Assortative Preferences and Information Constraint in Choice of Major

Yigit Aydede
Economics, Saint Mary's University,
Canada

Work-in-progress
Prepared for IZA World Labor Conference 2018

June 24, 2018

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 one-to-two matches (one 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.

***Keywords*:** Intergenerational transmission of education; field-of-study homogamy; choice of field of study; occupational relatedness

***JEL Classification*:** J1, I2, D1

Canada ranks third after South Korea and Japan among OECD countries with close to 60 percent of population aged between 25 and 34 holding a university degree in 2011 (OECD 2013). This is a positive development if education is to continue to operate as an engine of the economic and social progress. Yet, educational decisions are no longer just about the quantity, but about specialization to pursue as well. 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 the progress. A question from the 2013 National Graduate Survey (NGS) reveals that about

24 percent of university graduates in Canada would not choose the same field of study had they studied again. The same is true for about 36 percent of graduates majoring in Physical and Life Sciences and Technologies. Boudarbat and Chernoff (2012) report that 35.1 percent of Canadian university graduates are in jobs that are not related to their education five years after graduation. College attrition is also a major concern. Recent numbers in the United States, for example, show that about half of the students entering the college do not earn a bachelor's degree within six years, and one-third of them drop out entirely. Arcidiacono et al. (2016) attributes this attrition problem to information frictions in schooling and work decisions that are built on the ability component gradually revealed to students as they cumulate more information.

Although the gap in lifetime income earnings between university majors in Canada is substantial, recent studies point to significant but quantitatively small elasticities of major choice to expected earnings (Beffy et al. 2012, Altnoji et al. 2015). Oreopoulos and Salvanes (2009) present evidence that non-pecuniary returns to schooling as large as pecuniary ones. Expectations about 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 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 2013, Arcidiacono et al. 2012).

The evidence also shows that parental approval is one of the most important determinants in major choices (Zafar 2011)[1]. According to the 2013 NGS, close to 60 percent of university graduates report that their parents' recommendations played a very important role in their choice of major. This is not surprising as the information and its value from different sources becomes more dispersed and questionable. Altonji et al. (2015), for example, documented that Princeton pushes students to consider departments with fewer students. Some postsecondary institutions prefer a distribution of students across majors that 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 realization 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.[2][3] Yet, a significant assortativity (a

[1]The peer effect on major choice has not been a subject of much attention. There are two notable exceptions: Sacerdote (2001) found no effect, De Giorgi et al. (2010) have found a significant effect.

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

[3]Although the system is different from one where the major is chosen at the university entrance

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 realization of the future major-specific outcomes is bounded by parents' fields of study.[4]

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 question how the field-of-study homogamy and whether the parents work in related occupations affects the magnitude of information asymmetry. While we try to address these questions through this paper, the answer would seem obvious if we change our example to one where the father is an accountant but the mother is a biochemist, and both are working in related occupations. The following two empirical questions need to be answered to assess the role of information asymmetry in 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 individual tastes embedded in the resemblance of majors between each parent and child? We will try to answer both questions in this paper.

Whatever the reason is, what would the presence of assortative preferences in choice of major, if they exist, suggest for a society?[5] It is often argued that there are two postwar trends simultaneously observed in the Western world: a rising number of post-secondary graduates who are increasingly sorting into homogamous marriages (Schwartz and Mare 2005) and an upsurge in household income inequality (Western et al. 2008). The leading explanation that connects these two empirical

with a centralized test or a threshold GPA required for each major as in England, France, Chile and Norway, students in Canada are accepted to universities at three general faculty levels: Arts, Science, and Business/Commerce. Each of these requires different courses completed in high schools with competing GPAs at grades 11 and 12. Therefore, although the majors are decided after the 2nd year, roughly after completing 14 to 16 core courses within each faculty, switching majors across faculties is costly.

[4]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 towards (or against) a specific field of study. Zafar (2011) investigates this issue in his recent paper titled, "Double majors: one for me, one for the parents?"

[5]Individually, biased choices may be optimal in the short term when students can benefit from a rising marginal productivity in their education as parents would be more productive helpers with schoolwork in similar fields of study. Yet, the long-term personal cost of decisions made with bounded information is obvious as the evidence suggests that the choice of major does limit a person's career choice.

facts is fairly simple. A shift in the relative demand in favor of skilled workers led to a higher skill premium (Acemoglu and Autor, 2011) and a more polarized earning distribution due to skill sorting in mating (Eika et al. 2017). Understanding the spillover effect of education, therefore, has become a fundamental policy matter, as intergenerational skill transfers might bolster skill stratifications in every generation through homogenous marriages leading to progressively dispersed earning distributions. One empirical challenge in this argument is to measure skill sorting in marriages and its transmission across generations. In the literature, it is implicitly assumed that assortative mating in education (men and women with the same education measured by earned degrees marrying more or less frequently than the random patterns) reflects stronger sorting of partners with respect to skills. Yet, an additional year in a Bachelor of Arts with a history major is quite different than that of engineering. Large earning and ability differences exist across majors and the returns to skill have been substantially increasing over time (Gemici and Wiswall 2014). Hence, understanding the degree of ability sorting in marriages by field-of-study homogamy and its intergenerational transmission by assortative preferences in children's choice of major provides an important policy tool, as it shows the extend to which information constraint drives the spillover effect of education.

This topic is also nested in the discussions on the genetic and environmental roots of educational inequality that the genome wide association studies have been increasingly able to address (Nature 2017). The evidence shows that intelligence is one of the most heritable behavioral traits and that assortative mating is greater for intelligence than for any other behavioural (personality or psychopathology) or physical (height or weight) traits (Plomin and Deary 2015). There is a growing consensus that both high heritability and assortative mating aspects of intelligence might pump additive genetic variance into the population, which in turn contribute to rising ability stratifications and social inequalities in every generation to come (Hugh-Jones et al. 2016). Thus, the role of information boundaries in assortative preferences in choice of major (in addition to the child's genetic endowments) is an important question to answer with significant policy consequences.[6]

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 field-of-study homogamy and whether parents work in their trained jobs within Canadian families. We use the confidential major file of the 2011 National Household Survey so that the size of the data and the availability of different levels of aggregation in the Classification of Instructional Programs (CIP) allows us to develop three indicators: the degree of children's attraction to their parents' field of study (FSA), the degree of field-of-study homogamy (FSH), and the degree of relatedness between each parent's field of study and occupation (FOR). To identify the role of information

[6] A good summary on this topic can be found in Heckman and Masso (2014).

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 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 with this problem, we define within-family transmission functions based on a non-parametric assortative matching model with one-to-two matches (one 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 field-of-study homogamy 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 the evidence of a substantial degree of gender heterogeneity in the preferences for each major (Wiswall and Zafar 2015). These findings, 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 a higher assortativity in each child-parent combination is strongly associated with a greater homogamy and field-of-study relatedness in parents' jobs. The empirical approach that we applied aims is 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 1 explains subjective expectations in choice specific models; Section 2 introduces the data, assortative preferences, homogamy, and occupational relatedness; the empirical framework is explained in Section 3; the estimation results are reported in Section 4; and we provide the concluding remarks in Section 5.

1 Conceptual background

This study brings together three different but interrelated fields in the literature: choice of field of study, assortative mating, and intergenerational transmission of education. To date, children's choice of field of study has been investigated as a separate subject mostly in relation to prospects of majors, such as expected wage earnings and employment opportunities (see Altonji, Arcidiacona, and Maurel 2015 for a comprehensive review). The non-pecuniary determinants, such as the effect of parental influence, have mainly been overlooked due to the lack of available data. Evidence in recent studies unambiguously indicates that gaining parental approval is the most important determinant of major choice, in addition to enjoying coursework. Baudarbat and Montmarquette (2009) also found that parents have a strong bias against or in favor of some fields of study that affects Canadian graduates when in their choice of major.

