Temporal Fusion Transformer Models for Predicting Stock Behavior

By Andrew Lys*

Your abstract here.

In portfolio design, a very common, and profitable strategy is called quantitative value investing. This strategy consists of designing some objective measure, ranking stocks based on this measure, and investing into the top ranked companies. The measure is derived from expert knowledge about how markets function, data, or some combination of the two. Portfolio design, in general, has been a field slow to adopt machine learning techniques, generally forgoing the data driven aspect of constructing these measures in exchange for more traditional financial measures. Examples of these measures include Trailing Twelve Month Returns (TTM Returns) and Earnings Before Interest and Taxes/Enterprise Value (EBIT/EV). These measures are often called value factors, as in the Fama-French value factor model. A clear drawback of this method is that there are countless different value factors an analyst could choose. For instance, what if it's the case that ranking stocks based on Trailing Eighteen Month Returns, instead of TTMs, is the secret to beating the market? Of course this example is silly, but in this manner, in their use as feature engineers, Neural Networks serve as an obvious next step to improve this strategy.

In traditional computer vision learning techniques, like boosted decision trees for facial recognition (Yoav Freund and Robert E. Schapire 1995), the main hurdle was always the development of better methods to extract relevant features. With the advent of convolutional neural networks, able to extract relevant features automatically, computer vision was effectively revolutionized (Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton 2017). Why can't Neural Networks do the same for factor engineering? Humans are bias-prone. When we pick factors, we are unintentionally choosing factors that are not statistically clean. Analysts have subconscious reasons for selecting factors, no matter how hard they can try to avoid these biases. For instance, if an analyst truly believes in a company, they might unintentionally choose factors that bias them towards that company, but will ultimately lead to lower returns overall. For reasons like this, Neural Networks have the capacity to revolutionize forecasting in the same way as they did image recognition.

Financial fundamental data is in the form of time series data, and thus admits a very natural sequence structure that can be leveraged by Recurrent Neural Networks (RNNs). (John Alberg and Zachary C. Lipton 2018) investigated the

* Lys: University of Chicago, andrewlys@uchicago.edu

use of Long Short Term Memory (LSTM) RNNS in forecasting fundamental data. These are RNN architectures where there are two streams of data passing through the recurrent neural network, the long term and short term streams. In this paper, we will investigate the use of Temporal Fusion Transformers (TFTs), as proposed by (Bryan Lim, Sercan O. Arik, Nicolas Loeff and Tomas Pfister 2020). We use the implementation provided by PyTorch Forecasting. TFTs are a novel architecture developed by Google in collaboration with the University of Oxford. They leverage the recent success of Transformers in natural language processing, and the natural sequence structure of time series data. (Lim et al. 2020) found promising results in forecasting, so we will investigate their ability to predict stock behavior.

So far, we have been careful to avoid forecasting. This is because forecasting is very difficult. (Alberg and Lipton 2018) found that there is far too much noise to be able to accurately generate forecasts that meaningfully outperform linear models. For this reason, we focus on a far more tractable problem. We will attempt to predict certain qualitative features of stock behavior. The main feature we will focus on is whether a stock is a value trap.

In the quantitative value investing strategy, we seek to select stocks that are, in some sense, undervalued. We approximate this true value of a stock by variuous value factors. The value factors we select maybe misleading, causing us to misunderstand the true value of a stock. In this way, we may select stocks that are not actually undervalued, but are instead value traps. A stock may perform very well on our value factors, but there may be some aspect we are missing in our analysis that will cause the stock to perform poorly in the future. For instance, a stock may have a high EBIT/EV ratio, but if the company is about to go bankrupt, this ratio will not be very useful.

This problem is significantly more tractable than forecasting, and is in fact, very valuable. (Alberg and Lipton 2018) have shown that identifying value traps can result in a 3.5% increase in annualized returns (See Figure 1).

In this paper, we will investigate the ability of TFTs to predict whether a stock is a value trap. The structure of the paper is as follows: in section 2, we will discuss the data and the pre-processing; in section 3, we will discuss the model framework; In section 4, we will present the experimental setup and the results; and in section 5, we will conclude.

I. Data and Data Pre-Processing

A. Raw Data

The data we use is from the Wharton Research Data Services (WRDS) database. We use the Compustat North America database, which contains fundamental data on companies. As undergraduate students, we are able to access this data through the University of Chicago's library, although, without a PI, we are only able to get a day pass.

