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

Girija Borker*

July 2020

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 3,800 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 7,500 (USD 110) 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.

*The World Bank, 1818 H Street NW, Washington, DC 20433 USA. Tel: +1 (202) 473 1583 Email: *gborker@worldbank.org*. 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, Simone Schaner, Tarun Jain, Gabriel Kreindler and Akhil Lohia for their helpful conversations and comments, and the participants at various seminars and workshops. I am grateful to Inderjeet S. Bakshi, Harjender Singh 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 for their support and to the team of dedicated survey enumerators for their hardwork and dedication 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 Data2x, the Population Studies and Training Center, Global Health Initiative, Brown India Initiative, Pembroke Center for Teaching and Research on Women, and the Department of Economics at Brown University. IRB approval #1511001373 for this project was granted by the Office of Research Integrity at Brown University on December 22, 2015.

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 to 49 report feeling unsafe in public spaces (ICRW 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 or the fear of it (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. 2020), 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. 2019).

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 that 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 estimates based on women-only public transportation (Aguilar, Gutierrez and Soto Villagran 2019, Kondylis et al. 2020).

I face three key data 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 data 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 mixed logit model to estimate the magnitude of the students' willingness to pay for travel safety. In my benchmark specification, students' indirect utility depends on college quality, route safety, travel costs, and travel time, and I estimate the model for women and men

separately. 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 from a top 20 percent college to a top 30 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 17,500 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 9,950 (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 strands 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, the school choice literature that assesses the factors influencing choice of a school, and the literature on spatial frictions driving gender disparities in acquisition of human capital. It has been shown that fear of imagined dangers affects individual behavior (Becker and Rubinstein 2011). There is evidence that harassment by strangers strongly affects women's perceptions of safety across social contexts (Macmillan, Nierobisz and Welsh 2000) and that women change their behavior in response (Keane 1998). Specifically, lack of safety has been found to affect women's mobility patterns (Hsu 2011, Porter et al. 2011) and is negatively correlated with women's labor force participation on the extensive margin (Chakraborty et al. 2018, Siddique 2018) and the intensive margin (Cook et al. Forthcoming). 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). 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. Access to schools and training centers through choice of their location, better roads or provision of transportation has been found to play a crucial role in women's take-up of opportunities (Mukherjee 2012, Burde and Linden 2013, Muralidharan and Prakash 2017, Jacoby and Mansuri 2015, Cheema et al. 2020). While most of this work alludes to women's own or their parents' safety concerns as a potential mechanism in affecting women's choices, this is the first paper to explicitly measure the extent to which safety affects women's educational choices, conditional on travel time.

I Institutional Setting: College Choice and Harassment

A Structure of DU

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.[4]

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,[5] 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] Based on the comparison of 2016-17 fees for general undergraduate courses.

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

average a college offers about 12 majors with most colleges having a large overlap in the majors they offer.[6] Each college has its own campus, staff, classes, and placements.[7] 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 cutoff scores from the the first cutoff list. Rank within a student's choice set ranks the colleges to which the student was admitted to 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. This implies that 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.[8] This is despite women scoring higher than men on exams at the end of high school, as shown in Figure A1.

Another feature of DU is that majority of the students (72 percent) enrolled at the University are residents of the Delhi National Capital Region (NCR). 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.[9] The students travel to college by either public or private

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

[7]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 (Ghosh, Sushmita. 2017. "DU cell shows dismal placement record". The Asian Age, May 5, , last accessed: May 25, 2020).

[8]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.

[9]Press Trust of India. 2015. "With Hostel Shortage in Delhi University, Students Demand Implementation of Rent Act". NDTV, June 14, ,last accessed: May 25, 2020.

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. Moreover, home ownership rates in the sample of Delhi residents is high, with 82 percent of them owning the homes they reside in, indicative of the high costs associated with changing their place of residence.

B 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 (*Three in Four Women Experience Harassment and Violence in UK and Global Cities* Action Aid 2016). Delhi, infamously known as the "rape capital" of India, is notorious for both verbal and physical harassment on public transportation.[10] In my sample, 89 percent of female college students have faced some form of harassment while traveling in Delhi. In particular, 64 percent of female students have experienced unwanted staring, 50 percent have received inappropriate comments, 33 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 71 percent of female students report avoiding an unsafe area, 67 percent avoid going out after dark, 30 percent move away from the harasser, and only 3 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 stan-

[10] Thomson Reuters Foundation News. 2014. "Most dangerous transport systems for women". October, 31. , last accessed: May 25, 2020.

dardized 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.

II 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 3a, 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 3b and 3c 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 = $\frac{g+n}{2}$. Figure 3d 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 3e, 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 3f 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. With this a female student on average attends a college with quality = $\frac{g+3n}{4} < \frac{g+n}{2}$. Another case is where the woman who can only attend the Not so good college by traveling through the unsafe area could have chosen to not attend college at all, as denoted by the thick arrow. This is beyond the scope of my study since I examine the choices of students currently enrolled in DU and 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. This is 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.

III Data

I use 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 A describes the student data, Section D describes the route creation using Google Maps, and Section E outlines the mobile app safety data.

A 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

I use detailed data on students in eight colleges at DU from a survey conducted in January - April 2016. As part of the survey, data was collected on 3,846 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, modes of travel and time of departure, 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 4 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 3,744 students with complete information and geocoded travel routes. Of these, 2,713 students (72.5 percent) are residents of Delhi who live with their families and travel to college every day.

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 the 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 conducted a short survey across 32 other colleges in DU to be able to compare students in my sample to students in other colleges. 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.

Data on 799 male and female students was collected through a combination of online (34 percent) and intercept (66 percent) surveys. From this survey, I have information on students' current and permanent residential locations and high school exam scores by subject. In the short survey data, 630 students (79 percent) live in Delhi with their family and travel to college every day and comprise 99.5 percent 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 5 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 A1 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. [11]

B 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.[12] The high school scores are based on the national Senior School Certificate Examinations.[13] There is a different cutoff score for each major on the basis of the seats available in a college, the number of applicants, the high school scores of applicants, and the cutoff score in previous years according to the Delhi University Standing Committee on Admissions 2015.[14] The cutoffs vary by social category, disability status, subjects studied in Class 12 and in some cases by gender of the student.[15] Following the release of the cutoff list, students have about three days to register at 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 to admit every student who approaches the college with a score above the released cutoff score.[16] After three days if there are seats available in a college then the college

[11] Social categories refers to the officially designated groups of historically disadvantaged people in India. These categories are used for the purposes of affirmative action. The categories are General category, Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Castes (OBC). 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. 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 OBC, and SC female students. 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

[12] The average for each student is calculated on a "best of four" basis or using scores of four of the five or six subjects that a student wrote exams for. Most colleges require students to include at least one language in this average.

[13] 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. There are other boards of education at the state level. In our sample over 96 percent of students' board of examination was the CBSE.

[14] Kohli, Gauri. 2015. "Want to join DU? Check out how cutoffs are calculated". The Hindustan Times, June 30, https://www.hindustantimes.com/education/want-to-join-du-check-out-how-cutoffs-are-calculated (last accessed: May 25, 2020).

[15] 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, most often for undergraduate courses in language. A sample cutoff list is shown in Figure A2. 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.

[16] 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 (Joshi, Mallica. 2013. "Some colleges flouting norms, ad-

revises its cutoffs downward and releases a 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.[17]

C 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. The cutoff that is applicable for each student based on their social category, gender, religion and high school subjects is used. I construct the choice sets cumulatively using all the cutoff lists released by every college, which is equivalent to using the lowest cutoff score across cutoff lists. As mentioned previously, in 2015 DU had 12 cutoff lists [18,19] 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 A3.

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 location, gender, social category and parental occupation.[20] I find 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.

---

mits DU". Hindustan Times, July, 4. http://www.hindustantimes.com/education/some-colleges-flouting-norms-admits-du, last accessed: May 25, 2020).

[17] 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, 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 the colleges.

[18] While all colleges were open for enrollment in History honors in the first cutoff list, 62.5 percent colleges were open for enrollment in the second cutoff list and ony 37.5 percent colleges had seats remaining in the third cutoff list.

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

[20] I was able to match 78 percent of the Delhi residents in my full survey sample to the administrative data for the one college, without any conflicts.

D 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.[21] 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.[22] The public transit routes are then broken into separate legs based on travel modes. 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.[23] 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.[24]

An example of route mapping is given in Figure A4. Figure A4a shows a student and the college he chose to attend. Figure A4b 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 A4c shows potential route options to the chosen college and Figure A4d shows the potential route options to each of the 32 colleges in this student's choice set.

E 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.

[21] 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 exact daily travel routes by mode, which I collect.

[22] The algorithm allows for extraction of four routes since a maximum of three routes per mode are suggested by Google maps in intercative mode.

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

[24] 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.

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 **??**. 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 crowdsourced and partially collected by trained auditors. The latter enables SafetiPin to have a wider and more representative coverage of the city (Viswanath and Basu 2015).

I have data on over 26,500 audits from November 2013 to January 2016, as shown in Figure 6a. In this sample, 98 percent of the contributors are 39 years or younger and 70 percent of the users are female.[25] I interpolate these audits to create a safety surface using Inverse Distance Weighting, this base level of area safety is shown in Figure 6b. Each pixel is 300 meters$\times$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 crowdsourced 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, 80 percent reported exclusively traveling in the ladies-only compartment. A large

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

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.

F Calculating Route Safety

I assign a safety score to the reported and potential travel routes by computing a weighted average of the area safety for the travel route, where the weights are the proportion of the route and harassment by travel mode (m) in each safety pixel (p). Specifically, the safety score for a route, such as the one shown in Figure 7 is calculated 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 area safety is from the SafetiPin data; route length in a safety pixel 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.[26] I check for robustness by using alternative safety measures.

Table 1 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.[27] 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.[28] 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.

[26] 69.6 percent of the variation in the safety score comes from variation in area safety and the remaining 30.4 percent of variation in harassment across modes.

[27] Table A3describes the sample of non-Delhi residents.

[28] 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.

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 II.

IV Descriptive Evidence: Response to Changes in Choice Set

We could get an insight into students' preferences if we observed their response to different choice sets. The ideal experiment for this exercise would require evaluating students' responses to a random allocation of college choice sets. I exploit DU's admissions process to approximate this ideal experimental design. I use the fact that students' high school exam scores combined with colleges' cutoff scores completely determine their choice set.

I compare the choices made by males and females relative to other students of the same gender from their neighborhood, with the same major and admission year as their relative exam scores change. Given the discrete cutoffs, a change in the student's relative exam score, or score gap, also changes their relative choice set. A student with a higher 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 exam score. A neighborhood is defined as a 1.5 km radius around the index. studentI have 1,228 unique pairs that use information on 1,232 students in my sample.

To better understand how analyzing students' choices to a change in their relative choice set helps in deducing 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 college chosen by the index student would increase with 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 of the index student's chosen route relative to her neighbor's chosen route with an increase in the score gap.[29] The relative college quality of the chosen college, the relative travel time, and cost could move in any direction with a increase in the score gap.

Figure 8 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

[29]The safety difference would remain constant with an increase in the score gap 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.

school exam score and their neighbor's, separately for 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 gap implies that the index student faces a larger 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 8a, we can see that there is a positive relation between safety difference and the score gap 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 8b, the positive relation between quality difference and the score gap 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 8c 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 8d.

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.

V Model of College Choice

A Estimating the Trade-off between Route Safety and College Quality

To estimate the partial effects and measure students' willingness to pay for different choice attributes, I use an additive random utility framework with a rational, utility maximizing student i (McFadden et al. 1977, McFadden 1978). In this framework, each student i faces a choice of N_i

mutually exclusive colleges denoted by $\mathbf{C_i} = \{C_{i1}, C_{i2}, \ldots, C_{iN_i}\}$ 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.

I assume that each student i maximizes an indirect utility function of the form:

$$\begin{aligned} U^c_{ir} &= \beta_i' X^c_{ir} + \varepsilon^c_{ir} \\ &= \beta_{iq} Q^c_i + \beta_{is} S^c_{ir} + \beta_{it} T^c_{ir} + \beta_{ip} P^c_{ir} + \varepsilon^c_{ir} \end{aligned} \tag{1}$$

where r and c denote the travel route and college respectively, Q^c_i is quality of college c, S^c_{ir} is safety of the travel route to college, T^c_{ir} is the travel time to college, P^c_{ir} is the travel cost to college, and the respective, $\beta_i$represent the weight student i places on the respective attribute. ε^c_{ir} is the unobserved part of utility that captures the effect of unmeasured variables, personal idiosyncracies, maximization error, etc. Student i's chooses college c and route r ($d^c_{ir} = 1$) such that the choice maximizes his or her utility over all possible colleges and routes in their choice set:

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

The main variable of interest is the trade-off between route safety and college quality, as measured by the marginal rate of substitution representing the college quality a student is willing to give up for an additional unit of perceived safety while traveling to college. This trade-off can be denoted by

$$MRS^{QS}_i \equiv \frac{\triangle Q^c_i}{\triangle S^c_{ir}} = \frac{\beta_{is}}{\beta_{iq}} \tag{2}$$

I estimate this model in a mixed logit framework with random coefficients to estimate variation in preference across the population and recover the full distribution of the MRS_{QS} for male and female students. Mixed logit model is appropriate in this setting for several reasons. First, it allows relaxation of the Independence of Irrelevant Alternatives (IIA) assumption that is imposed by logit and generalized extreme value (GEV) models and allows for flexible substitution patterns. The IIA assumption is potentially problematic in this case since there are several routes to every college in the student's choice set. The IIA assumption implies that the relative odds of choosing between two routes to a college remains constant when a new route option is introduced, say with a mode composition similar to one of the existing routes. Second, logit and GEV models assume that all agents in the population have the same preferences whereas mixed logit allows for random taste variation and enables explicit estimation of parameter heterogeneity. This is relevant since 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 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 high importance on college quality, some students may face high decision making costs, due to individual or household constraints, 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 some students are inherently averse to harassment while others may also dislike harassment but due to differential exposure to harassment in their lifetime are less sensitive to it. These different sources of heterogeneity cannot be separately identified in this analysis because they result in observationally equivalent choice behavior.

