Courant Research Centre
'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 248

Gender Segregation in Education and Its Implications for Labour Market Outcomes: Evidence from India

Soham Sahoo and Stephan Klasen

June 2018

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Gender Segregation in Education and Its Implications for Labour Market Outcomes: Evidence from India

Soham Sahoo[1] and Stephan Klasen[2]

June 2018

Abstract

This paper investigates gender-based segregation across different fields of study at the post-secondary level of schooling, and how that affects subsequent labour market outcomes of men and women. Using a nationally representative longitudinal data-set from India, we provide evidence that there is substantial intra-household gender disparity in the choice of study stream at the higher-secondary level of education. A household fixed effects regression shows that girls are 20 percentage points less likely than boys to study in technical streams, namely science (STEM) and commerce, vis-à-vis arts or humanities. This gender disparity is not driven by gender specific differences in mathematical ability, as the gap remains large and significant even after controlling for individuals' past test scores. Our further analysis on working-age individuals suggests that technical stream choice at higher-secondary level significantly affects the gender gap in labour market outcomes in adult life, including labour force participation, nature of employment, and earnings. Thus our findings reveal how gender disparity in economic outcomes at a later stage in the life-course is affected by gendered trajectories set earlier in life, especially at the school level.

Keywords: Post-secondary education; STEM; Gender; Labour market; India

JEL Codes: I20; J16; J24

[1] Indian Institute of Management Bangalore. Email: soham.sahoo@iimb.ac.in

[2] University of Göttingen, Germany. Email: sklasen@uni-goettingen.de

Acknowledgements: We would like to thank participants of GrOW Workshop 2016 at Stellenbosch University, Contemporary Issues in Development Economics Conference 2016 at Jadavpur University, PEGNet Conference 2017 at ETH Zurich, and GREThA International Conference on Economic Development 2018 at University of Bordeaux for helpful comments. We gratefully acknowledge funding from the Growth and Economic Opportunities for Women (GrOW) initiative, a multi-funder partnership between the UK's Department for International Development, the Hewlett Foundation and the International Development Research Centre.

1 Introduction

There are various forms of gender inequality observed in many parts of the developing world. The motivation behind this study stems from two recent and pressing issues of gender disparity in many developing countries, including India. First, female labour force participation which is viewed as one of the important indicators of inclusive development, has remained very low, stagnant and sometimes declining in India, despite the nation's rapid economic growth, female educational expansion, and fertility decline in the last two decades (Klasen and Pieters, 2015). Second, occupational and sectoral segregation of employment by gender is remarkably persistent and is a key issue behind perpetuating female disadvantage, such as the gender pay gap in the labour market (Borrowman and Klasen, 2017). It is possible that gender gap in economic participation in adult life is determined by gendered trajectories set earlier in life, especially at the school level. Against this backdrop, this study seeks to answer the following two questions: (a) Is there educational segregation prevailing at the school level? Specifically, we want to identify the gender gap in stream choice at the post-secondary level of education, and (b) Do the gender differences at the school level link to labour market outcomes later in life? To reflect on this question, we investigate how post-secondary stream choice affects adult life employment and earnings.

There exists a considerable literature on post-secondary stream choice, its determinants and its implications (Fuller et al., 1982; Beffy et al., 2012). Many studies also recognize that educational segregation and occupational segregation go hand in hand. For instance, Schneeweis and Zweimüller (2012), in the context of Austria, argues that gender segregation in employment is explained by women's reluctance to choose technical occupations. However, they also show that the foundations for career choices are laid much earlier. Most of the studies in this literature are based on developed countries, with only a few exceptions such as Sookram and Strobl (2009) who investigate the relationship between educational and occupational segregation for Trinidad and Tobago.

We seek to contribute to this literature by investigating this issue in the Indian context. Over the last few decades, the Indian government has been able to provide better access to schooling to the population. This has resulted in narrowing of the gender gap in overall school enrollment and completion rates. However, gender disparity has manifested in terms of some more nuanced indicators, such as a female disadvantage in private school choice (Maitra et al., 2016; Sahoo, 2017). In this study we focus on the choice of study stream at the higher-secondary level of education. After secondary schooling which finishes at age 15, students are given the option of choosing from arts (humanities), commerce, science, engineering/vocational, and other streams for higher-secondary education (which lasts for another 2 years). This is a crucial juncture in their career because once they make the stream choice, most of them continue to pursue studying subjects in that particular stream subsequently in college and university.[3] This choice also influences the nature of jobs that they may obtain in the future. We especially focus on the choice between technical (science/commerce/engineering/vocation) versus non-technical (arts/humanities) streams.[4]

We use a nationally representative household level panel data that track same households and individual members at two time points: 2005 and 2012. The novelty of this dataset is that it asks all individuals about their performance in the secondary school leaving certificate (SSLC) examination, and subsequently what stream they studied in higher-secondary level. Additionally, individuals aged 15-18 years (which is the official age for higher-secondary schooling) in 2012 can be matched with information on their prior skills in mathematics, reading and writing from an independent test that was conducted in the earlier round of survey in 2005. Thus, we have a unique setting to investigate higher-secondary stream choice of individuals after controlling for their past academic performances that serve as reasonable proxies of their cognitive ability.

We estimate a household fixed effects model to take into account unobserved household-specific tastes and preferences. Our estimates show a significant intra-household gender disparity of around 20 percentage points in the choice of technical stream at the higher-secondary level among youth aged 15-18 years. The gender gap remains unchanged even after controlling for SSLC exam performance and lagged test scores from the previous

---

[3] Estimates from the data we use suggest that 93 (85) percent of students who are currently studying engineering (science) in college have studied a technical/STEM stream at the higher-secondary level. 85 percent of students who are studying arts in college have studied arts in higher-secondary as well.

[4] Our categorization of subjects is closely related to STEM versus non-STEM subject groups. While traditionally STEM includes science, technology, engineering and mathematics, recently subjects such as accountancy or finance from the commerce stream are also considered close to STEM subjects as they involve mathematical tools. Another reason for considering commerce in the "technical stream" category is that many students in the commerce stream study mathematics and statistics. In later analysis, we also consider each subject separately in our analysis.

survey, suggesting that it is not caused by any difference in cognitive ability between boys and girls.

While we find substantial gender segregation in school level stream choice, does this affect economic outcomes later in life? We reflect on this question by analyzing the relationship between higher-secondary stream choice and labour market outcomes of working-age individuals. Here also we include individual's performance in the SSLC exam as a proxy for innate ability – omission of which may result in biased estimates if individuals who choose technical stream are also better performers in the labour market because they have higher ability. We find that women have higher chances of participating in the labour force, getting salaried employment, choosing a male-dominated occupation, and having higher earnings when they have studied a technical stream in higher-secondary education. This particularly leads to a reduction of gender gap within household in terms of all these economic outcomes. Exploring this further, we find this relationship between education choice and labour market outcomes to be more prominent in urban areas. Also, among the technical subjects, science appears to have the most significant effect. We also show that the results remain robust to inclusion of additional control variables that partially capture women's personality traits.

Our study contributes to the literature in several ways. First, the importance of stream choice at the school level is an under-researched area especially in the context of developing countries. Second, in such settings it is rare to have information on school level stream choice, adult-life outcomes, and especially measures of ability on the same individuals. Having all three types of information from the same dataset allows us to investigate gender segregation in education as well as its connection to labour market outcomes, while mitigating the concern of omitted ability bias. Thus, our study is the first to credibly quantify the extent of gender gap in post-secondary stream choice in India. Further, it provides evidence that gender segregation in education determines occupational segregation and earnings gap in the labour market. This study is of significant policy relevance as it emphasizes the need for promoting gender equality of opportunities at various stages of life. The issue of stream choice becomes increasingly important with rising enrollment rates of girls in secondary and post-secondary education. To achieve greater parity in labour market outcomes, policies need to focus on girls' access to technical education that can help them build human capital at par with boys. Such policies will be effective in enhancing women's participation in the labour market and in appropriate jobs, leading to efficient utilization of human resources and helping the economy derive greater returns from the “demographic dividend” associated with falling fertility and a low dependency ratio.

The remaining of the paper is organized as follows. In section 2, we discuss the background and related literature. We explain the dataset and summary statistics in section 3. Section 4 contains the analysis of gender disparity in stream choice among individuals in higher-secondary age-group. This discussion includes both empirical model and results of the analysis. Section 5 shows how higher-secondary stream choice affects labour market outcomes of men and women. Section 6 concludes.

2 Background and related literature

In the last few decades, there has been considerable progress in terms of bridging the gender gap in educational attainment around the world. At the same time, female labour force participation has increased in most of the developing countries. But, trends have been rather uneven with South Asia actually experiencing declining female labor force participation rates (Klasen, 2017). Moreover, a more nuanced analysis of this trend reveals that while the extent of participation by women in education and labour market has increased, the nature of participation has not changed over time. Using data from 69 countries between 1980 and 2011, Borrowman and Klasen (2017) find that women have continued to be employed predominantly only in few sectors and occupations. Using data from India for approximately the same period, Duraisamy and Duraisamy (2014) find that occupational segregation has increased over time. These studies also find that this perpetuating trend in occupational and sectoral segregation is a major reason behind the existence of male-female earnings gap. Even in the context of US, Blau and Kahn (2017) have found that gender differences in industry and occupation are more important determinants of gender pay gap than conventional measures of human capital.

On the education front, there has been a spectacular rise in girls’ school participation during the era when many countries implemented the “Education for All” program. India has made tremendous progress in narrowing down the gender gap in school enrollment during the last few decades. Figure 1 illustrates the trends in boys’ and girls’ age-specific enrollment rates using multiple rounds of nationally representative data from India. While there was a substantial gender gap in the mid 1980s and 1990s, it has practically disappeared in the recent times. It is particularly noteworthy that this trend is true even at the secondary and higher-secondary levels of education, suggesting that adolescent girls are now as likely as boys to attend school at this level.

The choice of field of study at the higher-secondary level is an important decision in an individual's career because it is the first step towards further specialization at the subsequent levels of education and also for labour market outcomes such as occupational choice. There are various studies, mostly on developed countries, which analyze the determinants of post-secondary field of study choice. The literature also highlights that educational choices at this level and labour market outcomes are closely linked. On one hand, the choice of a particular field of study is found to be affected by the expected future earnings from different fields (Boudarbat, 2008; Beffy et al. 2012). On the other hand, such educational choices also cause much of the variation in earnings later in life (Dustmann, 2004; Joensen and Nielsen, 2009).

Focusing on gender, studies have commonly found the existence of gender disparity in the choice of study-streams – girls are especially under-represented in Science, Technology, Engineering and Management (STEM) at the post-secondary and tertiary levels of education almost all over the world (Hill et al., 2010; Flabbi, 2011). The World Development Report 2012 points out that "the seeds of segregation are planted early" and "gender differences in education trajectories shape employment segregation" (World Bank, 2012, p. 216). Various studies have sought to analyze the extent and causes of gender streaming in education and its relation to occupational segregation. Decomposing the overall gender segregation in the Flemish labour market, Van Puyenbroeck et al. (2012) find that the choice of study field has a larger effect on overall segregation than sectoral choice which itself is partly explained by educational choice. In the context of Bulgaria, Bieri et al. (2016) documents that stream choice at the tertiary level affects men's but not women's choice of male versus female-dominated occupation. Using a sample of OECD countries, Flabbi (2011) finds that gender is a significant determinant of the chosen field of study, that females are less likely than men to choose science and social sciences, and more likely to choose humanities. Further, this choice significantly explains the gender-wage gap and the relationship between field of study and labour market outcomes varies between men and women.

Since girls are less likely to choose STEM subjects, one potential reason that has been explored in the literature is whether boys have a comparative advantage in mathematics. Various studies have found evidence which contradicts this hypothesis. For instance, Friedman-Sokuler and Justman (2016) finds in Israel that gendered choices remain unperturbed even after conditioning on differences in perceived mathematical ability captured by prior achievement. The received wisdom from the literature is that inherent gender difference in cognitive ability is almost non-existent; rather, other societal, psychosocial and preference related factors play a larger role in explaining the under-representation of women in math-intensive STEM subjects (Antecol and Cobb-Clark, 2013; Zafar, 2013; Buser et al., 2014; Kahn and Ginther, 2017).

Most of the evidence on field of study choice is limited to advanced economies. It is perhaps not surprising that very few studies have analyzed this issue in the context of underdeveloped countries which, until recent times, have struggled even to raise enrollment rates. However, with a substantial increase in school participation, now the focus is shifting from quantity to quality aspects of education.[5] Since gender based segregation in stream choice can have important implications for labour market outcomes, therefore, this study seeks to analyze this issue in the context of India.

3 Data and descriptive analysis

We use a nationally representative two period longitudinal dataset called the India Human Development Survey (IHDS).[6] The first round of data was collected in 2004-05 on 41,554 households in 1503 villages and 971 urban neighborhoods across India. In 2011-12, the second round of survey re-interviewed 83 percent of the same households and for those households which could not be tracked, a replacement sample was used. Thus, the second round of survey covered 42,152 households across India. For the purpose of brevity we will refer to the first round as 2005 data and the second round as 2012 data. IHDS is a multi-topic survey consisting of detailed information at the levels of individuals, households, and communities. The analysis in this paper mainly uses the sample from the 2012 survey, and uses the 2005 survey to account for past characteristics of the same individuals.

