# FINANCIAL FORECASTING THROUGH MACHINE LEARNING AND TRADITIONAL ECONOMETRICS: A NOVEL HYBRID APPROACH

Georgiou Catherine[1], Fassas Athanasios[2] , Gkonis Vasileios[3] , Tsakalos Ioannis[4]
[1]*Dr*
*aikategeorgiou@uth.gr*

***Abstract:*** The ongoing interest on exposing the predictive components in returns, the necessity for absolute accuracy and reliability in forecasting along with the impressive advancement in computing power, have transferred the attention in the use of machine learning techniques for forecasting purposes. However, studies on the field are yet to include pure finance predictors when forecasting returns through AI. Our research tries to cover this gap by employing both the simple and the modified versions of the most well-known predictors. Our analysis goes a step further by integrating additional predictors, that is combining behavioral and stress indexes with the basic dividend-price ratio and earnings-price ratio, grounded in the long-run equilibrium relationships present in the examined variables. Our data set focuses on the S&P 500 index. Our key findings include the forecasting superiority of the modified ratios in all forecasting horizons, regardless of the methodological route followed.

***Keywords*:** *return predictability, machine learning, deep learning, big data*

## INTRODUCTION

The ongoing interest on exposing the predictive components in returns, the necessity for absolute accuracy and reliability in forecasting along with the impressive advancement in computing power and the availability of data, have transferred the attention in the use of machine learning techniques for forecasting purposes. A typical use of such approaches is as ensembles of several network models that tackle with modelling and sampling uncertainties with adverse impacts on accuracy and robustness in our forecasts. Over the last two decades, research efforts on such issues have intensified including multiple successful applications (see the discussion in Zhang et al., 1998). Artificial Intelligence (AI) approaches provide the flexibility of working on nonlinear data driven modelling whose forecasting properties are rather appealing. Indeed, it is proven that they are universal approximators, able to fit any underlying data generating process (Hornik, Stinchcombe and White, 1989; Hornik, 1991), and it is empirically supported that they can predict both linear and nonlinear time series (Zhang, 2001; Zhang et al., 2001). The emergence of several types and applications of AI techniques is in fact due to the aforementioned interesting properties.

However, studies on the field are yet to include pure finance predictors when forecasting returns through AI. Our research efforts try to cover this gap in the literature by employing both the simple and the modified versions of the most well-known predictors in traditional finance. Using a neutral network, we forecast returns and proceed on a straightforward comparison with the equivalent findings that are obtained through the traditional route as dictated in financial econometrics. To the best of our knowledge, while AI techniques have been used for forecasting in finance, they have not fully incorporated certain crucial financial variables as potential predictors. Our analysis goes a step further by integrating additional predictors. This integration involves combining behavioral and stress indexes with the basic dividend-price ratio and earnings-price ratio, grounded in the long-run equilibrium relationships present in the examined variables. Consequently, we propose novel predictors, developed from the ground up, which are anticipated to offer enhanced forecasting benefits.

Therefore, the primary focus of the study is to introduce new predictors that can potentially increase return predictability benefits. Modern literature is now attempting to econometrically alter conventional predictors to enhance their predictive capacity.

Our overall research plan can be divided in the following two strands; on the one hand, our methodology resembles the traditional finance empirical routes (see objectives 1-5 below) and on the other hand, we employ artificial intelligence (AI) techniques in an attempt to predict US market returns. That way we manage to build bridges between the more traditional financial econometric techniques already employed by both practitioners and academics, and the more modern computing tools that are gaining ground in finance.

Overall, we aim to accomplish the following objectives:

1. We propose proper modifications to the classical dividend-price ratio and earnings-price ratio, by cyclically adjusting dividend and earnings as a moving average of the last 10 years following the rationale of Campbell and Shiller (1988) in the construction of their Cyclically-Adjusted Price-Earnings (CAPE) which we believe addresses the statistical concerns and strengthens the predictive capacity of the simple ratios. Therefore, we construct the cyclically adjusted dividend-price (cadp) and the cyclically-adjusted earnings-price (caep) ratios, as well as the total return cadp and caep (trcadp and trcaep respectively).
2. We proceed on comparing the predictive performance of the cyclically adjusted ratios to their simple versions.
3. By identifying cointegration relationships within the basic ratios, we aim to econometrically modify the construction of both simple and cyclically adjusted ratios. Our hypothesis is that these modified ratios will exhibit superior forecastability and deliver enhanced forecasting quality compared to the basic ratios. This improvement is anticipated not only for return forecasts but also for the growth rates of dividends and earnings.
4. We test all included ratios' predictive performance both in-sample and out-of-sample (employing both a recursive and a full-sample approach as indicated by empirical literature- see the discussion in Georgiou et al., 2022; Georgiou, 2023).
5. Utilizing neural network techniques, we forecast returns, including dividend and earnings growth, by using both the conventional ratios and their cyclically adjusted and modified counterparts. This approach serves as an additional, more potent test of the reliability and robustness of the results traditionally obtained through financial econometric methods.

