences that are lexicographic in quality. These results are suggestive of important differences between men and women's preferences for safety and quality. However, it is unlikely that students consider each attribute in isolation, hence we need to compute partial effects or the effect of each choice attribute conditional on other attributes. Based on this evidence we also cannot ascertain the magnitude of the trade-offs. I address both these issues in the utility model of college choice presented in the next section.

6 Model of College Choice

To estimate the partial effects and measure students' willingness to pay for different choice attributes, I structurally estimate students' indirect utility function by gender. This section lays out the structural model of college choice, which is estimated in Section 9. I follow an additive random utility framework with a rational, utility maximizing student i (McFadden 1977, McFadden 1978) to estimate revealed preference valuation of safety. In this framework, each student i faces a choice of N_i mutually exclusive colleges denoted by $C_{i1},\ldots,C_{iN_i}$ and travel routes to each college in her choice set $r^1_{i1},\ldots,r^1_{iR_1},\ldots,r^{N_i}_{i1},\ldots,r^{N_i}_{iR_{N_i}}$ where r^c_{iR} is the R^{th} route that student i can take to college c. The entire set of routes r is partitioned according to their destination colleges into N_i non-overlapping subsets denoted $\mathbf{C_i} = \{C_{i1},C_{i2},\ldots,C_{iN_i}\}$ to give the tree structure shown in Figure 10.

Students are assumed to maximize an indirect utility function of the form:

$$\begin{aligned} U^c_{ir} &= V^c_{ir} + \varepsilon^c_{ir} \\ &= W^c_i + Y^c_{ir} + \varepsilon^c_{ir} \end{aligned} \tag{1}$$

where r and c denote the travel route and college respectively, V^c_{ir} denotes the deterministic component of the utility function, and ε^c_{ir} is the unobserved part of utility that captures the effect of unmeasured variables, personal idiosyncracies, maximization error, etc. Each student i chooses the college c and route r ($d^c_{ir} = 1$) that maximizes his or her utility over all possible colleges and routes in their choice set, such that:

$$\begin{aligned}
&d^c_{ir} = 1 \quad if\ and\ only \quad if \quad U^c_{ir} > U^b_{is} \qquad \forall b \neq c \qquad \forall r \neq s \\
&d^c_{ir} = 0 \quad otherwise
\end{aligned}$$

The empirical implementation of the choice model requires specific functional form assumptions about the deterministic component of the utility function as well distributional assumptions about the error term. Following the literature on discrete choice, I assume that V^c_{ir} is a linear function of college, route, and student characteristics that is composed of two additively separable components – one that is constant for all routes to a college and depends only on the college attributes (W^c_i), and another that varies over routes to each college (Y^c_{ir}). These components can be further broken into the following components:

$$\begin{aligned}
U^c_{ir} = & \quad W^c_i + Y^c_{ir} + \varepsilon^c_{ir} \\
U^c_{ir} = & \quad \overbrace{\gamma_q Q^c_i + \delta_q Q^c_i \times Fem_i}^{W^c_i} + \overbrace{\gamma_s S^c_{ir} + \delta_s S^c_{ir} \times Fem_i}^{Y^c_{ir}} \\
& + \overbrace{\gamma_p p^c_{ir} + \delta_p p^c_{ir} \times Fem_i + \gamma_t T^c_{ir} + \delta_t T^c_{ir} \times Fem_i}^{Y^c_{ir}} + \varepsilon^c_{ir}
\end{aligned} \tag{2}$$

$X^c_{ir} = W^c_i + Y^c_{ir}$ is a set of characteristics for student i, route r, and college c where Q^c_i is quality of college c, S^c_{ir} is safety of the travel route to college, p^c_{ir} is the travel cost to college, T^c_{ir} is the travel time to college and Fem_i indicates whether the student is female.

With respect to the error term ε_i, I assume that it follows the generalized extreme value distribution with a cumulative distribution function of the following form

$$\varepsilon^c_i \sim exp\left(-\Sigma^{N_i}_{c=1}\left(\Sigma_{r\in C_c} exp\left(\frac{-\varepsilon_{ir}}{\lambda_c}\right)\right)^{\lambda_c}\right) \tag{3}$$

The marginal distribution of each ε^c_{ir} is univariate extreme value but the ε^c_{ir} are correlated within nests i.e.

$$
\begin{aligned}
Cov(\varepsilon_{i\ell}^{c}, \varepsilon_{im}^{b}) \quad &\neq 0 \qquad if \quad c = b \\
&= 0 \qquad if \quad c \neq b \;\; for\ any\ \ell \in C_c and\ m \in C_b
\end{aligned}
$$

The parameter λ_c is the inclusive value coefficient for college c. It measures the degree of independence in unobserved utility among travel route alternatives in nest c. A higher value of λ_c means greater independence.

The probability that student i chooses route r to college c is given by the standard condition:

$$P_{ir}^{c} = Pr(U_{ir}^{c} > U_{is}^{b}) \qquad \forall b \neq c \qquad \forall r \neq s$$

This probability can also be written as a product of two standard logit probabilities, the conditional probability of choosing r given that a route in nest C_c is chosen ($P_{ir|C_c}$) and the marginal probability of choosing a route in nest C_c (P_{iC_c}).

$$P_{ir}^{c} = P_{ir|C_c} \times P_{iC_c} \tag{4}$$

Based on the distribution for the unobserved components of utility, we have the following:

$P_{ir|C_c} = \frac{e^{Y_{ir}^{c}/\lambda_c}}{\Sigma_{s \in C_c} e^{Y_{is}^{c}/\lambda_c}}$ and

$P_{iC_c} = \frac{e^{W_i^{c} + \lambda_c \mathrm{I}_{ic}}}{\Sigma_{b \in \mathbf{C_i}} e^{W_i^{b} + \lambda_b \mathrm{I}_{ib}}}$

where $\mathrm{I}_{ic} = ln \Sigma_{s \in C_c} e^{Y_{is}/\lambda_c}$

These probabilities form the log-likelihood function:

$$LL_{nl}(X, \gamma) = \sum_{i} \sum_{c} \sum_{r} d_{ir}^{c} \, log\{P_{ir|C_c} \times P_{iC_c}\}$$

The nested structure of the student choice model places less structure on the college and route selection process than the simple multinomial logit models. It does so by relaxing the assumption of independence of irrelevant alternatives (IIA). The nested logit structure assumes instead that choices within each nest are similar in unobserved factors, so that IIA holds for any pair of alternatives within each nest but not for the entire choice set. The relaxation of the IIA property translates into more plausible substitution patterns, enabling the econometrician to capture student specific responses to unobserved characteristics that are common to travel routes to a specific college. For example, if there are two equally preferred safe routes and one lesser preferred dangerous route to college 1 in Figure 10, the nested logit error structure assumes that if we remove one of the safe routes then there will be proportionate substitution to the remaining safe and the dangerous route to college 1. But there will be no substitution to travel routes to college 2 or college 3 in student i's choice set. Identification is based on the assumption that the difference between men and women's unobserved preferences for a college and the travel route to the college is uncorrelated with observed perceived safety, college quality, travel costs, and travel time. The explanatory variation comes from the spatial variation in students' locations, destination colleges, area safety, travel route and mode choices. Estimation is done using full information maximum likelihood.

Daly and Zachary (1978), and McFadden (1978) identify a set of conditions on parameter values under which the nested logit model equation 4 is globally consistent with utility maximization. One of these conditions is that the choice probabilities, P_{ir}^{c}, must have non-negative even and non-positive odd mixed partial derivatives with respect to components of V other than V_i, where V is the deterministic component of the indirect utility function in equation 1. This condition ensures that implied probability density function will be non-negative.[24] For this condition to hold globally, Daly and Zachary (1979), and McFadden (1979) show that the inclusive value coefficients must lie in the unit interval ($0 < \lambda_c \leq 1,\ \forall c$). Börsch-Supan (1990) derives a set of conditions under which a nested logit model is consistent with utility maximization even when the inclusive value coefficients lie outside the unit interval. These conditions were later extended and applied by Herriges and Kling (1996) and Kling and Herriges (1995). Consistency with random utility maximization implies the

[24] The other conditions are that the probabilities be non-negative, the probabilities over all alternatives sum to one, the probabilities depend only on the differences in utilities, and that the cross derivatives of the probabilities with respect to the arguments be symmetric.

following necessary conditions for all nests with two or more alternatives:

$$\lambda_c \leq U_{1c}(V) \equiv \frac{1}{1 - P_{iC_c}} \qquad c = 1, \ldots, N \tag{5}$$

Since $0 < P_{iC_c} \leq 1$, this conditions are always satisfied for any $\lambda_c \in [0,1]$. For the condition to be satisfied when $\lambda_c > 1$, P_{iC_c} must be sufficiently large. In general, for a nest with n alternatives, there will be $n-1$ necessary conditions for each order of mixed partial derivatives. However, Kling and Herriges (1995) note that in practice even when a model has many alternatives within each nest, given the errors implicit in model estimation, satisfaction of the necessary condition (equation 5) may be considered adequate. The local consistency condition are checked at the mean of the sample, i.e., $\hat{\lambda}_c \leq U_{1c}(\overline{V})$, where $\overline{V}$ is the value of the indirect utility function at the means of explanatory variables.

After estimating the nested logit, which does not allow for parameter heterogeneity across students, I follow Train (2003), Kremer et al. (2011) and others in explicitly estimating heterogeneity using a mixed logit model with random coefficients on college quality, route safety, and travel time in the student's indirect utility function. The weight students place on college quality and route safety may vary idiosyncratically and with observable student characteristics such as socio-economic status and high school academic achievement. The weight male and female students place on college quality may vary for two reasons. First, some students may simply place an inherently high value on institutional quality. Second, even if all students place low importance on college quality, some students may face high decision making costs leading them to place lower expressed weight on quality when determining their expected utility and selecting a college (Hastings, Kane and Staiger 2009). Similarly, the weight students place on route safety may vary because either students are inherently averse to harassment or because they face external pressure for example from parents, to travel by safe routes. These different sources of heterogeneity cannot be separately identified in this analysis because they result in observationally equivalent choice behavior.

By introducing individual heterogeneity in logit coefficients, the mixed logit model allows for flexible substitution patterns albeit by imposing more structure on the distribution of preferences.

The mixed logit model can approximate any random utility model, given appropriate mixing distributions and explanatory variables (McFadden and Train 2000). With random coefficients, equation 2 becomes:

$$\begin{aligned} U_{ir}^{c} &= \phi_i' X_{ir}^{c} + \varepsilon_{ir}^{c} \\ &= \phi_{iq} Q_i^c + \phi_{is} S_{ir}^c + \phi_p P_{ir}^c + \phi_{it} T_{ir}^c + \varepsilon_{ir}^c \end{aligned} \tag{6}$$

I assume that ε_{ir}^c is distributed i.i.d. extreme value and that the idiosyncratic portions of preferences are drawn from a multivariate normal mixing distribution, i.e.,$\phi \sim f(\phi|\mu,\nu)$, where $\mu$and ν denote the mean and variance parameters. Given these assumptions, the probability that student i chooses route r to college c is:

$$\mathscr{P}_{ir}^{c} = \int \left(\frac{exp(\phi' X_{ir}^{c})}{\Sigma_{c=1}^{N_i} \Sigma_{r \in C_c} exp(\phi' X_{ir}^{c})} \right) f(\phi|\mu,\nu)\, d\phi$$

where X_{ir}^c is as defined before, and $f(\cdot)$ is the mixing distribution. These probabilities form the log-likelihood function:

$$LL_{mxl}(X,\mu,\nu) = \sum_i \sum_c \sum_r d_{ir}^c \, log\{\mathscr{P}_{ir}^c\}$$

I estimate the model separately for men and women. I allow mean preferences for route safety and college quality to vary with students' SES and their high school exam scores. I assume $f(\cdot)$ to be the normal distribution for the route safety coefficient and the college quality coefficient, and the negative lognormal distribution for the travel time coefficient so that all students dislike longer commute time. The coefficient of travel costs is assumed to be fixed, following the literature with fixed price coefficients (Train 2003). Since the log-likelihood function does not have a closed form solution, simulation methods are used to generate draws of ϕ from $f(\cdot)$ to numerically integrate over the distribution of ϕ. Estimation is done by the method of maximum simulated likelihood, using 200 Halton draws of ϕ from $f(\cdot)$ for each student in the data set. The results are not sensitive to increasing the number of draws.

7 Identification

Several aspects of the context and data help to identify the parameters in the model of college choice. First, in addition to the lack of on-campus housing at DU, it is the norm that students live at home with their parents. That parents are unlikely to base their residential choices on the location of their children's future preferred colleges, helps to identify values placed on travel times and travel safety separately from residential sorting by focusing on the sample Delhi residents. Residential sorting could overstate the importance of travel time and safety for students located near to their preferred colleges.

Second, the colleges in DU are spread out across the city and are located in neighborhoods with varying characteristics and with students of both genders and all socio-economic groups. Each student faces a host of college and route choices only determined by the student's high school exam score and the colleges' cutoff score. Figure 11 shows the characteristics of students and the area around each college. Each bar represents a college and the colleges are in ascending order of quality.[25] Figure 11a and Figure 11b show that students with all levels of high school scores and both genders live near colleges across quality levels. There is also no sorting of colleges by quality according to the socio-economic status or safety of neighborhoods, as can be seen from Figure 11c and 12d. Hence, I have wide variation in both college and student locations, providing variation in route safety for students of both genders and colleges of all quality.

Third, college cutoff scores do not seem to take into account women's safety concerns. If travel safety affects the pool of students who enroll in a college, such that the number of high achieving female students who enroll is less than what a college anticipated, it maybe that the cutoff scores for women decrease or the advantage given to them increases the following year. This could bias the safety estimate. However, I find that observable characteristics of a college are unable to predict the advantage given to women, as shown in Table A3.

[25] Based on the cutoff score for Bachelors in Political Science, as shown in Figure 6(a).

8 Empirical Specification

The benchmark specification for the nested logit model estimates parameters of the following indirect utility function, as specified in equation 2:

$$
\begin{aligned}
U_{ir}^{c} = & \gamma_q Q_i^c + \delta_q Q_i^c \times Fem_i + \gamma_s S_{ir}^c + \delta_s S_{ir}^c \times Fem_i \\
& + \gamma_p p_{ir}^c + \delta_p p_{ir}^c \times Fem_i + \gamma_t T_{ir}^c + \delta_t T_{ir}^c \times Fem_i \\
& + \varepsilon_{ir}^c
\end{aligned}
$$

Q_i^c is quality of college c, measured by the cutoff score of college c to capture the selectivity of the college. I use the cutoffs for general category male students for co-educational colleges and for general category female students for women only colleges. I use these cutoffs for two reasons, first to ensure comparability across colleges because general category cutoffs are available for all colleges while some other social category cutoffs are not,[26,27] and second, using cutoffs for female students would, by construction, lower the quality of colleges that give an advantage to female students. Figure A4 shows the correlation between the cutoff score and proportion of accepted students who enrolled in a college. As expected we see a strong positive relationship. S_{ir}^c is the safety of the travel route to college measured in standard deviations (SD) from the mean. The safety score for each route is computed as explained previously. p_{ir}^c is the monthly travel cost to college in thousands of Indian Rupees and T_{ir}^c is the travel time to college in minutes, as computed by Google Maps. I use monthly costs here to replicate the monthly payments students make for bus travel which also lends a more relevant interpretation to the time coefficient, i.e., the marginal utility from a unit increase in travel time keeping the total monthly travel cost fixed. The use of travel time improves on previous estimations using travel distance to proxy for duration of travel. Students' choice variable is an indicator equal to 1 for the reported daily travel route to their chosen college, and 0 otherwise. The ratio of the coefficient estimate on route safety to the coefficient estimate

[26] For example, colleges that are recognized as Sikh minority institutions do not release a separate cutoff for students belonging to the OBC social category.

