*Yigit Aydede**

Assortative preferences in choice of major

Abstract

The primary objective of this study is to examine the contribution of available information constrained by parents' fields of study to the observed assortative preferences in their children's choice of major. Comparable to panel models, we define within-family transmission functions with 1-to-2 matches (1 for each parent). Using the confidential major file of the 2011 National Household Survey from Canada, the results show that children's choice of field of study exhibits significant assortative preferences isolated from ability sorting and unobserved differences across majors and other family characteristics. With some caution, we attribute this persisting assortative tendency to the information asymmetry across alternative majors built on by parents' educational backgrounds within families.

Current version: July 16, 2020
Keywords: field-of-study homogamy, choice of field of study, occupational relatedness
JEL-codes: J1, I2, D1
Corresponding author: Yigit Aydede
yigit.aydede@smu.ca

Department of Economics, Saint Mary's University, Nova Scotia B3H 3C3, Canada.

© The Author(s). 2020 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
Cite as: Aydede. *IZA Journal of Labor Economics* (2020) 9:6.
https://doi.org/10.2478/izajole-2020-0006

1 Introduction

Evidence shows that expectations about earnings, employment opportunities, marriage options, job–family balance, enjoying course work, social status of available jobs, and own ability to successfully complete the study associated with each major are fundamental factors in the choice of field of study. The evidence also shows that there is a substantial error in beliefs (subjective expectations) about the population values of these determinants (Stinebrickner and Stinebrickner, 2014). When students are provided correct information, they update their beliefs and their choice of field of study (Wiswall and Zafar, 2015; Arcidiacono et al. 2012).

According to the 2013 National Graduate Survey, close to 60% of university graduates in Canada report that their parents' recommendations played a very important role in their choice of major. This is not surprising because the information and its value from different sources become more dispersed and questionable. Altonji et al. (2015), for instance, documented that Princeton pushes students to consider departments with fewer students. Some postsecondary institutions prefer a distribution of students across majors in such a way that it correlates with the distribution of faculty members in those majors. Departments in high (low) demand make their own field of study less (more) attractive when counseling students in their choice of major. As complex education choices are made under uncertainty about the achievement of choice-specific outcomes and personal preferences and abilities, parents become the least costly and most trustworthy channels of information, especially in Canada, where switching majors is not costless.[1,2] Yet, a significant assortativity (a child predictably becomes a teacher because it is his father's and/or mother's job) could also suggest systemic biases in decision-making, specially when the information about the achievement of the future major-specific outcomes is bounded by parents' fields of study.[3]

What then is the parents' role in the belief formation? To understand that information is not distributed symmetrically across majors with the same value and volume, imagine that both parents are accountants and working in the finance industry. The cost of obtaining the same level of information about other majors, say on biochemistry, is obvious. This brings us to the question of how the field-of-study homogamy (FSH) and whether the parents work in related occupations affect the magnitude of information asymmetry. The following 2 empirical questions need to be answered to assess the role of information asymmetry in the choice of major more formally: How can we quantify the resemblance of fields of study between parents and children beyond a binary proposition that reflects the assortative tendency, an association that exposes the attraction of each child to their own parents' majors? How can we identify the role of information asymmetry in this assortative tendency, after removing the other factors that are not observed by the researcher, such as implicit randomness, ability sorting, and

1 As expected, studies (e.g., Hoxby and Avery, 2013) show that less well-educated parents with no specialization would not be good transmitters of information.

2 Although the system is different from one where the major is chosen at entry into the university through a centralized test or using a threshold grade point average (GPA) required for each major, students in Canada are usually accepted to universities at three main faculty levels: Arts, Science, and Business/Commerce. Each of these requires different courses to be completed in high school with competitive GPAs at grades 11 and 12. Therefore, although the majors are decided after the second year, roughly after completing 14–16 core courses within each faculty, switching majors across faculties imposes a significant cost on students and parents.

3 In addition to information asymmetry, parents could also impose their preferences on their child's educational attainment by their willingness to use financial transfers to "distort" their child's choice toward (or against) a specific field of study. Zafar (2012) investigates this issue in his recent paper titled, "Double majors: one for me, one for the parents?"

individual tastes, embedded in the resemblance of majors between each parent and child? We will try to answer both questions in this paper.

This study's primary objective is to investigate the role of information asymmetry in children's attraction to their parents' field of study reflected by assortative tendencies in child–parent matches. We apply conventional intergenerational transmission functions that relate the children's assortative preferences to FSH and whether parents work in their trained jobs within Canadian families. We use the confidential major file of the 2011 National Household Survey (NHS) so that the size of the data and the availability of different levels of aggregation in the Classification of Instructional Programs (CIP) allow us to develop 3 indicators: the degree of children's attraction to their parents' field of study (field of study attraction or FSA), the degree of FSH, and the degree of relatedness between each parent's field of study and occupation (field-of-study relatedness or FOR). To identify the role of information asymmetry in assortative patterns, we define quasi-likelihood transmission functions, where the response variables take on fractional values of FSA between each child (son/daughter) and parent (father/mother) as a function of FSH and FOR. Similar to the difficulties in identifying the role of expected earnings in college major choice, the challenge here is also to control for selection into each major. To tackle this problem, we define within-family transmission functions based on an assortative matching model with 1-to-2 matches (1 for each parent), inspired by Diamond and Agarwal (2016). Comparable to panel models, this allows us to reduce unobserved heterogeneity so that the results provide new and more direct evidence about the intergenerational association of field of study due to information asymmetry reflected in assortative tendencies, which is, to the best of our knowledge, the first of its kind in the literature.

The first part of our results shows that children's choice of field of study exhibits significant assortative preferences. This finding is a new contribution that reports intergenerational skill transfers as opposed to educational mobility. We also find that the assortative tendency is the highest between fathers and sons relative to all other pairs, namely, father–daughter, mother–son, and mother–daughter. This evidence becomes even stronger when we use more disaggregated CIP codes and control for educational degrees. A significant skill sorting in mating is also revealed by the FSH measures, which also indicate gender differences in the attractiveness of each major in marriage. This finding is consistent with the evidence that the gain from the marriage could be different for each spouse (Choo and Siow, 2006) and with the evidence of a substantial degree of gender heterogeneity in the preferences for each major (Wiswall and Zafar, 2015). These findings, on significant intergenerational skill transfers and greater assortative mating for skills, are in line with the concerns about the possible progressive skill stratifications and earning inequalities in societies.

In the second part, the estimation results show that higher assortativity in each child–parent combination is strongly associated with greater homogamy and field-of-study relatedness in parents' jobs. The empirical approach that we apply here aims to identify the role of information boundaries in this relationship. Our findings indicate that asymmetric information is a significant contributor to children's assortative tendencies in their choice of major.

The remainder of the paper is organized as follows: Section 2 introduces the data, homogamy, assortative preferences, and occupational relatedness; Section 3 introduces the conceptual background that links the subjective expectations in choice of major to information

entropy; the empirical framework is explained in Section 4; the estimation results are reported in Section 5; and we provide the concluding remarks in Section 6.

2 Data, assortative preferences, homogamy, and occupation match

2.1 Data

This study uses the confidential major file of the 2011 NHS. We restricted the data to include only non-Aboriginal native-born individuals living in 10 provinces. We also dropped non-degree-holder parents (i.e., those with no education or an education degree that does not grant a major) and those whose field of study contains <10 workers. After these restrictions, we obtained about 2.3 million observations. The 2011 NHS enables the classification of individuals' major field of study in which the highest postsecondary certificate, diploma, or degree was granted. Statistics Canada classifies the major fields of study by using the CIP, which includes 1,688 instructional program classes.[4]

One major challenge in identifying children's choice of field of study in relation to their parents' educational background is the availability of data. There is no survey in Canada in which respondents are directly asked about their parents' field of study. Although parents' schooling years are more accessible, many studies on educational transmission face the same challenge. In a recent study, e.g., Chevalier et al. (2013) use a subsample from a pool of Labour Force Surveys in the UK, which include children aged 16–18 years and living at home, so the parental information can be matched to the child's record. In order to identify field-of-study resemblance between parents and children, we use the same approach and create a subsample that is composed of children living at home. Although this restriction reduces the total sample size, it becomes less severe for the comparable age groups between 16 and 25 years of age. For example, while there are 122,000 females with an identified field of study between the ages of 19 and 21 years in the whole sample, our subsample includes 26,000 children who live with their parents. Moreover, we use this subsample only for FSA calculations, while the indices for parental homogamy and occupational relatedness use the full sample. We are aware that using a subsample of observations raises a question of selectivity. To ensure that the final sample is representative of the population, we first compare the distribution of parents (fathers and mothers, separately) living with their children to that of the whole sample across CIP codes classified into 12 and 41 groups based on 5-year age classes. We applied the same comparison for children based on gender and age. The results seem to confirm that the distribution of children and parents across fields of study by age and gender in our restricted subsample mirrors the same distributions in the full sample.[5]

Although we are forced to study only those children who live with their parents, this issue has to be well thought out. Our sample's distributional representativeness of population by age

4 For more information on CIP classification, see www.statcan.gc.ca/concepts/classification-eng.htm. The most aggregated level classifies CIP codes into 12 major groups. This aggregation is reduced to 41 and 372 groups and is classified down to the most detailed level, where all majors are presented with 1,688 CIP codes.

