

**Leveraging Deep Learning for Asset Pricing:
A Comparative Study of Non-Linear and Linear Models**

An-Sing Chen

Ting-Hsuan Chien

Jui-Hsiang Lin

December 2025

Abstract

This study examines the predictive power of industry-adjusted employment change and consecutive earnings increases in forecasting U.S. stock returns from 1998 to 2024. Using traditional linear models like OLS, we find that while these characteristics show limited statistical significance in predicting stock returns, they still contribute to positive portfolio returns when incorporated into cross-sectional momentum strategies. In contrast, deep learning models, particularly Neural Networks with two and four hidden layers (NN-2 and NN-4), capture non-linear patterns in these same variables and significantly outperform OLS benchmarks, both in terms of raw returns and risk-adjusted metrics like the Sharpe ratio. This study highlights the limitations of linear models in capturing non-linear relationships in financial data and demonstrates the superior performance of deep learning techniques in asset pricing. The findings underscore the value of incorporating machine learning methods to enhance portfolio performance and uncover hidden market inefficiencies that traditional models may overlook.

Keywords: Asset Pricing, Momentum, Deep Learning, Neural Networks, Financial Markets

JEL classification: G11, G12, C45, C63, D53

1. Introduction

The U.S. stock market plays a central role in empirical asset pricing, with traditional models successfully identifying cross-sectional anomalies such as momentum and value (Kelly & Pruitt, 2013). However, these models often struggle to capture the complex, non-linear relationships inherent in market dynamics. Recent advancements in Machine Learning (ML) and Deep Learning (DL) have provided powerful tools to address these limitations. Techniques like Principal Component Analysis (PCA) and cross-sectional momentum have been widely used for predicting stock returns (Kelly et al., 2019; Yan & Yu, 2023). Moreover, ML and DL models, such as random forests and neural networks, offer enhanced predictive capabilities by modeling intricate non-linearities and interactions that traditional linear models, like OLS and VWLS, may overlook (Li et al., 2025; DeMiguel et al., 2023). These methods present new opportunities for improving return predictability and uncovering hidden patterns in financial data.

Traditional asset pricing frameworks, such as the Fama-French models, employ cross-sectional OLS regressions or Value-Weighted Least Squares (VWLS) to construct portfolios and identify market anomalies (Fama & MacBeth, 1973). These models have been instrumental in detecting systematic risks and pricing anomalies, but their inherently linear nature may constrain their ability to capture more complex, non-linear relationships in the market. For instance, momentum strategies, as discussed by Moskowitz, Ooi, and Pedersen (2012), and Pitkäjärvi et al. (2020), reveal market trends that are not fully explained by traditional linear models. The reliance on OLS-based Long-Short strategies as the benchmark for factor validity further underscores the limitations of these models in fully exploiting intricate market dynamics. Thus, while linear models serve as useful benchmarks, their "linearity" may impede a deeper understanding of more sophisticated market behaviors.

This study follows Green, Hand, and Zhang (2017) by utilizing two key predictive characteristics: *chempia* (industry-adjusted change in employees) and *consec_increase* (number of consecutive earnings increases), both of which GHZ (2017) identified as independent predictors of US stock returns since 2003. These variables are particularly relevant as they have been shown to provide significant information about stock returns beyond traditional factors. Additionally, following the methodology of Hou, Xue, and Zhang (2015), we exclude micro-cap

companies, which constitute a small portion of the market and could distort the results. The data is further analyzed using Fama's five-factor asset pricing model (Fama & French, 2015) to capture systematic risk and asset pricing anomalies. This rigorous data treatment ensures the robustness and tradability of the findings.

Recent advancements in deep learning highlight its capacity to address limitations of traditional linear models, particularly in cases where linear regression fails to capture complex patterns. The Universal Approximation Theorem demonstrates that neural networks (NNs) can approximate any continuous function, overcoming the constraints of linear models (Hornik, Stinchcombe, & White, 1989; Cybenko, 1989). Specifically, neural networks with multiple hidden layers, such as Neural Network with 2 Hidden Layers (NN-2) and Neural Network with 4 Hidden Layers (NN-4), can model non-linear relationships that traditional methods cannot handle (Lotfi & Akbarzadeh-T, 2014). These capabilities enable neural networks to solve problems that linear regression struggles with, offering a more flexible and powerful approach for modeling complex data and uncovering underlying patterns.

