Rejoinder

Applied Stochastic Models in Business and Industry

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2230 Published online in Wiley Online Library

Rejoinder to 'Deep learning for finance: deep portfolios'

We would like to thank all the discussants for their valuable insights and comments. Forbes and Maneesoonthon point out that the main challenge of deep learning for finance is coherently synthesizing the framework with traditional financial theory and financial econometric methods. Their discussion goes a long way in addressing this challenge. They also ask a number of important questions which we will address in our rejoinder.

- How can, in a deep learning framework, risk be measured and controlled effectively as in the familiar and intuitive Markowitz setup?
- How are the characteristic features of a time series (as for example positive autocorrelation, jumps, or volatility clustering) represented?
- How readily can a deep portfolio theory be implemented, especially considering that derivative structures may be required to represent the non-linear transaction of the deep neural net?

Sokolov provides a very clear description of the connection between modern-day deep learning and non-parametric statistical models. He also highlights an important link between deep auto-encoders and singular value decomposition when mining for low-dimensional statistical structure. Sokolov also stresses that what makes deep learning so applicable to finance is that there is a lot of returns data at very high frequencies (a.k.a. big data). We enjoyed reading his analysis of the IBB (iShares Nasdaq Biotechnology ETF) index, and it is re-assuring that singular value decomposition can uncover the similar sets of communal and non-communal sets of stock holdings. These *model-free* techniques that are solely *data driven* provide a powerful set of tools for modern-day indexing. Sokolov reminds us that the theory of deep learning lies in the functional estimation literature – this is formalized by the Kolmogorov–Arnold theorem (see section 2 of our paper). Another related area of current interest is using trees as layers. Applications to credit risk and merging of different data sources is also a fruitful avenue of exploration not yet covered.

To address the discussions, we focus on two issues: first, the difference between the data mining approach and traditional statistics metrics (such as R^2, t-values, p-values), and second, the link between traditional finance factor models and deep learning.

We close with a brief additional comment on the newly introduced deep frontier.

1. Data mining

The paradigm that deep learning provides for data analysis is very different from the traditional statistical modeling and testing framework that is commonplace in empirical finance that has been around for over 50 years. In sum, traditional modeling and fit metrics (R^2, $t-$values, p-values) and the notion of *statistical significance* has been replaced in the machine learning literature by out-of-sample forecasting and understanding the bias-variance trade-off. This might come as a large shock to a number of literature, particularly finance and accounting based empirical literatures.

Deep learning is data driven and focusing on finding structure in large data sets. The main tools are regularization and assessing the bias-variance trade-off in statistical estimators. Out-of-sample predictive performance helps gauge the optimal amount of regularization. There is a very Bayesian flavor to the modeling procedure. There are two key steps.

1. Training phase: pair the input with expected output, until a sufficiently close match has been found. Gauss' original least squares procedure is a common example.
2. Validation and test phase: assess how well your deep learner has been trained for out-of-sample prediction. This depends on the size of your data, the value you would like to predict, the input, etc., and various model properties including the mean error for numeric predictors and classification errors for classifiers.

Often, the validation phase is split into two parts.

2.a First, estimate the out-of-sample accuracy of all considered approaches (=validation).
2.b Second, compare your models and select the best performing approach based on the validation data (=verification).

If you do not need to select an appropriate model from several rivaling approaches (which means you are skipping 2.b), then you can just re-partition your data set such that you basically only have training and test set, without performing the verification of your trained model.

2. Traditional finance models

Forbes and Maneesoonthon ask about the links with traditional finance models. We would like to point out that many classical finance models are a special case of deep architectures, and we, in the succeeding text, briefly discuss traditional factor models as such an example. Specifically, this means that the powerful transparency and simplicity provided by many traditional finance models comes at the expense of a thorough exploitation of all possible models in 2.b earlier.

With the regard to the special characteristics of a time-series mentioned by Forbes and Maneesoonthon, we believe that, if desired, these can easily be incorporated through suitably chosen data sets in validation and verification. We agree with Forbes and Maneesoonthon that, for a simulated deep portfolio to be directly tradable, a special simulation architecture is required (in particular for the realistic inclusion of derivative costs). In fact, such a special simulation (or rather trading) architecture is inevitable if the data-driven creation of path-dependent portfolios is desired.

Rosenberg and McKibben [1] pioneered factor models in empirical finance. They arise as two-layer deep learning models with linear structures. Of course, this means that deeper structures and non-linearities might get missed, but the advantage is great interpretability. For example, given returns $\{r_1, r_2, \ldots, r_N\}$ on a benchmark asset (e.g. the SP500), we can learn a dictionary, denoted by $\{F_1, F_2, \ldots, F_K\}$, of K factors such that we can recover the output variable in-sample as

$$r_n = \sum_{k=1}^{K} B_{nk} F_k \quad \forall\, n = 1, \ldots, N,$$

where F_k are pricing factors. The regularization problem solves

$$\arg\min_{B,F} \sum_{n=1}^{N} \| r_n - \sum_{k=1}^{K} B_{nk} F_k \|^2 + \lambda \sum_{k,n=1}^{N,K} |B_{nk}|.$$

The first term is a reconstruction error (a.k.a. accuracy term), and the second is a regularization penalty to gauge the variance-bias trade-off to obtain good out-of-sample predictive performance and to avoid over-fitting. In sparse coding, the B_{nk} are mostly zeros. As we increase λ, the solution obtains more zeros, which highlights the inevitable trade-off in training for good out-of-sample prediction. There is an extremely fast scalable algorithm to solve this model: given the factors F, we solve for the weights (a.k.a. betas) using the L^1-lasso. And given the betas B, we solve for the latent factors using quadratic programming.

These models are special cases of deep learning – they are two-layer linear models. Deep learning generalizes this framework to multi-layer non-linear structures.

3. The deep frontier

In its purest form, Markowitz's *efficient frontier* formulates a *quid pro quo* in decision making, a *compromise* between risk and return. A particular strength of Markowitz's approach is its transferability and accessibility, and its inherent standardization.

The *deep frontier* generalizes Markowitz's concept by saying that, rather than plotting model mean against model standard deviation, we should plot validation performance against relevant constraints (where the constraints can originate in either implementation or validation). At its core, the deep frontier simply highlights that in Markowitz's approach, the assumption is made that the model is descriptive of the future, whereas the deep frontier demands a focus on validation and out-of-sample performance. (At its simplest, this is given when plotting validation mean against validation standard deviation, a basic data-driven comparison but normally much less clear than Markowitz's model.)

We recognize the truth in Forbes and Maneesoonthon's observation that, with the increased generalization and lack of standardization, a deep frontier may (at first) be challenging to include in professional investor's decision making routines. However, we would feel that out-of-sample performance (rather than ease-of-use) should be the guiding decision making principle, and we would expect the industry to adapt with time.

To summarize, a bright future for deep learning methods in finance.

Applied Stochastic Models in Business and Industry

James B. Heaton
Bartlit Beck Herman Palenchar & Scott LLP, and GreyMaths Inc.
Chicago, IL, USA

Nicholas Polson
University of Chicago
5807 S. Woodlawn Ave, Chicago, IL 60637, USA
Email: *ngp@chicagobooth.edu*

Jan H. Witte
Mathematical Institute, University of Oxford, and GreyMaths Inc.
Oxford, Oxfordshire, UK

Reference

1. Rosenberg B, McKibben W. The prediction of systematic risk in common stocks. *Journal of Financial and Quantitative Analysis* 1973; **8**:317–333.