What is the parents' role in forming subjective expectations about choice-specific outcomes? In choice models, forward-looking individuals derive their current utility from each major and choose one among many alternatives that maximizes their utility, which is a function of two vectors of outcomes, $\mathbf{a}$ and $\mathbf{c}$, realized in college and after graduation, respectively. Successfully completing a major (graduating), enjoying the coursework, and parental approval are the outcomes among those realized in college. Among several other outcomes, the ones that are realized after graduating are related to income, employability, and the social status of available jobs (Zafar 2011). Individual i with a set of characteristics, $\mathbf{Z}_{it}$, chooses major k (for all $k \in \mathrm{C}_i$) based on his subjective beliefs (expectations), $\mathrm{Pr}_{ikt}(\mathbf{a}, \mathbf{c})$, on outcomes that are uncertain at time t. If he chooses major m, the standard revealed preferences argument implies that:

$$m \equiv \arg\max_{k \in C_i} \int U_{it}\left(\mathrm{a}, \mathrm{c}, \mathrm{Z}_{it}\right) d\mathrm{Pr}_{ikt}\left(\mathrm{a}, \mathrm{c}\right). \tag{1}$$

Given his preferences, the ex-ante treatment effect, $\mathrm{Pr}_{imt}(\mathbf{a}, \mathbf{c}) - \mathrm{Pr}_{ikt}(\mathbf{a}, \mathbf{c})$, will be reflected in his choice of m, which means that two individuals with identical utility functions and characteristics can choose two different majors if they hold different beliefs about the choice-specific outcomes. But, it also means that the person's choice of major, m, could be suboptimal due to a lack of information on major k and he would update his choice as he obtains new information. His persistency on his current choice of major m depends not only on the cost of switching to a different major, but also the initial state of his beliefs. Zafar (2011) finds that students update their beliefs for various outcomes for pursued as well as non-pursued majors. Yet, he also finds that students who switch majors are mainly responding to new information about their major. This is an important point because if the student chooses major m based on more and reliable information associated with his parents' own fields of study, his response to new information on major k would not make a difference on his choice of major m, even if it is optimal for him to switch to k. Consistently, the evidence also shows that students who are more uncertain about the major-specific outcomes when choosing their majors make greater revisions in their beliefs. In other words, if the student is equally uncertain about major k and m (i.e. more symmetric information on k and m), updating his beliefs and switching from m to k would be more likely.[7]

Although theoretical work incorporates the uncertainty in schooling decisions, most empirical studies assume that individuals are rational and use realized (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, Hastings et al. 2015, Arcidiacono et al. 2012, Stinebrickner and Stinebrickner 2014) address this identification problem by directly eliciting subjective beliefs from a sample of university students. Although the evidence in these studies reveals that subjective expectations on major-specific outcomes greatly varies across individuals, there is a lack of evidence as to why beliefs are so dispersed around the true population values. This is exactly what we try to answer in this study.[8] Once the expectations are assumed to be subjective, they could vary based on parental background in education, which in turn lead to differential choice sets. In other words, a set of alternative majors for those whose parents are engineers would be different than for those whose parents are biochemists due to the information asymmetry.

In this study, we want to understand the role of parents' educational background in the process of expectations formation by looking at assortative preferences that result from asymmetric information. The main driver of child i's attraction to major m revealed in his choice is the expected lifetime utility from the vector of future outcomes ($\mathbf{X}$) of a specific human capital endowment if major m is chosen as defined below:

$$E_i V_{i,m} = \sum_{t=1}^{T} \beta^{t-1} \int U(\mathbf{X}) \mathrm{dG_i} (\mathbf{X} \mid m, t). \tag{2}$$

This implies that the appeal of major m would be different than that of major k for individual i due to differences in beliefs reflected in the subjective joint probability distribution, $G(\mathbf{X}|m,\ t)$, even if the majors have identical distributions in terms of their observed outcomes, that is $F(\mathbf{X}|m,\ t) = F(\mathbf{X}|k,\ t)$. 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 first introduced by Shannon (1948) defines the basic model of a data communication system with three

[7] This could be the case, for example, when both parents are less well-educated without a field of study.

[8] To test for information asymmetry, one could also elicit beliefs about future outcomes of majors and then see if the students who have parents with the majors in question are more knowledgable about the (average) future outcomes.

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 was generated by the source, based on the signal it receives through the (potentially noisy) channel. Differences in the "channel capacity" make the same data communicated differently, thus introduce an additional uncertainty ("noise" or entropy) to its value. In the information theory, a channel is defined as "the medium over which the signal is sent" and its capacity can be defined as the "mutual information" between the message sent and the message received. Shannon describes channel capacity (to reliably carry the information) by using probability distributions of the channel's input and the output. More specifically, if $\boldsymbol{I}$ and $\boldsymbol{O}$ are the random variables representing the input and output of the channel, respectively, the conditional probability of $\boldsymbol{O}$ given $\boldsymbol{I}$, $\Pr(\boldsymbol{O}|\boldsymbol{I})$, which is an inherent fixed property of the channel, differentiates its capacity from others as expressed below:

$$C = \sum_{I,O} \log \left[\Pr(\boldsymbol{I}) \left[\Pr(\boldsymbol{O} \mid \boldsymbol{I}) - \Pr(\boldsymbol{O}) \right] \right]. \tag{3}$$

The key element in this expression is the distance between $\Pr(\boldsymbol{O}|\boldsymbol{I})$ and $\Pr(\boldsymbol{O})$, which measures the "mutual information", the output driven by the input rather than the "noise" in the channel. When $\Pr(\boldsymbol{O}|\boldsymbol{I}) = \Pr(\boldsymbol{O})$, the whole output represents a white noise (Gaussian noise) suggesting that the channel has a zero-capacity, C , for the input, $\boldsymbol{I}$.[9]

This idea provides a convenient framework in our context: the difference between observed distributions of population values of choice-specific outcomes (the information on $\mathbf{X}$ of major m at the source, $F_i(\mathbf{X}|m, t)$) and the subjective beliefs on them (the information on $\mathbf{X}$ at the receiver ($G_i(\mathbf{X}|m, t)$) can be considered as information entropy due to the channel capacity. More specifically, the information on $\mathbf{X}$ of m can be defined as a set of random events, x_i, (hourly wage earnings 5 years after graduation, unemployment durations, and so on) with their joint probability distribution. Hence, the information on $\mathbf{X}$ sent with joint probability distribution F_i and received with joint probability distribution G_i will be different if G_i diverges from F_i. This difference, also called "relative entropy" or Kullback-Leibler (D_{KL}) divergence, reflects an amount of information lost when G_i is used to approximate F_i and can be expressed as follows.[10]

$$G_i \left(\mathbf{X} \mid m, t \right) - F_i \left(\mathbf{X} \mid m, t \right) = \Theta_{i,m}. \tag{4}$$

[9]The channel capacity, the maximum mutual information between $\boldsymbol{I}$ and $\boldsymbol{O}$, is the upper limit of C over all possible choices of $\Pr(\boldsymbol{I})$.

[10]The complete formula of D_{KL} for discrete probability distributions would be:

$$D_{KL} = -\sum \mathrm{F}(\boldsymbol{x}) \log \left[\mathrm{G}(\boldsymbol{x}) - \mathrm{F}(\boldsymbol{x}) \right].$$

In our context, parents serve as communication channels, not as the source of data, in transmitting publicly available information on observed choice-specific outcomes to their child, the "receiver".[11] Their capacity 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 nuclear physics or accounting. When we accept that the channel capacity on major m can be represented by $E_{i,m}(O|I,t)$, and is bounded by the parents' educational background, we can approximate it with the level of entropy, $\Theta_{i,m}$, given by (4), as a function of parents' fields of study (FOS for **F**ather and **M**other) and whether they practice their majors (FOR):

$$E_{i,m}(O|I,t) = \Theta_{i,m} = g_i(FOS^F, FOS^M, FOR^F, FOR^M, \text{z}). \tag{5}$$

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 $\mathbf{z}$.[12]

Although modeling information flows is a complex task, one simple way to approximate the behavior of Θ in our context expressed by (5) is to define it as conditional on one of the parent's major. Thus, assuming linearity in $g(.)$, the channeling capacity of the parents can be represented as follows:

$$E_{i,m}(O|I, FOS^F = m, t) = \gamma + \propto FSH^M + \delta FOR^F + \beta FOR^M FSH^M + \rho \text{z}, \tag{6}$$

which implies that, when $FOS^F = m$, the expected level of information received by the child on major m is equal to an index number (γ, how less "noisy" a channel the father is for the information on his own major) plus how compatible the mother's major is to the father's major (FSH^M) and the degree of relatedness between the parent's fields of study and their occupations (FOR). The key element in this expression is FSH^M. As explained later, it reflects the degree of relatedness (normalized between 0 and 1) between each parent's field of study. Suppose that the mother's major is the least-related major to her husband's major ($FSH^M = 0$). 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

[11]The information on choice-specific outcomes ($\mathbf{X}$) is not generated by the parents. These outcomes are population values and the information on them is publicly available.

