AI Alpha Lab – Our DNA

By CIO Mikkel Petersen, AI Alpha Lab ApS

At AI Alpha Lab we strive to optimize the intersection between financial domain knowledge and machine learning with the purpose of making better forecasting models. We are driven by the fact that the investment decision in its very essence is a decision under uncertainty and we want to help investors with the two main problems they are facing.

1) Data – Investors are forced to navigate in an abundance of unstructured data and the variables investors try to forecast is non-stationary resulting in low signal-to-noise models.
2) Too simple models – The data problem is being exaggerated by the models that investors are applying to the data, since these were never meant to provide optimal decision making under uncertainty.

The investment problem is an extremely complex problem to solve. The problem investors' face tomorrow is not the same problem they faced yesterday and as such the variables of interest to investors; future returns, future volatility and future correlations, are highly time-varying. In order to solve this problem investors have resided to either high variance regression or parameter sensitive maximum likelihood (ML) optimized models. Fundamental investors, exposing their portfolios to risk premia and behavioral anomalies, are typically underfitted due to simple regression like model fitting. They have an approach that captures some of the structural information in data, but not all. The result is a long-term excess return expectancy, but with substantial time variation along the way. More quantitative managers typically optimize their exposure to the same risk premia or behavioral factors through ML and end up with models that are overfitted. The result is excellent historical test results, but poor out-of-sample performance due to the model being fitted to the stochastic noise in data. Our objective in AI Alpha Lab is to create models that are neither underfitted nor overfitted, but fitted to the causal structures available in data and not to spurious correlations and noise. Machine learning i.e. computer power, has facilitated a new way of fitting models that don't rely on data mined optimizations or high variance regressions, but on efficient sampling from a large number of generated models that all are capable of explaining data sufficiently well.

Machine learning in itself is complex analysis driven by computer power. AI is a great buzzword and an excellent marketing phrase, but let's be clear: Today, we do not have artificial intelligence. We have complex models that can solve very complex domain specific tasks, much better than humans, but the models do not understand the physical reality in which they live. They are totally dependent upon human intelligence in order to specify the objective and the framework for solving the problem at hand. The uniqueness of AI Alpha Lab's AI inference model comes from the team of quantum physicists sitting behind the model - and not from machine learning. Quantum physics is tailor-made for solving problems associated with significant uncertainty and uses probability to describe unobservable quantities, like future return on financial assets. Machine learning is what facilitates the use of concepts from scientific fields, such as quantum physics, enabling us to build models that, contrary to the models used today, are suited for solving complex problems like the investment problem.

Bayesian Inference

At AI Alpha Lab we employ a Bayesian multi-layer perceptron (neural network) to do our inference on financial assets. By doing the full Bayesian inference we move from a ML paradigm postulating that the model that best fit the data equals the true model and instead postulating that the simplest model that generalizes the best is the most likely to be the true model. Furthermore, the sampling estimation not only provides us with a more robust forecast of the future, we are also provided with the uncertainty surrounding the forecast, the predictive uncertainty.

Maximum likelihood paradigm

$$P(H|e) \neq P(e|H)$$

Full Bayesian inference paradigm

Likelihood
How probable is the evidence given that our hypothesis is true?

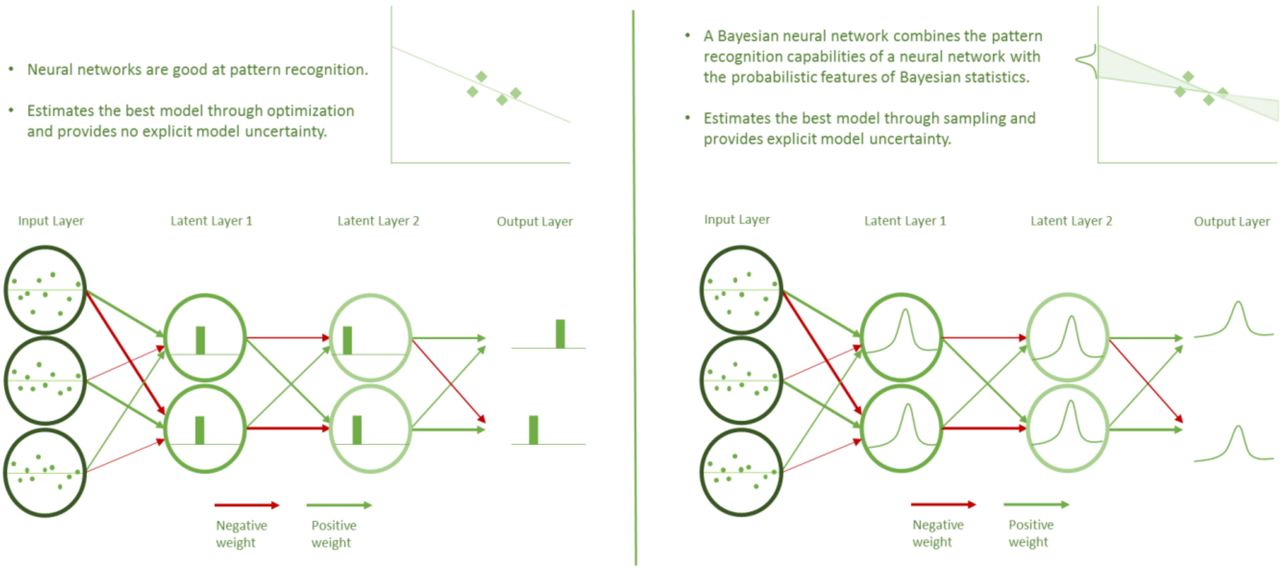
Prior
How probable was our hypothesis before observing the evidence?

