Online Appendix for College Major Choice and the Gender Gap

Basit Zafar*

A.1 Comparison with *Stated* Preferences

It would be interesting to see how the estimated preference parameters (in section 5) compare with what respondents claim were important determinants in their choice of major. In a follow-up survey, respondents were asked to assign an integer between zero and 100 to a list of reasons such that the numbers added up to a 100. Table A1 shows the average weights assigned to the various reasons given by males and females. I interpret these numbers as the relative preference for the given reason in the choice of major. Enjoying work at the jobs and learning more about things that interest them were the two most important reasons for choosing a major for both males and females. However, females, on average, assign higher weights to these reasons (the gender difference is significant). For males, the third most important stated reason for choosing a major is getting a high-paying job. Conversely, doing well in the coursework is the third most important reason for females. These stated preferences for various outcomes are consistent with the parameter estimates discussed in section 5.

A.2 Estimation with Heterogeneous Preferences

Preferences for the different outcomes could depend on individual characteristics (other than gender). Several empirical studies have documented the influence of family and society in the endogenous formation

*Research and Statistics, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045. E-mail: Basit.Zafar@ny.frb.org

of preferences (Fernandez et al., 2004; Guiso et al., 2006). Such heterogeneity, if not accounted for, may bias the estimates presented in section 5. I now relax the assumption of section 5.1 that the utility for each binary outcome $u_r(b_r)$ and the constants γ_q for the continuous outcomes do not depend on individual characteristics other than gender. Though I have relatively rich demographic information on the respondents, it is not possible to account for heterogeneity in all outcomes because of the small sample size. I therefore consider heterogeneity along the following dimensions:

1. An individual might be more inclined to ensure that her parents approve of her choice of major if she relies on them for college support. Moreover, concern for parents' approval might depend on the individual's cultural and ethnic background. I allow for heterogeneity in the utility for approval of parents by incorporating the financial support an individual receives from her parents when in college and whether her parents are foreign-born or not. Section A.4 of this appendix provides discussion of what majors parents are more likely to approve.

2. Children growing up in divorced or separated households make different choices than other individuals (Gruber, 2004). Here, I consider the effect of growing up in such a household on the individual's preference for being able to reconcile work and family.

3. An individual's preference for the social status of jobs may vary by her cultural background. This heterogeneity is accounted for by taking into account whether the individual's parents are foreign-born.

4. If individuals have declining marginal utility of consumption, and preferences are separable in consumption and non-pecuniary outcomes, an individual from a low-income family will value the income profiles associated with the majors more. I account for this heterogeneity by including information on parents' annual income. I also allow for heterogeneity by taking into account whether an

individual's parents are foreign-born or not.

The enriched utility function for individual i is:

$$U(X_i, \{P_{im}(b_r), E_{im}(d_q)\}_{r\in\{1,.,7\},q\in\{1,.,4\}})$$

$$= \sum_{r=\{1,2,3,5,6,7\}} P_{im}(b_r = 1)\triangle u_r + \triangle u_4[\textit{parents'\_support}_i \times (1\text{-}\textit{Foreign}_i) \times P_{im}(b_4 = 1)]$$

$$+\widetilde{\triangle u_4}\ [\textit{parents'\_support}_i \times \textit{Foreign}_i \times P_{im}(b_4 = 1)] + \widetilde{\triangle u_7}\ \ [\textit{divorced}_i\ \times\ P_{ijt}(b_7 = 1)]$$

$$+ \textstyle\sum_{q=1}^{2} \gamma_q E_{im}(d_q) +\ \gamma_3\ [(1\text{-}\textit{Foreign}_i)\ \times E_{im}(d_3)]\ +\ \widetilde{\gamma_3}\ [\textit{Foreign}_i\ \times E_{im}(d_3)] + \gamma_4 E_{im}(d_4)$$

$$+ \gamma_4^{HI}[E_{im}(d_4) \times\ (1\text{-}\textit{low\_inc}_i) \times (1\text{-}\textit{Foreign}_i)]\ \ +\ \widetilde{\gamma_4^{HI}}[E_{im}(d_4) \times\ (1\text{-}\textit{low\_inc}_i) \times \textit{Foreign}_i]$$

