

Osikominu, Aderonke; Grossmann, Volker; Osterfeld, Marius

Conference Paper

Are Sociocultural Factors Important for Studying a Science University Major?

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik
- Session: Education II, No. B18-V1

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Osikominu, Aderonke; Grossmann, Volker; Osterfeld, Marius (2014) : Are Sociocultural Factors Important for Studying a Science University Major?, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Education II, No. B18-V1, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/100404>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Are Sociocultural Factors Important for Studying a Science University Major?

Volker Grossmann,^{*}Aderonke Osikominu,[†]and Marius Osterfeld[‡]

February 14, 2014

Abstract

This paper examines the role of the sociocultural background of students as measured by both religious denomination and political attitudes towards gender equality and science-related issues in their home environment for choosing STEM fields in university. Our empirical estimates are based on a structural Roy model which accounts for differences in costs (distance to the next technical university) and earnings across majors as well as for selection bias. Our findings suggest that male students from conservative and Catholic municipalities are more likely to study a STEM field, whereas the sociocultural background plays little role for the major choice of females.

Key words: Gender differences; Selection bias; Sociocultural environment; STEM fields.

JEL classification: I20, C81.

^{*}University of Fribourg, Switzerland; CESifo, Munich; Institute for the Study of Labor (IZA), Bonn. Postal address: University of Fribourg, Department of Economics, Bd. de Pérolles 90, CH-1700 Fribourg. E-mail: volker.grossmann@unifr.ch.

[†]Corresponding author. University of Hohenheim, Germany; CESifo, Munich; Institute for the Study of Labor (IZA), Bonn. Postal address: University of Hohenheim, Department of Economics (520B), D-70593 Stuttgart. E-mail: a.osikominu@uni-hohenheim.de.

[‡]University of Fribourg, Switzerland. Postal address: University of Fribourg, Department of Economics, Bd. de Pérolles 90, CH-1700 Fribourg. E-mail: marius.osterfeld@unifr.ch.

1 Introduction

Modern growth theory suggests that economic growth is ultimately driven by technological progress. In fact, empirical evidence strongly supports the hypothesis that cross-country differences in per capita income is primarily driven by differences in total factor productivity (e.g., Jones and Romer, 2010). Technological change, in turn, relies on innovations which are created primarily by scientists and engineers. There is consequently a widespread consensus that the so-called *STEM* (science, technology, engineering, and mathematics) skills are of major importance to sustain innovation and growth. Recent evidence also suggests that college graduates in STEM fields exert much larger human capital externalities to local labor markets than other college graduates (Winters, 2013). According to OECD (2013; Tab. C3.3a), however, the average share of new entrants into STEM fields in OECD countries in the year 2011 was only about 25 percent. There is a striking and widely recognized gender difference. Whereas 39 percent of male students choose a STEM field, only 14 percent of female students do so (OECD, 2013; Tab. C3.3b).

A recent literature discusses the role of cultural factors for economic growth (e.g., Barro and McCleary, 2003, 2006; Guiso, Sapienza and Zingales, 2006; Tabellini, 2008; Becker and Woessmann, 2009). In light of the salient role of STEM fields for R&D-based growth, an interesting hypothesis is that a potentially important mechanism for the culture-growth relationship works through the university major choice. It is interesting to know, for instance, to which degree gender- and science-related progressive versus conservative political attitudes and religious denomination affect the decision of students to choose a STEM field. At the first glance, one may think that a progressive environment contributes to developing a taste for science. Alternatively, however, STEM fields may preferred in conservative environments vis-à-vis, for instance, social sciences which may be considered to be oriented to left-wing political attitudes. One may also hypothesize that conservatism is related to a low fraction of females choosing a STEM field by creating an environment where certain fields and occupations are perceived as being typically male or female. Interestingly, however, this hypothesis is

at odds with evidence at the macro level. In fact, the fraction of females choosing a STEM field in religiously conservative countries like Turkey and Saudi Arabia, is very similar to, say, Germany, the UK or Scandinavian countries, which are known to emphasize gender equality.¹

This paper proposes a structural framework which allows us to examine if and how sociocultural factors determine the probability that a student chooses a STEM field in tertiary education and to what extent they explain gender differences in the major choice. We exploit a unique micro data set which contains rich panel information from Swiss Graduate Surveys five years after graduation from college. The variables include earnings of a graduate, parental education, age, gender, and the municipality in which a graduate lived before entering college (referred to as "home municipality").

Our goal is to account for differences between STEM fields and other university majors in both individual tastes and pecuniary benefits/costs. We develop a theoretical model to motivate a structural econometric framework. Individuals value material consumption, related to earnings over the life-cycle, and directly derive utility from their field of study. Tastes for study majors depend on the sociocultural environment and the geographical proximity of a student's home municipality to the next technical university, given the distance to the next university. Geographical distance between the home municipality and the location of universities is truly exogenous to an individual, being salient for identification. It determines differences in pecuniary and non-pecuniary costs of studying across fields.

The main pecuniary motivation for university major choice are earnings. Thus, we need to account for earnings differences across majors. However, we observe earnings of a university graduate only for the field he or she self-selected into. To address the potentially important selection bias, we construct the *expected* earnings difference across fields by alluding to the notion that rational individuals choose their major according to comparative advantage. We borrow from the literature on the choice of college participation, which is plagued by similar self-selection problems as the university major

¹The fraction of female students choosing a STEM field is 17 percent in Turkey, 16 percent in Sweden, 15 percent in Saudi Arabia and Germany, 14 percent in Finland, and 12 percent in Sweden and the UK (OECD, 2013; Tab. C3.3b).

choice.²

The main innovation of our study is to capture the sociocultural background of a student in the home municipality, based on a structural Roy model. We exploit the unique opportunity of the direct democratic system in Switzerland to characterize the environment with respect to conservative or progressive attitudes on gender equality and science-related issues by results from nation-wide referenda. We select the following referenda which took place in the period between 1981 and 2005 in order to estimate their effects on the probability to study a STEM field in university: on introducing equal rights of men and women in the constitution (referendum held in 1981), on providing addicts with medical prescriptions of heroin (1999), on the regulation of stem cell research (2004), and on the civil union of homosexual couples (2005), providing similar rights than to married couples (except for the right to adopt children and access to in vitro fertilization). We consider the latter referendum to be science-related because a large body of research has severely questioned the argument typically put forward by the religiously conservative that homosexuality is "unnatural". Scientific research which has contradicted this notion has largely affected the public discussion about equal rights for homosexual partnerships. Also the referendum on novel ways to cope with criminal activity of heroin addicts is an example of science-based changes in political attitudes. In addition to referenda results, we measure the impact of the fraction of votes which accrued to left-wing parties at the national parliamentary elections of the year 1995, roughly the year most graduates in our sample started their university studies, on major choice. Moreover, we examine the role of the distribution of religious denomination in a municipality for major choice. A principal component analysis suggests that results of the four referenda we consider and the vote share for left-wing parties are primarily loaded on the first principal component, which may therefore be interpreted as measuring progressive political attitudes. Religious denomination is primarily loaded on the second principal component.

Our main results are as follows. First, male students from municipalities with a

²Seminal papers include Heckman (1976) and Willis and Rosen (1979). See Heckman, Lochner and Todd (2006) for a survey.

high fraction of voters in favor of gender equality, the science-related issues and left-wing parties are significantly less likely to enter a STEM field in university, whereas in municipalities with a high share of Catholics, students are more likely to enter a STEM field. Thus, somewhat surprisingly, our results suggest that the probability of male students in tertiary education to choose a STEM field is higher when a student is socialized in a more conservative or Catholic background.³ However, in none of our specifications we find that women in a conservative or Catholic environment are less likely to study a STEM major than women in a progressive environment. In fact, sociocultural factors apparently have negligible effects for female students. As they matter for the major choice of men to a non-negligible extent, those factors contribute to understanding gender differences in the decision to study a STEM field. Finally, consistent with previous research, our results also suggest that female students are less motivated than men by earnings differences across STEM fields and other majors.

The paper is organized as follows. Section 2 discusses related literature. In section 3 we present theoretical considerations which motivate the structural econometric model. The econometric model and the identification strategy is discussed in section 4. Section 5 presents the data. Section 6 shows the results. The last section concludes.

2 Related Literature

Our paper contributes to the literature on college major choice by employing a unique data set which allows us to examine the role of sociocultural indicators based on a structural model. The main determinants of college major choice suggested by the previous literature were quantitative abilities (e.g., Arcidiacono, 2004; Wang, 2013), parental background (e.g., Boudarbat and Montmarquette, 2009; Sonner, 2009), expectations on future labor force participation (e.g., Polachek, 1978; Blakemore and Low, 1984), and lifetime earnings (e.g., Berger, 1988; Eidea and Waehrer, 1998). Arcidiacono, Hotz

³In Switzerland, Catholics were typically associated with more conservative religious values compared to Protestants, which has been attributed to the Reformation process in Switzerland itself (e.g., Gordon, 2002). The relation between conservatism and Catholicism is not so clear, however, for modern times (Altermatt, 1979).

and Kang (2010) use surveys to elicit measures of expectations and counterfactuals. They argue that, by accounting for students' expectations, ability and expected earnings are important determinants of college major choice. Boudarbat (2008) argues that lifetime earnings are more important if students have gained work experience before attending college. Moreover, Boudarbat and Montmarquette (2009) find that lifetime earnings seem to play a smaller role for the college major choice if the parent of the same gender as the student enjoyed university education. In our sample, the education of parents does not affect college major choice.

With respect to gender differences in college major choice, Zafar (2009) finds that the gender gap can partly be attributed to a higher emphasis of men on pecuniary outcomes (see also Montmarquette, Cannings and Mahseredjian, 2002) and higher emphasis of women on enjoying coursework and employment in potential jobs, and gaining the approval of parents. Our results are consistent with the hypothesis that earnings are less important for men than for women. Carell, Page and West (2010) find that the gender of the professor plays a role for both females' performance in basic math and science classes and the choice of women with high quantitative skills to graduate from a STEM field. Their results suggest that role models are important for females with respect to study STEM fields, but not for males. Kane and Mertz (2011) show that gender equality as measured by gender gaps in income and political participation is positively correlated with math performance of both male and female high-school pupils. However, gender inequality cannot explain the gender gap in math performance. None of these studies examine the question whether gender differences in major choice can be attributed to differential effects of sociocultural factors across genders.

3 Theoretical Considerations

We first develop a simple theoretical model of individual university major choice in order to motivate our structural estimation approach. We focus on the binary decision to study a STEM field (alternative A) or a non-STEM field (alternative B). This

allows us to employ an identification strategy which is inspired by the literature on the (binary) college participation decision.

