Extended Abstract
Please do not add your name or affiliation

Paper/Poster Title	Paper/Poster Title

Abstract prepared for presentation at the 96th Annual Conference of the Agricultural Economics Society, K U Leuven, Belgium

4th – 6th April 2022

Abstract	***200 words max***

The evaluation of the effect a certain policy has presents several difficulties, from the justification of the identifying assumptions to the characterization of the heterogeneity in the impact. Over the years, several statistical approaches have been proposed to tackle the problem, but also many missing points have been found. Moreover, a conspicuous number of techniques has started flowing from the Machine Learning dimension to the causal inference domain. However, these techniques have never been validated in the Agricultural economic field. The objective of this paper is then to provide a set of meaningful comparison to understand which model is suitable for a vast range of scenarios tuned on the European agricultural sector. To achieve this, we make use of a simulated dataset, based on the FADN, where we set both the treatment effect and assignment, as well as a number other factors, to then assess which method is able to uncover the true causal effect. Eventually, the results from this paper will provide the validation to use Machine Learning model in our sector, as well as pointing which model is suited for each occasion.

Keywords	Causal Inference, EU Agricultural Policy, Machine Learning, Simulation, Econometrics
JEL Code	Mathematics and Quantitative Methods: C1 Econometric and Statistical Methods and Methodology: General see: www.aeaweb.org/jel/guide/jel.php?class=Q)

Introduction	***100 – 250 words***

The impact assessment of the European agricultural policies poses several challenges to a researcher. It is not possible to control the treatment assignment of the programmes, which are often targeted towards specific products or regions, and mostly work on a voluntary base. Therefore, farmers are selected or select themselves into the programmes expected to grant them greater benefits

These issues invalidate standard causal analysis and introduce estimation bias. Moreover, the methods that have been traditionally used to evaluate European agricultural policies as been recently shown to fall short of expectations in addressing this obstacle. In the meantime, an abundant literature on ML for causal inference has begun to bloom in recent years: these methods do not help in determining whether a particular treatment effect can be identified, but rather promise the researcher the possibility to perform an increased set of tasks with a data-driven approach. By relying

on the data for the selection of the model parameters and functional form, it is possible to minimize the number of decisions that the researcher has to make, and therefore also decrease research discretion and the model dependence of the results [but none of them has been validated for use in the agricultural sector specifically. This is particularly true for those newly proposed Machine Learning methods suggested in recent years for causal inference.

Starting from highlighting the theoretical assumptions at the base of the different models, we compare their accuracy in retrieving a simulated treatment effect in a simulated dataset.

Methodology	***100 – 250 words***

To deliver the afore-introduced results, we simulate a dataset on the basis of the European Farm Accountancy Data Network. Essentially, we exploit the pre-existing correlation among variables in the FADN to mimic the European Agricultural sector, and then proceed with simulating both the treatment effect and assignment. In doing this, we tweak different parameters in order to generate a wide range of scenarios to obtain a broader perspective of where each model is more suited. In particular, we focus on: noise level in the outcome; heterogeneity of the treatment effect; percentage of treated observations; magnitude of the treatment effect; strength of confounding, number of covariates in respect to number of observations.

We are going to evaluate 2 models from the Machine Learning domain, comparing them with 3 more established one from the statistical world: Bayesian Addittive Regression Trees and Causal Forests for the former; Propensity score matching, Diff-in-Diff and Coarsened Exact Matching for the latter.

Results	***100 – 250 words***

The results are still to be retrieved, as at the moment the research is in its preliminary stage. What we expect from this analysis is to provide guidance over the choice of which technique to use in different causal inference settings in agricultural economics. The models in fact will be detailed both from a theoretical perspective, minding the assumptions as well as the functioning, but also from a practical perspective, ranking them in the different scenarios according to defined measure of accuracy computed on the point estimate as well as the coverage interval of the causal effect. In addition to this, the simulated dataset as well as the code for the analysis will be made available for further research.

Discussion and Conclusion	*100 – 250 words*
The challenges that we have to face in estimating causal effect in the agricultural sector are well-known. The same cannot be stated about the most recent techniques for causal inference. The reason for this submission it to highlight the importance of these new advancements in our fields, and to discuss about what would be the best procedure to facilitate their adoption. A better understanding of the effect of the policies regulating our agricultural system allows to tackle the multitude of problems affecting it from an evidence-based perspective. In this way, the whole policy-making procedure is affected, resulting eventually in better policies and better outcomes.	