Consistent with traditional empirical practice, we evaluate the predictive power of the characteristics through cross-sectional return tests grounded in linear asset-pricing frameworks widely used in the literature (Fama and MacBeth, 1973, Asness et al., 2013, Freyberger et al., 2020). The Full-Sample OLS model (1998–2024) confirms the expected patterns, with the long–short portfolio producing the strongest performance, consistent with established cross-sectional evidence. However, when OLS is estimated only on the in-sample period (1998–2010) and applied to the out-of-sample period (2011–2024), its predictive ability deteriorates significantly. The long-only strategy remains statistically significant—generating a monthly mean return of 0.0131 and a CAGR of 0.1478—while unexpectedly outperforming the long–short strategy. The short-only portfolio fails to produce negative returns, causing the long–short spread to collapse. These results highlight the fragility and overfitting tendencies of linear models, establishing OLS as a weak benchmark for evaluating more flexible predictive approaches.

Under the same out-of-sample period (2011–2024) and using identical firm-level characteristics, the Deep Learning models in this study markedly outperform the OLS benchmark.

The NN-2 Hidden Layer model achieves a CAGR of 0.23, while the NN-4 model records 0.21, both substantially exceeding the long-run US equity return reported by Chambers et al. (2024). In terms of risk-adjusted performance, the Long-Only Sharpe Ratios of NN-2 (0.3337) and NN-4 (0.3147) also surpass that of the OLS Test Set (0.2415), confirming superior predictive ability. Consistent with Gu, Kelly, and Xiu (2020) and Leippold, Wang, and Zhou (2022), the slightly weaker NN-4 results suggest that deeper architectures are more prone to overfitting, reducing generalization despite greater functional flexibility. Overall, these findings highlight the clear advantage of Neural Networks in modeling nonlinear return structures and underscore their effectiveness relative to traditional linear approaches.

The remainder of the paper is structured as follows. Section 2 develops the data and methodology underlying the OLS and neural network frameworks. Section 3 presents the empirical results and discussion. Section 4 provides concluding remarks.

2. Data and Methodology

2.1 Data Description and Sample Construction

This study constructs a monthly panel dataset of U.S. stocks from Compustat North America over the period 1998–2024, covering firms with available price, return, and accounting information. To avoid distortions caused by extremely small firms, we follow Hou, Xue, and Zhang (2015) and exclude microcap stocks, defined as firms with market capitalization below the NYSE 20th percentile in each month. Although microcaps represent a large portion of listed firms numerically, they account for only about 3% of total U.S. market value, and including them would overweight small, illiquid firms in cross-sectional analyses. After standard data cleaning procedures, the sample includes an unbalanced panel comprising 953,876 firm-months across 3,000–4,000 distinct firms per year, as summarized in Table 1, which reports the annual distribution of firm-month observations, the number of distinct firms, the average monthly return, and the median log market capitalization.

[Insert Table. 1. Approximately Here]

2.2 Variable Construction

Building on the empirical findings of Green, Hand, and Zhang (2017), who document that only two firm characteristics remain reliably independent determinants of expected returns in the post-2003 period, we construct these two predictors using the definitions established in prior accounting research. The first variable, **chempia**, captures a firm's industry-adjusted change in annual employment and follows the construction in Soliman (2008). For each firm i in year t , we compute its employment change ($emp_{i,t} - emp_{i,t-1}$) and subtract the average employment change among firms in the same two-digit SIC industry:

$$chempia_{i,t} = (emp_{i,t} - emp_{i,t-1}) - \frac{1}{|J(sic2_i, t)|} \sum_{j \in J(sic2_i, t)} (emp_{j,t} - emp_{j,t-1}). \quad (1)$$

The second variable, **consec_increase**, measures the number of consecutive quarters in which earnings exceed those from the same quarter in the previous year, following Barth, Elliott, and Finn (1999). We first define an indicator for earnings improvement relative to the same quarter four periods prior, and then accumulate the length of consecutive improvements, resetting the count when the streak breaks and top-coding the maximum at eight quarters:

$$ibq_increase_{i,t} = \begin{cases} 1, & \text{if } ibq_{i,t} > ibq_{i,t-4}, \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

$$consec_increase_{i,t} = \min \left(\sum_{k=0}^{\infty} \prod_{j=0}^k ibq_increase_{i,t-j}, 8 \right). \quad (3)$$

Together, these two characteristics—industry-adjusted employment growth and sustained earnings improvement—represent the post-2003 independent determinants emphasized by Green, Hand, and Zhang (2017). Summary statistics for all variables are presented in Table 2.