[5] In the post 2015 agenda of the United Nations, the Sustainable Development Goals aim to ensure free, equitable and quality primary and secondary education for all boys and girls.

[6] IHDS was carried out jointly by the University of Maryland and the National Council of Applied Economic Research, New Delhi. The dataset is publicly available. More details can be found here: https://ihds.umd.edu/

At first, we seek to explore whether there is a gender bias within households in the choice of study stream at the higher-secondary level. In India, the official school entry age is 6 years, and the (lower) secondary level ends at the completion of 10 years of schooling. In the IHDS sample, the enrollment rate of children who are in the secondary age-group (14-15 years) is 87 percent and the gender gap in enrollment rate is only 2 percentage points. The succeeding level after secondary consists of two years of schooling, called "higher-secondary" (or senior-secondary, or upper-secondary) is the level we want to consider here. Therefore, we concentrate on the sample of individuals who are in the age-group of 15-18 corresponding to the higher-secondary level of education.[7] There are 14,845 children in this age-group. Information on stream choice at the higher-secondary level is available only for those individuals who have passed secondary level and enrolled in the subsequent level of education. The secondary pass rate for our sample is around 40 percent; it is 39.4 percent for males and 40.6 percent for females. A t-test reveals that the gender difference in secondary pass rate is statistically not significant. After dropping observations on missing values, the final sample of analysis consists of 5213 children.

The first step towards specialization begins at the higher-secondary level of education where students have to choose a stream mainly from the following options: arts, commerce, science, engineering/vocational, and others (home science, craft, design, etc.). As evident from Figure 2, more than 50 percent of individuals of all age groups who have studied at this level have chosen the arts stream. The next popular streams are commerce and science, followed by engineering/vocational and others which are chosen by very few. This general pattern is similar for both males and females, and has stayed more or less the same over time. Figure 3 depicts the proportion of males and females who have studied each of the streams. Females are always more likely than males to study arts, and less likely to study science, commerce, and engineering/vocational. Consistent with the existing studies on other countries, this crude measure of gender gap in stream choice suggests that girls are under-represented in the STEM subjects also in India.

[7] Strictly speaking, the age band corresponding to higher-secondary level should be 16-17 years. However, we include one year below and above to allow for the possibility that some children may finish secondary level earlier, and some children later. The enrollment rate among children in 16-17 year age is 74 percent, however, many of them are yet to complete secondary level schooling.

Is there a link between post-secondary stream choice and subsequent labour market outcomes? For this analysis we consider adult individuals aged 25-60 years. After dropping missing observations on the relevant variables, we have a sample of 80,302 individuals who are included in the analysis of labour market outcomes. The secondary pass rate in the adult sample is 21 percent, as secondary school participation was much lower in the earlier decades. Among those who have passed secondary education, 39 percent men and 28 percent women opted for technical streams at the higher-secondary level, the remaining mostly studied arts or humanities. The overall labour force participation rate of 60 percent masks a huge gender gap, as only 26 percent women against 96 percent men have participated in the labour force. Among those who participate as salaried employment and casual wage labour, we have information on their annual earnings and the number of hours they worked in the last year. Figure 4 shows the kernel density estimates of log of annual and hourly earnings of adult individuals based on what they studied at the higher-secondary level. The figure indicates that earnings of those who studied a technical stream clearly dominate the earnings of those who studied a non-technical stream. We also find that there is a substantial gender gap in the whole distribution of earnings, as shown by Figure 5. In a later part of our study, we investigate whether higher-secondary stream choice can be connected to these observed differences in the labour market outcomes.

4 Analysis of gender gap in technical stream choice

In this section we lay out an econometric model to investigate if girls are systematically less likely than boys to choose STEM subjects at the higher-secondary level of education. In the following part of the analysis, we also explore if the choice of study stream affects labour market outcomes of individuals, particularly the gender gap in earnings.

4.1 Empirical model to identify intra-household gender gap in stream choice

Our first objective is to identify if there is an intra-household gender disparity in the choice of technical stream versus non-technical stream (arts / humanities). It is important to consider various observable and unobservable factors at the level of individual, household, and community to analyze this issue. Therefore we use the following econometric model:

$$Prob(Techedu_{ih} = 1) = \alpha + \beta Female_{ih} + \gamma X_{ih} + \theta Ability_{ih} + \phi_h + \varepsilon_{ih} \quad (1)$$

We estimate a linear probability model where the dependent variable is a binary indicator of whether an individual aged between 15 and 18 years, corresponding to the higher-secondary level of education, has chosen to study in technical stream ($Techedu = 1$) or arts/humanities ($Techedu = 0$). Essentially, this reflects the choice between STEM (Commerce, Science, Engineering and Vocational) versus arts or other non-technical subjects. The subscript i denotes individual and h denotes household. The main explanatory variable is gender reflected by a dummy variable ($Female$) to denote whether the individual is female. In addition, we control for age, birth order, number of siblings, mother's years of education, and father's years of education.

There are various household level factors that potentially affect the stream choice. If studying technical subjects is more expensive than studying non-technical subjects, then households which are wealthier may have higher likelihood of children being enrolled in such type of education. Other background characteristics, such as having a family member who has studied technical education may also influence the decision of the child to choose this stream. Furthermore, household's taste and preference towards technical education is not observable in the data, but can be important in this choice. Therefore, we control for household level heterogeneity by including household fixed effects (ϕ_h) in the regression.[8] This also enables us to compare between boys and girls in the same household to identify the gender gap. This is especially important in the context of India where household's unobserved preferences are correlated with various forms of gender inequality. Often female children are more likely to be found in larger families because fertility decisions are endogenously determined wherein parents keep having children until they have at least one boy (Yamaguchi, 1989; Clark, 2000; Basu and Jong, 2010). If technical education needs higher investments, then comparisons across households may artificially show a gender gap because girls belong to larger families where investments in the human capital of each child is less. Due to these reasons, existing

[8] Note that inclusion of household fixed effects implies that only those households with at least two individuals contribute to identification in this regression. In comparison with the overall sample, the number of such households, i.e. with at least two individuals in age-group 15–18 years, is much lower at 1085. Therefore, we present results both with and without household fixed effects while showing the gender gap in stream choice.

studies that investigate gender discrimination in educational investments advocate using household fixed effects (Jensen, 2002; Kingdon, 2005; Sahoo, 2017). Using fixed effects also imply that the effects of various observable household characteristics are subsumed by these fixed effects.

Within households, gender difference in the choice of technical education may be driven by the possibility that girls have lower cognitive ability than boys, especially in terms of their performance in mathematics which is an essential subject for technical education. There is an expansive literature that looks at gender gap in mathematics achievement and concludes that most of the observed gap is explained by background factors (Benbow and Stanley, 1980; Nollenberger et al, 2016). In the Indian case, because of systematic and continual under-investment on girls' human capital from early childhood, when girls reach the level of higher-secondary education, they may be lagging behind boys in terms of mathematical ability. A novel feature of our data allows us to account for this problem. In the 2005 IHDS survey, children who were in 8-11 years age-group were given cognitive tests on mathematics, reading, and writing ability. Incidentally, in the 2012 survey, these children are in the right age-group corresponding to higher-secondary level, and considered in the regression. Therefore, we are able to control for their past cognitive ability by including their performance in these tests.[9]

In addition to accounting for past test scores, children's performance in the secondary level board examination is potentially an important predictor of stream choice at the higher-secondary level. In India, there is a standardized examination conducted by the education board (state or national level) to which each school belongs. Every student has to pass this examination and obtain the secondary school leaving certificate (SSLC) to be able to continue higher-secondary levels of education. The results of this examination are typically categorized into divisions 1, 2, and 3, in the declining order of the quality of grade obtained by a child. In

[9] There is a substantial reduction in sample size by about 50 percent when we control for past test scores from the previous round of survey. This is due to many missing values in the variables capturing past test scores (as evident from Table 1 presenting summary statistics). Some individuals (about 11 percent of the sample) could not be found in the 2012 survey, and some other individuals may have misreported their age in the previous survey, leading to missing values in test scores for this age-group. To show that our estimates are not driven by variation in sample size, we present multiple specifications where we add covariates gradually.

addition to prior test scores, we also use this SSLC performance indicator to further control for cognitive ability of the child. Therefore, we control for achievement scores collected by two independent tests: one was conducted by IHDS enumerators in 2005 survey, and the other by the SSLC board. Hence we believe that our regression adequately captures the differences in abilities of children and identifies the gender gap in stream choice.

A potential concern that remains is the fact that stream choice is defined only for those individuals who have passed the secondary level and enrolled in higher-secondary level. Summary statistics presented in Table 1 shows that 40 percent of children passed secondary level. It is likely that these children are systematically different from those who have an education below the secondary level. However, disaggregating this pass rate by boys and girls, we find that there is no gender gap in secondary level pass rate. To further investigate this matter, we run a regression similar to Equation (1) but with the dependent variable being a binary indicator of whether a child has passed the secondary level (hence eligible for higher-secondary stream choice). The result of this regression is presented in Appendix Table A1. We find that the coefficient of gender is almost always insignificant and the magnitude is almost zero. Therefore, it suggests that the probability of selecting into the sample of our main regression (stream choice) does not vary by gender. Hence this is unlikely to confound the effect of gender in the regression of technical stream choice.

While the regression given in Equation (1) identifies the gender disparity in technical stream, i.e. STEM subject choice, we also investigate the gender gap for each subject separately. Towards this objective, we estimate a multinomial logit model where the outcome is one of the following stream choices: arts, commerce, science, engineering/vocational, and others. We use the same explanatory variables as described in Equation (1). Since one cannot include fixed effects directly in such a non-linear model, therefore we express the household fixed effects as linear functions of the household specific averages of the individual level explanatory variables, as suggested by Mundlak (1978). This analysis helps us to identify the gender gap in each of the subjects considered to define the technical stream.

4.2 Results on gender gap in stream choice

First we focus on intra-household gender differences in the choice of technical stream at the post-secondary level. The results from estimating Equation (1) are given in Table 2. The first column presents a model that includes only some basic individual level control variables, such as age, birth order, and parental education. Household fixed effects are included in the subsequent models. We add the control variables in sequence to check if the coefficient of gender is sensitive to varying specification. Column 2 considers the SSLC exam performance as past ability control, while column 3 accounts for prior mathematics score from the earlier survey round. The final column presents the full model including all measures of cognitive ability.

In all of the regressions, there is a statistically significant female disadvantage in the choice of technical stream. Girls are approximately 20 percentage points less likely than boys to study a technical stream vis-à-vis humanities. Among all boys and girls, 50 percent are enrolled in the technical stream. This implies a gender gap of a magnitude of 40 percent which is very substantial.

Turning to the other variables, education of both parents has a positive effect, with the effect of mother's education being substantially larger. But, unsurprisingly, it becomes imprecise once household fixed effects are included as little variation is left to identify effects. As expected we find that students who score better in the secondary level board examination (SSLC) are more likely to study a technical stream at the higher-secondary level. Among the variables capturing the past test scores, mathematics has a significant effect. Considering that these independent tests were designed to assess very basic level of knowledge, it is not surprising that only the highest level of difficulty in mathematics, i.e. division, comes out to be the only significant predictor of choosing the technical stream. The effect of mathematical ability remains significant even when all the controls are included in the final model. However, the gender gap also remains significant and the effect size is almost unperturbed even after taking into account the variation in cognitive abilities. This finding strongly indicates that the gender gap in stream choice is not driven by the intrinsic ability of students.

Heterogeneity in gender gap

What drives the gender gap in stream choice? To explore this question, we analyze how the gender gap varies along with some other explanatory factors. One possibility is that studying in the technical stream is more expensive than studying arts or humanities. Technical subjects are likely to involve higher direct cost (e.g. school fees may be higher due to science lab requirements) or indirect cost (e.g. private tuition for mathematics, which is a widely practiced phenomenon in India). Chandrasekhar et al. (2016) find that the average spending on higher education by those who are enrolled in technical stream is nearly five times larger than in general stream in India. Since Indian households are likely to make greater educational investments on boys, therefore higher cost of technical stream may discourage them from enrolling girls in such streams, especially when households have limited resource for children's education.[10] To check if resource constraint leads to gender disparity, we interact the gender dummy in our model with household income (per capita). We mitigate the potential endogeneity in household income by taking baseline income from the earlier round, instead of contemporaneous income. The results are given in columns 1 and 2 of Table 3. The first column, which does not include household fixed effects, shows that household income has a positive effect on a male student's likelihood of selecting technical stream, but it has no significant effect on the gender gap. The second model includes household fixed effects and provides the same finding that household income does not affect the gender gap. We arrive at a similar result if we use household assets (durable goods) instead of income.