$$+ \gamma_4^{LI}[E_{im}(d_4) \times\ \textit{low\_inc}_i \times (1\text{-}\textit{Foreign}_i)]\ +\ \widetilde{\gamma_4^{LI}}[E_{im}(d_4) \times \textit{low\_inc}_i \times \textit{Foreign}_i]\ + \varepsilon_{im}$$

$$\forall\ m = 1, ., 8$$

where *low_inc* is a dummy variable that equals one if the individual's parents earn less than $150,000 annually; *parents'_support* captures the financial support an individual receives from her parents,[1] *Foreign* is a dummy that equals one if either of the individual's parents is foreign-born, and *divorced* is a dummy that equals one if the individual's parents are either separated or divorced. In this specification, the parameter $\triangle u_4$ ($\widetilde{\triangle u_4}$) is the gain in utility from parent's approval as a function of financial dependence on parents for students with US-born parents (at least one foreign-born parent); γ_3 ($\widetilde{\gamma_3}$) is the utility gain from the social status of jobs for respondents with US-born parents (at least one foreign-born parent); and the various γ_4's are parameters on expected income.

[1] It is increasing in the financial support an individual receives from her parents. Parents' support = 1 if no education expenses are paid by one's parents; equals 2 if they pay less than $5,000; equals 3 if they pay between $5,000 and $10,000; equals 4 if they pay between $10,000 and $15,000; equals 5 if they pay between $15,000 and $25,000; and equals 6 if they pay more than $25,000.

I continue to assume that the random terms $\{\varepsilon_{ik}\}$ are independent for every individual i and choice k. Column (1) of Table A2 presents the maximum-likelihood estimates of this model for the pooled sample using stated preference data. Estimated coefficients of the outcomes for which heterogeneity is not considered are similar to those in the specification with homogenous preferences (column (1) of Table 5). With this enriched specification, the difference in utility levels for parents' approval is 0.32 for individuals with US-born parents who do not receive college support from them, and 1.90 for individuals who annually receive more than $25,000 in college support from their parents. This is consistent with the hypothesis that approval of parents matters more to individuals who depend on their parents for college funding. However, I don't find support for this hypothesis for individuals with foreign-born parents. The difference in utility levels for reconciling work and family continues to be insignificant. Introducing heterogeneity for the status outcome gives an interesting result. Status of the available jobs, an important determinant in the choice in the earlier specifications, is only significant (and much larger in magnitude when compared to earlier specifications) for students with US-born parents. This suggests that the large positive coefficient on the social status of jobs in earlier specifications is being driven by the preferences of individuals with foreign-born parents in the sample. The coefficient on income at age 30 is still not significantly different from zero. However, there is weak support for the hypothesis that individuals from low-income households value the future earnings profile more in their choice.

Columns (2) and (3) of Table A2 present the results of the heterogeneous choice model for the male and female sub-samples, respectively. In order to gain an insight into the magnitude of these parameters, Table A3 shows the results of the decomposition methodology outlined in equation (8). Except for males with foreign-born parents, non-pecuniary attributes explain more than half of the choice. For individuals with US-born parents, more than two-thirds of the choice is driven by non-pecuniary motivations; the non-pecuniary outcomes at college are of utmost importance to this group. For individuals with foreign-born

parents, pecuniary outcomes at the workplace are of greatest value in the choice.

The analysis in this section indicates that demographic characteristics bias preferences in favor of certain outcomes.

A.3 Robustness Check

The model estimated in section 5 assumes that all individuals have homogeneous preferences for various outcomes. Individuals with different characteristics are very likely to have different preferences. Moreover, the assumption that the random terms $\{\varepsilon_{ik}\}$ are independent for every individual i and choice k might be very strong. Though a model with limited heterogeneity in preferences is estimated in section A.2, any unaccounted or unobserved heterogeneity may bias the model estimates. In this section, I specify a random parameters logit model to account for these issues (see Revelt and Train, 1998, for a discussion of mixed logit models). One could allow heterogeneity in preferences for all outcomes, but I focus on the most important outcomes: I consider a model in which the differences in utility levels for graduating with a GPA of at least 3.5, enjoying the coursework, gaining approval of parents, enjoying work at the available jobs, and the parameter for social status of the available jobs are allowed to vary in the population with a specified distribution. The utility that individual i receives from choosing major m is:

$$U(X_i, \{P_{im}(b_r), E_{im}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\}})$$
$$= \sum_{r=\{1,5,7\}} P_{im}(b_r = 1)\triangle u_r + \sum_{s=\{2,3,4,6\}} P_{im}(b_s = 1)\triangle u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{im}(d_q) + \gamma_{3i} E_{im}(d_3) + \varepsilon_{im}$$

where $\triangle u_{si}$ for $s = \{2, 3, 4, 6\}$ and γ_{3i} are allowed to vary in the population according to a specified parametric distribution, and ε_{im} is an iid random term that is extreme value distributed. I denote the vector of parameters $\{\triangle u_{2i}, \triangle u_{3i}, \triangle u_{4i}, \triangle u_{6i}, \gamma_{3i}\}$ by $\boldsymbol{\beta}_i$, and the density of these parameters $f(\boldsymbol{\beta}_i|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$

are the parameters of the distribution. The probability of i choosing the major m *conditional* on $\boldsymbol{\beta}_i$ is:

$$\Pr(m|\boldsymbol{\beta}_i\,) = \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\,k\in C_i}, \boldsymbol{\beta}_i\,) =$$

$$= \frac{\exp(\sum_{r=\{1,5,7\}} P_{im}(b_r=1)\triangle u_r + \sum_{s=\{2,3,4,6\}} P_{im}(b_s=1)\triangle u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{im}(d_q) + \gamma_{3i} E_{im}(d_3))}{\sum_{k\in C_i} \exp(\sum_{r=\{1,5,7\}} P_{ik}(b_r=1)\triangle u_r + \sum_{s=\{2,3,4,6\}} P_{ik}(b_s=1)\triangle u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{ik}(d_q) + \gamma_{3i} E_{ik}(d_3))}$$

The unconditional probability of choosing m is the integral of this conditional probability over all possible values of $\boldsymbol{\beta}_i$ and depends on the parameters $\boldsymbol{\theta}$ of the distribution of $\boldsymbol{\beta}_i$. The *unconditional* probability for i choosing m is:

$$P_{im}(\theta) = \int \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\,k\in C_i}, \boldsymbol{\beta}_i\,) f(\boldsymbol{\beta}_i|\boldsymbol{\theta}) d\boldsymbol{\beta}_i$$

This integral is approximated through simulation since it cannot be calculated analytically.[2] The log-likelihood function $\sum_i \ln(\Pr_i\,)$ is approximated by the simulated log-likelihood function $\sum_i \ln(\widehat{P_i(\theta)})$, and the estimated parameters are those that maximize the simulated log-likelihood function. I assume that the coefficients for graduating with a GPA of at least 3.5, enjoying the coursework, gaining the approval of parents, enjoying work at the available jobs, and valuing the social status of the available jobs are independently log-normally distributed.[3]

Columns (1a)-(1c) in Table A4 present the estimates of the mixed logit specification for the model with

[2]For a given value of the parameter vector $\boldsymbol{\theta}$, a value of $\boldsymbol{\beta}_i$ is drawn from its distribution. Using this draw, I calculate the conditional probability. This process is repeated for D draws, and the average is taken as the approximate choice probability:

$$\widehat{P_{im}(\theta)} = \frac{1}{D}\sum_{d=1}^{D} \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,3\};\,k\in C_i}, \boldsymbol{\beta}_i^d)$$

[3]I use a log-normal distribution instead of a normal distribution for these parameters since these are all outcomes that one would expect to be desirable to an individual. The normal distribution allows coefficients of both signs and implies that some share of the sample has negative coefficients for those outcomes, whether or not it is true. The log-normal assumption ensures that each respondent in the sample has a positive coefficient for these outcomes.

The difference in utility levels for an outcome k that is assumed to vary in the population is expressed as $\triangle u_k = \exp(\overline{\triangle u_k} + \sigma_k \mu_k)$, where μ_k is a standard normal deviate. The parameters $\overline{\triangle u_k}$ and σ_k, which represent the mean and standard deviation of $\log(\triangle u_k)$, are estimated. The mean and standard deviation of $\triangle u_k$ are $\exp(\overline{\triangle u_k} + \frac{\sigma_k^2}{2})$ and $\exp(\overline{\triangle u_k} + \frac{\sigma_k^2}{2}) * \sqrt{(\exp(\sigma_k^2) - 1)}$, respectively.

