Book Project:
Econometrics with Machine Learning

Editors: Felix Chan and Laszlo Matyas

1 Motivation of the Volume

The success of Machine Learning techniques is reflected by its widespread commercial and academic applications. Apart from their abilities to provide accurate predictions, Machine Learning techniques have also been identified as effective tool-kits for extracting complex patterns from 'big data'. Their feature extraction abilities coincide with the fundamental purpose of most econometric investigations and therefore, Machine Learning provides a potential new avenue to advance econometric methodology. Such investigation is becoming prominent also due to the availability of 'big data' in economics and finance. Despite the obvious links, the differences in objectives and approaches, but also in vocabulary and notation between Machine Learning and econometrics creates many challenges in making their techniques readily applicable in economic analyses, on economic data. While the main objective of Machine Learning is to produce accurate prediction, econometricians also tend to focus on statistical inference and identification of causal relationships. This difference in objectives leads to several interesting questions that need to be resolved in order for the potential of Machine Learning to be fully realised in econometrics. The basic questions that need to be answered:

1. What are the similarities between existing econometric and machine leaning techniques?
2. To what extant can Machine Learning techniques assist econometric investigation? Specifically, how robust or stable is the prediction from Machine Learning algorithms given the ever change nature of human behaviour?
3. Can Machine Learning techniques assist in testing statistical hypotheses and identifying causal relationships in 'big data'?
4. How can we extend existing econometric techniques by incorporating Machine Learning concepts?

5. How can we elaborate new econometric tools and approaches based on Machine Learning techniques?
6. And the other way round: Is it possible to develop Machine Learning techniques further and make them even more readily applicable in econometrics?

In addition to these questions, computational difficulty remains a challenge as the complexity of data structures and models grows despite the sophistication of many Machine Learning algorithms. As the data structures in economic and financial data becomes more complex and models become more sophisticated, there appears to be many fruitful opportunities in developing both disciplines in conjunction rather than in isolation.

2 Structure and Main Topics of Volume

The main purpose of this volume is to help and promote the use of Machine Learning tools and techniques in econometrics, to show how Machine Learning can enhance and expand our econometrics toolbox in theory and in practice.

The main target audience of this volume are all those who are using econometrics in theory or in practice. Basically, volume may contain three types of contributions:

1. A comprehensive survey chapter on each of the topics listed below, from an econometric perspective.
2. Empirical applications which show how Machine Learning techniques can be applied to empirical econometric problems.
3. Any new results, both theoretical or empirical, would also be a welcome chapter.

The focus of the volume is to consider the following preliminary list of topics:

1. Model selection in linear models including variable and functional form selection as well as moment selection in the context of Generalised Methods of Moments.
2. Model selection in non-linear models including selection between linear and non-linear models and its relation to neural network and deep learning.
3. Econometric inference and causality with Machine Learning
4. Econometric forecasting and prediction
5. Policy evaluation including the modelling of treatment effects
6. Modelling non-stationary data with Machine Learning
7. Networks and Machine Learning

Topics 1 and 2 have a long history in econometrics. Some of the estimators in Machine Learning, such as the ridge estimator, also appear in conventional econometrics. However, the applications of the more general regularization and its connection to other penalty functions in econometrics still require some careful investigation. Novel applications of regularization in other econometric problems are also emerging in the literature, see for examples Liao (2013); Cheng and Liao (2015), which proposed LASSO type methods for moment selection in the GMM context. Despite its theoretical contribution, the empirical performance of these methods still require further examination. More importantly, what other novel applications are there? In terms of nonlinear models, the connection between some nonlinear time series models, such as the smooth transition autoregressive model, and the artificial neural network is well known but the identification on the number of layers and nodes in the context of economic or finance data is still an open area of research.

Topic 3 covers issue concerning hypothesis testing and identification of causality using Machine Learning techniques. While there are some results on the distributional properties of LASSO type estimators, see for examples, Knight and Fu (2000); Lockhart et al. (2014), the theoretical properties of Machine Learning techniques are often lacking which makes statistical inference difficult. Issues such as pre-testing have also just started being considered, see for example, Lee et al. (2016). Research on causal inference using Machine Learning also appears to be challenging with very little progress, see Pearl (2019). However, causal inference is at the crust of policy evaluation which is the focus of Topic 5. The aim of the topic is to provide an opportunity for researchers to develop further techniques for causal inference using Machine Learning beyond those proposed in Kreif et al. (2019).

Topic 4 is perhaps where the synergy is most natural. Though, related concepts in forecasting have developed independently between the two disciplines. One such example is the concept of ensemble forecasts in Machine Learning which shared a lot similarities to forecast combinations in time series econometrics. This topic aims to investigate further on the connection and between these related concepts. The topic also aims to distinguish between prediction and forecast. While forecast appears to be a concept concerning time, i.e., forecast future events, the definition of prediction is more ambiguous. While one can ‘predict’ future events, one can also predict consumer behaviour using discrete choice type models with cross sectional data. Thus, this topic will also investigate the synergy between classification problems in Machine Learning and Discrete Choice Models in econometrics.

Least squares is a popular loss function in Machine Learning techniques and a natural question is the impact of non-stationarity on these techniques. Given most macro-economics variables are non-stationary, this topic investigates what modifications, if any, are required for Machine Learning Techniques to be useful in modelling macro-economic relations.

Research on network is also becoming popular in economics and finance. At the core of the problem is the understanding of Graph, especially dynamic graphs where the structure evolves over time. While there are many Machine Learning algorithms for modelling static graphs, techniques for dynamic graphs are still relative scarce, see for example Kazemi et al. (2020) and it is the aim of this topic to provide further insights into this area.

REFERENCES

Cheng, X. and Z. Liao (2015). Select the valid and relevant moments: An information-based LASSO for GMM with many moments. *Journal of Econometrics 186*(2), 443–464.

Kazemi, S. M., R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart (2020). Representation learning for dynamic graphs: A survey. *Journal of Machine Learning Research 21*, 1–73.

Knight, K. and W. Fu (2000). Asymptotics for LASSO-Type Estimators. *The Annals of Statistics 28*(5), 1356–1378.

Kreif, N., K. DiazOrdaz, N. Kreif, and K. DiazOrdaz (2019). Machine learning in policy evaluation: New tools for causal inference.

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor (2016). Exact post-selection inference, with application to the LASSO. *Annals of Statistics 44*(3), 907–927.

Liao, Z. (2013). Adaptive GMM shrinkage estimation with consistent moment selection. *Econometric Theory 29*(05), 857–904.

Lockhart, R., J. Taylor, R. J. Tibshirani, and R. Tibshirani (2014). A significance test for the LASSO. *The Annals of Statistics 42*(2), 413–468.

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM 62*(3), 54–60.