If the gender gap does not vary with household's affluence, then does it vary with some non-pecuniary characteristic of the household? A relevant aspect to consider would be household's attitude towards gender equality in education. We capture this aspect through a measure of educational parity between parents. Usually mothers have lower education level than fathers. We posit that if the difference between mother's and father's educational attainment is smaller, then it signifies higher educational parity between parents, and hence it can reduce gender disparity in their children's education as well. Therefore, we interact female dummy with the difference in years of education between mother and father. Columns 3 and 4 of Table 3 show that a greater parity in parental education significantly reduces the gender gap in stream choice. On average, a mother has 1.7 years lesser education than father; if this figure improves and there is equality in educational attainment between the parents, then it reduces the gender gap in technical stream choice of their children by 3 percentage points.

In addition to demand side factors such as household income and preferences, we also seek to explore the role of supply side factors, especially access, in this context. Any variation in access to technical education is subsumed in household fixed effects in our main model. However, access may have a differential effect by gender. For instance, girls may be more discouraged to travel longer distance to get education than boys. Therefore, a pertinent

[10] On similar lines, Sahoo (2017) finds that due to higher fees in private schools, parents send boys to private schools and girls to government schools which are almost free.

question is whether gender disparity reduces when technical education is made more accessible. To investigate this question, we use a variable that measures the total number of science and technical colleges in the district at the time when the choice of technical stream was made.[11] These colleges would create avenues for the students to continue their technical education at the tertiary level. As Mukhopadhyay and Sahoo (2016) point out, the possibility of continuation to higher levels of education is an important determinant of participation at lower levels of schooling. Thus, on one hand the number of such colleges represents the overall access to technical education in the district. On the other hand, they reveal the possibility of further pursuing technical education at higher levels. While we are unable to separate out these two effects, yet, this variable gives us a plausible measure of access. When we interact this variable with gender dummy, the results show that districts with higher number of colleges that provide science or technical education not only attract more students to study technical stream, but also it helps reduce the gender gap in stream choice. A standard deviation increase in the number of science/technical colleges per million population in the district is associated with a reduction of 13.5 percentage points in the gender gap in higher-secondary stream choice.

Gender gap in the choice of individual streams

As mentioned before, technical stream includes STEM and commerce streams, while non-technical stream represents arts or humanities. To understand the gender gap in each stream separately, we estimate a multinomial logit model and present the marginal effects in Table 4. Girls are 22 percentage points more likely than boys to choose arts than other subjects. The gender gap is highest for science (9.5 percentage points) followed by commerce (8.8 percentage points). It is not significant in the choice of engineering/vocational or other streams, which anyway have very low overall enrollment rates – only about 5 percent individuals study in these streams.

[11] We use All India Survey of Higher Education (AISHE) data to calculate this number. AISHE includes information on the courses offered by and the year of inception of each institute. We use a lagged measure of number of colleges with respect to an individual's admittance to the higher-secondary level of education, when the decision of stream choice would be made. We normalize the total number of science/technical colleges with respect to the total population of the district, before using the measure in the regression.

5 Exploring relationship between stream choice and labour market outcomes

In this part of the paper, we delve into exploring whether the choice of different subjects at the post-secondary schooling affects labour market outcomes of adults. Our empirical model is also designed to illustrate the implication of stream choice on the gender gap in economic outcomes.

5.1 Empirical model for labour market outcomes

For this analysis we consider all individuals in the age-group of 25-60 years. The dependent variable Y_{ih} is a specific measure of labour market outcome of individual i from household h. Our model for estimation is given below:

$$\begin{aligned} Y_{ih} = \delta + \eta\, Female_{ih} + \lambda\, Female_{ih} * Sec_{ih} * Techedu_{ih} + \psi\, Female_{ih} * Sec_{ih} \\ + \pi\, Sec_{ih} * Techedu_{ih} + \sum_k \tau_k Eduyear_{ihk} + \rho\, Ability_{ih} + \zeta X_{ih} + \sigma_h \\ + \epsilon_{ih} \end{aligned} \tag{2}$$

Among the main explanatory variables, we have gender of the individual ($Female_{ih}$), an indicator for whether the individual has passed secondary level of education (Sec_{ih}), and whether the individual studied in a technical stream ($Techedu_{ih}$) in the higher-secondary level. Note that stream choice is applicable only for individuals who have passed secondary level, hence we include $Techedu_{ih}$ interacted with Sec_{ih}. Therefore, π captures the effect of technical stream choice for men given that he has at least secondary education degree. Similarly, λ gives us the additional effect of technical stream choice for females given that her education level is at least secondary. It is possible that having a secondary level education itself has a differential effect on earnings by gender. To ensure that the effect of technical stream choice is not confounded by the effect of secondary level education, we also interact the gender dummy with the secondary-pass dummy. This helps us to identify any differential effect of stream choice over and above any effect of secondary level of education. The variable Sec_{ih} does not appear separately in the equation because it is subsumed in the set of dummy variables indicating years of education ($Eduyear_{ihk}$).

This analysis essentially is very similar to estimation of returns to education where the usual omitted ability bias is a concern. Individuals who choose a technical stream in their higher-secondary education may have better ability because of which they may also perform better in the labour market later in life. If this is true, then the effect of technical stream choice may be overestimated. The data on secondary school leaving certificate (SSLC) examination results are available for all individuals including the adult sample. Therefore, we include this variable as a proxy for cognitive ability in this regression.[12] Since the results of the SSLC examination are available only for those who have passed this level, therefore this variable is interacted with the secondary pass dummy. Thus, in essence, ability is controlled for those individuals who have passed secondary level.

We control for other individual level factors (X_{ih}) such as marital status and relationship to head of the household, which are important determinants of labour market participation especially for women. Note that we also include household fixed effects (σ_h) in the regression. Therefore, the coefficients of gender and its associated interaction terms would measure the gender disparity in economic outcomes within households. The importance of using a household fixed effects model lies in the fact that much of the variation in female labour force participation is caused by household level determinants in India (Sarkar et al., 2017). Unobservable preferences such as family status concerns determine what kind of employment is deemed fit for a woman. Fixed effects would absorb all observable and unobservable household specific factors that may be relevant in this context.

The model allows us to answer two questions that we are particularly interested in. The first question focuses on women and examines if the outcome varies depending on whether she has studied in a technical stream or non-technical stream at the higher-secondary level. This is basically the returns to technical stream choice for women, and it is captured by ($\lambda + \pi$), as illustrated below:

$$E(Y \mid Female = 1, Sec = 1, Techedu = 1) - E(Y \mid Female = 1, Sec = 1, Techedu = 0)$$
$$= \lambda + \pi$$

The second question seeks to identify if the returns to technical stream choice vary between men and women. If women can reap greater benefits from studying technical field than men, then it may reduce the gender gap in economic outcomes. To test whether this is true, we

[12] Azam et al. (2013) use a similar strategy to examine returns to English language skills in India using the first round of IHDS data.

focus on the coefficient λ, which demonstrates whether studying in technical stream at the higher-secondary level has any differential effect between men and women's outcome:

$$[E(Y \mid Female = 1, Sec = 1, Techedu = 1) - E(Y \mid Female = 1, Sec = 1, Techedu = 0)]$$
$$-[E(Y \mid Female = 0, Sec = 1, Techedu = 1) - E(Y \mid Female = 0, Sec = 1, Techedu = 0)] = \lambda$$

In the tables containing the results of our analysis, we present the main coefficients, and include $(\lambda + \pi)$. We also estimate Equation (2) by replacing $Techedu$ with individual subjects to explore the effect of different subjects separately.

5.2 Results: effect of stream choice on adult life outcomes

Next we focus on results of estimating Equation (2) which reflects whether the choice of study stream matter for gender gap in various labour market outcomes in the adult life. We present the regressions for the overall sample, and also separately for a rural and urban sample acknowledging the possibility that the labour market may be structurally different between rural and urban areas. We discuss the effect on different outcomes as follows.

Labour force participation

Table 5 presents the results of the regression where the outcome variable is whether an individual participates in the labour force, according to the principal activity status during the last one year from the date of survey. We investigate whether technical stream choice affects female labour force participation (FLFP). Since men's labour force participation is almost universal (96 percent), therefore having a technical degree may not make any difference for them. However, women's labour force participation, as corroborated by the recent literature, is very low at 26 percent in our sample, and the female dummy is large, negative, and highly significant; interestingly, it is even slightly larger for those women with a secondary pass, in line with the literature showing a u-shape effect of education on female participation. One of the reasons behind low FLFP is that not all jobs are considered suitable for women to participate, which is partly due to social stigma and family status concerns (Klasen and Pieters, 2015). Having a technical education may enable a woman to gain access to better-quality jobs – which may encourage her to take up employment. Without a technical education, she may only be able to get a lower-quality job and hence not participate at all, as returns from household production or status concerns may prevent her from taking up that job.

However, these mechanisms are unlikely to apply for men, as almost all of them participate in the labour market. It is therefore possible that women benefit more from studying in a technical stream than men, resulting in a reduction of gender gap in labour force participation.

Let's first consider the total effect on women as captured by the term $(\lambda + \pi)$, which is statistically significant and positive in the overall, rural and urban samples. The probability that a woman will participate in the labour market increases by around 9 percentage points when she has studied in a technical stream as compared to arts or humanities at higher-secondary level. As expected, there is no effect of technical stream choice on men's participation. Hence, the coefficient λ that captures the differential effect on women versus men, is also significant and positive. This suggests that among those who have studied at the higher-secondary level, the choice of technical stream significantly reduces the gender gap in labour force participation. We decompose the effect of the technical stream into various individual streams and run the same regression. Appendix Table A2 reveals that the effect is mainly driven by science and engineering/vocational subjects.[13]

Salaried employment versus farm work

In the next analysis, we examine whether women are indeed able to participate in better-paying jobs when they have studied in technical stream. The survey records the following types of labour market engagement of each individual: salaried employment, casual wage labour, family business, and family farm and animal rearing. We find that salaried employment is the most remunerative activity, where the average annual income of participating individuals is Rs. 102,012. The next best activity in terms of remuneration is casual wage labour which, having the average annual earning at Rs. 31,112, lags far behind. Salaried employment also has much lower female representation – among the employed men, 27 percent are salaried employee while that number is only 13.17 percent for employed women. Almost 45 percent women participate in family farm and animal work which, along with family business, is the least remunerative activity. These observations motivate us to test whether having a technical education helps women to participate more in salaried employment vis-à-vis other less-paying activities.

[13] Among other variables, having scored in the first division in the SSLC exam has a significant positive effect on labour force participation especially in urban area. However, the effect of secondary education itself on female participation is negative, which is consistent with the U-shaped education-participation relationship found in other papers as well (Klasen and Pieters, 2015; Sarkar et al., 2017).

We present the results of regression where the outcome variable is a binary indicator of whether the individual participated in salaried employment as compared to other types of employment, restricting ourselves to the sample of employed individuals. From Table 6 we find that technical stream choice significantly increases the probability of females being employed in salaried jobs by 10.2 percentage points in the overall sample. The effect for men is null. The positive effect for women is found in both rural and urban areas.

When we estimate the model for other types of employment outcomes (Appendix Table A3), we find that females who have technical education are significantly less likely to work on the family farm and in animal husbandry. Therefore, the findings show that technical education helps women to move away from traditional engagement in the family farm, and be employed in a salaried job. Disaggregating the effect of technical education, we find that the increased participation in salaried employment is mainly driven by the choice of the science stream (Appendix Table A4).

Participation in male-dominated occupations

A major reason behind pervasive occupational segregation is that even when women's labour force participation increases, many participate in traditionally female-dominated occupations (Borrowman and Klasen, 2017). Does technical education help them break this barrier and enter into male-dominated occupations? To answer this question, we compute the share of males in each occupation category and mark those occupations as male-dominated where the share of males is greater than the median share in all occupations.[14] We estimate the model with an outcome variable to indicate whether the individual participated in a male-dominated occupation or in one of the other occupations.

Table 7 reveals that in the overall sample, women with technical education ($\lambda + \pi$) have a significant 7.1 percentage points higher likelihood of participating in a male-dominated occupation than women who have non-technical education. This effect is even more prominent in urban areas where it also leads to a significant reduction in the gender gap in participation in male-dominated occupations. Appendix Table A5 shows that the effect is driven by both commerce and science subjects. There is no such effect in rural areas, which is plausible, because rural areas usually offer less opportunity in terms of diversity in

[14] These occupations are categorized into two-digit National Occupation Classification (NCO) codes. There are about 90 such occupation categories in the data.

occupations. Besides, gender norms are usually much stronger in rural areas, making it even more difficult for women to participate in occupations dominated by men.

Earnings

The final labour market outcome we consider is the annual earnings of all working individuals. The earlier results show that technical education promotes not only women's participation in the labour force, but also the probability that she will be a salaried employee and take up an occupation which is male-dominated. Salaried employment happens to offer substantially better remuneration than other types of employment. Figure 6 reveals that occupations which have higher shares of male employees, also offer higher levels of income. Therefore, participation in a male-dominated occupation is likely to translate into higher earnings as well. To test if this is indeed the case, we estimate the regression with logarithm of annual earnings for all workers as a dependent variable. We also consider annual work intensity and hourly earnings in the later part of this analysis.

A well-known problem in estimating earnings equation is that individual level earnings are not observable for all, leading to a potential sample selection problem. In our case, individual specific earnings are recorded only for those engaged in casual wage or salaried employment. We take this issue into account by following Heckman's two-step selection correction model. In the first step, a probit equation modeling participation is estimated, and the selection correction term (Inverse Mills ratio) is calculated, which is then included as an additional control in the main earnings equation.[15] The standard errors are bootstrapped to avoid the problem of generated regressor.