3.1 Set Up

Consider an infinitively-living individual i which chooses university major $j \in \{A, B\}$ in period 0 and earns wage income Y_{ijt} in period $t = 1, 2, 3, \dots$ thereafter.

For simplicity and following standard arguments, suppose individuals cannot borrow against future income while attending university in period 0. We assume that, in period 0, individual i possesses and uses resources y_i^0 for consumption.⁴ Moreover, suppose the growth rate of wage income in period $t \geq 1$ is time-invariant and independent of the university major (otherwise income differences across majors would diverge infinitely), but possibly is individual-specific. Let g_i denote the wage growth rate for individual i .⁵ Moreover, let us denote the earnings of individual i in period 1, obtained initially after finishing university studies in field $j \in \{A, B\}$, by y_{ij} . Thus, earnings in period $t \geq 1$ read

$$Y_{ijt} = y_{ij}(1 + g_i)^{t-1}. \quad (1)$$

Initial earnings y_{ij} are given by some major-specific function f_j , which depends on a vector of observable characteristics, \mathbf{X}_i , and an individual- and major-specific (unobservable) "ability" component, ξ_{ij} , which is unobservable for an econometrician. We specify

$$y_{ij} = f_j(\mathbf{X}_i, \xi_{ij}) = \xi_{ij} \exp(\mathbf{X}_i' \boldsymbol{\beta}_j + \kappa_j), \quad (2)$$

where $\boldsymbol{\beta}_j$ is a parameter vector and κ_j is a major-specific shift parameter. Defining

⁴Sources of income in period 0 could be income from a sideline during study, scholarships, transfers by parents or prior savings. Individuals may save in period 0 but provided that future income is sufficiently higher than available resources in period 0, which we implicitly assume, they optimally choose zero savings.

⁵We do not have good data on the growth rate of wage income. We observe individual earnings one year and five years after graduation from university. As some graduates acquired post-graduate education in-between, we shall not attempt to compute the growth rate of earnings over the life-cycle of a worker based on these observations. Fortunately, as will become apparent, g_i will not affect the university major choice as long as we assume that earnings growth does not depend on field j .

$u_{ij} \equiv \log \xi_{ij} + \kappa_j$, we can write

$$\log y_{ij} = \mathbf{X}_i' \boldsymbol{\beta}_j + u_{ij}, \quad j \in \{A, B\}. \quad (3)$$

(3) gives us a familiar linear form for log earnings of an individual, which in our context captures that earnings may depend on the university major chosen, possibly in interaction with individual ability.

In periods $t \geq 1$, individuals can freely borrow and lend at the time-invariant (world market) interest rate r . For technical reasons, suppose that r exceeds the growth rate of earnings g_i for all i . Individuals draw utility from their consumption stream. Importantly, each individual i also draws utility h_{ij} from studying (or graduating from) major j . Total intertemporal utility of individual i when choosing field j is given by

$$U_{ij} = \sum_{t=0}^{\infty} \rho^t \log c_{ijt} + h_{ij}, \quad (4)$$

where $\rho \in (0, 1)$ is the discount rate and c_{ijt} denotes the level of material consumption in t of an individual i which has chosen major j . We assume that h_{ij} depends on observable characteristics, \mathbf{Z}_i of individual i and on an unobservable "taste" component, τ_{ij} , of i for major j . We specify the linear form

$$h_{ij} = \mathbf{Z}_i' \boldsymbol{\gamma}_j + \tau_{ij}, \quad (5)$$

where $\boldsymbol{\gamma}_j$ is a parameter vector.

Our empirical identification strategy, explained in section 4, requires that \mathbf{Z}_i contains variables which are not included in \mathbf{X}_i . Taste characteristics in \mathbf{Z}_i which determine utility from studying a certain major but do not shape earnings differentials across majors may be thought of (i) the distance of the home municipality to the next technical university (given the distance to the next university), which is related to the costs of studying a STEM field, and (ii) the sociocultural environment in the home municipality at the time the major is chosen.

3.2 Consumption Profile

After choosing a major j , individual i solves the following maximization problem which smooths consumption over time:

$$\max_{\{c_{it}\}} \sum_{t=1}^{\infty} \rho^t \log c_{ijt} \text{ s.t. } \sum_{t=1}^{\infty} \frac{c_{ijt}}{(1+r)^t} = \sum_{t=1}^{\infty} \frac{Y_{ijt}}{(1+r)^t}. \quad (6)$$

This leads to the well-known Euler equation

$$c_{ijt+1} = (1+r)\rho c_{ijt}, \quad (7)$$

$t \geq 1$. Using (7) in (4) and observing that $c_{ij0} = y_i^0$ for all j , intertemporal utility of individual i conditional on major choice j reads as⁶

$$U_{ij} = \frac{\rho}{1-\rho} \log c_{ij1} + y_i^0 + \left(\frac{\rho}{1-\rho} \right)^2 \log[(1+r)\rho] + h_{ij}. \quad (8)$$

Combining (7) with the intertemporal budget constraint in (6) and using expression (1) for earnings Y_{ijt} we obtain the (optimal) level of consumption of individual i under college major choice j in the first working period ($t = 1$). Observing that $r > g_i$ by assumption, it is given by

$$c_{ij1} = \frac{(1-\rho)(1+r)}{r-g_i} y_{ij}. \quad (9)$$

3.3 Major Choice

Substituting both (5) and (9) for $j \in \{A, B\}$ into (8), we find that the difference in utility between studying a STEM major (alternative A) and a non-STEM field (alternative B) is given by

$$\Delta U_i = \alpha_1 [\log y_{iA} - \log y_{iB}] + \mathbf{Z}'_i \boldsymbol{\alpha}_z + \nu_i, \quad (10)$$

⁶Use $\sum_{t=0}^{\infty} \rho^t = (1-\rho)^{-1}$ and $\sum_{t=1}^{\infty} t\rho^t = \rho(1-\rho)^{-2}$.

where $\alpha_1 \equiv \frac{\rho}{1-\rho}$, $\alpha_z \equiv \gamma_A - \gamma_B$ and $\nu_i \equiv \tau_{iA} - \tau_{iB}$. As an implication of the logarithmic form of instantaneous utility from material consumption the expression for utility difference ΔU_i does neither depend on the growth rate of earnings, g_i , nor on the interest rate, r . Moreover, because income (and consumption) while attending university, y_i^0 , is independent of the major choice, it cancels out.

Substituting the expression for $\log y_{ij}$ in (3) for $j \in \{A, B\}$ into (10), we can thus write

$$\Delta U_i = \mathbf{W}'_i \boldsymbol{\pi} - \varepsilon_i, \quad (11)$$

where $\mathbf{W}'_i \equiv \left[\mathbf{X}'_i \quad \mathbf{Z}'_i \right]$, $\boldsymbol{\pi}' \equiv \left[\alpha_1(\boldsymbol{\beta}_A - \boldsymbol{\beta}_B)' \quad \boldsymbol{\alpha}'_z \right]$ and $-\varepsilon_i \equiv \alpha_1(u_{iA} - u_{iB}) + \nu_i$. Clearly, individual i prefers alternative A (STEM major) to B if $\Delta U_i > 0$, i.e., if $\mathbf{W}'_i \boldsymbol{\pi} > \varepsilon_i$.

Willis and Rosen (1979) derived expressions for ΔU_i which are formally equivalent to (10) and (11). Whereas they analyzed the choice of college participation, our framework captures the choice of the university major given that an individual goes to university. Willis and Rosen (1979) did not consider consumption smoothing and utility from attending university. In their theoretical model, family background enters by affecting the individual discount rate. The (log) linear forms in their model analogous to (10) and (11) are implied by a first-order Taylor-approximation.

4 Econometric Model

We now present the econometric model.

4.1 Selection Bias and Estimation

Ideally, we would like to estimate the probability that a STEM field is chosen, $\text{prob}\{\Delta U_i > 0\}$, on basis of (10); that is,

$$\text{prob}\{\Delta U_i > 0\} = \text{prob}\{\alpha_1[\log y_{iA} - \log y_{iB}] + \mathbf{Z}'_i \boldsymbol{\alpha}_z + \nu_i > 0\}. \quad (12)$$

However, we observe wages of an individual only for the major which has been actually chosen. Thus, we have to estimate wage equations (3) by correcting for selection bias and then plug the estimates for $\log y_{iA}$ and $\log y_{iB}$ into (12).

4.1.1 Stage 1

We assume that the error terms u_{iA} , u_{iB} in wage equation (3) and ε_i in expression (11) have mean zero and standard deviation σ_A , σ_B and σ_ε , respectively. In a first step, we estimate a “reduced form” probit function based on (11). We suppose that

$$\text{prob}\{\Delta U_i > 0\} = \text{prob}\{\mathbf{W}'_i \boldsymbol{\pi} > \varepsilon_i\} = \Phi\left(\mathbf{W}'_i \boldsymbol{\pi} / \sigma_\varepsilon\right), \quad (13)$$

where Φ denotes the standard normal c.d.f.; its p.d.f. is denoted by φ .