5 Hilger (2016) has developed a new method to adjust the data to recover the outcomes of "missing" independent children. However, their educational outcome is measured in terms of years of schooling. We have also applied the inverse probability weights method to our subsample to address the possible selectivity problem. The results on FSA calculations do not change significantly.

and gender is just one aspect of the matter. It is possible, for instance, that those individuals who live at home are more likely to attend their closest higher education institution and thus their program choice set would be limited by a regional concentration in a particular industry, in which the parents work. Although we address this problem by controlling for unobserved regional heterogeneity, another issue would be that children who have a good relationship with their parents are more likely to stay at home and may therefore hold their parents as likely role models and follow in their footsteps. If this is the case, using our sample would lead to an upward bias in the parent–child assortativeness of field of study.[6] More descriptive information about the data and our samples are provided in the following sections, which explain FSH, FSA, and FOR.

2.2 **FSH measures**

Assortative mating has long been documented by demographers using nonparametric log–linear models based on contingency tables of ethnicity, education, religion, and other attributes (Schwartz, 2013). Following Becker's (1973, 1974) theory on marriage markets, economists have also investigated assortative mating in relation to match gains and returns to marriage. Chiappori and Salanie (2016), for instance, show that educational homogamy of posterity is likely to be reinforced by increases in the human capital of parents, who are matched homogamously themselves. Bicakova and Jurajda (2016) are the first to analyze mating by field of study for European countries.

Unlike joint or conditional probabilities that define the likelihood of a match, we choose the following identity that recognizes the randomness inherent in the matching process and specifies to what extent the match is driven by assortative mating on the field of study and to what extent it reflects the marginal distributions of each major:

$$\mathrm{FSH} = \Pr(\boldsymbol{M})\left[\Pr(\boldsymbol{F}|\boldsymbol{M}) - \Pr(\boldsymbol{F})\right], \tag{1}$$

where F and M are indicators of fields of study for the female and male mates, respectively, in matching couples. As recognized in the literature, the observed matches in a marriage market are jointly determined by the preferences of both partners. For example, Choo and Siow (2006) argue that the observed marriage patterns positively depend on the gross gains to marriage in which the individual returns could be different for each spouse, reflecting a spousal "appreciation" or "attraction" of each field of study in mating.

With the number of matches, m_{ij}, where i and j reflect the husband's and wife's major in each row and column in the resulting contingency table, the FSH matrix can be calculated by Equation (2):

$$\frac{m_{ij}}{T} - \frac{m_i m_j}{T^2}, \tag{2}$$

where m_i and m_j represent the row and the column totals, respectively, and T is the total number of pairs. While the FSH matrix reveals the assortativeness between, e.g., a male accountant and a female historian in mating, it would be quite possible that a male accountant's attraction to a

6 Gratefully, this point has been brought to our attention by one of our referees. Another point raised by the referee is that the financial crisis before the 2011 Census is likely to have an impact on choices of field of study. Issues like oversensitivity to uncertainty and a higher level of risk aversion may have strong effects on the choice set of majors for the cohort of students that we investigate in our study.

female historian would be different from her attraction to a male accountant. In order to reflect a differential "appreciation" of each field of study for each spouse in mating, we use a simple horizontal (vertical) normalization for each row (column) of the FSH matrix between 0 and 1. The results are reported in Table 1.

For example, the match between a male accountant (Business) and a female historian (Humanities) is ranked at 0.30 in terms of its assortativity, among all possible matches available for a male accountant with other different major holders. The same match is ranked at 0.53 among those available for a female historian reported in the bottom section of the table. These indices simply order each partner's appeal by his/her field of study and do not impose cardinal restrictions. It is obvious from the diagonal of both sections of the table that the evidence supports a strong FSH. Although not reported here, FSH becomes even stronger

Table 1 Field-of-study homogamy (FSH) by prime CIP codes (weighted)

		Wife's major										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Husband's major		**Normalized by husband's major**										
Education	1	**1.00**	0.25	0.27	0.19	0.00	0.27	0.26	0.25	0.26	0.18	0.19
Arts	2	0.18	**1.00**	0.42	0.42	0.00	0.29	0.31	0.29	0.30	0.01	0.28
Humanities	3	0.43	0.43	**1.00**	0.42	0.00	0.38	0.36	0.32	0.35	0.14	0.25
Law	4	0.37	0.36	0.47	**1.00**	0.00	0.34	0.32	0.28	0.31	0.01	0.17
Business	5	0.23	0.26	0.30	0.33	**1.00**	0.28	0.27	0.22	0.23	0.00	0.11
Science	6	0.47	0.44	0.58	0.43	0.00	**1.00**	0.85	0.42	0.47	0.34	0.26
Math/Comp.	7	0.25	0.51	0.51	0.61	0.39	0.50	**1.00**	0.43	0.42	0.00	0.30
Engineering	8	0.00	0.35	0.17	0.17	1.00	0.35	0.45	**0.66**	0.44	0.70	0.79
Agriculture	9	0.30	0.19	0.09	0.16	0.00	0.30	0.19	0.22	**1.00**	0.36	0.21
Health	10	0.28	0.29	0.30	0.26	0.00	0.34	0.29	0.29	0.31	**1.00**	0.25
Services	11	0.00	0.32	0.18	0.25	0.54	0.28	0.36	0.35	0.38	0.63	**1.00**
		Normalized by wife's major										
Education	1	**1.00**	0.29	0.50	0.29	0.00	0.37	0.01	0.07	0.14	0.18	0.10
Arts	2	0.37	**1.00**	0.57	0.47	0.23	0.41	0.15	0.18	0.23	0.20	0.26
Humanities	3	0.43	0.48	**1.00**	0.46	0.14	0.45	0.15	0.13	0.20	0.17	0.18
Law	4	0.42	0.46	0.67	**1.00**	0.09	0.47	0.12	0.11	0.20	0.04	0.08
Business	5	0.37	0.31	0.53	0.47	**0.97**	0.38	0.04	0.00	0.00	0.00	0.00
Science	6	0.41	0.39	0.59	0.42	0.20	**1.00**	0.87	0.20	0.30	0.28	0.19
Math/Comp.	7	0.37	0.44	0.55	0.48	0.32	0.48	**1.00**	0.20	0.23	0.17	0.22
Engineering	8	0.00	0.00	0.00	0.00	1.00	0.00	0.13	**1.00**	0.16	0.68	1.00
Agriculture	9	0.41	0.35	0.47	0.40	0.27	0.45	0.07	0.18	**1.00**	0.34	0.26
Health	10	0.38	0.31	0.49	0.36	0.05	0.46	0.00	0.12	0.17	**1.00**	0.17
Services	11	0.30	0.33	0.42	0.37	0.39	0.30	0.11	0.17	0.23	0.42	**0.68**

Notes: The sample used in this table contains all working spouses, regardless of whether they have children with or without an identified CIP code. Because of the very few observations, the table does not report "Others" classified under CIP code 12. The details of majors are as follows: (1) Education, (2) Visual and performing arts, and communication technologies, (3) Humanities, (4) Social and behavioral sciences and law, (5) Business, management and public administration, (6) Physical and life sciences and technologies, (7) Mathematics, computer and information sciences, (8) Architecture, engineering, and related technologies, (9) Agriculture, and natural sources and conservation, (10) Health and related fields, and (11) Personal, protective, and transportation services.

when we use 41 CIP codes matrix calculated based on 12 and 41 major CIP codes. The results indicate a slightly increasing FSH as we use more detailed CIP codes.[7]

2.3 **Quantifying assortative preferences: FSA**

The FSA index compares the field of study of each parent to that of each child in a family and calculates the degree of attraction between the 2 based on the probability distributions. We create 4 contingency tables using the restricted subsample explained earlier. Each table reports the number of field-of-study matches between sons and fathers, daughters and fathers, sons and mothers, and daughters and mothers. Similar to Equation (1), to identify the observed matching patterns, we choose the following identity that reflects the differences between observed and expected frequencies under independence:

$$\mathrm{FSA} = \Pr(\boldsymbol{P})\left[\Pr(\boldsymbol{K}|\boldsymbol{P}) - \Pr(\boldsymbol{K})\right] \tag{3}$$

where P and K are indicators of fields of study for parents and children in matching, respectively. When it is normalized for each parental field of study between 0 and 1, the resulting measures imply the attraction of children to their parents' majors evaluated by the observed distribution of all possible matches between parents and children. The number of different matching possibilities between the parent and the child comes from the fact that it is the child who faces many different alternatives before making a decision on a major.

The assortativity exposed by FSA reflects only the child's preferences as they are defined over children, not over parents in matches. While we use 12, 41, and 137 major groups of CIP, in the 4 match tables, we report only the sons' match calculated with 12 major CIP codes in Table 2. The higher values of FSA on the diagonal indicate that the most likely matches happen between the same fields of study. In each row, for any given major that the parent holds, the normalized FSA indicates the son's attraction to all other majors relative to the most likely match. The premise of this measure is that the child's attraction to each parent's major could be different even if the parents have the same field of study. Intuitively, the same major could be more(or less) attractive for the son, e.g., if it is held by his father, which may reflect not only the differences between maternal and paternal influence but also gender differences in occupational distributions. While dissimilarities in each cell between the upper and lower parts of the table may expose this fact, the presence of a strong assortativity indicates that parents' field of study is a fundamental factor in children's choice of field of study.

Although we refrain from using more space to interpret the results here, one may ask to compare the extent of field-of-study attraction between parents and children across 4 match tables. We use an index that computes the ratio of 2 diagonal shares of a match matrix as follows:

$$H = 100\left(\frac{\sum m_{ij}/T}{\sum m_i m_j/T^2} - 1\right) \tag{4}$$

which is the sum of the joint probabilities on the diagonal relative to the sum of the products of their marginal probabilities. Hence, it provides the ratio of the actual share of matches with

7 We use H-index (explained in the next section) to compare two FSH contingency tables. The index provides the ratio of the actual share of matches with the same field of study (on the diagonal) to the share of matches that one would expect under the random matching assumption.

