#Assignment 9

MACS 30000 **Dr. Richard Evans**

Referee Report with Extension

I have selected Susan Athey's paper "The Impact of Machine Learning on Economics" for assignment nine.

Question 1.

The author doesn't answer any particular research question and instead provides a comprehensive review of the role of Machine Learning (ML) in the field of Economics. She not only makes a thorough assessment of the contributions made by ML in economics, till date, but also predicts about its future contributions. Overall, she lays down a detailed study of "The Impact of Machine Learning on Economics".

Question 2.

The author compellingly answers about "The Impact of Machine Learning on Economics" in a nice and structured way. Firstly, she succinctly talks about the structure of her paper and the different themes relating ML to economics. For example, the second theme talks about how the empirical analysis in ML involves an algorithmic approach to compare estimates originating from different alternative models. On the other hand, in Economics, the researcher picks up a model based on principles and, at-least in principle, estimate it only once.

After introduction, she then defines ML as "a field that develops algorithms designed to be applied to datasets, with the main areas of focus being prediction (regression), classification, and clustering or grouping tasks." The author then talks about the two main branches of ML: Unsupervised ML and Supervised ML. Unsupervised ML involves clustering of observations into similar groups based on different algorithms like k- means clustering, community detection methods for networks, etc. Unsupervised ML is commonly used for analysing text, images and videos. Supervised ML deals with predicting an outcome variable, Y using a set of features (X). The author then discusses the usage of "off-the-shelf" ML techniques in traditional economics research and "prediction policy problems". The covariates created using ML techniques play an important role in standard econometric analyses. She then talks about the growing literature at the intersection of ML and causal inference. Out of the many different themes at the intersection of ML and causal inference which are booming with research, she discusses four of the themes at length. Then at last before concluding her paper, she talks about the broader predictions about the role of ML on Economics and how it would also change the nature of research questions being asked and the degree of inter-disciplinary collaboration. Overall, she puts up a nice structured and detailed answer which studies "The Impact of Machine Learning on Economics".

Question3.

The methods of answering the research questions are appropriate and sufficient. The author provides a comprehensive overview of what is ML and its different types. While talking about the use of "off-the-shelf" ML techniques in economics research, she took a balanced

approach by first portraying the research problem from the lenses of ML and Economics individually. She then would also draw out contrasts between the different approaches, if any. She defines the research problem well which makes it easy read for the person coming from a different background and without a rich exposure to Machine Learning techniques. In the section, wherein she talks about the growing research themes at the intersection of ML and Economics, the exposure is sufficiently detailed. She picks up on the four leading research areas. For each of the research area, she sets out a context either using the previous literature or defining the research problem well. She then talks about the recent literature and how the methods being developed help in solving different research problems. Wherever deemed necessary, she draws parallels between the different research works under the same theme (in terms of intersection between ML and Economics). Towards the end, she talks about "Broader Predictions about the Impact of Machine Learning on Economics". Here also, there is no hand waving and she thoroughly discusses each of her predictions. For example, she links the growing research involving panel data and ML techniques to "digitization". She briefly reasons out, how digitization is leading to creation of many panel datasets with records about individual behaviour and the amalgamation of ML techniques with Econometric research methods would contribute to the insightful study of these datasets.

Question 4.

The author has made a balanced use of citations and their paper is at the forefront of the current research work occurring at the intersection of ML and economics. The author has referred to old citations whenever necessary, for example, she informs the reader how application of ML's neural nets method in Economics is not entirely new and was tried by H. White (White, 1992).[1] She points to a large amount of recent literature while discussing about the current themes at the intersection of ML and economics which have got to researchers attention. Given the task of summarising "The Impact of Machine Learning on Economics" and making predictions about the future, I think their use of numerous citations is justified and the author has put them to a good use. On one hand, there are citations which are relatively old and are used to provide a context or explain the application of ML technique to an economics problem in a better way. On the other hand, there are very recent citations to capture the ongoing research work at the intersection of ML and economics and also to provide basis to the author's predictions about the role of ML on Economics in future.

Question 5.

The following grammatical, spelling, or style errors were found in the research paper:

"Next, I review some applications of ML in economics where ML can be used off-the-shelf: the use case in economics is essentially the same use case that the ML tools were designed an optimized for."

There's a typo at the end of the sentence and it should be "and" instead of "an".