[12]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.

also means that parents become less efficient channels for other majors with a rising relative entropy. Therefore, the level of field-of-study homogamy defines the level of information asymmetry in a family.[13]

This example becomes less intuitive when we compare two 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 values of FSH^M would be different, it appears that, these two 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 not only be determined by 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 collectively reflected in FSH^M.[14]

In what follows, we first define and calculate assortative preferences by measuring the attraction of children to their parents' majors (FSA). To reflect the differences in available information on each major across families, we also define two indices, the degree of field-of-study homogamy (FSH) and the degree of relatedness between each parent's field of study and occupation (FOR). Our empirical approach to test whether a higher level of available information in terms of its reliability, cost, and volume could be the main driver of children's assortative preferences towards their parents' fields of study is explained in Section 3.

2 Data, assortative preferences, homogamy, and occupation match

2.1 Data

This study uses the confidential major file of the 2011 National Household Survey (NHS) only available in Canadian Research Data Centers. We restricted the data to include only non-Aboriginal native-born individuals living in 10 provinces. We also dropped nondegree-holder parents (that is, those with no education or an education degree that does not grant a major) and those whose field of study contains fewer than 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.

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

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

Statistics Canada classifies the major fields of study by using the Classification of Instructional Programs (CIP), which includes 1,688 instructional program classes with finer breakdowns provided with up to six-digit codes. Unlike earlier censuses, NHS also includes variables that group CIP codes into 4 different levels. The most aggregated level classifies CIP codes in 12 major groups. This aggregation is reduced to 41 and 372 groups, and down to the most detailed level where all majors are presented with 1,688 CIP codes.[15]

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, for example, Chevalier et al. (2013) use a subsample from a pool of Labour Force Surveys in the U.K. that include children aged 16 to 18 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 thousand females with an identified field of study between the ages of 19 and 21 in the whole sample, our subsample includes 26 thousand who live with their parents. Moreover, we use this subsample only for FSA calculations, while 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 was representative of the population, we first compared the distribution of parents (fathers and mothers, separately) living with their children to that of the whole sample across CIP codes classified at 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.[16] More descriptive information about the data and our samples will be provided in the following sections that explain FSH, FSA, and FOR.

[15] For more information on CIP classification see www.statcan.gc.ca/concepts /classification-eng.htm.

[16] Hilger (2015) develops a new method to adjust the data to recover the outcomes of "missing" independent children. However, their educational outcome is measured in 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 did not change significantly.

Table 1: FSA - Sons attraction to father's and mother's major by prime CIP codes (weighted)

		Son's major										
Father's major		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Education	1	**0.939**	0.797	1.000	0.813	0.795	0.741	0.735	0.000	0.703	0.879	0.631
Arts	2	0.319	**1.000**	0.370	0.349	0.306	0.356	0.321	0.000	0.209	0.352	0.381
Humanities	3	0.597	0.715	**1.000**	0.729	0.509	0.586	0.571	0.000	0.557	0.587	0.476
Law	4	0.647	0.715	0.871	**1.000**	0.773	0.682	0.655	0.000	0.598	0.677	0.528
Business	5	0.546	0.541	0.636	0.688	**1.000**	0.614	0.573	0.000	0.502	0.556	0.429
Science	6	0.639	0.626	0.674	0.726	0.597	**1.000**	0.694	0.000	0.512	0.627	0.352
Math/Comp.	7	0.639	0.759	0.875	0.629	0.730	0.982	**1.000**	0.000	0.524	0.636	0.543
Engineering	8	0.171	0.144	0.004	0.023	0.000	0.082	0.181	**1.000**	0.197	0.143	0.211
Agriculture	9	0.228	0.098	0.012	0.111	0.015	0.199	0.155	0.000	**1.000**	0.259	0.512
Health	10	0.694	0.658	0.757	0.748	0.632	1.000	0.574	0.000	0.569	**0.843**	0.588
Services	11	0.247	0.208	0.078	0.062	0.000	0.120	0.246	0.514	0.237	0.234	**1.000**
Mother's major												
Education	1	**0.977**	0.839	0.944	1.000	0.856	0.842	0.772	0.000	0.771	0.849	0.463
Arts	2	0.378	**1.000**	0.651	0.624	0.402	0.451	0.417	0.000	0.371	0.415	0.148
Humanities	3	0.624	0.709	**1.000**	0.782	0.702	0.739	0.674	0.000	0.568	0.581	0.460
Law	4	0.518	0.585	0.694	**1.000**	0.752	0.532	0.467	0.000	0.554	0.543	0.466
Business	5	0.212	0.138	0.064	0.000	**0.367**	0.167	0.286	1.000	0.254	0.157	0.383
Science	6	0.723	0.644	0.696	0.658	0.752	**1.000**	0.704	0.000	0.595	0.635	0.565
Math/Comp.	7	0.386	0.437	0.505	0.377	0.000	0.959	**1.000**	0.449	0.299	0.198	0.047
Engineering	8	0.135	0.188	0.097	0.093	0.000	0.274	0.231	**1.000**	0.171	0.328	0.252
Agriculture	9	0.393	0.432	0.273	0.402	0.000	0.418	0.324	0.692	**1.000**	0.337	0.422
Health	10	0.447	0.177	0.000	0.067	0.023	0.416	0.341	0.840	0.569	**1.000**	0.957
Services	11	0.210	0.265	0.058	0.000	0.018	0.206	0.256	1.000	0.279	0.222	**0.595**

Notes: Due to 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, (11) Personal, protective and transportation services.

2.2 Quantifying assortative preferences: field-of-study attraction - 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 two based on the probability distributions. We create four 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.

Economists have investigated the identification and estimation of preferences in large matching markets where the matching is positive assortative along a single dimensional index (Diamond and Agarwal 2016, Chiappori et al. 2012, Choo and Siow 2006). Among the measures used to identify observed matching patterns, we choose the following identity that reflects the differences between observed and expected frequencies under independence:

$$\text{FSA} = \text{P}(\boldsymbol{P})[\text{P}(\boldsymbol{K} \mid \boldsymbol{P}) - \text{P}(\boldsymbol{K})],$$

where $\boldsymbol{P}$ and $\boldsymbol{K}$ are indicators of fields of study for parents and children in matching, respectively. Weighted by the marginal probability of the parent's choice of major i, $\text{P}(\boldsymbol{P} = i)$, the term in brackets shows how much the children's choice of major j is pulled ($\text{P}(\boldsymbol{K}{=}j|\boldsymbol{P}{=}i) > \text{P}(\boldsymbol{K}{=}j)$) by or pushed ($\text{P}(\boldsymbol{K}{=}j|\boldsymbol{P}{=}i) < \text{P}(\boldsymbol{K}{=}j)$) from the parent's field of study. The index is calculated for each cell of the P-K table, which is a square matrix of the same number of fields of study, i and j, both for parents and children, respectively.[17] 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.[18] While we use 12, 41, and 137 major groups of CIP, in the four match tables, we report only the sons' match calculated with 12 major CIP codes in Table 1. The higher values of FSA on the diagonal show 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

[17] With the number of matches, m_{ij}, in each cell of the match table, FSA can be calculated by m_{ij}/T - $m_i m_j/T^2$, where T is the total number of pairs.