Table 2 Field-of-study homogamy (FSH) by prime CIP codes (weighted)

		Son's major										
Father's major		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Education	1	**0.94**	0.80	1.00	0.81	0.80	0.74	0.74	0.00	0.70	0.88	0.63
Arts	2	0.32	**1.00**	0.37	0.35	0.31	0.36	0.32	0.00	0.21	0.35	0.38
Humanities	3	0.60	0.72	**1.00**	0.73	0.51	0.59	0.57	0.00	0.56	0.59	0.48
Law	4	0.65	0.72	0.87	**1.00**	0.77	0.68	0.66	0.00	0.60	0.68	0.53
Business	5	0.55	0.54	0.64	0.69	**1.00**	0.61	0.57	0.00	0.50	0.56	0.43
Science	6	0.64	0.63	0.67	0.73	0.60	**1.00**	0.69	0.00	0.51	0.63	0.35
Math/Comp.	7	0.64	0.76	0.88	0.63	0.73	0.98	**1.00**	0.00	0.52	0.64	0.54
Engineering	8	0.17	0.14	0.00	0.02	0.00	0.08	0.18	**1.00**	0.20	0.14	0.21
Agriculture	9	0.23	0.10	0.01	0.11	0.02	0.20	0.16	0.00	**1.00**	0.26	0.51
Health	10	0.69	0.66	0.76	0.75	0.63	1.00	0.57	0.00	0.57	**0.84**	0.59
Services	11	0.25	0.21	0.08	0.06	0.00	0.12	0.25	0.51	0.24	0.23	**1.00**
Mother's major												
Education	1	**0.98**	0.84	0.94	1.00	0.86	0.84	0.77	0.00	0.77	0.85	0.46
Arts	2	0.38	**1.00**	0.65	0.62	0.40	0.45	0.42	0.00	0.37	0.42	0.15
Humanities	3	0.62	0.71	**1.00**	0.78	0.70	0.74	0.67	0.00	0.57	0.58	0.46
Law	4	0.52	0.59	0.69	**1.00**	0.75	0.53	0.47	0.00	0.55	0.54	0.47
Business	5	0.21	0.14	0.06	0.00	**0.37**	0.17	0.29	1.00	0.25	0.16	0.38
Science	6	0.72	0.64	0.70	0.66	0.75	**1.00**	0.70	0.00	0.60	0.64	0.57
Math/Comp.	7	0.39	0.44	0.51	0.38	0.00	0.96	**1.00**	0.45	0.30	0.20	0.05
Engineering	8	0.14	0.19	0.10	0.09	0.00	0.27	0.23	**1.00**	0.17	0.33	0.25
Agriculture	9	0.39	0.43	0.27	0.40	0.00	0.42	0.32	0.69	**1.00**	0.34	0.42
Health	10	0.45	0.18	0.00	0.07	0.02	0.42	0.34	0.84	0.57	**1.00**	0.96
Services	11	0.21	0.27	0.06	0.00	0.02	0.21	0.26	1.00	0.28	0.22	**0.60**

Notes: See the notes to Table 1 for the full description of majors.

the same field of study (on the diagonal) to the share of matches that one would expect under the random matching assumption.[8] The indices calculated for the 41 major CIP codes are as follows: 119.56 for Father–Son, 31.28 for Mother–Son, 60.02 for Father–Daughter, and 48.88 for Mother–Daughter. These sharp differences, tested by 95% bootstrapped confidence intervals, indicate that a randomly picked father–son pair with the same field of study is about twice as likely than would be predicted under random matching. Moreover, a very low index for mother–son pairs suggests that the overall attraction of sons to their mother's major is slightly higher than what would be predicted if sons randomly pick their majors. Although these observations are very informative, they would not provide answers that explain the underlying reasons. In Section 3, we will attempt to confront this challenge.

2.4 **Field-of-study occupation relatedness - FOR**

The evidence shows that when people do not work in their trained jobs, the value of their field of study diminishes (Aydede and Dar, 2016; Robst, 2007). A recent study by Lemieux (2014) finds that this wage penalty varies by each field of study in the range of 16% for engineers

8 When the children's choice of major is not affected by their parents' field of study, each joint probability on the diagonal (nominator) approaches the product of its marginal probabilities; then the whole index becomes zero. Thus, any departure from zero indicates the tendency toward the same field-of-study matches.

and 5.7% for degree holders in the Humanities. The quality of parents' occupational match would also contribute to the formation of subjective expectations about the major-specific outcomes. An accountant working as a chef, for instance, would be a less-reliable channel of information on the prospects of an accounting major than one who works as a certified public accountant.

To measure FOR beyond a binary proposition, related or not, we use the following continuous index suggested by Aydede and Dar (2016):

$$\mathrm{FOR}_{\mathrm{of}} = \frac{L_{of} / L_f}{L_o / L_T} \tag{5}$$

where L is the number of workers, o is the occupation, f is the field of study, and T denotes the whole workforce. Given the large sample at our disposal, we use the frequency distribution of 41 fields of study across 40 occupations classified according to the National Occupational Classification (NOC–2011), which gives us 1,640 cells to calculate FOR. For each of the 41 fields of study, when we normalize FOR between 0 and 1 by using the highest FOR as numeraire, the resulting index, NFOR, reveals the ranking of each occupation for each major based on the native-born workers' distribution. To provide a descriptive summary for FOR, we classify the NFOR in 2 class intervals (1–0.8 and 0.8–0.0) and report the distribution of spouses across these classes and 11 major fields of study in Table 3. If, for any given field of study, we consider the occupations with NFOR between 1.0 and 0.8 as relatively better-matching occupations, we see that 32% of husbands work in related occupations, with the same ratio slightly lower for wives. As expected, the ratio varies across majors from 10% for wives in humanities to 57% for husbands in education.

Finally, to see the relationship between parents' education–job relatedness and children's attraction to their parents' field of study, we summarize the FSA for each child–parent pair by

Table 3 Distribution of fathers and mothers by NFOR and 12 prime CIP codes (% and weighted)

		Father			**Mother**		
		NFOR		Major's	**NFOR**		Major's
Majors		**1.0–0.8**	**0.8–0.0**	share	**1.0–0.8**	**0.8–0.0**	share
Education	1	57.32	42.68	7.41	56.80	43.20	8.74
Arts	2	28.50	71.50	3.44	28.33	71.67	3.55
Humanities	3	11.03	88.97	4.86	10.05	89.95	5.16
Law	4	23.89	76.11	9.62	23.51	76.49	11.15
Business	5	16.34	83.66	20.02	14.98	85.02	22.94
Science	6	26.53	73.47	3.33	26.28	73.72	3.21
Math/Comp.	7	32.11	67.89	3.61	29.01	70.99	3.27
Engineering	8	42.99	57.01	26.26	42.10	57.90	16.79
Agriculture	9	25.39	74.61	2.93	23.51	76.49	2.55
Health	10	36.15	63.85	11.91	33.22	66.78	16.08
Services	11	36.91	63.09	6.61	34.58	65.42	6.58
	Total	**32.15**	**67.85**		**29.62**	**70.38**	

Notes: (1) See the notes to Table 1 for the full description of majors. (2) The sample used in this table contains all working spouses, regardless of whether they have children with or without an identified CIP code.

the parents' occupational relatedness. In the first row, both father (F) and mother (M) work in occupations that are related to their majors. This relatedness is reflected with a binary variable, NFORC, which is 1 if NFOR is between 1 and 0.2; and 0, otherwise. Although this classification is arbitrary, it seems that, in all parent–child pairs, a higher FSA is associated with a greater NFOR. More interestingly, the highest average FSA in each column is observed when the matching parent works in a related job irrespective of the other parent's occupational relatedness.

For example, in the first column of Table 4, the average FSA is much higher (0.466 and 0.469) when the father's NFOR is 1 and not affected by the mother's field-of-study relatedness. This observation recurring in each column implies that the FSA calculated for each child–parent pair is strongly related to the matching parent's occupational relatedness but not to that of the other parent. If this positive relationship is statistically meaningful, which we investigate in the following sections, it also implies that FSA indices properly retrieve parental differences in assortativity.

3 Conceptual background

Although theoretical work incorporates the uncertainty in schooling decisions, earlier empirical studies assume that individuals are rational and use the achieved (observed) outcomes to infer decision rules. The recent literature shows that this is not a valid assumption and the difference between beliefs on choice-specific outcomes and their true population values is not trivial. A few recent studies (Wiswall and Zafar, 2015; Zafar, 2012, 2013; Arcidiacono et al., 2012; Stinebrickner and Stinebrickner, 2014) address this identification problem by directly eliciting subjective beliefs from a sample of university students. While the evidence in these studies reveals that subjective expectations on major-specific outcomes greatly vary across individuals, there is a lack of evidence as to why beliefs are so dispersed around the true population values.

In this study, we want to understand the role of parents' educational background in the process of expectation formation by looking at the assortative preferences that result from asymmetric information. The main driver of child i's attraction to major m revealed in his/her choice is the expected lifetime utility from the vector of future outcomes ($\mathbf{Z}$) of a specific human capital endowment with the subjective joint probability distribution, $G(Z|m,t)$, at time t, defined as follows:

$$E_i V_{i,m} = \sum_{t=1}^{T} \beta^{t-1} \int U(\mathrm{Z}) dG_i(\mathrm{Z}|m,t) \tag{6}$$

Table 4 Average FSA by NFOR based on 41 major CIP codes (weighted)

NFORC	Father–son	Mother–son	Father–daughter	Mother–daughter
F = 1, M = 1	0.466	0.534	0.525	0.445
F = 1, M = 2	0.469	0.515	0.527	0.425
F = 2, M = 1	0.418	0.535	0.502	0.443
F = 2, M = 2	0.417	0.517	0.492	0.428

Note: The sample used in this table contains all working spouses regardless of whether they have children with or without an identified CIP code.