[27] The results do not change if I use cutoffs for other social categories.

on college quality is the marginal rate of substitution between safety and quality (MRS_{QS}). This gives the value of safety in terms of percentage points of the cutoff score. I allow the students' preferences for quality, travel costs, time costs and hence valuation of safety to vary by gender, by including interactions between Fem_i, an indicator for whether the student is female.

I expect the cost and time coefficients to be negative for both men (γ_p, γ_t) and women ($\gamma_p + \delta_p, \gamma_t + \delta_t$), indicating that all students prefer routes that are cheaper and shorter. I also expect the quality coefficient to be positive for both men (γ_q) and women ($\gamma_q + \delta_q$), indicating that all students prefer higher quality colleges. Based on my hypothesis, I would expect the total safety coefficient to be positive women ($\gamma_s + \delta_s$) such that women prefer routes that are safer. Most importantly, I expect the marginal rate of substitution between quality and safety for women to be greater than that for men ($MRS^F_{QS} > MRS^M_{QS}$). In other words, I expect women to be willing to forego a higher level of college quality for an additional SD of travel safety, compared to men. Similarly, I also expect women to have a higher willingness to pay for an additional unit of safety in terms of travel costs and travel time.

9 Results

The estimation results with the benchmark specification and the augmented specification are reported in Panel A and B in Table 4. For the estimation I restrict the sample to students who choose a college that is ranked eight or better in their choice set. Over 90 percent of students in the full survey sample choose a college that is ranked eighth or better in their choice set. I also restrict all students' choice set to the top eight colleges. This helps reduce the estimation time by a significant amount. The results are not sensitive to increasing the rank threshold.

As expected, the coefficient on cost and time is negative, and the coefficient on quality is positive for both men and women. The coefficient on safety is positive for both men and women, but significantly greater for women.[28] The positive safety coefficient for men most likely captures

[28]Using these coefficients, I predict the optimal route for each college in a student's choice set. For the chosen college, the chosen route is one of the top three predicted routes in 45 percent of the cases. Next, assuming that the

the amenity value of a safe route, i.e., better lighting, better access to transport etc. Based on the coefficient estimates in the benchmark specification, I find students' valuation of safety in terms of college quality, route travel costs, and travel time by gender. Women are willing to attend a college that is 13.04 percentage points lower in quality for an additional SD of safety. This is equivalent to choosing a college that is 8.5 ranks lower.[29] To better understand the meaning of one additional SD of travel safety, I translate perceived safety to actual safety using district level rape data from the National Crime Record Bureau. I estimate that one additional SD of route safety while walking is equivalent to a 3.1 percent decrease in the rapes reported annually.[30,31] Men on the other hand are willing to attend a college that is only 1.37 percentage points (or 0.9 ranks) lower in quality for an additional SD of safety. In terms of travel costs, women are also willing to travel by a route that costs Rs. 20,000 (USD 310) more per year as long as it is one SD safer. Men are willing to spend an additional Rs. 1,200 (USD 19). This shows that women are willing to spend 16 times more than men in terms of travel costs for an additional unit of safety. The difference of Rs. 18,800 is equal to almost double the average annual tuition at DU and 110 percent of the average annual travel costs in this context. Women are also willing to travel an additional 40 minutes daily for a route that is one SD safer. Men are willing to increase their travel time by four minutes for an additional SD of safety. All of the aforementioned safety valuations are measured in terms of the SD of route safety across the predicted route alternatives in a students' choice set, which is 26.5 percent lower than the overall SD in route safety.[32]

In Panel B, I include additional college level variables that might influence a students choice. These include for every college the annual tuition, the area safety within a 1.5 km around the college, the number of majors offered, an indicator for whether the college offers boarding facilities,

predicted route for a college is the route that a student would take if the student were to attend that college, I predict the college a student would choose. In 82 percent of the cases, the rank of the chosen college is the same as the rank of one of the top three predicted colleges.

[29]Conversion to rank is based on the regression of absolute rank on cutoff score for all general education undergraduate colleges in DU for the three years. The regression includes major and year fixed effects (not reported).

[30]This estimate is based on a district level regression of log of rapes in 2013 on average area safety and log of the number of the 15 to 34 year old females (not reported).

[31]Rape is the most feared crime by women younger than 35 years of age. Additionally, for women, the perceived seriousness of a rape is approximately equal to the perceived seriousness of murder (Fairchild and Rudman 2008).

[32]For example, $MRS^{M}_{QS} = (1-0.265) \times \frac{\gamma_s}{\gamma_q}$.

the total number of students enrolled, and an indicator for whether the college is women only. Addition of these controls does not change the results significantly and all additional variables have coefficients that are not significantly different from zero. The coefficient on annual tuition (not reported) is negative in line with the findings of previous studies in the US (Neill 2009 and Buss, Parker, and Riverburg 2004). However, the coefficient is close to zero and imprecisely estimated. College fees in this context is highly subsidized by the government, does not account for the majority of costs of college, and are basically a measure of the amenities in a college. It is expected that the inclusion of a different inclusive value coefficient for each college nest captures most of the tuition effect and hence the result is not surprising. The coefficient on the indicator for whether a college is women only is also negative and imprecisely estimated.

In the benchmark specification, 12 of the 56 inclusive value coefficients are greater than one. The coefficients are given in Table A4. Following Kling and Herriges (1995), I check for first-order and second-order consistency conditions by testing the hypothesis $H_{j0} : \hat{\lambda}_c \leq U_{jc}(\overline{V})$ against the alternative hypothesis $H_{jA} : \hat{\lambda}_c > U_{jc}(\overline{V})$ for each c, $j = 1,2$. Table A4 reports the t-ratios associated with the test statistic $\hat{z}_{jc} \equiv \hat{\lambda}_c - U_{jc}(\overline{V})$ for the one-tailed test. In my sample, students can have a different number of colleges and routes in their choice sets, hence, simply computing the first-order and second-order conditions at mean values of the explanatory variables does not yield a unique P_{iC_c} value for each college. So for these tests I use the mean value of $P_{iC_c}(\overline{V})$ for every college.[33] In both cases, negative t-values signify failure to reject the null hypothesis of consistency with utility maximization. Based on this criteria, all colleges satisfy the first-order conditions. These results can be considered conclusive if satisfaction of the first-order conditions is felt to be adequate.

Table 5 presents results from the mixed logit estimation by gender. I report estimates for the mean of each logit coefficient, along with the standard deviations and correlations for the random parameter distributions. I allow the mean weight that a student places on route safety and college quality to depend on the student's own high school exam score and SES (interaction terms with

[33] Using the median or modal value does not change the results.

route safety and college cutoff score). Students' high school exam scores and SES index are constructed as deviations from the mean, so that the coefficients on route safety and college cutoff scores represent the respective weights placed by a student with average high school exam scores and SES.

Broadly speaking, the estimates are consistent with the nested logit results. Focusing first on the means of the coefficients, both female and male students are more likely to choose a college that has a higher cutoff score, a route that is safer, cheaper, and shorter. In terms of heterogeneity by observable characteristics, the positive interactions of route safety and college cutoff scores with students' SES indicate that the average weight placed on route safety and college quality increases with SES for both men and women, consistent with findings of previous literature (Hastings, Kane and Staiger 2009, Satija and Datta 2015). Similarly, the positive interaction of college cutoff scores with students' high school exam scores shows that, as expected, the average weight placed on college quality increases with students' baseline exam scores for both men and women. The interaction of route safety and students' high school exam scores is positive but imprecisely estimated for both men and women.[34]

For women the random coefficient on route safety and college quality are negatively correlated. Hence, female students who choose safer routes are more likely to attend lower quality colleges. This shows that for the average woman, selecting a safe travel route will require her to choose a college that is lower quality, leading to tradeoffs between travel safety and college quality. On the other hand, the random coefficient on route safety and college quality are positively correlated for men. The random coefficient on college quality and travel time are negatively correlated for women and positively correlated for men. Hence, male students who choose a higher quality college are more likely to have longer travel times.

Similar to the nested logit results, the ratio of coefficients can be used to calculate the tradeoffs in relative terms. Based on the ratio of the mean route safety coefficient to the mean college cutoff

[34] In addition to the variation related to observable characteristics, idiosyncratic preferences contribute substantial variation to the weight places on travel route and college factors, with the estimated standard deviation for most of the random coefficients ranging between 0.4 to 2.5 times of the mean for each coefficient.

score, women are willing to attend a college that is 9.66 percentage points lower in quality for an additional SD of safety. Men on the other hand are willing to attend a college that is only 0.71 percentage points lower in quality for an additional SD of safety. These results, shown in Table 6, are lower but similar to the estimated MRS in the nested logit model.[35]

9.1 Robustness Checks for the Choice Model

I am working on conducting a variety of robustness checks for the benchmark specification. These include:

Alternative Construction of Route Safety: The area safety index is constructed using principal component analysis, with the nine parameters in SafetiPin as inputs. I drop one parameter each time and reconstruct the area safety index. I find that the results do not change significantly across these different safety indices. In Table A5, each panel reports results based on these alternative measures of safety. Here I report the estimates from two alternative safety indices that are the most different from my benchmark results. For example, in Panel A, the safety index excludes the subjective feeling parameter. The estimates with these alternative measures of safety for the benchmark specification are similar to what I obtain by using the safety measure using all parameters for area safety. There is a robust positive coefficient on travel safety for women and they have a high willingness to pay for an additional SD of safety in terms of college quality, travel costs, and time, relative to men.

Other Margins of Choice: It could be that students' jointly consider college and major choice. The benchmark analysis is conditional on major choice but it maybe that the choice of major is affected by students' safety consideration. From the full survey data I know the majors that each student considered at the time of application. I find that students apply for up to five majors, with majority of students submitting either one (41 percent) or two (20 percent) majors. The distribution is shown in Table A6. I find that there is significant overlap in related majors which students tend

[35]The safety valuation, as before, is measured in terms of the SD of route safety across the predicted route alternatives in a students' choice set, which is 45 percent lower than the overall SD in route safety for both men and women.

to consider together, at the time of application. As shown in table A7, the overlap in choice sets varies between 76 percent to as high as 96 percent.

9.2 Policy Experiments

There can be several policy responses to reduce street harassment or to alleviate its effects on women's economic choices. Using the structural estimates, I can evaluate the effects of some policies that are currently used and those that could be implemented. The first best policy intervention would make public spaces safe and accessible for women such that they would face no harassment. I assess the effect of such a first best policy where all travel routes for women have the highest level of safety observed in the data. Increasing safety of all travel routes for women improves the expected rank of their predicted college choice and women travel by cheaper but longer routes. Specifically, it reduces the average rank gap between men and women by five percent. Women's cost of travel decreases by almost 19.8 percent and their travel time increases by four percent. As we saw previously, men travel by longer and cheaper routes than women, the cost reduction from an increase in travel safety closes the travel cost gap between men and women by over 28 percent.

An alternative policy lever could empower women to fight back, thereby influencing women's preferences for safety. This can be done in many ways, either through self-defense training, such as initiatives taken by SheFighter in Egypt, or by encouraging women to reclaim public spaces as in the WalkAlone campaign by Blank Noise in India, or through information campaigns like 'Stop Telling Women to Smile' in several cities across the world. I estimate this in the model by assuming that women and men have the same preferences for safety. While the first best policy changes the environmental safety factor, this policy changes the preferences for different levels of environmental safety. I find that the effect of such policies is similar to the first-best policy. The gap between men and women's college rank reduces by four percent, and the travel cost gap decreases by 21 percent. Another intervention could be to subsidize the modes of transport for *some* women.[36] To simulate the effect of this policy experiment, I assume that women pay the

[36]The reason I consider subsidizing travel for some but not all women is to disregard the general equilibrium effects

highly subsidized bus fare on all modes of transport, this reduces the rank gap between men and women by two percent and ensures that women spend significantly less on travel compared to men, by design. While easier to implement, this intervention has half the effect of a policy that empowers women to be less affected by the lack of safety.

10 Conclusion

Street harassment is a serious problem around the world especially in rapidly urbanizing developing countries. While there is qualitative evidence on the negative effects of street harassment on women's economic mobility, this is the first study to quantitatively assess the long term economic consequences of street harassment. By combining unique data that I collected from the University of Delhi, with route mapping from Google Maps, and mobile app safety data, I study the trade-offs women face between college quality and travel safety, relative to men. I find that women face significant trade-offs and are willing to attend a college that is 25 percentage points lower in the quality distribution for a route that is perceived to be one SD safer. Men are only willing to attend a college that is 5 percentage points lower in the quality distribution for a route that is one SD safer. Additionally, I find that women are willing to spend an additional Rs. 18,800 on annual travel, relative to men, for a route that is one SD safer. This amount is more than double the average annual fees at DU. Using estimates from Sekhri (2014), on the labor-market earnings advantage from attending a public college, I estimate that women's willingness to pay for safety translates to a 20 percent decline in the present discounted value of their post-college salaries. These results show that street harassment is an important mechanism that could perpetuate gender inequality in both education and lifetime earnings.

Using structural econometric methods, I evaluate and interpret the impact of different existing and potential interventions that can alleviate the impact of street harassment on economic choices of women. I find that the first-best policy of making all travel routes safe for women reduces the average rank gap between men and women by five percent. While it is the ultimate hope,

of the policy change.

such an ambitious policy is hard to design and implement in the short or even medium-term. The desired outcome is most likely to be achieved through a series of multi-dimensional interventions over the long-term. I find that affecting women's preferences for safety through empowerment, has the same effect on women's quality of higher education as the first-best policy, and it is more effective than subsidizing travel by safer modes for them. This shows that one of the most effective and sustainable class of policies would empower women to deal with harassment and hold the perpetrators accountable, with support from the wider society. These multipronged policies can be in the form of self-defense training or information campaigns that make women aware of their rights and also involve the community.

