

Leveraging prediction tools and information disclosure policies to improve school choice

Andreas Bjerre-Nielsen, David Dreyer Lassen and Emil Chrisander

Motivation Every year Danish universities and colleges admit around 65,000 students to higher education. However, roughly 1 out of 3 admitted students drop out prematurely [1]. What is more, a considerable share of students end up with poor employment prospects after graduation [2]. Considering the fact that Danish universities spent 8.9 billion DKK on education in 2018 and the state another 3.4 billion DKK on unemployment benefits to graduates [3, 4], one may argue that university dropouts and post-graduation unemployment are problems of great economic importance. In this project, we will examine how to alleviate these problems by integrating them into a school choice framework [5] that explicitly accounts for the use of matching procedures to allocate students to schools. That is, we will study a *matching problem* that seeks to (i) match students with study programs that lower students' dropout probability and (ii) increase students' post-graduation employment prospects by guiding them to better decisions.

Projects and contribution to literature We propose two sub-projects aimed at assessing new policy instruments in the context of admission to Danish higher education. The first sub-project will examine the potential of using prediction tools to estimate the match quality of future students before they are admitted. The sub-project will contribute to the recent academic literature on understanding of prediction tools in economics [6–8] and how they can be combined with matching procedures based on mechanism design. The second sub-project investigates how the release of relevant information may help applicants to evaluate the potential matching quality of study programs and thus improve their decision making. The sub-project will contribute to the knowledge of how information disclosures may affect short- and long-term decisions [9]. Jonas Zielke Schaarup, who is Head of Office for analysis at the Ministry of Higher Education and Research, has expressed that these projects are of high interest for the ministry, see attached 'Declaration of Interest'.

Sub-project 1: Prediction tools for match quality. In the first sub-project, we will build predictive tools to forecast subsequent student outcomes after admission. In particular, we will predict student first-year attrition, completion of studies, as well as subsequent employment likelihood and income. These tools will be integrated into the Deferred Acceptance matching procedure [10] currently used in *Den Koordinerede Tilmelding* (KOT), the Danish tertiary education admissions system.

Prediction tools are often deemed 'sufficiently good' when they outperform humans in the same task, and such advances have recently taken place in some tasks of image recognition [11] and complex decisions [12]. We will compare the predictive performance of our machine learning models against that of costly, manual screening of applicants in the 'Quota 2'-screening which relies on reading resumes, assessing extra-curricular activities, and conducting interviews. These manual assessments are available from the Ministry of Higher Education and Research. A separate, important aim is to investigate how prediction models fare with respect to bias, including whether they are unbiased in terms of ethnic background, socioeconomic status, etc. If more successful on prediction quality and fairness, one could begin using such models as decision support systems or, with the appropriate legal basis, as full replacements for admission officers.

To illustrate the efficacy of the predictive models we will analyze a simple policy: exclude a student from admission if her attrition risk is above a given threshold. We plan to calibrate our predictive models on data from 2008 to 2019. This allows us to use data from 2020 for policy evaluations and measure outcomes after admission. This is possible because we are only interested in the prediction part of the policy and assume that admitted students have no externalities on each other. We will demonstrate that this policy works in practice by leveraging a differences-in-differences model of student outcomes combined with local changes to admission criteria, e.g. the requirement of 6.0 GPA from high school at the University of Copenhagen from 2018 onwards.

This sub-project will contribute to a nascent literature in economics and behavioral science more generally on prediction policy problems [6]. Important examples within the literature include using prediction models to forecast drug addiction before opioids are initially prescribed [8] or combining prediction models with a framework of economic decision making to examine the efficiency and bias

of court rulings [7]. Previous work in educational admissions has examined pre-admission likelihood of attrition [e.g. 13–20]; however, no studies have had access to the size, breadth and scope of the Danish registry data for the full population. Moreover, although existing work has leveraged prediction tools in a matching context for refugee resettlement [21], no previous work has attempted to integrate prediction tools within matching procedures based on mechanism design such as those underlying the Danish admissions system. Consequently, these procedures do not have the same real-life desirable properties as Deferred Acceptance, notably stability [10] and non-manipulability [22]. As a consequence, our project would be the inaugural paper for combining prediction models with stable mechanisms, and moreover do this in the context of Danish administrative applications. We expect that a paper from this sub-project will be of general interest and may be published in a top-ranked journal depending on the strength of the results, or at least a top journal within applied work such as AEJ Applied Economics or AEJ Economic Policy.