This implies that the appeal of major m would be different from that of major k for individual i due to differences in beliefs reflected in the subjective joint probability distributions $G(Z|m,t)$ and $G(Z|k,t)$, even if the majors have identical distributions in terms of their observed outcomes, $F(Z|m,t) = F(Z|k,t)$. What makes the uncertainty on the same major different for each individual? Or what makes the uncertainty different for each major for the same individual?

The concept of information entropy in computer science, introduced by Shannon (1948) and used in economics by Sims (2003), argues that people have limited information-processing capacity, which alters the information for each individual and thus differentiates their behaviors. The "fundamental problem of communication" is for the "receiver" (user of the information) to be able to identify what data were generated by the "source", based on the signal it receives through the (potentially noisy) "channel". Sim's "noisy information model" provides a convenient framework in our context because the information flow is modeled by the discrepancy in probability distributions of the same event at the source and the receiver.

In our context, parents serve as communication channels, not as the source of data, in transmitting publicly available information on choice-specific outcomes to their child, the "receiver". The parents' capacity (the level of complexity in their communication and the amount of time for them to convey the data) will be determined and bounded by their own majors. To understand the differences in this capacity and related entropy, one can imagine a biochemist father obtaining, carrying, and sustaining the information on possible outcomes of choosing nuclear physics or accounting as opposed to a father who is a nuclear physicist or an accountant.

The channel capacity, which reflects the information entropy on major m defined by the Kullback–Leibler (D_{KL}) divergence (the difference between the subjective and objective probability distributions of the future outcomes, $\mathbf{Z}$), can be expressed as follows when the father's field of study (FOS) is set to m:

$$E_i\left(D_{KL}\,|\,m,\mathrm{FOS}^{F}=m,t\right)=\gamma+\alpha\mathrm{FSH}^{M}+\delta\mathrm{FOR}^{F}+\beta\mathrm{FOR}^{M}\mathrm{FSH}^{M}+\rho z, \tag{7}$$

where superscripts F and M denote father and mother, respectively. Equation (7) implies that when $\mathrm{FOS}^{F} = m$, the expected level of information received by the child on major m is equal to an index number, g, the father's level of information-processing capacity on his own major plus how compatible the mother's major is with the father's major (FSM^{M}), and the degree of relatedness between the parents' fields of study and their occupations (FOR). The key element in this expression is FSM^{M}, which reflects the degree of relatedness (normalized between 0 and 1) between the fields of study of the parents. Suppose that the mother's major is the least-related major to her husband's major ($\mathrm{FSM}^{M} = 0$). It implies that she is not a "high capacity" channel for the information on major m but becomes one on her own major. Hence, a higher degree of field-of-study resemblance between parents makes them more efficient channels (less noisy) for more reliable information on major m by decreasing information entropy. However, a greater homogamy also means that parents become less efficient channels for other majors, with rising relative entropy. Therefore, the level of FSH defines the level of information asymmetry in a family.[9]

9 It could be argued that parents are not the only channels in accessing the information on majors. We assume that the information obtained from all other channels (child's peers, councillors in his/her school, his/her close relatives, and the parents of his/her best friends) that a child would receive would be filtered through parents. This assumption is in line with the evidence that parental approval is the most important factor in the choice of major (Zafar, 2012). However, this assumption is not required in our empirical setting, as will be evident later.

This example becomes less intuitive when we compare 2 cases where the father is an accountant in both cases but the mother is a biochemist in the first case and a historian in the second. How different would the parents be in terms of channeling reliable information on accounting? Although the values of FSM^M would be different, it appears that these 2 cases should be the same in terms of available information on accounting, especially relative to the case where both parents are accountants. However, one has to think that the information entropy on a major in a family will be determined by not only the fact that it is the major of one of the parents, but also how much the major (accounting) is appreciated, shared, understood, and discussed within the family, which is collectively reflected in FSM^M.[10] The details of the conceptual framework summarized here can be found in the Appendix.

4 Empirical framework

The key challenge in understanding the potential contribution of information asymmetry to the observed assortative patterns is to control for other characteristics that are not observed by the researcher but aggregated in FSA. To address this issue, we use a conventional intergenerational transmission framework, wherein we define quasi-likelihood functions with the response variables that take on fractional values of FSA between each child (son/daughter) and parent (father/mother) as a function of the spousal "appreciation" of each partner's major and field-of-study relatedness. Intergenerational transmission refers to a process that outlines the transfer of individual characteristics, including abilities, preferences, and outcomes, from parents to their children, which we choose as our empirical framework. For example, an intergenerational model of schooling estimated in the literature (Becker and Tomes, 1979; Solon, 2013; Black and Devereux, 2010; Becker et al., 2015) can be expressed as follows:

$$S^c = \alpha_0 + \alpha_1 S^P + \alpha_2 h^P + \alpha_2 f^P + e^c \tag{8}$$

This reduced-form equation explains the child's schooling (S^c) as a function of the parent's schooling (S^P), heritable attributes that parents may genetically pass on to children (h^P), parenting skills and preferences (f^P), and child-specific characteristics (e^c) independent from S^P, h^P, and f^P. Coefficient α_1 reflects the causal effect of the parent's schooling on the child's schooling joined with, among others, the income effect that more education would be associated with better parental education. It can be shown that, if Equation (8) reflects the true model, a direct estimation of Equation (8) with unobserved h^P and f^P cannot identify α_1, unless one assumes that endowments, h^P and f^P, are unrelated to S^P.[11] Hence, an estimation of Equation (8) without controlling for ability sorting and better parenting reveals the intergenerational elasticity between parent–child years of schooling, a summary measure of correlational associations between children's outcome and parental educational background. Although it cannot answer whether more educated parents have more educated children because of their education, the

10 It is true that more and better information on a major would not necessarily make it more attractive.

11 Holmlund et al. (2011) investigate the findings of a large number of studies to answer the following question: do more educated parents have more educated children because of their education? They show that the evidence is inconsistent across the other strategies (twins, adoptions, and Instrumental Variables models) and they could also encounter problems in obtaining bias-free estimates of causal intergenerational coefficients.

intergenerational elasticity of schooling is a fundamental metric that has been used to measure the mobility across generations.[12]

Inspired from this literature, we propose a different identification strategy and start with 4 reduced-form matching functions that use the child's assortative preferences aggregated in FSA as an outcome of transmission, a process that is built on available information based on the parents' educational background.

$$\mathrm{FSA}_{\mathrm{F,S}} = \alpha_0 + \alpha_1 \mathrm{NFSH}^{\mathrm{M}} + \alpha_2 \mathrm{NFOR}^{\mathrm{F}} + \alpha_3 h^{\mathrm{M}} + \alpha_4 f^{\mathrm{M}} + \alpha_5 h^{\mathrm{F}} + \alpha_6 f^{\mathrm{F}} + e^{\mathrm{S}}, \tag{9}$$

$$\mathrm{FSA}_{\mathrm{M,S}} = \beta_0 + \beta_1 \mathrm{NFSH}^{\mathrm{F}} + \beta_2 \mathrm{NFOR}^{\mathrm{M}} + \beta_3 h^{\mathrm{M}} + \beta_4 f^{\mathrm{M}} + \beta_5 h^{\mathrm{F}} + \beta_6 f^{\mathrm{F}} + \mu^{\mathrm{S}}, \tag{10}$$

$$\mathrm{FSA}_{\mathrm{F,D}} = \delta_0 + \delta_1 \mathrm{NFSH}^{\mathrm{M}} + \delta_2 \mathrm{NFOR}^{\mathrm{F}} + \delta_3 h^{\mathrm{M}} + \delta_4 f^{\mathrm{M}} + \delta_5 h^{\mathrm{F}} + \delta_6 f^{\mathrm{F}} + \varepsilon^{\mathrm{D}}, \tag{11}$$

$$\mathrm{FSA}_{\mathrm{M,S}} = \beta_0 + \beta_1 \mathrm{NFSH}^{\mathrm{F}} + \beta_2 \mathrm{NFOR}^{\mathrm{M}} + \beta_3 h^{\mathrm{M}} + \beta_4 f^{\mathrm{M}} + \beta_5 h^{\mathrm{F}} + \beta_6 f^{\mathrm{F}} + \mu^{\mathrm{S}}, \tag{12}$$

where scripts M, F, S, and D denote mother, father, son, and daughter, respectively. With the normalized FSH (NFSH) and FOR (NFOR), these equations reflect the idea that a child's assortative tendencies observed in his/her choice of major is related to the FSH and the degree of relatedness between each parent's field of study and occupation within a family.[13]

As long as a higher homogamy (and occupational match) suggests a greater limitation in available information on alternative majors, the coefficients of NFSH (NFOR) capture the underlying field-of-study transmission that relates the children's assortative preferences to the level of information asymmetry. The variable $\mathrm{NFSH}^{\mathrm{M}}$ in Equation (9), for instance, is bounded between 0 and 1. It reflects a perfect homogamy as it approaches 1. Intuitively, the α_1 coefficient reveals how much the son's preference for his father's major will be affected by the extent to which his mother's field of study becomes comparable. This reminds us of the earlier example: how much the son's aspiration for his father's major, accounting, will be affected if his mother was a biochemist instead of an accountant. Similarly, a positive and significant coefficient of NFOR validates the transmission as the parents would be more reliable transmitters of information when they work in their trained jobs. Hence, the presence of intergenerational transmission requires that the coefficients of NFSH and NFOR in those 4 equations should be positive, with dissimilarities reflecting the difference between maternal and paternal influences.