4.1.2 Stage 2

From (3) and (11), we find that the expectation of log wage income resulting from alternative A (STEM field) of an individual is conditional on the self-selection to that field and given as follows (see Willis and Rosen, 1979):

$$\begin{aligned} E(\log y_{iA} | \Delta U_i > 0) &= \mathbf{X}'_i \boldsymbol{\beta}_A + E(u_{iA} | \mathbf{W}'_i \boldsymbol{\pi} > \varepsilon_i) \\ &= \mathbf{X}'_i \boldsymbol{\beta}_A + \frac{\text{cov}(u_{iA}, -\varepsilon_i)}{\sigma_\varepsilon} \lambda_{iA}, \end{aligned} \quad (14)$$

where $\text{cov}(u_{iA}, -\varepsilon_i)$ denotes the covariance between u_{iA} and $-\varepsilon_i$ and we defined the inverse mills ratio

$$\lambda_{iA} \equiv \frac{\varphi(\mathbf{W}'_i \boldsymbol{\pi} / \sigma_\varepsilon)}{\Phi(\mathbf{W}'_i \boldsymbol{\pi} / \sigma_\varepsilon)}. \quad (15)$$

Analogously, we have

$$E(\log y_{iB} | \Delta U_i < 0) = \mathbf{X}'_i \boldsymbol{\beta}_B + \frac{\text{cov}(u_{iB}, \varepsilon_i)}{\sigma_\varepsilon} \lambda_{iB}, \quad (16)$$

where

$$\lambda_{iB} \equiv \frac{\varphi(\mathbf{W}'_i \boldsymbol{\pi} / \sigma_\varepsilon)}{1 - \Phi(\mathbf{W}'_i \boldsymbol{\pi} / \sigma_\varepsilon)}. \quad (17)$$

Let $\hat{\boldsymbol{\pi}}$ and $\hat{\sigma}_\varepsilon$ denote the estimates of $\boldsymbol{\pi}$ and σ_ε from estimating (13) at stage 1, respectively. Let $\hat{\lambda}_{iA}$ and $\hat{\lambda}_{iB}$ denote the resulting values for λ_{iA} and λ_{iB} when evaluating the right-hand sides of (15) and (17), respectively, at $\hat{\boldsymbol{\pi}}$ and $\hat{\sigma}_\varepsilon$. As suggested by (14) and (16), at stage 2, we estimate wage equations

$$\log y_{iA} = \mathbf{X}'_i \boldsymbol{\beta}_A + \delta_A \hat{\lambda}_{iA} + \eta_{iA}, \quad (18)$$

$$\log y_{iB} = \mathbf{X}'_i \boldsymbol{\beta}_B + \delta_B \hat{\lambda}_{iB} + \eta_{iB}, \quad (19)$$

where δ_A, δ_B denote the coefficients on correction terms $\hat{\lambda}_{iA}$ and $\hat{\lambda}_{iB}$, respectively, and η_{iA}, η_{iB} are standard error terms. Suppose, for instance, that the estimate of δ_A is positive (and significant). According to (14), we then conclude that (unobserved) parameter u_{iA} which affects earnings of individual i when graduating from major A, is positively related to parameter $-\varepsilon_i = \alpha_1(u_{iA} - u_{iB}) + \tau_{iA} - \tau_{iB}$ which consists of (unobserved) differences in ability and tastes of i between major A and the alternative major B. A positive estimate for δ_B has the analogous interpretation of a positive relation between u_{iB} and ε_i . In these cases, there is selection bias in the sense that individuals choose according their comparative advantage for a major.

As shown by Heckman (1976), we obtain consistent estimates $\hat{\boldsymbol{\beta}}_A$ and $\hat{\boldsymbol{\beta}}_B$ for coefficients $\boldsymbol{\beta}_A$ and $\boldsymbol{\beta}_B$, respectively. The estimated OLS standard errors of the estimated coefficients in (18) and (19) are incorrect, however, when including $\hat{\lambda}_{iA}$ and $\hat{\lambda}_{iB}$ constructed based on the first-stage estimates as regressors in stage two. Therefore, we bootstrap the full three-step estimation procedure using 499 bootstrap replications. Specifically, we apply the weighted bootstrap suggested by Barbe and Pertail (1995). For each person in our dataset we generate 499 weights based on random draws from a gamma distribution with shape and scale parameters equal to one. Thus, the bootstrap weights are non-integer and the probability that a weight exactly equals zero is zero. With a binary dependent variable and a number of discrete regressors, this

bootstrap procedure has the advantage that we avoid having to repeat the sampling if in a given resample the maximum likelihood estimation fails to converge or certain covariate settings perfectly predict the dependent variable (see also Fitzenberger and Muehler, 2011, for a similar argument).

4.1.3 Stage 3

In the final stage 3 (major choice), we estimate the structured probit equation (12) by replacing $\log y_{iA} - \log y_{iB}$ by

$$\log \hat{y}_{iA} - \log \hat{y}_{iB} = \mathbf{X}'_i(\hat{\boldsymbol{\beta}}_A - \hat{\boldsymbol{\beta}}_B). \quad (20)$$

Again, standard errors of all coefficients need to be corrected by bootstrapping because of the inclusion of an estimated term.

4.2 Identification

4.2.1 Earnings Regressions

According to (2) in the theoretical framework, variables in \mathbf{X}_i are those which potentially affect the earnings of an individual i early in the professional career (time $t = 1$) given the major choice (variable *LogEarnings*). They enter the estimated equations (18) and (19) at the second stage. (As usual, we also include a constant term.) We employ a dummy variable which indicates gender (equal to 1 if i is female) to capture potential discrimination of women in the labor market (variable *Female*). Moreover, we control for the age (in logs) of an individual at the time of the survey (five years after graduating from university) to capture work experience (variable *LogAge*). Some older graduates may have received vocational training. Some may have gained work experience prior or during attending university, the latter possibly prolonging their study duration to the benefit of higher (initial) earnings. We also account for the fact whether an individual has participated in a post-graduate education program for a period of at least six months (dummy variable *PostEdu*). We expect individuals who

have participated in such a program to earn significantly less than those who have not because, for a given age, they are likely to have shorter work experience at the time of observation. For instance, the work experience five years after receiving a master degree of an individual which participated in a doctoral program is relatively low (often literally zero). Even if the growth rate of income, g_i , of such an individual i could be higher on average than for individuals without a doctorate but five years of work experience,⁷ initial income y_{ij} for any major j is typically lower. Failing to account for this fact could heavily confound our second-stage estimates (earnings regressions), from which we construct the (log) earnings differences (20). These are denoted by *LogEarnDiff* and enter at the third stage of the estimation procedure (major choice). Finally, we also include a dummy variable indicating whether at least one parent attended university (*ParentUni*). We thereby account for a possible intergenerational transmission of cognitive ability.⁸ Note, however, that the education of parents may be less important for success in the labor market within the group of university graduates as compared to the whole population.

4.2.2 Exclusion Restrictions and Major Choice

Identification requires that we find convincing exclusion restrictions. That is, we need variables which are included in taste parameters \mathbf{Z}_i (affecting major choice) but not in characteristics \mathbf{X}_i (affecting earnings capability) of a student i . Moreover, our structural estimation approach dictates that only those variables enter \mathbf{Z}_i which are known to an individual i at the time of major choice. We thus employ *Female* as only variable (in addition to the constant) which enters both stage 2 and stage 3 estimations.

Sociocultural Characteristics and Gender Differences Our first identifying assumption is that variables which capture the sociocultural background of students affect

⁷In fact, our data set, graduates with good grades are more likely to have participated in a post-graduate program. Recall that g_i does not enter the estimated equations according to our structural model such that we do not have to observe it.

⁸As robustness checks, we included the education of the mother and father separately and constructed various indicators of their level of education. We found no important differences in the estimation results compared to the parental education measure we use to reach the reported results.

the decision whether to study a STEM field but do not affect the differences of (log) earnings across majors after graduating from university. We include in \mathbf{Z}_i the respective shares of yes-votes in the "home" municipality a college graduate i lived before going to university of a referendum held in 1981 on introducing equal rights of men and women in the constitution (variable *GenderEqu*), on providing drug addicts with medical prescriptions of heroin to addicts held in 1999 (variable *Heroin*), on regulating stem cell research held in 2004 (variable *StemCells*), and on a civil union of homosexual couples held in 2005 (variable *HomoCouples*). As a different kind of political indicator, we also employ the support for left wing parties in the election for the National Parliament of 1995 (variable *LeftWing*). We also aim to capture the cultural background by the share of Catholics in a graduate's home municipality (*ShareCath*). Although Catholicism and political attitudes may have been related to cognitive skills in the 19th century (e.g. Becker and Woessmann, 2009; Boppart et al., 2013), both kind of variables are unlikely to affect contemporaneous differences in the individual earnings potential across study fields.

As reported in an online appendix, a principal component analysis with these indicators suggests that the referenda results and the left-wing vote share can be attributed to the first principal component. It is therefore reasonable to interpret these variables as indicators of progressive political attitudes on gender equality and science-related issues. The first principal component is denoted by *FirstComp*. Religious denomination, by contrast, is more highly related to the second principal component, *SecondComp*.

In order to examine whether these sociocultural characteristics contribute to explaining gender differences with respect to choosing a STEM field in university, in an extension of our basic major choice probit estimations, we include interaction effects between sociocultural characteristics and gender (variable *GenderEqu* \times *Female* etc.). Moreover, we would like to check the well-known hypothesis that female students are less motivated by earnings than males. We therefore include an interaction effect at stage 3 between the (constructed) earnings difference across fields as given by (20) and gender, variable *LogEarnDiff* \times *Female*.

Geographical Characteristics Our second identifying assumption is that the distances of the home municipality to the next technical and the next general university, respectively, do not affect earnings capability for a given major choice. We take logs to capture that the marginal impact of an additional kilometer on major choice is decreasing with distance. Variables are denoted by *LogDistETH* and *LogDistNextUni*, respectively. The technical schools, called "Eidgenössische Technische Hochschulen" (ETH), are the only federal universities in Switzerland. They are located in Lausanne and Zurich. All other universities are financed at the cantonal (i.e., state) level.

Majority Language Finally, in Switzerland, the motivation to study a STEM field may depend on the majority language (German, French, Italian)⁹ of an individual's home environment. For instance, regional differences in institutions and industry structure may be affected by trade relations of Swiss regions to neighboring countries with a common language. Of course, language may also be perceived as cultural characteristic. It is not our goal, however, to generalize its impact on major choice beyond Switzerland but rather include it to mitigate omitted variable bias in the Swiss context. We include dummy variables *French* and *Italian*, i.e., German as majority language is the left-out category.

5 Data

5.1 Individual Characteristics

We employ the 'Swiss Graduate Survey' of the Federal Statistical Office of Switzerland (FOS), a unique dataset on the cohorts of 2001 and 2003 who earned a university diploma. At that time, there was no differentiation between a bachelor and a master degree, such that the diploma was the first graduate degree from university. According to the length of study and exigencies, it was comparable to a master degree nowadays.

⁹In Switzerland, Rhaeto-Romansh is the fourth official language. Very few graduates come from municipalities with Rhaeto-Romansh as the majority language, however. In order not to be able to infer from our data the identity of a graduate, we were obliged to remove these individuals from the dataset.

All graduates of Swiss universities were requested to fill out a questionnaire one year after graduation. If they agreed, the graduates were surveyed again five years after finishing their studies. The survey contains a large array of individual characteristics including labor income and working time five years after graduation, major at university, gender, the level of education of mother and father, and the home municipality before entering university (see e.g. FOS, 2008, 2009, 2012). Since our theoretical model implies that expected earning differences between fields of study may have an important impact on study major choice, the availability of individual earnings several years after graduation is crucial for our estimation strategy. As a consequence, we restrict our dataset to participants of the second poll, which includes earnings five years after graduation. Attendance was voluntary and the response rate was about 60 percent in the first poll and 65 percent in the second poll. The ‘Swiss Information System of Universities’ provides information on the distribution of the graduates across the different universities, majors and cantons of origin which allows us to correct for different response behavior across universities, fields of study, and regions (see FSO, 2012). We do this by employing the probability weights provided by the Federal Statistical Office of Switzerland (FSO, 2008, 2009).