Table 8 presents results of estimation without selection correction in columns 1–3, and with selection correction in columns 4–6. The Inverse Mills ratio is statistically significant in the selection-correction model for the overall, rural and urban samples. Without selection

[15] For ease of identification, the participation equation should contain some extra variable(s) not included in the main earnings equation. We follow the existing literature and use household size as an identifying variable because it is likely to affect income of an individual only through the incentives to participate (Nielsen and Westergård-Nielsen, 2001; Doud, 2005; Agrawal and Agrawal, 2018). We also include an interaction of female dummy with household size to allow the effect to vary by gender. Household size significantly increases participation of males, but there is no differential effect on females.

correction, the main coefficients are slightly underestimated. We find that the female dummy which reflects gender gap in earnings for those who have education below the secondary level is always significantly negative, revealing the large gender gap in earnings that prevails in the labour market. The interaction between female and secondary pass is positive and significant, suggesting that having a secondary level education helps women more than men in terms of earnings gain. Furthermore, we find that the interaction between female, secondary pass, and technical stream is positive and significant in the urban sample. This implies that among those individuals who have at least secondary level of education, having studied in a technical stream rather than arts reduces the gender gap in earnings by 28.2 percentage points. The reduction in gender gap accrues from a significant positive effect of technical education on women's earnings, while the same effect on men is insignificant.[16]

Appendix Table A6 decomposes the effect of technical education into various individual streams. The effect is seen to be completely driven by science. In fact, studying in science stream, as compared to arts, reduces the gender gap in earnings both in the overall and the urban samples. Studying engineering/vocational has a positive effect on the earnings of only men. For both men and women, stream choice does not affect earnings in the rural areas. Cognitive ability plays a role: those who have passed in the first or second division in secondary school leaving exam have significantly higher earnings as compared to those who have passed in the third division. However, the effect of cognitive ability is significant only in the urban areas, suggesting that earning opportunities in rural and urban areas are structurally different.

While we find a positive effect of technical education on women's annual earnings, we need to look into the intensity of their work-participation to interpret this effect appropriately. Technical education may boost up female labour force participation not only at the extensive margin, but also at the intensive margin. If employed individuals work for more number of days, that may lead to higher annual earnings. On the other hand, annual earnings may increase due to a rise in the wage rate which may reflect participation in better quality jobs. To separate out these two effects, we consider two additional outcome variables: individual's work intensity measured by the total number of hours worked in the last year, and hourly earnings. The results are presented in Appendix Table A7. We find that women who have

[16] Note that without household fixed effects, men with technical education earn more, suggesting that this premium is due to unobserved heterogeneity at the household level.

studied in a technical stream have higher work-intensity, and this effect is stronger in the urban sample. There is also a positive and significant effect of technical education on hourly earnings. This effect is found only in the urban areas, which is again consistent with our previous results. These findings imply that the pathway through which annual earnings rise when women have technical education, is partly increased work intensity and partly higher hourly income.[17]

5.3 Interpretation and robustness of results

The results of our analysis show a strong positive association between technical stream choice at the higher-secondary level and subsequent labour market outcomes of women in working-age. A pertinent question is whether the estimated relationship indeed captures the causal effect of technical education. If there are unobserved individual specific characteristics that propel women to self-select into technical education, and also drive their labour market performance, then the estimated effect is not causal. The existing literature on the confounding effect of omitted ability emphasizes that it causes an upward bias in the estimate. To mitigate this issue we use SSLC examination performance as a proxy for ability. The results indicate that this measure of cognitive ability used in our regression is quite meaningful, although it may not comprise all the dimensions of ability. Especially, what we cannot capture in our analysis is the effect of non-cognitive skills. A growing literature shows that non-cognitive or behavioral traits are important determinant of both stream choice and labour market outcomes. Existing studies such as Buser et al. (2014) emphasize that gender differences in competitiveness substantially explain why boys choose more math- and science-intensive streams. Most of these studies are based on laboratory experiments from advanced economies. Since we deal with observational data that do not have any direct measure of non-cognitive skills, therefore we are unable to examine whether the gender gap in stream choice is caused by any gender difference in non-cognitive characteristics.

What is also critical is the implication of omitted behavioral factors in the regression that links stream choice to labour market outcomes. Insofar as these behavioral aspects are not captured

[17] This explanation is also supported by an additional analysis, where we include work intensity as a control variable in the regression considering annual earnings (results not presented). The effect of technical education remains significant and positive for the urban sample, however, the magnitude diminishes.

by the included proxy for ability or the household fixed effects, the estimates may not represent the causal effect. In that case we cannot rule out the possibility that women who break the gender barrier and indeed choose technical stream have a particularly strong personality and hence achieve better economic outcomes later in life. Although due to data limitation we cannot deal with this problem completely, yet, we seek to address it by including a few additional control variables that partially capture personality traits of women. For a subset of women in the sample, IHDS collected additional information on their intra-household decision making and participation in various social and political activities in the community.[18] This subset consists of 74 percent of the women who are included in our sample of adult individuals. Using principal component analysis on individual variables, we construct two indices. The first index reflects to what extent the woman participates in decision making about day-to-day matters in the household.[19] The second index captures her involvement in community organizations: membership in women's group, self-help or microcredit group, any political organization, and participation in public meetings. Johnston et al. (2016) show that there is a strong connection between non-cognitive skills and intra-household decision making of individuals. Therefore, these indices can be considered as a partial measure of women's personality traits. We include them as additional controls in our regression. Women for whom we do not have this information are dropped from the sample for this robustness analysis. Note that we do not have information on men's personality; therefore these variables are included as an interaction with the female dummy.[20] We present the results for the main two outcomes, labour force participation and earnings, in Appendix Table A8. The indices of decision making as well as social and political participation have a statistically significant and positive effect on the outcome. However, the effect of technical stream choice still remains significant and there is a negligible change in its magnitude. Therefore, it seems that the relationship is unlikely to be mostly confounded by personality traits of individuals.

[18] The eligible women for additional information in the 2012 survey are those who were present in the 2005 survey, and an ever-married woman aged 15-49 years.

[19] These questions ask whether the woman decides what to cook on a daily basis, purchase of expensive items, decision on the number of children, decision about health seeking behaviour, purchase of land and properties, etc.

[20] Since the majority of households have one eligible woman, therefore an implication of not having personality information for men is that the variation in women's personality used in the regression is coming from across households, despite having household fixed effects in the regression.

6 Conclusion

In this study, we provide the first quantitative estimate of the extent of gender segregation in post-secondary stream choice, and its potential impact on gender inequality prevailing in the labour market of India. We show that within-households, girls are 20 percentage points less likely than boys to choose a technical stream, as compared to arts or humanities, in higher-secondary education. This gender difference is not caused by any possible variation in cognitive ability between boys and girls. The size of the gender gap is quite substantial considering the average participation in technical stream. Further, the gender gap is not affected by household's affluence; rather, educational parity between parents and better access to technical education help reduce the gender gap.

In the second part of our paper, we show that women who have studied a technical stream, especially science, have significantly better labour market outcomes in adult life than those who studied humanities. In many cases, women are able to reap larger benefits from studying technical or STEM subjects than men. Thus, STEM choice by women also significantly reduces the intra-household gender gap in adult-life economic outcomes, such as labour force participation and earnings. Exploring channels for these impacts, we find that STEM choice enables women to participate in salaried employment and in occupations which are usually male-dominated. This result highlights the connection between educational segregation in terms of stream choice at the school level and occupational segregation in the labour market.

In all the analyses presented in this paper, we try to control for cognitive ability of individuals to mitigate any omitted ability bias in our estimates. However, a caveat in our dataset is that it does not have any direct measure of non-cognitive skills which can also be an important determinant of both stream choice and labour market outcomes. We attempt to address this issue by including variables capturing women's personality traits, as reflected by their decision making power within the household and extent of participation in social and political activities outside the household. The results are robust to inclusion of these variables, suggesting that our main findings are not driven by unobservable confounders. Nonetheless, we acknowledge that our analysis is based on observational data and only includes partial measures of ability of individuals; hence it would require more information, especially on non-cognitive skills, to be more certain of causality in the relationship between technical education and adult-life economic outcomes.

While we acknowledge the potential role of behavioral characteristics of individuals, yet, factors such as access to technical education or attitude towards women's education and economic participation still remain crucial issues in a developing economy. Our analysis suggests that in the context of India, being a girl translates into choosing the type of education which is less rewarding in the labour market. Educational policies should aim at creating incentive for girls to break this barrier and be free to choose subjects which are traditionally dominated by boys. This will potentially have far reaching benefits in reducing the gender gap in labour market outcomes.

References:

Agrawal, T., & Agrawal, A. (2018). Who Gains More from Education? A Comparative Analysis of Business, Farm and Wage Workers in India. *The Journal of Development Studies*, 1-18.

Antecol, H., & Cobb-Clark, D. A. (2013). Do psychosocial traits help explain gender segregation in young people's occupations?. *Labour Economics*, *21*, 59-73.

Azam, M., Chin, A., & Prakash, N. (2013). The returns to English-language skills in India. *Economic Development and Cultural Change*, 61(2), 335-367.

Basu, D., & de Jong, R. (2010). Son targeting fertility behavior: Some consequences and determinants. *Demography*, 47(2), 521–536.

Beffy, M., Fougere, D., & Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter?. *Review of Economics and Statistics*, 94(1), 334-347.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: fact or artifact?. *Science*, 210(4475), 1262-1264.

Bieri, F., Imdorf, C., Stoilova, R., & Boyadjieva, P. (2016). The Bulgarian educational system and gender segregation in the labour market. *European societies*, *18*(2), 158-179.

Blau, Francine D., & Lawrence M. Kahn. (2017). The Gender Wage Gap: Extent, Trends, and Sources. *Journal of Economic Literature*, 55 (3), 789-865.

Borrowman, M., & Klasen, S. (2017). Drivers of gendered sectoral and occupational segregation in developing countries. Courant Research Centre: Poverty, Equity and Growth-Discussion Papers NO 222.

Boudarbat, B. (2008). Field of study choice by community college students in Canada. *Economics of Education Review*, 27(1), 79-93.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, *129*(3), 1409-1447.

Chandrasekhar, S., Rani, P. G., & Sahoo, S. (2016). Household Expenditure on Higher Education in India: What do we know & What do recent data have to say? IGIDR Working Paper No. 2016-030. Indira Gandhi Institute of Development Research, Mumbai, India.

Clark, S. (2000). Son preference and sex composition of children: Evidence from India. *Demography*, 37(1), 95–108.

Daoud, Y. (2005). Gender gap in returns to schooling in Palestine. *Economics of Education Review*, *24*(6), 633-649.

Duraisamy, P., & Duraisamy, M. (2014). Occupational segregation, wage and job discrimination against women across social groups in the Indian labor market: 1983–2010. Mimeo.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56(2), 209-230.

Flabbi, L. (2011). Gender Differentials in Education, Career, Choices and Labour Market Outcomes on a Sample of OECD Countries. World Bank, Washington, DC.

Friedman-Sokuler, N., & Justman, M. (2016). Gender streaming and prior achievement in high school science and mathematics. *Economics of Education Review*, *53*, 230-253.

Fuller, W. C., Manski, C. F., & Wise, D. A. (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources*, 477-498.

Hill, C., Corbett, C., & St Rose, A. (2010). Why so few? Women in science, technology, engineering, and mathematics. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Jensen, R. (2002). Fertility preferences and female disadvantage: Equal treatment, unequal outcomes? Cambridge: Mimeo, John F. Kennedy School of Government, Harvard University.

Joensen, J. S., & Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes?. *Journal of Human Resources*, *44*(1), 171-198.

Johnston, D. W., Kassenboehmer, S. C., & Shields, M. A. (2016). Financial decision-making in the household: Exploring the importance of survey respondent, health, cognitive ability and personality. *Journal of Economic Behavior & Organization*, *132*, 42-61.

Kahn, S., & Ginther, D. (2017). Women and STEM. NBER Working Paper No. 23525. National Bureau of Economic Research.

Kingdon, G. (2005). Where has all the bias gone? Detecting gender bias in the intrahousehold allocation of educational expenditure. *Economic Development and Cultural Change*, 53(2), 409–451.

Maitra, P., Pal, S., & Sharma, A. (2016). Absence of Altruism? Female Disadvantage in Private School Enrollment in India. *World Development*, 85, 105-125.

Mukhopadhyay, A., & Sahoo, S. (2016). Does access to secondary education affect primary schooling? Evidence from India. *Economics of Education Review*, 54, 124–142.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1), 69-85.

Nielsen, H. S., & Westergård-Nielsen, N. (2001). Returns to schooling in less developed countries: new evidence from Zambia. *Economic Development and Cultural Change*, *49*(2), 365-394.

Nollenberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. *The American Economic Review*, 106(5), 257-261.

Sahoo, S. (2017). Intra-household gender disparity in school choice: Evidence from private schooling in india. *The Journal of Development Studies*, *53*(10), 1714-1730.