$D = 5,000$. Estimates of various outcomes are similar to those obtained in the corresponding model with no heterogeneity (column 1 of Table 5). The mean coefficient of enjoying coursework is still largest in absolute value and significant. The estimated standard deviations of the (random) coefficients are highly significant, indicating that these parameters do indeed vary in the sample. Standard deviations for coefficients of graduating in 4 years and social status of available jobs are especially very large, indicating that there is substantial heterogeneity in how these outcomes are valued in the sample. Another point of note is that the mean coefficients in the mixed logit model are larger than the corresponding fixed coefficients in Table 5. This is because, in the mixed logit, some of the stochastic portion of the utility is captured in $\boldsymbol{\beta}_i$ rather than in ε_i. Since the utility is scaled so that ε_i has the variance of an extreme value, the parameters are scaled down in the standard model relative to the mixed logit model (the same result is obtained by Revelt and Train, 1998).

One might wonder about the extent to which the variation in the parameters in the mixed logit model can be explained by including demographic characteristics. Columns (2a) through (2c) in Table A4 present estimates of the mixed logit model with demographic variables that were used in the heterogeneous model described in section A.2. The estimates are similar to those in column (1) of Table A2, though they are larger in magnitude, which is expected. The standard deviations are still large and significant, which indicates that the demographic variables considered in section A.2 capture only some of the heterogeneity exhibited by the individuals. Nonetheless, the fact that the relative magnitude of the estimates is similar to previous results is reassuring.

A.4 Parents' Approval

Though section A.2 shows that one channel through which parents' approval matters is the individual's reliance on them for college support, it is not clear which majors parents are more likely to approve or what criteria they use for approving a major. Since only the beliefs of students are observed, I can only study

the relationship between students' beliefs about parents' approval of a major and their own beliefs about other outcomes associated with the choice.[4] Controlling for the individual's major, I regress respondent i's beliefs about her parents' approval for major j on her beliefs about the other outcomes associated with j. More specifically, I consider the following regression model:

$$P_{ij}(b_4 = 1) = \delta_i + \lambda_j + \boldsymbol{\alpha}^{'} X_{ij} + \boldsymbol{\beta}^{'} \left[\sum_{\substack{c=1 \\ c \neq 4}}^{7} P_{ij}(b_c = 1) + \sum_{q=1}^{4} E_{ij}(d_q) \right] + \varepsilon_{ij}$$

where δ_i is an individual fixed-effect, λ_j is a field-fixed effect, X_{ij} is a vector of individual-specific controls, and $\boldsymbol{\beta}$ is the vector of interest. The results are presented in Table A5. Students' beliefs about parents' approval for a given major increase in their beliefs of finding a job upon graduation, enjoying work at potential jobs, and the social status of jobs. Expectation of parents' approval for a major increases by nearly 3 points (on a scale of zero to 100) if the probability of finding a job upon graduation in that major increases by 10 points. This effect is even stronger for students with foreign-born parents: Students believe that switching to a major with a 10-point higher probability of getting a job upon graduation is likely to increase parents' approval by nearly 5 points. A positive and significant effect, half in magnitude to that of finding a job, is found for the social status of the jobs. Again, the effect is stronger for students with foreign-born parents. The only other outcome that affects beliefs about parents' approval, particularly for female students, is the expectation of enjoying work at the jobs. Another notable point is that, for females only, parents' approval is higher by about 5 points for one's chosen major.

[4] It could be that parents have subjective beliefs about the outcomes that are very different from those of the student. However, I can only analyze the relationship the student *believes* exists between her expectation of parents' approval and her subjective expectations of the various choice-specific outcomes.