5.2 Sociocultural and Geographical Characteristics

For our analysis, a salient piece of information in the ‘Swiss Graduate Survey’ is the residence of each university graduate before entering university. We draw on this variable to characterize a student’s sociocultural background at the time of major choice. According to our structural approach, these characteristics enter at the third but not at the second stage of the estimation procedure.

First, we characterize the majority language and the religious environment (share of Catholics) of a graduate in the home municipality from the ‘Swiss Census’ in 1990. The ‘Swiss Census’ takes place every 10 years. It is a complete population survey, aimed at revealing the demographic, economic, social and cultural structure of Switzerland and its development over time. Every household in Switzerland receives a household

specific questionnaire and individual specific questionnaires for each person living in the household. All households and individuals in Switzerland are obliged by law to respond (see FSO, 1996; FSO, 2005).

Second, the distances from the center of the home municipality to the next technical school, ETH Zurich or ETH Lausanne, and the distance to the next university are constructed from Google Maps.

Third, the total vote share which accrued to left-wing parties in the Swiss national election 1995 and the results from the nationwide referenda to capture how progressive views were on gender equality and science-related issues in the home environment of students are provided by ‘Swissvotes’. Some details of these referenda are outlined next.

The referendum on gender equality in 1981 intended to give equal rights to women with respect to professional life, family life, and education. A particular focus of the initiative was to ensure by law equal pay for equal work. Opponents feared that the proposal would interfere in wage negotiations and endanger private autonomy of families. The referendum passed in 17 of 26 cantons. The overall support for the proposal was 60.3 percent. A nationwide representative poll revealed that opponents of the initiative, in general, disapproved equal rights for men and women.¹⁰

After the isolation of embryonic stem cells in 1998, advances in stem cell engineering prompted the debate on how to deal with the use of embryonic stem cells legally and ethically. The Swiss government proposed a law allowing scientists to take stem cells only from embryos left over from in vitro fertilization procedures. In 2004, the electorate accepted the law with a majority of 66.4 percent. All cantons voted in favor of the proposal. Polls showed that 40 percent of the opponents said that ethical concerns were the principal reason for voting against the proposal. About 50 percent of adversaries expressed that doubts about the merits of scientific research in general and fears of unwelcome consequences were deciding factors to oppose the law.

In a scientific pilot project set up by the Swiss Government in 1994, drug addicts

¹⁰For such background information and further discussion of the referenda used in our study, see Linder, Bolliger and Rielle (2010).

were entitled to receive heroin from a physician for free. Evaluation of this program suggested positive effects on both the health and the social situation of drug addicts. Subsequently, the Swiss government was seeking to enlarge the set of therapies by a state-controlled distribution of heroin all over Switzerland. Adversaries argued, in particular, that the state would financially and morally support drug addicts by this law. In 1999, the people approved the proposal by a majority of 54.4 percent. Ten cantons refused the law.

Finally, in 2005, a referendum was held on the introduction of equal rights for homosexual couples in civil law. Registered homosexual partnerships were supposed to get the same rights as married couples except for the right to adopt children and access to in vitro fertilization. 58 percent of voters approved the law. In seven of the 26 cantons the majority refused it. A nationwide poll found that voters based their decision on their fundamental conviction whether homosexual partnerships should be legally and socially recognized.¹¹

5.3 Summary Statistics

Tab. 1 provides summary statistics of the employed variables. In order to address a "typical" career in the employed dataset, we leave out data of the top 10 percent of graduates according to their length of study. Still the maximum duration of study in the employed sample is 15 semesters. Moreover, we restrict our attention to individuals who report positive earnings from employment five years after graduating from university.

<Table 1>

¹¹Because of city splits and mergers, the town codes differ between our data sources. By means of the 'Swiss Historical Town Register', we transform these codes to the classification system used in 2010 by the Federal Statistical Office of Switzerland. In the case of a split, we take the observation of the old municipality as value for the new municipalities. In the case of a merger, for voting results, we add up the number of yes-votes over all merged municipalities and divide them by the total number of votes in the merged municipalities to calculate in the share of yes-votes for the new municipality. To calculate the share of the Catholics in the new municipality, we add up the number of Catholics over all merged municipalities and divide them by total population size in the new municipality in the year 1990. We construct the shares of the mother tongue in new municipalities similarly and from this assign dummies for the majority language.

Our dataset contains the number of working hours in addition to monthly wage income. From this, we compute earnings as full-time equivalents. In fact, work contracts in Switzerland are quite flexible with respect to working hours and part-time work is common to various degrees also for men.¹² In order to address potential measurement error, we leave out individuals which reported earnings in the top and bottom percentile of the wage distribution of the raw data as received from the ‘Swiss Graduate Survey’. This leaves us with monthly earnings in the range of CHF 2,012,- and CHF 17,572,-. Median and mean full-time equivalent earnings are about CHF 6,500,- per month. Men earn, on average, about CHF 200,- per month more than women. 30.5 percent of the graduates studied a STEM field, 41.6 percent among men and 17.8 percent among women. Both of these figures are slightly above the OECD average as reported by OECD (2013; Tab. C3.3b) for the year 2011. About 55 percent of the respondents participated in post-graduate education and slightly less than two-fifths of the graduates report having at least one parent with an academic background.

6 Estimation Results

We present two sets of results. First, we show the outcomes of estimations without any interaction effects with gender of sociocultural characteristics and earnings differences across majors. Second, we include these interaction effects to learn about gender differences.

6.1 Results without Interaction Effects

6.1.1 First Stage

We start with the results on the reduced form Probit estimations (13) of stage 1, presented in Tab. 2.

<Table 2>

¹²In our data set, 10.2 percent of men and 26.6 percent of women work less than 35 hours per week. In Switzerland, the weekly full-time work load is 42 hours.

As expected from the summary statistics, women are significantly less likely to study a STEM field. Moreover, the larger the (log) distance to the next technical university, the higher the probability to study a STEM field; coefficients are significantly different from zero at least at the five percent level. The (log) distance to the next university enters positively and significantly, suggesting that the proximity to the next ETH *relative* to the proximity to the next university positively affects the likelihood to choose a STEM major.¹³ As such geographic characteristics should not matter for earnings *per se* and hence can be left out at stage 2 estimations, these results well support our identification strategy.

Columns (1)-(6) of Tab. 2 vary in the sociocultural indicator we control for. Not all of the coefficients are significantly different from zero, but overall they provide a consistent picture. Individuals from more conservative environments are more likely to choose a STEM field. For instance, the higher the vote share for left-wing parties (column (1)) in a municipality and for supporting the referenda on extending gender equality (column (2)) as well as allowing stem cell research (column (3)), the lower the probability that a student chooses a STEM major (with significance at the one percent level). The point estimates of the other coefficients indicate the same conclusion, i.e., a negative sign for supporting heroin distribution to addicts (column (4)) and marriage of homosexuals (column (5)). The share of Catholics enters with a positive but insignificant sign (column (6)).

Instead of using these six sociocultural indicators separately, column (7) enters the two principal components extracted from these indicators. As argued above, the first principal component (*FirstComp*) may well be interpreted as context-specific measure of "progressivism" of the students' political background in their home environment. Consistent with the results from columns (1)-(5), column (7) suggests that a higher degree of context-specific progressivism reduces the probability to study a STEM field.

The other first stage results are less interesting. Students from Francophone environments are considerably less likely to study STEM fields. The result may be related

¹³This is confirmed when we include the log of the ratio of the two distance measures rather than including the measures separately (not reported). We choose to stick to the latter specification to allow for differential effects.

to a stronger pharmaceutical sector in the German-speaking part of Switzerland, for instance. Moreover, STEM students are typically younger than other students five years after graduation, possibly because vocational training is less common for this group. They are also less likely to have post-graduate education. Whether parents have attended university does not seem to be important for major choice. Recall that we include variable *ParentUni* at stage 2 but not at stage 3 of our estimation procedure.

6.1.2 Second Stage

Results of the earnings regressions at stage 2 are given in Tab. 3 and Tab. 4. Tab. 3 relates to STEM fields and Tab. 4 to humanities. The columns correspond to those of the first stage. This is why the estimates for the correction terms $\hat{\lambda}_{iA}$ and $\hat{\lambda}_{iB}$ (constructed from stage 1 estimates to account for selection bias; see (18) and (19)) are different across the columns within Tab. 3 and Tab. 4, respectively. In the tables, correction terms are denoted by *CorrectSTEM* and *CorrectHum* for STEM fields and Humanities respectively. We find that their coefficients are positive and significant at the five percent and one percent level for STEM and non-STEM graduates, respectively. As expected from the proposed Roy model, they suggest that individuals self-select according to their comparative advantage. Females earn about 6-8 percent more than males when graduating from a non-STEM field (Tab. 3). By contrast, there is little evidence for gender earnings differences among STEM university graduates. The point estimate on the *Female* dummy in Tab. 4 is non-significant with a small and mostly negative sign.

<Table 3>

As expected, older (more experienced) graduates earn more. The coefficients on *LogAge* are comparable between the two panels and highly significant. Also as expected, individuals with post-graduate education earn significantly less five years after graduation. If *PostEdu* = 1 (rather than 0), earnings are about 18 percent lower for STEM field graduates and four percent lower for non-STEM field graduates. Interestingly, earnings do not seem to be affected by the education level of parents. This is not

implausible. The education of parents would certainly matter for earnings in a sample with both graduates and non-graduates but there is not much reason for such effect when restricting focus on those who attended university.

<Table 4>

6.1.3 Third Stage

Results for stage 3, presented in Tab. 5, basically confirm those of stage 1. The only difference to Tab. 2 is that at stage 3 we employ our measure of constructed differences in (log) earnings across fields (20) rather than controlling for those variables which affect earnings at stage 2. As expected, an increase in variable *LogEarnDiff* positively affects the probability to choose a STEM field. According to the bootstrapped standard errors, the coefficients are significantly different from zero at the five percent level for all specifications. Also as expected, the coefficient on *Female* is negative and highly significant. The distance measures and sociocultural variables basically keep their significance levels from the stage 1 results and also the sizes of the coefficients are similar in magnitude to Tab. 2.

<Table 5>

We are interested in the quantitative impact of the factors which affect major choice. For this, we first calculate for each individual the effect of a change in a variable on the probability of studying a STEM field by using the point estimates at stage 3; secondly, we average the effects over all individuals.