While this study focuses on the role of street harassment in explaining women's choice of college, the findings are relevant for other economic decisions made by women that could be affected by their propensity to avoid harassment. For instance, the global labor force participation rate for women is 26.7 percentage points lower than the rate for men in 2017 and the largest gender gap in participation rates is faced by women in emerging countries (World Employment Social Outlook 2017). The results of this paper suggest that street harassment could help explain part of this gender gap. In the context of India, labor force participation rates for women aged 25-54 have stagnated at about 26 to 28 percent in urban areas, between 1987 and 2011. The fact that this is the case, despite the economic and demographic conditions that ordinarily would lead to rising female labor-force participation rates, remains a long-standing puzzle (Klasen and Pieters 2015). This is an important issue for India's economic development. With a high share of working-age population, labor force participation, savings, and investment can boost per capita growth rates. However, if a majority of women do not participate, say because of the fear of harassment, then the effect will be never be as strong.

References

Aguilar, Arturo, Emilio Gutierrez and Paula Soto. 2016. "Women-only Subway Cars, Sexual Harassment, and Physical Violence: Evidence from Mexico City." *Working Paper* .

Asiedu, Elizabeth, Isaac Kalonda-Kanyama, Leonce Ndikumana and Akwasi Nti-Addae. 2013. "Access to credit by firms in Sub-Saharan Africa: How relevant is gender?" *The American Economic Review* 103(3):293–297.

Azevedo, Eduardo M and Jacob D Leshno. 2016. "A supply and demand framework for two-sided matching markets." *Journal of Political Economy* 124(5):1235–1268.

Bansal, Prateek, Kara M Kockelman, Will Schievelbein and Scott Schauer-West. 2016. "Indian Vehicle Ownership and Travel Behaviors: A Case Study of Bangalore, Delhi and Kolkata.".

Becker, Gary and Yona Rubinstein. 2011. Fear and the Response to Terrorism: An Economic Analysis. Technical report Centre for Economic Performance, LSE.

Black, Dan A and Jeffrey A Smith. 2004. "How robust is the evidence on the effects of college quality? Evidence from matching." *Journal of Econometrics* 121(1):99–124.

Borker, Girija, Jan Eeckhout, Nancy Luke, Shantidani Minz, Kaivan Munshi and Soumya Swaminathan. 2017. "Wealth, Marriage, and Sex Selection." *Working Paper* .

Börsch-Supan, Axel. 1990. "On the compatibility of nested logit models with utility maximization." *Journal of Econometrics* 43(3):373–388.

Brewer, Dominic J, Eric R Eide and Ronald G Ehrenberg. 1999. "Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings." *Journal of Human Resources* pp. 104–123.

Burgess, Simon, Ellen Greaves, Anna Vignoles and Deborah Wilson. 2015. "What parents want: School preferences and school choice." *The Economic Journal* 125(587):1262–1289.

Buser, Thomas, Muriel Niederle and Hessel Oosterbeek. 2014. “Gender, competitiveness, and career choices.” *The Quarterly Journal of Economics* 129(3):1409–1447.

Buss, Christian, Jeffrey Parker and Jon Rivenburg. 2004. “Cost, quality and enrollment demand at liberal arts colleges.” *Economics of Education Review* 23(1):57–65.

Card, David and Alan B Krueger. 1992. “Does school quality matter? Returns to education and the characteristics of public schools in the United States.” *Journal of Political Economy* 100(1):1–40.

Carneiro, Pedro, Jishnu Das and Hugo Reis. 2013. “Parental Valuation of School Attributes in Developing Countries: Evidence from Pakistan.” *London, UK http://ipl.econ.duke.edu/bread/papers/1013conf/carneiro.pdf* .

Cattaneo, Matias D, Rocío Titiunik, Gonzalo Vazquez-Bare and Luke Keele. 2016. “Interpreting regression discontinuity designs with multiple cutoffs.” *The Journal of Politics* 78(4):1229–1248.

Daly, Andrew and Stanley Zachary. 1978. “Improved multiple choice models.” *Determinants of travel choice* 335:357.

Definitions of Street Harassment. 2015.
URL: *http://www.stopstreetharassment.org/resources/definitions/*

Eide, Eric, Dominic J Brewer and Ronald G Ehrenberg. 1998. “Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance.” *Economics of Education Review* 17(4):371–376.

Fairchild, Kimberly and Laurie A Rudman. 2008. “Everyday stranger harassment and women’s objectification.” *Social Justice Research* 21(3):338–357.

Ferrant, Gaëlle, Luca Maria Pesando and Keiko Nowacka. 2014. “Unpaid Care Work: The missing link in the analysis of gender gaps in labor outcomes.”.

Gallego, Francisco A and Andrés Hernando. 2009. “School Choice in Chile: Looking at the Demand Side.”.

Hastings, Justine S, Thomas J Kane and Douglas O Staiger. 2009. Heterogenous Preferences and the Efficacy of Public School Choice. Technical report National Bureau of Economic Research.

Herriges, Joseph A and Catherine L Kling. 1996. "Testing the consistency of nested logit models with utility maximization." *Economics Letters* 50(1):33–39.

Horsman, Jenny. 2013. *Too scared to learn: Women, violence, and education*. Routledge.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones and Peter J Klenow. 2016. "The Allocation of Talent and U.S. Economic Growth." *Working Paper* .

Hsu, Hsin-Ping. 2011. "How Does Fear of Sexual Harassment on Transit Affect Women's Use of Transit?" *Women's Issues in Transportation* .

Keane, Carl. 1998. "Evaluating the influence of fear of crime as an environmental mobility restrictor on women's routine activities." *Environment and Behavior* 30(1):60–74.

Klasen, Stephan and Janneke Pieters. 2015. "What Explains the Stagnation of Female Labor Force Participation in Urban India?" *The World Bank Economic Review* 29(3):449–478.

Kling, Catherine L and Joseph A Herriges. 1995. "An empirical investigation of the consistency of nested logit models with utility maximization." *American Journal of Agricultural Economics* 77(4):875–884.

Kremer, Michael, Jessica Leino, Edward Miguel and Alix Peterson Zwane. 2011. "Spring cleaning: Rural water impacts, valuation, and property rights institutions." *The Quarterly Journal of Economics* 126(1):145–205.

Langton, Lynn and Jennifer L Truman. 2014. *Socio-emotional impact of violent crime*. US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Livingston, Beth. 2015. "Hollaback! International Street Harassment Survey Project.".
URL: *https://www.ihollaback.org/cornell-international-survey-on-street-harassment/*

Macmillan, Ross, Annette Nierobisz and Sandy Welsh. 2000. “Experiencing the streets: Harassment and perceptions of safety among women.” *Journal of Research in Crime and Delinquency* 37(3):306–322.

McFadden, Daniel. 1978. “Modeling the choice of residential location.” *Transportation Research Record* (673).

McFadden, Daniel and Kenneth Train. 2000. “Mixed MNL Models for Discrete Response.” *Journal of applied Econometrics* pp. 447–470.

McFadden, Daniel et al. 1977. *Quantitative methods for analyzing travel behavior of individuals: some recent developments*. Institute of Transportation Studies, University of California.

Mitra-Sarkar, Sheila and P Partheeban. 2011. “Abandon All Hope, Ye Who Enter Here: Understanding the Problem of "Eve Teasing" in Chennai, India.” 2(46).

Most dangerous transport systems for women. 2014.
URL: *http://news.trust.org//spotlight/most-dangerous-transport-systems-for-women/*

Muralidharan, Karthik and Nishith Prakash. 2017. “Cycling to School: Increasing Secondary School Enrolment for Girls in India.” *American Economic Journal: Applied Economics* 29(3):321–350.

Neill, Christine. 2009. “Tuition fees and the demand for university places.” *Economics of Education Review* 28(5):561–570.

Niederle, Muriel and Lise Vesterlund. 2007. “Do women shy away from competition? Do men compete too much?” *The Quarterly Journal of Economics* 122(3):1067–1101.

Niederle, Muriel and Lise Vesterlund. 2011. “Gender and competition.” *Annual Review of Economics* 3(1):601–630.

Pain, Rachel H. 1997. “Social geographies of women’s fear of crime.” *Transactions of the Institute of British geographers* pp. 231–244.

Pascarella, Ernest T and Patrick T Terenzini. 2005. *How college affects students*. Vol. 2 Jossey-Bass San Francisco, CA.

Rakodi, Carole. 2014. "Expanding Women's Access to Land and Housing in Urban Areas.".

Rubinstein, Yona and Sheetal Sekhri. 2013. "Value-Added versus Positive Sorting In Higher Education: Evidence from General Education Sector in India." *Working Paper* .

Sanders, Nicholas J. 2012. "What doesn't kill you makes you weaker prenatal pollution exposure and educational outcomes." *Journal of Human Resources* 47(3):826–850.

Satija, Shivani and Amrita Datta. 2015. "Crime against Women and Children in Delhi." *Economic & Political Weekly* 50(9).

Schneider, Mark and Jack Buckley. 2002. "What do parents want from schools? Evidence from the Internet." *Educational evaluation and policy analysis* 24(2):133–144.

Sekhri, Sheetal. 2014. "Prestige Matters: Value of Connections Formed in Elite Colleges." *Working Paper* .

Some Colleges Flouting Norms, Admits DU. 2013.
URL: *http://www.hindustantimes.com/education/some-colleges-flouting-norms-admits-du/story-wl80DrnA8aDOLMj9VDyQxH.html*

Sudarshan, Ratna M and Shrayana Bhattacharya. 2009. "Through the magnifying glass: Women's work and labour force participation in urban Delhi." *Economic and Political Weekly* pp. 59–66.

The Annual Report of the University of Delhi 2013-14. 2014. Technical report.

The Times Higher Education BRICS and Emerging Economies Rankings. 2015.
URL: *https://www.timeshighereducation.com/world-university-rankings/2015*

Three in Four Women Experience Harassment and Violence in UK and Global Cities. 2016.
URL: *https://www.actionaid.org.uk/latest-news/three-in-four-women-experience-harassment-and-violence*

Train, Kenneth. 2003. "Discrete Choice Methods with Simulation." *Cambridge University Press* .

Unsafe: An Epidemic of Sexual Violence in Delhi's Public Spaces: Baseline Findings from the Safe Cities Delhi Programme. 2013. Technical report UN Women and ICRW.

Viscusi, W Kip and Joseph E Aldy. 2003. "The Value of a Statistical Life: a Critical Review of Market Estimates Throughout the World." *Journal of Risk and Uncertainty* 27(1):5–76.

Viswanath, Kalpana and Ashish Basu. 2015. "SafetiPin: an innovative mobile app to collect data on women's safety in Indian cities." *Gender & Development* 23(1):45–60.

Warr, Mark. 1985. "Fear of rape among urban women." *Social problems* 32(3):238–250.

Winston, Gordon and David Zimmerman. 2004. Peer effects in higher education. In *College choices: The economics of where to go, when to go, and how to pay for it*. University of Chicago Press pp. 395–424.

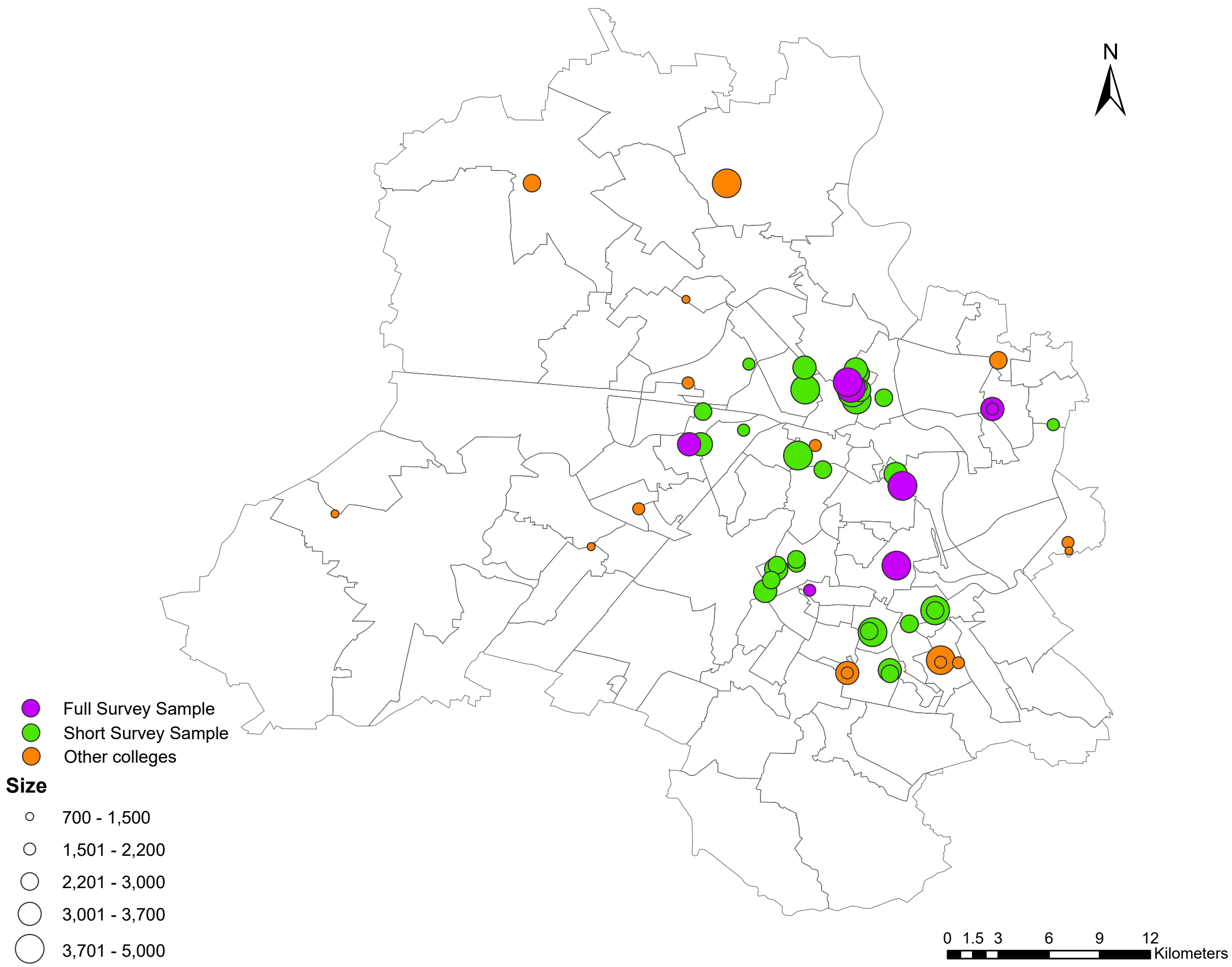
With Hostel Shortage in Delhi University, Students Demand Implementation of Rent Act. 2015. [Online, posted June 14, 2015].
URL: *https://www.ndtv.com/delhi-news/not-enough-hostel-seats-in-delhi-university-students-clamour-for-rent-act-771460*

World Employment Social Outlook - Trends for Women. 2017. Technical report International Labour Office, Geneva.

Zhang, Liang. 2005. "Advance to graduate education: The effect of college quality and undergraduate majors." *The Review of Higher Education* 28(3):313–338.