T

Table A1: Stated reasons for choosing a major

How imp. were the following reasons in choosing a major:[a]	Males	Females
My parents wanted me to	6.02[b]	5.33
	(9.60)	(11.13)
A mentor/ role model encouraged me to	7.31*	4.27
	(12.00)	(7.99)
My siblings made the same choice	1.80**	0.29
	(5.45)	(1.17)
My high school friends and peers made the same choice	1.43	1.21
	(3.51)	(5.27)
The societal reputation of the choice	7.75	7.71
	(10.01)	(11.74)
To be able to get a high-paying job	14***	7.92
	(11.80)	(10.43)
To be able to get a job where I could balance work & family	8.76*	6.06
	(8.92)	(7.64)
To be able to get a job in a field where people of my gender are not discriminated against	0**	0.80
		(2.34)
To get a job that I would enjoy	18.68*	23.15
	(13.73)	(15.40)
To get training for a specific career	7.24	7.57
	(9.69)	(9.19)
To learn more about things that interest me	18.96**	25.44
	(16.23)	(18.16)
To be able to do well in the coursework of the major	7.05	8.45
	(7.72)	(9.48)
Fraction of ppl of my gender teaching classes in the major	0	0.18
		(0.89)
Fraction of people of my gender taking classes in the major	0	0.076
		(0.62)
Fraction of ppl of my gender in jobs related to the major	0.29	0.15
	(1.19)	(0.86)
Other Reasons	0.69	1.36
	(4.90)	(6.53)

[a] The exact question was: "In deciding your major, how important to you was each of the following reasons? For this question you need to assign an integer between 0 and 100 to each of the following reasons. Moreover, the responses SHOULD ALL SUM TO 100."

[b] Each cell is the AVERAGE contribution of the reasons for the choice of majors

Standard deviation in parentheses

* gender diff is significant at 10%; ** sig at 5%; *** sig at 1% (2-tailed T-test)

Table A2: Estimation of heterogeneous preferences using Stated Preference

	All	Males	Females
	(1)	(2)	(3)
Δu_1 for graduating within 4 years	0.164	-0.530	1.033
	(0.590)	(0.703)	(0.829)
Δu_2 for graduating with a GPA of at least 3.5	0.647*	0.381	1.025*
	(0.372)	(0.477)	(0.550)
Δu_3 for enjoying the coursework	3.180***	3.030***	3.389***
	(0.337)	(0.583)	(0.407)
γ_1 for hours/week spent on coursework[a]	0.0137	0.0216	0.0118
	(0.0091)	(0.0132)	(0.0120)
Δu_4 for parents approv $\times$ parents'_supp[d] $\times$ (1-Foreign[e])	0.317***	0.709***	0.0753
	(0.102)	(0.166)	(0.135)
$\widetilde{\Delta u_4}$ for parents approval $\times$ parents'_support $\times$ Foreign	0.0324	-0.124	0.244*
	(0.096)	(0.129)	(0.125)
Δu_5 for finding a job upon graduation	0.331	0.403	0.243
	(0.326)	(0.449)	(0.479)
Δu_6 for enjoying work at the available jobs	1.338***	0.537	1.916***
	(0.301)	(0.551)	(0.408)
Δu_7 for reconciling family and work at available jobs	0.0272	0.360	-0.257
	(0.398)	(0.565)	(0.529)
$\widetilde{\Delta u_7}$ for reconciling family & work $\times$ divorced[f]	0.725	0.614	1.182
	(0.653)	(0.988)	(0.910)
γ_2 for hours/week spent at work[c]	-0.0016	-0.0029	0.0065
	(0.0075)	(0.0118)	(0.0091)
γ_3 for social status of the available jobs[b] $\times$ (1-Foreign)	0.277	0.942	0.0843
	(0.308)	(0.488)	(0.421)
$\widetilde{\gamma_3}$ for social status of jobs $\times$ Foreign	1.966***	2.929***	0.945**
	(0.382)	(0.596)	(0.423)
γ_4^{HI} for exp. Inc at 30 $\times$ (1- low_inc[g]) $\times$ (1-Foreign)	9.49e-08	1.39e-06	-2.34e-07
	(8.32e-07)	(3.00e-06)	(6.05e-07)
$\widetilde{\gamma_4^{HI}}$ for exp Inc at 30 $\times$ (1-low_income) $\times$ Foreign	-5.97e-07	-3.18e-06	2.90e-07
	(1.31e-06)	(5.01e-06)	(4.25e-07)
γ_4^{LI} for exp. Income at 30 $\times$ low_inc $\times$ (1-Foreign)	1.38e-06	-4.63e-06	3.82e-06*
	(2.56e-06)	(3.14e-06)	(2.12e-06)
$\widetilde{\gamma_4^{LI}}$ for expected Income at 30 $\times$ low_inc $\times$ Foreign	1.00e-06	8.08e-06*	1.37e-07
	(1.54e-06)	(4.50e-06)	(8.34e-07)
Log-Likelihood	-1396.05	-598.42	-756.96
No. of individuals	161	69	92