We find that women are expected to have a 20.5 percentage points lower fraction of STEM field graduates than men. This reflects the widely-discussed substantial gender differences in university major choice we see in most OECD countries, including Switzerland, as also reflected in Tab. 1. (We further explore the gender effect in the next subsection.) Moreover, an increase in (log) earnings differences from the 25th to the 75th percentile of the distribution of *LogEarnDiff* is expected to increase the fraction of STEM major graduates by almost five percentage points. This suggests

that pecuniary incentives are important for major choice. An 25th to 75th percentile increase in $LogDistETH$, which is equal to an increase in the distance to the next technical university from 30.5 km to 98.3 km (see Tab. 1), reduces the expected fraction of those choosing a STEM major by about two percentage points. The (positive) impact of a larger distance to the next university ($LogDistNextUni$), for a given distance to the next ETH, is even somewhat higher in magnitude.

The impact of the sociocultural background of a student as measured by referenda results on gender equality and science-related issues (conservatism versus progressivism) is typically non-negligible. If the support for liberalizing stem cell research rises from 63.4 percent (25th percentile) to 76 percent (75th percentile), then the fraction of STEM field graduates falls by almost four percentage points. Compared to the mean fraction of STEM field graduates of 30.5 percent, this is a sizable effect. It is comparable to the effect of an increase in (log) earnings differences across fields of study from the 25th to the 75th percentile. A similar, negative quantitative effect arises if the vote share at the municipality level in favor of gender equality rises from 56 to 71.6 percent. An increase from the 25th to 75th percentile in vote share for left-wing parties reduces the percentage of those choosing a STEM field by 2.7 points. Somewhat smaller but still negative effects come from similar increases of the vote share in support for a civil union of homosexual couples and heroin distribution to addicts. According to the insignificant point estimates, the reduction in the expected fraction of those choosing a STEM major is 1.8 percentage points in both cases.

Raising the share of Catholics in a municipality from 29.9 to 65.1 percent raises the fraction on STEM graduates on average by a rather moderate 1.5 percent points. The small and insignificant effect may be attributed to the fact that, in Switzerland, Catholics and Protestants contemporaneously do not differ in terms of conservatism than may have been the case in former times (Altermatt, 1979).

6.2 Interaction Effects with Gender

The results presented in Tab. 2-5 do not allow for differential effects of sociocultural characteristics and pecuniary incentives for study major choice between men and women. In this subsection, we include interaction effects with gender.

6.2.1 First Stage

We again start with the results from the reduced form estimation (stage 1). Tab. 6 shows that those variables which are not interacted with dummy variable *Female* enter very similarly as in Tab. 2. Most notably, the distance variables affect major choice significantly in the expected fashion. Graduates from Francophone municipalities are less likely to study a STEM field (compared to students with a German-speaking background). Moreover, STEM field graduates are on average younger and less likely to have participated in post-graduate education. Parental education does not seem to matter.

<Table 6>

The only interaction effect which is significantly different from zero is that between *Female* and the vote share in support of provision of heroin to drug addicts (variable *Heroin*). It suggests that the negative effect of higher support for a progressive view on this issue on the probability to choose a STEM field – as found in column (4) of Tab. 2 – holds for men only. Similarly, the coefficients on the interaction term of *Female* and the support of left-wing parties (column (1)), gender equality (column (2)) and equal rights for homosexuals (column (5)) is positive. The coefficient on *Female* \times *StemCells* (column (3)) is negative but insignificant. Column (6) suggests that the share of Catholics has a positive effect on the probability to study a STEM field for men but not for women; coefficients on *ShareCath* and *ShareCath* \times *Female* have opposite signs, although both being insignificant.

6.2.2 Second Stage

The results at stage 2 are basically identical to those presented in Tab. 3 and 4 and therefore relegated to the online-appendix. They again suggest that students self-select according to their comparative advantage. The coefficients on *LogAge* give us the elasticity of earnings with respect to age. They suggest that an increase in age by 10 percent (i.e., by about three years in view of the median age of 31 years reported in Tab. 1) raises earnings of STEM field graduates by 5.7-6.0 percent and for other graduates by 7.0-7.3 percent, depending on the specification. Again, postgraduate education reduces earnings early in the career, whereas education of parents does not seem to matter for earnings of graduates.

6.2.3 Third Stage

According to Tab. 7, the quantitative effects of distance variables and majority language are similar to those presented in Tab. 5. More interestingly, Tab. 7 provides a consistent picture how sociocultural variables determine the major choice conditional on gender. While coefficients on the vote share for left-wing parties and in support of progressive views on gender equality and science-related issues are negative and significantly different from zero, the interaction effects with *Female* are positive. In contrast to column (3) of Tab. 5, this is now also true for the coefficient on *Female* \times *StemCells*. The effect of Catholicism is again positive but insignificant for men and negligible for women (column (6)). Thus, the results suggest that the sociocultural background matters considerably for males but not for females. This is confirmed by column (7). The effect of first principal component (our measure of progressive values on gender equality and science-related issues) again suggests a negative effect on the probability to study a STEM field for men but not for women. The coefficient on *FirstComp* is negative and significantly different from zero at the one percent level whereas the interaction effect *FirstComp* \times *Female* is positive (but insignificant) and almost similar in magnitude than the coefficient on *FirstComp*.

<Table 7>

Quantitatively, for men, the impact of a 25th to 75th percentile increase in the vote share in favor of progressive changes in context-specific political attitudes reduces the fraction of STEM field graduates significantly by 3.5-5.7 percentage points, depending on the referendum. The corresponding impact of an increase in the vote share accruing to left-wing parties is 4.5 percentage points. By contrast, these variables typically have negligible and insignificant effects for women. Moreover, an increase in the share of Catholics from the 25th to the 75th percentile raises the fraction of STEM field graduates by 2.5 percentage points for men (although the effect is not significantly different from zero) and literally has no impact of the major choice of women.

The effect of pecuniary incentives to study a STEM field is different for men and women, too. For men, the expected increase in the fraction of STEM major graduates from an increase in (log) earnings differences across from the 25th to the 75th percentile of the distribution of *LogEarnDiff* is a sizeable and highly significant 7.2-7.7 percentage points. However, the coefficient of the interaction effect $LogEarnDiff \times Female$ is negative (although non-significantly different from zero). This amounts to a much lower quantitative effect of earnings differences across majors (STEM versus non-STEM) for women of less than 2 percentage points. This effect is not even significantly different from zero. It is thus safe to conclude that pecuniary incentives to study a STEM major matter considerably less for women than for men.

7 Conclusion

We have examined the role of the sociocultural background of students as measured by both religious denomination and political attitudes towards gender equality and science-related issues in their home environment for choosing a STEM field in university. The motivation to focus on the formation of STEM skills was rooted in their salient role for the process of long run economic growth. We exploited regional variation within Switzerland at the municipal level. The unique opportunity for our research comes from the frequently held national referenda in the Swiss direct democratic system. We based the empirical identification on a structural Roy model which accounts

for differences in costs (distance to the next technical university) and earnings across majors as well as for selection bias.

Our findings suggest that male students from conservative and Catholic municipalities are more likely to study a STEM field, whereas the sociocultural background plays little role for the major choice of females. That the sociological background matters in a quite sizable way for men but not for women suggests that it can contribute to understand the widely discussed gender gap in STEM major choice. Consistent with previous studies, we find that female students are less motivated by earnings than men.

Why men from more conservative environments seem to be more motivated to study STEM fields than men from progressive environments certainly deserves further research. One possibility could be that social sciences are generally considered to be oriented to left-wing political attitudes. Differentiating among the non-STEM fields seems welcome to dig deeper in this phenomenon. As this would probably require modelling of major choice among more than two alternatives, identification will be an important challenge which is beyond the scope of the current paper. The differential impact of the sociocultural environment on males and females certainly deserves further attention in future research as well.

References

Altermatt, Urs (1979). Conservatism in Switzerland: A Study in Antimodernism, *Journal of Contemporary History* 14, 581-610.

Arcidiacono, Peter (2004). Ability Sorting and Returns to College Major, *Journal of Econometrics* 121, 343-375.

Arcidiacono, Peter, V., Joseph Hotz and Songman Kang (2010). Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals, *Journal of Econometrics* 166, 3-16.

Barbe, Philippe and Patrice Bertail (1995). *The Weighted Bootstrap*, Springer, New York.

Barro, Robert J. and Rachel M. McCleary (2003). Religion and Economic Growth,

American Sociological Review 68, 760-781.

Barro, Robert J. and Rachel M. McCleary (2006). Religion and Political Economy in an International Panel, *Journal for the Scientific Study of Religion* 45, 149-175.

Becker, Sascha O. and Ludger Woessmann (2009). Was Weber Wrong? A Human Capital Theory of Protestant Economic History, *Quarterly Journal of Economics* 124, 531-596.

Berger, M.C. (1988). Predicted Future Earnings and Choice of College Major, *Industrial and Labour Relations Review* 41 418-29.

Blakemore, A.E. and S.A. Low (1984). Sex Differences in Occupational Selection: The Case of College Majors, *Review of Economics and Statistics* 66, 157-63.

Boudarbat, Brahim (2008). Field of Study Choice by Community College Students in Canada, *Economics of Education Review* 27, 79-93.

Boudarbat, Brahim and Claude Montmarquette (2009). Choice of Fields of Study of University Canadian Graduates: The Role of Gender and their Parents' Education, *Education Economics* 17, 185-213.

Carrell, Scott E., Marianne Page and James E. West (2010). Sex and Science: How Professor Gender Perpetuates The Gender Gap, *Quarterly Journal of Economics* 125, 1101-1144.

Eidea, Eric and Geetha Waehrer (1998). The Role of the Option Value of College Attendance in College Major Choice, *Economics of Education Review* 17, 73-82.

Federal Office of Public Health of Switzerland (1999). Die heroingestützte Behandlung. Argumentarium zur Volksabstimmung über den dringlichen Bundesbeschluss zur ärztlichen Heroinverschreibung (heroingestützte Behandlung) 13. Juni 1999, Bern.

Federal Statistical Office of Switzerland (1996). Evaluation der Volkszählung 1990: Erhebungsmethoden und Alternativen, Neuchâtel.

Federal Statistical Office of Switzerland (2005). Abschlussbericht zur Volkszählung 2000, Neuchâtel.

Federal Statistical Office of Switzerland (2008). Schlüsselkompetenzen der Schweizer Hochschulabsolvent/innen: Thematischer Sammelband mit empirischen Ergebnissen der Absolventenstudie, Neuchâtel.