Sarkar, S., Sahoo, S., & Klasen, S. (2017). Employment transitions of women in India: A panel analysis. Courant Research Centre: Poverty, Equity and Growth-Discussion Papers No. 239.

Schneeweis, N., & Zweimüller, M. (2012). Girls, girls, girls: Gender composition and female school choice. *Economics of Education Review*, 31(4), 482-500.

Sookram, S., & Strobl, E. (2009). The role of educational choice in occupational gender segregation: Evidence from Trinidad and Tobago. *Economics of Education Review*, 28(1), 1-10.

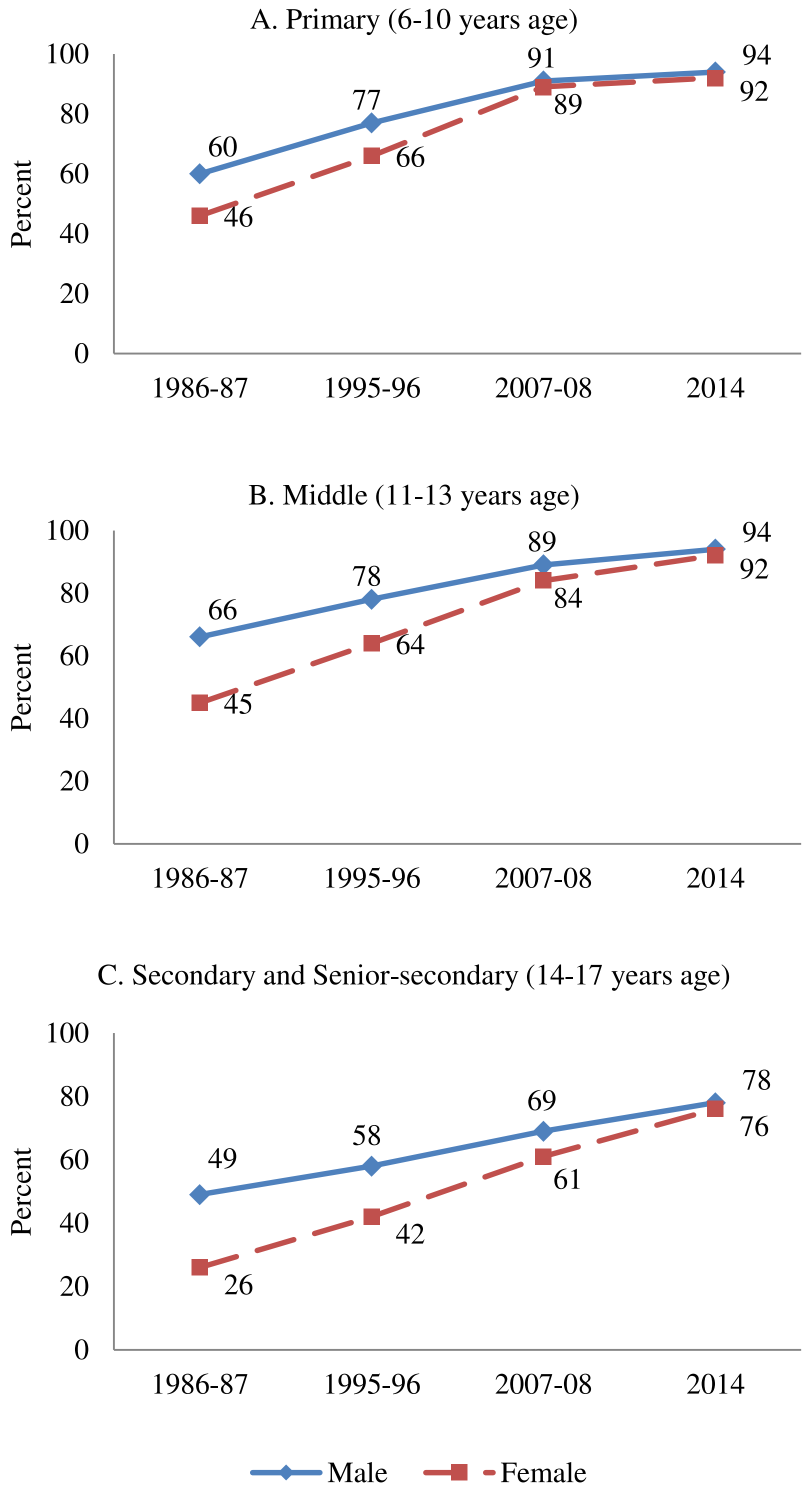
Van Puyenbroeck, T., De Bruyne, K., & Sels, L. (2012). More than 'Mutual Information': Educational and sectoral gender segregation and their interaction on the Flemish labor market. *Labour Economics*, *19*(1), 1-8.

World Bank (2012). World Development Report 2012: Gender Equality and Development. World Bank, Washington DC.

Yamaguchi, K. (1989). A formal theory for male-preferring stopping rules of childbearing: Sex differences in birth order and in the number of siblings. *Demography*, 26(3), 451–465.

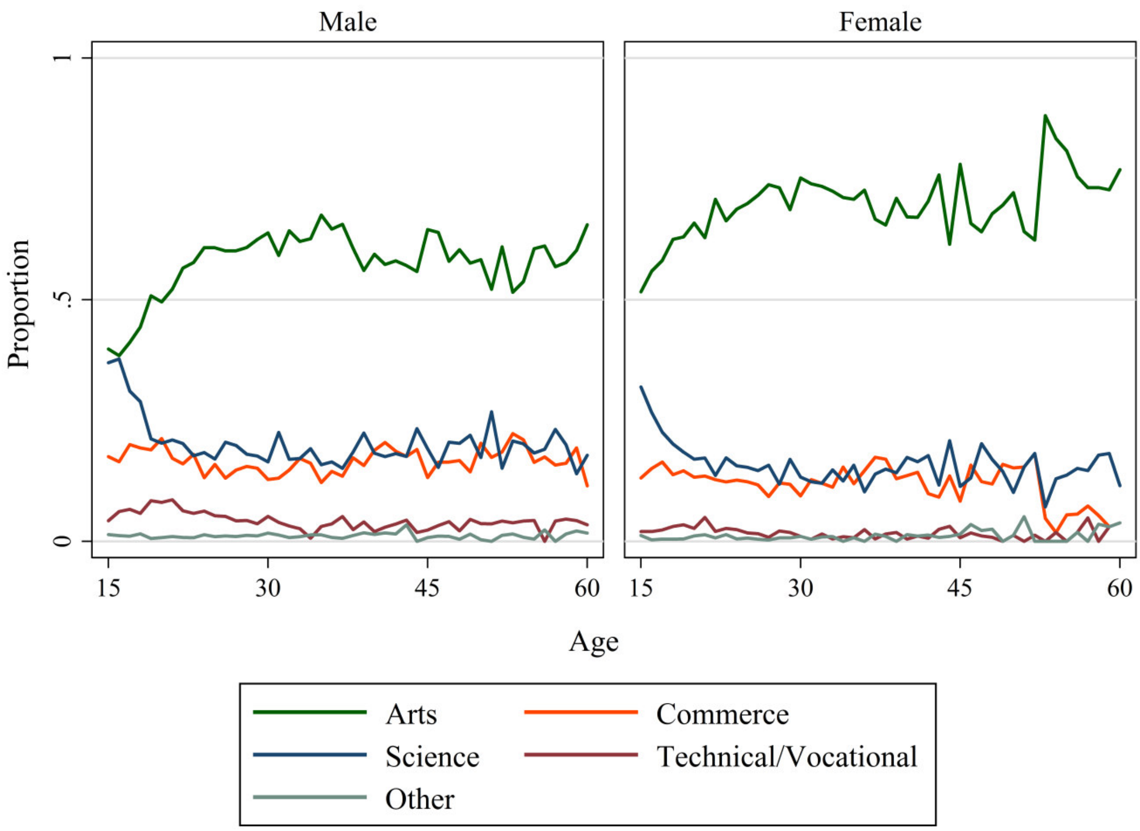
Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources*, *48*(3), 545-595.

Figure 1: Trends in enrollment rates in India: Estimates from National Sample Surveys



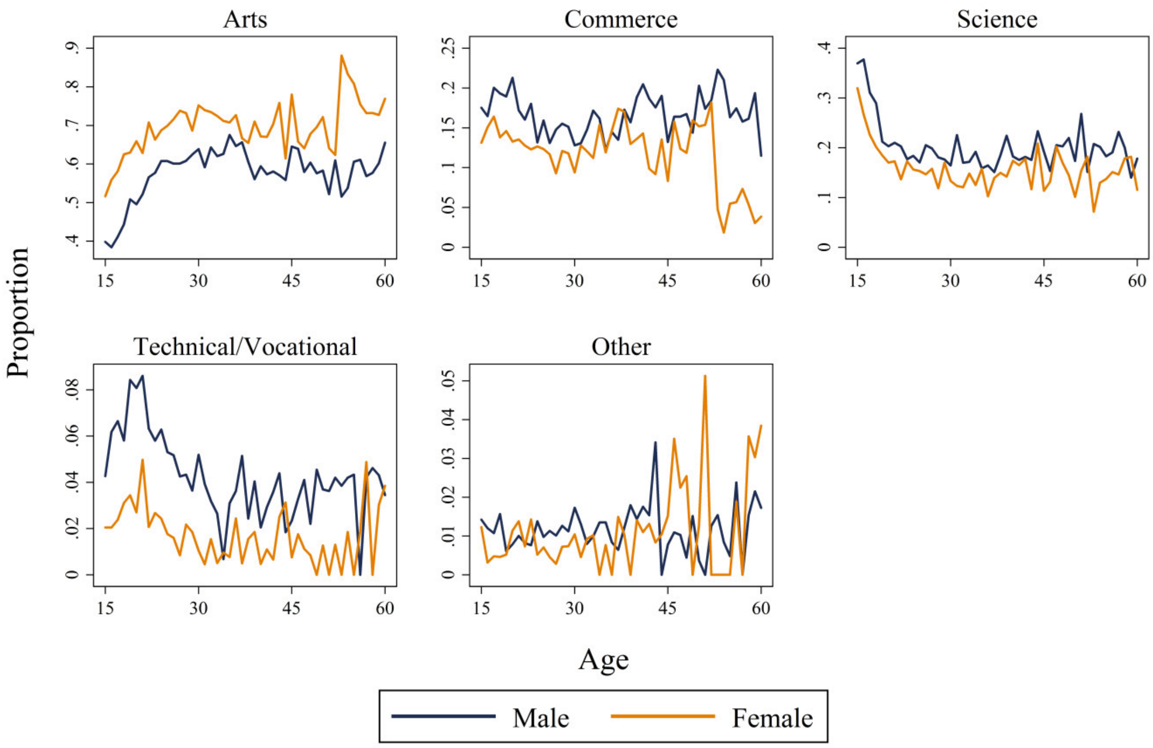
Source: Authors' calculation from multiple rounds of National Sample Survey data.

Figure 2: Proportion of males and females (aged 15-60 years) in different study streams at the higher-secondary level



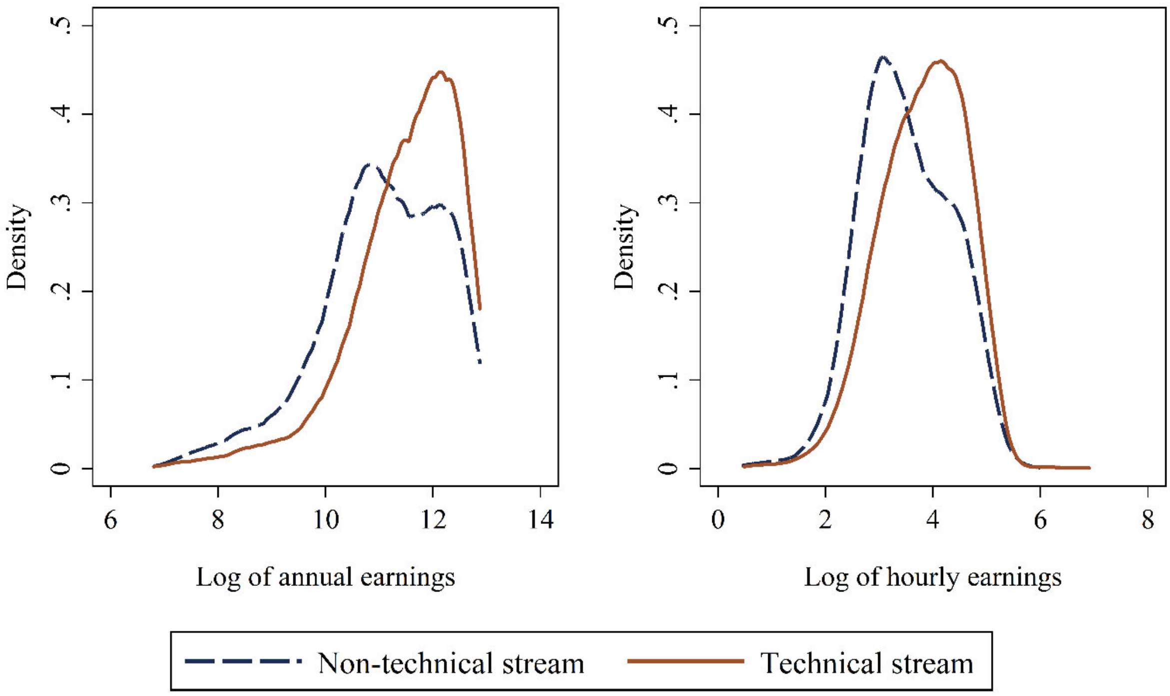
Source: Authors' calculation from IHDS 2011-12 data.