Yet, the identification of transmission due to information asymmetry across alternative majors requires controlling for ability sorting and unobserved heterogeneity. Defining each child's FSA separately for each parent provides an opportunity to create a setting similar to panel models. Since we observe 2 matches for each child, when we take the difference between them, the dependent variables in these matching functions better reflect the assortative tendency because the omitted heterogeneity across children are differenced out from the equations as shown below.

$$\mathrm{FSA}_{\mathrm{M,S}} - \mathrm{FSA}_{\mathrm{F,S}} = \omega_0 + \omega_1 \mathrm{NFSH}^{\mathrm{F}} - \omega_2 \mathrm{NFSH}^{\mathrm{M}} + \omega_3 \mathrm{NFOR}^{\mathrm{M}} - \omega_4 \mathrm{NFOR}^{\mathrm{F}} + \tau^{\mathrm{c}}, \tag{13}$$

12 There are several studies examining intergenerational education and income mobility (elasticity) in Canada: Turcotte (2011), Aydemir et al. (2013), McIntosh (2010), Corak (2001, 2017).

13 Given the parent's major, the FSA reflects the child's decision on a major that maximizes his/her expected utility. The theoretical foundation of this decision-making process is well-defined in the literature (Altonji et al. 2015). For now, we omit other child, parent, and family-specific attributes in Equations (9)–(12).

$$\mathrm{FSA}_{\mathrm{M,D}} - \mathrm{FSA}_{\mathrm{F,D}} = \sigma_0 + \sigma_1 \mathrm{NFSH}^{\mathrm{F}} - \sigma_2 \mathrm{NFSH}^{\mathrm{M}} + \sigma_3 \mathrm{NFOR}^{\mathrm{M}} - \sigma_4 \mathrm{NFOR}^{\mathrm{F}} + \upsilon^{\mathrm{c}}, \tag{14}$$

A similar identification method is also recognized and applied by Diamond and Agarwal (2016) by using the repeated measurements made available when each agent on one side of the market is matched to at least 2 agents on the other side. The intuition is that the same value of the unobservable characteristic of an agent determines multiple matches of that agent and can be differenced out in a measurement error model (Hu and Schennach, 2008).[14] Unlike in other matching markets, this is particularly effective in our case because the assortativity revealed by FSA reflects only the child's preferences defined over children, not over parents in matches.

These equations with within-parents differencing suggest that the difference $\mathrm{FSA}_{\mathrm{M,S}}$ and $\mathrm{FSA}_{\mathrm{F,S}}$, for instance, should be smaller when $\mathrm{NFSH}^{\mathrm{M}}$ decreases, holding other covariates constant. Intuitively, if the mother married to an accountant holds a degree in biochemistry, $\mathrm{NFSH}^{\mathrm{M}}$ approaches its lower limit.[15] As the mother becomes another channel of information on an alternative major, viz., biochemistry, the family information boundaries expand. Unlike the case when the mother was an accountant, this increase in the level of available information in turn reduces the son's bias toward his father's major, i.e., accounting. Therefore, $\mathrm{FSA}_{\mathrm{F,S}}$ (the son's attraction to his father's major) should be smaller when $\mathrm{NFSH}^{\mathrm{M}}$ (resemblance of the mother's major to her husband's, measured by spousal differences in the appeal of their majors) becomes lower. Hence, the differences in ω_1 and ω_2, as well as σ_1 and σ_1, will provide information about the difference in transmission between fathers and mothers. However, the value (and the volume) of the available information provided by the homogamy measures in the family depends on whether the parents work in related occupations. This could be better understood if we change the accountant–biochemist example to one where the father works as a chartered accountant while the biochemist mother works as a branch manager in a bank, which diminishes the value of information on biochemistry from the mother. Since the parents would be a better channel of information conditional on the quality of their occupational match, an increasing $\mathrm{NFOR}^{\mathrm{M}}$ in Equation (13) should have both a negative impact on $\mathrm{FSA}_{\mathrm{F,S}}$ and a positive effect on $\mathrm{FSA}_{\mathrm{M,S}}$. Hence, a positive and significant coefficient of $\mathrm{NFOR}^{\mathrm{M}}$ indicates the existence of a transmission of field of study reinforced by expanding the reliable information within the family.

As outlined earlier, in addition to the level of information asymmetry built on the parents' fields of study, children's assortative preferences could also reflect ability sorting. The suggested within-family specifications can address this identification problem conditional on the assumption that the effects of unobserved parental traits in Equations (9) and (10), as well as in Equations (11) and (12), are statistically similar. Without this assumption and excluding NFOR for now, Equation (13) can be expressed as follows:

$$\Delta FSA_{\mathrm{S}} = \omega_0 + \omega_1 NFSH^{\mathrm{F}} - \omega_2 NFSH^{\mathrm{M}} + \omega_3 h^{\mathrm{M}} + \omega_4 f^{\mathrm{M}} + \omega_5 h^{\mathrm{F}} + \omega_6 f^{\mathrm{F}} + \tau^{\mathrm{c}}, \tag{15}$$

where $\omega_3 = (\beta_3 - \alpha_3)$, $\omega_4 = (\beta_4 - \alpha_4)$, $\omega_5 = (\beta_5 - \alpha_5)$, and $\omega_6 = (\beta_6 - \alpha_6)$. When estimated by ordinary least squares (OLS), identification of ω_2 (ω_1) requires either that $\mathrm{NFSH}^{\mathrm{M}}$ ($\mathrm{NFSH}^{\mathrm{F}}$) is independent of unobserved parental traits or that ω_3, ω_4, ω_5, and ω_6 are 0, as shown below.

14 Models based on many-to-one matches are not new and are well-discussed in the literature (Roth and Sotomayor, 1992). The consequence of possible measurement errors in the dependent variable in our case may not result in attenuation bias but may inflate the standard errors of the estimates.

15 This statement is justified based on a strong field-of-study homogamy reported in Section 2.2.

$$\operatorname{plim}\widehat{\omega_{2ols}} = \omega_2 + \omega_3 \frac{\operatorname{cov}\left(\mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{M}}\right)}{\operatorname{var}\left(\mathrm{NFSH}^{\mathrm{M}}\right)} + \omega_4 \frac{\operatorname{cov}\left(\mathrm{NFSH}^{\mathrm{M}}, f^{\mathrm{M}}\right)}{\operatorname{var}\left(\mathrm{NFSH}^{\mathrm{M}}\right)} +$$

$$\omega_5 \frac{\operatorname{cov}\left(\mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{F}}\right)}{\operatorname{var}\left(\mathrm{NFSH}^{\mathrm{M}}\right)} + \omega_6 \frac{\operatorname{cov}\left(\mathrm{NFSH}^{\mathrm{M}}, f^{\mathrm{F}}\right)}{\operatorname{var}\left(\mathrm{NFSH}^{\mathrm{M}}\right)}. \tag{16}$$

First, we think that parents' child-rearing skills, f^{M} and f^{F}, should not be significantly correlated with the homogamy measures $\mathrm{NFSH}^{\mathrm{M}}$ and $\mathrm{NFSH}^{\mathrm{F}}$. It would be hard to find a systemic reason why individuals who choose their spouses in the same field of study would also be the future parents with more skills in rearing their children. Thus, a possible bias in the estimate of ω_2 should mostly originate from heritable traits, i.e., h^{M} and h^{F}, and their correlation with homogamy measures. To the extent that a field of study reveals the person's overall ability endowments, it would be reasonable to question the role of ability sorting in field-of-study matches. But, it is ambiguous how this possibility translates into nonzero $\operatorname{cov}(\mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{F}})$ and $\operatorname{cov}(\mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{M}})$.

If we assume that h represents heritable mathematical skills, for instance, a higher $\mathrm{NFSH}^{\mathrm{M}}$ could be related to a higher and a lower h^{F} (or h^{M}) at the same time.[16] To test this ambiguity, we can use matches where both spouses have at least a university degree in one of the following majors: science, technology, engineering, and math (STEM). Hence, what we observe by a higher or lower NFSH among STEM majors should be the differences in assortative preferences isolated from ability sorting. In other words, if NFSH is relatively higher for electrical/computer engineers, it means that they mostly choose their partners in similar fields instead of in theoretical statistics or chemical engineering, which are otherwise comparable in terms of ability requirements. The size of the data enables us to reduce the effect of $\operatorname{cov}(\mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{F}})$ and $\operatorname{cov}(\mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{M}})$ on the bias by estimating specifications (13) and (14) only for families that have similar ability endowments. Hence, as shown below, introducing a binary variable – STEM, which is 1 if both parents hold at least a university degree in one of the STEM majors, and 0 otherwise – into Equation (15), would help us address a possible bias in the transmission coefficients.