Figures

Figure 1: Colleges in Delhi University

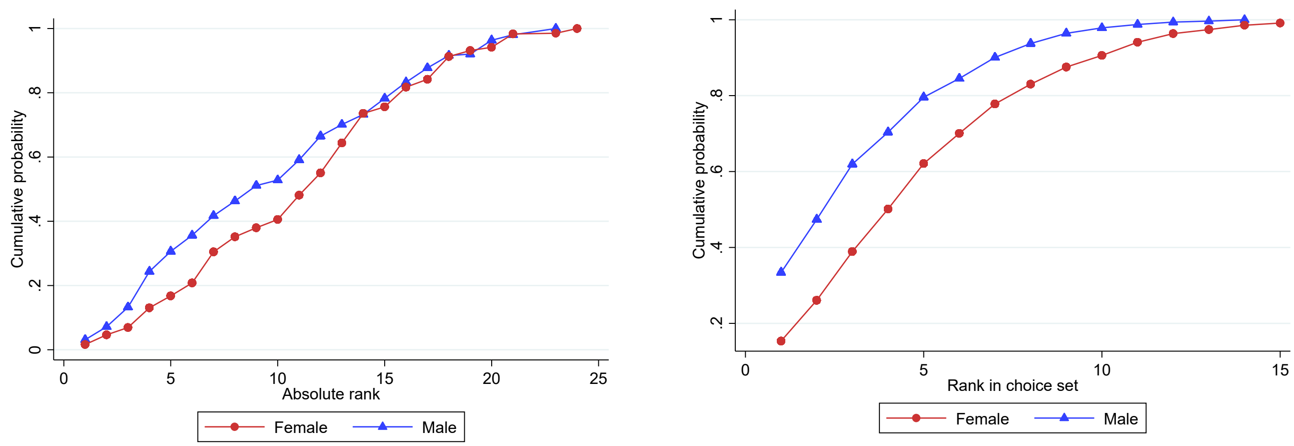


Notes: The map shows the 58 general education colleges in DU. Eight colleges are in the full survey sample and 32 colleges are in the short survey sample.

Figure 2: Cumulative Distribution Function of Rank for Women and Men

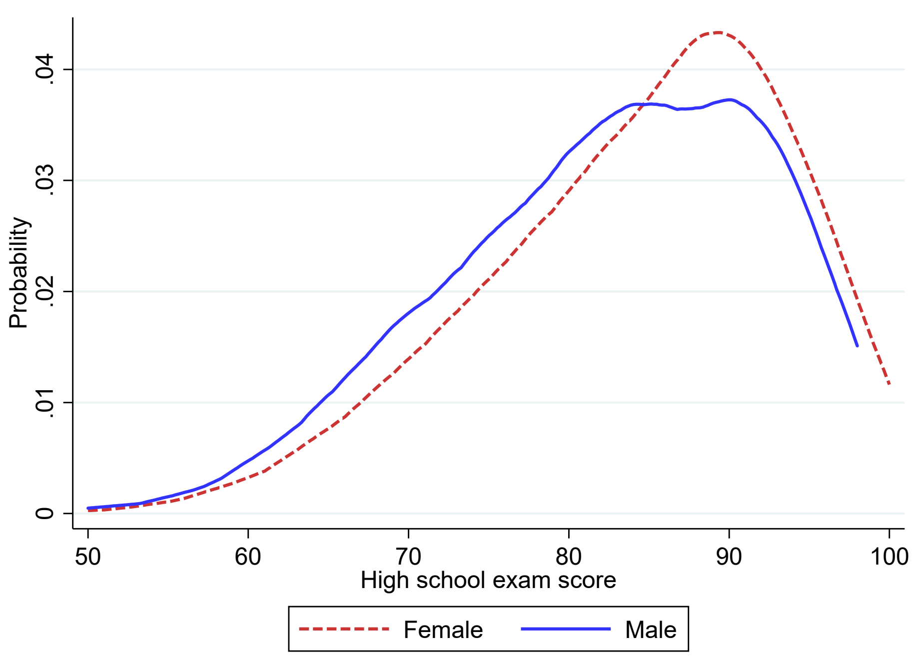
(a) Absolute rank

(b) Rank within choice set



Notes: Rank is based on cutoff scores of a college for the student's major and admission year from the first cutoff list for general category male students. A higher rank indicates a lower cutoff score or worse quality. The absolute rank in Panel (a) ranks college within a major and admission year using the first cutoff list for that year. Rank within a student's choice set in Panel (b) ranks the colleges that the student was eligible to attend, by their cutoff score for the student's major and admission year. The CDF is for colleges chosen by students in the full survey sample and short survey sample, who are Delhi residents and live at home.

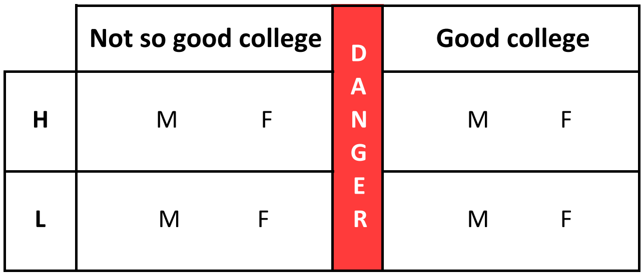
Figure 3: High School Exam Scores for Women and Men



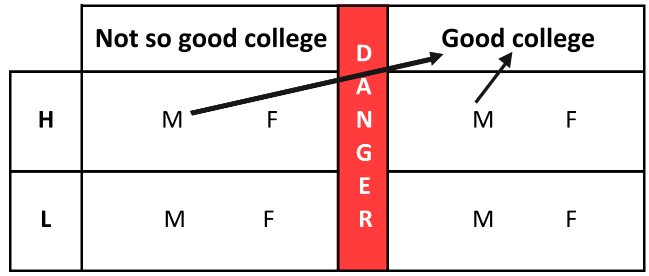
Notes: The figure shows the probability distribution function of high school exam scores for students in the full survey sample and short survey sample, who are Delhi residents and live at home.

Figure 4: Stylized Example

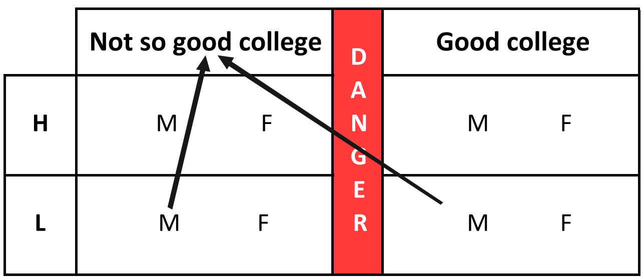
(a) The set-up



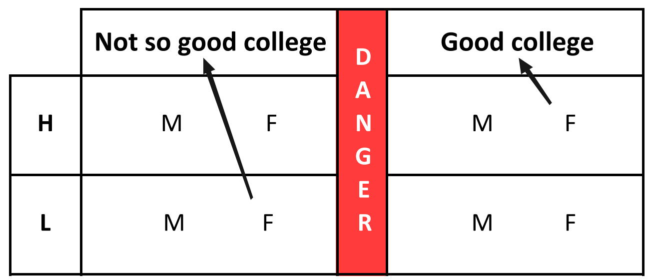
(b) High-scoring males



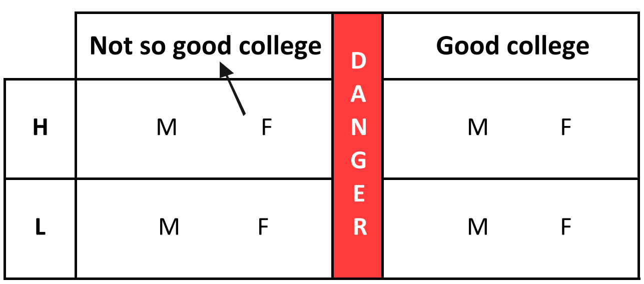
(c) Low-scoring males



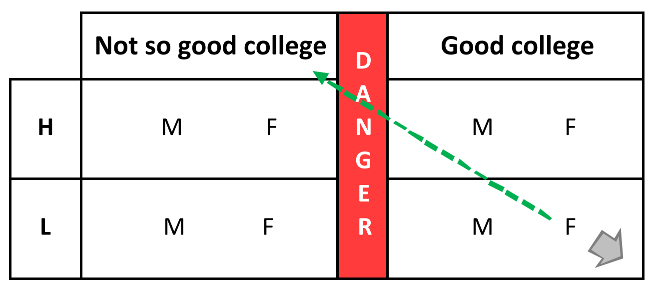
(d) Females who do not face a safety-quality trade-off



(e) High-scoring female chooses lower quality college

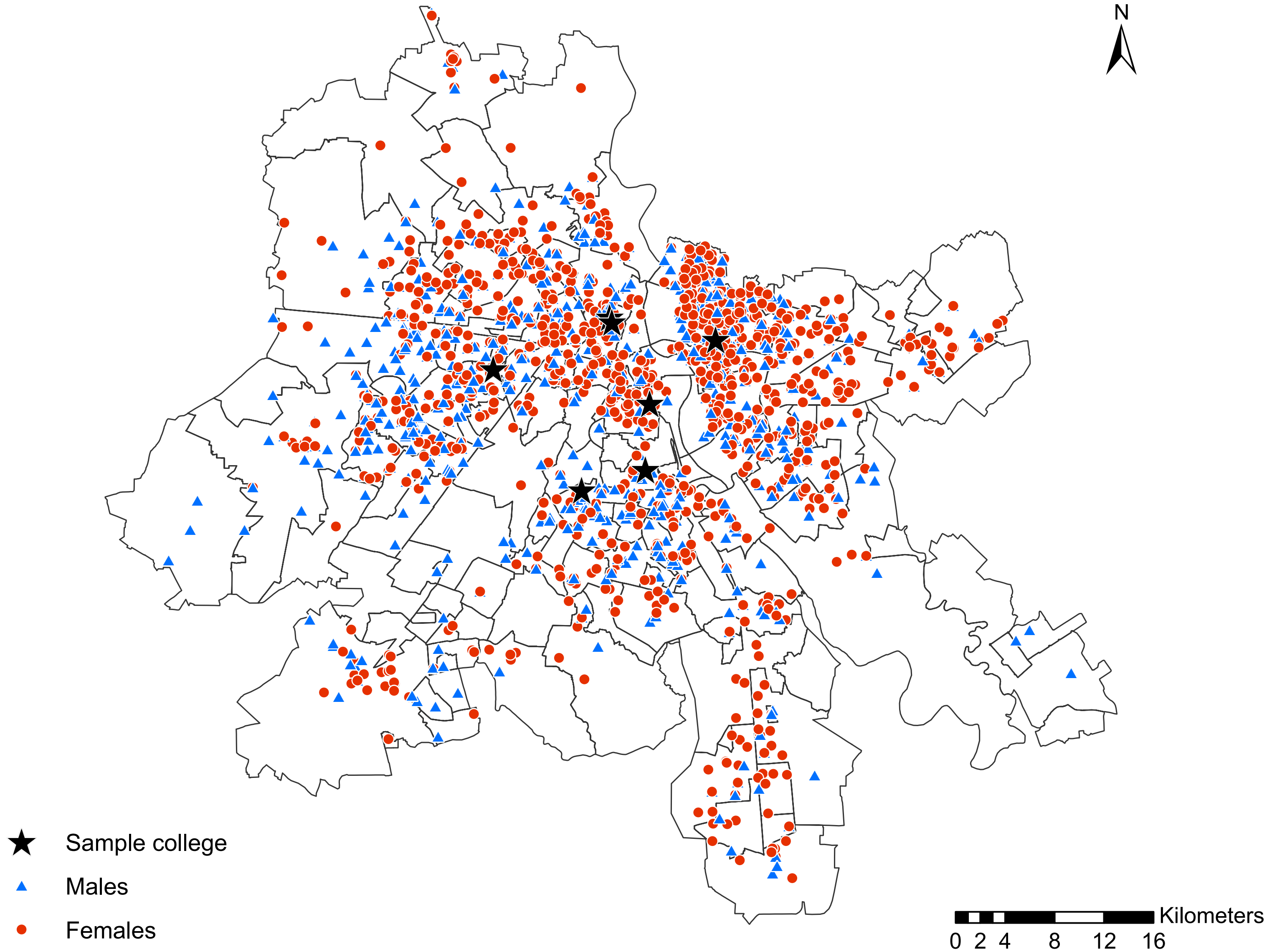


(f) Female uses more expensive mode of transport



Notes: This figure shows the college and route choice of students by gender and high school exam scores. H and L denote students' high school exam scores. M and F denote a male and female student respectively. The thin arrows denote a travel route. A route is considered unsafe if it crosses the red "danger" area. The green dashed arrow denotes a route using a more expensive mode of transport. The thick grey arrow denotes the choice of not attending college at all.

Figure 5: Students in Full Survey Sample

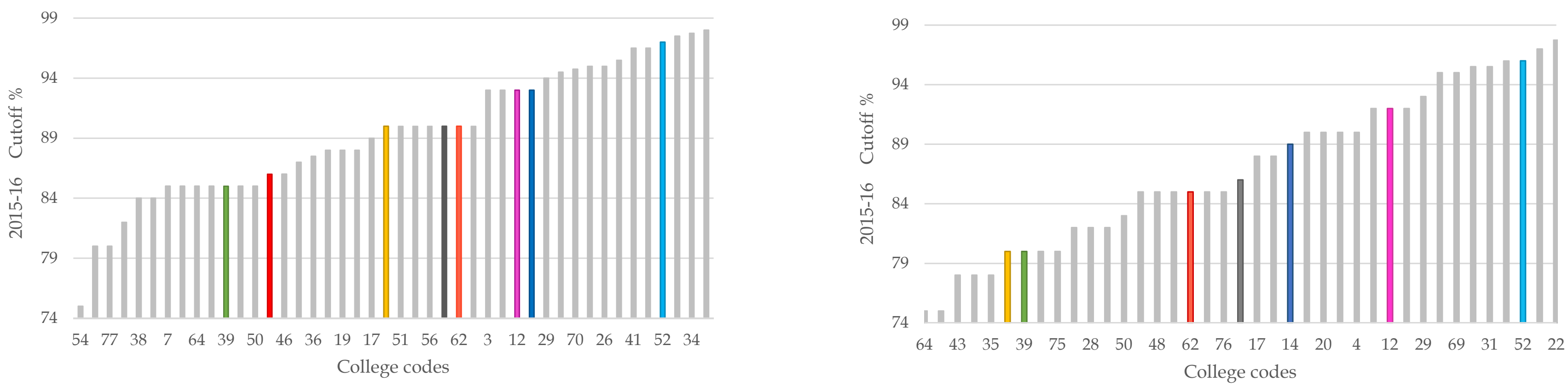


Notes: This map shows the residential location of students who are Delhi residents from the full survey sample, who live at home and travel to college everyday.