Sub-project 2: Information disclosure. In this sub-project we will analyze an information disclosure aimed at targeting the prospective students’ preferences rather than changing the matching mechanisms. The disclosure was the launch of ‘UddannelsesZoom’, an online tool for comparing study programs in Denmark, which was launched in April 2015. This tool allows for easy and precise information about various study programs, including the ability to compare and measure students’ post-graduation employment prospects, income and personal well-being.

To assess whether the tool affected the choices of applicants, we leverage the timing of implementation. Applicants in KOT may apply as ordinary admission (Quota 1) where only GPA is considered or send a special application (Quota 2). As the tool was implemented in April, i.e. after deadline for Quota 2 on March 15 but before the deadline for Quota 1 on July 1, we can exploit that applications through Quota 2 had no exposure to the tool, whereas applications through Quota 1 had exposure. This allows us to examine whether students who applied through Quota 2 changed their submitted ranking in their Quota 1 application. We will assess whether students responded to new information about post-graduation labor-market outcomes for individual study programs. Applicants are allowed to change their preferences and the reversals in the year of 2015 should be particularly affected while other years can act as a control. In particular, we want to see whether the reversals in 2015 were driven by certain study features, e.g. labor-market outcomes and teaching quality.

While the approach outlined above may establish an effect, it may underestimate the effect size as some applicants were likely not aware of the tool. Therefore, we will survey a subset of applicants from 2013-2020 to measure awareness of the tool. The survey will allow us to correct the measured response to the information disclosure for compliance to compute a local average treatment effect. The survey will also allow us to see whether awareness and compliance have grown over time. Moreover, the survey will also enable us to examine whether those who responded to the information disclosure had lower dropouts and higher wages after graduation. The sub-project will contribute to the information disclosures which has been limited to smaller field experiments [9].

Perspectives. The sub-projects outlined above have the potential to open up a number of new directions for research and policy. The first sub-project will develop a framework of prediction policies in admission for higher education. The simple policy described in the sub-project can be extended to more complex policies if coupled together with a structural model of students’ preferences and application behavior. We intend to enrich our research in this direction later on. In addition, the prediction tools can also be applied directly in other policies, e.g. interventions targeted at students predicted to be at-risk. Finally, the framework could be of broader strategic interest for Danish policymakers and adapted to other contexts, e.g. combating fraud or making personalized medical treatments. The sub-project on information disclosure could also be integrated into a more elaborate structural framework. In addition, the survey, we intend to conduct, can be used to identify areas around Denmark where information about the tool is lacking and may need some promotion.

Timeline and organization. If funded, the project will begin immediately by setting up the project at Statistics Denmark. We would survey applicants to Danish universities in the summer of 2020 and applicants for earlier years in the fall. For both projects, the first draft should be ready within 1-1.5 years and also published within 2-3 years. In terms of project participants, everyone will participate in the first sub-project. The second sub-project will be conducted by Andreas and Emil.

