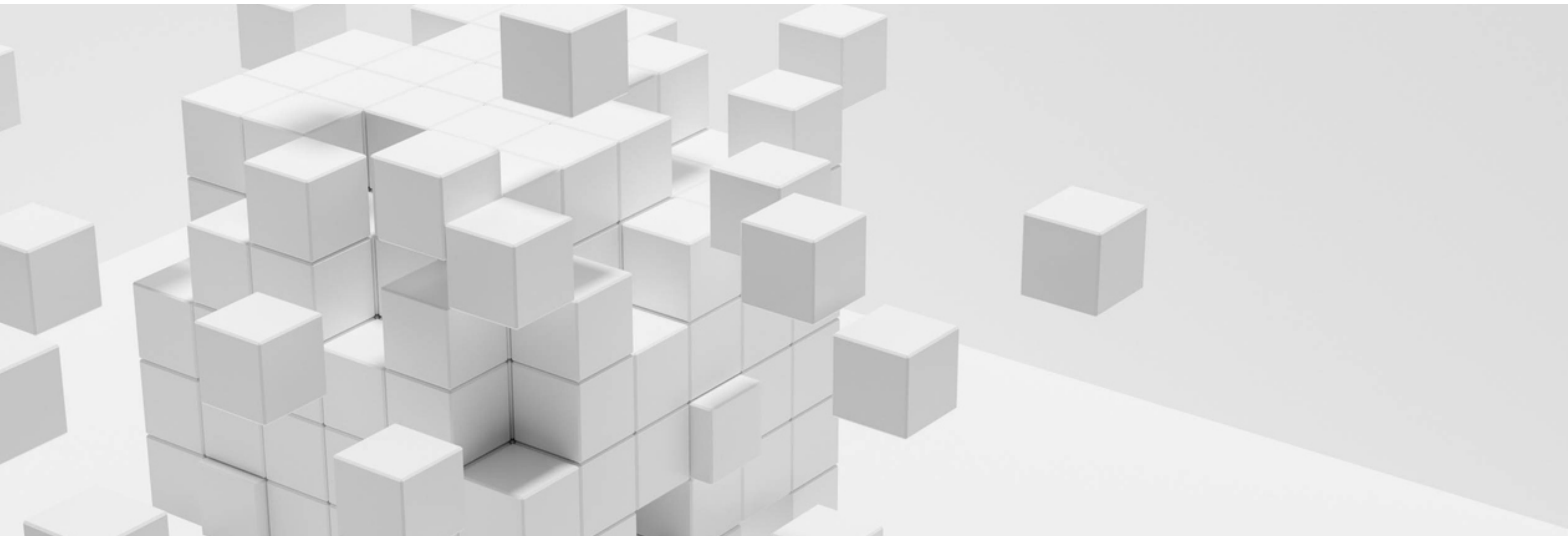
TD Global Investment Solutions

Whitepaper 10 Minutes
June 2023



Recent Enhancements to the TD Epoch Core Model

The TD Epoch Core Model is a central tool in our investment process and a key focus of our innovation efforts. The insights from our Core Model complement the deep fundamental analysis that we perform on candidate stocks for our portfolios. Over the past three years, we have made enhancements to our Core Model. These enhancements include implementing more sophisticated machine learning techniques, creating better ways to visualize and interact with our models, and adding a powerful alpha signal.

How We Have Enhanced Our Machine Learning Models

As discussed in our paper, The TD Epoch Core Model: An Evolving Tool, we believe that more complex and dynamic versions of the TD Epoch Core Model can complement the forecasts from our Standard, static model. We tested several modeling techniques and implemented three approaches in June 2021: Elastic Net Regressions, Random Forest Regressions, and Gradient Boosted Regression Trees.

Rather than select a single model, we have chosen to continue to use our Standard model as the primary model and corroborate its predictions with those from our machine learning models. Specifically, we look for agreement or disagreement between the models. We find that when our models agree, our predictions are more reliable.

Dynamic factor weights capture changing market regimes

All three machine learning versions of our Core Model have dynamically changing weights because they are re-estimated monthly, using only the most recent sixty months of data. This allows them to be more responsive to changes in the market environment. In contrast, our Standard Core Model has intuitive, but static, factor weights since it is re-estimated once every few years.

Tree-based algorithms capture non-linear relationships

Two of the three machine learning models – Random Forest Regressions and Gradient Boosted Regression Trees – are decision tree-based algorithms. Unlike our Standard and Elastic Net Models which assume a linear relationship between factors and stock returns, these tree-based algorithms have non-linear structures and can exploit complex relationships between factors and stock returns.