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)}$$

Posterior
How probable is our hypothesis given the observed evidence?

Marginal
How probable is the evidence under all possible hypotheses?

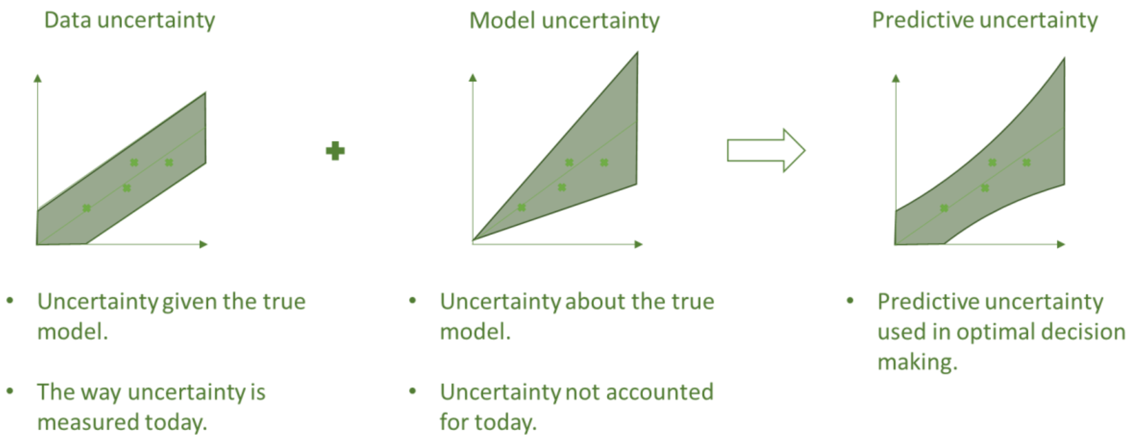
The challenge with a Bayesian inference is that the neural network becomes exponentially more complex since all parameter values are treated as distributions and not point estimates. The desired posterior distribution of models, from which we want to do inference, is intractable and therefore has to be approximated through efficient sampling algorithms. This is where our AI team stands out and our complex sampling techniques not only ensures that we fit causality in data, not correlations, but also provides a natural hair-cut of models through a Bayesian Occam's razor procedure, to ensure that we don't overfit.



A note on uncertainty

Optimal decision making under uncertainty requires a set of possible scenarios and the associated probabilities with which these scenarios are realized. The scenarios or outcome space is in itself useless if we can't quantify the uncertainty surrounding each future state, known as predictive uncertainty.

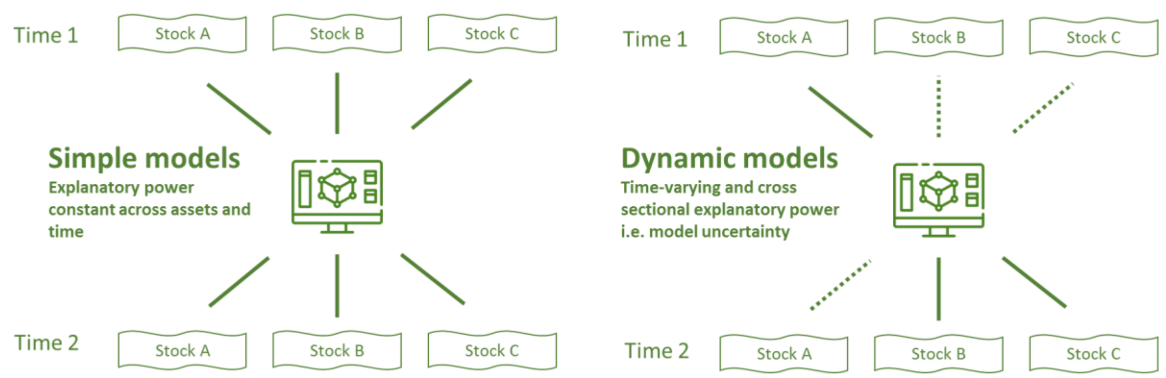
Predictive uncertainty is composed of two types of uncertainty, data uncertainty and model uncertainty. Data uncertainty is unstructured noise in data that has no causal explanatory power. Model uncertainty is a measure of how uncertain the model of choice is with regards to its ability to infer structure in data. Today, investors (probably unknowingly) approximates data uncertainty as being predictive uncertainty and are therefore not equipped with the necessary information for doing optimal investment decisions. Through the use of Bayesian neural networks, we can quantify model uncertainty and incorporate it into the investment process. Data uncertainty can to a large extent be diversified, but model uncertainty cannot. Today, model uncertainty is a large uncompensated risk inherent in most portfolios.