[Insert Table. 2. Approximately Here]

The time-series patterns of the final sample highlight important changes in firm coverage and market characteristics over the 1998–2024 period. As shown in Figure 1, the number of firm-month observations and the number of distinct firms exhibit substantial variation over time, reflecting both entry and exit dynamics as well as changes in the availability of accounting and price data. Firm-month counts fluctuate between roughly 24,000 and 39,000 per year, while the number of distinct firms ranges from about 2,800 to 4,000, with a gradual decline during the mid-2000s followed by a modest recovery in the 2010s. Figure 2 further illustrates the evolution of market conditions by plotting the annual mean return and the median log market capitalization. Mean returns display considerable year-to-year volatility, consistent with known business-cycle patterns, whereas median firm size shows a steady upward trend, reflecting long-run growth in market capitalization and shifts in the composition of publicly traded firms. Together, these trends provide context for the empirical tests by documenting how sample breadth, firm composition, and key financial characteristics evolve over time.

[Insert Figure. 1. Approximately Here]

[Insert Figure. 2. Approximately Here]

2.3 Model Specification

2.3.1 Baseline OLS Model

We establish the baseline relationship between firm characteristics and subsequent stock returns, we estimate a cross-sectional Ordinary Least Squares (OLS) model commonly used in the empirical asset-pricing literature, including Fama and French (1992) and subsequent characteristic-based studies. The model relates next month’s excess return to the two post-2003 independent determinants identified by Green, Hand, and Zhang (2017)—industry-adjusted employment growth and consecutive earnings increases. Specifically, for each firm i in month t , we estimate

$$y_{i,t+1} = \beta_{0,i} + \beta_{1,i} \text{chempia}_{i,t} + \beta_{2,i} \text{consec_increase}_{i,t} + \varepsilon_{i,t+1} , \quad (4)$$

where $y_{i,t+1}$ denotes the firm's next-month excess return, $chempia_{i,t}$ captures the industry-adjusted change in employment following Soliman (2008), and $consec_increase_{i,t}$ measures the number of consecutive quarters of earnings improvement based on Barth, Elliott, and Finn (1999). This specification provides a parsimonious benchmark for assessing whether these two characteristics have incremental predictive power for future returns before incorporating nonlinear structures or more flexible learning models in subsequent sections.

2.3.2 Deep Learning Model

To capture potential nonlinear relationships between firm characteristics and future stock returns, we employ feed-forward neural networks that map the two firm-level predictors, $chempia$ and $consec_increase$, into a nonlinear forecasting function of next-period returns. Formally, the prediction model is written as

$$pred_{ret_{i,t+1}} = f_{NN}(chempia_{i,t}, consec_{increase_{i,t}}), \quad (5)$$

where $f_{NN}(\cdot)$ denotes a multilayer perceptron (MLP). This modeling approach allows the network to learn flexible nonlinear transformations of the input variables, enabling it to capture complex patterns in the cross-section of expected returns that may not be well-approximated by linear functional forms.

The hidden layers employ the Rectified Linear Unit (ReLU) activation function, following Nair and Hinton (2010), who show that ReLU alleviates gradient saturation and improves the efficiency of feature extraction relative to sigmoid-based activations. Although more recent alternatives such as Gated Linear Units (GLU) (Dauphin et al., 2017) and Gaussian Error Linear Units (GELU) (Devlin et al., 2019) have demonstrated improved gradient propagation in natural language and convolutional architectures, ReLU remains the preferred choice for fully connected networks applied to tabular financial data due to its simplicity, computational stability, and empirical robustness. The activation functions are defined as

$$a = \max(0, z) \text{ and } \hat{y} = z, \quad (6)$$

where the first applies to all hidden layers and the second preserves the continuous nature of the output in stock return forecasting.

We evaluate two network architectures that differ in depth but share the same input structure. Both models take two input features, *chempia* and *consec_increase*, and propagate them through fully connected dense layers. The first specification, denoted NN-2, contains two hidden layers with 64 and 32 units, respectively. The second specification, NN-4, increases network depth by incorporating four hidden layers with 64, 64, 64, and 32 units. All hidden layers are fully connected, meaning each neuron receives inputs from every neuron in the previous layer. The architecture details, including output dimensionality and parameter counts for each layer, are summarized in Table 3, while Figure 3 illustrates the corresponding network structures. These configurations allow us to compare whether greater model capacity, achieved through additional layers and parameters, enhances return predictability.

