Asian Journal of Economics and Banking

ISSN 2615-9821

http://ajeb.buh.edu.vn/Home

Two Cultures are Merging and Some Applications in Finance

Son Phuc Nguyen†

Department of Mathematics, University of Economics and Law, Vietnam National University, Ho Chi Minh city, Vietnam

Article Info

Received: 05/02/2020
Accepted: 16/3/2020
Available online: In Press

Keywords
Bayesian Inference, Business, Economics, Econometrics, Machine Learning, Probabilistic programming, Statistics, Uncertainty

JEL classification

C1, C4, C5, C6, C8

MSC2020 classification

62C12, 62F18, 68T37, 6207, 91B82, 91B74, 91B84

Abstract

Nearly two decades ago, prominent statistician Leo Breiman wrote an influential paper titled *"Statistical Modeling: The Two Cultures"* [17] where he presented fundamental ideas of generative models (statistical) and algorithmic model in studying complex datasets. Breiman urged statisticians to adopt a more diverse set of tools for data modeling. As of 2019, the algorithmic model has witnessed a rapid development for more than 20 years and this trend will be continuing for years to come. Nowadays, it is usually referred to as machine learning, or in particular, neural network, deep learning etcetera. The good news is that with substantial investments from leading technology companies like Google, Facebook, Uber, the Breiman's goal of diversifying data modeling tools can be pushed forward even further: the infrastructure now supports getting the best of the two worlds, namely, researchers, practitioners are able to flexibly incorporate various components of both generative model and algorithmic model into a hybrid framework which allows to prioritize between predictive power or interpretability of the models.
In this paper, first, the two cultures are briefly surveyed. Second, basic ideas of integration techniques are discussed together with supporting programming languages. Third, some exemplary implementations of integration with concrete applications are presented. Finally, certain recent relevant work in business and finance are introduced.

†Corresponding author: Son Phuc Nguyen, Department of Mathematics, University of Economics and Law, Vietnam National University, Ho Chi Minh city, Vietnam. Email: sonnp@uel.edu.vn

1 INTRODUCTION

In the year 2001, renowned statistician Leo Breiman wrote the sort of "provocative" article [17] in *Statistical Science* claiming

> There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown.

He went on with

> The statistical community has been committed to the almost exclusive use of data models. This commitment has led to irrelevant theory, questionable conclusions, and has kept statisticians from working on a large range of interesting current problems. Algorithmic modeling, both in theory and practice, has developed rapidly in fields outside statistics. It can be used both on large complex data sets and as a more accurate and informative alternative to data modeling on smaller data sets. If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools.

It has been almost twenty years since then and the statistical community has, to some extent, accepted the data science and machine learning revolution that Breiman referred to as the algorithmic modeling cultures. Eventually, all researchers, as well as practitioners, would like to use data to solve meaningful problems. The amount of data nowadays is hundreds of times larger than twenty years ago. Besides, data come from numerous new sources with far greater complexities. The contemporary situation strongly validates Breiman's arguments.

In contrast, as machine learning's applications extend beyond traditional fields like computer vision, text processing or robotics, there has been an increasingly greater demand for effective methods for decision making under uncertainties. Economic, business and financial practitioners are particular interested in capturing and representing uncertainties because a large number of decisions are high-stakes and usually involve a lot of stake-holders. In other words, risks are primary concerns; they have to be classified and estimated for any solution to be accepted.

In a recent review [4] (on page 687), Athey et al. stress

> The ML techniques often require careful tuning and adaptation to effectively address the specific problems that economists are interested in. Perhaps the most important type of adapta-

tion is to exploit the structure of the problems, e.g., the causal nature of many estimands; the endogeneity of variables; the configuration of data such as panel data; the nature of discrete choice among a set of substitutable products; or the presence of credible restrictions motivated by economic theory, such as monotonicity of demand in prices or other shape restrictions.

Sections 2 and 3 contain brief reviews of the main ideas of statistical modeling and machine learning. In section 4, we first present the reasons why an integration of the two modeling approaches are necessary for modern research. Next, we quickly go over some frameworks as well as programming platforms having been developed recently to address the need. The paper concludes with Section 5 containing a few applications in finance and banking.