$$\Delta\mathrm{FSA}_{\mathrm{S}} = \omega_0 + \omega_1 \mathrm{NFSH}^{\mathrm{F}} - \omega_2 \mathrm{NFSH}^{\mathrm{M}} + \omega_3 \mathrm{STEM} + \omega_4 (\mathrm{STEM} \times \mathrm{NFSH}^{\mathrm{F}}) -$$

$$\omega_5 (\mathrm{STEM} \times \mathrm{NFSH}^{\mathrm{M}}) + \omega_6 \mathrm{NFOR}^{\mathrm{M}} - \omega_7 \mathrm{NFOR}^{\mathrm{F}} +$$

$$\omega_8 f^{\mathrm{F}} + \omega_9 h^{\mathrm{M}} + \omega_{10} f^{\mathrm{M}} + \omega_{11} h^{\mathrm{F}} + \tau^{\mathrm{c}}. \tag{17}$$

The coefficients of interaction terms will reveal the differences in the sons' assortative preferences in STEM families. With the within-family specification, 2 factors will shrink the bias on these coefficients: first, the differential effects of unobservables, $\omega_8 = (\beta_3 - \alpha_3)$ and $\omega_{10} = (\beta_5 - \alpha_5)$ in Equation (17), as opposed to their levels in specifications (9)–(12), will diminish their size; and second, $\operatorname{cov}(\mathrm{STEM} \times \mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{F}})$ and $\operatorname{cov}(\mathrm{STEM} \times \mathrm{NFSH}^{\mathrm{M}}, h^{\mathrm{M}})$ will be close to 0 for a subsample as specified by Equation (17). The definition of the bias in the estimate of ω_5, for instance, can be expressed as follows:

16 While we could observe a high NFSH for engineers and historians, they would have different mathematical skill endowments.

$$\operatorname{plim}\widehat{\omega_{5ols}} = \omega_5 + (\beta_3 - \alpha_3)\frac{\operatorname{cov}(\text{STEM}\times\text{NFSH}^{\text{M}}, h^{\text{M}})}{\operatorname{var}(\text{STEM}\times\text{NFSH}^{\text{M}})} +$$

$$(\beta_5 - \alpha_5)\frac{\operatorname{cov}(\text{STEM}\times\text{NFSH}^{\text{M}}, h^{\text{F}})}{\operatorname{var}(\text{STEM}\times\text{NFSH}^{\text{M}})}. \tag{18}$$

Hence, the size and the significance of the coefficients ω_5 and ω_6 will reveal whether cov(NFSH$^{\text{M}}$,h^{F}) and cov(NFSH$^{\text{M}}$,h^{M}) can reasonably be assumed to be 0. The next section will provide the results.

5 Estimation results

5.1 Without within-family differencing

We start with the 4 equations from Equation (9) to Equation (12). To reduce the unobserved heterogeneity across families, we expand the equations by controlling for household income, provincial fixed effects, first spoken official language, household size, and whether the family resides in an urban or rural area. We also control for homogamy in terms of parents' highest educational degree.[17] After these additions, Table 5 reports 2 sets of estimation results for the selected variables.[18] The first 4 columns report the estimation results, which include NFSH for each parent without accounting for parents' occupational relatedness. We control for FSH in the last 4 columns as a binary variable –1 if both parents have the same field of study, and 0 otherwise – and add FOR, for both father and other, as a categorical variable, FORC, which is equal to 1 if the normalized FOR is <0.2 and 0 otherwise. The first 4 specifications use larger subsamples because they exclude FOR, which can be identified only if the person's occupation is known.

The results reported in Table 5 are informative as they reflect the maternal and paternal differences in children's assortative preferences in choosing majors. The robust and positive NFSH coefficients provide evidence for the existence of what we call intergenerational transmission of field of study. As outlined before, the results reflect the combination of ability sorting, differences in parenting skills, and unobserved heterogeneity in individual and family characteristics, in addition to the limited information accessibility constrained by the parents' fields of study. The first 2 columns show that the son's attraction to his parents' majors is strongly related to the FSH, measured by spousal "appreciation" of each parent's major. A comparison of the coefficients (0.10 and 0.05) indicates that the paternal influence is more dominant in educational transmission for sons. A similar gap is not observed for daughters reported in the third and fourth columns. The robust NFSH coefficients still suggest that daughters will also be attracted to their parents' field of study, yet mothers have more influence on daughters.

17 Education-degree homogamy (EDH) is calculated similar to FSH by using Equation (1). A total of 11 major granting educational degrees are identified in the 2011 NHS: trades, registered apprenticeship, college - <1 year, college - 1–2 years, college - >2 years, university - below bachelor's, bachelor's, above bachelor's - less than master's, medicine-dentistry-veterinary, master's, and PhD.

18 Since our specifications have fractional response variables that have values ranging between zero and one, their linearity in this range becomes a question. To address this issue, we have also estimated all specifications in this section with quasi-likelihood methods where the response variables are transformed to log odds using the binomial distribution (Papke and Wooldridge, 1996). Since the results are almost the same, we report here only the linear specifications estimated by OLS.

Table 5 Intergenerational transmission of field of study with 41 major CIP codes

	1				2			
	FSA–Son		**FSA–Daughter**		**FSA–Son**		**FSA–Son**	
	Father	**Mother**	**Father**	**Mother**	**Father**	**Mother**	**Father**	**Mother**
FSH = 1 (if same major)					0.055	0.002	0.025	0.015
					0.001	0.008	0.001	0.062
$NFSH^M$	0.050		0.051					
	0.006		0.005					
$NFSH^F$		0.100		0.038				
		0.003		0.002				
$FORC^F = 1$					−0.044		−0.027	
					0.002		0.001	
$FORC^M = 1$						−0.020		−0.016
						0.001		0.003
EDH	0.046	0.027	0.018	0.025	0.040	0.045	0.025	0.026
	0.003	0.002	0.01	0.007	0.001	0.003	0.001	0.002
Household income	0.009	0.002	0.006	0.001	0.007	−0.002	0.005	0.001
	0.001	0.009	0.001	0.002	0.001	0.009	0.001	0.004
Constant	0.347	0.529	0.464	0.434	0.390	0.615	0.464	0.434
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of Obs.	23,748	23,748	22,254	22,254	21,101	20,016	20,195	19,159

Notes: (1) Dependent variables are indicated in each column's heading. (2) Standard errors reported under the coefficients are adjusted by using the two-way clustering method (Cameron et al., 2011) at the individual and household levels. (3) EDH reflects education-degree homogamy and is a continuous variable normalized between 0 and 1. HH Income is the annual disposable income for the household. Other variables that are not reported in the table control for household size, first spoken official language, whether the family is in rural area, and provincial fixed effects. (4) We also ran the regressions with and without the parental age variables. The results are insensitive to the inclusion of parental age variables. (5) When we control for field-of-study fixed effects, the results do not change significantly.

In the last 4 specifications, we distinguish the parents who have the same field of study and control for their occupational match. The results are consistent with those of the first 4 specifications. The effect of having homogamous parents on the son's attraction to his father's major (0.055) is much higher than his attraction to his mother's field of study (0.002). Again, the same significant but smaller difference can be observed for daughters. The second channel to identify the transmission is the relatedness of parents' field of study to their occupation, which is controlled by FORC in the last 4 estimations. The results confirm a strong and positive relationship between the parents' occupational match and the children's attraction to their parents' majors. The parental difference in this effect is also noticeable and in line with the earlier findings with FSH: the paternal effect is greater than the maternal influence for sons, while the same difference is less magnified for daughters. When it comes to other factors, a higher homogamy in terms of educational degree (EDH) is positively and significantly associated with FSA. Similarly, a higher household income has a positive effect on FSA. Among the other variables not reported in Table 5, only the urban–rural distinction in households' location is significant. Children from families in larger cities experience higher FSA.[19]

19 To test the robustness of the results in Table 5, we also used different levels of the CIP and occupation classifications available in the 2011 NHS. The results are not sensitive to using larger or smaller dimensions of match tables.

5.2 **Within-family differencing**

We address the identification with a within-family differencing as described in the previous section and report the results in Table 6.

The first 2 columns show the estimation results of Equation (13) with the same dependent variable, the difference in the son's attraction to his parents' majors. The first column reports the estimation results of the restricted version of Equation (13). The estimation results for daughters based on Equation (14) are reported in the last columns. The restricted specifications in the first and the third columns use a new binary variable, NFORC, which reflects the difference in NFOR in 3 categories; the base category refers to the case that both parents have the same field-of-study relatedness. Either both work in related jobs (e.g., NFOR is between 1.0 and 0.2 for both parents) or in unrelated jobs (e.g., NFOR is between 0.2 and 0.0 for both parents). The second category indicates that while the father works in a matching occupation, the mother does not. The third category specifies the opposite situation. Hence, the effect of parental differences in field-of-study relatedness can be captured by the last 2 categories.[20,21]

Table 6 Transmission of field of study by within-family specification with 41 major CIP codes

	$(FSA_{M,S}) - (FSA_{F,S})$		$(FSA_{M,D}) - (FSA_{F,D})$	
	1	**2**	**1**	**2**
$(NFSH^F) - (NFSH^M)$	0.153		−0.014	
	0.002		0.012	
DFORC				
1	Base		Base	
2	−0.033		−0.027	
	0.001		0.001	
3	0.035		0.163	
	0.001		0.068	
$NFSH^M$		−0.101		−0.005
		0.001		0.023
$NFSH^F$		0.162		−0.021
		0.001		0.018
$FORC^M = 1$		−0.016		−0.016
		0.001		0.071
$FORC^F = 1$		0.051		0.029
		0.001		0.001
Constant	0.069	0.026	−0.076	−0.070
	0.001	0.002	0.001	0.001
Number of Obs.	21,018	21,018	19,942	19,942

Notes: (1) Dependent variables are indicated in each column's heading. (2) Standard errors reported under the coefficients are adjusted by using the two-way clustering method (Cameron et al., 2011) at the individual and household levels.

