Major Switching Patterns of the Czech University Students with the Focus on STEM

Microeconometrics I, CERGE-EI 2022

Vít Illichmann

1 Introduction

STEM majors are highly competitive and demanding field of study. Although in the past successful STEM graduates were properly compensated for the invested effort, rapid technological development of the past two decades significantly reduced the premium on STEM degrees. Deming and Noray (2019) show, that STEM jobs became increasingly more skill intensive compared to any other field (comparing development from **2007** to **2017**). As a result, wage growth of STEM degree holders turned flat, compared to employees in other fields, as their skill set became outdated.

Astorne-Figari and Speer (2019) study major switching patterns of **US** students and conclude that the students who switch from STEM are driven out, among other factors, by the high levels of competitiveness and bad grades. In line with the anecdotal evidence, the most common switching paths identified by the researchers are **computer science to business** and **engineering to business** (for male students). Further, STEM programs seem to be biggest net losers, meaning that students switch out of STEM but do not switch in (e.g. education seems to be on the opposite end of spectrum as a study field). The model used by authors focuses largely on the social characteristics of each field and personal preferences of each student in terms of the courses taught within each field of study.

Data on job skill requirements presented by Deming and Noray (2019) show, that although the ICT skills required by employers experienced by far most rapid growth within STEM line of work, the similar requirements in other fields experienced fast paced growth as well. I posit that patterns observed by Astorne-Figari and Speer (2019) are not necessarily only result of personal academic preferences but are connected to the optimisation of future expected wage as well. With a STEM-like skills being increasingly required outside of their original field, STEM students can expect significant premium from switching to a less tech-heavy field where they have a comparative advantage with respect to their new peers and at the same time do not experience such a fast paced depreciation of their skill set as described by Deming and Noray (2019).

Astorne-Figari and Speer (2019) use data from NLSY, which is a large-scale longitudinal survey, containing rich individual-level data. This allows the authors to identify switching patterns with relative ease. To my best knowledge there is no comparable dataset in the Czech republic. To identify the migration patterns of students across different fields of study I use strategy inspired by Souza, Rasul and de Paula (2019). With a use of the panel data containing amounts of students within 36 distinct fields of study I attempt to recover interaction patterns among majors based on covariance relationship between the individual time series, controlling for long term demographic trends, time persistence and changes in admission policy.

Similarly to Souza, Rasul and de Paula (2019) I assume sparsity of the estimated parameter vector and use penalised **LASSO** regression to estimate my model.

2 Data Analysis

2.1 Dataset

Analysed panel data cover university students in the Czech republic and span across the **2001-2020** period. The data included aggregate amounts of students in **36** distinct fields of study classified according the **CZ -ISCED-F-2013** classification system. For the purposes of the analysis three items were removed from the original panel, because of the overwhelming amount of zero entries. Namely, **Interdisciplinary programs related to services**, **Hygiene and preventive health care** and **Fishery**. The data cover programs in all levels of study: **Bachelors**, **Masters** and **PhDs**. If a student majors in two (or more) distinct fields the data contain double entry, sum of students over all study fields is therefore not equivalent to the amount of students in total. This disparity makes up $\sim 2.1\%$ of all students in the given year, across the data (averaged over time). Interdisciplinary study programs are classified based on the content of the courses taught.

2.2 Preprocessing of the Data

An interactive data visualisation for exploratory purposes is accessible here.
After visual inspection of the data, it is clear that time series of students display similar long-term pattern across fields (with several exceptions). This probably result of underlying demographic trends influencing relative size of the student cohorts entering higher education, changes in educational policies and transition towards a more skill intensive labour market. To dispose of the variability induced by the common underlying trends I first differenced the data, see results in **Figure 1** (changes in individual fields of study are plotted by the grey opaque lines).