[18] Obviously, children cannot choose their parents as partners to match as in the marriage market but they choose a major that matches them to their parents. This type of two-sided matches is well-recognized in the literature (Roth and Sotomayor 1992).

even if the parents have the same field of study. Intuitively, the same major could be more (or less) attractive for the son, for example, 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, a couple of interesting observations are worth mentioning. It seems that engineering is the most avoided major by all sons, unless their parents hold it. Moreover, the assortativity is very distinct when the parents' field of study is engineering, arts, or agriculture. However, when the parent is a teacher, the sons' attraction to their parent's major is not so distinctand considerably dispersed among others. Finally, to compare the extent of field-of-study attraction between parents and children across four match tables, we use an index, $\boldsymbol{H}$-index, suggested by Bicakova and Jurajda (2016). The index computes the ratio of two diagonal shares of a match matrix as follows:

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

where both the nominator and the denominator are calculated for $j = i$, 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 the same field of study (on the diagonal) to the share of matches that one would expect under the random matching assumption. When the children's choice of major is not affected by their parents' field of study, each joint probability on the diagonal (nominator) approaches to the product of its marginal probabilities, then the whole index becomes zero. Hence any departure from zero indicates the tendency towards the same field-of-study matches. $\boldsymbol{H}$-indices calculated for 41 major CIP codes are: 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.

2.3 Field-of-study homogamy - FSH

Assortative mating has long been documented by demographers using non-parametric 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 et al. (2016), for example, show that educational homogamy of posterity is likely to be reinforced by increases in parents' human capital who are matched homogamously themselves. Bicakova and Jurajda (2016) are the first to analyze mating by field of study for European countries. Using the European Labor Force Survey, they show that there is a great extent of FSH among couples formed by college graduates in the twenty-four EU countries.

Unlike joint or conditional probabilities that define the likelihood of a match, we use the same approach applied in FSA 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:

$$\mathrm{FSH} = \mathrm{P}(\boldsymbol{M})[\mathrm{P}(\boldsymbol{F} \mid \boldsymbol{M}) - \mathrm{P}(\boldsymbol{F})] = \mathrm{P}(\boldsymbol{F})[\mathrm{P}(\boldsymbol{M} \mid \boldsymbol{F}) - \mathrm{P}(\boldsymbol{M})],$$

where $\mathbf{F}$ and $\mathbf{M}$ are indicators of fields of study for female and male mates in matching couples. There are three possible outcomes suggested by this measure calculated for each cell of the F-M table, which is a square matrix of the same number of fields of study, i and j, both for females and males, respectively. First, if the conditional probability of one partner's field of study, $\mathrm{P}(\mathbf{F} = i|\mathbf{M} = j)$ or $\mathrm{P}(\mathbf{M} = j|\mathbf{F} = i)$, gets closer to its marginal probability, $\mathrm{P}(\mathbf{F} = i)$ or $\mathrm{P}(\mathbf{M} = j)$, the whole term approaches to zero, which indicates a complete neutrality in attraction between two fields of study, i and j. When $\mathrm{P}(\mathbf{F}|\mathbf{M})$ is higher than $\mathrm{P}(\mathbf{F})$, the value reflects the magnitude of a positive pull between i and j in mating. A negative value, on the other hand, suggests a rising aversion between them. With the number of matching couples, m_{ij}, in each cell of the match table, FSH can be calculated as follows:

$$FSM = \frac{m_{ij}}{T} - \frac{m_i m_j}{T^2}, \tag{7}$$

where T is the total number of couples and m_i and m_j indicate the total number of female and male partners in fields of study of i and j, respectively. This approach was criticized by Choo and Siow (2006) on the grounds that it negates the equilibrium effects of policy interventions or changes in market structure. More specifically, equation (7) ignores the individuals who choose not to marry and it assumes that the number of marriages between two types of individuals is unaffected by changes in the number of other types of individuals. Instead, they offer a new measure, CS,

Table 2: 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		**Overall normalization**										
Education	1	1.000	0.377	0.394	0.323	0.169	0.391	0.384	0.373	0.387	0.322	0.331
Arts	2	0.374	0.556	0.427	0.428	0.334	0.399	0.402	0.398	0.401	0.337	0.396
Humanities	3	0.425	0.426	0.636	0.422	0.269	0.408	0.402	0.387	0.396	0.320	0.361
Law	4	0.424	0.421	0.475	0.749	0.235	0.412	0.398	0.381	0.396	0.242	0.321
Business	5	0.366	0.384	0.409	0.429	0.853	0.393	0.388	0.358	0.366	0.217	0.289
Science	6	0.410	0.404	0.435	0.402	0.308	0.527	0.494	0.401	0.411	0.383	0.365
Math/Comp.	7	0.367	0.416	0.416	0.435	0.395	0.415	0.510	0.402	0.400	0.320	0.377
Engineering	8	0.000	0.305	0.148	0.146	0.877	0.311	0.399	0.575	0.390	0.618	0.694
Agriculture	9	0.408	0.392	0.376	0.387	0.363	0.409	0.392	0.397	0.515	0.418	0.395
Health	10	0.378	0.384	0.385	0.362	0.207	0.411	0.383	0.384	0.391	0.809	0.359
Services	11	0.303	0.387	0.351	0.369	0.444	0.376	0.397	0.395	0.401	0.466	0.563
		Normalized by husband's major										
Education	1	**1.000**	0.250	0.271	0.185	0.000	0.267	0.259	0.245	0.262	0.184	0.195
Arts	2	0.180	**1.000**	0.419	0.423	0.000	0.293	0.306	0.288	0.302	0.014	0.279
Humanities	3	0.425	0.428	**1.000**	0.417	0.000	0.379	0.362	0.322	0.346	0.139	0.251
Law	4	0.368	0.362	0.467	**1.000**	0.000	0.344	0.317	0.284	0.313	0.014	0.167
Business	5	0.234	0.263	0.302	0.333	**1.000**	0.277	0.269	0.222	0.234	0.000	0.113
Science	6	0.466	0.438	0.580	0.429	0.000	**1.000**	0.849	0.425	0.470	0.342	0.260
Math/Comp.	7	0.247	0.505	0.505	0.605	0.395	0.500	**1.000**	0.432	0.421	0.000	0.300
Engineering	8	0.000	0.348	0.169	0.166	1.000	0.355	0.455	**0.656**	0.445	0.705	0.791
Agriculture	9	0.296	0.191	0.086	0.158	0.000	0.303	0.191	0.224	**1.000**	0.362	0.211
Health	10	0.284	0.294	0.296	0.257	0.000	0.339	0.292	0.294	0.306	**1.000**	0.252
Services	11	0.000	0.323	0.185	0.254	0.542	0.281	0.362	0.354	0.377	0.627	**1.000**
		Normalized by wife's major										
Education	1	**1.000**	0.287	0.504	0.294	0.000	0.370	0.008	0.069	0.141	0.177	0.104
Arts	2	0.374	**1.000**	0.572	0.468	0.233	0.407	0.150	0.184	0.235	0.203	0.264
Humanities	3	0.425	0.482	**1.000**	0.458	0.141	0.449	0.150	0.134	0.201	0.174	0.178
Law	4	0.424	0.462	0.670	**1.000**	0.093	0.468	0.118	0.106	0.201	0.042	0.079
Business	5	0.366	0.315	0.535	0.469	**0.966**	0.380	0.039	0.000	0.000	0.000	0.000
Science	6	0.410	0.394	0.588	0.425	0.196	**1.000**	0.874	0.198	0.302	0.280	0.188
Math/Comp.	7	0.367	0.442	0.549	0.479	0.319	0.481	**1.000**	0.203	0.228	0.174	0.217
Engineering	8	0.000	0.000	0.000	0.000	1.000	0.000	0.126	**1.000**	0.161	0.677	1.000
Agriculture	9	0.408	0.347	0.467	0.400	0.274	0.454	0.071	0.180	**1.000**	0.340	0.262
Health	10	0.378	0.315	0.486	0.358	0.054	0.463	0.000	0.120	0.168	**1.000**	0.173
Services	11	0.303	0.327	0.416	0.370	0.388	0.301	0.110	0.171	0.235	0.421	**0.677**

Notes: (1) See the notes to Table 1 for the full description of majors. (2) The last major, 12 Other, is ignored in the table as it includes very few observations. (3) The sample used in this table contains all working spouses regardless of whether they have children with or without an identified CIP code.

a scale-free "marriage matching function" based on a transferable utility model, as follows:

$$\text{CS} = \ln\left[\frac{m_{ij}}{\sqrt{m_{i0}m_{0j}}}\right],$$

where m_{i0} and m_{0j} indicate the total number of female and male unmarried individuals in fields of study of i and j, respectively. Their approach offers a behavioral interpretation of CS as a marriage gain, a log odds-ratio of matched vs. unmatched, which can represent a couple's systematic gains to marriage relative to remaining single. While we applied both methods by using 12 and 41 major groups of CIP, to avoid reporting multiple large tables in our limited space, we only provide the results based on FSH calculated with 12 major CIP codes. The top section of Table 2 shows normalized FSH measures calculated by equation (7).