[Insert Figure. 3. Approximately Here]

Both neural network models are trained using the Adam optimizer with a learning rate of 0.001, a configuration commonly adopted in forecasting applications due to its adaptive gradient updates and numerical stability. The loss function is mean squared error (MSE), reflecting the continuous nature of return predictions and ensuring that the models minimize squared deviations between predicted and realized returns. Model performance during training is monitored using the loss and additional compile metrics reported in Table 3. This training procedure provides a consistent basis for comparing the predictive performance of the neural networks with the linear OLS benchmark.

[Insert Table. 3. Approximately Here]

2.4 Portfolio Construction

To evaluate the economic content of the predicted returns, we construct cross-sectional strategies following standard portfolio-formation procedures widely used in the momentum and machine-learning asset-pricing literature (Jegadeesh and Titman, 1993, 2001; Han et al., 2024; Goulding, Harvey, and Mazzoleni, 2023). At the end of each month t , firms are sorted into

quantiles based on their predicted next-month return $\hat{y}_{i,t}$. We focus on three portfolio strategies that summarize the directional and cross-sectional dispersion in the prediction signal. The long-only strategy invests equally in firms in the highest decile of predicted returns,

$$r_{t+1}^L = \frac{1}{|\{i \in S_t: \hat{y}_{i,t} \geq Q_{0.9}(\hat{y}_t)\}|} \sum_{i \in S_t: \hat{y}_{i,t} \geq Q_{0.9}(\hat{y}_t)} r_{i,t+1}. \quad (7)$$

The short-only strategy invests equally in the lowest decile,

$$r_{t+1}^S = \frac{1}{|\{i \in S_t: \hat{y}_{i,t} \leq Q_{0.1}(\hat{y}_t)\}|} \sum_{i \in S_t: \hat{y}_{i,t} \leq Q_{0.1}(\hat{y}_t)} r_{i,t+1}. \quad (8)$$

The long-short strategy, which represents the factor return implied by the prediction signal, takes a long position in the top decile and a short position in the bottom decile:

$$r_{t+1}^{LS} = \frac{\sum_{i: \hat{y}_{i,t} \geq Q_{0.9}(\hat{y}_t)} r_{i,t+1}}{|\{i: \hat{y}_{i,t} \geq Q_{0.9}(\hat{y}_t)\}|} - \frac{\sum_{i: \hat{y}_{i,t} \leq Q_{0.1}(\hat{y}_t)} r_{i,t+1}}{|\{i: \hat{y}_{i,t} \leq Q_{0.1}(\hat{y}_t)\}|}. \quad (9)$$

These portfolios provide a model-agnostic, economically interpretable benchmark for assessing whether the prediction model generates meaningful cross-sectional variation in expected returns.

2.5 Performance Evaluation

We assess the performance of the portfolio strategies constructed from the predicted signals, we evaluate monthly returns using a set of standard metrics widely adopted in the asset-pricing and portfolio-management literature. Specifically, we report the mean return, volatility, Sharpe ratio, maximum drawdown (MDD), and compound annual growth rate (CAGR) for each strategy and model configuration. Mean return and volatility summarize the first- and second-moment properties of the return distribution, while the Sharpe ratio provides a normalized measure of risk-adjusted performance. Maximum drawdown captures the worst cumulative decline from a historical peak to a subsequent trough and is computed as

$$\text{MDD} = \min_t \left(\frac{P_t - \max_{\tau \leq t} P_\tau}{\max_{\tau \leq t} P_\tau} \right), \quad (10)$$

where P_t denotes the cumulative portfolio value at time t . The CAGR measures the long-run growth rate implied by the cumulative return series and is calculated as

$$\text{CAGR} = \left(\frac{P_T}{P_0} \right)^{1/N} - 1, \quad (11)$$

where N is the number of years in the sample. These metrics collectively provide a comprehensive evaluation of performance across levels, risk profiles, and long-horizon growth characteristics.

To ensure consistency across models, all performance statistics are computed from the same monthly rebalanced portfolios constructed using the formation procedure described in Section 2.4. For each strategy (Long Only, Short Only, and Long–Short), the portfolio return in month $t + 1$ is calculated as the simple average of the constituent stock returns. The resulting cumulative return series provides the basis for computing mean return, volatility, Sharpe ratio, maximum drawdown, and CAGR. This unified evaluation framework ensures that differences in performance are attributable to the forecasting models rather than to variations in portfolio formation or rebalancing mechanics.