2 STATISTICAL MODELING

Statistical modeling in the context of business and finance is studied under the name of econometrics (see [2], [35] and [64] for some leading textbooks). The common theme is to specify a target, an estimand which is a function of a joint distribution of the data. The data are typically some sets of variables (which usually include both categorical and quantitative variables). The target is then one parameter of a statistical model that describes the (conditional/unconditional) distribution of the data. The statistical model is often specified by a set of parameters which will be estimated from a random sample of the population by selecting the parameter values that best fit the full sample, using an objective function such as sum of squared errors or the likelihood function. The emphasis is placed on the quality of the estimators of the target, classically measured through the large sample efficiency. Researchers usually find point estimates and confidence intervals in terms of standard errors.

One example is the ordinary least square covered in any econometric book where an outcome Y_i is modeled conditioned on some regressor X_i. Suppose we know that

$$Y_i|X_i \sim N(\alpha + \beta^T X_i, \sigma^2)$$

There are two parameters $\theta = (\alpha, \beta)$ to be estimated. Without much explanation, a least squared error is used, i.e.

$$(\hat{\alpha}, \hat{\beta}) = \text{argmin}_{\alpha,\beta} \sum_{i=1}^{N} (Y_i - \alpha - \beta^T X_i)^2$$

If the model is correct, the least-square estimator is essentially the best estimator with all proven excellent properties like unbiasedness, consistency, i.e. best linear unbiased estimator (BLUE). In addition, it is also the maximal likelihood estimator and it has large sample efficiency properties. Some textbooks, for instance [2], barely mention the model building step assuming data to be in a good shape for methods like OLS. This might be the case in the past, but, certainly not the case today.

Statistical modeling, especially the bayesian framework, is, however, very

strong on expressing uncertainties. In many circumstances, it offers valuable insights, and in some cases, causality explanations for the phenomena of interest. This is where the algorithmic modeling approach is originally not designed for and where statistics complements algorithms effectively. Subsequent sections will clarify the point.

3 MACHINE LEARNING

In constrast to section 2, machine learning originates from computer science, thus, is all about designing algorithms (see [65] for a widely cited textbook). The ultimate goal is to make predictions about some target variables given other observed variables (usually called features). The predictions can be continuous (such as determining price, remaining useful life or customer lifetime value and so on) or discrete (such as classification of handwritten digits, customer churn, fraud detection etcetera). One common scenario is where the available information is noisy as well as limited, which means feature engineering (constructions of new variables) is required, since the raw data is by itself not enough to produce satisfactory results.

One of the difficulties for statisticians and econometricians in reading machine learning literature is new terminology. Instead of model being estimated, it is being *trained*. Well-known statistical terminology such as regressors, covariates, predictors are referred to as *features*. Unlike econometrics, model assumptions are not big concerns (at least in practice). Traditionally, machine learning has two major branches: supervised learning and unsupervised learning.

In supervised learning, data come with labels. In other words, we know the ground truth, for instance, in the customer churn problem, there are characteristics of a customer such as number of transactions, total values of transactions and so on, and the ground truth is whether she or he leaves the service/business or not (see [38] for more details). Another example is stock price prediction (back testing), features (characteristics) of a particular stock include past stock prices (before current date), past financial reports, headlines/news related to the ticker symbol etc., and the ground truth is the today price, that is, assuming stock information up to the previous day, we predict the stock price of today (see [23], [54] for comprehensive reviews of techniques). The former example is an instance of the classification problem and the latter is an instance of the regression problem.

In unsupervised machine learning, the goal is to understand the structures of data. A great number of data contain intrinsic structures, for instance, images are not arbitrary collections of pixels, but are composed of objects like humans, animals, landscapes etc, and time series or any temporal data contain dynamics or dependencies, not just random ups and downs. For an overview of unsupervised machine learning, see [29]. One classical problem in this areas is clustering where we try to group data into clusters sharing important characteristics. For clustering algorithm reviews, see [42], [48], [26], [66]. For appli-

cations of clustering, see [61], [45]. Another increasingly important problem in unsupervised learning is data representation. Actually, in one way or another, this problem has been studied long before the term "unsupervised learning" was invented. For example, in business and finance, there are a wide variety of categorical variables which need to be converted into some sort of numeric values in order to be fed into models. Straightforwardly, in econometrics, dummy variables, and in machine learning, one hot encoding are used for this purpose. Classical encoding methods suffer from many shortcomings, not to mention the inability to be used in many other situations such as encoding sentences, paragraphs and documents in natural languages. Therefore, there is a great demand for new representation methods. It should be emphasized that not only categorical variables but also numeric variables need representations. Dimensionality reduction has been studied for a very long time with algorithms like principal component analysis (PCA) etc. A wide range of modern algorithms have been invented to introduce nonlinearity and other properties which PCA and other earlier algorithms cannot capture. For example, autoencoders based on artificial neural network represent data much more flexibly, see [16], [32], [52] and [58]. Besides, better representations have been researched for lots of other purposes, see [34], [22], [3].