Figure 6: Variation in Quality of the Full Survey Sample Colleges

(a) Political Science

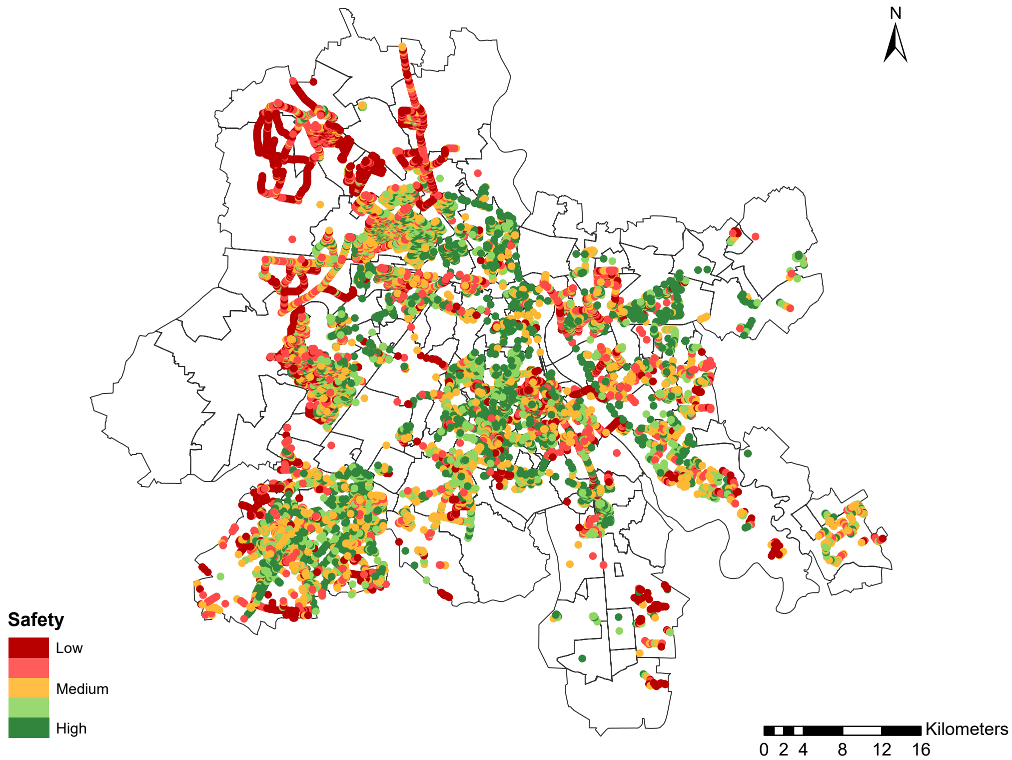
(b) History



Notes: This figure shows the cutoff scores for male, general category students from the first cutoff list in 2015-16. Panel (a) shows the cutoffs for Bachelor of Arts in Political Science and Panel (b) shows the cutoffs for Bachelor of Arts in History. Alternate bars are labelled.

Figure 7: Area Safety Data from SafetiPin

(a) Safety Audit Data



(b) Safety Surface

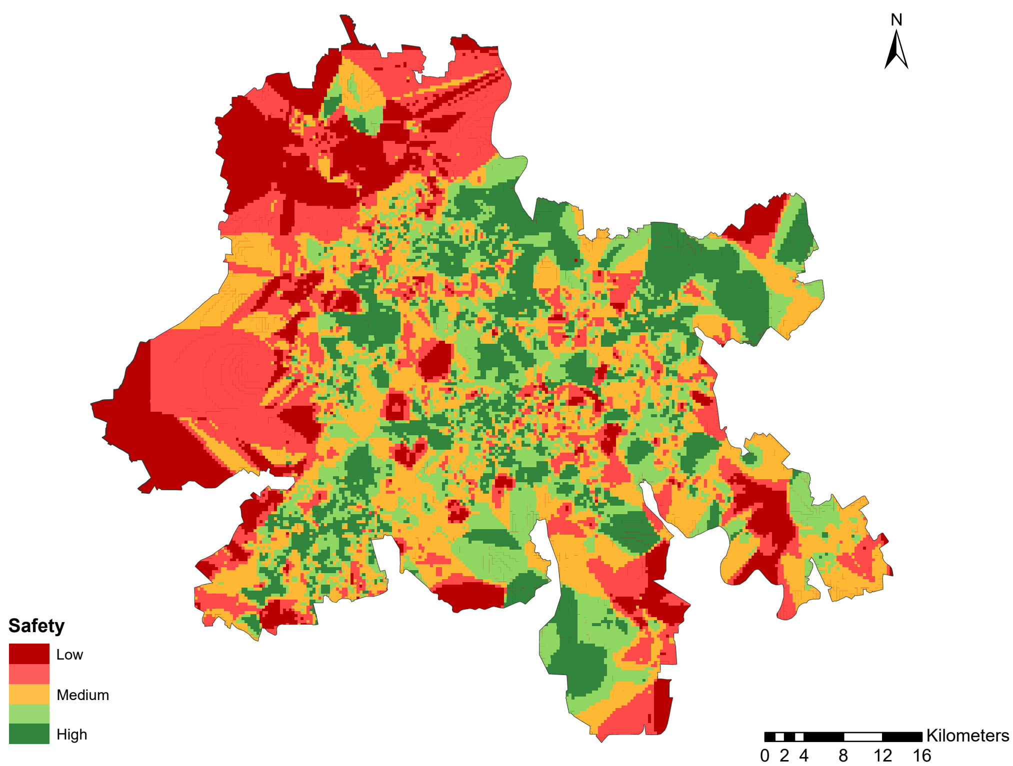
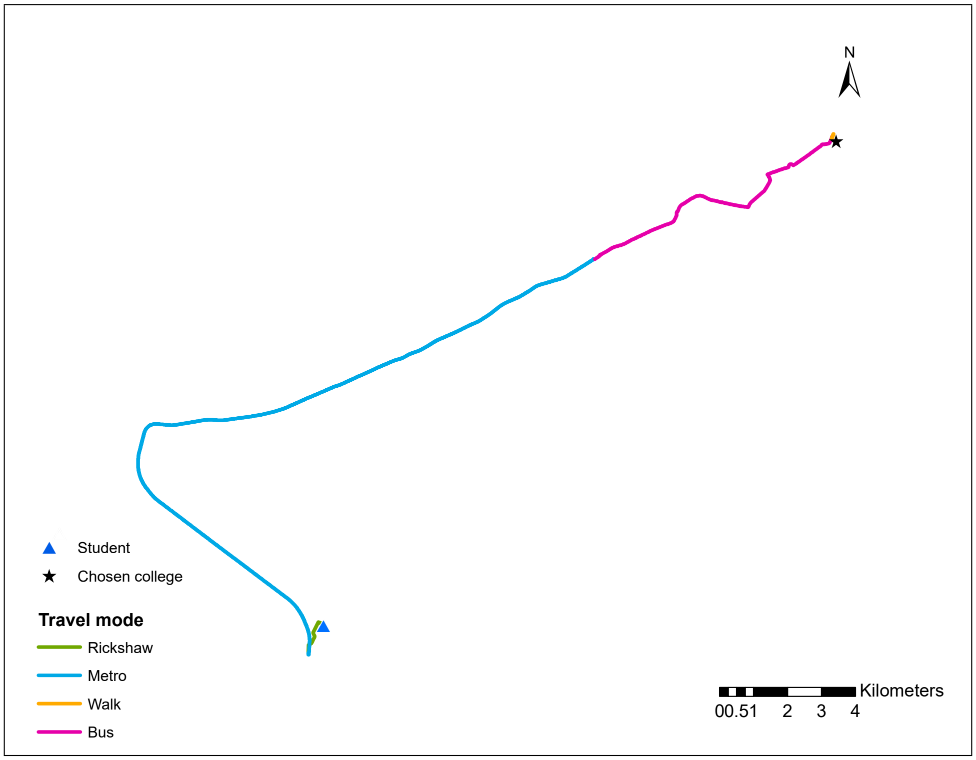
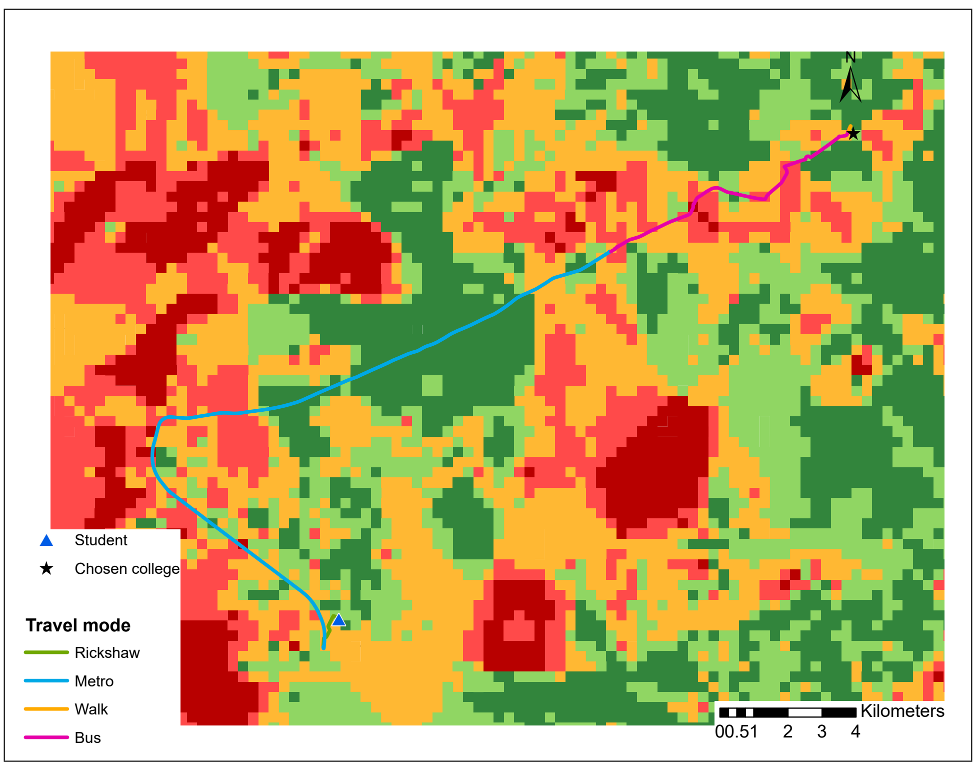


Figure 8: Calculating Route Safety

(a) Example Travel Route



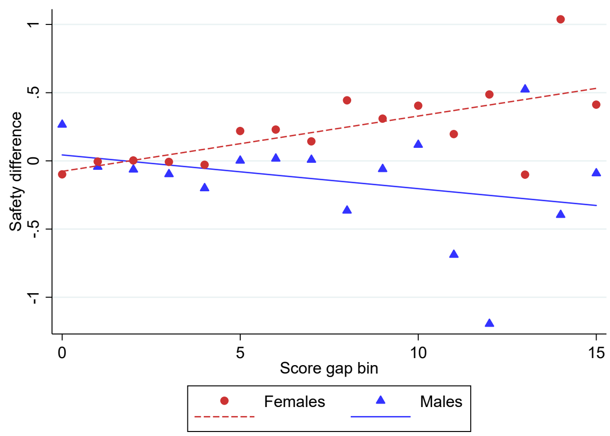
(b) Travel Route Over the Safety Surface



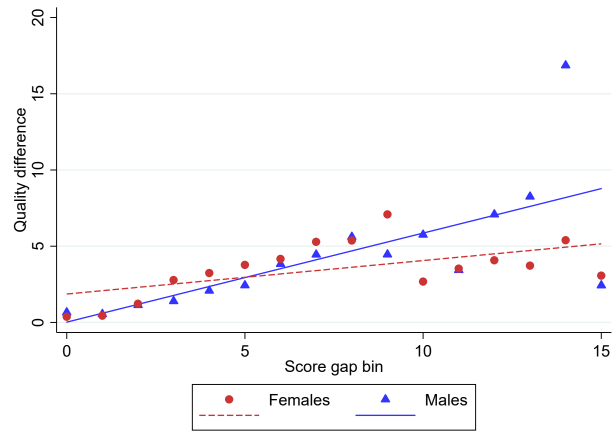
Notes: These maps exhibit how safety score of a travel route is computed as a combination of area safety from the SafetiPin mobile app data and harassment by mode from the Safecity data.

Figure 9: Choice Relative to Neighbors

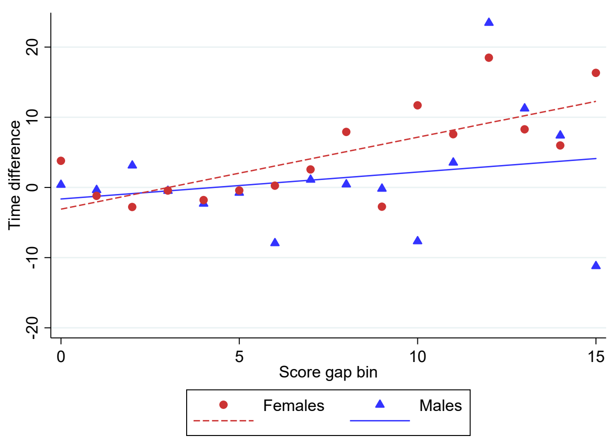
(a) Safety difference



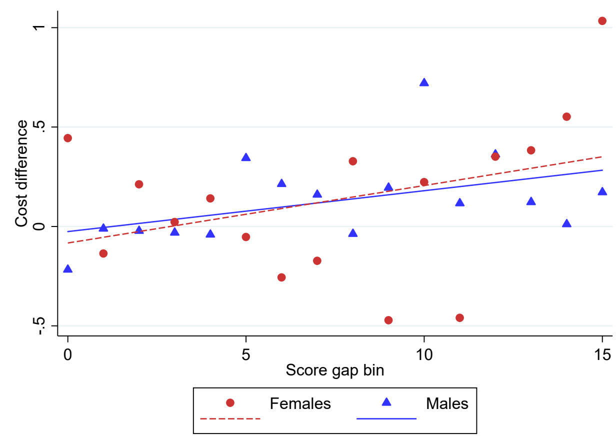
(b) Quality difference



(c) Time difference

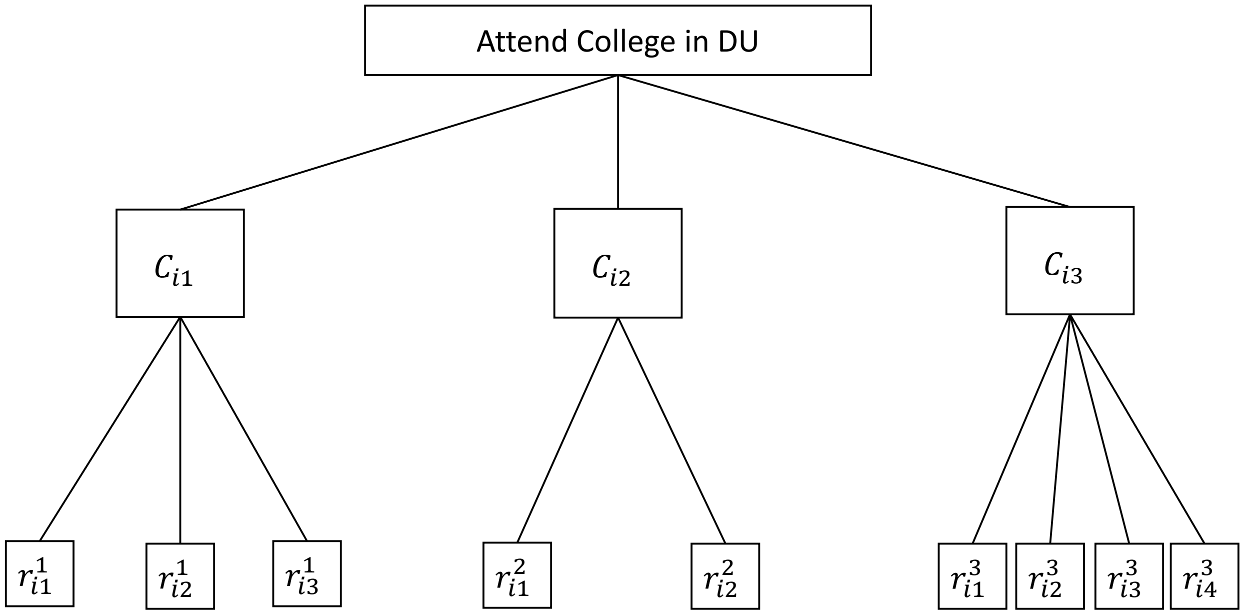


(d) Cost difference



Notes: The figures plot binned scatter plots of difference in travel safety, college quality, travel time, and travel cost between the index student and their neighbor's choice. Index student is the student who scores higher. Score gap bin is the two point bin of score difference between the index student and the neighbor. A neighbor is a student living within a 1.5 km radius of the index student and has the same gender, major, and admission year.