3. Result and Discussion

3.1 Performance of Full-Sample OLS Model (1998–2024)

The full-sample OLS model, estimated over the period from 1998 to 2024, confirms the predictive validity of our key predictors—industry-adjusted change in employees (`chempia`) and consecutive earnings increases (`consec_increase`)—in forecasting stock returns. As shown in Figure 4, the Long-Short portfolio strategy exhibits the strongest performance, with cumulative returns growing exponentially over the sample period. The Long Top 10% strategy (green line) outperforms the Short Bottom 10% (red line), with the long-short strategy (blue line) delivering superior returns. These results align with Green, Hand, and Zhang (2017), who demonstrate the significance of these variables in cross-sectional return predictability.

[Insert Table. 4. Approximately Here]

[Insert Figure. 4. Approximately Here]

3.2 Test Set OLS Model (2011–2024)

When the OLS model, trained on data from 1998–2010, is applied to the test period (2011–2024), the performance of the Long-Short strategy significantly deteriorates, as illustrated in Figure 5. The model’s predictive power declines, with the short-only strategy failing to generate negative returns and resulting in a collapsed long-short spread. These results underscore the overfitting tendencies of traditional linear models, highlighting their limitations in adapting to changing market dynamics. Despite this, the Long-Only strategy remains significant, albeit with a reduced performance, achieving a monthly mean return of 0.0131 and a CAGR of 0.1478.

[Insert Figure. 5. Approximately Here]

3.3 Neural Network Models with Two and Four Hidden Layers (2011–2024)

In contrast to the OLS models, the deep learning approaches, particularly the Neural Network models with two hidden layers (NN-2) and four hidden layers (NN-4), exhibit superior performance during the out-of-sample test period (2011–2024). Figures 6 and 7 show that both NN-2 and NN-4 outperform the OLS benchmark, with the Long-Only strategy producing the highest cumulative returns. The NN-2 model achieves a compound annual growth rate (CAGR) of 0.2388, while NN-4 yields a slightly lower CAGR of 0.2192. These results are consistent with Gu, Kelly, and Xiu (2020), who highlight the ability of neural networks to model complex, non-linear relationships in stock returns that linear models cannot capture.

[Insert Figure. 6. Approximately Here]

[Insert Figure. 7. Approximately Here]

However, the slightly weaker performance of the NN-4 model suggests that deeper architectures, while offering greater flexibility, are more prone to overfitting. This is evident in the slight deterioration of the Long-Short strategy's predictive power for NN-4 compared to NN-2.

This phenomenon highlights the trade-off between model capacity and generalization, underscoring the importance of balancing depth with the ability to generalize across various market conditions.

3.4 Comparison of Risk-Adjusted Performance

In terms of risk-adjusted performance, the NN-2 model's Long-Only Sharpe Ratio of 0.3337 significantly surpasses that of the OLS Test Set (0.2415), confirming its superior ability to generate returns relative to risk. The NN-4 model follows closely with a Sharpe Ratio of 0.3147, further illustrating the strong risk-adjusted returns achieved by neural networks. These findings reinforce the efficacy of deep learning techniques in capturing complex patterns in financial data, outperforming traditional models in both raw and risk-adjusted performance metrics.

[Insert Table. 5. Approximately Here]

3.5 Nonlinear Prediction with Neural Networks

While the OLS results suggest that *chempia* and *consec_increase* lack a significant linear relationship with stock returns, the performance of the neural network models demonstrates that these features contain valuable predictive power when modeled in a non-linear framework. Neural networks, particularly those with multiple hidden layers, can uncover complex, non-linear relationships between these variables and future stock returns. As shown in Figures 6 and 7, both NN-2 and NN-4 models are able to better capture these non-linear patterns, thereby improving the predictive power and portfolio performance compared to the traditional OLS approach.

The NN-2 model, in particular, outperforms OLS in terms of cumulative returns, highlighting the ability of deep learning models to leverage non-linearities that OLS cannot. This finding emphasizes the importance of non-linear modeling approaches in finance, where traditional linear models often fail to capture the full complexity of market dynamics. Moreover, while NN-4 performs slightly worse than NN-2, it still surpasses OLS, underscoring the advantage of neural networks in financial forecasting.

3.6 Implications for Asset Pricing and Investment Strategies

The results from this study have significant implications for asset pricing and portfolio construction. Traditional OLS-based strategies have been widely used in the literature to predict stock returns and construct portfolios. However, the declining performance of the OLS model in out-of-sample tests, especially during periods of market structural changes, highlights its limitations. Conversely, the superior performance of deep learning models, particularly those with two hidden layers, suggests that neural networks are better suited for capturing the dynamic, non-linear relationships inherent in financial markets.