Partial constraints produce intuitive and robust models

Earlier this year, we implemented partial monotonicity constraints on our tree-based models to make them more intuitive and robust. Monotonicity constraints ensure that a set of factors are used correctly in a model, e.g., that high free cash flow yield predicts higher stock returns. We found that imposing monotonicity constraints on all factors in our Core Model was too restrictive – it did not allow our tree-based models to fully exploit dynamic changes in factor performance and correlations. In contrast, imposing monotonicity on the subset of model factors which are most representative of our free cash flow philosophy provided us with an attractive balance of predictive power and economic intuition. They prevent our models from being fooled by statistical artifacts in the data or by short-term trends. This is an important and innovative feature of our machine learning models.

Hyperparameter tuning helps us to find the optimal model

An important step in machine learning modeling is to find the combination of settings (i.e., hyperparameters) that produces an optimal and sensible model. These hyperparameters control the learning process of an algorithm. However, they cannot be determined directly from data, can be computationally expensive to find, and can easily lead to overfitted models which perform poorly in real life.

Version 1: Our custom grid search algorithm. For our first-generation machine learning models, we created a custom grid search algorithm. Conventional grid search approaches allow users to fine-tune model parameters by optimizing statistical measures such as accuracy and recall for classification tasks, or mean squared error (MSE) or R-squared for regression tasks.[1]

Our approach to hyperparameter tuning has two distinctive features. First, we aim to optimize information coefficient, which we believe is a far more relevant metric for improving the accuracy and predictive power of an equity stock selection model. Second, we tune our hyperparameters on monthly cross-sectional data in keeping with our Fama-MacBeth approach to model building. This is a departure from the conventional approach which pools data cross-sectionally and across time. We believe our approach leads to forecasts which are less susceptible to outlier events.

Version 2: Bayesian optimization, a faster and more flexible approach. Our second generation of machine learning models, implemented in March 2023, use a different tuning approach.[2] While our custom grid search approach produced good results, the tuning process was computationally expensive and time-consuming. After extensive research, we decided to use a Bayesian Optimization approach as implemented in the Optuna library.

Unlike grid search which tries to find the optimal model setting across all possible options, Bayesian Optimization uses past evaluation results in an informed manner and thus can find optimal model settings in far fewer iterations (and less time!) We are also able to use Bayesian Optimization for a wider range of machine learning algorithms. This is an important consideration as we continue to investigate other modeling approaches.

Caveat: More frequent hyperparameter tuning is not necessarily better. Theoretically, tuning hyperparameters more frequently (e.g., on a monthly or annual basis) may be beneficial if there are frequent shifts in market regimes. In reality, we found little improvement in the performance of our models from tuning at this level of frequency. Tuning too often may also lead to over-fitting and excessive turnover generated by model changes. As such, we have chosen to take a more pragmatic approach, monitoring model performance and updating our model hyperparameters only when necessary.

[1] An example is the GridSearchCV implementation in scikit-learn, a popular machine learning library.

[2] Our second-generation machine models use the same algorithms as our first generation but contain three enhancements to our tree-based models, Random Forest and Gradient Boosted Trees. First, we updated the hyperparameters for both models using the Bayesian Optimization approach discussed in this section. Second, we applied partial monotonicity constraints in the manner described above. Third, for the Gradient Boosted model, we switched from scikit-learn's implementation to the LightGBM algorithm to take advantage of faster computation with similar performance.

How We Better Understand Our Machine Learning Models

The machine learning versions of our Core Model are obviously more complex than our Standard model, but they do not have to be black boxes. We use a variety of techniques to interpret their predictions.

Feature importance scores tell us the relative importance of individual features. A feature's importance score is calculated by measuring the change in a model's accuracy when the feature is removed or varied. In other words, it tells us how much individual features influence a model's predictions. For a given model, feature importance scores are normalized to sum up to 100% so that individual feature scores can be easily compare to factor weights in a linear model.

While feature importance scores are simple to understand, they have a limitation in that they do not tell us how a feature is used. For example, a 5% feature importance score suggests the feature contributes 5% to a model's prediction but we do not know whether the feature is used as a positive or negative contributor to the prediction.