Hence, it should have been: "Next, I review some applications of ML in economics where ML can be used off-the-shelf: the use case in economics is essentially the same use case that the ML tools were designed and optimized for."

[1] H. White. Artificial neural networks: approximation and learning theory. Blackwell Publishers, Inc., 1992.

“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, a literature that is providing insights and theoretical results that are novel from the perspective of both ML and statistics/econometrics.”
The use of ‘the’ twice in the starting of the sentence is wrong. It should be corrected to:
“Then, I provide an overview of the questions considered and early themes of the emerging literature in econometrics and statistics combining machine learning and causal inference, a literature that is providing insights and theoretical results that are novel from the perspective of both ML and statistics/econometrics.”

“In other words, the goal is to construct $\hat{\mu}(x)$, which is an estimator of $\mu(x) = E[Y|X = x]$, in order to do a good job predicting the true values of Y in an independent dataset”
The punctuation mark (.) is missing at the end of the sentence. The correct sentence is:
“In other words, the goal is to construct $\hat{\mu}(x)$, which is an estimator of $\mu(x) = E[Y|X = x]$, in order to do a good job predicting the true values of Y in an independent dataset.”

“An instrument in this case might be an input cost for the firm that shifts over time, and is unrelated to factors that shift consumer’s demand for the product (such demand shifters can be referred to as “confounders” becaues the affect both the optimal price set by the firm and the sales of the product).”
“Because” is misspelled as “because” and it should be followed with “they” in place of “the”.
The correct sentence is: “An instrument in this case might be an input cost for the firm that shifts over time, and is unrelated to factors that shift consumer’s demand for the product (such demand shifters can be referred to as “confounders” because they affect both the optimal price set by the firm and the sales of the product).”

“There is a small literature in ML referred to as “inverse reinforcement learning” (Ng et al., 2000) that has a similar approach to the structural estimation literature economics; this ML literature has mostly operated independently without much reference to the earlier econometric literature.”
There’s a missing “in” when referring to the structural estimation literature. The correct sentence should have been:
“There is a small literature in ML referred to as “inverse reinforcement learning” (Ng et al., 2000) that has a similar approach to the structural estimation literature in economics; this ML literature has mostly operated independently without much reference to the earlier econometric literature.”

“In the application of using mobile data to do credit scoring, a concern is that consumers may be able to maniplate the data observed by the loan provider (Bjorkegren and Grissen, 2015).”
“Manipulate” is spelled wrongly. It should have been:
“In the application of using mobile data to do credit scoring, a concern is that consumers may be able to manipulate the data observed by the loan provider (Bjorkegren and Grissen, 2015).”

“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.”
There’s a style error as there’s no space left between “time” and “at”. The correct version is:
“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 time at doing counterfactual predictions when the world changes.”

“The advantages of the forest weighting function are that is data-adaptive as well as model-adaptive.”
The sentence is incorrect because of missing “it” before “data-adaptive”. The correct form is:
“The advantages of the forest weighting function are that it is data-adaptive as well as model-adaptive.”

“This article has al discussed the first three predictions in some detail; I will now discuss each of remaining predictions in turn.”
Possibly an extra “al”, the sentence could be corrected to:
“This article has discussed the first three predictions in some detail; I will now discuss each of remaining predictions in turn."

Question 6.
In section 4.6 of the paper, “Factor Models and Structural Models”, the author talks about structural models of consumer choice which are widely studied by researchers in the field of economics and marketing. The author informs how ML has developed techniques to obtain estimates like Bayesian method, Markov Chain Monte Carlo. These models have been used to understand consumer choices and consumer’s response to changes in price. My extension of research method involves using ML for image and text analysis. Suppose, in addition to consumer choices over menu items in different restaurants we also have data from his social feed, say Facebook or Twitter. The large data from the social feeds then can be analysed using ML image and text analysis techniques to inform our structural model of consumer choice. In this manner, we would have added another useful dimension to our consumer choice model and would expect better predictions.

References:

Athey, Susan. "The impact of machine learning on economics." In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press, 2018.

D. Bjorkegren and D. Grissen. Behavior revealed in mobile phone usage predicts loan repayment. 2015.

A. Y. Ng, S. J. Russell, et al. Algorithms for inverse reinforcement learning. In Icml, pages 663–670, 2000.