These findings support the growing trend of integrating machine learning and deep learning techniques into financial modeling. By leveraging the predictive power of deep learning, investors can enhance portfolio performance and uncover inefficiencies in the market that traditional models may miss. However, caution is necessary when using deeper models, as overfitting can reduce their ability to generalize, particularly in volatile market environments.

4. Conclusion

This study highlights the limitations of traditional linear models, such as OLS, in capturing the complex, non-linear relationships inherent in financial markets. While OLS-based strategies have been widely used for predicting stock returns, our results demonstrate that these models struggle to maintain predictive power in out-of-sample tests, particularly during periods of structural change. Despite this, the use of industry-adjusted employment change (*chempia*) and consecutive earnings increases (*consec_increase*), even when they do not show significant linear relationships, still leads to positive portfolio returns when integrated into cross-sectional momentum strategies.

The superior performance of Neural Network models (NN-2 and NN-4), which capture non-linear patterns in these same variables, underscores the growing importance of machine learning and deep learning techniques in asset pricing. The neural network models, particularly NN-2, significantly outperform the OLS benchmarks, both in terms of raw return and risk-adjusted metrics like the Sharpe ratio, demonstrating that deep learning models are better suited for capturing the dynamic, non-linear relationships that define financial market behavior.

Moreover, while the deeper architecture of NN-4 provides greater functional flexibility, it is also more prone to overfitting, which slightly reduces its generalization ability in comparison to NN-2. These results emphasize the trade-off between model capacity and generalization, particularly in financial markets that are volatile and subject to frequent structural changes.

Overall, this study contributes to the literature by demonstrating that even simple predictive characteristics, when modeled non-linearly, can provide valuable information for asset pricing and portfolio construction. The findings suggest that traditional linear models may not fully exploit the potential of these variables, and that incorporating machine learning techniques like neural networks can offer more robust and accurate predictions. This approach opens new avenues for future research in asset pricing, particularly in exploring other machine learning techniques and their applications in financial markets.

References

- Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The journal of finance*, 68(3), 929-985.
- Barth, M. E., Elliott, J. A., & Finn, M. W. (1999). Market rewards associated with patterns of increasing earnings. *Journal of accounting research*, 37(2), 387-413.
- Burnside, C., Eichenbaum, M., & Rebelo, S. (2011). Carry trade and momentum in currency markets. *Annu. Rev. Financ. Econ.*, 3(1), 511-535.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4), 303-314.
- Dauphin, Y. N., Fan, A., Auli, M., & Grangier, D. (2017, July). Language modeling with gated convolutional networks. In *International conference on machine learning* (pp. 933-941). PMLR.
- DeMiguel, V., Gil-Bazo, J., Nogales, F. J., & Santos, A. A. (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, 150(3), 103737.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3), 607-636.
- Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2), 427-465.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of financial economics*, 116(1), 1-22.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5), 2326-2377.

Goulding, C. L., Harvey, C. R., & Mazzoleni, M. G. (2023). Momentum turning points. *Journal of Financial Economics*, 149(3), 378-406.

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average US monthly stock returns. *The Review of Financial Studies*, 30(12), 4389-4436.

Han, Y., He, A., Rapach, D. E., & Zhou, G. (2024). Cross-sectional expected returns: new Fama–MacBeth regressions in the era of machine learning. *Review of Finance*, 28(6), 1807-1831.

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

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3), 650-705.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1), 65-91.

Jegadeesh, N., & Titman, S. (2011). Momentum. *Annu. Rev. Financ. Econ.*, 3(1), 493-509.

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, 68(5), 1721-1756.

Kelly, B. T., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3), 501-524.

Li, B., Rossi, A. G., Yan, X. S., & Zheng, L. (2025). Machine learning from a “Universe” of signals: The role of feature engineering. *Journal of Financial Economics*, 172, 104138.

Lotfi, E., & Akbarzadeh-T, M. R. (2014). A novel single neuron perceptron with universal approximation and XOR computation properties. *Computational intelligence and neuroscience*, 2014(1), 746376.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of financial economics*, 104(2), 228-250.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).

Pitkäjärvi, A., Suominen, M., & Vaittinen, L. (2020). Cross-asset signals and time series momentum. *Journal of Financial Economics*, 136(1), 63-85.

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The accounting review*, 83(3), 823-853.

Yan, J., & Yu, J. (2023). Cross-stock momentum and factor momentum. *Journal of Financial Economics*, 150(2), 103716.