The mixed logit model can approximate any random utility model, given appropriate mixing distributions and explanatory variables (Train 2003). 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.,$\beta \sim f(\beta|\mu,\nu)$, 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:

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

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

$$LL(X,\mu,\nu) = \sum_{i} \sum_{c} \sum_{r} d_{ir}^{c}\, log\{\mathbb{P}_{ir}^{c}\}$$

I use the triangular distribution as the mixing distribution $f(\cdot)$ for the route safety and college quality coefficients, and the restricted triangular distribution for the travel time and cost coefficient so that all students dislike longer commute time and we have a negative price coefficient. The triangular and restricted triangular distributions have bounded support and are hence less sensitive to outliers.[30,31] I estimate the model separately for men and women. Since the log-likelihood function does not have a closed form solution, simulation methods are used to generate draws of β from $f(\cdot)$ to numerically integrate over the distribution of β. Estimation is done by the method of maximum simulated likelihood.

[30]Online Appendix Figure **??** presents an example of a triangular distribution which has positive density that starts at $b-s$, rises linearly to b, and then drops linearly to$b+s$, taking the form of a tent or triangle. The mean b and spread s is estimated. By constraining $s=b$, we can ensure that the coefficients have the same sign for all decision-makers (Train 2003). Kremer et al. 2011 and León and Miguel 2017 also use a restrcited triangular mixing distribution in their analysis.

[31]Estimation of the mixed logit models was carried out using Matlab code developed by Kenneth Train; see (accessed June 2020).

B Empirical Specification

In the empirical estimation, 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,[32,33] and second, using cutoffs for female students would, by construction, lower the quality of colleges that give an advantage to female students. Figure A5 shows the correlation between the cutoff score and proportion of accepted students who enrolled in a college, separately for male and female students. 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 in section F. P_{ir}^c is the monthly travel cost to college in thousands of Indian Rupees, its cal. T_{ir}^c is the daily 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 and how they receive travel allowances from their parents, it also lends a more relevant interpretation to the time coefficient, i.e., the marginal utility from a unit increase in daily 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 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 also compute the marginal rate of substitution between safety and travel time (MRS_{TS}) and marginal rate of substitution between safety and travel costs (MRS_{PS}) to highlight the potential costs women incur both in the short-term and long-term from the lack of travel safety.

I expect the distribution of marginal rate of substitution between quality and safety for women to lie to the left of the distribution for men. 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 the distribution of MRS for women to lie to the right to that of men, such that women have a higher willingness to pay for an additional unit of safety in terms of travel time and travel costs.

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

[33] The results do not change if I use cutoffs for other social categories (not shown).

VI 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. Additionally, home ownership rate among the Delhi residents is high, with 82 percent students living in owned houses.[34]

Second, the colleges in DU are spread out across the city and are located in neighborhoods with varying characteristics,with students of both genders across 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 9 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.[35] Figure 9a and Figure 9b 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 9c and 10d. 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 A4.

VII Results

Table **??** presents the main mixed logit model results. I regress the college route choice indicator on the safety score of the travel route (in SD from the mean), cutoff score that captures selectivity of the college as a measure of its quality (in percent), daily travel time (in minutes), and monthly travel cost (in '000 Indian Rupees). The model is estimated separately for female male students. I assume that the random coefficients associated with route safety and college quality follow a

[34] Another five percent of Delhi residents live in houses allotted to either parent by the parent's employer.

[35] Based on the cutoff score for Bachelors in Political Science, as shown in Figure 5(a).

triangular distribution and the time and cost coefficients follow a restricted triangular distribution.

As expected, both men and women prefer routes that are quicker and cheaper (column 1 and 3). The mean coefficient on route safety is positive for both men and women, additionally all men and women in the sample have a positive coefficient on safety. The positive safety coefficient for men most likely captures the amenity value of a safe route, i.e., better lighting, better access to transport etc. The mean coefficient on college selectivity is also positive for both men and women indicative of a preference for more selective colleges. However, 23 percent of women and 5 percent of men have a negative coefficient of quality, suggestive of decision making costs faced by some students. Following equation 2, I use the coefficient estimates on the route safety and college selectivity to estimate the average valuation of safety in terms of college quality, route travel costs, and travel time by gender. I find that women are willing to attend a college that is 8.8 percentage points lower in quality for an additional SD of safety. This is equivalent to choosing a college that is 5.8 ranks lower.[36] 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.[37,38] Men on the other hand are willing to attend a college that is only 2.11 percentage points (or 1.4 ranks) lower in quality for an additional SD of safety. 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. In terms of travel costs, women are also willing to travel by a route that costs Rs. 17,500 (USD 250) more per year as long as it is one SD safer. Men are willing to spend an additional Rs. 9,950 (USD 140). This shows that women are willing to spend 75 percent more than men in terms of travel costs for an additional unit of safety. The difference of Rs. 7,500 is equal to almost 70 percent of the average annual tuition at DU and 5 times the average monthly travel costs in this context. All of the aforementioned safety valuations are measured in terms of the SD of route safety across the predicted route alternatives within a students' choice set since the variation in route safety that matters is that *within* the students choice set of colleges. The within choice set variation is 38.1 percent lower than the overall SD in route safety for male students and 47.4 percent lower for female students.[39]

[36]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 shown).

[37]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). Correlation of area safety with other crimes againt women are shown in Figure A7 in the Online Appendix. These are all the crimes against women that could potentially take place in public spaces. As we can see, area safety is negatively correlated with all reported crimes against women except assaults against women, for which it is close to zero.

[38]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).

[39]For example, $\bar{MRS}_{QS} = (1 - 0.381) \times \frac{\bar{\beta}_s}{\bar{\beta}_q}$ for male students.

To control for other factors that may influnece students choice I include additional college level and route level variables. These include for every college the area safety within a 1.5 km radius around the college, the total number of students enrolled and an indicator for whether the college is women only. I also include an indicator for whether the majority of the travel route uses public transport (bus, train or metro), modes that are characterized by group travel and fixed schedules that has implications for travel cost (cheaper), travel time (longer) and perceived safety (considered unsafe). Addition of these controls does not change the results significantly and all additional variables have coefficients that are not significantly different from zero.

A 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 **??**, each panel reports results based on these alternative measures of safety. Here I report the MRS estimates from two alternative safety indices. For example, in Panel A, the safety index excludes the crowd parameter, which is a measure of the number of people a woman obsrves in the public space and has been found to have an inverse-U relationship with perceived safety with completely deserted areas and very crowded spaces making women feel unsafe. 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 A5. I find that there is significant overlap in related majors which students tend to consider together, at the

time of application. As shown in table A6, the overlap in choice sets varies between 76 percent to as high as 96 percent.

VIII 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 9 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 2 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. 7,500 on annual travel, relative to men, for a route that is one SD safer. This amount is almost 70 percent of the average annual fees at DU. Using estimates from Sekhri (2019), on the labor-market earnings advantage from attending a public college, I estimate that women's willingness to pay for safety translates to a 17 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.

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 (ILO 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 Villagran. 2019. "Benefits and Unintended Consequences of Gender Segregation in Public Transportation: Evidence from Mexico City's Subway System." *Economic Development and Cultural Change* .

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. 2020. "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.

Burde, Dana and Leigh L Linden. 2013. "Bringing education to Afghan girls: A randomized controlled trial of village-based schools." *American Economic Journal: Applied Economics* 5(3):27–40.

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.

Chakraborty, T., A. Mukherjee, S. R. Rachapalli and S. Saha. 2018. "Stigma of Sexual Violence and Women's Decision to Work." *World Development* 103:226–238.

Cheema, Ali, Asim I Khwaja, Farooq Naseer and Jacob N Shapiro. 2020. "Glass Walls: Experimental Evidence on Access Constraints Faced by Women.".

Cook, Cody, Rebecca Diamond, Jonathan Hall, John A List and Paul Oyer. Forthcoming. "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers." *Review of Economic Studies* .

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. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica* 87(5):1439–1474.

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

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

ILO. 2017. "World Employment Social Outlook - Trends for Women.".

Jacoby, Hanan G and Ghazala Mansuri. 2015. "Crossing boundaries: How social hierarchy impedes economic mobility." *Journal of Economic Behavior & Organization* 117:135–154.

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.

Kondylis, Florence, Arianna Legovini, Kate Vyborny, Astrid Zwager and Luiza Andrade. 2020. "Demand for "Safe Spaces": Avoiding Harassment and Stigma." *Working Paper* .

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.

León, Gianmarco and Edward Miguel. 2017. "Risky Transportation Choices and the Value of a Statistical Life." *American Economic Journal: Applied Economics* 9(1):202–28.

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/*

Mukherjee, Mukta. 2012. "Do Better Roads Increase School Enrollment? Evidence from a Unique Road Policy in India." *Evidence from a Unique Road Policy in India (August 28, 2012)* .

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.

Porter, Gina, Kate Hampshire, Albert Abane, Augustine Tanle, Alister Munthali, Elsbeth Robson, Mac Mashiri, Goodhope Maponya and S Dube. 2011. "Young people's transport and mobility in sub-Saharan Africa: the gendered journey to school." *Documents d'analisi geografica* 57(1):61–79.

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. Forthcoming. "Prestige Matters: Wage Premium and Value Addition in Elite Colleges." *American Economic Journal : Applied Economics* .

Siddique, Zahra. 2018. "Violence and Female Labor Supply." *IZA Discussion 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: Cambridge University Press.

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.

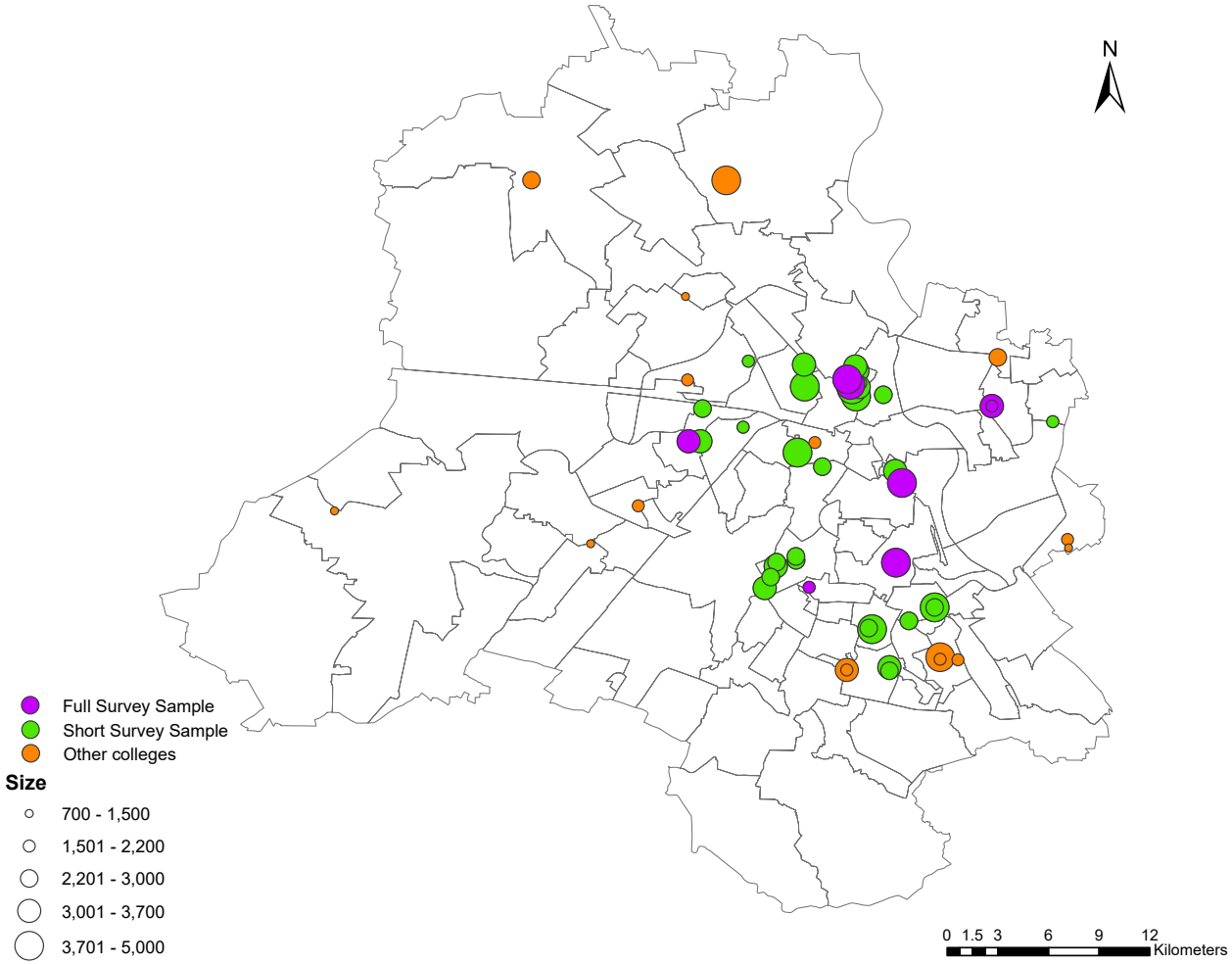
Warr, Mark. 1985. "Fear of rape among urban women." *Social problems* 32(3):238–250.