Figure 3: Raw gender differences in the choice of study streams at the higher-secondary level for individuals aged 15-60 years



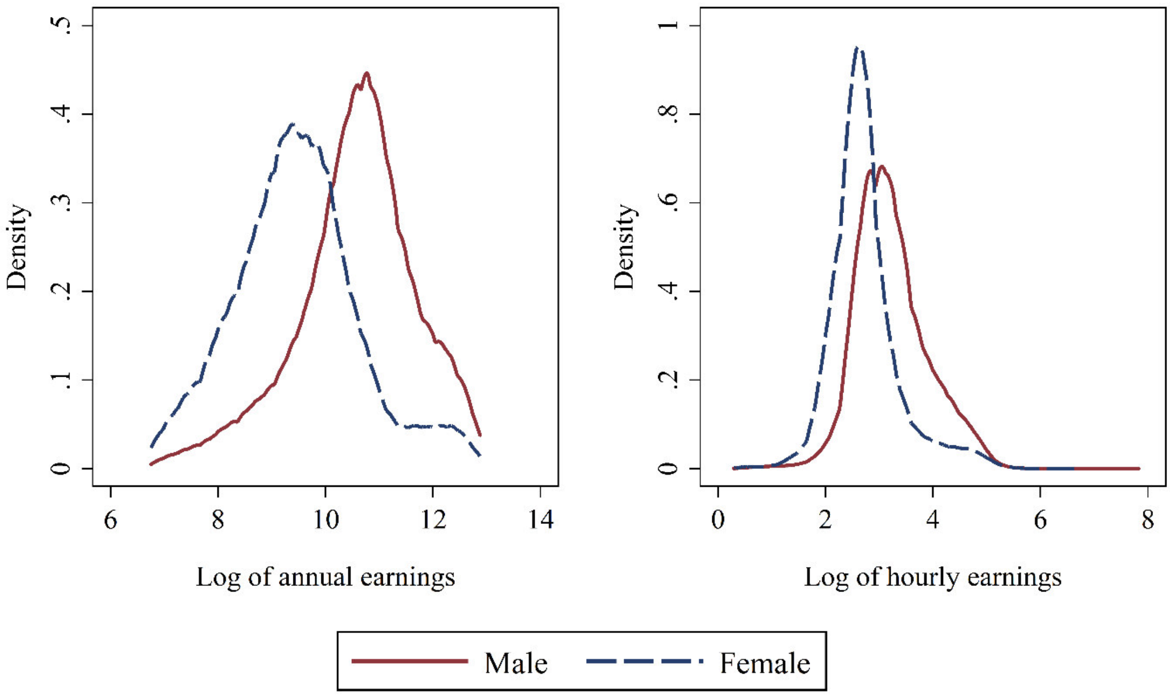
Source: Authors' calculation from IHDS 2011-12 data.

Figure 4: Kernel density estimates of log of annual and hourly earnings of individuals (aged 25-60 years) based on whether they studied a technical stream in higher-secondary level



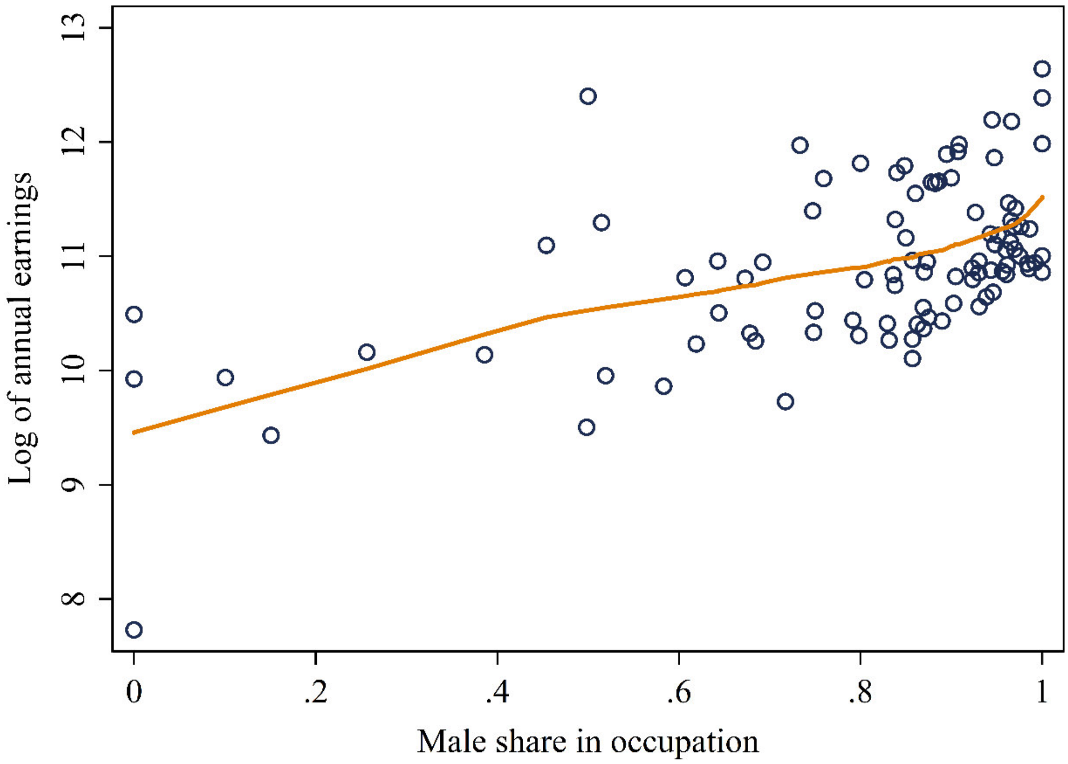
Source: Authors' calculation from IHDS 2011-12 data.

Figure 5: Kernel density estimates of log of annual and hourly earnings of males and females (aged 25-60 years)



Source: Authors' calculation from IHDS 2011-12 data.

Figure 6: Relationship between male share in occupation and average earnings in occupation (scatter plot and lowess smoother)



Source: Authors' calculations from IHDS data.

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.
Stream choice sample: age 15-18 years			
Female	14845	0.50	0.50
Age (years)	14845	16.49	1.11
Birth order	14845	1.19	0.42
Number of siblings	14845	0.40	0.55
Father's years of education	14845	4.93	4.93
Mother's years of education	14845	3.23	4.34
Secondary pass	14845	0.40	0.49
Technical subject	5213	0.50	0.50
Arts	5213	0.50	0.50
Commerce	5213	0.17	0.37
Science	5213	0.28	0.45
Engineering/Vocational	5213	0.04	0.20
Others	5213	0.01	0.09
Secondary result: 1st division	5928	0.41	0.49
Secondary result: 2nd division	5928	0.47	0.50
Secondary result: 3rd division	5928	0.12	0.33
Math: none	3005	0.04	0.21
Math: number	3005	0.20	0.40
Math: subtraction	3005	0.34	0.47
Math: division	3005	0.42	0.49
Reading: none	3034	0.01	0.11
Reading: letter	3034	0.05	0.22
Reading: word	3034	0.11	0.31
Reading: paragraph	3034	0.24	0.43
Reading: story	3034	0.59	0.49
Writing: none	2992	0.12	0.32
Writing: can write	2992	0.88	0.32
Household income per capita (baseline)	5235	9.68	12.23
Educational parity between parents	5934	-1.67	4.93
Science/technical colleges in district (number per million population)	5707	9.67	10.39
Labour market outcome sample: age 25-60 years			
Labour force participation	80302	0.60	0.49
Salaried employment	80302	0.15	0.35
Casual wage labour	80302	0.29	0.45
Family business	80302	0.08	0.27
Family farm and animal work	80302	0.19	0.40
Male dominated occupation	37348	0.57	0.50
Log of annual earnings	36559	10.16	1.24
Female	80302	0.52	0.50
Age (years)	80302	40.40	10.33
Years of education	80302	5.60	5.23

Secondary pass	80302	0.21	0.41
Secondary pass * Arts	80302	0.14	0.34
Secondary pass * Commerce	80302	0.03	0.17
Secondary pass * Science	80302	0.04	0.19
Secondary pass * Engineering/Vocational	80302	0.01	0.08
Secondary pass * Others	80302	0.002	0.05
Secondary result: 1st division	80302	0.06	0.23
Secondary result: 2nd division	80302	0.12	0.33
Secondary result: 3rd division	80302	0.03	0.17
Marital status: Single	80302	0.07	0.25
Marital status: Married	80302	0.83	0.37
Marital status: Other	80302	0.10	0.30
Relation: Household head	80302	0.36	0.48
Relation: Spouse of head	80302	0.34	0.47
Relation: Son/daughter	80302	0.16	0.37
Relation: Child-in-law	80302	0.08	0.28
Relation: Grandchild	80302	0.003	0.05
Relation: Other	80302	0.05	0.22

Notes: The first set of summary statistics are for the sample of individuals from the stream choice analysis (15-18 years age). The second set corresponds to the labour market outcome regression which focuses on individuals in 25-60 years age-group.

Table 2: Effect of gender on higher-secondary stream choice (probability of choosing technical subjects (Science / Engineering / Vocational / Commerce) versus humanities (Arts))

	Chose technical subjects in higher-secondary education			
	(1)	(2)	(3)	(4)
Female	-0.178***	-0.213***	-0.202***	-0.221***
	(0.013)	(0.031)	(0.063)	(0.062)
Age (years)	-0.024***	0.028	-0.069	-0.058
	(0.008)	(0.029)	(0.070)	(0.070)
Birth order	-0.036	0.035	-0.061	-0.065
	(0.025)	(0.054)	(0.122)	(0.122)
Number of siblings	-0.007	0.085	0.184	0.236
	(0.015)	(0.115)	(0.258)	(0.248)
Father's years of education	0.004**	-0.001	-0.031	-0.030
	(0.002)	(0.009)	(0.050)	(0.053)
Mother's years of education	0.022***	0.013	-0.004	0.007
	(0.002)	(0.017)	(0.043)	(0.045)
Secondary result: 1st division		0.240***		0.200
		(0.065)		(0.153)
Secondary result: 2nd division		0.127**		0.072
		(0.057)		(0.128)
Math: number			-0.016	0.011
			(0.225)	(0.188)
Math: subtraction			0.221	0.276
			(0.227)	(0.205)
Math: division			0.501**	0.563**
			(0.250)	(0.222)
Reading: word				0.265
				(0.312)
Reading: paragraph				0.056
				(0.320)
Reading: story				0.116
				(0.327)
Writing: can write				-0.108
				(0.134)
Constant	0.897***	-0.157	1.709	1.262
	(0.153)	(0.561)	(1.288)	(1.278)
Observations	5,213	5,207	2,656	2,634
R-squared	0.086	0.129	0.236	0.283
Household fixed effects	No	Yes	Yes	Yes
Number of households (fixed effects)		4,653	2,515	2,496

The results are from a linear probability model taking children in the age-group of 15-18 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Heterogeneity in the effect of gender on higher-secondary stream choice

	Chose technical subjects in higher-secondary education					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.168***	-0.196***	-0.166***	-0.181***	-0.251***	-0.345***
	(0.018)	(0.074)	(0.014)	(0.066)	(0.018)	(0.085)
Female * Household income per capita (baseline)	-0.001	-0.002				
	(0.001)	(0.004)				
Household income per capita (baseline)	0.003***					
	(0.001)					
Female * Educational parity between parents			0.007***	0.018**		
			(0.003)	(0.009)		
Educational parity between parents			0.018***	0.023		
			(0.002)	(0.047)		
Female * Science/technical colleges in district (number per million population)					0.008***	0.013**
					(0.001)	(0.005)
Science/technical colleges in district (number per million population)					0.003***	-0.006
					(0.001)	(0.016)
Constant	0.827***	1.331	0.896***	1.378	0.707***	-0.379
	(0.162)	(1.274)	(0.152)	(1.257)	(0.157)	(1.311)
Observations	4,622	2,634	5,213	2,634	4,998	2,583
R-squared	0.088	0.284	0.087	0.298	0.112	0.321
Other individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Secondary exam result	No	Yes	No	Yes	No	Yes
Past math score	No	Yes	No	Yes	No	Yes
Past reading & writing score	No	Yes	No	Yes	No	Yes
Household fixed effects	No	Yes	No	Yes	No	Yes
Number of households (fixed effects)		2,496		2,496		2,447

The results are from a linear probability model taking children in the age-group of 15-18 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Marginal effects from multinomial logit model for the effect of gender on probability of choosing different study streams at the higher-secondary level of education

