

REPORT ON MASTER THESIS
CENTER FOR ECONOMIC RESEARCH AND GRADUATE EDUCATION

STUDENT:	Aslan Bakirov
ADVISOR:	Paolo Zacchia
TITLE OF THE THESIS:	Revisiting Treatment Effects with Causal Forests

OVERALL ASSESSMENT

CONTRIBUTION: It is well-known that empirical research in economics has in recent decades benefited greatly from the “credibility” revolution grounded on statistical causal inference. More recently however, machine learning methods have been making a headway into the literature and practice of causal inference, aiming at improving the statistical properties and the interpretability of estimated causal effects. The diffusion of these methods in applied economic research is still in its infancy, and whereas the fruits from promoting it clearly hang low, any efforts towards this direction are worthwhile. This thesis offers such an attempt, specifically through the application of the now well-established “Causal Forests” model to the analysis of the gender wage gap. This is to the best of my knowledge a novel undertaking which, despite its limited scope in the present form, offers interesting insights and is fairly remarkable for a Master’s thesis.

METHODS: The empirical methods used in this piece of research are cutting-edge and certainly beyond the standard toolbox of applied empirical economists (though perhaps fairly standard in the machine learning domain). Despite some limitations, namely: the philosophical issues, which the author briefly mentioned in the thesis, about the “treatment” interpretation of gender; and the lack of a correction for selection into the labor force in the current version of the method (which the author also openly discusses), these methods offer novel insights about the main predictors of the gender wage gap, as random forests are well-suited to uncover patterns of heterogeneity.

LITERATURE: The thesis displays the author’s remarkable knowledge of the recent literature on machine learning and econometrics, in particular for causal inference. However, certain parts of Section 2 currently feel redundant. In particular, the introduction to causal inference as well as the discussion of machine learning methods other than tree-based ones (e.g. LASSO) feel partly out of place, though some readers may find them helpful for the sake of the big picture.

MANUSCRIPT FORM: The manuscript is well structured, although Section 2 could have been shortened and the final document could have benefited from some additional technical editing, with particular regard to the placement and size of the figures, and their accompanying notes. On the other hand, the mathematical equations and their notation are consistent and well-executed. It goes without saying that the manuscript displays a fairly good command of written English.

SUMMARY AND SUGGESTED QUESTIONS FOR THE DISCUSSION DURING THE DEFENSE: This is a strong thesis which displays the student’s potential for original research in applied statistics and economics, and his interest for frontier methodologies. In this regard, one

could ask the student-defendant whether he is aware of any recent developments of tree-based machine learning methods that allow to address the kind of endogeneity issues that economists are typically concerned with. I can anticipate that the student-defendant is likely aware of some such developments, yet his ability to effectively communicate them (possibly, with reference to the thesis' empirical application) may be revealing of his proficiency in this domain.

I recommend the thesis for defense.

TEXT ORIGINALITY CONTROL

I confirm that I acquainted myself with the report on the originality of the text of the thesis from

Theses Turnitin Ouriginal (Urkund)

The Turnitin report exclusively displays similarity indices in the order of 1-2%, with just a single Charles University December 2022 submission, likely related to the current manuscript, being close to 2%. Reassuringly, the submitted thesis is self-evidently original.

SUMMARY OF POINTS AWARDED

CATEGORY	POINTS
Contribution	28
Methods	30
Literature	18
Manuscript Form	19
TOTAL POINTS	95
GRADE	A

NAME OF THE REFEREE: Paolo Zacchia

DATE OF EVALUATION: August 28th, 2023



REFEREE SIGNATURE