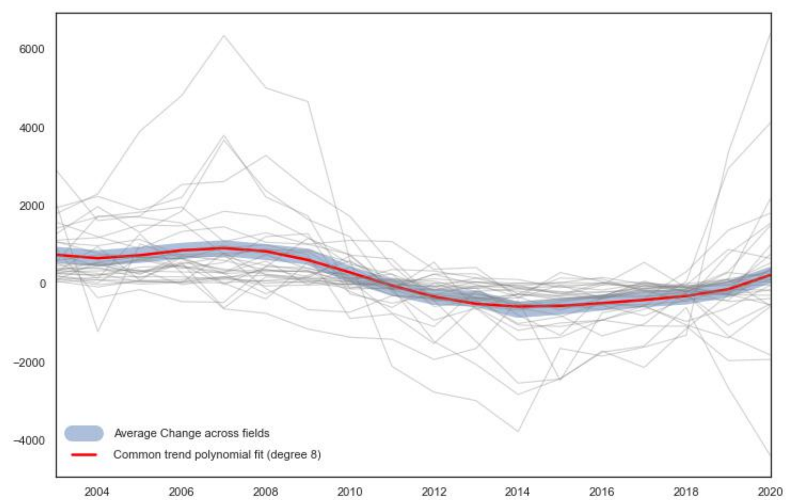

Figure 1: Yearly changes in amounts of students across study fields with fitted polynomial trend.

First differences clearly overwhelmingly follow long-term S shaped cycle. To remove the cycle I subtracted an 8-degree polynomial fit from each time series (results in Figure 2).

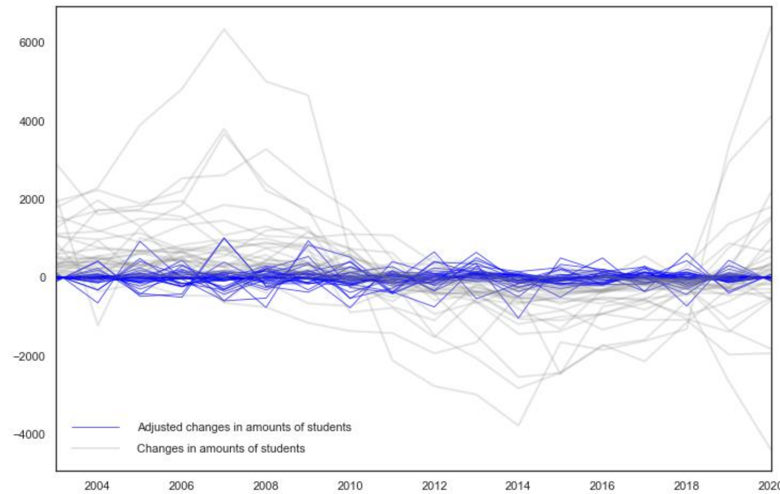

Figure 2: Yearly changes in amounts of students adjusted for long-term cycle.

2.3 Econometric Model and Estimation

Let $i = 1, 2, ..., N$ be an index denoting individual fields ($N = 33$) and t a time index ranging across the time span of the panel. To the estimate migration patterns of students from one field to another I propose a following relationship.

$$\Delta y_{i|t} = \sum_{j \neq i} \gamma_{ji} \Delta y_{j|t} + \beta_i \Delta y_{i|t-1} + \alpha_{i|t} + \epsilon_{i|t} \tag{1}$$

The $\Delta y_{i|t}$ is a change in the amount of students in the field i at time t. The sum of changes on the right hand side weighted by their respective coefficients γ_{ji} represents overall student migration into the ith field from other disciplines. Rapid changes of student numbers might display certain pattern of persistence over time, this is controlled for by the autoregressive term $\beta_i \Delta y_{i|t-1}$. Further, discrete change in admission policy by major universities might occur (e.g. simplified admission process during Covid-19 pandemic), for this reason I included fixed time effects $\alpha_{i|t}$ as well (results of estimation show they are overwhelmingly equal to zero). I theory, with enough control for outside influences the system could be regarded as closed with $\gamma_{ji} = -\gamma_{ij}$, however estimation results suggest that this not a valid identification strategy. Naturally,γ_{ii} are assumed to be zero.