What do we want our models to do?

The simple models, processes and analyzes used by investors today all rely on the same universal assumption: The explanatory power is constant across assets and across time. In other words, models and processes are expected to explain different assets equally well both today and tomorrow. When dealing with non-stationary data, this assumption is not true!

AI Alpha Lab has developed a model that is able to validate the forecast of every single asset and inform the investor about the quality of the forecast. As an example, the model can have the same future return estimate on two stocks, but very different confidence. In this way, investors can filter out insignificant forecasts and do fair comparisons on the remaining assets. The model we have developed in AI Alpha Lab do provide us with return forecasts that are better than most other investors, but the true value-add lies in the fact that our model tells us when to use these forecasts. That enables us to be selective in our inference.



Robust investment principles

Signal theory tells us that whenever we can't estimate with a certain degree of significance i.e. if we face low signal-to-noise in our models, we are better off not estimating at all.

This is very relevant since most real world problems are associated with a high degree of uncertainty and the investment decision is a prime example of a complex dynamic forecasting exercise. Financial data is filled with correlations and little causality and as a result investors are prone to make false conclusions about their forecasting abilities. The smart and lucky investor is at best equipped with a weak predictive edge that over time can be applied successively.

Within investing there is a lot of information contained in what we don't know. By being able to estimate and quantify our unknowns we can select investment on that basis just like most others choose their investments from what they think they know. The difference is that we can estimate what we don't know, our uncertainty, much better and more reliably than what we think we know. This reverse engineered process is what enables us to provide better and more robust return and risk performance than the rest of the market.

At AI Alpha Lab we work with some of the most complex inference models, Bayesian multi-layer perceptrons. These are well described in theory, but few besides us has been able to practically apply them. In itself this might be of little interest for investment professionals if it wasn't for the fact that these models do provide us with better forecast capabilities than the rest of the investment crowd.

What might surprise most investors is that we, despite our better estimation capabilities, do our outmost to rely as little on our estimations as possible. We employ ensemble methods, to everything we do in order to be conceptually exposed to our models and not exposed to specifications and parameter selection. This is very much in contrast to general practice within investing and only highlights the fragility of most active managed portfolios, putting way too much emphasis on low or even insignificant signals.

Models are imperfect reflections of an underlying truth and their explanatory efforts when applied to problems associated with uncertainty is time-varying at best.

At AI Alpha Lab we play by one primary rule: We know that we don't know!

We do estimate and forecast returns on financial assets, but we also estimate the uncertainty associated with the estimates in order to scale our reliance on any single estimate. Furthermore, we ensemble everything we do in order to be as little exposed to our own ignorance as possible. Diversification across assets, processes and implementation is what makes our investment solutions robust across time. Specification risk and implementation risk (read our white papers about these risks on our webpage) are just two of many uncompensated exposures that most managers ignore or don’t understand. These exposures require exceptional skills in order to generate long term performance and the underlying premise of these are, that the manager is better than the market. For most asset managers this is not the case.

As investors the best we can hope for is to add together small significant edges (in statistics referred to as adding weak predictive models), exploit the correlation structure between these and end up with an investment process that provides robust future return expectations.

This material is provided for information purposes only and does not constitute, and shall not be considered as, an offer, solicitation or invitation to engage in investment operations or as investment advice. All reasonable precautions have been taken to ensure the correctness and accuracy of the information. However, the correctness and accuracy are not guaranteed, and we accept no liability for any errors or omissions. The material may not be reproduced or distributed, in whole or in part, without our prior written consent.

It is emphasized that investment returns shown are simulated and do not represent actual performance of assets during a period. If the simulated strategy had been implemented during the period, the actual returns may have differed significantly from the simulated returns presented. Past performance, whether actual or simulated, is not a reliable indicator of future results and the return on investments may vary as a result of currency fluctuations.

AI Alpha Lab ApS

Univate

Njalsgade 76

DK-2300 Copenhagen

Denmark

VAT 40 41 55 99