As suggested by Choo and Siow (2006), one expects the systematic gains to marriage to be large for i, j pairs if one observes many i, j marriages. Hence each cell of the matrix reflects the level of this gain revealed by the assortativity embedded in the match relative to the most homogamous match. Except for three majors (business, engineering and services), the measures on the diagonal are the highest in their respective columns and rows indicating the presence of a strong field-of-study homogamy. Further, the comparison of measures on the diagonal reveals the differences in the degree of homogamy across all homogamous matches in which teachers (education) have the strongest field-of-study homogamy. Although the overall normalization in the top section of Table 2 exposes the ranking of each match in terms of its assortativity among all matches, a more interesting comparison would be the ranking of individual specific systematic gains in each match among the matches that are only possible with the male or female partner's major in each match.

As recognized in the literature, observed matches in a marriage market are jointly determined by the preferences of both partners. The match between a male accountant (business) and a female historian (humanities), for example, is ranked at 0.409 relative to the most homogamous match between teachers. While this comparison reveals the assortativeness between a male accountant and a female historian in mating, it would be quite possible that a male accountant's attraction to a female historian would be different than her attraction to a male accountant. For example, Choo and Siow 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 and identify the systematic gains to marriage for a type i male and a type j female in an ij marriage, respectively, as follows:

$$n_{ij} = \ln\left[\frac{m_{ij}}{\sqrt{m_{i0}}}\right],$$

$$N_{ij} = \ln\left[\frac{m_{ij}}{\sqrt{m_{0j}}}\right].$$

These measures reveal the differences in individual gains in an ij marriage reflecting a spousal "appreciation" or "attraction" of each of field of study in mating. One can compare not only the gain of each spouse in a given marriage ($n_{ij} \gtrless N_{ij}$) but also rank fields of study in terms of each spouse's gain with possible alternatives ($n_{ij} \lessgtr n_{ik}$ or $N_{ij} \gtrless N_{ik}$). A simple horizontal or vertical normalization of the FSH matrix for each row or column between 0 and 1 delivers a ranking similar to n_{ij} or N_{ij} measures. For example, the match between a male accountant and a female historian is ranked at 0.302 in terms of its assortativity among all possible matches available for a male accountant with other different major holders. The middle part of Table 2 shows this ranking based on the horizontal normalization of FSH measures reported in the top part. This process now allows us approximate the attractiveness of each field of study for men in mating. The same match is ranked at 0.535 among those available for a female historian reported in the bottom section of the table. Similarly, the vertical normalization reveals the relative attractiveness of other majors in women's eyes when they are compared to the most preferred one in mating. These indices simply order each partner's appeal by his/her field of study and do not impose cardinal restrictions.

The only limit in the data is the requirement of holding a degree in education that grants a major for both partners. It is obvious from the diagonal of both the middle and bottom sections of the table that the evidence supports a strong field-of-study homogamy. Although not reported here, FSH becomes even stronger when we use 41 CIP codes.[19] We use normalized FSH (NFSH) measures in our analysis to understand the differences in maternal and paternal effects on children's choice of major.[20]

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) reports that this wage penalty varies by each field of study in the range of 16 percent for engineers and 5.7 percent for degree holders in the Humanities. This fact underlines why the occupational relatedness of parents' major could play an important role in children's choice of major. When parents do not practice their profession, even if they earn a higher income, the attraction of their trained profession might become diluted in the eyes of their children. Children

[19] We apply the H-index (explained in the next section) to compare the magnitude of FSH in each match matrix calculated based on 12 and 41 major CIP codes. The results indicate a slightly increasing FSH as we use more detailed CIP codes.

[20] It would be very informative to expand this descriptive analysis of FSH further, as it is the first in Canada; however, it is beyond the scope of this paper.

would be deterred from their parents' major even more if a wage penalty is associated with their parents' educational mismatch, which is an important topic and has been extensively investigated in the literature (Aydede and Dar 2017). 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 example, 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):

$$FOR_{of} = \frac{L_{of}/L_f}{L_o/L_T},$$

where L is the number of workers, o is the occupation, f is the field of study and T denotes the whole workforce. This index measures the relatedness of occupation o in major f by calculating the percentage of workers in major f working in occupation o adjusted by the size of occupation o in the entire workforce. The 2011 NHS occupation data are classified according to the National Occupational Classification (NOC–2011), which is composed of four levels of aggregation. At the first 3 levels, there are 10 broad occupational categories containing 40 major groups that are further subdivided into 140 minor groups. At the most detailed level there are 500 occupation unit groups. Statistics Canada defines this classification as occupation unit groups that are formed on the basis of the education, training, or skill level required to enter the job, as well as the kind of work performed, as determined by the tasks, duties and responsibilities of the occupation.

Given the large sample at our disposal, we use the frequency distribution of 41 fields of study across 40 occupations, which gives us 1640 cells to calculate FOR. For each of the 41 fields of study, when we normalize FOR between 1 and 0 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 normalized FOR in two class intervals (1-0.8 and 0.8-0) 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 and 0.8 as relatively better matching occupations, we see that 32 percent of husbands work in related occupations, with the same ratio slightly lower for wives. As expected, the ratio varies across majors from 10 percent for wives in humanities to 57 percent 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 FSA for each child-parent pair by 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

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 last major, 12 Other, is ignored in the table as it includes very few observations. (3) The sample used in this table contains all working spouses regardless of whether they have children with or without an identified CIP code.

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, the average FSA is much higher (0.466 and 0.469) when the father's NFOR is 1 and not affected by mother's field-of-study relatedness. This observation recurring in each column implies that 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 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 where 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

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

Notes: (1) See the notes to Table 1 for the full description of majors. (2) the last major, 12 Other, is ignored in the table as it includes very few observations. (3) The sample used in this table contains all working spouses regardless of whether they have children with or without an identified CIP code.

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:

$$S^c = \alpha_0 + \alpha_1 S^p + \alpha_2 h^p + \alpha_3 f^p + e^c. \tag{8}$$

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, the bias in α_1 estimated by least-squares (OLS) can be expressed as follows:

$$\text{plim } \widehat{\alpha_{1ols}} = \alpha_1 + \alpha_2 \frac{\text{cov}(S^p, h^p)}{\text{var}(S^p)} + \alpha_3 \frac{\text{cov}(S^p, f^p)}{\text{var}(S^p)}.$$

With educational outcomes, such as schooling years, observed from samples of parents and their own birth children, a direct estimation of (8) cannot identify α_1, unless one assumes that unobserved endowments, h^p and f^p, are unrelated to S^p.[21] Hence, an estimation of (8) without controlling ability sorting and better parenting reveals the intergenerational elasticity between parent-child years of schooling,

[21] Homlund 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 IV models) and they could also encounter problems in obtaining bias free estimates of causal intergenerational coefficients.

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 intergenerational elasticity of schooling is a fundamental metric that has been used to measure the mobility across generations.[22]

Inspired from this literature, we propose a different identification strategy and start with four reduced-form non-parametric 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.

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

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

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

$$FSA_{M,D} = \theta_0 + \theta_1 NFSH^F + \theta_2 NFOR^M + \theta_3 h^M + \theta_4 f^M + \theta_5 h^F + \theta_6 f^F + \eta^D, \tag{12}$$

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 child's assortative tendencies observed in his choice of major is related to the field-of-study homogamy and the degree of relatedness between each parent's field of study and occupation within a family.[23] 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 $NFSH^M$ in (9), for example, is bounded between 0 and 1. It reflects a perfect homogamy as it approaches to 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 a more reliable transmitter 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 four equations

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

[23] Given the parent's major, 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 from (9) to (12).