To estimate the equation (1) I use a standard **LASSO** with penalization term in the loss function proportional to the L^1 norm of the estimated vector of coefficients. The **LASSO** algorithm i implemented as part of the **Sklearn** module in **Python** programming language. The estimation results are visualised in the **Figure 3**. Due to the large amount of the estimated interaction coefficients the static representation is inconvenient, thus interactive representation is available **here**

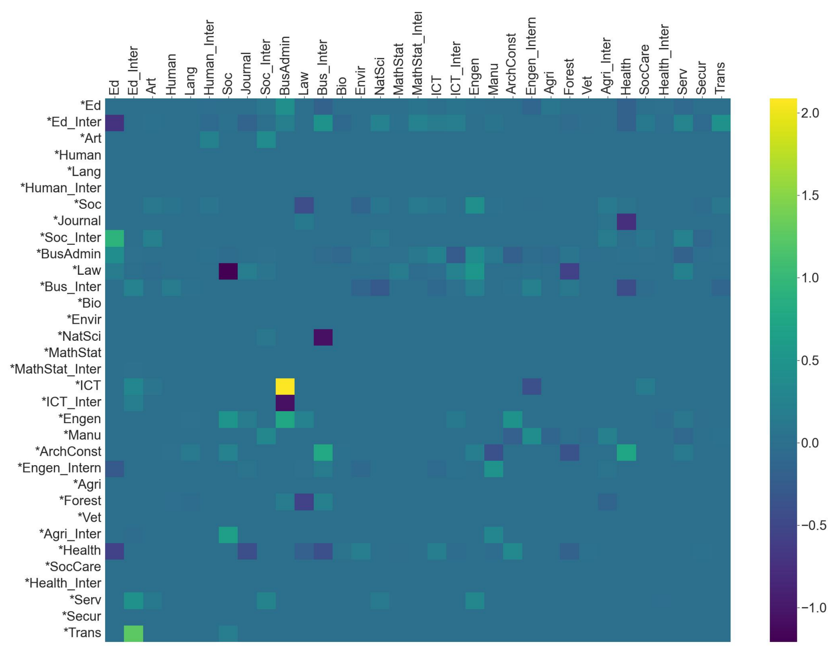

Figure 3: Visualisation of the estimated $\hat{\gamma}_{ij}$, columns of the *heath map chart* are depended variables (study fields), the rows with variables of the form $*X$ represent the independent variables in the regression equation.

The study fields are referred to by **code names**,their correspondence to the original names is summarised in **APENDIX I**, **Table 1**.

Most estimated interaction coefficients are zero or very close to zero (in line with the sparsity assumption). The influence of ICT majors on Business and Administration is striking compared to other relatively "weak" estimates the relationship in the opposite direction is positive as well but much weaker. This is in line with the results by Astorne-Figari and Speer (2019) cited in the introduction. The weak positive response in the opposite direction could be explained by presence high ability students taking on double majors in both fields.