One very popular machine learning technique in the last ten years is artificial neural network (ANN). This is actually composed of a vast number of different algorithms solving a wide range of problems with applications from robotics, automatic driven cars to marketing, management and finance. In a nutshell, neural networks are tunable nonlinear functions with many parameters.

Parameters θ of a neural network are called weights of the network. The output y is modeled as a nonlinear function of weights θ and inputs x

$$p(y^n = 1|x^{(n)}, \theta) = \sigma\Big(\sum_i \theta_i x_i^{(n)}\Big) \quad (1)$$

where σ is an activation function which introduces nonlinearity to the model.

Multi-layered neural network can be thought of as a composition of functions (which are called layers)

$$y^{(n)} = \sum_j \theta_j^{(2)} \sigma\Big(\sum_i \theta_{ji}^{(1)} x_i^{(n)}\Big) + \epsilon^{(n)} \quad (2)$$

Estimations of parameters are then computed by optimizing some variant of the gradient descent algorithm. The process is called training and usually corresponds to maximizing likelihood or penalized likelihood of the parameters.

The following formula summarizes basic ANN

ANN = nonlinear function + basic statistics + basic optimization

Recently, Deep learning (DL) has been mentioned so frequently. In short, it is a very large ANN with many architectural and algorithmic innovations. Usually, DL contains many layers, has various choices for the activation functions σ (RELU, softmax etc.), include some new tricks like dropout and so on. The major differences between a DL and a classical ANN include vastly

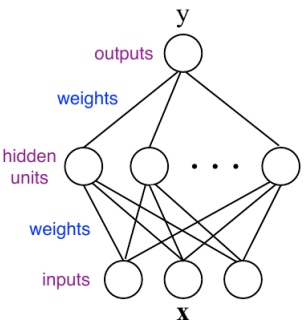

Fig. 1 Source: [30]

larger datasets, vastly larger computational resources (graphical processing unit-GPU, cloud), much better supports and tools from technology companies (Torch, Tensorflow), not to mention hugely increased business investment and media hype. For more information, see the article [46] and the book [33].

4 TWO CULTURES ARE MERGING

Statistics is more concerned about causality, interpretability while machine learning is more concerned about prediction. There has been much effort lately to obtain the best of the two worlds with contributions from many branches of science, not solely from statistical and artificial intelligence community. One example is the four consecutive workshops in bayesian deep learning (from 2016 to 2019) organized by one of the most important conference in machine learning, NeurIPS (Conference on Neural Information Processing Systems) to bring together scientists in every place working on the intersection of the two fields. In addition to complementing each other, machine learning and statistics indeed need to be integrated to tackle extremely important problems. In a lot of high-stakes decision making situations such as medical diagnosis or autonomous driven vehicles, caturing and reporting predictive uncertainty is essential since there is no single best model and a mistake can be a matter of life and death. Let us recall that predictive uncertainty consists primarily of:

1. *Aleatoric uncertainty:* comes from noisy data. For example, our observed labels might be noisy, maybe, as a result of measurement imprecision or human errors etc. This type of uncertainty is not reducible even when more data arrive.

2. *Epistemic uncertainty:* usually comes in two kinds:

   (a) structure uncertainty or model selection uncertainty: we do not know which model structure should be constructed, or more generally, what type of models is good to interpolate/extrapolate.

   (b) uncertainty in model parameters that best explain the

observed data: we do not know whether parameters that fit the current observed data will predict/generalize well with respect to unobserved data coming in the future.

3. *Distribution shift:* In [51], dataset shift is defined as a situation where the joint distribution of inputs and outputs differs between the training and test stages. This situation arises when the environment is non-stationary (i.e. the environment changes between the time to train models and the time to use models to predict). This is ubiquitous, especially in business and finance since markets are inherently non-stationary; conditions can change extremely fast (due to new policies, unforeseen disasters, human behaviors and so on). Distribution shift can be further classified into *covariate shift* (change in the distributions of independent variables), *prior probability shift* (change in the target variable) and *concept shift* (change in the relationships between the independent and the target variables).