Federal Statistical Office of Switzerland (2009). Von der Hochschule ins Berufsleben: Erste Ergebnisse der Absolventenbefragung 2007, Neuchâtel.

Federal Statistical Office of Switzerland (2012). Erhebungen, Quellen – Absolventenstudien Hochschulen, Steckbrief, Neuchâtel.

Fitzenberger, Bernd and Grit Muehler (2011). Dips and Floors in Workplace Training: Using Personnel Records to Estimate Gender Differences, ZEW Discussion Paper No. 11-023.

Gordon, Bruce (2002). The Swiss Reformation, Manchester University Press, Manchester, New York.

Guiso, Luigi, Paola Sapienza and Luigi Zingales (2006). Does Culture Affect Economic Outcomes, *Journal of Economic Perspectives* 20, 23-48.

Heckman, James J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement* 5, 475–492.

Heckman, James J., Lance J. Lochner, Petra E. Todd (2006). *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond*, in: Hanushek, E. and F. Welch (eds.), *Handbook of the Economics of Education*, Vol. 1, ch. 7, Elsevier.

Jones, Charles I. and Paul M. Romer (2010). The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital, *American Economic Journal: Macroeconomics* 2, 224–245.

Kane, Jonathan M. and Janet E. Mertz (2011). Debunking Myths about Gender and Mathematics Performance, *Notices of the American Mathematical Society* 59, 10-21.

Montmarquette, Claude, K. Cannings, and S. Mahseredjian (2002). How Do Young People Choose College Majors?, *Economics of Education Review* 21, 543-556.

OECD (2013). Education at a Glance, Paris.

Polachek, S.W. (1978). Sex Differences in College Major, *Industrial and Labor Relations Review* 31, 498–508.

Sonnert, Gerhard (2009). Parents Who Influence Their Children to Become Scien-

tists: Effects of Gender and Parental Education, *Social Studies of Science* 39, 927-941.

Tabellini, Guido (2008). Institutions and Culture, *Journal of the European Economic Association* 6, 255-294.

Wang, Xueli (2013). Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support, *American Educational Research Journal* 50, 1081–1121.

Willis, Robert J. and Sherwin Rosen (1979). Education and Self-Selection, *Journal of Political Economy* 87, S7-S36.

Winters, John V. (2013). STEM Graduates, Human Capital Externalities, and Wages in the U.S., IZA Discussion Paper No. 7830.

Zafar, Basit (2009). College Major Choice and the Gender Gap, Federal Reserve Bank of New York Staff Report no. 364.

Table1: Descriptive statistics, total and conditional on gender

	Total							Men							Women						
	Mean	Std. Dev	Min	1st Quartile	Median	3rd Quartile	Max	Mean	Std. Dev	Min	1st Quartile	Median	3rd Quartile	Max	Mean	Std. Dev	Min	1st Quartile	Median	3rd Quartile	Max
STEM Field	0.305	0.460	0.000	0.000	0.000	1.000	1.000	0.416	0.493	0.000	0.000	0.000	1.000	1.000	0.178	0.382	0.000	0.000	0.000	0.000	1.000
StemCells	0.694	0.099	0.314	0.634	0.674	0.760	0.970	0.690	0.097	0.314	0.630	0.672	0.747	0.964	0.699	0.101	0.314	0.635	0.674	0.774	0.970
GenderEqu	0.636	0.125	0.179	0.560	0.652	0.716	0.914	0.632	0.126	0.179	0.549	0.652	0.716	0.914	0.640	0.124	0.179	0.569	0.654	0.716	0.914
Heroin	0.550	0.102	0.217	0.492	0.563	0.620	0.766	0.552	0.101	0.223	0.492	0.564	0.621	0.766	0.549	0.102	0.217	0.492	0.562	0.620	0.766
ShareCath	0.469	0.230	0.018	0.299	0.411	0.651	0.974	0.472	0.231	0.018	0.298	0.416	0.669	0.974	0.465	0.229	0.018	0.301	0.407	0.607	0.974
LeftWing	0.332	0.122	0.021	0.246	0.321	0.410	0.868	0.328	0.121	0.021	0.241	0.316	0.410	0.868	0.337	0.123	0.029	0.248	0.323	0.410	0.868
HomoCouples	0.598	0.084	0.183	0.549	0.619	0.655	0.766	0.597	0.085	0.183	0.549	0.617	0.655	0.766	0.600	0.084	0.200	0.548	0.622	0.655	0.742
French	0.316	0.465	0.000	0.000	0.000	1.000	1.000	0.293	0.455	0.000	0.000	0.000	1.000	1.000	0.342	0.474	0.000	0.000	0.000	1.000	1.000
Italian	0.060	0.238	0.000	0.000	0.000	0.000	1.000	0.057	0.231	0.000	0.000	0.000	0.000	1.000	0.064	0.245	0.000	0.000	0.000	0.000	1.000
LogDistETH	3.821	1.258	0.000	3.418	4.213	4.588	5.537	3.843	1.233	0.000	3.472	4.238	4.597	5.497	3.797	1.286	0.000	3.388	4.202	4.584	5.537
LogDistNextUni	2.550	1.463	0.000	1.723	3.008	3.671	5.389	2.586	1.453	0.000	1.758	3.020	3.701	5.389	2.509	1.474	0.000	1.609	2.950	3.630	5.226
LogAge	31.335	2.930	27.000	30.000	31.000	32.000	57.000	31.327	2.544	27.000	30.000	31.000	32.000	54.000	31.345	3.317	27.000	30.000	31.000	32.000	57.000
PostEdu	0.549	0.498	0.000	0.000	1.000	1.000	1.000	0.546	0.498	0.000	0.000	1.000	1.000	1.000	0.551	0.498	0.000	0.000	1.000	1.000	1.000
ParentUni	0.379	0.485	0.000	0.000	0.000	1.000	1.000	0.362	0.481	0.000	0.000	0.000	1.000	1.000	0.398	0.490	0.000	0.000	0.000	1.000	1.000
Female	0.467	0.499	0.000	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
LogEarnings (STEM fields)	8.672	0.367	7.607	8.487	8.741	8.912	9.774	8.7	0.36	7.607	8.517	8.774	8.933	9.774	8.596	0.373	7.665	8.43	8.64	8.836	9.761
LogEarnings (Humanities)	8.759	0.309	7.65	8.595	8.781	8.949	9.74	8.802	0.315	7.65	8.638	8.83	8.997	9.74	8.725	0.301	7.65	8.573	8.748	8.91	9.69

Notes: For a reading example take the value of the variable "STEM field" in the 8th column. The value indicates that 41.6 percent of all male students in the survey graduated in a STEM field.

Table 2: First stage Probit estimates for the effects on major choice

Dependent Variable: Dummy for graduation in STEM fields							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LeftWing	-0.512** (0.22)						
GenderEqu		-0.714*** (0.227)					
StemCells			-0.908*** (0.313)				
Heroin				-0.372 (0.315)			
HomoCouples					-0.492 (0.368)		
ShareCath						0.129 (0.114)	
FirstComp							-0.057*** (0.019)
SecondComp							0.123** (0.058)
French	-0.366*** (0.055)	-0.318*** (0.058)	-0.276*** (0.065)	-0.436*** (0.073)	-0.414*** (0.061)	-0.403*** (0.057)	-0.174 (0.118)
Italian	-0.131 (0.1)	-0.031 (0.097)	-0.025 (0.097)	-0.099 (0.099)	-0.133 (0.109)	-0.122 (0.108)	-0.03 (0.118)
Female	-0.708*** (0.046)	-0.712*** (0.046)	-0.711*** (0.046)	-0.71*** (0.046)	-0.709*** (0.046)	-0.709*** (0.046)	-0.708*** (0.047)
LogDistETH	-0.045** (0.021)	-0.042** (0.02)	-0.057*** (0.021)	-0.046** (0.02)	-0.052** (0.021)	-0.048** (0.02)	-0.052** (0.02)
LogDistNextUni	0.037** (0.018)	0.026 (0.019)	0.042** (0.017)	0.045** (0.019)	0.045** (0.019)	0.053*** (0.017)	0.036* (0.02)
LogAge	-2.232*** (0.328)	-2.233*** (0.327)	-2.261*** (0.327)	-2.212*** (0.325)	-2.22*** (0.325)	-2.229*** (0.326)	-2.266*** (0.328)
PostEdu	-0.175*** (0.045)	-0.177*** (0.046)	-0.176*** (0.046)	-0.176*** (0.046)	-0.176*** (0.046)	-0.175*** (0.046)	-0.169*** (0.046)
ParentUni	-0.075 (0.047)	-0.064 (0.047)	-0.066 (0.047)	-0.072 (0.047)	-0.07 (0.047)	-0.072 (0.047)	-0.065 (0.047)
Constant	7.871*** (1.155)	8.147*** (1.165)	8.423*** (1.183)	7.84*** (1.165)	7.971*** (1.192)	7.607*** (1.141)	7.806*** (1.155)
Statistics							
Observations	4162	4162	4162	4162	4162	4162	4138
Pseudo R ²	0.0812	0.0821	0.0819	0.0804	0.0805	0.0804	0.0823

Notes: Bootstrapped standard errors of the coefficients are shown in parentheses. The stars indicate the significance level of a coefficient:
*** p<0.01, ** p<0.05, * p<0.1.