The "who is who" puzzle related to the identification of the most powerful return predictor has attracted huge conversations both in practitioners and academics for decades. It is simply not enough to retrieve a strong predictor that may explain time variations in returns in only one index; baseline is that we are looking for tools with strong predictive capacity and the assurance of reliability, accuracy and robustness in the results on a global scale.

Our work is strongly related to the daily challenges faced by financial analysts, portfolio and risk managers, investors but also fellow researchers in the field. We propose the use of predictors whose construction is simple and straightforward and thus, addresses several concerns (mainly by practitioners) regarding the practicality of advanced econometric tools. Also, our modified predictors manage to tackle certain econometric issues (such as the sample bias, the forward-looking bias, stationarity and so forth as mentioned above). The AI approach that we follow takes the analysis to the next level with enhanced predictive benefits but also even more reliable and robust results. The dominance of such techniques in the near future seems inevitable due to their flexibility, speed of adjustment, practicality and increased accuracy. The proposed analysis provides an alternative look on predictors' construction either through AI or the traditional econometrics, that improves the quality but also quantity of our forecasts.

## METHODOLOGY

Our overall analysis can be divided into two strands; first, we examine the existence of predictive components in returns, dividend and earnings growth rates through the traditional route (see the steps 1-5 below) and second, we implement an AI approach as an extra sensitivity check of our original findings. Specifically, the methodological steps of our study are as follows:

1. Confirm that we are dealing with AR(1) processes by conducting unit root testing.
2. Test for the existence of long-run equilibrium relationships based on the Johansen approach in a different set of vectors.
3. Create the altered versions of the traditional ratios.
4. Run in-sample regression models (both univariate and multivariate to better examine ratios' dynamics).
5. Perform out-of-sample testing (using both the recursive and the population techniques-see the discussion in Welch and Goyal, 2008; Campbell and Thompson, 2008; Lettau and Ludvigson, 2005).

6. AI and return forecasting. More specifically, machine learning provides a plethora of algorithms capable of identifying complicated patterns and making accurate conclusions. In this research, we utilize XGBoost and LSTM to precisely forecast returns. The use of metaheuristic algorithms for hyperparameter tuning is a novel aspect of our method. We assess a diverse range of metaheuristic algorithms, including the Artificial Gorilla Troop Optimizer (AGTO) (Abdollahzadeh, 2021), Electromagnetic Field Optimization (EFO) (Abedinpourshotorban et al.,2016) and Fox Optimizer (FOX) (Mohammed, 2023) in order to identify the most optimal approach. This work aims to provide a precise model for forecasting stock returns, while also making valuable contributions to the disciplines of machine learning and neural networks.

Our data refers to the S&P 500 index as available in CRSP. We employ returns with and without dividends, and obtain data from Welch's database. We also focus on Shiller's database and use his earnings, dividends and cpi (consumer price index) series. In order to estimate the cyclically-adjusted versions of the simple dp and ep, we originally re-estimate the dividends and earnings series as the moving average of the last 10 years of real dividends and real earnings respectively, by deflating via cpi at the end date of the sample, following the rationale of Campbell and Shiller in the construction of their cape.

Data on a monthly frequency is used to retrieve the cointegration relations as mentioned above, but then we transform our data set on an annual frequency to run the in and out-of-sample predictive regressions. We prefer using the annual frequency since it allows us to deal with the seasonality issues observed in dividends and earnings series as discussed in Lamont (1998) and Lettau and Ludvigson (2005). Our sample has the potential to range from 1926-2022. Table 1 below summarizes data's origin.

Table 1: Primary data sources

| Variables | Data source |
|---|---|
| S&P 500 data<br>(by Welch and Goyal) | https://www.ivo-welch.info/professional/goyal-welch/ |
| Data on dividends, earnings, cpi<br>(by Shiller) | http://www.econ.yale.edu/~shiller/data.htm |
| VIX | https://www.cboe.com/tradable_products/vix/vix_historical_data/ |
| ISI<br>(by Wurgler) | https://pages.stern.nyu.edu/~jwurgler/ |
| FSI | https://fred.stlouisfed.org/series/STLFSI4 |
| Money supply (M2-seasonally and non-seasonally adjusted) | https://fred.stlouisfed.org/series/M2SL |

We intend to include not only nominal ($r_t$), but also excess ($re_t$) and real returns ($rr_t$) based on similar approaches as in Welch and Goyal (2008), Chen (2009) and Cochrane (2011) among others. We also try to explain the predictive components in dividend and earnings growth (both in nominal and real values).