For the former two uncertainties, see [30], [27] for an overview. For distribution shift, see the book [51] for a comprehensive introduction.

In order to incorporate uncertainties, probability framework is needed in machine learning. To put it briefly, it can be thought as inferring *plausible* models to explain observed data. A machine is then used to make predictions about future data, and take decisions that are rational given these predictions. It should be emphasized that since data can be consistent with more than one model, there may not exist a single best model for future data prediction as well as future data decision making. For more information, see the classic texts [19], [8].

A straightforward approach is to allow uncertainties in the weights of the artificial neural network. In formula (2), a vanilla ANN is determined by its weights θ's which are constants after training. In a bayesian neural network (BNN), the weights θ's are random variables with some prior distributions. The bayesian inference framework is employed in training process to obtain posterior distributions for parameters θ's. Unfortunately, constructing posterior distributions for θ's is usually intractable. Even obtaining a posterior sample via a Markov Chain Monte Carlo (MCMC) usually takes a very long time for the algorithm to converge. This is the place where variational inference comes in handy. Starting with a pre-specified family of probability distributions, posterior distributions are approximated by a member of the family which is "closest" (smallest distance) to the true distribution. The distance is normally the Kullback-Leibler divergence (KL divergence or relative entropy) which is a similarity measure between two probability distributions. Minimizing the KL divergence is equivalent to minimizing the evidence lower bound (ELBO). ELBO can be considered as a loss function that

can be optimized by a back propagation algorithm in machine learning. This explains how bayesian statistics benefits from machine learning. For more information on variational inference, see [15], [62] and [63]. Note that to reduce the computational complexity of the training process, not all layers of the neural network need to be stochastic (weights are random variables), some layers can be deterministic (weights are constant), see [59] for information on how to insert stochastic layers to common architectures.

Since probabilistic model development and the derivation of inference algorithms is time-consuming and error-prone, there has been a lot of work on developing *probabilistic programming languages (PPL)* for expressing probabilistic models as computer programs that generate data. The ultimate goal is to derive *universal inference engines* for these languages that do inference over program traces given observed data (in other words, Bayes rule on computer programs). A wide variety of such languages have been developed such as BUGS, Infer.NET, BLOG, STAN, Church etc. A great deal of progress has been made with implementations of inference algorithms like Metropolis-Hastings, variational inference, particle filtering, slice sampling etc. For an overview, see Figure 2 and articles [14], [24]. For a comprehensive discussions, see dissertation [40].

Since around the year 2015, major technology companies like Google, Facebook and Uber have invested heavily on developing platforms for probabilistic programming with the final goal to deploy it into business applications. Google has created new API called tensorflow probability to integrate probabilistic modeling into its artificial intelligence platform tensorflow. Together with a complete platform, Google offers the a new probabilistic language named Edward as an inherent part of the package. Not lagging behind are Facebook and Uber whose joint effort results in the probabilistic language Pyro integrated into Facebook artificial intelligence package Torch. For this line of research, see [10], [60], [57] and [13].

In addition to expressing all types of uncertainties in machine learning, the bayesian framework can also be used to find global optimum for an uncertain function f, meaning there is no closed form formula for the objective f. The only thing possible is evaluating f (with error) at some particular input x. Finding global optimum is posed as a sequential decision theory problem: suppose after having evaluated at three points, measuring the values of f at those points, $[(x_1, f(x_1)), \quad (x_2, f(x_2)), \quad (x_3, f(x_3))]$, which point x should the algorithm evaluate next, and where does it believe the maximum/minimum to be ? This has long been studied in machine learning with applications ranging from drug design to robotics, to any problem involving the optimization of expensive functions (i.e. require substantial computational resources). See [55], [56] for more details.