20 We define the base category with two opposite cases, either both parents work in related jobs or unrelated jobs, because we want to estimate the effect of field-of-study relatedness for each parent. Given that the dependent variable is the difference in child's attraction to each parent's major, this effect can only be captured when parents' FOR is different.

21 The idea here is to identify the effect of FOR on the children's attraction to their parents' major, when the parents work in an unrelated occupation. Therefore, we actually tried to find the lowest cutoff point that realistically classifies the person's job completely unrelated to his/her training. We also applied higher thresholds up to 0.4. The results are still robust. This is mostly because the distribution of FOR is convex. Hence, increasing the threshold from 0.2 to 0.4 had a minor effect because relatively few people exist in the bin of 0.2–0.4.

The results are interesting and in line with the findings in our earlier estimations: the coefficient of $NFSH^F$ in the first column, 0.153, confirms that ($\mathbf{FSA}_{M,S}$) – ($\mathbf{FSA}_{F,S}$) is greater when the gap between the appeal of each spouse's major, ($NFSH^F$) – ($NFSH^M$), increases. Although this signifies the presence of intergenerational transmission, it does not offer an insight about the parental difference. This is because the gap could rise when $NFSH^F$ goes up, $NFSH^M$ goes down, or both occur simultaneously. The second column, based on Equation (15), helps us understand the difference. The sign of the coefficients on $NFSH^F$ and $NFSH^M$ are as expected. Since an increase in $NFSH^M$ has a positive impact on $\mathbf{FSA}_{F,S}$, it reduces the distance between $\mathbf{FSA}_{M,S}$ and $\mathbf{FSA}_{F,S}$. When the mother's major becomes similar to the father's, $NFSH^M$ rises. Because higher homogamy implies more constraint in the family in terms of available information on other majors, the son's bias toward his father's field of study rises. This is confirmed by the negative sign of the $NFSH^M$ coefficient. Equally, when $NFSH^F$ rises, the similarity between parents' major becomes higher. Constrained by less information being available regarding other majors, the son's attraction toward his mother's field of study increases. This is verified by the positive sign of the $NFSH^F$ coefficient: because a rise in $NFSH^F$ increases $\mathbf{FSA}_{MS}$, the distance between $\mathbf{FSA}_{MS}$ and $\mathbf{FSA}_{FS}$ becomes larger. The difference between these effects (0.101 and 0.162) again suggests that paternal influence is noticeably greater than maternal influence for sons. The same comparison for daughters in both specifications of Equation (14) would not offer the same evidence, which is also consistent with the relatively weaker effects for daughters reported in Table 5.

The existence of intergenerational transmission is also verified by the effect of the parents' field-of-study relatedness. In the first column, when evaluated against the base, the first category (fathers work in their trained job but mothers do not) has a negative effect on ($\mathbf{FSA}_{MS}$) – ($\mathbf{FSA}_{FS}$). Similarly, a significant positive effect is observed for the second category, wherein the mother works in her trained job but the father does not. These results are also confirmed with the unrestricted specification reported in the second column. Now, using FORC, if the mother's major is not a good fit for her occupation, the negative coefficient (–0.016) indicates that $\mathbf{FSA}_{MS}$ falls. Yet, when the father faces an educational mismatch in his job, the effect on ($\mathbf{FSA}_{MS}$) – ($\mathbf{FSA}_{FS}$) captured by a positive coefficient (0.051) becomes much greater. Interestingly, despite the insignificant effects of NFSH in the third and fourth columns, significant effects of FORC are observed for daughters, indicating the importance of parents' occupational matching in transmission.

5.3 **Within-family differencing among families with STEM majors**

With within-family differencing as specified by Equations (13) and (14), the other factors, such as the effects of siblings, neighborhoods, and peers, either observed or unobserved, are differenced out in the estimations. Hence, the results deliver better evidence about the role of information constraint in children's assortative preferences. However, as outlined earlier, the success of this identification strategy is conditional on the extent to which the FSH is driven by the ability sorting in parents' marriage.

One way to address this problem is to use a subsample that includes only those families in which both parents hold at least a university degree in one of the STEM majors so that the

difference in terms of their ability endowments would not be significant. Table 7 reports the estimations of the same specifications shown in the second and the last columns of Table 6 with STEM variables as expressed by Equation (17).

When one of the parents holds a degree in one of the non-STEM majors (or less than a bachelor's degree), the coefficients of $NFSH^M$ and $NFSH^F$ (–0.079 and 0.161) are almost identical to those reported in Table 6 for sons. This could be plausible given that the share of families where both parents have a STEM major with at least a university degree is <20% in the whole sample. The insignificant interaction terms indicate that ability sorting may not play a strong role in sons' assortative preferences. Hence, when the comparison is made only among STEM parents, the difference between paternal and maternal effects observed in field-of-study transmissions tends to remain similar to those found in our earlier results. The significant effect of STEM implies that the difference between $\mathbf{FSA_{MS}}$ and $\mathbf{FSA_{FS}}$ is lower for STEM families than for non-STEM families. None of the results are significant for daughters, except for FORC, which is in line with our earlier findings. There is a large literature on gender differences in occupational preferences and major choices. However, we do not have a satisfactory explanation why daughters' assortative preferences show no evidence of the link between parents' homogamy and their assortative preferences in their choice of major.

5.4 **Limitations**

The total elasticity of the assortative preferences in terms of parental homogamy can be expressed for sons by the sum of the coefficients $NFSH^M$ and $NFSH^F$, which is 0.24 (–0.079 and 0.161) from Table 7. This measure suggests an important role of information asymmetry in children's choice of major to the extent that the FSH reflects the level of constraint on the available information when children choose a major. It should be noted that the results reported here are conditional on a couple of assumptions. Although our sample, children living with their parents, is representative of the whole sample, there would still be a selection problem whereby children living with their parents may have different behavioral predispositions that affect their assortative preferences.

Table 7 Within-family specification for STEM parents with 41 major CIP codes

	$(FSA_{MS}) - (FSA_{FS})$		$(FSA_{MD}) - (FSA_{FD})$	
	Coef.	**Std. err.**	**Coef.**	**Std. err**
$NFSH^M$	−0.079	0.002	0.001	0.370
$NFSH^F$	0.161	0.001	−0.020	0.019
STEM	−0.151	0.001	0.019	0.176
STEM × $NFSH^M$	−0.042	0.101	−0.037	0.306
STEM × $NFSH^F$	−0.037	0.098	−0.103	0.423
$FORC^M = 1$	−0.017	0.001	−0.016	0.004
$FORC^F = 1$	0.042	0.002	0.027	0.001
Constant	0.027	0.003	−0.073	0.003
Number of Obs.	21,018		19,942	

Notes: (1) Dependent variables are indicated in each column's heading. (2) Standard errors are adjusted by using the two-way clustering method (Cameron et al., 2011) at the individual and household levels. Coef. = coefficient; Std. err. = standard error.

Second, our underlying model is static and uses data that includes children mostly with completed majors. The evidence in the literature is very clear that students update their beliefs in their first years of study and switch majors, if the cost is endurable. We believe that using data on completed majors leads to a downward bias in our estimations.

Third, the constraint on available information in a family measured by the FSH would not necessarily suggest a positive bias in children's choice toward their parent's majors. Although it is less likely, 2 accountant parents would not necessarily be in favor of their child taking up their major and may deter their children from their own majors. This possibility would also create a downward bias in our estimations.

Finally, as is very common in most empirical studies in the field of education economics, our attempt to remove a possible ability bias from our estimations has its own limits. We think that specifications that use within-family differencing and a proxy that groups families with similar ability endowments substantially shrink the bias. Still, within-family differencing may have some other complications in our estimations. For example, a son's attraction to both parents' majors may not be homogeneously comparable, when each parent's attraction to their own field of study strongly reflects their gender preferences. There is an extensive literature on gender differences in field-of-study preferences (Zafar, 2013). Our hypothesis in this study implies that, when his mother's field of study becomes distinct from his father's major, the son's attraction to his father's major will be affected negatively. This is because the field-of-study diversity in the family will expand the information boundaries and consequently his choice set on majors. This may not be true, i.e., he will not be less attracted to his father's major, if his mother's choice of field of study is strongly gender based. Our expectation is that a possible bias due to this issue would lead to underestimation of the true effect of information constraints.

With all these caveats, we still believe that the transmission coefficients provide very valuable information on the intergenerational field-of-study elasticity, which is the first in the literature, to the best of our knowledge.

6 Concluding remarks

The potential spillover effect of education is a fundamental public policy matter because it may lead to progressive skill stratifications and dispersed income distributions in every generation if ability sorting in mating and across generations is substantial. Most studies use years of schooling as the educational outcome for children, treating education as unidimensional. Yet, educational decisions are no longer just about the quantity but about the specialization to be pursued as well. This study quantifies assortative mating by estimating FSH and intergenerational transmission of skills by measuring assortative preferences in the choice of major. As uncertainty increases with the complexity of educational choices, misinformed decisions made by students in choosing their field of study or by administrators in allocating their limited resources across disciplines would curtail social and economic progress. This study's primary objective is to investigate children's attraction to their parents' field of study, reflected by assortative tendencies in child–parent matches as an outcome of information asymmetry.

To identify the role of information asymmetry in assortative patterns in each field-of-study match between parents and children, we define quasi-likelihood transmission functions, wherein the response variables take on fractional values of FSA between each

child (son/daughter) and parent (father/mother) as a function of the spousal "appreciation" of each partner's field of study. We use the confidential major file of the 2011 NHS so that the size of the data and the availability of different levels of aggregation in the CIP allow us to develop 3 indicators: the degree of children's attraction to their parents' field of study (FSA), the degree of FSH, and the degree of relatedness between each parent's field of study and occupation (FOR).