	Probability of choosing at higher-secondary level:				
	Arts	Commerce	Science	Engineering / Vocational	Others
	(1)	(2)	(3)	(4)	(5)
Female	0.220***	-0.088***	-0.095**	-0.025	-0.012
	(0.039)	(0.033)	(0.040)	(0.018)	(0.020)
Age (years)	-0.038	0.004	0.024	0.007	0.003
	(0.040)	(0.033)	(0.036)	(0.016)	(0.007)
Secondary result: 1st division	-0.281***	0.053	0.231***	-0.000	-0.002
	(0.070)	(0.062)	(0.067)	(0.025)	(0.012)
Secondary result: 2nd division	-0.050	0.024	0.020	-0.004	0.010
	(0.062)	(0.054)	(0.065)	(0.026)	(0.019)
Math: number	-0.158	-0.266**	-0.144	0.558***	0.009
	(0.128)	(0.116)	(0.129)	(0.062)	(0.026)
Math: subtraction	-0.359**	-0.203*	-0.107	0.649***	0.019
	(0.142)	(0.123)	(0.137)	(0.070)	(0.034)
Math: division	-0.396***	-0.142	-0.087	0.597***	0.028
	(0.151)	(0.138)	(0.143)	(0.069)	(0.046)
Reading: word	-0.085	0.119	-0.039	0.030	-0.026
	(0.141)	(0.105)	(0.154)	(0.053)	(0.050)
Reading: paragraph	-0.093	0.125	-0.040	0.015	-0.008
	(0.151)	(0.104)	(0.160)	(0.061)	(0.026)
Reading: story	-0.133	0.155	-0.008	0.008	-0.022
	(0.156)	(0.114)	(0.165)	(0.067)	(0.038)
Writing: can write	0.288***	-0.150**	-0.094	-0.022	-0.021
	(0.101)	(0.075)	(0.095)	(0.045)	(0.034)
Observations	2,634	2,634	2,634	2,634	2,634
Additional controls	Yes	Yes	Yes	Yes	Yes
Household fixed effects (Mundlak)	Yes	Yes	Yes	Yes	Yes

The results are marginal effects estimated from a multinomial logit model taking children in the age-group of 15-18 years. Control variables included are according to column 4 of Table 3. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Effect of higher-secondary technical stream choice on adult life labour force participation

	Probability of labour force participation		
	(1)	(2)	(3)
	Overall	Rural	Urban
Female	-0.586***	-0.566***	-0.643***
	(0.009)	(0.012)	(0.014)
Female * Secondary pass * Technical stream (λ)	0.090***	0.107***	0.088***
	(0.016)	(0.030)	(0.019)
Secondary pass * Technical stream (π)	0.005	-0.013	0.001
	(0.009)	(0.014)	(0.011)
Female * Secondary pass	-0.040***	-0.065***	0.027**
	(0.009)	(0.014)	(0.013)
Secondary result: 1st division	0.024*	-0.012	0.045***
	(0.012)	(0.019)	(0.017)
Secondary result: 2nd division	0.005	-0.014	0.020
	(0.010)	(0.014)	(0.014)
Constant	1.052***	1.082***	0.999***
	(0.018)	(0.023)	(0.030)
Effect of Technical stream on females (λ + π)	0.095***	0.094***	0.089***
	(0.015)	(0.029)	(0.018)
Observations	80,302	51,976	28,326
R-squared	0.665	0.646	0.704
Years of education dummies	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Number of households (fixed effects)	38,656	25,205	13,451

The results are from a linear probability model taking individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Effect of higher-secondary technical stream choice on salaried employment versus other employment in adult life

	Probability of salaried employment as compared to other employment		
	(1)	(2)	(3)
	Overall	Rural	Urban
Female	-0.006	-0.023**	0.063***
	(0.010)	(0.011)	(0.024)
Female * Secondary pass * Technical stream (λ)	0.115***	0.094*	0.098**
	(0.030)	(0.050)	(0.040)
Secondary pass * Technical stream (π)	-0.013	0.020	-0.031
	(0.016)	(0.022)	(0.024)
Female * Secondary pass	0.029	0.018	-0.002
	(0.018)	(0.023)	(0.030)
Secondary result: 1st division	0.098***	0.077***	0.130***
	(0.023)	(0.030)	(0.037)
Secondary result: 2nd division	0.056***	0.048**	0.072**
	(0.018)	(0.021)	(0.032)
Constant	0.159***	0.076***	0.304***
	(0.024)	(0.026)	(0.058)
Effect of Technical stream on females (λ + π)	0.102***	0.115**	0.067*
	(0.029)	(0.048)	(0.037)
Observations	57,735	41,170	16,565
R-squared	0.095	0.113	0.08
Years of education dummies	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Number of households (fixed effects)	34,622	23,367	11,255

The results are from a linear probability model taking employed individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Effect of higher-secondary technical stream choice on participation in male-dominated occupations in adult life

	Probability of participating in male-dominated occupations as compared to other occupations		
	(1)	(2)	(3)
	Overall	Rural	Urban
Female	-0.213***	-0.164***	-0.383***
	(0.017)	(0.019)	(0.034)
Female * Secondary pass * Technical stream (λ)	0.049	-0.021	0.123**
	(0.043)	(0.077)	(0.052)
Secondary pass * Technical stream (π)	0.022	0.048	-0.017
	(0.023)	(0.039)	(0.029)
Female * Secondary pass	-0.165***	-0.124***	-0.057
	(0.026)	(0.040)	(0.038)
Secondary result: 1st division	-0.022	-0.053	0.000
	(0.033)	(0.053)	(0.043)
Secondary result: 2nd division	-0.014	-0.049	0.022
	(0.027)	(0.038)	(0.036)
Constant	0.730***	0.702***	0.835***
	(0.039)	(0.050)	(0.066)
Effect of Technical stream on females (λ + π)	0.071*	0.027	0.105**
	(0.042)	(0.075)	(0.051)
Observations	37,552	25,742	11,810
R-squared	0.160	0.122	0.304
Years of education dummies	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Number of households (fixed effects)	25,954	17,252	8,702

The results are from linear probability model taking employed individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Effect of higher-secondary technical stream choice on adult life earnings

	Log of annual earnings					
	Without selection correction			With selection correction		
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Rural	Urban	Overall	Rural	Urban
Female	-0.774***	-0.785***	-0.793***	-1.100***	-0.983***	-1.485***
	(0.033)	(0.041)	(0.056)	(0.118)	(0.127)	(0.270)
Female * Secondary pass * Technical stream (λ)	0.073	-0.177	0.154*	0.118	-0.137	0.282***
	(0.082)	(0.176)	(0.088)	(0.083)	(0.183)	(0.094)
Secondary pass * Technical stream (π)	0.007	0.045	0.049	-0.004	0.011	0.014
	(0.048)	(0.081)	(0.056)	(0.051)	(0.082)	(0.058)
Female * Secondary pass	0.367***	0.440***	0.406***	0.348***	0.441***	0.470***
	(0.057)	(0.098)	(0.070)	(0.058)	(0.102)	(0.083)
Secondary result: 1st division	0.139**	0.096	0.251***	0.251***	0.150	0.452***
	(0.066)	(0.106)	(0.078)	(0.078)	(0.111)	(0.105)
Secondary result: 2nd division	0.113**	0.075	0.179***	0.155***	0.092	0.266***
	(0.056)	(0.083)	(0.067)	(0.056)	(0.083)	(0.075)
Inverse Mills ratio				-1.945***	-1.305*	-3.328***
				(0.694)	(0.783)	(1.282)
Constant	10.413***	10.240***	10.869***	11.715***	11.143***	12.862***
	(0.075)	(0.093)	(0.136)	(0.457)	(0.548)	(0.776)
Effect of Technical stream on females (λ + π)	0.080	-0.132	0.203**	0.114	-0.126	0.297***
	(0.078)	(0.171)	(0.079)	(0.081)	(0.173)	(0.086)
Observations	36,760	25,382	11,378	36,677	25,350	11,327
R-squared	0.371	0.387	0.352	0.371	0.387	0.355
Years of education dummies	Yes	Yes	Yes	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of households (fixed effects)	25,511	17,078	8,433	25,452	17,055	8,397

This regression considers salaried/casual wage employees in the age-group of 25-60 years. Robust standard errors (bootstrapped for columns 4–6) clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix

Appendix Table A1: Effect of gender on the probability of secondary level completion

	Completed secondary level of education			
	(1)	(2)	(3)	(4)
Female	0.012*	-0.013	0.018	0.023
	(0.007)	(0.013)	(0.026)	(0.026)
Age (years)	0.134***	0.089***	0.089***	0.086***
	(0.004)	(0.011)	(0.026)	(0.027)
Birth order	-0.024*	-0.103***	-0.094*	-0.085
	(0.012)	(0.023)	(0.052)	(0.053)
Number of siblings	-0.003	0.074**	0.242	0.272*
	(0.010)	(0.038)	(0.151)	(0.148)
Father's years of education	0.015***	0.007	-0.022*	-0.020*
	(0.001)	(0.005)	(0.012)	(0.012)
Mother's years of education	0.027***	0.001	-0.017	-0.016
	(0.001)	(0.011)	(0.014)	(0.014)
Math: number			0.021	0.017
			(0.044)	(0.047)
Math: subtraction			0.106**	0.069
			(0.050)	(0.058)
Math: division			0.338***	0.285***
			(0.058)	(0.066)
Reading: letter				0.007
				(0.057)
Reading: word				-0.038
				(0.057)
Reading: paragraph				0.005
				(0.068)
Reading: story				0.064
				(0.077)
Writing: can write				0.027
				(0.041)
Constant	-1.953***	-1.012***	-0.976**	-0.982*
	(0.069)	(0.212)	(0.496)	(0.502)
Observations	14,845	14,845	7,408	7,348
R-squared	0.208	0.207	0.326	0.330
Household fixed effects	No	Yes	Yes	Yes
Number of households (fixed effects)		11,636	6,560	6,512

The results are from a linear probability model taking children in the age-group of 15-18 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A2: Effect of higher-secondary subject choice on adult life labour force participation

	Probability of labour force participation		
	(1)	(2)	(3)
	Overall	Rural	Urban
Female	-0.587***	-0.566***	-0.645***
	(0.009)	(0.012)	(0.014)
Female * Secondary pass * Commerce	-0.009	0.046	-0.019
	(0.021)	(0.042)	(0.024)
Female * Secondary pass * Science	0.160***	0.123***	0.178***
	(0.022)	(0.041)	(0.026)
Female * Secondary pass * Engineering/Vocational	0.217***	0.359***	0.137
	(0.072)	(0.114)	(0.091)
Female * Secondary pass * Others	0.003	-0.095	0.086
	(0.075)	(0.119)	(0.098)
Secondary pass * Commerce	0.021*	-0.037*	0.026*
	(0.012)	(0.022)	(0.014)
Secondary pass * Science	0.001	0.015	-0.014
	(0.011)	(0.018)	(0.015)
Secondary pass * Engineering/Vocational	-0.027	-0.084**	-0.006
	(0.026)	(0.041)	(0.033)
Secondary pass * Others	-0.037	0.041	-0.118*
	(0.041)	(0.048)	(0.065)
Female * Secondary pass	-0.040***	-0.065***	0.027**
	(0.009)	(0.014)	(0.013)
Secondary result: 1st division	0.023*	-0.014	0.044***
	(0.013)	(0.019)	(0.017)
Secondary result: 2nd division	0.004	-0.016	0.021
	(0.010)	(0.014)	(0.014)
Constant	1.053***	1.082***	1.001***
	(0.018)	(0.023)	(0.030)
Observations	80,302	51,976	28,326
R-squared	0.666	0.647	0.705
Years of education dummies	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Number of households (fixed effects)	38,656	25,205	13,451

The results are from linear probability model taking individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A3: Effect of higher-secondary technical stream choice on different types of employment

	Probability of choosing:								
	Casual wage			Family business			Family farm and animal work		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Overall	Rural	Urban	Overall	Rural	Urban	Overall	Rural	Urban
Female	-0.173***	-0.167***	-0.190***	-0.015*	-0.012	-0.032*	0.194***	0.202***	0.159***
	(0.012)	(0.014)	(0.022)	(0.008)	(0.008)	(0.019)	(0.011)	(0.014)	(0.015)
Female * Secondary pass * Technical stream (λ)	-0.002	-0.019	-0.016	0.018	0.047	-0.004	-0.131***	-0.123**	-0.078***
	(0.019)	(0.037)	(0.021)	(0.024)	(0.035)	(0.035)	(0.021)	(0.050)	(0.017)
Secondary pass * Technical stream (π)	-0.031***	-0.026	-0.027**	-0.009	-0.027	0.024	0.053***	0.033	0.034***
	(0.011)	(0.017)	(0.014)	(0.013)	(0.017)	(0.021)	(0.013)	(0.022)	(0.012)
Female * Secondary pass	0.096***	0.060***	0.134***	-0.070***	-0.073***	-0.028	-0.054***	-0.005	-0.104***
	(0.012)	(0.018)	(0.020)	(0.014)	(0.016)	(0.025)	(0.015)	(0.024)	(0.017)
Secondary result: 1st division	-0.028	-0.037	-0.021	-0.046**	0.000	-0.091***	-0.025	-0.040	-0.018
	(0.017)	(0.025)	(0.022)	(0.018)	(0.021)	(0.032)	(0.018)	(0.030)	(0.015)
Secondary result: 2nd division	-0.030**	-0.032	-0.033	0.004	0.019	-0.022	-0.029*	-0.034	-0.017
	(0.015)	(0.020)	(0.020)	(0.014)	(0.015)	(0.028)	(0.015)	(0.022)	(0.014)
Constant	0.635***	0.723***	0.414***	0.105***	0.070***	0.250***	0.101***	0.130***	0.032
	(0.029)	(0.034)	(0.055)	(0.019)	(0.020)	(0.047)	(0.027)	(0.034)	(0.036)
Effect of Technical stream on females (λ + π)	-0.033*	-0.045	-0.043**	0.009	0.020	0.020	-0.078***	-0.089*	-0.044**
	(0.018)	(0.035)	(0.020)	(0.023)	(0.032)	(0.032)	(0.021)	(0.049)	(0.017)
Observations	57,735	41,170	16,565	57,735	41,170	16,565	57,735	41,170	16,565
R-squared	0.085	0.092	0.069	0.022	0.024	0.036	0.211	0.219	0.194
Years of education dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of households (fixed effects)	34,622	23,367	11,255	34,622	23,367	11,255	34,622	23,367	11,255