SHapley Additive exPlanations (SHAP) values shine additional light on complex models. Introduced in 1951 as a method in cooperative game theory, SHapley Additive exPlanations (or SHAP) values are used today to reverse-engineer the outputs of complex predictive models. Similar to feature importance scores, SHAP values measure the marginal contribution of each feature to the final prediction of a model. We use SHAP values to create powerful images such beeswarm plots to visualize for each feature in a model, the magnitude of its contribution, its distribution of values, and most importantly, the direction of the relationship between the feature and the prediction.

These insights provide much needed transparency into the behavior of our models. For example, our research on monotonicity constraints was motivated in part by our observation that the SHAP values in our beeswarm plots showed patterns what were not consistent with the way we want our tree-based models to use key model features. We also used these types of plots to verify that we successfully implemented partial monotonicity constraints.

Our proprietary platform allows us to interact with our Core Model versions

Our proprietary platform is a web-based portal which aims to deliver the right information at the right time to portfolio managers and research analysts at TD Epoch. Over the past two years, we have developed eight customized dashboards including one dedicated to the Epoch Core Model. This dashboard provides convenient access to scores and rankings produced by the Standard and machine learning versions of our Core Model. We can use the dashboard to screen for stocks within our investable universe, audit feature calculations, and perform scenario analysis on final model outputs.

Where We Have Looked For New Sources of Alpha

We are always searching for new sources of alpha to boost our Core Model. In this section, we discuss two recent research efforts which focus on market-based signals.

Short Interest-Based Signals. In June 2021, we added a new market-based signal to our Core Model. The signal, ***Active Adjusted Utilization***, compares the demand for selling short a security (net of any hedging activities) to the stock of available, lendable supply for that security. We compute this measure using information from a proprietary database of activity in the securities lending market. We found that our ***Active Adjusted Utilization*** measure is a powerful predictor of future stock returns and complementary to the sell-side analyst-based measures in the Investor Behavior component of our Epoch Core Model.

Option-Based Signals. In 2022, we investigated another group of market-based signals: measures based on options on stocks. We hypothesized that informed investors may choose to express views in the securities lending and/or options markets, depending on which market allows them to implement their trades cheaply and efficiently. If this is the case, we may be able to extract information based on their activities to forecast future stock returns.

After surveying thirty years of research on options-based alpha signals, we identified twelve potential measures for further testing. While two of the measures demonstrated some predictive power over the length of our investment horizon (i.e., at least twelve months), their results were not sufficiently compelling to recommend future research.

Current and Future Research

We have an ambitious research agenda for the next two years. We continue to explore enhancements to our machine learning models, including intelligent ways to combine model forecasts. We are also actively investigating text-based data using natural language processing techniques. These initiatives are part of our drive to continuously innovate and improve our investment tools and processes.

Research Agenda

2017

- Introduced several new and enhanced factors
- Used Ridge and LASSO regression estimation techniques
- Globalized peer groups
- Implemented accounting adjustments for R&D, operating leases, under-funded pension liabilities
- Introduced Banks version

2018

- Introduced interactive tool for Core Model drivers and history
- Introduced customizable Core Model screening tool for idea generation
- Used higher frequency reporting data
- Implemented stubbed methodology for calculating debt issuance in the case of bad data

2019

- Introduced Insurance version
- Introduced REIT version
- Implemented Standard Model for insurance brokers and real estate management firm
- Implemented pro forma adjustments to model universe for index constituent changes

2020

- Switched to Axioma equity beta calculations from Barra
- Implemented changes in operating lease calculation given IFRS 16 adoption
- Conducted research on potential short interest-based alpha signals

2021

- Implemented dynamic, machine learning-driven versions of the Core Model
- Incorporated new short interest factor
- Updated model weights to incorporate more recent data
- Launched our proprietary platform
- Implemented Core Model tools in the proprietary platform to improve accessibility and ease of use

2022

- Conducted research on potential option-based alpha signals
- Implemented tools in the proprietary platform to make the machine learning versions of the Core Model more interpretable
- Conducted research on the impact of monotonicity constraints on tree-based versions of our Core Model

2023

- Implemented partial monotonicity constraints on tree-based versions of our Core Model
- Evaluate machine learning versions of our Banks, Insurance, and REIT models
- Conduct research on news-based analytics for potential alpha signals
- Investigate techniques for ensembling forecasts from different versions of the Core Model

2024

- Investigate free cash flow forecasting approaches
- Conduct research on other potential alpha factors, e.g., management quality
- Add more tools to our proprietary platform to explain and visualize our models

Learn more about TD Global Investment Solutions

www.tdgis.com