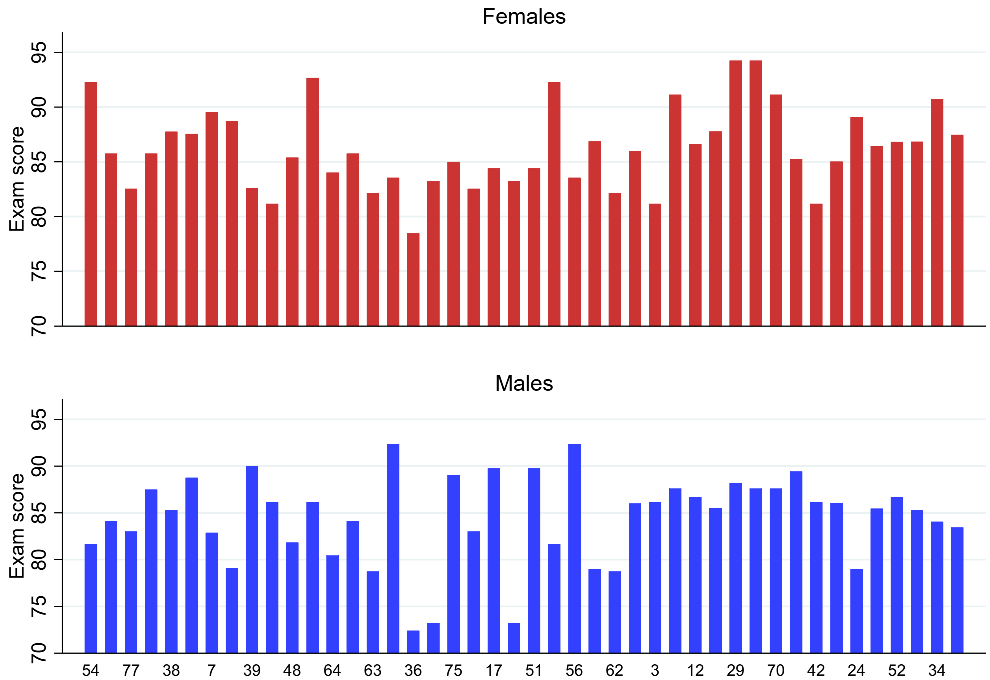
Figure 10: Student Choice Model



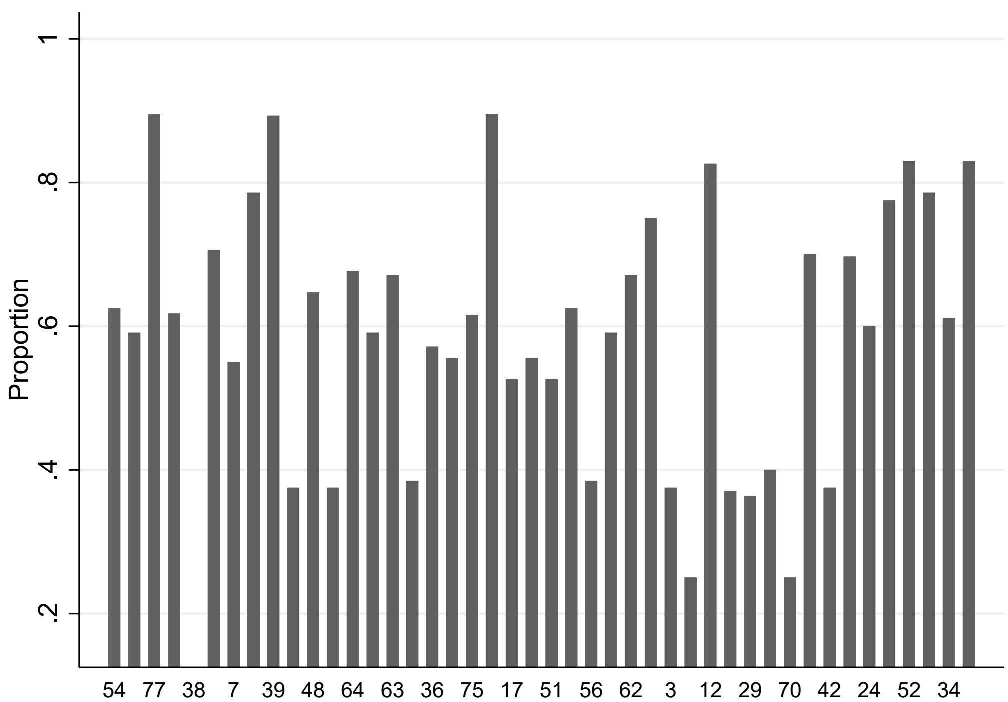
Notes: Nests are colleges $C_{i1}, C_{i2}, and C_{i3}$. Routes to each college are given by r_{iR}^c

Figure 11: Variation of Student and Area Characteristics Around Colleges

(a) Average high school exam score by gender



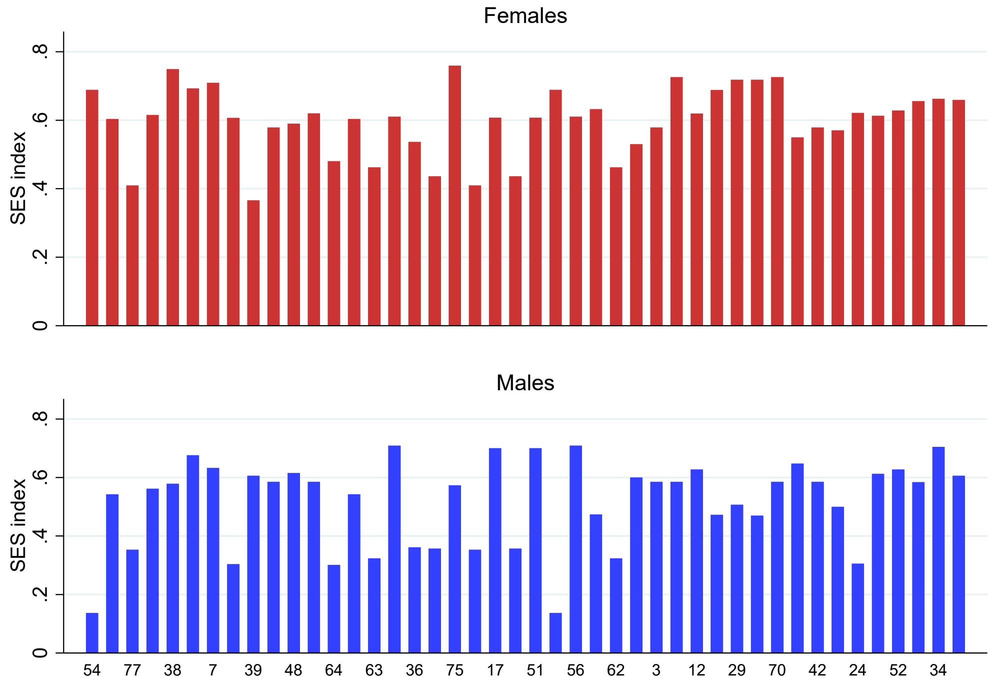
(b) Proportion of female students

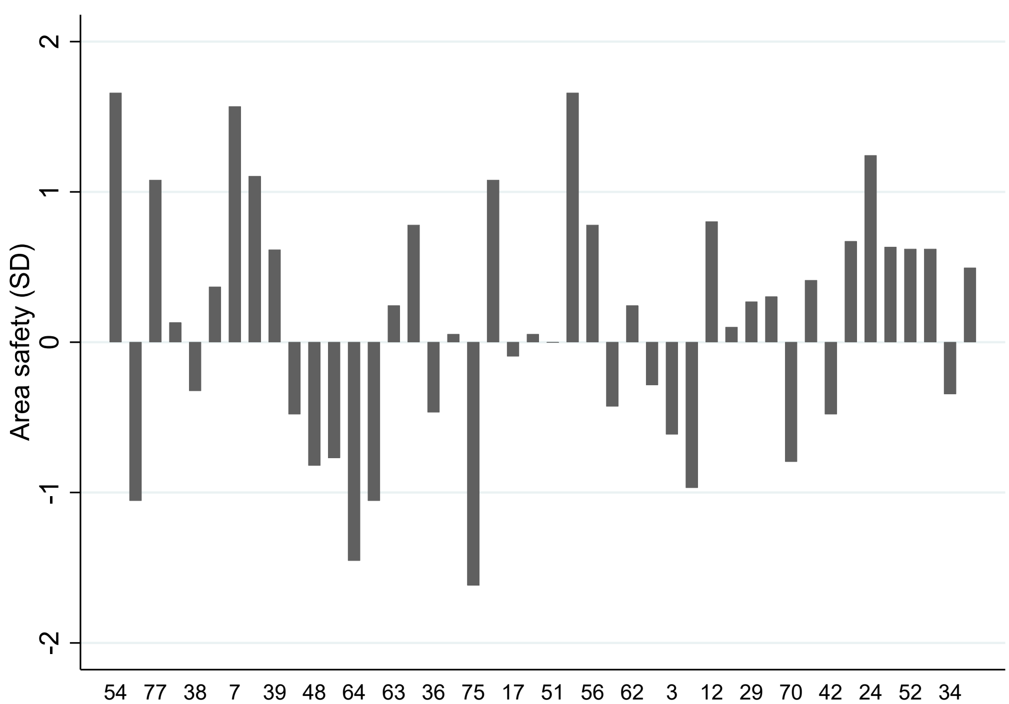


Notes: The figures show characteristics of students living within a 1.5km radius around each general education college in DU that offers a Bachelor of Arts in Political Science. Each bar reprsents a college. The colleges are in ascending order of quality. The quality measure used here is the cutoff scores for Bachelor of Arts in Political science applicable for male, general category students from the first cutoff list in 2015-16, as shown in Figure 6(a). Alternate bars are labelled.

Figure 11: Variation of Student and Area Characteristics Around Colleges

(c) Average socio-economic status of students by gender





(d) Average area safety

Notes: The figures show characteristics of students living within a 1.5km radius around each general education college in DU that offers a Bachelor of Arts in Political Science. Each bar reprsents a college. The colleges are in ascending order of quality. The quality measure used here is the cutoff scores for Bachelor of Arts in Political science applicable for male, general category students from the first cutoff list in 2015-16, as shown in Figure 6(a). Alternate bars are labelled.

Tables

Table 1: Comparison of Student Characteristics

	Female			Male		
Data	Full Survey	Short Survey	Admin. Data	Full Survey	Short Survey	Admin. Data
Delhi residents	1,757	454	11,450	938	169	8,288
Proportion of surveyed	0.74	0.82	0.80	0.67	0.62	0.68
Social Category						
General	0.75	0.70	0.73	0.56	0.59	0.51
		[0.97]	[1.30]		[0.37]	[2.56]
SC	0.12	0.11	0.14	0.20	0.14	0.19
		[0.37]	[2.43]		[1.20]	[0.42]
ST	0.01	0.01	0.02	0.02	0.02	0.03
		[1.10]	[2.36]		[0.20]	[1.61]
OBC	0.12	0.18	0.11	0.22	0.25	0.27
		[1.64]	[0.52]		[1.03]	[3.37]
High school exam score (%)	84.37	83.89	–	82.42	82.01	–
	[1.01]			[0.51]		
Distance to college (kms.)	13.11	11.96	13.47	13.14	15.58	14.50
		[1.03]	[0.43]		[1.42]	[1.48]
Distance to center (kms.)	15.50	11.93	15.08	16.04	15.58	16.64
		[4.05]	[1.32]		[0.27]	[0.90]

Notes: Distance to college is the shortest travel distance to the chosen college. Distance to city center is the shortest distance to India Gate, a central landmark in Delhi. Test statistic for two sample t-tests is reported in brackets where sample mean for each data set is compared with sample mean of full survey data. General = General unreserved category SC= Scheduled Caste, ST = Scheduled Tribe, OBC = Other Backward Castes. These social categories are officially designated groups used for the purposes of affirmative action.