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.

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

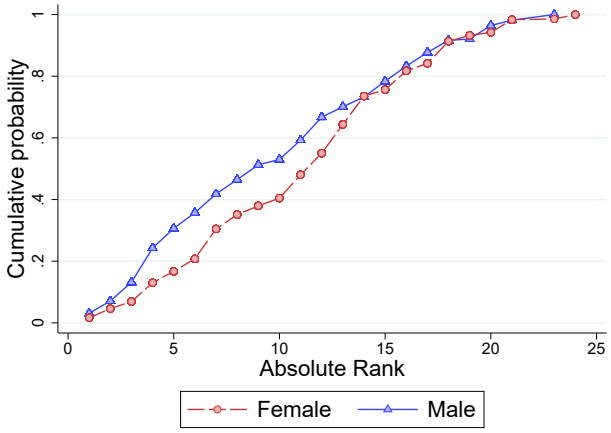
Figure 1: Colleges in Delhi University



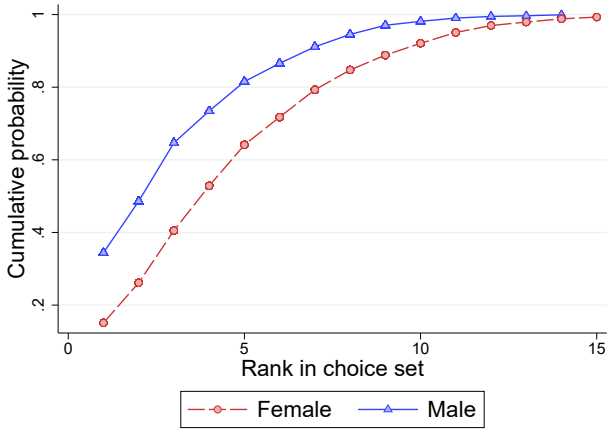
Notes: The map shows the 58 general education colleges in DU. Eight colleges are in the full survey sample and 32 colleges are in the short survey sample.

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

(a) Absolute rank



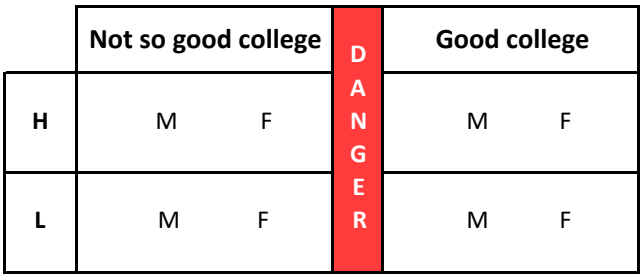
(b) Rank within choice set



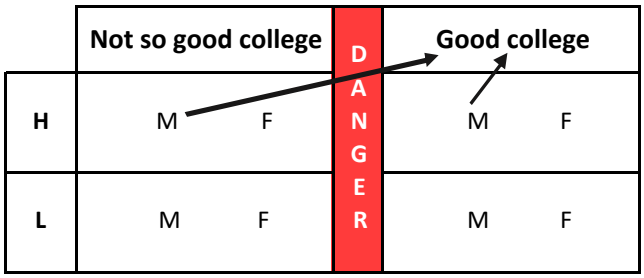
Notes: Rank is based on cutoff scores of a college for the student's major and admission year from the first cutoff list for general category male students. A higher rank indicates a lower cutoff score or worse quality. The absolute rank in Panel (a) ranks college within a major and admission year using 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: Stylized Example

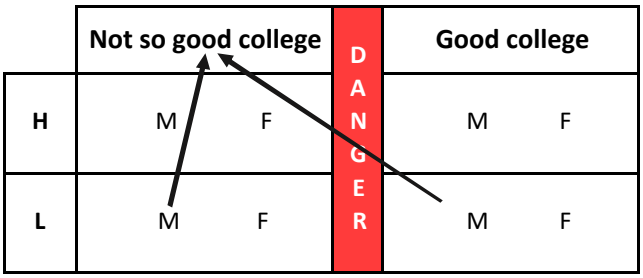
(a) The set-up



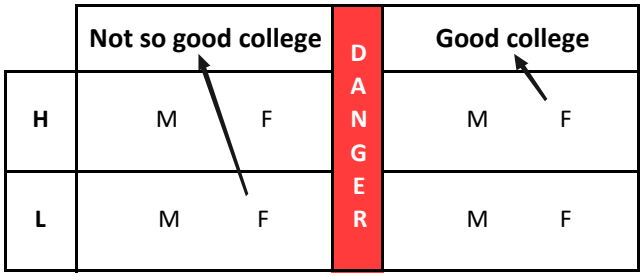
(b) High-scoring males



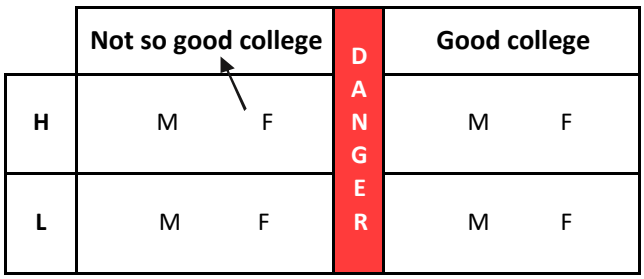
(c) Low-scoring males



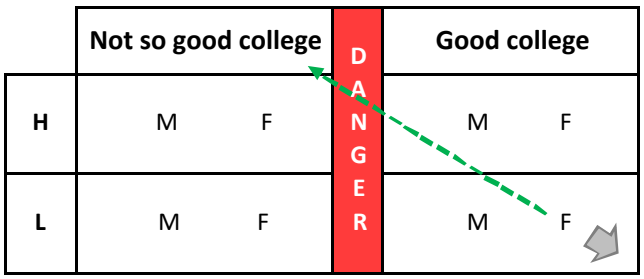
(d) Females who do not face a safety-quality trade-off



(e) High-scoring female chooses lower quality college

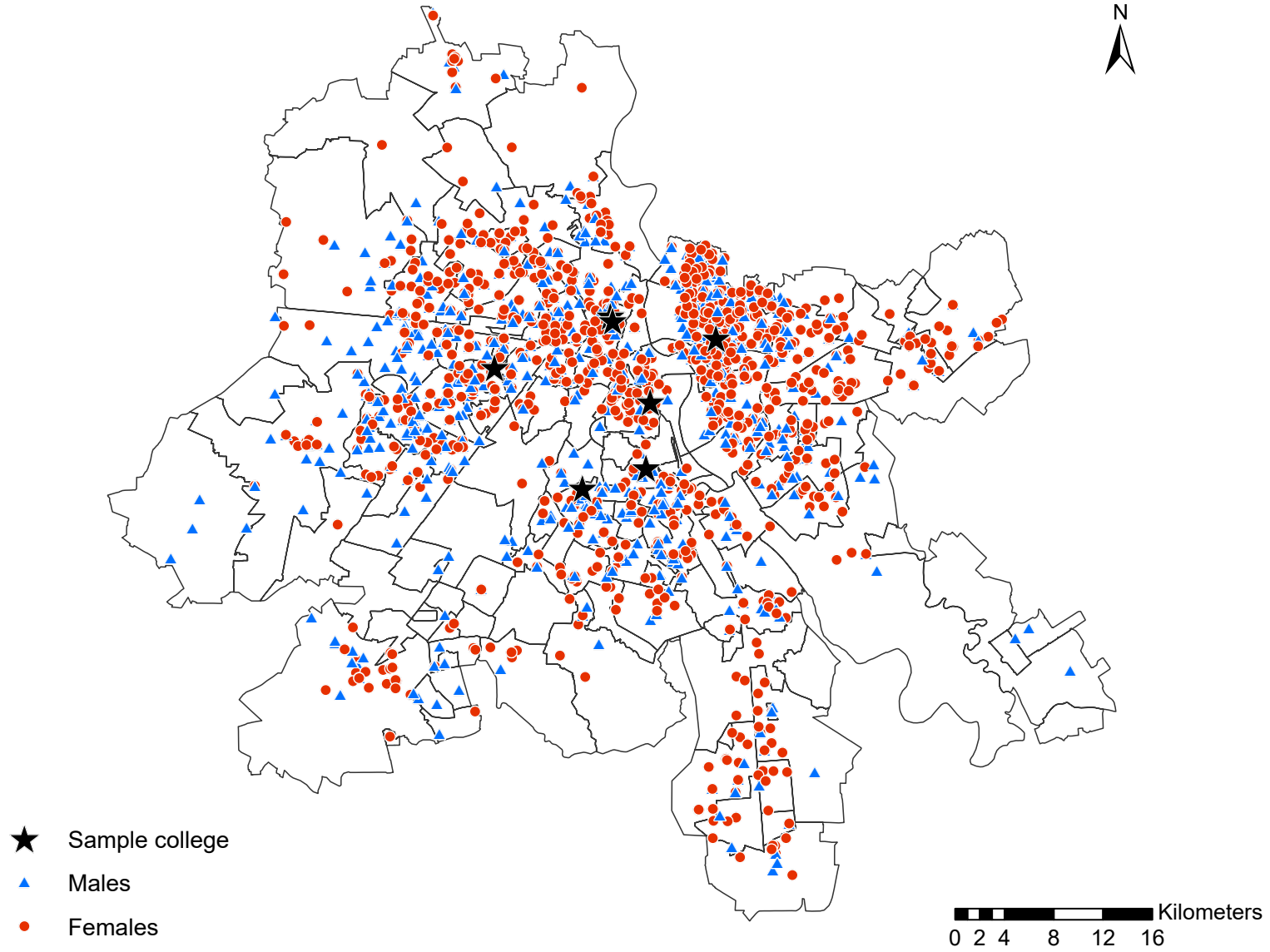


(f) Female uses more expensive mode of transport



Notes: This figure shows the college and route choice of students by gender and high school exam scores. M and F denote a male and female student respectively. H and L denote students' high school exam scores. 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.

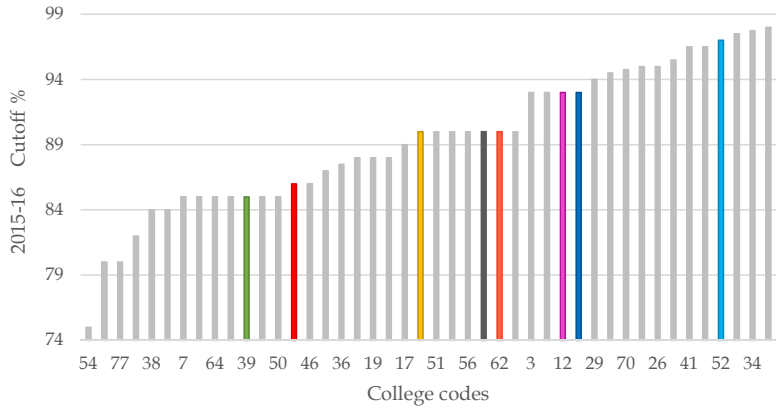
Figure 4: Students in Full Survey Sample



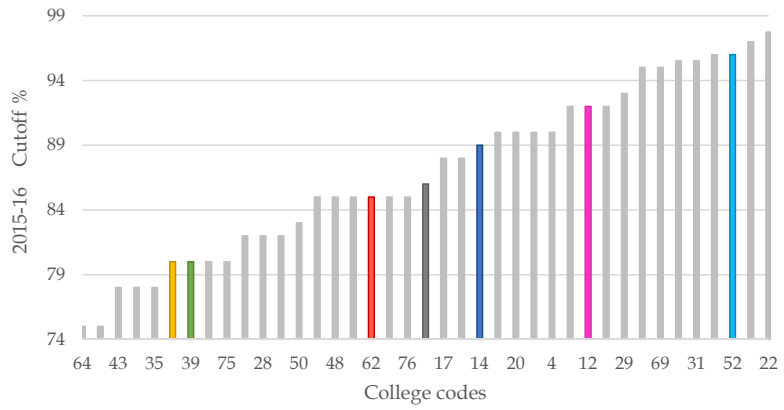
Notes: This map shows the residential location of students who are Delhi residents from the full survey sample, who live at home and travel to college everyday.