Comparable to panel models, we define within-family transmission functions with 1-to-2 matches (1 for each parent). The results show that children's choice of field of study exhibits significant assortative preferences isolated from ability sorting and unobserved differences across majors and other family characteristics. We also find that the assortative tendency is the highest between fathers and sons relative to all other pairs, namely, father–daughter, mother–son, and mother–daughter. This evidence becomes even stronger when we use more disaggregated CIP codes and control for the educational degrees. With some caution, we attribute this persisting assortative tendency to the information asymmetry across alternative majors built on by parents' educational backgrounds within families.

References

Altonji, J. G.; P. Arcidiacono; A. Maurel (2015): The Analysis of Field of Study and Graduate School: Determinants and Wage Effects. NBER Working Paper 21655.

Arcidiacono, P.; G. Foster; N. Goodpaster; J. Kinsler (2012): Estimating spillovers using panel data, with an application to the classroom. Quantitative Economics, 3, 421-470.

Aydede, Y.; A. Dar (2016): The Cost of Immigrants' Occupational Mismatch and the Effectiveness of Postarrival Policies in Canada. IZA Journal of Migration 5(9), 1-23.

Aydemir, A.; W. Chen; M. Corak (2013): Intergenerational Education Mobility Among the Children of Canadian Immigrants. Canadian Public Policy 39, S108-S122.

Becker, G. S. (1973): A Theory of Marriage: Part I. Journal of Political Economy 81, 813-846.

Becker, G. S. (1974): A Theory of Marriage: Part II. Journal of Political Economy 82, S11-S26.

Becker, G. S.; N. Tomes (1979): An Equilibrium Theory of the Distribution Income and Intergenerational Mobility. Journal of Political Economy 87(6), 1153-1189.

Becker, G. S.; S. D. Kominers; K. M. Murphy; J. L. Spenkuch (2015): Theory of Intergenerational Mobility. University Library of Munich, Germany in its Series MPRA Paper with Number 66334. https://mpra.ub.uni-muenchen.de/66334/1/MPRA_paper_66334.pdf.

Bicakova, A.; S. Jurajda (2016): Field-of-Study Homogamy. IZA Discussion Papers 9844.

Black, S.; P. J. Devereux (2010): Recent Developments in Intergenerational Mobility. NBER Working Paper 15889.

Cameron, A. C.; J. Gelbach; D. Miller (2011): Robust Inference with Multi-Way Clustering. Journal of Business and Economic Statistics 29(2), 238-249.

Chevalier, A.; C. Harmon; V. O' Sullivan et al. (2013): The impact of parental income and education on the schooling of their children. IZA Journal of Labor Economics 2, 8.

Chiappori, P. A.; B. Salanie (2016): Econometrics of Matching Models. Journal of Economic Literature 54(3), 832-861.

Choo, E.; A. Siow (2006): Estimating A Marriage Matching Model with Spillover Effects. Demography 43(3), 463-490.

Corak, M. (2001): Are Kids All Right? Intergenerational Mobility and Child Well-Being in Canada. Centre for Living Standard & Institute for Research on Public Policy. The Review of Economic Performance and Social Progress V1. http://www.csls.ca/repsp/1/13-corak.pdf.

Corak, M. (2017): Divided Landscapes of Economic Opportunity: The Canadian Geography of Intergenerational Income Mobility, Working Papers 2017-043, Human Capital and Economic Opportunity Working Group.

Diamond, W.; N. Agarwal (2016): Latent Indices in Assortative Matching Models. http://economics.mit.edu/files/11242.

Hilger, N. G. (2016): The Great Escape: Intergenerational Mobility in the United States Since 1940. NBER Working Papers 21217.

Holmlund, H.; M. Lindahl; E. Plug (2011): The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods. Journal of Economic Literature 49(3), 615-651.

Hoxby, C.; C. Avery (2013): The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students. Brookings Papers on Economic Activity, 2013, 44(1) (Spring), 1-65.

Hu, Y.; S.M. Schennach (2008): Instrumental Variable Treatment of Nonclassical Measurement Error Models. Econometrica, 76, 195-216.

Lemieux, T. (2014): Occupations, Fields of Study, and Returns to Education. Canadian Journal of Economics 47(4), 1047-1077.

Mankiw, H. G.; R. Reis (2002): Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve. Quarterly Journal of Economics 117(4), 1295-1328.

McIntosh, J. (2010): Educational Mobility in Canada: Results from the 2001 General Social Survey. Empirical Economics 38, 457-470.

Papke, L. E.; J. M. Wooldridge (1996): Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. Journal of Applied Econometrics 11, 619-632.

Robst, J. (2007): Education and Job Match: The Relatedness of College Major and Work. Economics of Education Review 26, 397-407.

Roth, A. E.; M. A. O. Sotomayor (1992): Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Econometric Society Monographs No.18. Cambridge University Press, Cambridge, United Kingdom. http://web.stanford.edu/sim$alroth/papers/92_HGT_Two-SidedMatching.pdf

Schwartz, C. R. (2013): Trends and Variations in Assortative Mating: Causes and Consequences. Annual Review of Sociology 39, 451-470.

Shannon, C. E. (1948): A Mathematical Theory of Communication. The Bell System Technical Journal 27, 379-423.

Sims, C. (2003): Implications of rational inattention, Journal of Monetary Economics, 50(3), 665-690.

Solon, G. (2013): Theoretical Models of Inequality Transmission Across Multiple Generations. NBER Working Papers 18790.

Stinebrickner, R.; T. Stinebrickner (2014): A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout. Review of Economic Studies 81(1), 426-472.

Turcotte, M. (2011): Intergenerational Education Mobility: University Completion in Relation to Parents' Education Level. Statistics Canada, Canadian Social Trends. Catalogue No. 11-008-X.

Wiswall, M.; B. Zafar (2015): Determinants of College Major Choice: Identification Using an Information Experiment. Review of Economic Studies 82(2), 791-824.

Zafar, B. (2012): Double Majors: One for Me, One for the Parents? Economic Inquiry 50(2), 287-308.

Zafar, B. (2013): College Major Choice and the Gender Gap. Journal of Human Resources 48(3), 545-595.

Appendix on Equation (7)

The concept of information entropy in computer science was introduced by Shannon in 1948. The theory defines the basic model of information processing with 3 elements: a source of data, a communication channel, and a receiver. The "fundamental problem of communication" is for the receiver to be able to identify what data were generated by the source, based on the signal it receives through the (potentially noisy) channel.

Different from the "sticky price model" (Mankiw and Reis, 2002), in which the agents receive the data with delay yet without error, Sims uses Shannon's information theory and argues that people have limited information-processing capacity, which alters the information for each individual and thus differentiates their behaviors. His model, called the "noisy information model", provides a convenient framework in our context because the information flow is modeled by probability distributions. The difference between the joint distribution of population values of choice-specific outcomes (the "input" of the channel, $F_i(Z|m,t)$, typically representing the "true" distribution of data) and the subjective beliefs on them (the "output" of the channel, $G_i(Z|k,t)$ typically representing an "approximation" of $F_i(Z|m,t)$ can be considered as the information loss due to the channel capacity. The divergence of $G_i(Z|k,t)$ from $F_i(Z|m,t)$, also called "relative entropy" or Kullback–Leibler (D_{KL}) divergence, reflects the amount of information lost when G_i is used to approximate F_i and can be expressed as follows:

$$D_{KL,i} = \sum f_i(z) \log \left[f_i(z) / g_i(z) \right]$$

Since f_i represents the "true" (joint, marginal, or conditional) distribution of the data, differences in D_{KL} across individuals can only be explained by varying g_i, different approximations of f_i for any given value, z.

This example illustrates a more general point: the optimal prediction of f_i with g_i that minimizes the information loss depends on the information-processing capacity of each channel. Therefore, there must exist 1 optimal channel among the many options for each f_i that maximizes the mutual information measured by D_{KL}. Yet, when these options are not readily and equally available, the optimization problem leads to suboptimal choices that fit the channel capacity. Sims (2003) suggests that the nature of the "noise" that quantifies the information-flow constraint in each channel does not need to be exogenous as in physics. Instead, the available information capacity delivers a model for the "noise" in many applications in economics.

In our context, parents serve as communication channels, not as the source of data, in transmitting publicly available information on choice-specific outcomes to their child, the "receiver". The information on choice-specific outcomes is not generated by the parents. These outcomes are random variables, and the information on them is publicly available. The parents' capacity can also be defined by their ability to reach out to available sources. Sims (2003) calls this concept as "individuals' information-processing constraints".

The channel capacity on major m, defined by D_{KL}, can be expressed as a function of the parents' fields of study (FOS) and whether they practice their majors (FOR):

$$E_i\left(D_{KL} \mid m,t\right) = f\left(\mathrm{FOS}^F, \mathrm{FOS}^M, \mathrm{FOR}^F, \mathrm{FOR}^M, s\right)$$

All other factors, including public domains that are accessible by the parents of individual *i* and may affect their capacity on major *m,* are included in vector *s*. For example, even if both parents are teachers (major *k*), the parents' friends (or close relatives) who are dentists (major *m*) would reinforce the parents' capacity and reduce the entropy in transmitting the information on major *m*. Although modeling information flows is a complex task, one simple way to approximate the above expression is to define it as conditional on one of the parent's major. Thus, assuming linearity in *f*(.), the level of information-processing constraint can be expressed as Equation (7) in the text.