Table 2: Travel Mode Choice and Harassment

	Total (1)	Female (2)	Male (3)	(Female - Male) (4)	Harassment Incidents (5)
Auto Rickshaw	0.31	0.36	0.22	0.14***	0.07
Bus	0.38	0.33	0.48	-0.15***	0.40
Car or Uber	0.08	0.07	0.09	-0.02*	0.14
Metro	0.42	0.49	0.29	0.20***	0.16
Ladies compartment		0.85			
Train	0.03	0.03	0.03	0.00	0.05
Walk	0.68	0.66	0.71	-0.04**	–

*Notes:*Sample is Delhi residents from the full survey who live with their family. Harassment by travel mode is from Safecity analytical data. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: Summary Statistics

	Total (1)	Female (2)	Male (3)	(Female - Male) (4)
A. Student Characteristics				
Delhi resident sample	2,695	1,757	938	
Proportion of surveyed	0.71	0.74	0.67	
Proportion of Delhi residents	0.99	0.99	0.99	
Females	0.65			
High school exam score (%)	83.69	84.37	82.42	1.95***
SES index	0.49	0.53	0.43	0.10***
B. College Characteristics				
First cutoff score	87.98	87.59	88.70	1.11***
Absolute Rank	10.85	11.14	10.31	0.83***
Rank in choice set	4.52	5.02	3.59	1.43***
Distance (kms.)	13.12	13.10	13.14	0.04
Annual Tuition (thousand Rs.)	10.57	10.86	10.03	0.83***
Size of college	3,569	3,763	3,206	557***
Number of majors	13.69	14.11	12.91	1.20***
Boarding College	0.21	0.26	0.12	0.14***
Women only		0.44		
C. Route Characteristics				
Route Safety (SD)	0.42	0.52	0.23	0.29***
Monthly Travel cost (thousand Rs.)	1.39	1.54	1.12	0.42***
Travel time (mins.)	67.50	66.37	69.61	-3.23**

Notes: Sample is Delhi residents from the full survey who live with their families. The cutoff score is the first cutoff score for male general category students. Annual college fees are for 2016 admissions. SES index is calculated using principal component analysis using information on whether a student's family is a home owner, student has own laptop or computer, the number of cars, scooters and motorcycles in household, price of the most expensive car in household, student's monthly expenses excluding travel expenses, whether student attended private school, and parents' years of education. Route characteristics are for the student reported travel route. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Bechmark Specification: Nested Logit Model Results

	Female (1)	Male (2)	Female - Male (3)	SE of Difference (4)
Panel A: Benchmark specification				
Route safety (SD)	0.695	0.168	0.527	(0.0948)
Cutoff score	0.039	0.090	-0.051	(0.0295)
Monthly travel cost (thousand Rs.)	-0.306	-1.267	0.961	(0.1058)
Route time (mins.)	-0.013	-0.033	0.020	(0.0044)
MRS (Safety, Score)	13.04	1.37	11.67	
Percentage points per SD of safety				
MRS (Safety, Cost)	-1.67	-0.10	-1.57	
Thousand Rs. per SD of safety				
MRS (Safety, Time)	-39.24	-3.73	-35.51	
Minutes per SD of safety				
Panel B: With additional controls				
Route safety (SD)	0.798	0.211	0.587	(0.2135)
Cutoff score	0.067	0.094	-0.028	(0.0567)
Monthly travel cost (thousand Rs.)	-0.337	-1.134	1.008	(0.2054)
Daily travel time (mins.)	-0.013	-0.036	0.024	(0.0094)
MRS (Safety, Score)	8.85	1.65	7.20	
Percentage points per SD of safety				
MRS (Safety, Cost)	-1.75	-0.12	-1.64	
Thousand Rs. per SD of safety				
MRS (Safety, Time)	-46.77	-4.30	-42.47	
Minutes per SD of safety				
Observations	111,190	54,434		

Notes: Estimation allows for different inclusive value coefficients for each college. Additional controls are annual college fees, college's neighborhood safety, number of majors offered by college, indicator for boarding facilities, total number of students enrolled in college, and an indicator for women only colleges. All coefficients are significant at conventional levels of significance. All additional controls are not significantly different from zero at conventional levels of significance. MRS are measured in terms of the SD of route safety across the predicted route alternatives *within* a students' choice set, which is 26.5 percent lower than the overall SD in route safety in Panel A and 26 percent lower in Panel B.

Table 5: Results from the Mixed Logit Model of College and Route Choice

Variable	Parameter	Female (1)	Male (2)
Preferences for travel safety			
Route safety (SD)	Mean	0.768	0.482
		(0.059)	(0.091)
	St. Dev.	0.594	0.611
Route safety × SES	Mean	0.251	0.332
		(0.068)	(0.091)
	St. Dev.	–	–
Route safety × Exam score	Mean	0.093	0.081
		(0.060)	(0.067)
	St. Dev.	–	–
Preferences for college quality			
Cutoff score	Mean	0.038	0.229
		(0.017)	(0.029)
	St. Dev.	0.064	0.176
Cutoff score × SES	Mean	0.030	0.127
		(0.018)	(0.029)
	St. Dev.	–	–
Cutoff score × Exam score	Mean	0.022	0.033
		(0.008)	(0.010)
	St. Dev.	–	–
Other preferences			
Monthly travel cost (thousand Rs.)	Mean	-0.250	-0.330
		(0.027)	(0.042)
	St. Dev.	–	–
Daily travel time (mins.)	Lognormal Mean	-0.011	-0.019
	St. Dev.	0.010	0.033
	Normal Mean	-4.790	-4.701
		(0.462)	(0.419)
Estimated correlation coefficients			
Safety, Cutoff score		-0.660	0.714
Safety, Time		0.764	0.950
Cutoff score, Time		-0.106	0.896
Observations		111,190	54,434

Notes: Estimates generated by Simulated Maximum Likelihood using 200 Halton draws. Estimates for mean and standard deviations of distribution of preferences for route and college characteristics. Monthly travel cost is fixed. Distribution of preferences on time is constrained to follow a negative lognormal distribution. All coefficients are significant at conventional levels of significance, other than the coefficient for Route Safety × Exam Score for both females and males.

Table 6: Willingness to Pay Estimates from the Mixed Logit Model

	Female (1)	Male (2)	Female - Male (3)
MRS (Safety, Score) pp per SD of safety	9.06	0.96	8.10
MRS (Safety, Cost) '000 Rs. per SD of safety	-1.36	-0.67	-0.70
MRS (Safety, Time) Minutes per SD of safety	-31.57	-11.85	-19.71

Notes: MRS are measured in terms of the SD of route safety *within* a student's choice set, which is 44.4 percent lower than the overall SD in route safety for females and 45.7 percent lower for males.

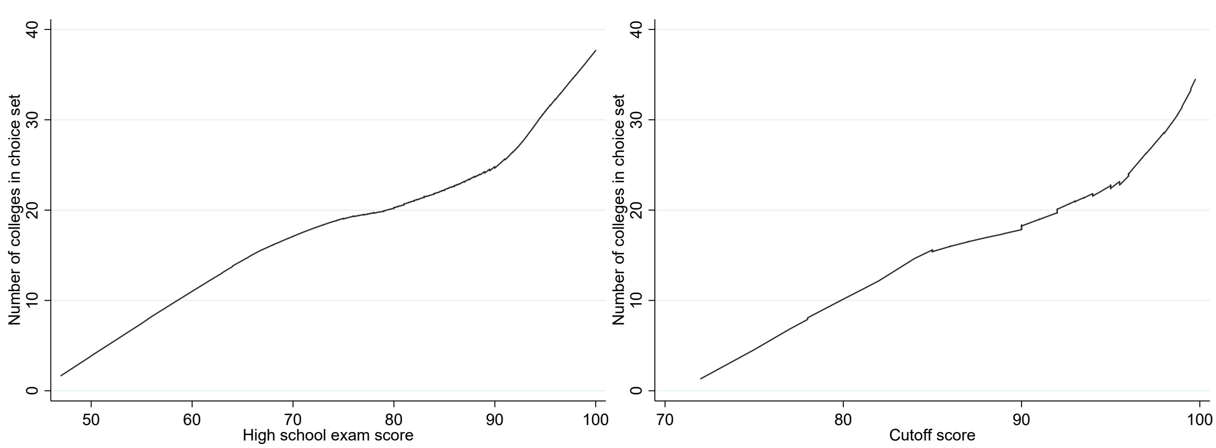
Appendix

Figure A1: Sample Cutoff List

S. No.	COURSE	GENERAL	OBC	SC	ST	PH	Remarks
1	B.A. Prog.	82%	78%	72%	72%	70%	2% less for girls.
2	B.A. (H)Economics	95%	92%	90%	85%	88%	1% less for Girls.
3	B.A.(H) English	91%	86%	82%	82%	80%	----------------------------------
4	B.A.(H) Hindi	75%	70%	65%	65%	65%	3% less for girls.
5	B.A.(H) History	85%	80%	78%	78%	78%	2% less for girls
6	B.A.(H) Pol. Science	85%	83%	80%	80%	80%	2% less for girls
7	B.A.(H) Sanskrit	70%	68%	65%	65%	65%	3% less for Girls.
8	B. Com. (H)	94%	90%	85%	85%	80%	2% less for girls.
9	B.Sc. (H) Chemistry	92% PCM	90% PCM	85% PCM	80% PCM	80% PCM	----------------------------------
10	B.Sc.(H) Electronics	90% PCM	87% PCM	80% PCM	70% PCM	70% PCM	3% less for Girls.
11	B.Sc. (H) Mathematics	92% (One Language, Mathematics and best two elective subjects)	87% (One Language, Mathemati cs and best two elective subjects)	80% (One Language, Mathemati cs and best two elective subjects)	80% (One Language, Mathemati cs and best two elective subjects)	80% (One Language, Mathematics and best two elective subjects)	2% less for Girls. (One Language, Mathematics and best two elective subjects)

Notes: This is scanned copy of a sample cutoff list from a college in DU for 2015-16. The rows specify the major, and the columns show the category of admission. Each cell has the relevant cutoff score. The remarks specify additional conditions.

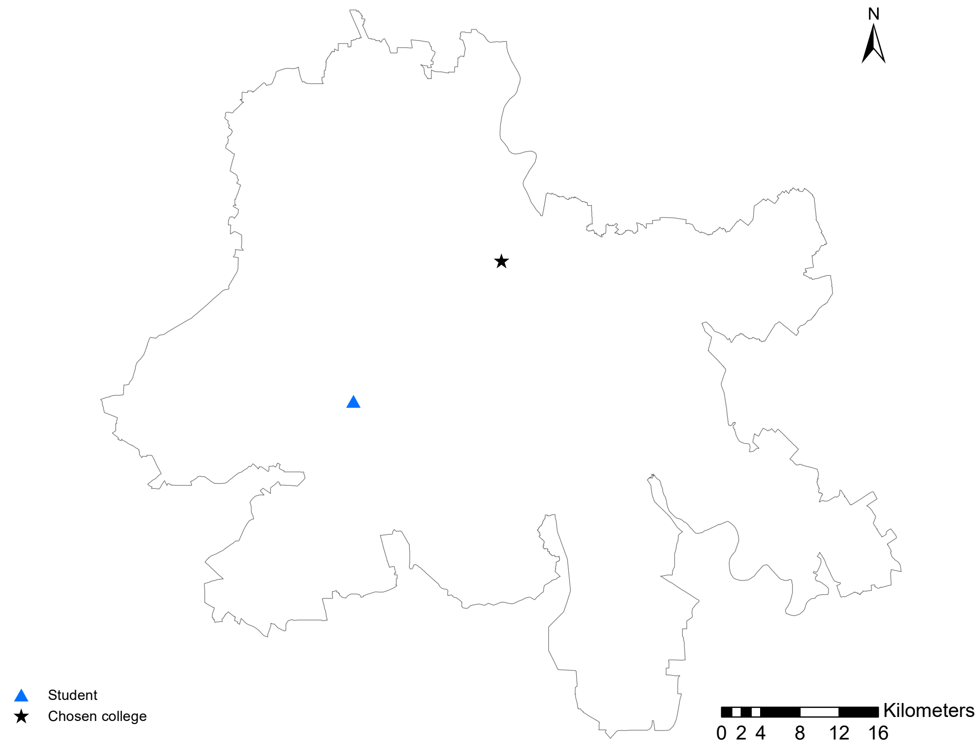
Figure A2: Correlation Between the Number of Choice Set Colleges and Students High School Exam Score



Notes: The figure shows the correlation between the number of colleges in a student's choice set and their high school exam score in the left panel, and the cutoff score for their chosen college in the right panel.

Figure A3: An example of route mapping

(a) Student and chosen college



(b) Reported route to chosen college

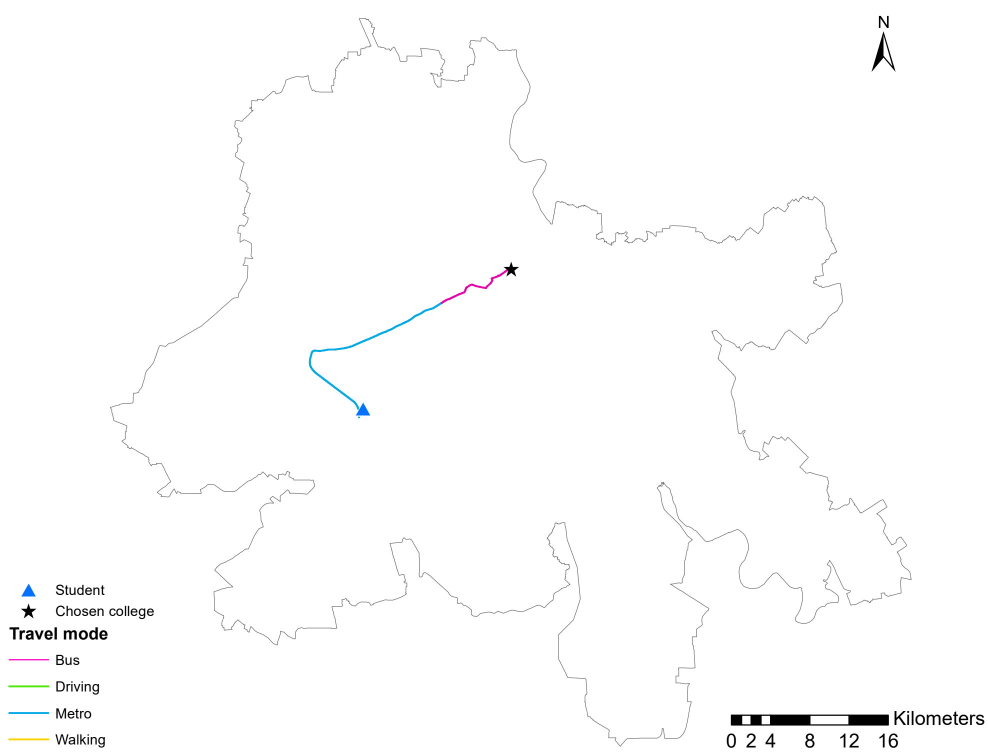
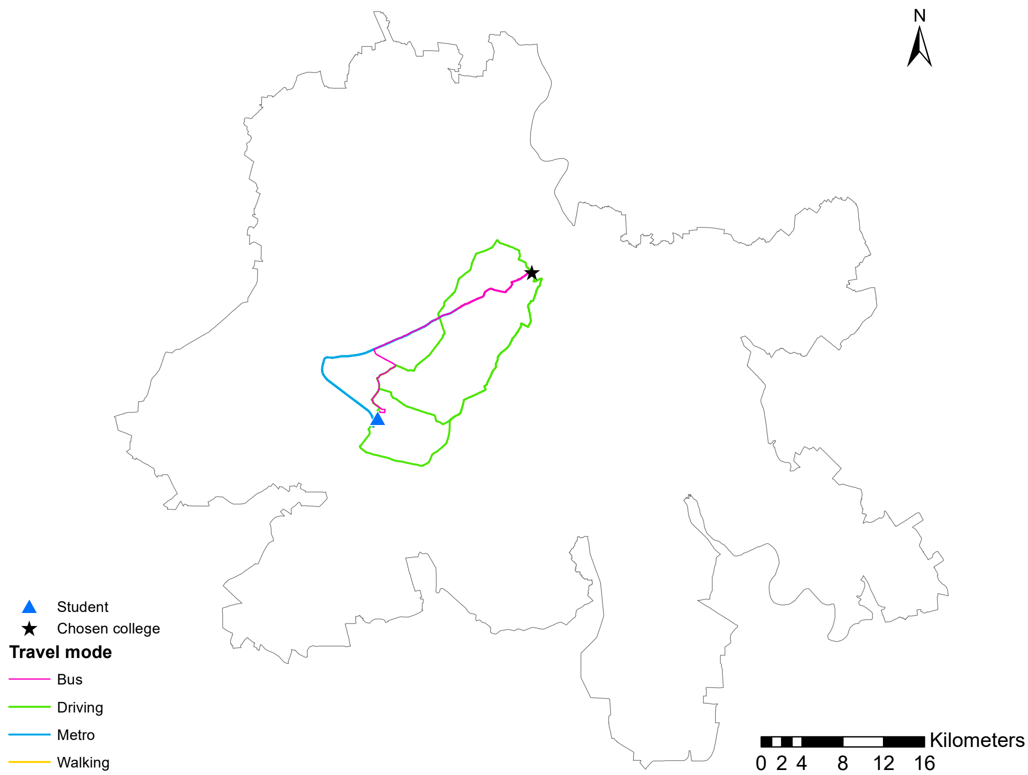