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 two matches for each child, when we take the difference between them, the dependent variables in these matching functions better reflect assortative tendency because the omitted heterogeneity across children are differenced out from the equations as shown below.

$$FSA_{M,S} - FSA_{F,S} = \omega_0 + \omega_1 NFSH^F - \omega_2 NFSH^M + \omega_3 NFOR^M - \omega_4 NFOR^F + \tau^c, \quad (13)$$

$$FSA_{M,D} - FSA_{F,D} = \sigma_0 + \sigma_1 NFSH^F - \sigma_2 NFSH^M + \sigma_3 NFOR^M - \sigma_4 NFOR^F + \vartheta^c. \quad (14)$$

A similar non-parametric 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 at two 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).[24] 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 between $FSA_{M,S}$ and $FSA_{F,S}$, for example, should be smaller when $NFSH^M$ decreases, holding other covariates constant. Intuitively, if the mother married to an accountant holds a degree in biochemistry, $NFSH^M$ approaches to its lower limit.[25] As the mother becomes another channel of information on an alternative major, 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 towards his father's major, accounting. That is why $FSA_{F,S}$ (the son's attraction to his father's major) should be smaller when $NFSH^M$ (resemblance of the mother's major to her husband's, measured by spousal differences in the appeal of their majors) gets lower. Hence, the differences in ω_1 and ω_2 as well as σ_1 and σ_2 will provide information about the difference in transmission between fathers and mothers. However, the value (and the volume) of the available information provided

[24] Models based on many-to-one matches are not new and 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 an attenuation bias but may inflate the standard errors of the estimates.

[25] This statement is justified based on a strong field-of-study homogamy reported in Section 3.1.

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 is 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 his mother. Since the parents would be a better channel of information conditional on the quality of their occupational match, an increasing $NFOR^M$ in (13) should have both a negative impact on $FSA_{F,S}$ and a positive effect on $FSA_{M,S}$. Hence, a positive and significant coefficient of $NFOR^M$ indicates the existence of a transmission of field of study reinforced by expanding the reliable information within the family.

As outlined before, 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 specification, equation (13) for example, can address this identification problem conditional on the assumption that (β_3-α_3), (β_4-α_4), (β_5-α_5), and (β_6-α_6) are not statistically significant. Without this assumption and excluding NFOR for now, equation (13) can be expressed as follows:

$$\Delta FSA_S^{M,F} = \omega_0 + \omega_1 NFSH^F - \omega_2 NFSH^M + \omega_3 h^M + \omega_4 f^M + \omega_5 h^F + \omega_6 f^F + \tau^c, \quad (15)$$

where ω_3= (β_3-α_3), ω_4= (β_4-α_4), ω_5= (β_5-α_5), and ω_6= (β_6-α_6). When estimated by OLS, identification of ω_2 (ω_1) requires either that $NFSH^M$ ($NFSH^F$) is independent of unobserved parental traits or that ω_3, ω_4, ω_5, and ω_6 are zero, as shown below.

$$\text{plim}\ \widehat{\omega_{2ols}} = \omega_2 + \omega_3 \frac{\text{cov}\left(NFSH^M, h^M\right)}{\text{var}\left(NFSH^M\right)} + \omega_4 \frac{\text{cov}\left(NFSH^M, f^M\right)}{\text{var}\left(NFSH^M\right)}$$
$$+\omega_5 \frac{\text{cov}(NFSH^M, h^F)}{\text{var}(NFSH^M)} + \omega_6 \frac{\text{cov}(NFSH^M, f^F)}{\text{var}(NFSH^M)}.$$

First, we think that parents' child rearing skills, f^M and f^F should not be significantly correlated with homogamy measures, $NFSH^M$ and $NFSH^F$. 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, 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 cov($NFSH^M$, h^F) and cov($NFSH^M$, h^M).

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

	1				**2**			
	FSA (Son)		**FSA (Daughter)**		**FSA (Son)**		**FSA (Daughter)**	
	Father	**Mother**	**Father**	**Mother**	**Father**	**Mother**	**Father**	**Mother**
$FSH = 1$ if same major					0.055	0.002	0.025	0.015
					0.000	0.771	0.000	0.033
$NFSH^M$	0.050		0.051					
	0.000		0.000					
$NFSH^F$		0.100		0.038				
		0.000		0.000				
$FORC^F = 1$					-0.044		-0.027	
					0.000		0.000	
$FORC^M = 1$						-0.020		-0.016
						0.000		0.001
EDH	0.046	0.027	0.018	0.025	0.040	0.045	0.025	0.026
	0.000	0.000	0.013	0.002	0.000	0.000	0.001	0.002
HH Income	0.009	0.002	0.006	0.001	0.007	-0.002	0.005	0.001
	0.000	0.758	0.000	0.210	0.000	0.788	0.000	0.442
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 are adjusted by using the 2-way clustering method (Cameron et al. 2011) at individual and household levels. (3) The numbers under the coefficients report P>|t|. (4) 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. (5) We also ran the regressions with and without the parental age variables. The results are insensitive to the inclusion of parental age variables. (6) When we control for field-of-study fixed effects the results do not change significantly.

If we assume that h represents heritable mathematical skills, for example, a higher $NFSH^M$ could be related to a higher and a lower h^F (or h^M) at the same time.[26] To test this ambiguity, we can use matches where both spouses have at least a university degree with one of the STEM majors (science, technology, engineering, and math). 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 cov($NFSH^M$, h^F) and cov($NFSH^M$, h^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 with one of the STEM majors, 0 otherwise — into (13), would help us address a possible bias in the transmission coefficients.

$$
\begin{aligned}
FSA_{M,S} - FSA_{F,S} = \omega_0 + \omega_1 NFSH^F - \omega_2 NFSH^M + \omega_3 STEM \\
+\omega_4 STEM \text{ x } NFSH^F - \omega_5 STEM \text{ x } NFSH^M + \omega_6 NFOR^M \\
-\omega_7 NFOR^F + \omega_8 h^M + \omega_9 f^M + \omega_{10} h^F + \omega_{11} f^F + \tau^c.
\end{aligned} \tag{16}
$$

The coefficients of interaction terms will reveal the differences in the sons' assortative preferences in STEM families.

With the within-family specification, two factors will shrink the bias on these coefficients: first, the differential effects of unobservables, ω_8= (β_3-α_3) and ω_{10}= (β_5-α_5), in (16), as opposed to their levels in specifications (8) - (11), will diminish in their size; and second, cov($STEM$ x $NFSH^M$, h^F) and cov($STEM$ x $NFSH^M$, h^M) will be close to zero for a subsample as specified by (16). The definition of the bias in the OLS estimate of ω_5 in specification (16), for example, can be expressed as follows:

$$
\begin{aligned}
\text{plim } \widehat{\omega_{5ols}} = \omega_5 + (\beta_3 - \alpha_3) \frac{\text{cov}\left(STEM \text{ x } NFSH^M, h^M\right)}{\text{var}\left(STEM \text{ x } NFSH^M\right)} \\
+ (\beta_5 - \alpha_5) \frac{\text{cov}\left(STEM \text{ x } NFSH^M, h^F\right)}{\text{var}\left(STEM \text{ x } NFSH^M\right)}.
\end{aligned}
$$

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

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

4 Estimation results

We start with the four equations from (8) to (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.[27] After these additions, Table 5 reports two sets of estimation results for selected variables.[28] The first four columns report the estimation results that include NFSH for each parent without accounting for parents' occupational relatedness. We control for FSH in the last four columns as a binary variable — 1 if both parents have the same field of study, 0 otherwise — and add FOR for both father and mother as a categorical variable, FORC, that is one if the normalized FOR is less than 0.2 and 0 otherwise. The first four 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 two columns show that the son's attraction to his parents' majors is strongly related to 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.

In the last four 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 four 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

[27]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—less than 1 year, college—1 to 2 years, college—more than 2 years, university—below bachelor's, bachelor's, above bachelor's—less than master's, medicine-dentistry-veterinary, master's, PhD.