Appendix

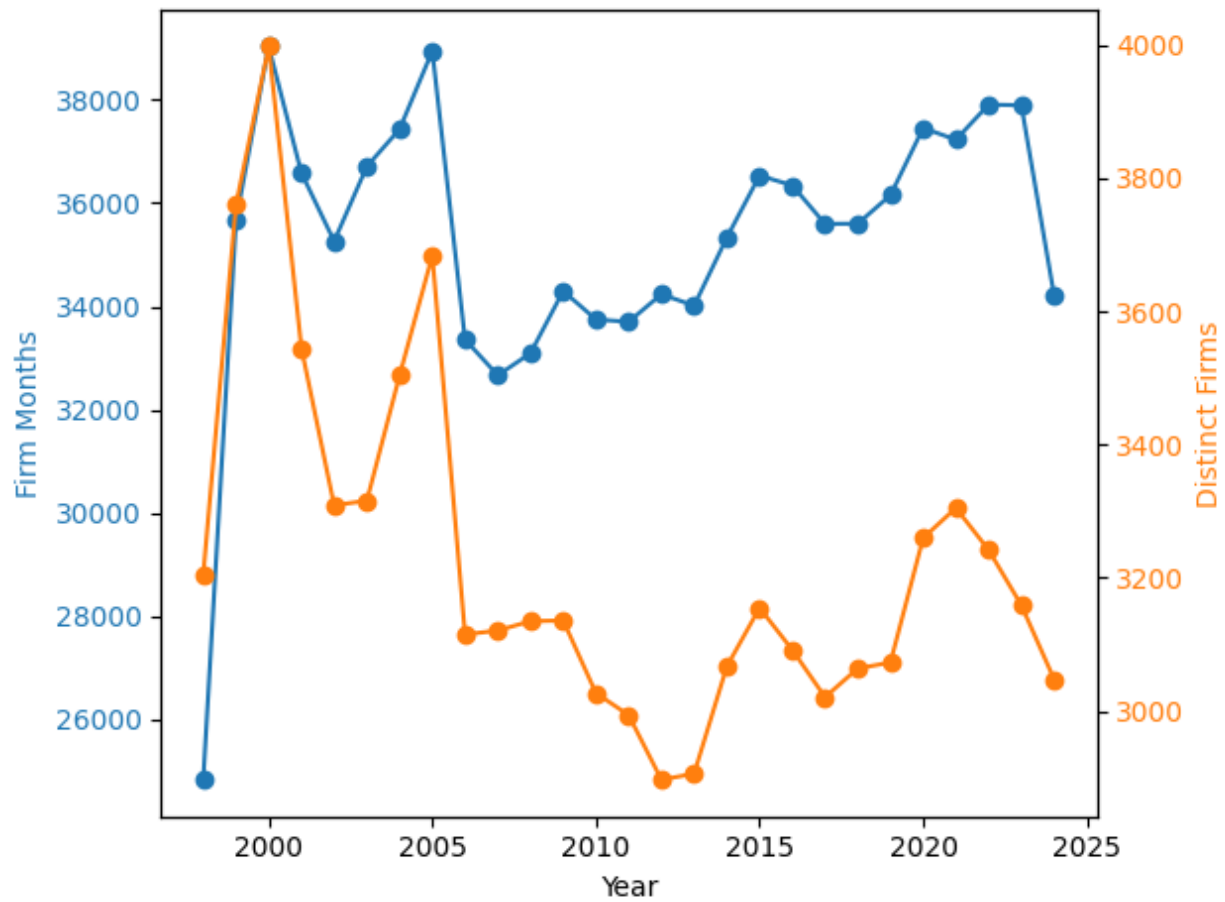


Figure. 1. Trends in Firm-Level Data (1998–2024)

This figure visualizes the annual trends of key firm-level variables from 1998 to 2024. The blue line represents Firm Months, showing the total number of months with valid data for each distinct firm in a given year. The orange line depicts Distinct Firms, indicating the number of firms with available data each year. This visualization offers insights into the evolution of firm participation and the amount of data available over time, emphasizing the trends in data coverage and firm activity.