## FINDINGS

Our findings can be summarized as follows:

(a) There is siginificant evidence of long-run equilibrium relationships in a variety fo vectors that are also unique.
(b) The proposed modified ratios provide more powerful forecasting outcomes regardless the forecasting horizon (short, medium, large) both in and out-of-sample.
(c) We also provide significant evidence on dividend and earnings growth predictability.
(d) Findings on excess and real returns follow the same predictive pattern as nominal returns even thought the results are rather lower in value.
(e) The use of AI modelling provides respective predictability findings, indicating the forecasting superiority of the altered versions of the simple predictors.

## CONCLUSIONS

In this paper we propose econometric adjustments to simple ratios that have been traditionally employed for their predictive capacity. Our results indicate that by simply adjusting these ratios' construction and by creating new predictors from ground-up based on observed long-run equilibrium relationships, we obtain stronger forecasting outcomes. Our modified ratios increase both reliability and accuracy due to their construction. AI techniques employed reconfirm predictability findings as obtained through the traditional financial econometrics approach.

The paper's primary limitations are that the observed extra complexity in predicting dividend and earnings growth remains harder to justify at this stage. Also, we could add applicability and robustness to our findings if we could widen our econometric analysis on international markets. We leave both venues for future research endeavors.

Overall, the findings of this paper have the following implications for both practitioners and fellow researchers: (a) the introduction of new predictors whose construction is straightforward, without the involvement of any complex econometric techniques; (b) practitioners who employ our proposed predictors can increase their forecasting gains both in and oos and thus, boost their ability to manage portfolio risks or address market-timing and asset allocation strategic issues.

Finally, our work is strongly related to the daily challenges faced by financial analysts, portfolio and risk managers and investors. Our study aims to shed more light to the predictive components of time-varying expected returns and underline the increased predictive potentials of already identified predictors under certain econometric modifications.

## REFERENCES


Abdollahzadeh, B., Gharehchopogh, F. S., and Mirjalili, S. (2021). Artificial gorilla troops optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems. *International Journal of Intelligent Systems*, 36, 5887-5958.

Abedinpourshotorban, H., Shamsuddin, S.M., Beheshti, Z., and Jawawi, D.N. (2016). Electromagnetic field optimization: A physics-inspired metaheuristic optimization algorithm. *Swarm and Evolutionary Computation*, 26, 8-22.

Campbell J. Y., and Shiller, R. (1988). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *Review of Financial Studies*, 1 (3), 195-228.

Campbell, J.Y., and Thompson S.B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average?. *Review of Financial Studies*, 21, 1509-1531.

Chen, L. (2009). On the reversal of return and dividend growth predictability: A tale of two periods. *Journal of Financial Economics*, 92 (1), 128-151.

Cochrane, J.H. (2011). Presidential address: Discount rates. *The Journal of Finance*, 66 (4), 1047-1108.

Georgiou, C., Neokosmidis, I., and Polimenis, V. (2022). Modified ratios and the cyclically-adjusted price-earnings ratio. *Global Business and Economics Review*, 27 (2), 209-231.

Georgiou, C. (2023). Interpreting Return Variability via the Dividend-Price-Earnings Ratio. *International Journal of Computational Economics and Econometrics*, 13 (4), 423-445.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251–257.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.

Lamont, O. (1998). Earnings and Expected Returns. *The Journal of Finance*, 53 (5), 1563-1587.

Lettau, M., and Ludvigson S.C. (2005). Expected returns and expected dividend growth. *Journal of Financial Economics*, 76 (3), 583-626.

Mohammed, H., and Rashid, T. (2023). FOX: a FOX-inspired optimization algorithm. *Applied Intelligence*, 53, 1030–1050.

Stambaugh, R. (1999). Predictive regressions. *Journal of Financial Economics*, 54 (3), 375-421.

Welch, I., and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21 (4), 1455-1508.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14 (1), 35–62.

Zhang, G. P. (2001). An investigation of neural networks for linear time-series forecasting. *Computers and Operations Research*, 28(12), 1183–1202.

Zhang, G. P., Patuwo, B. E., and Hu, M. Y. (2001). A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers and Operations Research*, 28 (4), 381–396.