For a relative new direction of research, let us explore the impact of machine learning on economics and econometrics. In 2019, the American Eco-

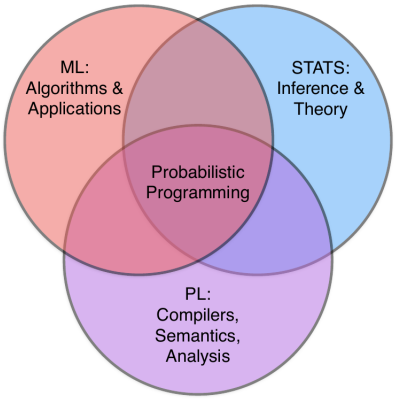

Fig. 2 Probabilistic Programming

nomic Association (AEA) has organized a course on machine learning and econometrics to educate and encourage economists, especially econometricians to adopt machine learning and data science. This can be seen as an effort of AEA just like that of Leo Breiman in 2001 (see [17]) to convince statisticians to bring algorithmic modeling into their work. The content (with numerous extensions) of the activities above is published in [5], [4] and [6]. In the annual review of economics in 2019 for AEA, Athey et al. have compared and constrasted the goals and methodologies of machine learning and economics. The highlight is that the adoptions should go in both directions. On the one hand, there are insights that machine learning might miss when focusing everything on predictive ability. On the other hand, in order to take full advantage of the availability of all sort of data, some of which have not been utilized effectively by economists before, economists should accept machine learning and data science as one of their standard methods. In fact, experts from both sides have worked diligently to combine the two methodologies to tackle many important problems, especially where deep insights into the phenomena (such as causality) are needed.

5 SOME APPLICATIONS IN FINANCE AND BANKING

Traditionally, statistics, in particular econometrics, has been used extensively for many decades. In the past, it is relatively successful in providing predictions and certain insights into financial markets as well as assisting banking operations. However, recently, with the incorporation of technology into every sector of any economy, the financial and banking services have been evolving at unprecedented speed, thus, create great demands for novel techniques, novel tools both for predictive analytics and for understanding of the intrinsic dynamics of the industries.

Our first example is the classical asset pricing problem in finance. Earlier approaches include stochastic differential equations and probability theory, see [7], [41], [47]. When the world economy was not highly connected and

markets were not linked closely, these methods proved to be adequate for basic understanding of the pricing process. Nevertheless, modern economies are so inter-connected that a change in one place can have impacts worldwide; hence, market dynamics are a lot more complex and harder to understand. That is to say new ideas and new technology are required to capture subtle information undetected by econometrics and other traditional techniques. In [37], Gu et al. utilize the feature engineering procedure in data science to construct new latent factors which are combinations (can be both linear and non-linear) of asset characteristics. To be more precise, the authors provide estimates of nonlinear conditional exposures and the associated latent factors. The workhorse is a polpular dimensionality reduction tool in machine learning, namely autoencoder neural network. Besides a smaller out-of-sample errors compared to other factor models, the machine learning framework allows the implementation of the economic restriction of no-arbitrage. See [21], [36], [43] for more work along this line of research.

The second example is the forecast problem in financial markets. Stock markets attract a great deal of attention around the world. Not only researchers but also investors relentlessly seek new methods to stay ahead in the forecasting game. State-of-the-art econometrics and machine learning are employed with the deployments of large infrastructure to gain advantages in this area. Early adoptions of machine learning algorithms are in [25], [39] and [9] where classical architectures like multi-layer perception are experimented with some successes. These efforts are further investigated in [53] and [28], especially in [28] with all cutting-edge sequence machine learning models such as recurrent neural network, long-short-term network (LSTM) and convolutional neural network (CNN). See [54] for a most recent review. As for the Vietnamese stock markets, the work in [49], [50] considers bayesian average techniques in a novel hybrid model to better predict the market stock indices.

The third example comes from banking industry. Chakraborty et al. in [18] give a review of machine learning in the context of central banking and policy analyses. Policy analysis is customarily built around causal inference where causality is obtained by comparing the effect of policy with a randomized control group. Actually, a policy problem in the twenty-first century can be divided into a *prediction* and a *causal inference* component. In this sense, machine learning and econometrics can be seen as mutual extensions of each other where econometrics has developed elaborate tools like instrumental variables, regression discontinuity, difference-in-difference analysis etcetera for causal inference while machine learning is all about prediction with correlations between variables. More interactions in banking industry can be seen in [1], [12], [13], [44].

6 AN APPLICATION WITH THE VIETNAMESE STOCK MARKET

In order to illustrate the benefits of integrating techniques from the two cultures, we consider a problem of forecasting market index in the Vietnamese stock market, in this case, the VN30 index. The data are obtained from January 5th, 2015 to February 21st, 2020. Besides the index itself, we collect close prices from the same period for the top three companies, namely Vingroup (VIC), Vietcombank (VCB) and Vinamilk (VNM). These three stocks serve as signals to help predict the VN30 index, in other words, they are covariates in the time-dependent linear regression with response variable VN30. For the historical index data, we use a time series decomposition into a local linear trend component, a seasonality component and a simple autoregressive with lag 1 component. The framework is the bayesian multivariable time series. All codes are written in Python with the Google tensorflow probability platform.

