

## Report on the PhD dissertation thesis of Marin Drlje

A growing literature in education economics makes use of centralized student-school matching markets to provide causal estimates of various types of policy-relevant effects. This literature relies on the fact that in such markets, small differences in application test scores result in quasi-random assignments of students to schools. The dissertation produced by Marin Drlje provides three innovative steps forward within this literature; it combines econometric tools with empirical explorations and forms a cohesive whole of studies centered on this relatively new type of research design.

In the first essay, Marin revisits the standard studies in this literature, which employ the regression discontinuity design (RDD) or similar instrumental-variable approaches leading to local average treatment effect (LATE) estimates, based on students whose application test scores are located near school-specific admission thresholds. An example of an outcome variable in this literature are wages as a function of what type of school (field of study) a given student has attended. Marin argues both theoretically and empirically (based on Croatian data that unfortunately lacks a measure of outcome such as wages) that the samples used in this literature do not satisfy key assumptions of these methods required for valid causal inference. This insight is generated based on a novel idea of viewing the entire market structure as of the application moment as generating ex ante uncertainty in whether a student will be offered admission to a given program. Such randomness occurs even far away from admission thresholds due to applicants near one threshold bumping other applicants further up or down within the deferred acceptance matching algorithm employed in these matching markets. This ex ante uncertainty depends on who else applies to a given program in a given year and what the admission scores of these other applicants are. Marin quantifies this uncertainty by re-sampling from the applicant population and shows that these ex-ante admission probabilities indeed can dramatically differ between treated and non-treated students within typical RDD bandwidths, i.e., between students offered admission to a given program and those not offered such admission near an application threshold. At first glance, this would imply reducing the RDD bandwidths and sample sizes, but Marin points out that there are many quasi-random student-school assignments outside of a narrow bandwidth centered on school admission thresholds. He then proposes a new estimator that avoids the potential pitfalls of the standard RDD studies. Overall, this is an innovative approach that in future work should ideally lead to an empirical illustration comparing LATE estimates across these various sources of identifying variation.

The second essay in this thesis considers the consequences of non-compliance costs for typical LATE estimates. Within the context of the dissertation, non-compliance costs are for example associated with students applying to new schools one year after their initial assignment. They cover direct (re-)application costs and the costs of delaying by one year the entry in the first year of the desired program. Delaying entry implies delaying graduation and having less experienced than compliers at some given horizon after the initial assignments. Non-compliance costs challenge the exclusion restriction of the identification strategy that uses cutoff rules as instruments for program graduation. The paper formalizes non-compliance costs in a potential outcome framework. It highlights the identification threat of non-compliance costs, and proposes an alternative identification strategy under one key alternative assumption. The key alternative assumption is that non-compliance costs incurred by always-takers in the group with  $Z=0$  (where  $Z$  is the binary instrument) are the same as the costs incurred by never-takers in the group  $Z=1$ . This seems a reasonable assumption at least in the context of university program evaluation using admission cutoff rules as IV. Under the

key alternative assumption, the paper provides bounds for the LATE. The set identification is based on the method of Lee (2009). In a second part, the paper analyzes data from the Croatian admission process, and shows that the share of never-takers in the group with  $Z=1$  differs from the share of always-takers in the group with  $Z=0$ . This provides motivation for the previous econometric framework. This paper is sufficiently well developed to satisfy the requirements of awarding a doctoral degree. For the purpose of eventual journal submissions, I would suggest addressing the following concerns. First, an alternative strategy to solve the problem of non-compliance is to change the treatment of interest. In the context of school program evaluation, instead of estimating the effect of graduation, this means estimating the effect of starting in the first year of a given program. The paper already discusses this alternative strategy, but it would be nice to further explain the pros and cons of each alternative strategy. For example, it is clear that when one estimates the effect of being admitted in the first year, there are gains in terms of identification credibility, but the economic relevance of the treatment effect may be less convincing. There are probably other aspects that applied researchers should consider when choosing their econometric model. Second, the paper could convincingly show it can be applied in other empirical settings. For example, when programs are offered through randomized list and people can apply to several lotteries, this is certainly an issue (see the review in de Chaisemartin and Behaghel 2019; and in particular Le Barbanchon et al 2019). More generally, I think it is a general problem in program evaluation when individuals drop out of the program before its end. Third, the paper should show an empirical application where the outcome of interest is observed, so that the proposed estimators can be computed and compared to naïve estimators. Fourth, I think there is a bit of more work to be done about statistical precision. Standard errors need to account for the randomness of the trimming in the always-taker group when  $Z=0$ .

Finally, the third, empirical essay explores the importance of family linkages for college application choices. Much research studies college application choices and shows that these are frequently sub-optimal from the applicants' point of view. The estimation provided in this essay, which is based on standard RDD techniques, implies that college application choices are causally affected by education experiences and choices of one's older sibling(s). When an older sibling applies to and barely gets into a college, their younger sibling is more likely to attend this college and the same field of study. The paper not only presents new and interesting estimates of sibling causal effects, but it also considers several potential mechanisms underlying these effects. Sharing information about college experiences turns out to be the most likely channel. The paper is well developed and carefully executed. I have a few questions that might be addressed before the journal submission stage, need not be reflected in the approved thesis, but perhaps could be discussed as part of the defense: Are the results robust to switching from using the oldest sibling to the older sibling closest in age? Can you provide a clearer discussion of the extent to which major choices overlap with college choices in the results that are presented? I would also appreciate more detail on the practical policy recommendations that one can draw from this research – surely these have to do with the informational environment in which students make college application choices. What have we learned here that adds to our knowledge of how this information environment can be amended and improved?

Overall, this dissertation fully satisfies the requirements of Charles University for awarding the Ph.D. degree. It is on the forefront of this exiting new literature and one hopes that it will lead to high-impact publications. Marin's dissertation would be a well regarded thesis in many US or EU-15 schools. I congratulate Marin on his achievement.

Thomas Le Barbanchon