Figure 5: Variation in Quality of the Full Survey Sample Colleges

(a) Political Science



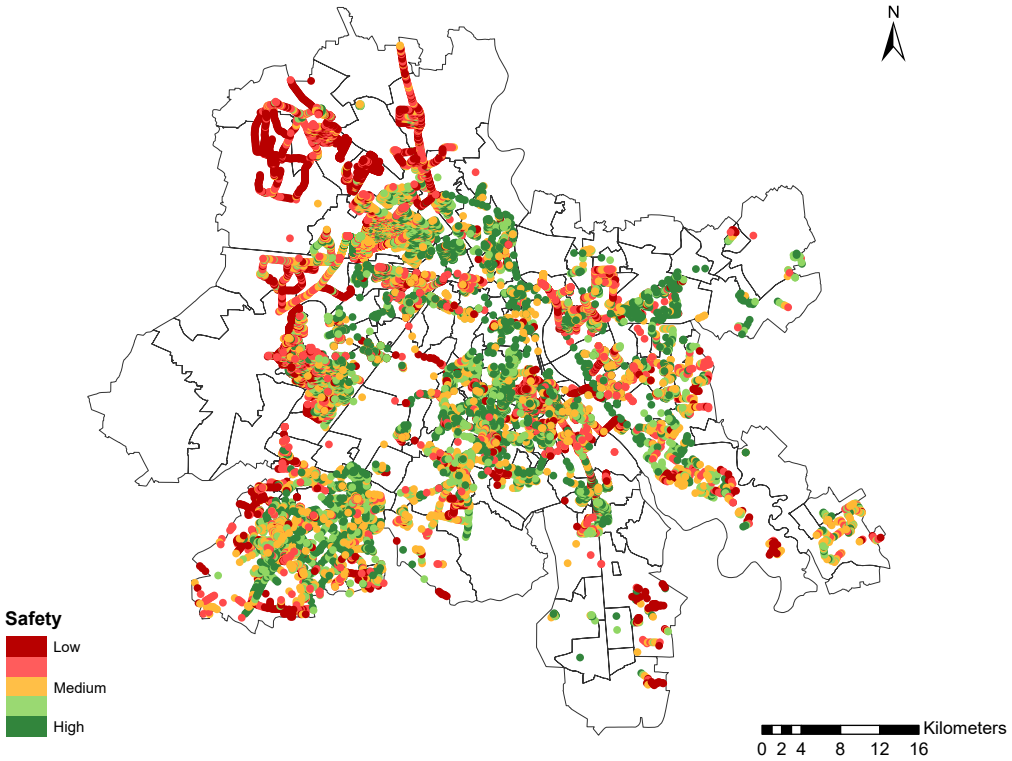
(b) History



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 with college codes.

Figure 6: Area Safety Data from SafetiPin

(a) Safety Audit Data



(b) Safety Surface

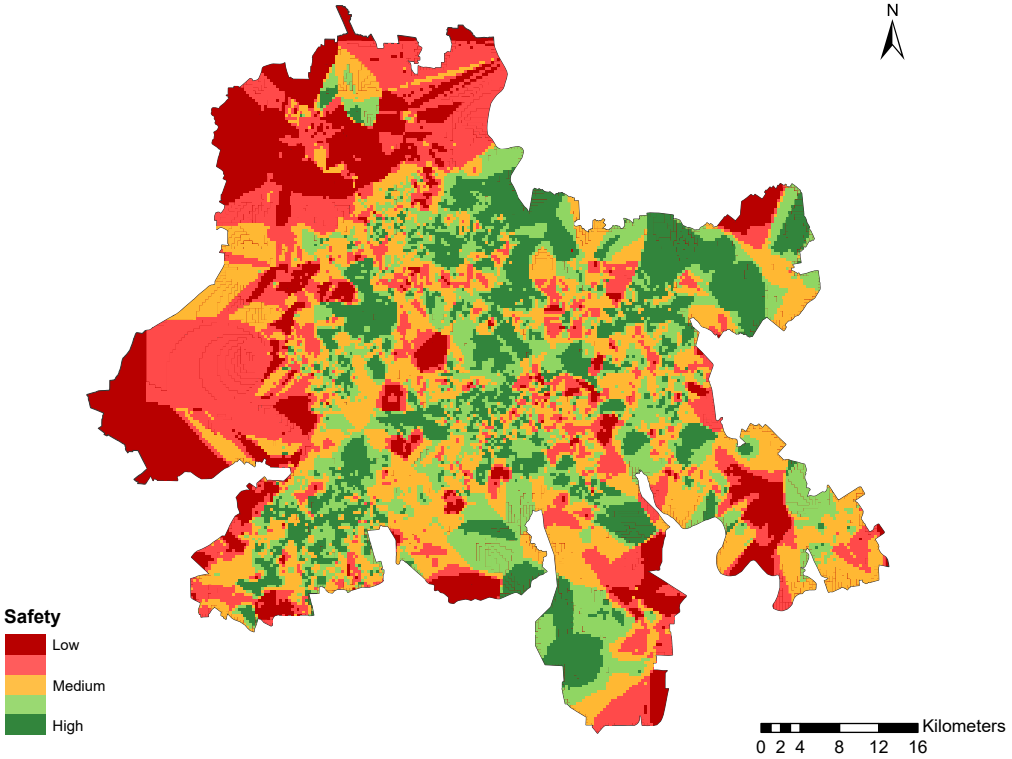
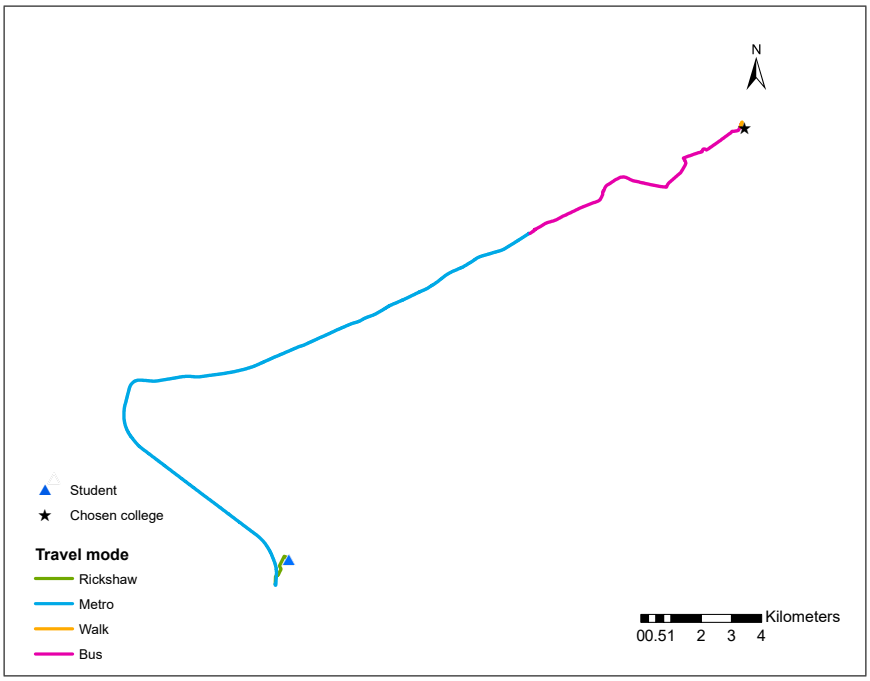
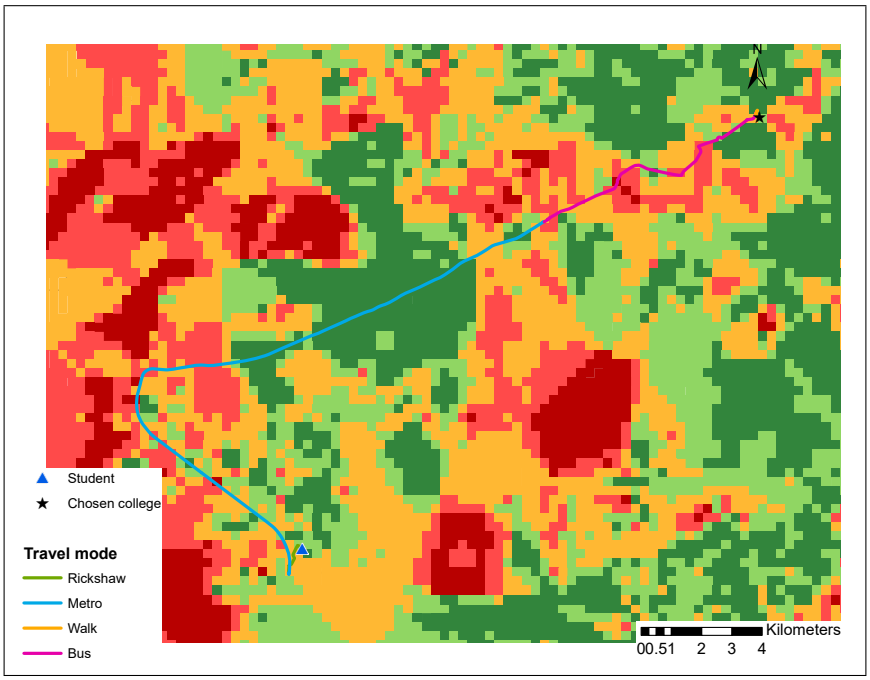


Figure 7: Calculating Route Safety

(a) Example Travel Route

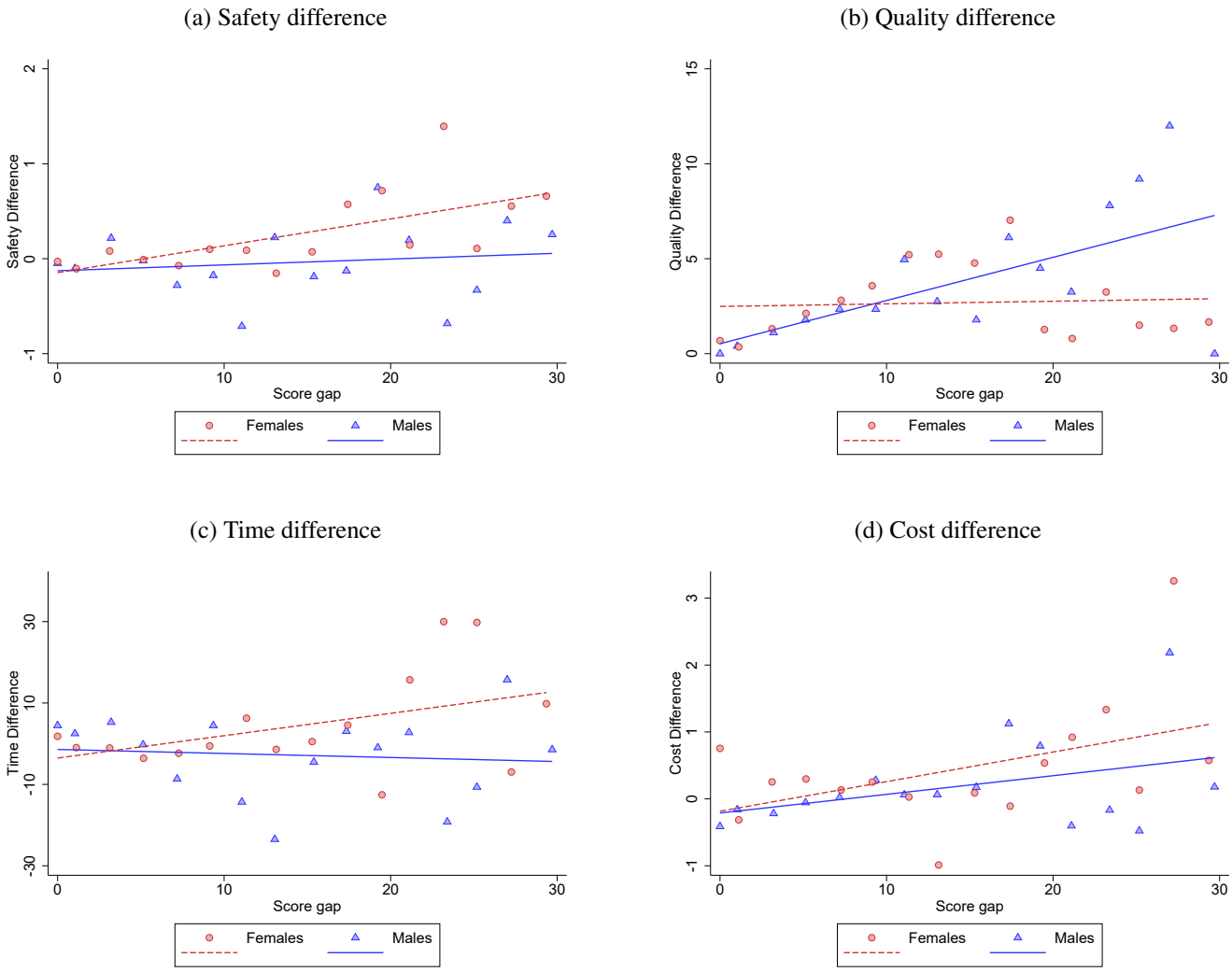


(b) Travel Route Over the Safety Surface



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.