Table 3: Second stage estimates of the earnings equation in STEM majors corrected for self-selection

Dependent Variable: LogEarnings in STEM fields							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LogAge	0.577*** (0.174)	0.6*** (0.176)	0.582*** (0.172)	0.565*** (0.174)	0.572*** (0.174)	0.568*** (0.174)	0.601*** (0.173)
PostEdu	-0.179*** (0.026)	-0.176*** (0.026)	-0.178*** (0.026)	-0.18*** (0.026)	-0.18*** (0.026)	-0.18*** (0.026)	-0.177*** (0.026)
ParentUni	-0.028 (0.023)	-0.027 (0.023)	-0.028 (0.023)	-0.029 (0.023)	-0.028 (0.023)	-0.029 (0.023)	-0.028 (0.023)
Female	-0.008 (0.049)	0.002 (0.049)	-0.005 (0.049)	-0.013 (0.05)	-0.011 (0.051)	-0.012 (0.05)	0.006 (0.048)
CorrectSTEM	0.199** (0.078)	0.218*** (0.078)	0.205*** (0.077)	0.189** (0.081)	0.194** (0.081)	0.193** (0.081)	0.225*** (0.076)
Constant	7.007*** (0.566)	6.946*** (0.572)	6.994*** (0.559)	7.043*** (0.564)	7.022*** (0.565)	7.034*** (0.564)	6.95*** (0.564)
Statistics							
Observations	1270	1270	1270	1270	1270	1270	1260
R ²	0.1186	0.12	0.1192	0.1178	0.1181	0.118	0.1214
R ² adjusted	0.1151	0.1165	0.1157	0.1143	0.1147	0.1145	0.1179

Notes: The different specifications vary in the cultural variable included at the first stage. The order of cultural variables included at the first stage corresponds to the order in Table 2. Bootstrapped standard errors are shown in parentheses. The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Second stage estimates of the earnings equation in humanities corrected for self-selection

Dependent Variable: LogEarnings in Humanities							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LogAge	0.709*** (0.084)	0.701*** (0.084)	0.7*** (0.083)	0.727*** (0.086)	0.724*** (0.085)	0.73*** (0.087)	0.688*** (0.082)
PostEdu	-0.039** (0.016)	-0.04** (0.016)	-0.041** (0.016)	-0.037** (0.016)	-0.038** (0.016)	-0.037** (0.016)	-0.043*** (0.016)
ParentUni	-0.007 (0.015)	-0.008 (0.015)	-0.008 (0.015)	-0.006 (0.016)	-0.006 (0.015)	-0.006 (0.016)	-0.008 (0.015)
Female	0.067** (0.029)	0.063** (0.029)	0.062** (0.03)	0.076** (0.031)	0.075** (0.031)	0.078** (0.031)	0.055* (0.028)
CorrectHum	0.422*** (0.084)	0.412*** (0.083)	0.408*** (0.084)	0.45*** (0.089)	0.447*** (0.087)	0.455*** (0.088)	0.393*** (0.081)
Constant	6.117*** (0.323)	6.15*** (0.323)	6.158*** (0.321)	6.036*** (0.334)	6.048*** (0.328)	6.022*** (0.336)	6.209*** (0.314)
Statistics							
Observations	2892	2892	2892	2892	2892	2892	2878
R ²	0.0649	0.0649	0.0641	0.0664	0.0662	0.0669	0.0645
R ² adjusted	0.0633	0.0632	0.0625	0.0648	0.0646	0.0652	0.0628

Notes: The different specifications vary in the cultural variable included at the first stage. The order of cultural variables included at the first stage corresponds to the order in Table 2. Bootstrapped standard errors are shown in parentheses. The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Third stage Probit estimates for the effects on major choice

Dependent Variable: Dummy for graduation in STEM fields							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LeftWing	-0.526** (0.227)						
GenderEqu		-0.735*** (0.239)					
StemCells			-0.878*** (0.323)				
Heroin				-0.46 (0.338)			
HomoCouples					-0.559 (0.395)		
ShareCath						0.136 (0.123)	
FirstComp							-0.057*** (0.019)
SecondComp							0.105* (0.059)
French	-0.282*** (0.062)	-0.224*** (0.064)	-0.19*** (0.07)	-0.373*** (0.081)	-0.34*** (0.068)	-0.329*** (0.066)	-0.107 (0.121)
Italian	-0.044 (0.106)	0.059 (0.102)	0.064 (0.103)	-0.019 (0.108)	-0.053 (0.116)	-0.036 (0.118)	0.041 (0.121)
LogEarnDiff	1.186** (0.583)	1.197** (0.585)	1.199** (0.592)	1.172** (0.586)	1.16** (0.585)	1.164** (0.588)	1.177** (0.583)
Female	-0.651*** (0.062)	-0.655*** (0.061)	-0.654*** (0.061)	-0.653*** (0.061)	-0.653*** (0.061)	-0.652*** (0.061)	-0.655*** (0.061)
LogDistETH	-0.048** (0.021)	-0.043** (0.021)	-0.058*** (0.021)	-0.05** (0.021)	-0.056*** (0.021)	-0.051** (0.021)	-0.053*** (0.02)
LogDistNextUni	0.048** (0.019)	0.035* (0.019)	0.054*** (0.017)	0.055*** (0.02)	0.055*** (0.02)	0.066*** (0.017)	0.043** (0.021)
Constant	0.157 (0.155)	0.443** (0.208)	0.585** (0.273)	0.251 (0.261)	0.344 (0.315)	-0.103 (0.119)	-0.008 (0.141)
Impacts							
Impact of Female	-0.205***	-0.206***	-0.205***	-0.205***	-0.205***	-0.205***	-0.205***
χ^2 -Statistic	210.289	213.444	212.544	211.666	211.314	210.91	212.956
Impact of FirstComp	-0.027**	-0.036***	-0.038***	-0.018	-0.018	0.015	-0.039***
χ^2 -Statistic	5.185	10.301	8.236	2.212	2.462	1.444	8.819
Impact of SecondComp							0.028**
χ^2 -Statistic							3.856
Impact of LogEarnDiff	0.048***	0.048***	0.048***	0.047***	0.047***	0.047***	0.047***
χ^2 -Statistic	12.966	13.123	13.242	12.721	12.437	12.602	12.804
Impact of French	-0.087***	-0.069***	-0.059***	-0.114***	-0.104***	-0.101***	-0.033
χ^2 -Statistic	33.384	18.49	10.027	33.93	41.912	42.642	1.045
Impact of Italian	-0.014	0.019	0.02	-0.006	-0.017	-0.011	0.013
χ^2 -Statistic	0.185	0.334	0.386	0.033	0.235	0.108	0.116
Impact of LogDistETH	-0.017**	-0.015**	-0.02***	-0.017**	-0.019***	-0.018***	-0.018***
χ^2 -Statistic	5.998	4.812	8.786	6.431	7.912	6.899	7.269
Impact of LogDistNextUni	0.031**	0.023*	0.035***	0.035***	0.036***	0.043***	0.028**
χ^2 -Statistic	6.031	3.226	9.223	7.508	7.714	14.181	4.16
Statistics							
Observations	4162	4162	4162	4162	4162	4162	4138
Pseudo R ²	0.071	0.072	0.071	0.07	0.07	0.07	0.072

Notes: Reading examples for impacts: in column 1, the effect of being female is estimated to reduce the probability of studying natural sciences by 20.5 percent points. Bootstrapped standard errors are shown in parentheses. The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.

Table 6: First stage Probit estimates for the effects on major choice with gender interaction terms

Dependent Variable: Dummy for graduation in STEM fields							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LeftWing	-0.613** (0.258)						
Female x LeftWing	0.274 (0.369)						
GenderEqu		-0.79*** (0.266)					
Female x GenderEqu		0.224 (0.364)					
StemCells			-0.844** (0.363)				
Female x StemCells			-0.174 (0.45)				
Heroin				-0.724** (0.367)			
Female x Heroin				1.016** (0.446)			
HomoCouples					-0.789* (0.43)		
Female x HomoCouples					0.869 (0.54)		
ShareCath						0.177 (0.14)	
Female x ShareCath						-0.134 (0.204)	
FirstComp							-0.068*** (0.022)
Female x FirstComp							0.03 (0.03)
SecondComp							0.09 (0.063)
Female x SecondComp							0.1* (0.054)
Female	-0.799*** (0.131)	-0.854*** (0.236)	-0.591* (0.316)	-1.27*** (0.249)	-1.229*** (0.324)	-0.645*** (0.108)	-0.777*** (0.06)
French	-0.365*** (0.055)	-0.317*** (0.058)	-0.276*** (0.065)	-0.433*** (0.073)	-0.412*** (0.061)	-0.403*** (0.057)	-0.167 (0.119)
Italian	-0.129 (0.1)	-0.031 (0.097)	-0.025 (0.097)	-0.094 (0.1)	-0.123 (0.111)	-0.118 (0.109)	-0.021 (0.12)
LogDistETH	-0.045** (0.02)	-0.042** (0.02)	-0.057*** (0.021)	-0.046** (0.02)	-0.051** (0.02)	-0.047** (0.02)	-0.052** (0.02)
LogDistNextUni	0.037** (0.018)	0.027 (0.019)	0.042** (0.017)	0.046** (0.019)	0.045** (0.019)	0.053*** (0.017)	0.036* (0.02)
LogAge	-2.228*** (0.328)	-2.227*** (0.328)	-2.268*** (0.327)	-2.219*** (0.325)	-2.222*** (0.326)	-2.23*** (0.326)	-2.286*** (0.328)
PostEdu	-0.174*** (0.045)	-0.176*** (0.046)	-0.177*** (0.046)	-0.175*** (0.046)	-0.173*** (0.046)	-0.174*** (0.046)	-0.168*** (0.046)
ParentUni	-0.075 (0.047)	-0.064 (0.047)	-0.066 (0.047)	-0.075 (0.047)	-0.07 (0.047)	-0.072 (0.047)	-0.067 (0.047)
Constant	7.888*** (1.156)	8.172*** (1.167)	8.403*** (1.187)	8.054*** (1.17)	8.15*** (1.19)	7.584*** (1.144)	7.892*** (1.156)
Statistics							
Observations	4162	4162	4162	4162	4162	4162	4138
Pseudo R ²	0.0813	0.0821	0.0819	0.0814	0.081	0.0804	0.0833

Notes: Bootstrapped standard errors of the coefficients are shown in parentheses. The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Third stage Probit estimates for the effects on major choice with gender interaction terms

