Referee Report of "The Impact of Machine Learning on Economics"

Research Question & Methods

In Athey (2018), the author states her research question clearly at the start of the Abstract part. The research question, which is adequately defined, could be summarized as: what is the early contributions of machine learning to economics and what is its future contributions?

To answer this question, the author conducts literature review of broad literatures from both economics and machine learning fields. The developing path of the paper can be summarized as follows:

The author starts with the discussion of the definition of machine learning and its early use cases. She states the narrow definition of machine learning as "a field that develops algorithms designed to be applied to datasets" (Athey, 2018, p. 3) with three main tasks divided into two branches: supervised and unsupervised machine learning. Then she introduces the task, use cases and common techniques of the two branches respectively with a focus on how these techniques could be used in economics research.

Next, the author contrastes machine learning with causal inference based on their key concerns and the way to address these concerns. To start with, she uses instrumental variable as an intuitive example to compare the difference in the aim of the two approaches: while machine learning pursuits "high explanatory power" (Athey, 2018, p. 5), causal models pursuit "an unbiased estimate of a causal parameter of interest" (Athey, 2018, p. 5). Then she analyzes the different key concerns of the two approaches. According to the author, the key concern in predictive models is "the tradeoff between expressiveness and overfitting" (Athey, 2018, p. 5). Thus, goodness of fit is an important evaluation standard for a predictive model. However, in causal models, there are a series of different concerns including spurious parameter estimates, "whether the assumptions required to "identify" a causal effect are satisfied" (Athey, 2018, p. 5) and behavioral and statistical assumptions. The author also illustrates how machine learning could address some of the concerns in economics along the way.

Then the author moves to evaluate applications of machine learning models to public policy. Besides discussing about the current application examples, the author also raises several concerns of the application, including "interpretability of models", "the question of fairness and nondiscrimination", "stability and robustness" and "manipulability" (Athey, 2018, pp. 8 - 9). She believes these concerns could be defined more formally and addressed by social scientists.

The author then makes a prediction of the creation of more new methods combining machine learning and causal inference. She lists a variety of problems in causal inference that could be tackled with machine learning methods. Among these problems, she discussed "average treatment effects", "heterogeneous treatment effects and optimal policies", "contextual bandits", "robustness and supplementary analysis", "panel data and difference-in-difference models" and "factor models and structural models" in detail (Athey, 2018, pp. 12 - 20).

In the end, the author makes broader predictions about how machine learning could impact economics in the future, not only including the way empirical work could be conducted, but also potential changes in various aspects of economics research and teaching.

Based on the above analysis of the structure of the paper, I can state that the author compellingly answers the research question with rich citation, appropriate organization of the existing literature as well as reasonable prediction of the potential future impact. The author clearly possesses a wealth of knowledge and unique perspectives in both machine learning and economics fields and her analysis is comprehensive and thorough.

The author mainly uses literature review and summarization to answer the research question and I believe it is appropriate and sufficient. The aim of the paper is to provide an overview of the impact of machine learning on economics, thus it is appropriate to develop the paper with analysis of broad existing literature in the two fields.

Literature

In this paper, the author broadly cites a variety of literatures in economics and machine learning. As the topic of the paper is large, the author admitts that she doesn't "attempt to conduct a comprehensive survey or reference every application in economics" (Athey, 2018, p. 1). Within the topics the author focuses on, she makes appropriate and sufficient citations.

The author uses public policy as an example of an economics field to illustrate the application of machine learning to economics. But besides using predictive models in public policy, there are many other applications of machine learning to other fields of economics. If the author considers extending the topics in her paper, there are the following additional citations that I would recommend:

In macroeconomics, Cook and Hall (2017) uses deep neural networks to make macroeconomic indicator forecasting. The authors use "models based on four different neural network architectures" to predict civilian unemployment and their best-performing model "outperforms benchmark models at every forecast horizon" (Cook and Hall, 2017, p. 1).

In game theory, Agrawal and Jaiswal (2012) discuss the application of machine learning to problems of adaptation in multi-agent systems. The authors propose an algorithm to maximize "the long-term expected payoff" in The Prisoner's Dilemma and they also uses "these ideas to develop an intelligent agent for a zero-sum game" (Agrawal and Jaiswal, 2012, p. 1).