References

- [1] Uddannelses-og Forskningsministeriet. Frafald blandt akademiske bachelorstuderende fra universiteterne inden for 4. år efter studiestart, jf. danske universitetersopgørelse, 2018. <https://ufm.dk/uddannelse/statistik-og-analyser/frafald/notat-om-frafald-og-studieskift.pdf>.
- [2] Styrelsen for Forskning og Uddannelse. Bruttoledigheden for nyuddannede på de videregående uddannelser fra 2012 til og med 2016, 2018. <https://ufm.dk/uddannelse/statistik-og-analyser/faerdiguddannede/aktuel-ledighed/aktuel-ledighed-2016.pdf>.
- [3] Uddannelses-og Forskningsministeriet. Nøgletal for universiteternes omkostninger, 2018. <https://ufm.dk/uddannelse/videregaende-uddannelse/universiteter/okonomi/nogletal-for-universiteternes-omkostninger>.
- [4] Beskæftigelsesministeriet. Nøgletal for udgifter til ledige dimittender, 2019. <https://www.ft.dk/samling/20191/almdel/ufu/spm/24/svar/1612650/2114261/index.htm>.
- [5] Atila Abdulkadiroğlu and Tayfun Sönmez. School Choice: A Mechanism Design Approach. *American Economic Review*, 93(3):729–747, 5 2003. ISSN 0002-8282. doi: 10.1257/000282803322157061. URL <http://pubs.aeaweb.org/doi/10.1257/000282803322157061>.
- [6] Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. Prediction Policy Problems. *American Economic Review - Papers & Proceedings*, 105(5):491–495, 5 2015.
- [7] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237–293, 2018.
- [8] Justine S Hastings, Mark Howison, and Sarah E Inman. Predicting high-risk opioid prescriptions before they are given. *Proceedings of the National Academy of Sciences*, 117(4):1917–1923, 2020.
- [9] Justine Hastings, Christopher A Neilson, and Seth D Zimmerman. The effects of earnings disclosure on college enrollment decisions. Technical report, National Bureau of Economic Research, 2015.
- [10] D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9, 1 1962. ISSN 00029890. doi: 10.2307/2312726. URL <https://www.jstor.org/stable/2312726?origin=crossref>.
- [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [12] Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456):885–890, 2019.
- [13] Serge Herzog. Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New Directions for Institutional Research*, 2006. ISSN 02710579. doi: 10.1002/ir.185.
- [14] Gerben W Dekker, Mykola Pechenizkiy, and Jan M Vleeshouwers. Predicting Students Drop Out: A Case Study. 2009. URL <http://www.win.tue.nl/~mpechen/research/edu.html>.
- [15] Dursun Delen. Predicting Student Attrition with Data Mining Methods. *Journal of College Student Retention: Research, Theory & Practice*, 13(1):17–35, 5 2011. ISSN 1521-0251. doi: 10.2190/CS.13.1.b. URL <http://baywood.comhttp://journals.sagepub.com/doi/10.2190/CS.13.1.b>.
- [16] Dursun Delen, Kazim Topuz, and Enes Eryarsoy. Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research*, 3 2019. ISSN 03772217. doi: 10.1016/j.ejor.2019.03.037. URL <https://linkinghub.elsevier.com/retrieve/pii/S0377221719302954>.

- [17] Sattar Ameri, Mahtab J Fard, Ratna B Chinnam, and Chandan K Reddy. Survival Analysis based Framework for Early Prediction of Student Dropouts. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16*, pages 903–912, New York, New York, USA, 2016. ACM Press. ISBN 9781450340731. doi: 10.1145/2983323.2983351. URL <http://dx.doi.org/10.1145/2983323.2983351><http://dl.acm.org/citation.cfm?doid=2983323.2983351>.
- [18] Dech Thammasiri, Dursun Delen, Phayung Meesad, and Nihat Kasap. A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition. *Expert Systems with Applications*, 41(2):321–330, 2 2014. ISSN 09574174. doi: 10.1016/j.eswa.2013.07.046. URL <https://linkinghub.elsevier.com/retrieve/pii/S0957417413005332>.
- [19] Lovenoor Aulck, Nishant Velagapudi, Joshua Blumenstock, and Jevin West. Predicting Student Dropout in Higher Education. 6 2016. URL <http://arxiv.org/abs/1606.06364>.
- [20] Anne-Sophie Hoffait and Michaël Schyns. Early detection of university students with potential difficulties. *Decision Support Systems*, 101:1–11, 9 2017. ISSN 01679236. doi: 10.1016/j.dss.2017.05.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S0167923617300817>.
- [21] Kirk Bansak, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. Improving refugee integration through data-driven algorithmic assignment. *Science*, 359(6373):325–329, 2018. ISSN 10959203. doi: 10.1126/science.aao4408. URL <http://science.sciencemag.org/>.
- [22] Alvin E Roth. The economics of matching: Stability and incentives. *Mathematics of operations research*, 7(4):617–628, 1982.