Dependent Variable: Dummy for graduation in STEM fields							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LeftWing	-0.71** (0.289)						
Female x LeftWing	0.454 (0.393)						
GenderEqu		-0.938*** (0.318)					
Female x GenderEqu		0.498 (0.405)					
StemCells			-0.978** (0.41)				
Female x StemCells			0.24 (0.487)				
Heroin				-0.852** (0.412)			
Female x Heroin				0.976** (0.472)			
HomoCouples					-0.912* (0.488)		
Female x HomoCouples					0.923 (0.57)		
ShareCath						0.19 (0.161)	
Female x ShareCath						-0.131 (0.216)	
FirstComp							-0.073*** (0.025)
Female x FirstComp							0.043 (0.031)
SecondComp							0.082 (0.066)
Female x SecondComp							0.066 (0.055)
LogEarnDiff	1.672** (0.801)	1.717** (0.83)	1.682** (0.828)	1.596** (0.791)	1.618** (0.787)	1.623** (0.8)	1.572* (0.805)
Female x LogEarnDiff	-1.123 (0.74)	-1.166 (0.772)	-1.104 (0.754)	-1.046 (0.722)	-1.062 (0.724)	-1.081 (0.73)	-0.927 (0.731)
Female	-0.949*** (0.174)	-1.125*** (0.293)	-0.963*** (0.368)	-1.328*** (0.273)	-1.344*** (0.35)	-0.731*** (0.151)	-0.848*** (0.111)
French	-0.287*** (0.064)	-0.23*** (0.067)	-0.194*** (0.072)	-0.374*** (0.083)	-0.341*** (0.07)	-0.335*** (0.068)	-0.097 (0.122)
Italian	-0.042 (0.109)	0.058 (0.105)	0.063 (0.105)	-0.014 (0.109)	-0.044 (0.12)	-0.034 (0.121)	0.053 (0.123)
Log DistETH	-0.048** (0.021)	-0.044** (0.021)	-0.059*** (0.021)	-0.049** (0.021)	-0.055** (0.021)	-0.052** (0.022)	-0.052** (0.02)
Log DistNextUni	0.048** (0.02)	0.036* (0.02)	0.054*** (0.018)	0.055*** (0.02)	0.055*** (0.021)	0.066*** (0.018)	0.043** (0.021)
Constant	0.274 (0.187)	0.633** (0.272)	0.71** (0.341)	0.513* (0.305)	0.603 (0.37)	-0.076 (0.145)	0.056 (0.155)
Impacts							
Impact of Female	-0.206***	-0.207***	-0.206***	-0.206***	-0.206***	-0.206***	-0.205***
χ^2 -Statistic	217.333	220.961	216.294	213.348	215.792	214.936	212.426
Impact of FirstComp (Male)	-0.045***	-0.057***	-0.047***	-0.041**	-0.035**	0.025	-0.062***
χ^2 -Statistic	6.87	11.773	7.167	5.792	4.864	1.968	10.489
Impact of FirstComp (Female)	-0.01	-0.015	-0.028*	0.004	0	0.005	-0.015
χ^2 -Statistic	0.622	2.008	3.722	0.094	0.001	0.123	1.345
Impact of SecondComp (Male)							0.042
χ^2 -Statistic							2.072
Impact of SecondComp (Female)							0.057**
χ^2 -Statistic							5.816
Impact of LogEarnDiff (Male)	0.075***	0.078***	0.077***	0.072***	0.074***	0.073***	0.072***
χ^2 -Statistic	14.104	14.417	13.949	12.725	13.226	13.514	11.912
Impact of LogEarnDiff (Female)	0.018	0.018	0.019	0.018	0.019	0.018	0.021
χ^2 -Statistic	1.359	1.366	1.52	1.312	1.376	1.321	1.843
Impact of French (Male)	-0.106***	-0.085***	-0.072***	-0.137***	-0.125***	-0.123***	-0.036
χ^2 -Statistic	34.185	19.079	10.33	33.687	41.843	43.71	0.838
Impact of French (Female)	-0.066***	-0.053***	-0.045***	-0.084***	-0.077***	-0.076***	-0.023
χ^2 -Statistic	34.159	19.326	10.542	33.245	41.318	42.998	0.853
Impact of Italian (Male)	-0.016	0.022	0.024	-0.005	-0.016	-0.013	0.02
χ^2 -Statistic	0.166	0.333	0.384	0.018	0.151	0.093	0.191
Impact of Italian (Female)	-0.01	0.014	0.015	-0.003	-0.01	-0.008	0.013
χ^2 -Statistic	0.17	0.321	0.369	0.018	0.155	0.095	0.185
Impact of DistETH (Male)	-0.02**	-0.018**	-0.024***	-0.02**	-0.022***	-0.021***	-0.021***
χ^2 -Statistic	6.059	4.916	8.92	6.371	7.779	7.004	7.051

Impact of DistETH (Female)	-0.013**	-0.012**	-0.016***	-0.013**	-0.015***	-0.014***	-0.014***
χ^2 -Statistic	6.085	4.94	8.951	6.402	7.8	7.036	7.08
Impact of DistNextUni (Male)	0.035**	0.026*	0.04***	0.041***	0.041***	0.049***	0.032**
χ^2 -Statistic	5.99	3.279	9.189	7.478	7.667	14.209	4.065
Impact of DistNextUni (Female)	0.024**	0.018*	0.027***	0.028***	0.028***	0.034***	0.022**
χ^2 -Statistic	5.997	3.291	9.205	7.493	7.689	14.209	4.083
Statistics							
Observations	4162	4162	4162	4162	4162	4162	4138
Pseudo R ²	0.0717	0.0727	0.072	0.0716	0.0713	0.0707	0.0732

Notes: The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.

Online Appendix

(Not intended for publication)

In this appendix we first show the results of a Principal Component Analysis (PCA) for all sociocultural variables included in our sample: *GenderEqu*, *StemCells*, *Heroin*, *HomoCouples*, *ShareCath*, *LeftWing*, *ShareCath*.

Table A.1: Eigenvalues of the Principal Component Analysis on all used sociocultural variables

Component	Eigenvalue	Proportion	Cumulative
Comp1	2.61982	0.4366	0.4366
Comp2	1.03686	0.1728	0.6094
Comp3	0.977546	0.1629	0.7724
Comp4	0.640539	0.1068	0.8791
Comp5	0.481539	0.0803	0.9594
Comp6	0.243701	0.0406	1

Notes: The table shows in column 2 the eigenvalues for each component of a PCA on all cultural variables included in the paper. The eigenvalues allow us to calculate the proportion of the variance explained by a component (column 3). For example, the first component *Comp1* captures 43.66 percent of the total variance in the dataset. The last column shows the cumulative proportion explained by the first components. For example, the first four components explain 87.91 of the total variance in the data.

Table A.2: Eigenvectors of the Principal Component Analysis on all used sociocultural variables

Variables	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6
<i>GenderEqu</i>	0.4137	-0.3415	0.4093	0.2409	-0.6837	-0.1386
<i>StemCells</i>	0.417	-0.5142	0.1832	-0.4353	0.3754	0.4446
<i>Heroin</i>	0.3494	0.7361	0.1882	-0.0868	-0.1779	0.5114
<i>HomoCouples</i>	0.5233	0.2592	0.0331	-0.307	0.2429	-0.7104
<i>ShareCath</i>	-0.2704	0.0962	0.8409	0.2256	0.3894	-0.0893
<i>LeftWing</i>	0.4311	-0.0256	-0.2351	0.7745	0.3865	0.0948

Notes: The table presents the eigenvectors of each component of a PCA on all cultural variables included in the paper. The eigenvectors show how the variables “load” on each component. For a reading example, take the eigenvector of the first component. The results show that all referenda results *GenderEqu*, *StemCells*, *Heroin*, *HomoCouples*, *ShareCath* and *LeftWing* are positively correlated with the first component whereas *ShareCath* is negatively correlated with the first component.

Next, in Tab. A.3 and A.4 we provide the results from the estimates at stage 2 based on the specifications with gender differences which are analogous to Tab. 3 and 4, respectively.

Table A.3: Second stage estimates of the earnings equation in STEM fields corrected for self-selection with gender interaction terms

Dependent Variable: LogEarnings in STEM Fields							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LogAge	0.576*** (0.174)	0.596*** (0.176)	0.585*** (0.172)	0.577*** (0.174)	0.582*** (0.175)	0.571*** (0.174)	0.621*** (0.175)
PostEdu	-0.179*** (0.026)	-0.177*** (0.026)	-0.178*** (0.026)	-0.179*** (0.026)	-0.178*** (0.026)	-0.18*** (0.026)	-0.175*** (0.026)
ParentUni	-0.028 (0.023)	-0.027 (0.023)	-0.027 (0.023)	-0.028 (0.023)	-0.027 (0.023)	-0.028 (0.023)	-0.026 (0.024)
Female	-0.009 (0.049)	0 (0.049)	-0.004 (0.05)	-0.012 (0.05)	-0.007 (0.05)	-0.01 (0.05)	0.01 (0.049)
CorrectSTEM	0.197** (0.077)	0.213*** (0.077)	0.207*** (0.079)	0.194** (0.08)	0.201** (0.08)	0.195** (0.08)	0.236*** (0.077)
Constant	7.01*** (0.564)	6.956*** (0.571)	6.987*** (0.561)	7.004*** (0.565)	6.994*** (0.566)	7.026*** (0.564)	6.891*** (0.57)
<i>Statistics</i>							
Observations	1270	1270	1270	1270	1270	1270	1260
R ²	0.1186	0.1199	0.1192	0.1182	0.1187	0.1182	0.1223
R ² adjusted	0.1151	0.1164	0.1158	0.1147	0.1152	0.1147	0.1188

Notes: The different specifications vary in the cultural variable included on the first stage. The order of cultural variables included on the first stage corresponds to the order in Table 6. Bootstrapped standard errors are shown in parentheses. The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Second stage estimates of the earnings equation in humanities corrected for self-selection with gender interaction terms

Dependent Variable: LogEarnings in Humanities							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	b/se	b/se	b/se	b/se	b/se	b/se	b/se
LogAge	0.705*** (0.083)	0.701*** (0.084)	0.7*** (0.083)	0.716*** (0.083)	0.715*** (0.084)	0.727*** (0.087)	0.678*** (0.079)
PostEdu	-0.04** (0.016)	-0.041** (0.016)	-0.041** (0.016)	-0.039** (0.016)	-0.039** (0.016)	-0.038** (0.016)	-0.044*** (0.015)
ParentUni	-0.008 (0.015)	-0.008 (0.015)	-0.008 (0.015)	-0.007 (0.015)	-0.007 (0.015)	-0.007 (0.015)	-0.009 (0.015)
Female	0.065** (0.029)	0.063** (0.029)	0.062** (0.03)	0.072** (0.031)	0.071** (0.03)	0.076** (0.031)	0.051* (0.028)
CorrectHum	0.417*** (0.082)	0.413*** (0.082)	0.408*** (0.084)	0.438*** (0.084)	0.434*** (0.084)	0.451*** (0.087)	0.381*** (0.077)
Constant	6.133*** (0.319)	6.151*** (0.321)	6.156*** (0.318)	6.081*** (0.32)	6.088*** (0.322)	6.034*** (0.334)	6.252*** (0.301)
<i>Statistics</i>							
Observations	2892	2892	2892	2892	2892	2892	2878
R ²	0.0647	0.065	0.0642	0.0665	0.0657	0.0666	0.0642
R ² adjusted	0.0631	0.0634	0.0626	0.0649	0.0641	0.0649	0.0626

Notes: The different specifications vary in the cultural variable included at the first stage. The order of cultural variables included at the first stage corresponds to the order in Table 6. Bootstrapped standard errors are shown in parentheses. The stars indicate the significance level of a coefficient: *** p<0.01, ** p<0.05, * p<0.1.