Traditionally, certain Monte Carlo Markov Chain (MCMC) algorithm will be employed to generate samples for the posterior distributions of the model parameters. For comparison, we have tried the Hamiltonian Monte Carlo (HMC) (see [11]), one of the state-of-the-art algorithms in the MCMC family. However, it takes so long to run; we cannot get the parameters to converge although HMC has been running on a Google colab virtual machine stronger than any personal computer. Therefore, we are obliged to switch to the machine learning framework using variational inference instead of random walk like the HMC. The advantage is that minimizing the Kullback-Leibler divergence (KL divergence or relative entropy) in variational inference is equivalent to maximizing the Evidence lower bound (ELBO) (see [67]) which is a loss function that can be optimized by some backpropagation algorithm in machine learning (see [20]). The time needed to finish the training process is under 30 minutes, so the algorithm has the ability to scale up to much larger size problem than the current experiment.

Figure 3 contains the density plots for the posterior parameters. In this study, the sample size for each parameter is 2000. The samples are used to generate the plots as well as to construct the forecast distribution for the VN30 index. The forecast horizon (number of days to forecast into the future) is 12 days.

Figure 4 consists of the density forecasts for the VN30 index for every 12 days into the future. The vertical red line in each graph is the true index value from the test set (12 last days in our dataset). As can be observed in Figure 4, the maximum a posterior probability (MAP) estimate is a pretty good estimate in these forecasts with the average root mean square errors for the 12 days being around 27 points, approximately 3% of the true daily index value.

For uncertainty evaluation, naturally, the standard errors increase from day 1 to day 12 as shown in the plots. Note that there are plenty of rooms for improvements in this regard. In subsequent papers, we will explore and develop mixtures of techniques which cap-

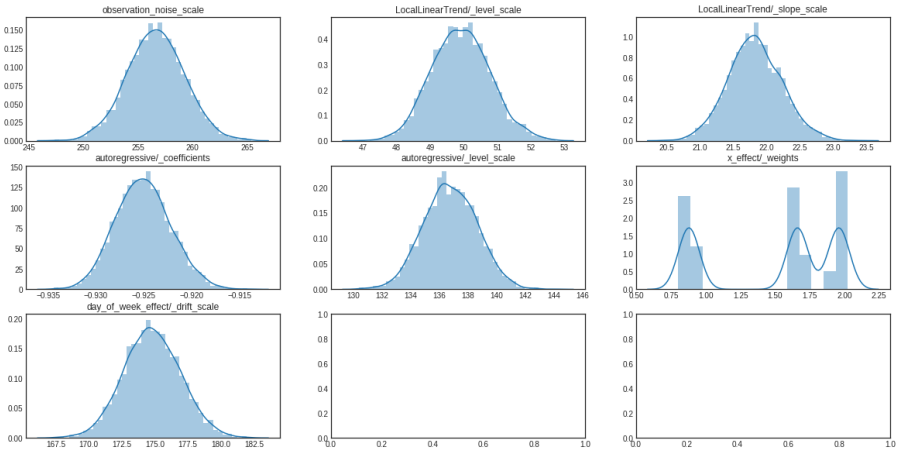

Fig. 3 Posterior Parameter Densities

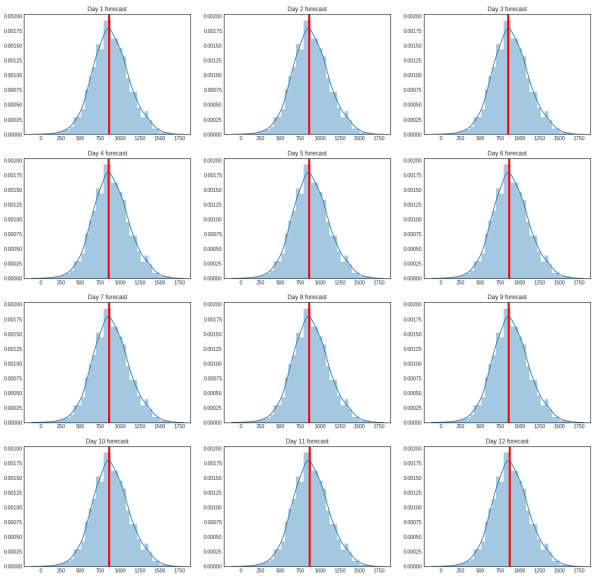

Fig. 4 Forecast Densities for 12 days

ture uncertainties much better, and in general, produce better predictions.