The results are from linear probability model taking employed individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A4: Effect of higher-secondary subject choice on salaried employment versus other employment in adult life

	Probability of salaried employment as compared to other employment		
	(1)	(2)	(3)
	Overall	Rural	Urban
Female	-0.007	-0.024**	0.062***
	(0.010)	(0.011)	(0.024)
Female * Secondary pass * Commerce	0.073	0.008	0.074
	(0.047)	(0.079)	(0.059)
Female * Secondary pass * Science	0.157***	0.195***	0.123***
	(0.035)	(0.060)	(0.046)
Female * Secondary pass * Engineering/Vocational	-0.020	-0.054	0.012
	(0.091)	(0.130)	(0.130)
Female * Secondary pass * Others	0.235	0.430*	0.127
	(0.153)	(0.227)	(0.207)
Secondary pass * Commerce	-0.028	-0.004	-0.035
	(0.021)	(0.032)	(0.030)
Secondary pass * Science	-0.002	0.036	-0.030
	(0.020)	(0.030)	(0.029)
Secondary pass * Engineering/Vocational	0.029	0.096	-0.025
	(0.042)	(0.060)	(0.058)
Secondary pass * Others	-0.078	-0.036	-0.126
	(0.067)	(0.076)	(0.127)
Female * Secondary pass	0.027	0.016	-0.004
	(0.018)	(0.023)	(0.030)
Secondary result: 1st division	0.095***	0.072**	0.129***
	(0.023)	(0.030)	(0.037)
Secondary result: 2nd division	0.055***	0.046**	0.071**
	(0.018)	(0.021)	(0.032)
Constant	0.158***	0.076***	0.302***
	(0.024)	(0.026)	(0.058)
Observations	57,735	41,170	16,565
R-squared	0.096	0.115	0.081
Years of education dummies	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Number of households (fixed effects)	34,622	23,367	11,255

The results are from a linear probability model taking employed individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A5: Effect of higher-secondary subject choice on participation in male-dominated occupations in adult life

	Probability of participating in male-dominated occupations as compared to other occupations		
	(1)	(2)	(3)
	Overall	Rural	Urban
Female	-0.214***	-0.163***	-0.386***
	(0.017)	(0.019)	(0.034)
Female * Secondary pass * Commerce	0.190***	0.085	0.271***
	(0.062)	(0.141)	(0.070)
Female * Secondary pass * Science	0.029	-0.076	0.121*
	(0.053)	(0.088)	(0.065)
Female * Secondary pass * Engineering/Vocational	-0.093	0.020	-0.178
	(0.138)	(0.220)	(0.159)
Female * Secondary pass * Others	-0.109	-0.117	-0.092
	(0.194)	(0.260)	(0.242)
Secondary pass * Commerce	0.053*	0.087	0.016
	(0.029)	(0.054)	(0.034)
Secondary pass * Science	-0.040	-0.004	-0.089**
	(0.031)	(0.053)	(0.038)
Secondary pass * Engineering/Vocational	0.182***	0.154	0.190***
	(0.051)	(0.095)	(0.056)
Secondary pass * Others	-0.089	-0.061	-0.146
	(0.094)	(0.091)	(0.171)
Female * Secondary pass	-0.170***	-0.123***	-0.067*
	(0.026)	(0.040)	(0.038)
Secondary result: 1st division	-0.018	-0.047	0.002
	(0.033)	(0.053)	(0.043)
Secondary result: 2nd division	-0.010	-0.045	0.023
	(0.027)	(0.039)	(0.036)
Constant	0.729***	0.704***	0.835***
	(0.039)	(0.050)	(0.065)
Observations	37,552	25,742	11,810
R-squared	0.163	0.123	0.311
Years of education dummies	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes
Number of households (fixed effects)	25,954	17,252	8,702

The results are from a linear probability model taking employed individuals in the age-group of 25-60 years. Robust standard errors clustered at the household level are given in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix Table A6: Effect of higher-secondary subject choice on adult life earnings

	Log of annual earnings					
	Without selection correction			With selection correction		
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Rural	Urban	Overall	Rural	Urban
Female	-0.774***	-0.787***	-0.792***	-1.104***	-0.998***	-1.489***
	(0.033)	(0.041)	(0.056)	(0.118)	(0.126)	(0.274)
Female * Secondary pass * Commerce	0.040	-0.289	0.152	0.016	-0.255	0.143
	(0.121)	(0.277)	(0.128)	(0.118)	(0.276)	(0.122)
Female * Secondary pass * Science	0.165*	0.014	0.203*	0.273***	0.112	0.416***
	(0.094)	(0.183)	(0.104)	(0.102)	(0.202)	(0.127)
Female * Secondary pass * Engineering/Vocational	-0.520	-0.711	-0.455	-0.449	-0.619	-0.345
	(0.362)	(0.728)	(0.278)	(0.366)	(0.823)	(0.304)
Female * Secondary pass * Others	0.200	-0.188	0.363	0.286	-0.089	0.503
	(0.587)	(0.897)	(0.575)	(0.649)	(1.021)	(0.684)
Secondary pass * Commerce	-0.051	0.006	-0.020	-0.041	-0.052	-0.002
	(0.056)	(0.096)	(0.067)	(0.058)	(0.106)	(0.072)
Secondary pass * Science	0.010	0.033	0.055	-0.023	-0.008	-0.007
	(0.064)	(0.114)	(0.072)	(0.066)	(0.121)	(0.074)
Secondary pass * Engineering/Vocational	0.149	0.046	0.313***	0.167*	0.056	0.311***
	(0.101)	(0.192)	(0.104)	(0.102)	(0.188)	(0.108)
Secondary pass * Others	0.089	0.387	-0.197	0.026	0.353	-0.369*
	(0.202)	(0.335)	(0.182)	(0.205)	(0.354)	(0.210)
Female * Secondary pass	0.367***	0.440***	0.404***	0.346***	0.437***	0.473***
	(0.057)	(0.098)	(0.071)	(0.058)	(0.102)	(0.084)
Secondary result: 1st division	0.134**	0.091	0.239***	0.249***	0.148	0.444***
	(0.066)	(0.107)	(0.077)	(0.078)	(0.113)	(0.105)
Secondary result: 2nd division	0.115**	0.080	0.176***	0.159***	0.099	0.267***
	(0.056)	(0.083)	(0.066)	(0.056)	(0.083)	(0.075)
Inverse Mills ratio				-1.974***	-1.382*	-3.351***
				(0.693)	(0.777)	(1.293)
Constant	10.407***	10.236***	10.861***	11.729***	11.192***	12.869***
	(0.075)	(0.093)	(0.136)	(0.458)	(0.544)	(0.787)
Observations	36,760	25,382	11,378	36,677	25,350	11,327
R-squared	0.371	0.388	0.354	0.372	0.388	0.357
Years of education dummies	Yes	Yes	Yes	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of households (fixed effects)	25,511	17,078	8,433	25,452	17,055	8,397

This regression considers salaried/casual wage employees in the age-group of 25-60 years. Robust standard errors (bootstrapped for columns 4–6) clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A7: Effect of higher-secondary technical stream choice on adult life working-intensity and hourly earnings

	Work intensity (annual hours of work)			Log of hourly earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Rural	Urban	Overall	Rural	Urban
Female	-599.163***	-550.627***	-668.059***	-0.488***	-0.401***	-0.843***
	(100.248)	(106.377)	(227.688)	(0.073)	(0.069)	(0.184)
Female * Secondary pass * Technical stream (λ)	120.184*	29.647	211.411**	0.029	-0.162	0.146*
	(67.911)	(125.401)	(88.086)	(0.065)	(0.119)	(0.082)
Secondary pass * Technical stream (π)	4.990	-6.694	6.736	0.003	0.018	0.011
	(39.126)	(68.375)	(53.479)	(0.033)	(0.055)	(0.048)
Female * Secondary pass	204.730***	311.972***	220.495***	0.076*	0.111*	0.207***
	(48.233)	(74.256)	(71.219)	(0.043)	(0.063)	(0.059)
Secondary result: 1st division	78.895	194.424**	6.428	0.167***	0.016	0.382***
	(68.854)	(93.079)	(95.895)	(0.051)	(0.068)	(0.081)
Secondary result: 2nd division	110.160**	109.870	99.023	0.056	-0.000	0.164***
	(50.522)	(68.339)	(72.115)	(0.037)	(0.046)	(0.058)
Inverse Mills ratio	-1,145.381**	-772.252	-1,356.518	-0.767*	-0.351	-1.964**
	(572.115)	(660.639)	(1,073.254)	(0.413)	(0.420)	(0.853)
Constant	2,710.602***	2,306.489***	3,186.124***	3.515***	3.185***	4.331***
	(383.722)	(459.471)	(649.543)	(0.278)	(0.291)	(0.532)
Effect of Technical stream on females (λ + π)	125.174**	22.953	218.147***	0.032	-0.144	0.157**
	62.816	116.542	77.435	0.062	0.116	0.075
Observations	36,646	25,325	11,321	36,646	25,325	11,321
R-squared	0.232	0.249	0.209	0.238	0.276	0.222
Years of education dummies	Yes	Yes	Yes	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of households (fixed effects)	25,432	17,038	8,394	25,432	17,038	8,394

This regression considers salaried/casual wage employees in the age-group of 25-60 years. Bootstrapped standard errors clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A8: Robustness of results to inclusion of additional control variables capturing women's personality traits

	Labour force participation			Log of annual earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
	Overall	Rural	Urban	Overall	Rural	Urban
Female	-0.619***	-0.587***	-0.699***	-0.929***	-0.942***	-1.436***
	(0.012)	(0.016)	(0.020)	(0.123)	(0.130)	(0.295)
Female * Secondary pass * Technical stream (λ)	0.074***	0.094***	0.082***	0.098	-0.032	0.239**
	(0.018)	(0.036)	(0.022)	(0.094)	(0.183)	(0.117)
Secondary pass * Technical stream (π)	0.008	0.000	-0.003	-0.025	-0.015	-0.021
	(0.009)	(0.015)	(0.012)	(0.049)	(0.089)	(0.060)
Female * Secondary pass	-0.063***	-0.074***	-0.007	0.331***	0.414***	0.448***
	(0.010)	(0.015)	(0.014)	(0.066)	(0.111)	(0.089)
Secondary result: 1st division	0.027**	-0.005	0.049***	0.227***	0.241**	0.399***
	(0.014)	(0.020)	(0.019)	(0.081)	(0.118)	(0.115)
Secondary result: 2nd division	0.003	-0.021	0.024	0.151**	0.132	0.244***
	(0.011)	(0.015)	(0.015)	(0.060)	(0.088)	(0.080)
Index of intra-household decision making power	0.013***	0.015***	0.010***	0.011*	0.016**	0.002
	(0.001)	(0.002)	(0.002)	(0.007)	(0.007)	(0.019)
Index of social and political participation	0.043***	0.051***	0.022***	0.026*	0.040**	-0.009
	(0.003)	(0.003)	(0.004)	(0.014)	(0.017)	(0.019)
Inverse Mills ratio				-0.921	-1.116	-2.791**
				(0.721)	(0.825)	(1.352)
Constant	1.064***	1.082***	1.039***	10.955***	10.916***	12.603***
	(0.024)	(0.031)	(0.040)	(0.471)	(0.567)	(0.824)
Effect of Technical stream on females ($\lambda + \pi$)	0.082***	0.094***	0.079***	0.073	-0.047	0.219**
	0.018	0.034	0.021	0.089	0.173	0.108
Observations	69,565	45,287	24,278	34,158	23,627	10,531
R-squared	0.685	0.666	0.729	0.371	0.383	0.371
Years of education dummies	Yes	Yes	Yes	Yes	Yes	Yes
Other individual covariates	Yes	Yes	Yes	Yes	Yes	Yes
Household fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of households (fixed effects)	36,769	23,956	12,813	24,525	16,451	8,074

Robust standard errors (bootstrapped for columns 4-6) clustered at the household level are given in parentheses. *** p<0.01, ** p<0.05, * p<0.1