In behavioral economics, Cramer (2017) discusses how artificial intelligence and behavioral economics might interact. The author primarily discusses three ideas: the use of machine learning to search for new behavioral variables, the misunderstanding of perceiving "human prediction as imperfect machine learning" (Cramer, 2017, p. 13), and how the application of machine learning could both "overcome and exploit human limits" (Cramer, 2017, p. 1).

Grammatical, spelling, or style errors

I found the following grammatical, spelling or styles errors while reading this paper:

1. In the Introduction part on page 1, there is an "an" in the sentence "the use case in economics is essentially the same use case that the ML tools were designed an optimized for" (Athey, 2018, p. 1), I believe it should be an "and".

2. Also in the introduction part on page 1, in sentence "Then, I provide an overview of the questions considered and early themes of the the emerging literature in econometrics and statistics combining machine learning and causal inference" (Athey, 2018, p. 1), there is a duplicated "the", causing grammatical error. On page 3, there is another duplicated "the" in the sentence "and when a human watches the the largest group," (Athey, 2018, p. 3).

3. On page 5, there are two errors in a sentence in parenthesis - "such demand shifters can be referred to as "con- founders" becaues the affect both the optimal price set by the firm and the sales of the product". Firstly, the "becaues" is a wrong spelling of "because". Secondly, "the affect" is grammatically wrong and it could be modified as "it affects".

4. On page 10 and 11, the "timeat" in sentence "The place where the econometric model with a causal estimate would do better is at fitting what happens if the firm actually changes prices at a given point in timeat doing counterfactual predictions when the world changes" (Athey, 2018, pp. 10 - 11) is confounding. I believe the author wants to express either "at fitting what happens if the firm actually changes prices at a given point in time" or "at doing counterfactual predictions when the world changes".

5. The next sentence on page 11 is also confounding - "Techniques like instrumental variables seek to use only some of the information that is in the data the clean or exogenous or experiment-like variation in pricesacrificing predictive accuracy in the current environment to learn about a more fundamental relationship that will help make decisions about changing price" (Athey, 2018, p. 11). I believe adding a parenthesis in the middle of the sentence would help make sense - "Techniques like instrumental variables seek to use only some of the information that is in the data (the clean or exogenous or experiment-like variation in pricesacrificing predictive accuracy in the current environment) to learn about a more fundamental relationship that will help make decisions about changing price". Also it could be easier to read if adding a "-" between price and sacrificing.

6. On page 22, in sentence "Thus, there is no obvious benefit from ML in terms thinking about identification issues" (Athey, 2018, p. 22), I believe there is a lack of "of" after "in terms".

7. On page 26, in sentence "firms want to know the return on investment on advertising campaigns,[2]," (Athey, 2018, p. 26), there is a duplicated comma.

Extension

The author basically uses literature review to have an overview of the impact of machine learning on economics and make prediction on the future impact of machine learning to the field of economics. Within the topics she focuses in this paper, the literature cited are abundant and appropriate. If there would be any extension to the methodology of this paper, I would recommend the author use some data to support her analysis and prediction. For example, when discussing about the application of predictive models to public policy, to illustrate the increasing

number of successful applications, she could list and analyze the amount of relevant papers submitted to or accepted by relevant journals in recent years to have a clear overview. In her prediction part, when she is predicting the impact of machine learning on future economics research and teaching, she could also use some data to show the increased amounts of interdisciplinary researches conducted or the increased interdisciplinary undergraduate or graduate programs announced in universities in recent years.

REFERENCES

Smalter Hall, Aaron and Cook, Thomas R., Macroeconomic Indicator Forecasting with Deep Neural Networks (September 29, 2017). Federal Reserve Bank of Kansas City Working Paper No. 17-11. Available at
SSRN: https://ssrn.com/abstract=3046657 or http://dx.doi.org/10.2139/ssrn.3046657

Agrawal, A., & Jaiswal, D. (2012). When Machine Learning Meets AI and Game Theory. Available at
http://cs229.stanford.edu/proj2012/AgrawalJaiswal-
WhenMachineLearningMeetsAIandGameTheory.pdf

Camer, C. (2017). Artificial intelligence and behavioral economics. NBER working paper. Available at https://www.nber.org/chapters/c14013.pdf