7 CONCLUSION

In this paper, we have had a look back at the influential paper on the two cultures of Breiman in 2001. It can be said that Breiman's ideas stand the test of time and we are still in the process of realizing his brilliant vision. With contributions from a great number of scientists from all branches of science and the significant investments from large enterprises, the two modelings, the two cultures, will integrate more and more intensively hereinafter, generating remarkable products in every place. Furthermore, the amount of data continues to grow exponentially, and it is in-

creasingly harder and harder to find any industry or any branch of science that does not have a demand for utilizing data analysis and data science. Ubiquitous applications bring novel challenges which can be dealt with effectively if scientists from different cultures, with different expertise, join effort together to tackle.

ACKNOWLEDGEMENT

The author is thankful to the anonymous referee for the helpful suggestions.

References

[1] Aikman, D., Galesic, M., Gigerenzer, G., Kapadia, S., Katsikopoulos, K. V., Kothiyal, A., Neumann, T. (2014). Taking uncertainty seriously: simplicity versus complexity in financial regulation. *Bank of England Financial stability paper,* 28.

[2] Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

[3] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875.*

[4] Athey, S. (2018). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda*, 507-547. University of Chicago Press.

[5] Athey, S., & Imbens, G. (2019). Machine learning methods economists should know about, arxiv.

[6] Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives,* 31(2), 3-32.

[7] Back, K. (2010). *Asset pricing and portfolio choice theory.* Oxford University Press.

[8] Barber, D. (2012). *Bayesian reasoning and machine learning.* Cambridge University Press.

[9] Batres-Estrada, B. (2015). *Deep learning for multivariate financial time series.*

[10] Baudart, G., Hirzel, M., & Mandel, L. (2018). Deep probabilistic programming languages: A qualitative study. *arXiv preprint arXiv:1804-06458.*

[11] Betancourt, M. (2017). A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv: 1701.02434.*

[12] Bianchi, D., Buchner, M., & Tamoni, A. (2019). Bond risk premia with machine learning. *USC-INET Research paper,* 19-11.

[13] Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Goodman, N. D. (2019). Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research,* 20(1), 973-978.

[14] Blackwell, A., Kohn, T., Erwig, M., Baydin, A. G., Church, L., Geddes, J., others (n.d.). *Usability of probabilistic programming languages.*

[15] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association,* 112(518), 859-877.

[16] Bourlard, H., & Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological cybernetics,* 55(4-5), 291-294.

[17] Breiman, L., et al. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical science,* 16(3), 199-231.

[18] Chakraborty, C., & Joseph, A. (2017). *Machine learning at central banks.*

[19] Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). *Probabilistic models of cognition: Conceptual foundations.* Elsevier.

[20] Chauvin, Y., & Rumelhart, D. E. (2013). *Backpropagation: theory, architectures, and applications.* Psychology press.

[21] Chen, L., Pelger, M., & Zhu, J. (2019). Deep learning in asset pricing. *Available at SSRN 3350138.*

[22] Coates, A., & Ng, A. Y. (2012). Learning feature representations with k-means. In *Neural networks: Tricks of the trade*, 561-580). Springer.

[23] De Prado, M. L. (2018). *Advances in financial machine learning.* John Wiley & Sons.

[24] DeRaedt, L., & Kimmig, A. (2013). Probabilistic programming concepts. *arXiv preprint arXiv:1312-4328.*

[25] Devadoss, A. V., & Ligori, T. A. A. (2013). Forecasting of stock prices using multi layer perceptron. *International Journal of Web Technology,* 2(2), 49-55.

[26] Diday, E., & Simon, J. (1976). Clustering analysis. In *Digital pattern recognition,* 47-94. Springer.

[27] Gal, Y. (2016). Uncertainty in deep learning. *University of Cambridge,* 1(3).

[28] Gao, T., & Chai, Y. (2018). Improving stock closing price prediction using recurrent neural network and technical indicators. *Neural computation,* 30(10), 2833-2854.