Figure A3: An example of route mapping

(c) Route options to chosen college



(d) Route options to colleges in choice set

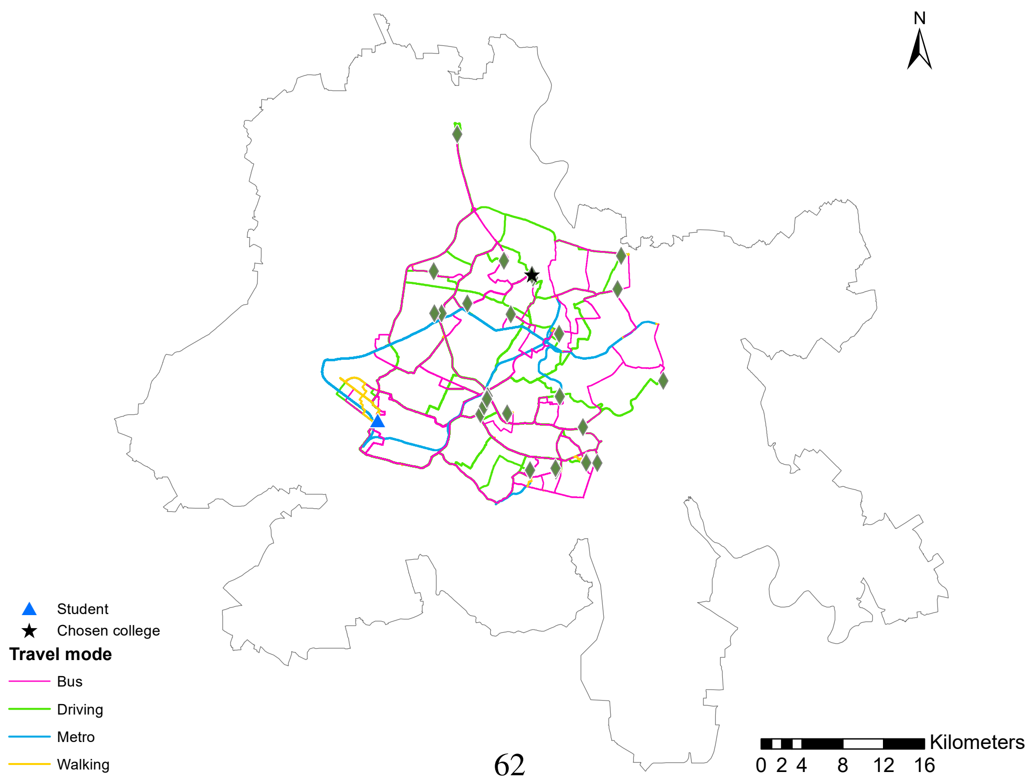
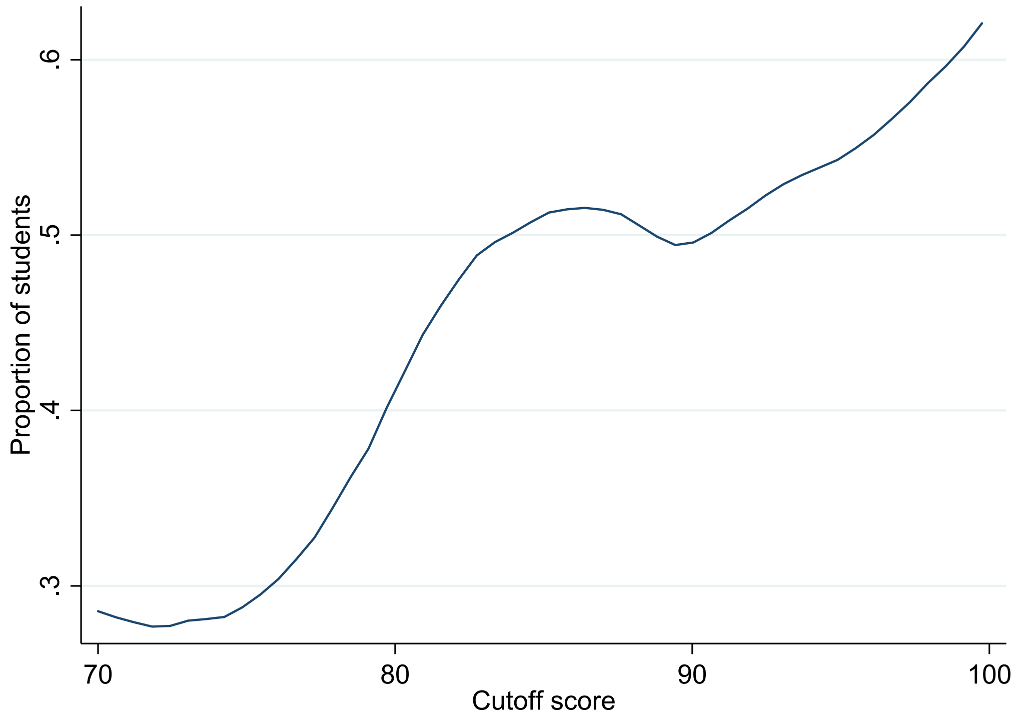


Figure A4: Proportion of Accepted Students who Enrolled



Notes: The figure shows the relationship between the proportion of accepted male students who enrolled and the cutoff score of the college. Results are from a local polynomial regression with a bandwidth of 3.5.

Table A1: SafetiPin Parameters

		0	1	2	3
1	Light (Night)	**None.** No street or other lights.	**Little.** Can see lights, but barely reaches this spot.	**Enough.** Lighting is enough for clear visibility.	**Bright.** Whole area brightly lit.
2	Openness	**Not Open.** Many blind corners and no clear sightline.	**Partly Open.** Able to see a little ahead and around.	**Mostly Open.** Able to see in most directions.	**Completely Open.** Can see clearly in all directions.
3	Visibility	**No Eyes.** No windows or entrances (to shops or residences) or street vendors.	**Few Eyes.** Less than 5 windows or entrances or street vendors.	**More Eyes.** Less than 10 windows or entrances or street vendors.	**Highly Visible.** More than 10 windows or entrances or street vendors.
4	People	**Deserted.** No one in sight.	**Few People.** Less than 10 people in sight.	**Some Crowd.** More than 10 people visible.	**Crowded.** Many people within touching distance.
5	Security	**None.** No private security or police visible in surrounding area.	**Minimal.** Some private security visible in surrounding area but not nearby.	**Moderate.** Private security within hailing distance.	**High.** Police within hailing distance.
6	Walk Path	**None.** No walking path available.	**Poor.** Path exists but in very bad condition.	**Fair.** Can walk but not run.	**Good.** Easy to walk fast or run.
7	Public Transport	**Unavailable.** No metro or bus stop, auto rickshaw or cycle rickshaw within 10 minutes walk.	**Distant.** Metro or bus stop, auto rickshaw or cycle rickshaw within 10 minutes walk.	**Nearby.** Metro or bus stop, auto rickshaw or cycle rickshaw within 2-5 minutes walk.	**Very Close.** Metro or bus stop, auto rickshaw or cycle rickshaw within 2 minutes walk.
8	Gender Usage	**Not Diverse.** No one in sight, or only men.	**Somewhat Diverse.** Mostly men, very few women or children.	**Fairly Diverse.** Some women or children.	**Diverse.** Balance of all genders or more women and children.
9	Feeling	**Frightening.** Will never venture here without sufficient escort.	**Uncomfortable.** Will avoid whenever possible.	**Acceptable.** Will take other available and better routes when possible.	**Comfortable.** Feel safe here even after dark.

Table A2: Changes in students' choice attributes with a change in their relative test score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Safety difference		Quality difference		Time difference		Cost difference	
Score difference	-0.0056	-0.0086	0.270***	0.172***	-0.0347	-0.404**	0.021**	-0.0024
	(0.00670)	(0.00801)	(0.0263)	(0.0226)	(0.137)	(0.163)	(0.00891)	(0.00977)
Score difference x Fem	0.0236***	0.0244***	-0.0650**	-0.0240	0.391**	0.458**	-0.0181*	0.0098
	(0.00813)	(0.00942)	(0.0320)	(0.0266)	(0.166)	(0.191)	(0.0108)	(0.0115)
Female	-0.0210		0.736**		-2.329		0.0903	
	(0.0810)		(0.318)		(1.657)		(0.108)	
Constant	-0.0278	-0.0260	0.279	1.282***	-0.0979	0.512	-0.0872	0.0108
	(0.0691)	(0.0345)	(0.272)	(0.0973)	(1.415)	(0.700)	(0.0919)	(0.0421)
Observations	2,951	2,951	2,951	2,951	2,951	2,951	2,951	2,951
Student FE	No	Yes	No	Yes	No	Yes	No	Yes

Notes: This table shows the regression equivalent of the binned scatter plots in Figure 9. A neighbor is a student who lives within a 1.5 km radius of the index student, is of the same gender, studies the same major, and has the same admission year. Score difference is difference between index student's high school exam score and the neighbor's exam score. Safety difference is difference between the safety of the travel route chosen by the index student and their neighbor. Quality difference, time difference, and cost difference are defined similarly. Standard errors in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A3: Predicting cutoff scores for women

	Advantage to women in 2014
Proportion female (2013)	-0.088
	(2.049)
Area safety around college (SD)	0.057
	(0.244)
Boarding college	-1.967***
	(0.626)
Number of majors	0.037
	(0.0541)
Size of college	0.000
	(0.0003)
Annual Tuition (thousand Rs.)	0.000
	(0.0001)
Constant	1.597
	(0.994)
Observations	41

Notes: Advantage to women is measured in percentage points advantage given to female students by the college. Size of college is the total number of students in college in 2013, both from DU's annual report from 2013-14. Boarding college is an indicator equal to 1 for colleges that have boarding facilities. Sample includes all general education co-educational colleges in DU.

Table A4: Testing first-order and second-order local consistency conditions

College code	$\hat{\lambda}_c$ (1)	$t - ratio_1$ (2)	Condition satisfied (3)
1	0.544	-0.391	1
2	1.767	0.558	1
3	0.413	-0.232	1
4	0.588	-0.39	1
6	1.262	0.076	1
7	0.332	-0.382	1
9	0.45	-0.951	1
11	0.68	-0.314	1
12	1.756	2.217	1
13	0.449	-0.65	1
14	1.464	1.473	1
17	0.31	-0.345	1
18	1.475	1.588	1
19	0.243	-0.248	1
20	1.006	-0.056	1
21	0.554	-0.315	1
22	0.678	-0.186	1
24	0.679	-0.38	1
26	0.168	-0.209	1
28	0.506	-0.351	1
29	0.19	-0.142	1
30	0.545	-0.413	1
31	0.615	-0.316	1
34	1.274	0.092	1
35	0.454	-0.544	1
36	0.416	-0.917	1
38	0.462	-0.239	1
39	1.678	1.038	1
41	0.743	-0.199	1
42	0.509	-0.535	1
43	0.534	-0.418	1
46	0.108	-0.22	1
47	0.57	-0.342	1
48	0.449	-0.778	1
50	0.175	-0.234	1
51	0.752	-0.485	1
52	1.68	1.748	1
53	0.439	-0.877	1
54	0.368	-0.922	1
56	0.469	-0.379	1
57	0.031	-0.142	1
60	1.44	1.526	1
61	0	0	1
62	1.368	0.791	1
63	1.313	0.228	1
64	0.685	-0.171	1
65	0.484	-0.64	1
66	0.472	-0.279	1
67	0.77	-0.186	1
68	0.651	-0.549	1
69	0.521	-0.649	1
70	0.367	-0.316	1
72	0.186	-0.531	1
75	0.998	-0.115	1
76	0.946	-0.21	1
77	0.612	-0.527	1

Notes: Condition satisfied indicates whether the t-ratios are less than the 99% critical value for the one-tailed test = 2.326. The condition for this test are calculated using the mean value of the explanatory variables and since students have different number of (college and route) alternatives in their choice set, I use the mean value of P_{iC_c} for each nest.

Table A5: Alternative Construction of Route Safety

	Female (1)	Male (2)	Female - Male (3)	SE of Difference (4)
Panel A: Excluding "Feeling"				
Route safety (SD)	0.619	-0.016	0.636	(0.1903)
Cutoff score	0.023	0.072	-0.049	(0.0518)
Monthly travel cost (thousand Rs.)	-0.256	-1.397	1.141	(0.2065)
Route time (mins.)	-0.014	-0.032	0.018	(0.0079)
MRS (Safety, Score)	17.21	-0.14	17.35	
MRS (Safety, Cost)	-1.54	0.01	-1.55	
MRS (Safety, Time)	-28.46	0.32	-28.78	
Panel B: Excluding "Light"				
Route safety (SD)	0.546	-0.010	0.556	(0.2615)
Cutoff score	0.012	0.079	-0.067	(0.0566)
Monthly travel cost (thousand Rs.)	-0.329	-1.382	1.052	(0.2657)
Daily travel time (mins.)	-0.015	-0.035	0.020	(0.0011)
MRS (Safety, Score)	26.72	-0.08	26.80	
MRS (Safety, Cost)	-1.01	0.00	-1.01	
MRS (Safety, Time)	-21.90	0.18	-22.07	
Observations	43,874			

Notes: Estimation allows for different inclusive value coefficients for each college. MRS are computed in terms of SD of safety *within* the students' choice set. Estimation results are based on a 20% random sample of students.

Table A6: Number of Majors Considered by Students

Number of Majors	% students
1	0.41
2	0.20
3	0.18
4	0.08
5	0.13

Notes: This table shows the number of majors students applied for at the time of admission and the proportion of students.

Table A7: Overlap in Choice Sets of Related Majors

Related Majors	Overlap in choice set
Arts General, Commerce General	0.96
Commerce, Commerce General	0.78
Commerce, Economics	0.84
History, English	0.77
Political Science, Hindi	0.93
Political Science, English	0.82
Political Science, History	0.76

Notes: This table shows the percentage overlap in college choice sets for related majors.