Research Statement

Ágoston Reguly

Current Research

My research field lies in empirical finance and econometrics focusing on improving methodology. In my ongoing research with Professor Sudheer Chava (Georgia Institute of Technology - GT, Financial Services and Innovation Lab - FSIL), we employ new matching methods to estimate the effect of corporate events (IPO, M&A, SEO, or dividend payment) of firms in the long-run. Our researched methodology not only identifies one matching firm that has similar characteristics to the event firm before the event and then controls for multiple factors (e.g. in Bessembinder, Cooper, and Zhang (2019)) but allows to weight multiple control firms with many characteristics through multiple periods. These weights are used to create synthetic event firms or one can view it as an alternative portfolio for investment. Preliminary results show that we can improve on the literature benchmark and explain the variation in contradictory results found in the literature.

I am also involved in research that aims to understand households' financial mistakes. As a partnership with Equifax (one of the large credit-score bureaus in the US), we have access to the majority of the US population's financial data (40 million + individuals). Usage of this rich dataset poses new challenges in causal inference and computational dimensions: heterogeneity analysis with unbalanced panel data and how to efficiently estimate the parameters of interests. In the long term at FSIL, we want to understand how life events affect the financial decisions of the household. As a recently formed project, we would like to investigate how retirement incentives some (poor) people to invest in risky assets (e.g. crypto).

I do research on the more traditional econometric problem: selection procedures of fixed effects (FEs) in panel data. This is joint work with László Mátyás (Central European University) and Felix Chan (Curtin University) and we investigate how assumed fixed effect structure (e.g. additive time and individual FEs) leads to biased parameter estimates and how a partially penalized regression method (e.g. LASSO) can estimate the unknown FE structure more flexibly resulting in an unbiased estimate.
Finally, I was also involved in writing several policy-analysis papers, mainly in my former job as an expert economic analyst. During this period, I worked with various databases on different topics mainly in the field of financial markets.

Future research plans

In the future, I intend to continue my research with FSIL members at GT, explore Hungarian data and in the area of machine learning methods and treatment effect estimation.

- I am interested in the implications of corporate events, thus not only analyzing the market performance of the firms but understanding better how corporate events affect employment and other firm dynamics (e.g. bankruptcy) in the long run.
- With Gábor Békes (CEU, CERS), I would like to explore the effects of similar corporate events using Hungarian data.
- At GT, I had the fortune to learn various deep learning and natural language processing methods. I believe the new data that these methods generate will play an important role in future research. I would like to start using these techniques and create useful economic/financial datasets based on Hungarian text data and explore how these new techniques can help answer questions.
- At the moment there is an on-hold project with Michael Knaus (University of Tübingen). We would like to explore how different methods are used to identify different types of CATE functions in different (quasi-) experimental designs, and we want to create a unified approach based on which these methods can be categorized. Such a paper would synthesize the rapidly growing literature in the light of the identification of the estimand and would help practitioners in finding a method that fits their question the best.
- In the medium term, I would like to start a team with colleagues in the spirit of the Machine Learning in Financial Team at FSIL that I have launched with Professor Sudheer Chava. The goal of this team would be a lab that does high-quality empirical research on Hungarian data with advanced tools. Similarly to US R1 school labs, this lab could initiate partnerships with industry, allowing the lab to access high-quality data. As I have been fortunate to see how these partnerships work at FSIL, I see some of its potential challenges (branding, non-academic work, legal issues, etc.) and many advantages (sponsorship,

high-quality data, knowledge transfer, etc). I believe in the long run we can build fruitful partnerships with Hungarian firms, where we have the opportunity of working with their data and they get benefits such as the prestige of public services and openly available resources from the Lab.

Finally, I am open to various applied economic questions researched by your university. I hope with my strong econometric and data skills I can contribute to the department and I am also excited to work on empirical problems related to Hungary.

Matured working papers

Heterogeneous Treatment Effects in Regression Discontinuity Designs

In economics, causal inference is often conducted using various (quasi-) experimental designs. Most of these designs lead to the identification of some overall average treatment effect parameters. However, the estimation of this parameter is often followed by some type of robustness and/or heterogeneity analysis to ensure internal or external validity. Machine learning methods present a unified approach to carrying out such complementary analysis in a systematic and reproducible way.
With Róbert Lieli, (Central European University), we work on the extension of my former job market paper, where we search for heterogeneous treatment effects in regression discontinuity designs with bandwidth selection. The paper consists of a theoretical and an empirical component. In the theoretical section, we propose a causal supervised machine learning algorithm to uncover treatment effect heterogeneity in classical regression discontinuity designs using regression trees. Groups with potentially different treatment effects are defined in terms of the values of a set of pre-treatment covariates. It is a priori unknown which covariates are relevant for capturing treatment effect heterogeneity, and it is the task of the algorithm to discover them, without invalidating inference. We study the performance of the method through Monte Carlo simulations and apply it to the data set compiled by Pop-Eleches and Urquiola (2013) to uncover various sources of heterogeneity in the impact of attending a better secondary school in Romania.
We contribute to the regression discontinuity design literature by offering a tool for heterogeneity analysis and to the causal machine learning literature by introducing a specialized criterion to search for and estimate conditional average treatment effects with leaf-by-leaf nonparametric regressions.

Discretized Variables

In Fall 2018, Professor László Mátyás (Central European University) drew my attention to a specific problem encountered by MaaSLab from the University College London, a research group focusing on public transportation. The problem centers on bias in parameter estimates when using discretized survey data.
Based on this issue, Professor László Mátyás, Professor Felix Chan, and I researched the econometric theory and applications of models where the dependent or explanatory variables are discretized due to surveys, but they are in fact continuous. The research resulted in two working papers. These papers propose a new data gathering method, called 'split-sampling', which allows point identification and consistent estimation of parameters in the linear regression model when variables are observed through a discretization process. The first paper discusses the split sampling data gathering method in detail and investigates the properties of the least squares estimator when the discretized variable is on the right-hand side. The second paper discusses the identification and estimation when the discretized variable is on the left-hand side. Throughout this research, I benefited from feedback from William Greene, Tom Wansbeek, and Richard Blundell.
At the current stage, we invited Professor Sudheer Chava to collaborate on this paper. We are planning to merge the two existing papers and create one paper that uses Equifax data as an empirical application. With Equifax data, a major problem is the anonymity of individuals for the income variable. Due to legal issues, we have no access to real income only a modeled income which introduces bias. Using one of our split-sampling methods would grant anonymity and allow us to solve for discretization bias. This would be not only beneficial for our paper, but for future research as well which uses Equifax's income variable.