† Estimates correspond to the estimation of a logit model on stated preference data
* significant at 10%; ** significant at 5%; *** significant at 1%; robust standard errors in parentheses
a (b) - number of hours spent per week on coursework (job) varies between 0 and 100;
c - social status is on a scale of 1-8 (8 being the highest social status); normalized to be between 0.1-0.8
all other variables (except income) are probabilities between 0 and 1
d - parents' supp = 1 if parents pay no education expenses; = 2 if pay $<$ \$5,000; = 3 if they pay \$5,000 -\$10,000; = 4 if they pay \$10,000- \$15,000; = 5 if they pay \$15,000-\$25,000; = 6 if they pay \$25,000+
e - Foreign is a dummy that equals 1 if either of the respondent's parents is foreign-born.
f - divorced = 1 if respondent's parents are divorced or separated; zero otherwise
g - low_income = 1 if parents' annual income is less than \$150,000; zero otherwise

Table A3: Decomposition Analysis

	Foreign-Born Parents		No Foreign-Born Parents	
	Males	**Females**	**Males**	**Females**
	(1)	(2)	(3)	(4)
Attributed to:				
Pecuniary Attributes	67.65%	41.40%	29.10%	7.50%
Non-Pecuniary Attributes	32.35%	58.60%	70.90%	92.50%
Attributed to:				
Parents' Approval + Enjoying Coursework	28.05%	32.30%	59.00%	54.25%
Coursework hrs/week + GPA + Graduating in 4 yrs	5.35%	12.60%	11.00%	14.05%
Finding a job + Job hrs/week + Income at 30 + Status of Job	60.95%	35.00%	24.30%	11.35%
Reconcile work & family + Enjoying Work	5.65%	20.10%	5.70%	20.35%

a Pecuniary attributes are the following pooled together: Graduating in 4 years; Graduating with GPA of at least 3.5; hrs/week spent on coursework; Finding a job upon graduation; Job hrs/week; Income at 30; Status of available jobs.
b The non-pecuniary attributes include all outcomes not included in a

Table A4: Mixed Logit Model Estimation

		Without Demographics			With Demographics		
			Mean, Dev Log-Normal[a]			Mean, Dev Log-Normal	
		Est. (Std Error)	Mean	Dev.	Est. (Std Error)	Mean	Dev.
		$(1a)$	$(1b)$	$(1c)$	$(2a)$	$(2b)$	$(2c)$
Fixed Coefficients							
Δu_2 for graduating within 4 years	Coeff	0.337 (0.462)	-	-	0.627 (0.508)	-	-
γ_1 hours/week on coursework	Coeff	0.016* (0.009)	-	-	0.019* (0.010)	-	-
$\widetilde{\Delta u_4}$ approval × parent_sup × For.	Coeff	-	-	-	-0.461*** (0.144)	-	-
Δu_5 finding a job upon graduation	Coeff	-0.110(0.315)	-	-	0.058 (0.323)	-	-
Δu_7 reconcile family & work at job	Coeff	0.213 (0.354)	-	-	-0.055 (0.401)	-	-
$\widetilde{\Delta u_7}$ reconcile family & work× Div	Coeff	-	-	-	0.647 (0.786)	-	-
γ_2 for hours/week spent at work	Coeff	-0.001 (0.006)	-	-	0.003 (0.007)	-	-
γ_4 - Expected Income at 30	Coeff	5.8e-7 (8.7e-7)	-	-	-1.2e-7 (1.1e-6)	-	-
Random Coefficients							
Δu_2 graduating with GPA≥3.5	$\overline{\Delta u}$	-0.887 (0.934)	1.80	7.12***	-1.358 (1.096)	2.44	23.85***
	σ	1.721*** (0.597)			2.112*** (0.616)		
Δu_3 for enjoying the coursework	$\overline{\Delta u}$	1.180*** (0.136)	4.26***	3.60***	1.238*** (0.137)	4.61***	4.09***
	σ	0.773*** (0.134)			0.762*** (0.142)		
Δu_4 for approval of parents	$\overline{\Delta u}$	-0.082 (0.429)	1.81	3.09***	-	-	-
	σ	1.162*** (0.296)					
Δu_4 approval × Parents'_supp	$\overline{\Delta u}$	-	-	-	-0.886*** (0.335)	0.60***	0.65***
	σ				0.874*** (0.257)		
Δu_6 for enjoying work at jobs	$\overline{\Delta u}$	0.237 (0.322)	2.06	2.62***	0.269 (0.332)	2.15	2.80***
	σ	0.983*** (0.284)			0.998*** (0.291)		
γ_3 for the social status of jobs	$\overline{\Delta u}$	-0.718 (0.453)	2.22	8.78***	0.358 (0.414)	-	-
	σ	1.746*** (0.277)			2.494*** (0.442)		
$\widetilde{\gamma_3}$ social status× Foreign	$\overline{\Delta u}$	-	-	-	2.506*** (0.665)	-	-
	σ				1.191 (1.201)		
$\widetilde{\gamma_4}$ Income at 30 × Low_Inc	$\overline{\Delta u}$	-	-	-	1.47e-6 (1.98e-6)	-	-
	σ				6.4e-7 (3.39e-6)		
Log-Likelihood		-1346.99			-1334.79		
No. of Groups		161			161		
No. of Choice Situations		1290			1290		
No. of Observations		5809			5809		