[29] Ghahramani, Z. (2003). Unsupervised learning. In *Summer school on machine learning*, 72-112.

[30] Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature,* 521(7553), 452.

[31] Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics,* 81, 709-727.

[32] Glorot, X., Bordes, A., & Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach.

[33] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT press.

[34] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, 2672-2680.

[35] Greene, W. H. (2000). Econometric analysis 4th edition. *International edition, New Jersey: Prentice Hall,* 201-215.

[36] Gu, S., Kelly, B., & Xiu, D. (2018). *Empirical asset pricing via machine learning* (Tech. Rep.). National Bureau of Economic Research.

[37] Gu, S., Kelly, B. T., & Xiu, D. (2019). Autoencoder asset pricing models. *Available at SSRN.*

[38] Hassouna, M., Tarhini, A., Elyas, T., & AbouTrab, M. S. (2016). Customer churn in mobile markets a comparison of techniques. *arXiv preprint arXiv:1607.07792.*

[39] Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. *arXiv preprint arXiv: 1402.7351.*

[40] Huang, D. E. (2017). *On programming languages for probabilistic modeling* (Unpublished doctoral dissertation).

[41] Hull, J., et al. (2009). *Options, futures and other derivatives/john c. hull.* Upper Saddle River, NJ: Prentice Hall.

[42] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR),* 31(3), 264 323.

[43] Kelly, B., Pruitt, S., & Su, Y. (2017). Some characteristics are risk exposures, and the rest are irrelevant. *Unpublished Manuscript, University of Chicago.*

[44] Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance,* 34 (11), 2767-2787.

[45] Kim, S.-Y., Jung, T.-S., Suh, E.-H., & Hwang, H.-S. (2006). Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert systems with applications,* 31(1), 101 107.

[46] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning, *nature,* 521(7553),436-444.

[47] Levy, H. (2011). *The capital asset pricing model in the 21st century: Analytical, empirical, and behavioral perspectives.* Cambridge University Press.

[48] Ng, A. Y., Jordan, M. I., & Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, 849-856.

[49] Nguyen, D. T., Nguyen, S. P., Nguyen, T. D., & Pham, U. H. (2018). On a new calibrated mixture model for a density forecast of the vn30 index. In *International econometric conference of Vietnam*, 466-473.

[50] Nguyen, S. P., Pham, U. H., & Nguyen, T. D. (2018). On a generalized method of combining predictive distributions for stock market index. In *International symposium on integrated uncertainty in knowledge modelling and decision making,* 253-263.

[51] Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (2009). *Dataset shift in machine learning.* The MIT Press.

[52] Salakhutdinov, R., & Hinton, G. (2009). Semantic hashing. *International Journal of Approximate Reasoning,* 50(7), 969-978.

[53] Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. K., & Soman, K. (2017). Stock price prediction using lstm, rnn and cnn-sliding window model. In 2017 *international conference on advances in computing, communications and informatics (icacci),* 1643-1647.

[54] Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies,* 7(2), 26.

[55] Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE,* 104(1), 148-175.

[56] Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, 2951-2959.

[57] Team, T., et al. (n.d.). Tensorflow probability, 2018-2019. *URL* https://github. *com/tensorflow/probability.*

[58] Torralba, A., Fergus, R., & Weiss, Y. (2008). Small codes and large image databases for recognition. In *2008 ieee conference on computer vision and pattern recognition*, 1-8.

[59] Tran, D., Dusenberry, M., Vander Wilk, M., & Hafner, D. (2019). Bayesian layers: A module for neural network uncertainty. In *Advances in neural information processing systems*, 14633-14645.

[60] Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M., Liang, D., & Blei, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787.*

[61] Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data mining techniques in erm: inside customer segmentation.* John Wiley & Sons.

[62] Wainwright, M. J., & Jordan, M. I. (2008). Introduction to variational methods for graphical models. *Foundations and Trends in Machine Learning,* 1, 1-103.

[63] Wainwright, M. J., Jordan, M. I., et al. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends®in Machine Learning,* 1(1-2), 1-305.

[64] Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data mit press. *Cambridge, MA, 108.*

[65] Wu, X., & Kumar, V. (2009). *The top ten algorithms in data mining.* CRC press.

[66] Xie, J., Girshick, R., & Farhadi, A. (2016). Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, 478-487.

[67] Yang, X. (2017). *Understanding the variational lower bound.*