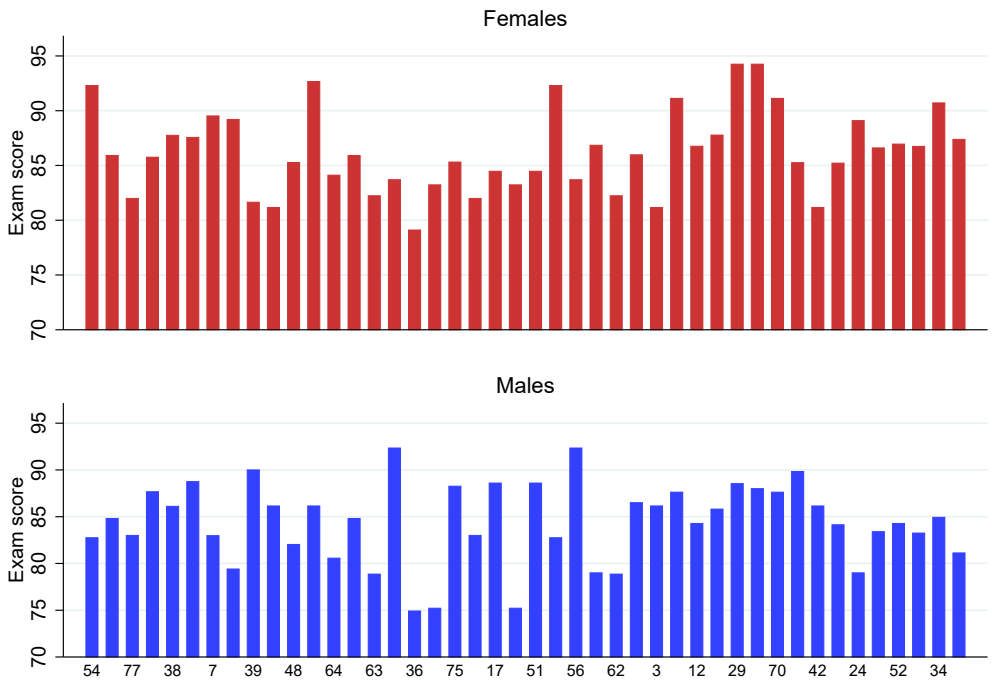
Figure 8: Choice Relative to Neighbors



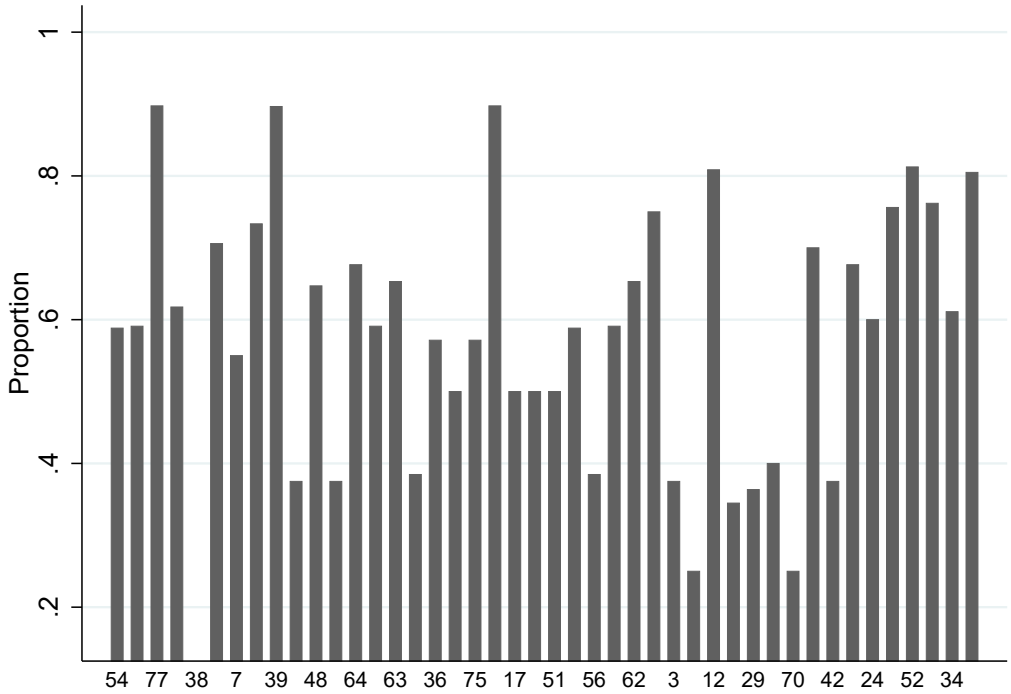
Notes: The figures plot binned scatter plots of difference in travel safety, college quality, travel time, and travel cost between the index student and their neighbor's choice. Index student is the student who scores higher. Score gap bin is the two point bin of score difference between the index student 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 9: Variation of Student and Area Characteristics Around Colleges

(a) Average high school exam score by gender



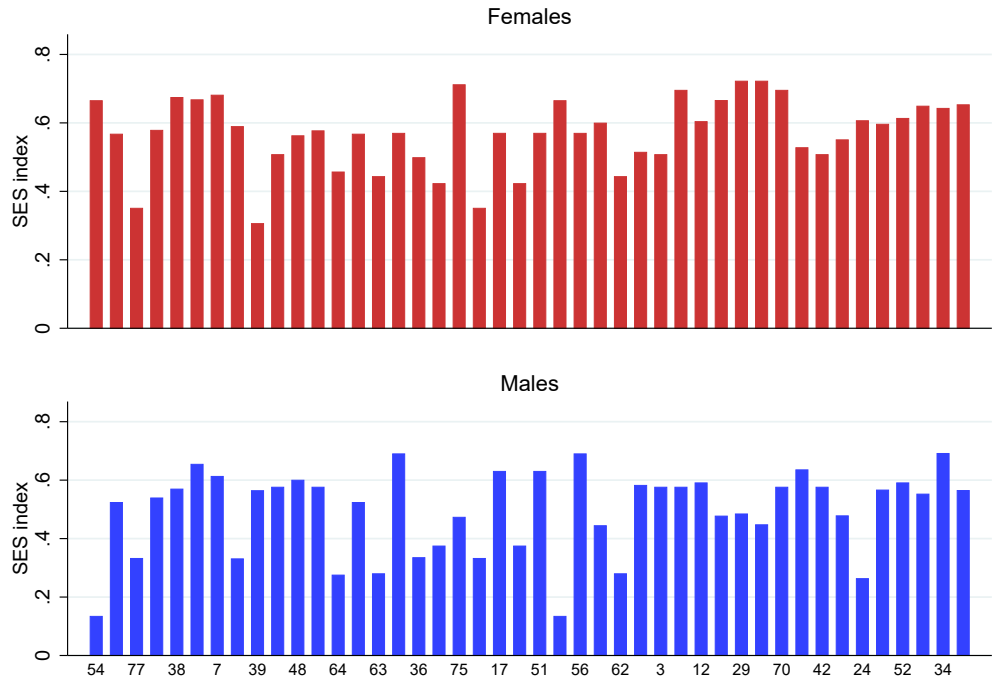
(b) Proportion of female students

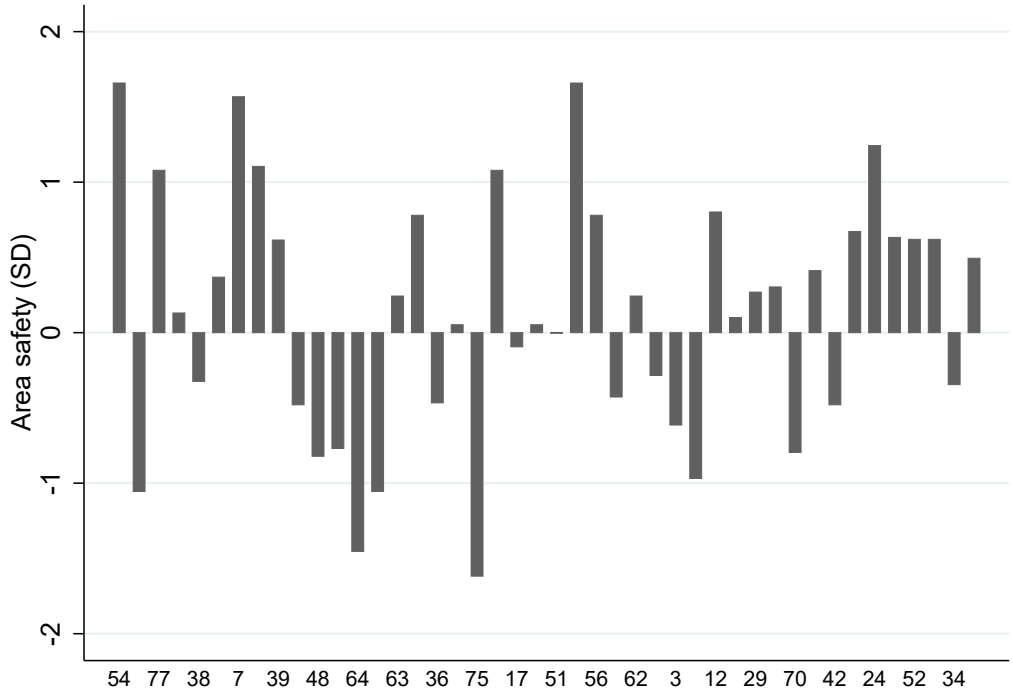


Notes: The figures show characteristics of students living within a 1.5km radius around each general education college in DU that offers a Bachelor of Arts in Political Science. Each bar reprsents a college. The colleges are in ascending order of quality. The quality measure used here is the cutoff scores 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 5(a). Alternate bars are labelled.

Figure 9: Variation of Student and Area Characteristics Around Colleges

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



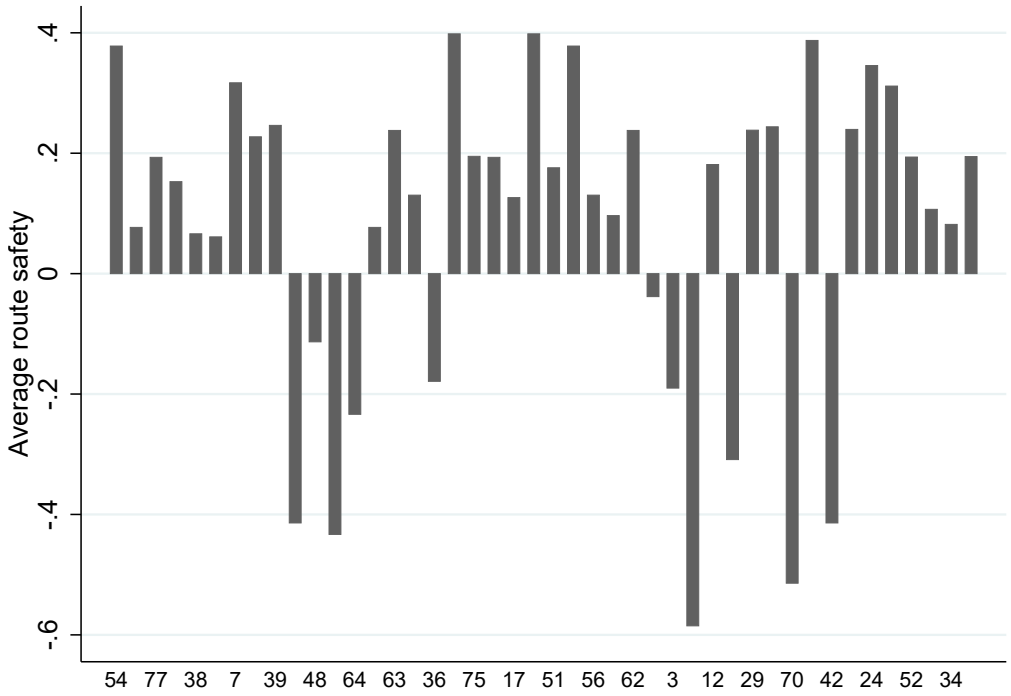


(d) Average area safety

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 5(a). Alternate bars are labelled.

Figure 9: Variation of Student and Area Characteristics Around Colleges

(e) Average route 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 5(a). Alternate bars are labelled.