a Mean and Standard Deviation of the Log-Normally distributed coefficients are calculated at the estimated $\overline{\Delta u}$ and σ.
* significant at 10%; ** significant at 5%; *** significant at 1%; standard errors in parentheses

Table A5: Best Linear Predictor of Expectation of Parents' Approval

Dependent Variable:Expectation of Parents' Approval†	Entire Sample		Males		Females	
	Estimates	Std. Error	Estimates	Std. Error	Estimates	Std. Error
	(1)	(2)	(3)	(4)	(5)	(6)
Expectation of:[a]						
status of jobs × (1- Parents_foreign[b])	0.084**	(0.035)	0.0611	(0.0622)	0.090**	(0.043)
status of the jobs × Parents_foreign	0.188***	(0.047)	0.125*	(0.091)	0.228***	(0.064)
graduating with a GPA of at least 3.5	−0.0466	(0.0467)	−0.003	(0.078)	−0.073	(0.056)
graduating in 4 years	0.0798	(0.067)	0.068	(0.096)	0.069	(0.092)
enjoying coursework	0.0013	(0.0013)	0.00046	(0.0019)	0.0016	(0.0018)
enjoying work at the jobs	0.114***	(0.041)	0.063	(0.0660)	0.145***	(0.053)
finding a job upon graduation	0.289***	(0.067)	0.279**	(0.122)	0.303***	(0.071)
finding a job × Parents_foreign	0.207**	(0.082)	0.219*	(0.124)	0.202*	(0.110)
income at 30 (in 10,000s)	0.000023	(0.00112)	0.0023	(0.0035)	−0.0006	(0.0009)
income at 30 (in 10,000s) × Low_Income[c]	0.0018	(0.0022)	−0.00082	(0.0048)	0.0028*	(0.0015)
Mother studied given major[d]	0.024	(0.018)	0.051	(0.031)	0.0055	(0.02)
Father studied given major[e]	0.032**	(0.015)	0.0364*	(0.022)	0.024	(0.022)
Studying Given Major[f]	0.0357***	(0.013)	0.021	(0.021)	0.048***	(0.016)
Respondent Fixed-Effects	Yes		Yes		Yes	
Major-Specific Dummies	Yes		Yes		Yes	
R-Squared	0.6943		0.6858		0.7113	
No. of Observations	1287		551		736	
No. of Clusters/ Individuals	161		69		92	

† Dependent variable is a response 0-1 to: "If you were majoring in [X], what do you think is the percent chance that your parents and other family members would approve of it?"

All regressions include major-specific dummies, and respondent fixed effects. (Constants not shown)

Parameter estimates correspond to the estimation of OLS model. Cluster errors in parentheses

* significant at 10%; ** significant at 5%; *** significant at 1%

a Expectations of outcomes except income are between 0 and 1; status is discrete on a scale of 0-0.9

b a dummy that equals one if either of the respondent's parents is foreign-born

c a dummy that equal one if respondent's parents' annual earnings are less than $150,000

d a dummy that equals one if mother's field of study is the same as the relevant question

e a dummy that equals one if father's field of study is the same as the relevant question

f a dummy that equals one if the respondent's intended major category is same as category X in the question