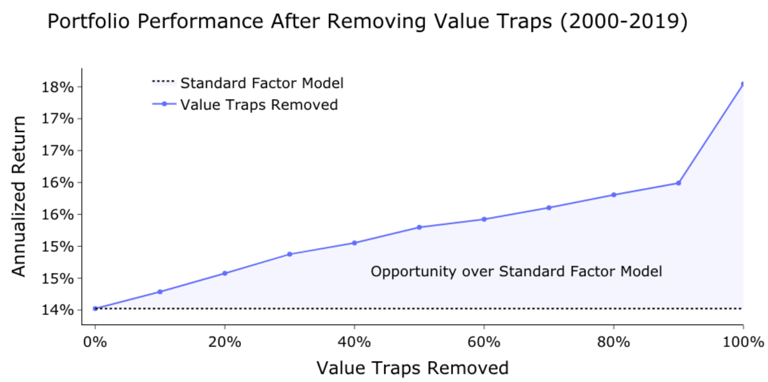

FIGURE 1. PORTFOLIO PERFORMANCE AS AN INCREASING NUMBER OF VALUE TRAPS ARE REMOVED.

We focused on the time frame 1970-2025, since before 1970, fundamentals reporting was not as standardized. Additionally, we only considered companies that are located in the United States. We filtered for companies only traded on either the NYSE, NASDAQ, or AMEX. We also filtered out financial services companies. The motivation for this is that the fundamentals reporting of financial services companies represent very different things than for industrial companies. Although this reasoning may be disputed, this is considered a common practice. The data was taken in quarterly intervals, since this is how often companies report fundamental data. This represents a total of approximately 12,000 companies, and a total of 876,347 observations.

Following (Alberg and Lipton 2018), we selected the following source fields, as show in Table 1.

B. Pre-Processing

There is a lot of pre-preprocessing that needs to be done before we can use this data in a model. We cannot simply feed the raw data into the model. For each company, there are very unique patterns in the raw data that the model will memorize. As a result, the model will not be able to learn. We need to apply some processing to the data to make sure this does not happen. Additionally, there are some other financial features that we know we want to include. Since this data is fairly noisy, there are some NaNs that we need to deal with. We will develop a NaN policy for dealing with missing data. Finally, following the reccommendation of (Kasun Bandara, Christoph Bergmeir and Slawek Smyl 2018), we will apply a seasonalization algorithm to these features. We will then feed the original features in alongside the seasonalized features. Additionally, we will need to apply some sort of normalization to the data before the seasonalization. These two will be discussed in the next section. We will end up with five categories of features:

Income Statement & Other Items	**Balance Sheet**
Sales (Revenue)	Cash & Cash Equivalents
Cost of Goods Sold	Receivables
Sales, General, and Admin Expenses	Current Assets
Operating Income After Depreciation	Property Plant & Equipment
Net-Income	Other Assets (Incl. Goodwill)
Capital Expenditures	Total Assets
Dividends	Debt in Current Liabilities
Common Shares Outstanding	Accounts Payable
Price per Share	Current Liabilities
	Long-Term Debt
	Other Liabilities
	Total Liabilities
	Minority Interest
	Preferred Stock
	Shareholders' Equity

Table 1—Selected Source Fields

Momentum features, Valuation Features, Fundamental Features, Missing Value Features, and Seasonal Features.

Momentum Features

Since it's quite likely that recent stock price changes will be indicative of future stock price changes, we will include some momentum features, in order to make it easier for the Neural Network to find these patterns. We will include 3, 9, 12, and 18 month percent changes in stock price. Additionally, we will provide the percentile rankings amongst stocks traded during that quarter of these features. These features will hopefully allow the Neural Network to find these patterns easier, and make use of them.

Valuation Features

We will include the following valuation features: Book to Market and Earnings Yield. These are common value factors used in the Fama-French model. They are defined as follows:

- bkmktq: Book to Market.

$$\text{Book to Market } = \frac{\text{Shareholder Equity}}{\text{Closing Price } \cdot \text{ Common Shares Outstanding}}$$

- evq: Enterprise Value

$$\begin{aligned}\text{Enterprise Value} = {} & \text{Closing Price} \cdot \text{Common Shares Outstanding} \\ & + \text{Long Term Debt} + \text{Debt in Current Liabilities} \\ & - \text{Cash and Short Term Investments}\end{aligned}$$

- eyq: Earnings Yield:

$$\text{Earnings Yield} = \frac{\text{Operating Income After Depreciation}}{\text{Enterprise Value}}$$

C. Transformation and Deseasonalization

II. Model Framework

A. Temporal Fusion Transformer

B. Quantile to Ordinal Conversion

III. Experimental Results

IV. Conclusions

REFERENCES

Alberg, John, and Zachary C. Lipton. 2018. "Improving Factor-Based Quantitative Investing by Forecasting Company Fundamentals."

Bandara, Kasun, Christoph Bergmeir, and Slawek Smyl. 2018. "Forecasting Across Time Series Databases using Recurrent Neural Networks on Groups of Similar Series: A Clustering Approach."

Freund, Yoav, and Robert E. Schapire. 1995. "A desicion-theoretic generalization of on-line learning and an application to boosting." 23–37. Berlin, Heidelberg:Springer Berlin Heidelberg.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2017. "ImageNet classification with deep convolutional neural networks." *Commun. ACM*, 60(6): 84–90.

Lim, Bryan, Sercan O. Arik, Nicolas Loeff, and Tomas Pfister. 2020. "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting."

Mathematical Appendix