Tables

Table 1: Summary Statistics

	Total *N = 2,713*		Female *N = 1,767*		Male *N = 946*		Female - Male	
	Mean	SD	Mean	SD	Mean	SD	Mean Diff.	SE
A. Student Characteristics								
Proportion surveyed	0.72		0.76		0.67			
Proportion of Delhi residents	0.99		0.99		0.99			
Female	0.65							
High school exam score (%)	83.88	[9.08]	84.52	[8.85]	82.68	[9.39]	-1.84	(0.36)
Socio-economic status	0.47	[0.29]	0.50	[0.28]	0.40	[0.30]	-0.10	(0.01)
B. College Characteristics								
First cutoff score	87.96	[8.82]	87.57	[9.20]	88.70	[8.00]	1.13	(0.35)
Absolute rank	10.87	[5.51]	11.15	[5.31]	10.34	[5.83]	-0.81	(0.22)
Rank in choice set	4.53	[3.23]	5.02	[3.35]	3.60	[2.75]	-1.42	(0.13)
Distance (km.)	13.05	[9.66]	13.06	[9.54]	13.05	[9.88]	-0.01	(0.39)
Annual Tuition ('000 Rs.)	10.57	[3.22]	10.86	[3.07]	10.02	[3.43]	-0.85	(0.13)
Size of college	3,566.14	[942.36]	3,761.39	[798.86]	3,201.44	[1,073.08]	-559.94	(36.42)
Number of majors	13.68	[4.20]	14.11	[3.66]	12.88	[4.96]	-1.23	(0.17)
Boarding college	0.21	[0.41]	0.26	[0.44]	0.12	[0.32]	-0.14	(0.02)
Women only college			0.44	0.50				
B. Route Characteristics								
Route safety (SD)	0.36	[1.09]	0.45	[1.04]	0.17	[1.15]	-0.28	(0.04)
Monthly travel cost ('000 Rs.)	1.59	[2.00]	1.78	[2.14]	1.24	1.66	-0.55	(0.08)
Travel time (minutes)	67.27	[35.71]	66.26	[34.42]	69.16	[37.95]	2.90	(1.44)

Notes: Based on sample of Delhi residents from the full survey who live at home. The socio-economic status is an index that captures the student relative wealth, its construction is explained in the Appendix SectionC. First cutoff score is for male students of general category. College characteristics describe the college chosen by the student. The absolute rank rates colleges within a major and admission year using cutoff scores from the the first cutoff list. Rank in choice set ranks the colleges to which the student was admitted to by their cutoff score for the students' major and admission year. Distance is the shortest distance from student's resident to their chosen college Annual tuition is for 2016. Size of the college is the number of students enrolled in the college.

Table 2: Travel Mode Choice and Harassment

	Total		Female		Male		(Female - Male)		Harassment
	Mean	SD	Mean	SD	Mean	SD	Mean	SE	Incidents
Auto Rickshaw	0.31	[0.46]	0.36	[0.48]	0.22	[0.41]	0.14	(0.02)	0.07
Bus	0.38	[0.49]	0.33	[0.47]	0.48	[0.50]	-0.15	(0.02)	0.40
Car or Uber	0.08	[0.27]	0.08	[0.26]	0.09	[0.29]	-0.02	(0.01)	0.14
Metro	0.42	[0.49]	0.49	[0.50]	0.29	[0.45]	0.20	(0.02)	0.16
Ladies compartment			0.80	[0.40]					
Train	0.03	[0.18]	0.04	[0.19]	0.03	[0.18]	0.00	(0.01)	0.05
Walk	0.68	[0.47]	0.66	[0.47]	0.71	[0.46]	-0.04	(0.02)	-

Notes: This table shows the mode usage by gender. Mode usage is for Delhi residents from the full survey who live at home. Harassment incidents show the proportion of harassment incidents by mode based on Safecity data.

Table 3: Bechmark Specification: Mixed Logit Model Results

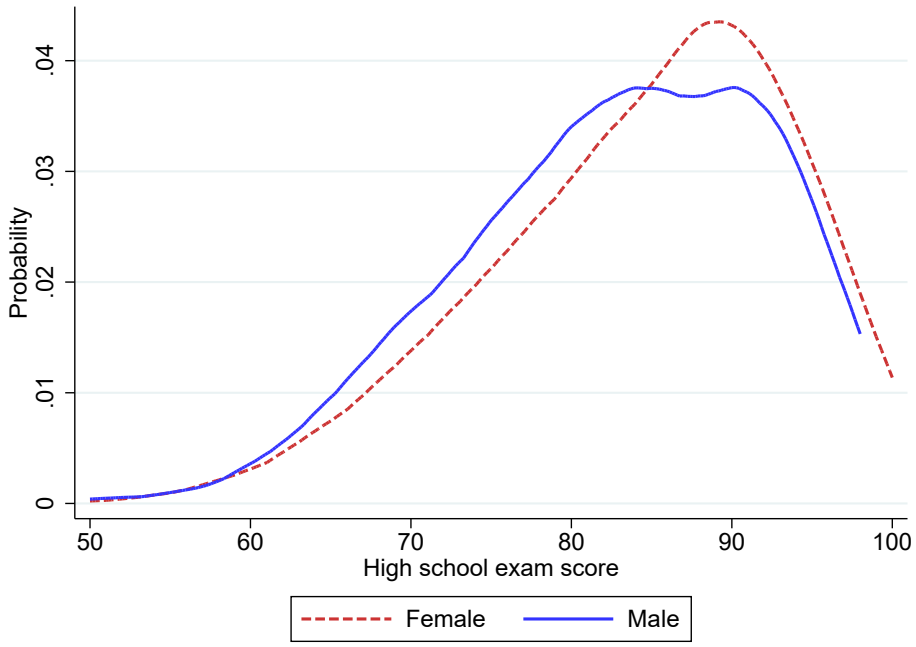
	Female		Male	
	(1)	(2)	(3)	(4)
Random coefficients				
Route safety	0.705	1.010	0.468	0.570
	(0.025)	(0.033)	(0.034)	(0.043)
Cutoff score	0.065	0.045	0.143	0.162
	(0.007)	(0.007)	(0.012)	(0.014)
Daily travel time	-0.014	-0.023	-0.014	-0.019
	(0.001)	(0.002)	(0.002)	(0.002)
Monthly travel cost	-0.256	-0.115	-0.352	-0.303
	(0.015)	(0.018)	(0.026)	(0.029)
Fixed coefficients				
College neighborhood safety		0.159		0.088
		(0.037)		(0.042)
Size of college		0.001		0.001
		(0.000)		(0.000)
Women's only college		0.591		-
		(0.056)		
Public transport mode		1.581		0.601
		(0.077)		(0.109)
Number of students	1,767	1,767	946	946
Observations	289,121	298,121	112,958	112,958
Log-likelihood	-7985.66	-6825.76	-3901.95	-3807.22
Mean MRS (Safety, Score)				
(pp per SD of safety)	-8.798	-13.518	-2.112	-2.932
2.5 percentile	-39.048	-28.790	-3.299	-12.204
97.5 percentile	-3.152	-7.694	-1.556	-1.189
Mean MRS (Safety, Time)				
(minutes per SD of safety)	26.773	22.698	20.798	18.135
2.5 percentile	24.630	19.698	19.022	16.108
97.5 percentile	30.468	27.744	24.409	22.622
Mean MRS (Safety, Cost)				
('000 Rs. per SD of safety)	1.453	4.473	0.829	1.127
2.5 percentile	1.329	4.279	0.760	1.042
97.5 percentile	1.749	4.857	1.052	1.392

Notes: The data comes from a survey of students at Delhi University in January-April 2016. Each observation is a unique student-route pair. The dependent variable is an indicator equal to one for the route reported by the student. The random coefficients associated with route safety, cutoff score are estimated using a triangular distribution and the daily travel time and monthly route cost coefficients are estimated using a restricted triangular distribution, which are both assumed to be non-positive. The other coefficients are assumed to be fixed. The MRS is the negative ratio of the coefficient estimates on route safety over score or cost or time, and its standard error is estimated using the delta method. MRS are measured in terms of the SD of route safety across the predicted route alternatives within a students' choice set, the calculation is explained in section **??** of the Appendix.

Online Appendix

A Additional Figures

Figure A1: High School Exam Scores for Women and Men



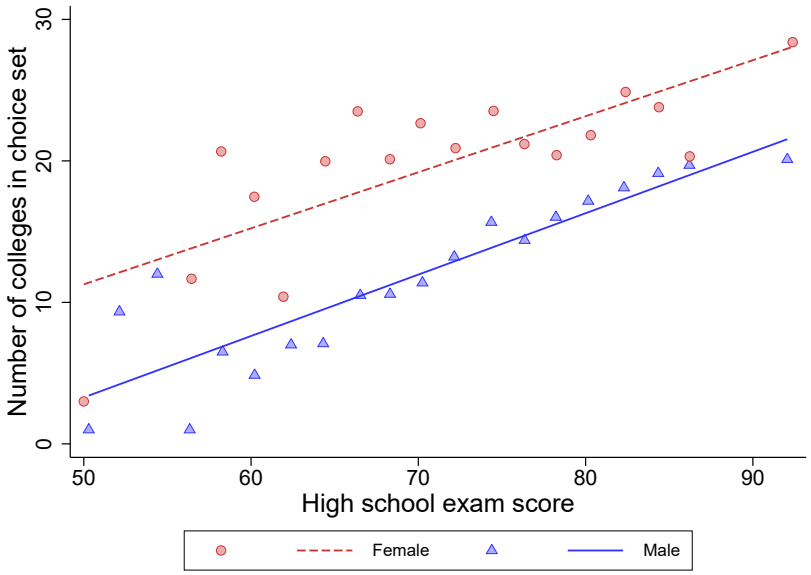
Notes: The figure shows the probability distribution function of high school exam scores for students in the full survey sample and short survey sample, who are Delhi residents and live at home.

Figure A2: 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, Mathematics and best two elective subjects)	80% (One Language, Mathematics and best two elective subjects)	80% (One Language, Mathematics 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.

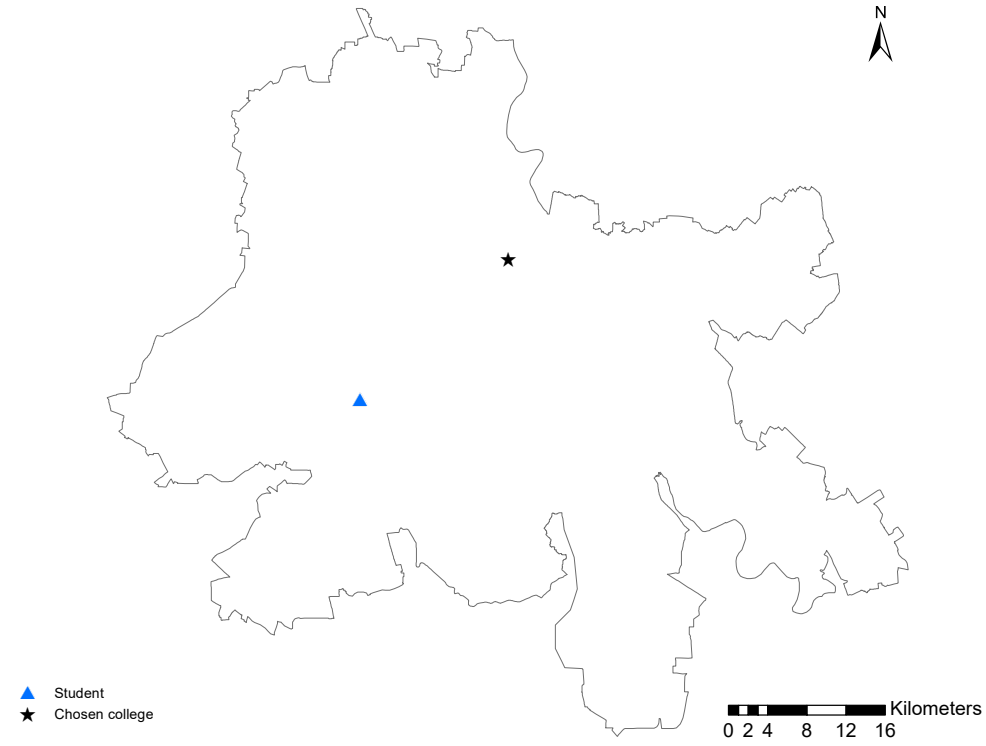
Figure A3: Correlation Between the Number of Choice Set Colleges and Students High School Exam Score



Notes: The figure shows the correlation between the number of colleges in a student's choice set and their high school exam score using a binned scatter plot. The binned scatter plot groups the high school exam score in 18 equal sized bins and plots the mean number of colleges in student's choice set for each bin.