[28]Since our specifications have fractional response variables that have values ranging between 0 and 1, 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 with the use of the binomial distribution (Papke and Wooldridge 1993). Since the results are almost the same, we report here only the linear specifications estimated by OLS.

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 four 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 a higher FSA. To test the robustness of the results in Table 5, we also used an alternative measure, CS (see Section 3.1), and recalculated the variables related to FSH. The estimations of the same specifications in Table 5 indicate that using the CS method would not make a substantial difference in the results. Furthermore, we also used different levels of the CIP and occupation classifications available in the 2011 NHS. Again, the results are not sensitive to using larger or smaller dimensions of match tables.

We address the identification problem stated in the previous section in Table 6. The first two 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 (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, that reflects the difference in NFOR in three 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 and 0.2 for both parents) or in unrelated jobs (e.g. NFOR is between 0.2 and 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 two categories.[29]

The results are interesting and in line with the findings in our earlier estimations: the coefficient of $NFSH^F$ in the first column, 0.1531, confirms that the distance between $FSA_{M,S}$ and $FSA_{F,S}$ 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

[29]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, this effect can only be captured when parents' FOR is different.

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

	$(\mathbf{FSA}_{MS})$ - $(\mathbf{FSA}_{FS})$		$(\mathbf{FSA}_{MD})$ - $(\mathbf{FSA}_{FD})$	
	1	**2**	**1**	**2**
$(NFSH^F)$ - $(NFSH^M)$	0.1531		-0.0143	
	0.0000		0.3050	
$DFORC$				
1	Base		Base	
2	-0.0333		-0.0271	
	0.0000		0.0000	
3	0.0350		0.1630	
	0.0000		0.0340	
$NFSH^M$		-0.1008		-0.0049
		0.0000		0.7390
$NFSH^F$		0.1624		-0.0210
		0.0000		0.1430
$FORC^M = \mathbf{1}$		-0.0164		-0.0162
		0.0000		0.0100
$FORC^F = \mathbf{1}$		0.0505		0.0276
		0.0000		0.0000
Constant	0.0686	0.0259	-0.0756	-0.0695
	0.0000	0.0000	0.0000	0.0000
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 are adjusted by using the 2-way clustering method (Cameron et al. 2011) at individual and household levels. (3) The numbers under the coefficients report P>|t|.

down, or both occur simultaneously. The second column based on equation (13) 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 $FSA_{F,S}$ it reduces the distance between $FSA_{M,S}$ and $FSA_{F,S}$. Similarly, because a rising in $NFSH^F$ increases $FSA_{M,S}$, the distance between $FSA_{M,S}$ and $FSA_{F,S}$ becomes larger. More importantly, though, the difference between these effects (0.1008 and 0.1624) again suggests evidence that paternal influence is noticeably greater than maternal influence for sons. The same comparison for daughters in both specifications of (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 the distance between $FSA_{M,S}$ and $FSA_{F,S}$ by reducing $FSA_{M,S}$ and increasing $FSA_{F,S}$. Similarly, a significant positive effect is observed for the second category where 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 (-0.0164) indicates that $FSA_{M,S}$ falls. Yet, when the father faces an educational mismatch in his job, the effect on the distance between $FSA_{M,S}$ and $FSA_{F,S}$ captured by a positive coefficient (0.0505) becomes much greater. Interestingly, despite the insignificant effects of NFSH in the third and fourth columns, the significant effects of FORC are observed for daughters, which also indicates the importance of parents' occupational matching in transmission.

With within-family differencing as specified by (13) and (14), the other factors, such as the effects of siblings, neighborhoods, and peers, observed or unobserved, are differenced out in estimations. Hence the results deliver better evidence about the role of information constraint in children's assortative preferences. When the mother's major gets similar to the father's, $NFSH^M$ rises. As a higher homogamy implies more constraint in the family in terms of available information on other majors, the son's bias towards his father's field of study rises. This is confirmed by the negative sign of the $NFSH^M$ coefficient: $NFSH^M$ has a negative effect on the difference between $FSA_{M,S}$ and $FSA_{F,S}$ by increasing $FSA_{F,S}$. Equally, when $NFSH^F$ rises, the similarity between parents' major gets higher. Confined with less available information on other majors, the son's attraction to his mother's field of study rises. This is verified by the positive sign of the $NFSH^F$ coefficient: $NFSH^F$ has a positive effect on the difference between $FSA_{M,S}$ and $FSA_{F,S}$ by increasing $FSA_{M,S}$.

As outlined before, conditional on the extent to which the field-of-study homogamy is driven by the ability sorting in parents' marriage, within-family differencing as specified by (13) and (14) may still have a possible bias in transmission coefficients. One way to address this problem is to use a subsample that includes

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

	$(\mathbf{FSA}_{MS})$ - $(\mathbf{FSA}_{FS})$		$(\mathbf{FSA}_{MD})$ - $(\mathbf{FSA}_{FD})$	
	Coef	P>|t|	Coef	P>|t|
$NFSH^M$	-0.0787	0.0000	0.0009	0.9510
$NFSH^F$	0.1609	0.0000	-0.0200	0.1650
$STEM$	-0.1506	0.0000	0.0191	0.6440
$STEM$ **x** $NFSH^M$	-0.0419	0.6140	0.0368	0.7340
$STEM$ **x** $NFSH^F$	-0.0372	0.6950	-0.1025	0.4110
$FORC^M$ $\mathbf{= 1}$	-0.0172	0.0020	-0.0162	0.0100
$FORC^F$ $\mathbf{= 1}$	0.0418	0.0000	0.0268	0.0000
Constant	0.0268	0.0000	-0.0733	0.0000
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 2-way clustering method (Cameron et al. 2011) at individual and household levels.

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 (16). 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 $NFSH^F$ (-0.0787 and 0.1609) 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 less than 20 percent in the whole sample. The insignificant interaction terms indicate that the 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 $FSA_{M,S}$ and $FSA_{F,S}$ is lower for STEM families than 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 about the link between parents' homogamy and their assortative preferences in their choice of major.

The total elasticity of sons' assortative preferences in terms of parental homogamy can be expressed for sons by the sum of coefficients of $NFSH^M$ and $NFSH^F$, which is 0.2396 (0.0787 + 0.1609). This measure suggests an important role of information asymmetry in children's choice of major to the extent that the field-of-study homogamy 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. 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 field-of-study homogamy would not necessarily suggest a positive bias in children's choice towards their parent's majors. Although it is less likely, two accountant parents would not necessarily be in favor of their majors and may deter their children from their own majors. This possibility would also create a downward bias in our estimations. Finally, as it 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. However, even with the bias, 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.

5 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 pursue as well. This study quantifies assortative mating by estimating field-of-study homogamy and intergenerational transmission of skills by measuring assortative preferences in 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 the social and economic progress. This study's primary objective is to investigate the 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 where 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 National Household Survey 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 three indicators: the degree of children's attraction to their parents' field of study (FSA), the degree of field-of-study homogamy (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 one-to-two matches (one 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.

Acknowledgement

The analysis presented in this paper was conducted at the Atlantic Research Data Centre, which is part of the Canadian Research Data Centre Network (CRDCN). The services and activities provided by the Atlantic Research Data Centre are made possible by the financial or in-kind support of the SSHRC, the CIHR, the CFI, Statistics Canada, and Dalhousie University. The views expressed in this paper do not necessarily represent those of the CRDCN or those of its partners. We thank our research assistant James Stutely for his help, Heather Hobson of ARDC for her support during our work at the center and, Mutlu Yuksel (Dalhousie University) for his comments on an earlier draft of our study. Responsibility for any shortcomings remains with us.

References

[1] Acemoglu, D. and Autor, D. 2011. "Skills, tasks and technologies: implications for employment and earnings," *Handbook of Labor Economcis*, 4: 1043-1171.

[2] Altonji, J. G., Blom, E. and Meghir, C. 2012. "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, Annual Reviews, 4(1): 185-223.

[3] Altonji, J. G., Arcidiacono, P. and Maurel, A. 2015. "The analysis of field of study and graduate school: determinants and wage effects." NBER Working Paper: 21655.