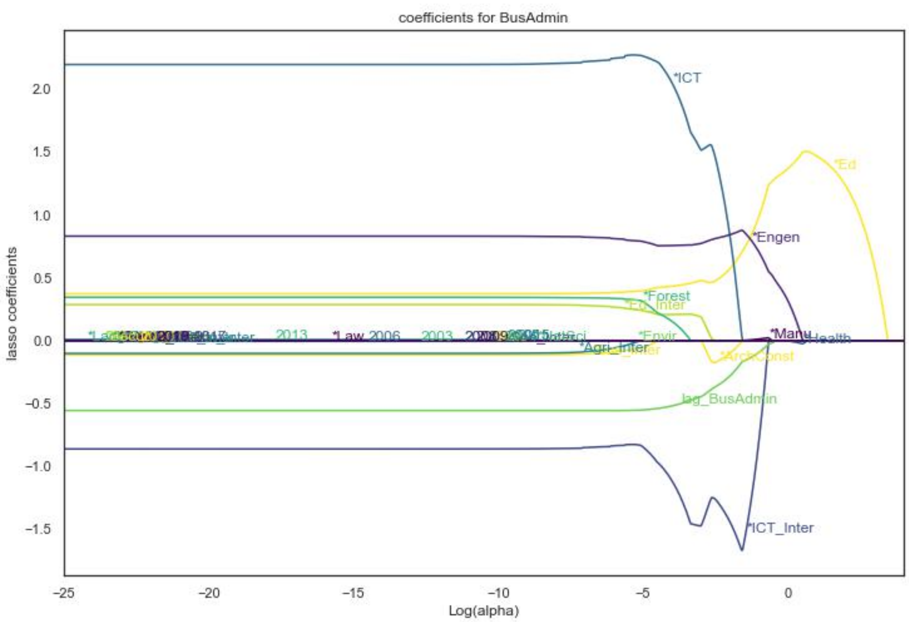

Figure 4: Sensitivity of the estimated **LASSO** coefficients on the changes in the penalisation weight α.

Naturally the value of estimated coefficients is highly sensitive to the weight put on the penalization term in the **LASSO** loss function. **Figure 4** shows dependence of the interaction coefficients for the **Business and Administration** majors on the logarithmic value of the penalisation weight α and shows that the positive influence of the **ICT** fields remains stable even for relatively high weights. Similar stability analysis was conducted for all dependent variables and exhibits similar pattern. That is, most weak dependencies die out for α slightly above e^{-5}, consequently weight $\alpha = e^{-5}$ to eliminate misleading estimates. Estimation results suggest that although the **STEM** degrees are not systematically subject to negative influences from other fields, that there is no persistent exodus out of STEM. There is an evidence of one sided migration which is in line with the patterns described by Astorne-Figari and Speer (2019). Unlike the **US** students who fund their education privately, Czech students have their cost covered by the state if they complete their studies within pre specified period of time, which makes decision to switch costly relative to sticking with the original choice.

Overall, most academic fields tend to influence those academically close to them which is in line with the intuition. However some strong negative dependencies seem to be counter intuitive and might be result of insensitive data manipulation or model misspecification. Some fields contain relatively small amounts of students who might have a limited choice in terms of universities offering such study programs and thus geo-spatial friction might play certain role (e.g. student who moved to Prague and lately decided to switch degree will probably choose new field of study in Prague, turning down perhaps better academic in a different city).

References

1. Astorne-Figari, C., & Speer, J. D. (2019). Are changes of major major changes? The roles of grades, gender, and preferences in college major switching. *Economics of Education Review*, 70, 75–93. https://doi.org/10.1016/j.econedurev.2019.03.005

2. Deming, D., & Noray, K. (2019). STEM Careers and the Changing Skill Requirements of Work. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3451346

3. Souza, P. C., Rasul, I., & de Paula, Á. (2019). Identifying network ties from panel data: theory and an application to tax competition. https://doi.org/10.1920/wp.cem.2019.5519

Non Academic References - Links

1. Sklearn LASSO documenation

2. QuantEcon Lecture on Regression and LASSO

APENDIX I

Table 1

	Field of Study	Code Name
0	Education & Pedagogy	Ed
1	Education - Interdiciplinary	Ed_Inter
2	Art	Art
3	Humanities	Human
4	Languages	Lang
5	Humanities - Interdisciplinary	Human_Inter
6	Social & Behavioral Sciences	Soc
7	Journalism	Journal
8	Social an Journalism - Interdisciplinary	Soc_Inter
9	Business & Admin	BusAdmin
10	Law	Law
11	Business & Law - Interdisciplinary	Bus_Inter
12	Biology	Bio
13	Environment	Envir
14	Natural Sciences	NatSci
15	Mathematics & Statistics	MathStat
16	Mathematics & Statistics - Interdisciplinary	MathStat_Inter
17	ICT	ICT
18	ICT - Interdisciplinary	ICT_Inter
19	Engeneering	Engen
20	Manufacturing	Manu
21	Architecture & Construction	ArchConst
22	Engeneering & Manufacturing - Interdisciplinary	Engen_Intern
23	Agriculture	Agri
24	Forestry	Forest
25	Veterinary care	Vet
26	Agriculture - Interdisciplinary	Agri_Inter
27	Healthcare	Health
28	Social care	SocCare
29	Healthcare - Interdisciplinary	Health_Inter
30	Services	Serv
31	Security Servicies	Secur
32	Transport	Trans