Figure A4: An example of route mapping

(a) Student and chosen college



(b) Reported route to chosen college

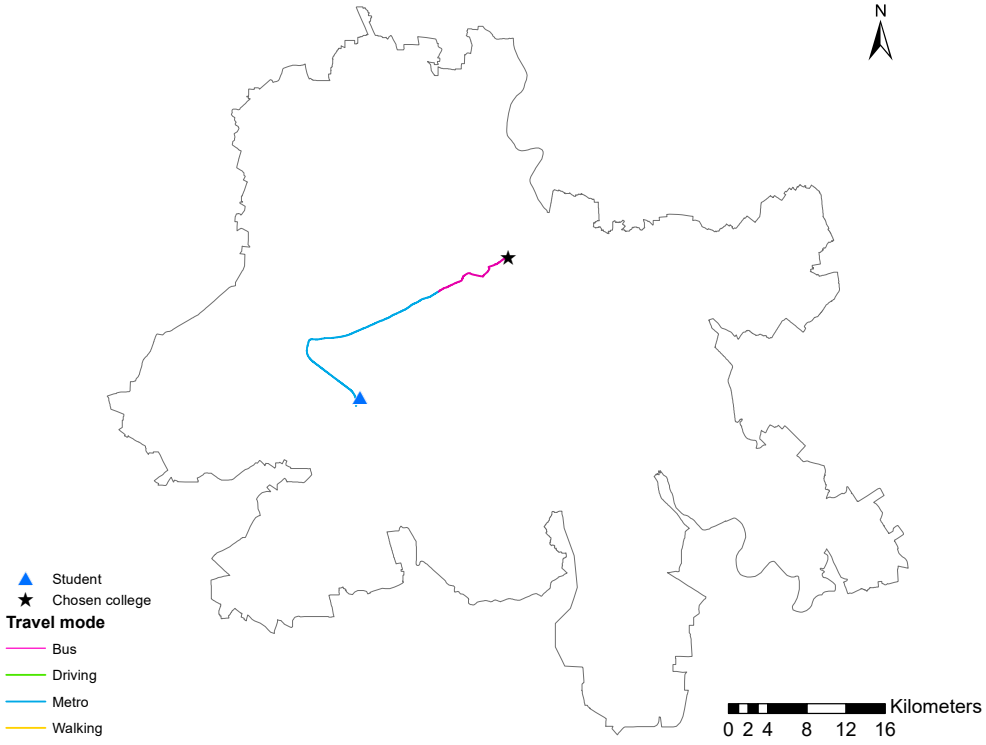
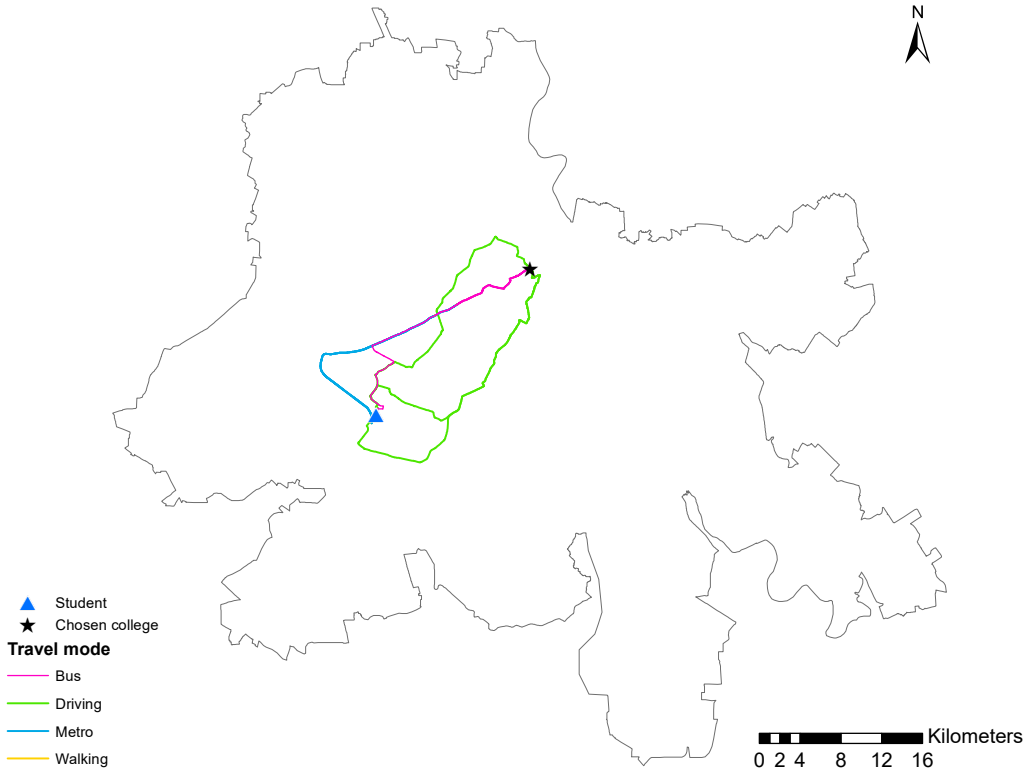


Figure A4: An example of route mapping

(c) Route options to chosen college



(d) Route options to colleges in choice set

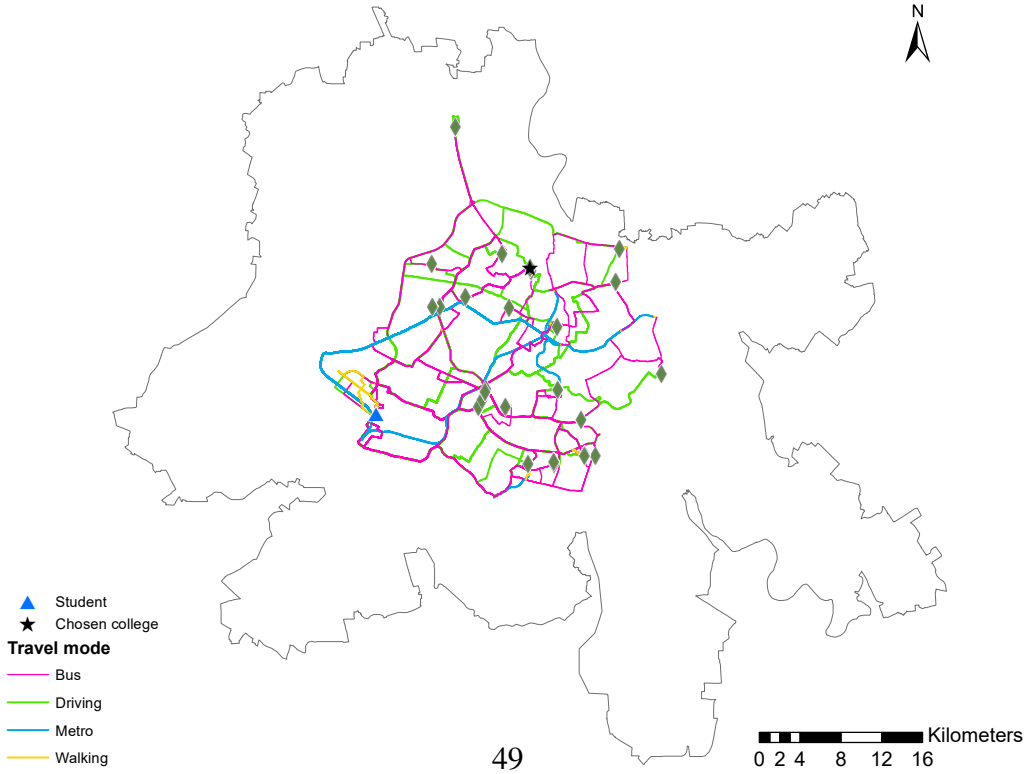
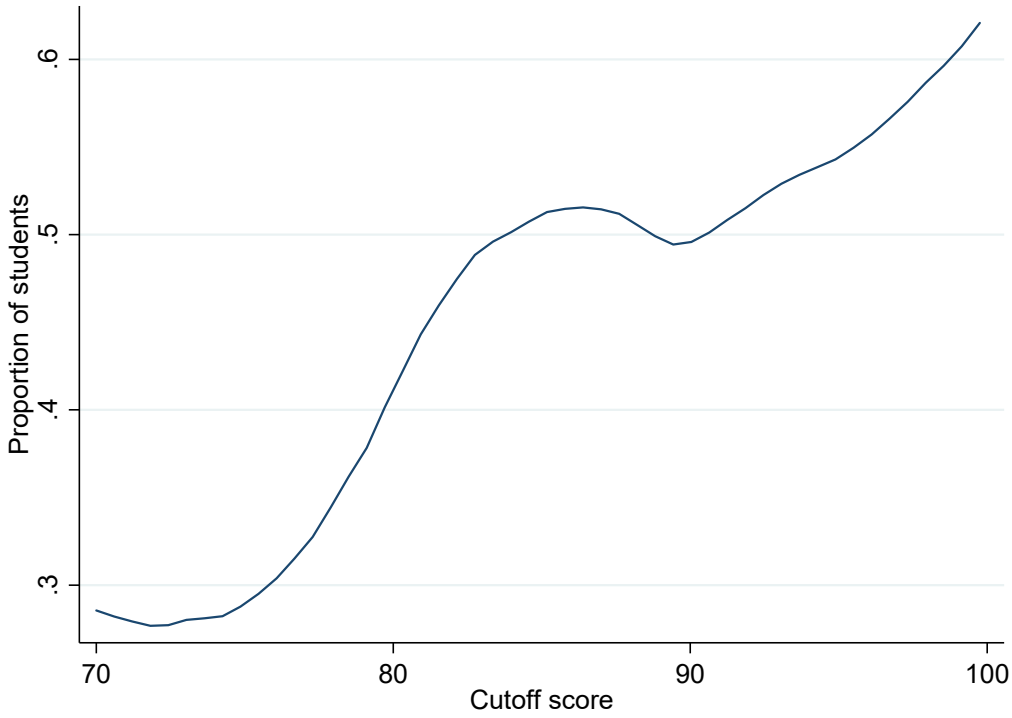
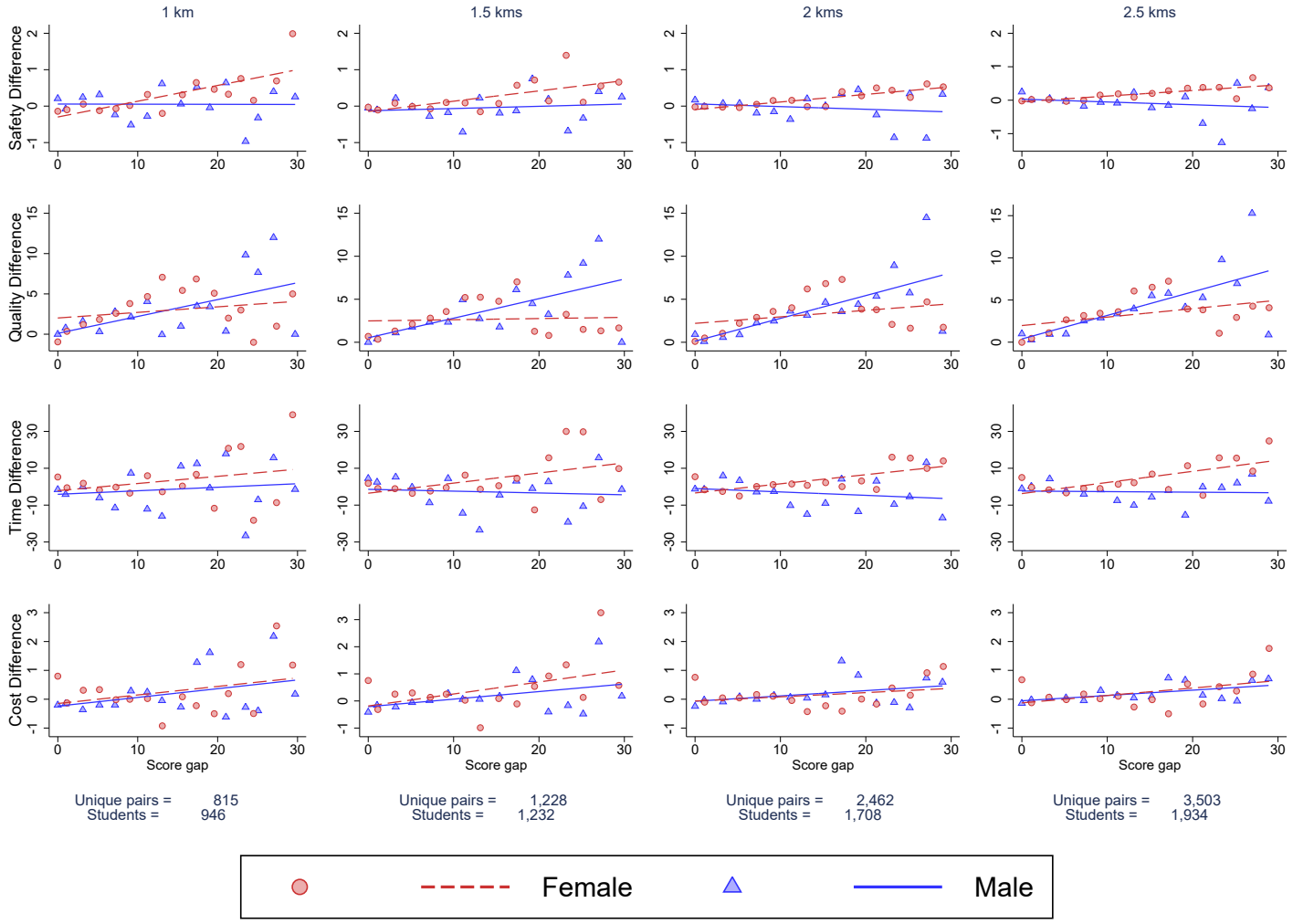