[4] Aydede, Y. and Dar, A. 2016. "The cost of immigrants' occupational mismatch and the effectiveness of postarrival policies in Canada," *IZA Journal of Migration,* 5(9): 1-23.

[5] Aydede, Y. and Dar, A. 2017. "Is the lower return to immigrants' foreign education a postarrival problem in Canada," *IZA Journal of Migration*, 6(4): 1-25.

[6] Aydemir, A., Chen, W. and Corak, M. 2013. "Intergenerational education mobility among the children of Canadian immigrants." *Canadian Public Policy*, 39: S108-S122.

[7] Becker, G. S. 1973. "A theory of marriage: Part I." *Journal of Political Economy*, 81: 813-846.

[8] Becker, G. S. 1974. "A theory of marriage: Part II." *Journal of Political Economy*, 82:S11-S26.

[9] Becker, G.S. and Tomes, N. 1979. "An equilibrium theory of the distribution income and intergenerational mobility." *Journal of Political Economy.* 87(6): 1153-89.

[10] Becker, G.S., Kominers, S.D., Murphy, K.M. and Spenkuch, J.L. 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.

[11] Bicakova, A. and Jurajda, S. 2016. "Field-of-study homogamy." IZA Discussion Papers: 9844.

[12] Black, S. and Devereux, P.J. 2010. "Recent developments in intergenerational mobility." NBER Working Paper 15889.

[13] Boudarbat, B. and Montmarquette, C. 2009. "Choice of fields of study of Canadian university graduates: the role of gender and their parents' education." *Education Economics,* 17(2): 185-213.

[14] Boudarbat, B. and Chernoff, V. 2012. "Education-job match among recent Canadian university graduates." *Applied Economics Letters,* 19(18): 1923-1926.

[15] Cameron, A.C., Gelbach, J, and Miller, D. 2011. "Robust inference with multi-way clustering" *Journal of Business and Economic Statistics*, 29(2): 238-249.

[16] Carneiro, P., Maghir, C. and Parey, M. 2013. “Maternal education, home environments, and the development of children and adolescents.” *Journal of the European Economic Association*, 11(1): 123-160.

[17] Chiappori, P.A., Salanie, B., Tillman, A., and Weiss, Y. 2008. “Assortative matching on the marriage market: a structural investigation.” Mimeo, Columbia University.http://adres.ens.fr/IMG/pdf/09022009.pdf.

[18] Chiappori, P.A. and Oreffice, S. and Quintana-Domeque, C. 2012. “Fatter attraction: anthropometrics and socioeconomic matching on the marriage market.” *Journal of Political Economy*, 120(4): 659-695.

[19] Chiappori, P.A. and Salanie, B. 2016 “Econometrics of matching models.” *Journal of Economic Literature*, 54(3): 832-861.

[20] Choo, E. and Siow, A. 2006. “Estimating a marriage matching model with spillover effects.” *Demography*, 43(3): 463-490.

[21] 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.

[22] Diamond, W. and Agarwal, N. 2016. “Latent indices in assortative matching models.” http://economics.mit.edu/files/11242.

[23] Eika, L., Mogstad, M. and Zafar, B. 2017. “Educational assortative mating and household income inequality.” Federal Reserve Bank of New York Staff Reports: 682.

[24] Goodman, L.A. 2007. “Statistical magic and/or statistical serendipity: an age of progress in the analysis of categorical data.” *Annual Review of Sociology*, 33: 1-19.

[25] Haegland, T., Lars, K., Oddbjorn, R. and Kjell G., S. 2010. “Why Children of College Graduates Outperform their Schoolmates: A Study of Cousins and Adoptees.” IZA Discussion Papers 5369.

[26] Handy, C. 2015. “Assortative mating and intergenerational persistence of schooling and earnings.” The 21^{st} SOLE 2016 Annual Meeting - Seattle USA: http://www.sole-jole.org/16095.pdf.

[27] Hilger, N.G. 2016. “The great escape: intergenerational mobility in the United States since 1940.” NBER Working papers: 21217.

[28] Holmlund, H., Lindahl, M. and Plug, E. 2011. “The causal effect of parents’ schooling on children’s schooling: a comparison of estimation methods.” *Journal of Economic Literature*, 49(3): 615-651.

[29] Hugh-Jones, D., Verweij, K. J. H., St.Pourcain, B. and Abdellaoui, A. 2016. "Assortative mating on educational attainment leads to genetic spousal resemblance for causal alleles." *Intelligence*, 1-6.

[30] Lemieux, T. 2014. "Occupations, fields of study, and returns to education." *Canadian Journal of Economics*. 47(4): 1047-1077

[31] Mare, R.D. 2016. "Educational assortative mating in two gilded ages: Evidence from inter-generational social mobility data." *The ANNALS of the American Academy of Political and Social Science*, 663: 117-139.

[32] Martin, N. 2018. "Getting to the genetic and environmental roots fo educational inequality." *NPJ Science of Learning*, 3(4): 1-33.

[33] Maurin, E. and McNally, S. 2008. "Vive la revolution! Long-term educational returns of 1968 to the angry students." *Journal of Labor Economics*, 26(1): 1-33.

[34] McIntosh, J. 2010. "Educational mobility in Canada: results from the 2001 general social survey." *Empirical Economics*, 38: 457-470.

[35] Nielsen, H.S. and Svarer, M. 2006. "Educational homogamy: preferences or opportunities?" IZA Discussion Paper: 2271.

[36] Nordin, M., Persson, I. and Rooth, D. 2010. "Education-occupation mismatch: Is there an income penalty?" *Economics of Education Review*, 29(6): 1047-1059.

[37] Oreopoulos, P., Page, M.E. and Stevens, A.H. 2006. "The intergenerational effects of compulsory schooling." *Journal of Labor Economics*, 24(4): 729-760.

[38] Papke, L. E. and Wooldrige, J. M. 1996. "Econometric methods for fractional response variables with an application to 401(k) plan participation rates." *Journal of Applied Econometrics*, 11: 619-632.

[39] Plomin, R. and Deary, I. J. 2015. 'Genetics and intelligence differences: five special findings." *Molecular Psychiatry*, 20: 98-108.

[40] Robst, J. 2007. "Education and job match: the relatedness of college major and work." *Economics of Education Review*, 26: 397-407.

[41] Roth, A.E. and Sotomayor, M.A.O. 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/~alroth/papers/92_HGT_Two-SidedMatching.pdf

[42] Schwartz, C.R. 2013. "Trends and variations in assortative mating: causes and consequences." *Annual Review of Sociology*, 39:451-470.

[43] Schwartz, C.R. and Mare, R. D. 2005. “Trends in educational assortative mating from 1940 to 2003.” *Demography*, 42(4):621-646.

[44] Shannon, C.E. 1948. “A mathematical theory of communication.” *The Bell System Technical Journal*, 27:379-423.

[45] Siow, A. 2015. “Testing Becker’s theory of positive assortative mating.” *Journal of Labor Economics*, 33(2): 409-441.

[46] Solon, G. 2013. “Theoretical models of inequality transmission across multiple generations.” NBER Working papers: 18790.

[47] Stinebrickner, R. and Stinebrickner, T. 2014. “A major in science? Initial beliefs and final outcomes for college major and dropout.” Review of Economic Studies, 81(1): 426-472.

[48] Turcotte, M. 2011. “Intergenerational education mobility: University completion in relation to parents’ education level.” Statistics Canada, Canadian Social Trends. Catalogue no: 11-008-X.

[49] Wiswall, M. and Zafar, B. 2015. “Determinants of college major choice: identification using an information experiment.” *Review of Economic Studies*, 82(2): 791-824.

[50] Western, B., Bloome, D. and Percheski, C. 2008. “Inequality among American families with children, 1975 to 2005.” *American Sociological Review*, 73(6): 903-920.

[51] Zafar, B. 2012. “Double majors: one for me, one for the parents?” *Economic Inquiry,* 50(2): 287-308.

[52] Zafar, B. 2013. “College major choice and the gender gap.” *Journal of Human Resources*, 48(3): 545-595.