Figure A5: Proportion of Accepted Students who Enrolled



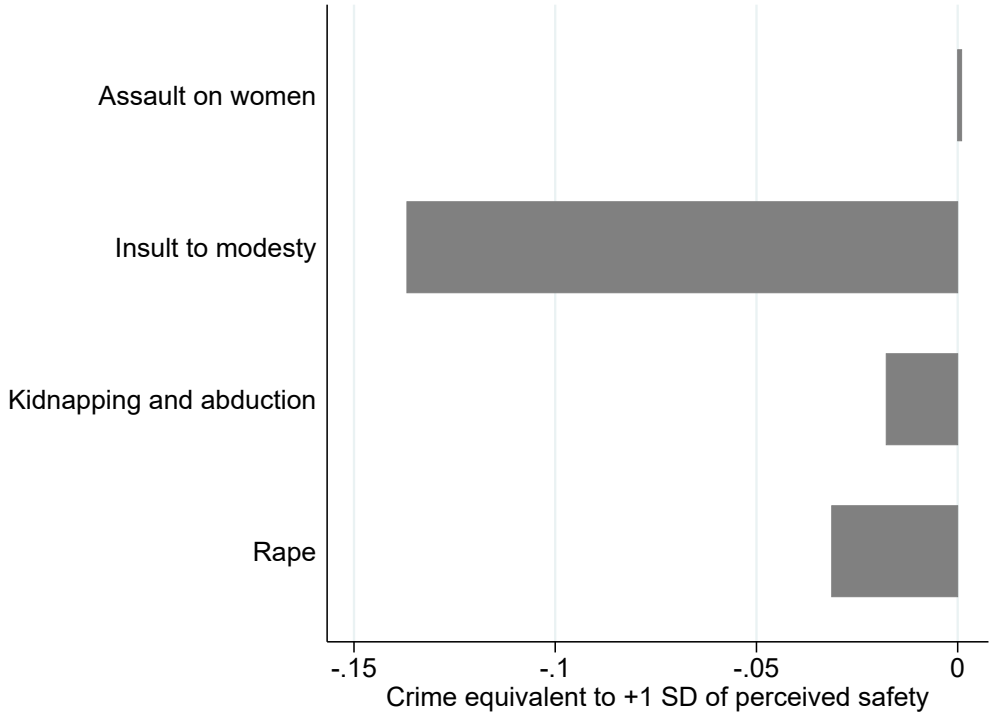
Notes: The figure shows the relationship between the proportion of accepted male students who enrolled and the cutoff score of the college. Results are from a local polynomial regression with a bandwidth of 3.5.

Figure A6: Robustness of Choice Relative to Neighbors



Notes: The figures plot binned scatter plots of difference in travel safety (SD), college quality (percentage), daily travel time (minutes), and monthly travel cost (thousand Rs.) 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 high school score difference between the index student and the neighbor. A neighbor is defined as a student living within a 1 km, 1.5 km, 2 km and 2.5 km radius of the index student and has the same gender, major, and admission year.

Figure A7: Reported Crime and Perceived Safety



Notes: The figure shows the coefficient from 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. Data on crimes is from the National Crime Records Bureau. The four types of crime are all the crimes against women that could potentially take place in public spaces.

B Additional Tables

Table A1: Comparison of Student Characteristics

Data	Female			Male		
	Full Survey	Short Survey	Admin. Data	Full Survey	Short Survey	Admin. Data
Delhi residents	1,767	459	11,450	946	171	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	0.50
		[2.19]	[2.25]		[0.81]	[-3.05]
SC	0.12	0.11	0.14	0.20	0.14	0.19
		[-0.39]	[2.39]		[-1.91]	[-0.79]
ST	0.01	0.01	0.02	0.02	0.02	0.03
		[-1.07]	[1.80]		[-0.22]	[2.07]
OBC	0.11	0.18	0.11	0.22	0.25	0.27
		[3.69]	[-0.94]		[0.97]	[3.34]
High school exam score (%)	84.88	83.90	-	82.70	82.03	-
		[-2.05]			[-0.84]	
Distance to college (kms.)	13.06	11.98	13.34	13.05	15.63	14.30
	[-2.14]	[1.28]		[2.87]	[3.68]	
Distance to center (kms.)	15.42	12.11	15.05	15.99	15.93	16.49
		[-7.96]	[-1.92]		[-0.09]	[1.65]

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 A2: SafetiPin Parameters

S.No.	Parameters	Score 0	Score 1	Score 2	Score 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.

Notes: This is the SafetiPin safety audit rubric that shows the options available to an app user when conducting an audit (Viswanath and Basu (2015)).

Table A3: Summary Statistics for non-Delhi residents

	Total		Female		Male		Female - Male	
	N = 1,031		*N = 564*		*N = 467*			
	Mean	SD	Mean	SD	Mean	SD	Mean Diff.	SE
A. Student Characteristics								
High school exam score (%)	89.15	7.37	90.57	6.33	87.44	8.14		-3.13
							(0.45)	
SES index	0.56	0.26	0.64	0.24	0.47	0.26		-0.17
							(0.02)	
B. Route Characteristics								
Route safety (SD)	1.17	1.26	1.41	1.14	0.87	1.34	-0.54	(0.08)
Monthly travel cost ('000 Rs.)	0.81	1.39	0.78	1.43	0.84	1.34	0.05	(0.09)
Travel time (minutes)	36.46	38.52	30.18	39.43	44.03	36.00	13.85	(2.37)

Notes: Based on sample of students who are not residents of Delhi from the full survey. The socio-economic status is an index that captures the student relative wealth, its construction is explained in the Appendix SectionC. First cutoff score is for male students of general category. College characteristics describe the college chosen by the student. The absolute rank rates colleges within a major and admission year using cutoff scores from the the first cutoff list. Rank in choice set ranks the colleges to which the student was admitted to by their cutoff score for the students' major and admission year. Distance is the shortest distance from student's resident to their chosen college Annual tuition is for 2016. Size of the college is the number of students enrolled in the college.

Table A4: Predicting cutoff scores for women

	Advantage to women in 2014
Proportion female in 2013	-0.088
	(2.049)
College neighborhood safety	0.057
	(0.244)
Boarding	-1.967
	(0.626)
Number of majors	0.037
	(0.054)
Size of college	0.000
	(0.000)
Annual tuition	-0.000
	(0.000)
Constant	1.597
	(0.994)
Mean of Y	1.415
Observations	41

Notes:Advantage to women is the average advantage given to female students across majors in a year by the college. It is measured in percentage points, for example if the advantage given to female students is 1pp then the cutoff score for females is 1pp lower compared to the cutoff score for male students. College neighborhood safety is the area safety of a 1.5km radius around the college. Boarding college is an indicator equal to 1 for colleges that have boarding facilities. Size of college is the total number of students in college in 2013, both from DU's annual report from 2013-14. Annual tuition is the tuition for college in 2016. Sample includes all general education co-educational colleges in DU.

Table A5: 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 A6: 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.

Table A7: Bechmark Specification: Mixed Logit Model Results

	Female (1)	Male (2)
Panel A: Excluding "Crowd"		
Route safety (SD)	0.504	0.394
	(0.027)	(0.039)
Cutoff score	0.062	0.142
	(0.006)	(0.012)
Route time (mins.)	-0.015	-0.015
	(0.001)	(0.002)
Monthly travel cost (thousand Rs.)	-0.173	-0.278
	(0.013)	(0.024)
Panel B: Excluding "Light"		
Route safety (SD)	0.496	0.325
	(0.027)	(0.039)
Cutoff score	0.060	0.139
	(0.006)	(0.011)
Daily travel time (mins.)	-0.015	-0.016
	(0.001)	(0.002)
Monthly travel cost (thousand Rs.)	-0.175	-0.279
	(0.013)	(0.024)
Number of students	1,767	946
Observations	289,121	112,958

C Data

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.

A Route mapping

Students reported their travel routes as a combination of n landmarks and $n-1$ modes of transport with the first landmark being the student's home address and final landmark being the college location. The route is split into "legs" by landmarks, in the data a maximum number of four legs are reported by any student. All routes are mapped using Google Maps API. The student reported routes are mapped as a sequence of legs. The routes take into account student's reported departure time. For the reported routes where student's travel by a rickshaw, the travel times are mapped as walking routes and the travel time is then adjusted to be 60 percent of the walking travel time.
For the potential travel routes, up to four routes are extracted per routing option available in Google Maps, these are the top suggested routes by Google. The routing options in Google Maps are driving, walking and public transit. The driving and walking routes are unimodal. The public transit routing options in Google Maps includes bus, subway, train, tram and rail. If this was to be done interactively in Google Maps then you would search for directions from origin (student's home) to destination (choice set college) and then record the different route options provided by Google Maps by each Google Maps routing option. The potential travel routes also take into account the student's departure time, the departure time to choice set colleges is assumed to be the same as the reported departure time to their chosen college if the chosen college and the choice set college are both morning/afternoon colleges. If not then for morning colleges in the choice set, the departure time is assumed to be 9am and for afternoon colleges it is assumed to be 2pm. For the potential travel routes, a date in the future has to be set for the student to travel to college, this is set to the Wednesday following the date the script is run. Thus if the script is run on Tuesday, the student will be modeled as traveling at their designated time the following day. Setting such a time in the future ensures consistency across choice set routes. Wednesday was chosen because it's the day least likely to be affected by any holidays or metro maintenance, but it could be any day of the week that you'd like. All travel times take into account traffic conditions (as opposed to a traffic-free journey).

B Travel Cost

Below I specify how the travel cost is calculated by travel mode, where C_i = cost in Rupees for model i, d = distance in kms and t = time in minutes.

- Auto: $C_a = 25 + 8(d-2)$, where Rs. 25 is charged at hire for the first two kilometers and Rs. 8 for every subsequent kilometer.
- Car:$C_c = \frac{d}{13} \times 60$, assuming an average mileage of 13km/liter and Rs. 60/liter as the cost of fuel, which was the average price of petrol in Delhi from September 2015 to August 2016.
- Carpool:$C_{cp} = \frac{d}{13} \times 60 \times \frac{1}{3}$, assuming that three people on average travel by carpool.
- Bus: Rs. 115 per month for the monthly student pass.
- Metro: The following fares are used, which are the official metro fares that were effective from November 13, 2009 to May 10, 2017.[40]

Table A8: Metro fares

Distance Zones (Kms.)	Fare (Rs.)
0-2	8
2-4	10
4-6	12
6-9	15
9-12	16
12-15	18
15-18	19
18-21	21
21-24	22
24-27	23
27-31	25
31-35	27
35-39	28
39-44	29
> 44	30

Notes: These fares were effective from November 13, 2009 to May 10, 2017. Source: Delhi Metro Rail Corporation. http://www.delhimetrorail.com/ (last accessed: May 28, 2020)

- Rickshaw: A flat rate of Rs. 10 per rickshaw ride as per the rate charged by E-rickshaws in Delhi.
- Motorcycle:$C_m = \frac{d}{50} \times 60$, assuming an average mileage of 50km/liter and Rs. 60/liter as the cost of fuel.

[40] As given here: (last accessed: May 28, 2020).

- Scooter: $C_s = \frac{d}{40} \times 60$, assuming an average mileage of 40km/liter and Rs. 60/liter as the cost of fuel.
- Taxi/Uber:$C_u = 40 + t + 6d$ based on UberGo fares.
- Walking: Rs. 0
- Train: The fares for monthly train passes are shown below, such a pass enables travel by second class. Following the official guidelines a 50 percent dicount is applied for all general category students and 75 percent discount for SC and ST students. Students are also assumed to purchase a quaterly train pass which gives them an additional dicount of 10 percent on the fare.[41]

[41] As given here: https://nfr.indianrailways.gov.in/(last accessed: May 28, 2020).

Table A9: Train fares for a monthly travel pass

Distance (Kms.)	Fare (Rs.)
1-10	60
11-15	75
16-20	90
21-30	105
31-35	120
36-40	135
41-45	150
46-50	165
51-55	180
56-60	195
61-65	195
66-70	210
71-75	225
76-80	240
81-85	235
86-95	270
96-100	285
101-110	300
111-115	315
116-125	330
126-135	345
136-140	360
141-150	375

Notes: These are monthly fares for Second class travel by a general category adult. Season ticket for general category students are at a discount of 50 percent and at a discount of 75 percent for SC / ST students. Quarterly tickets offer an additional discount of 10 percent.Source: https://nfr.indianrailways.gov.in (last accessed: May 28, 2020)

C Wealth Index

An index of the students' socio-economic status is created in two steps. First using principal component analysis a wealth index is created, the variables included are: dummy for residing in an owned house, dummy for owning a laptop and/or computer, amount of money spent every month, excluding travel expensed and rent), in INR, dummy for private school attended in Class 12, years of education completed by father, years of education completed by mother, number of cars owned by student's permanent household, number of scooters owned by student's permanent household,

and price of the most expensive car owned by student's permament household.[42]Next, a continuous measure of the student's relative wealth is created by ranking them according to this wealth index, from 0 (poorest) to 1 (wealthiest).

[42]The students' were asked to list the "type/make and model" of all the cars owned by their pemanent household, the price was extracted from https://www.carwale.com (last accessed June 25, 2017).