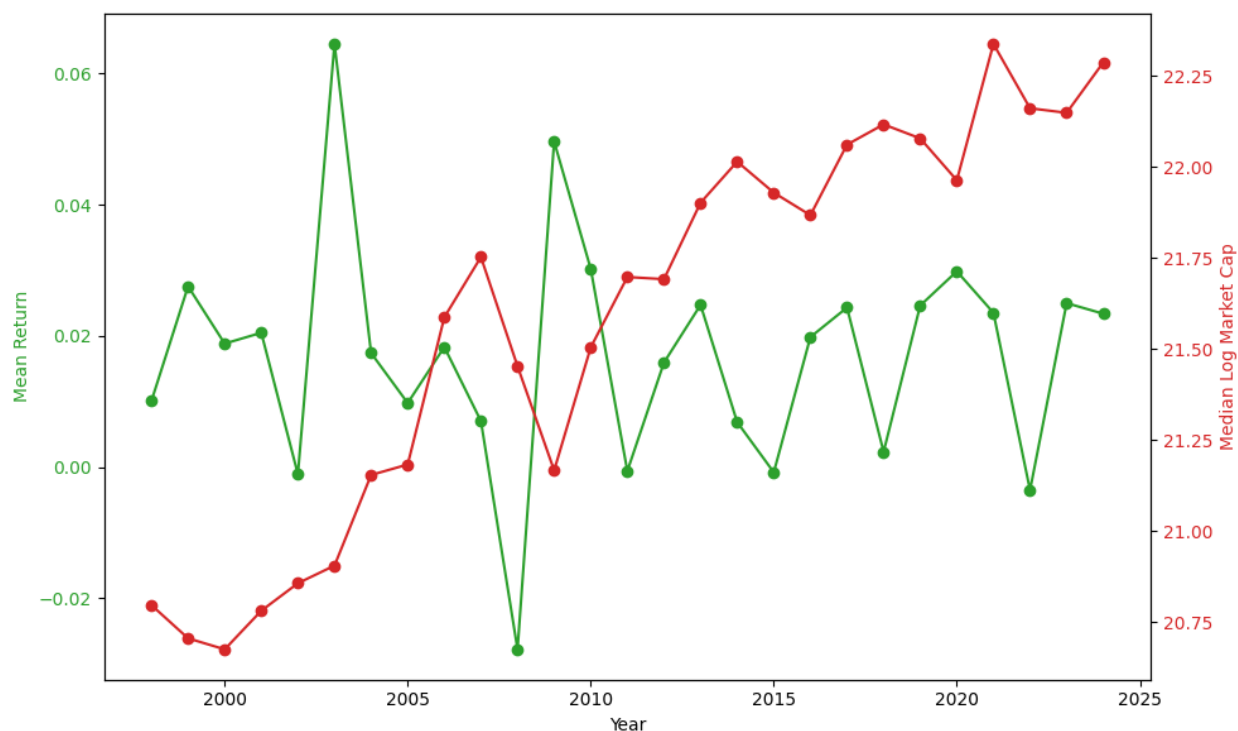


Figure. 2. Trends in Mean Return and Median Log Market Capitalization (1998–2024)

This figure visualizes the annual trends of Mean Return (green line) and Median Log Market Capitalization (red line) from 1998 to 2024. The Mean Return represents the average monthly return for firms in a given year, while the Median Log Market Cap reflects the median of the log-transformed market capitalization. The data visualization highlights the dynamic relationship between firm returns and market size over time.

Figure. 3A. Neural Network - 2 Hidden Layer

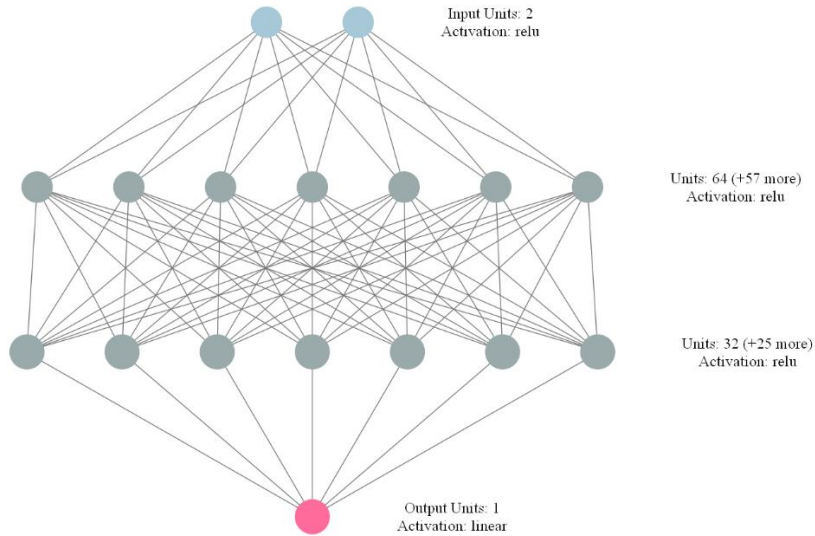


Figure. 3B. Neural Network - 4 Hidden Layer

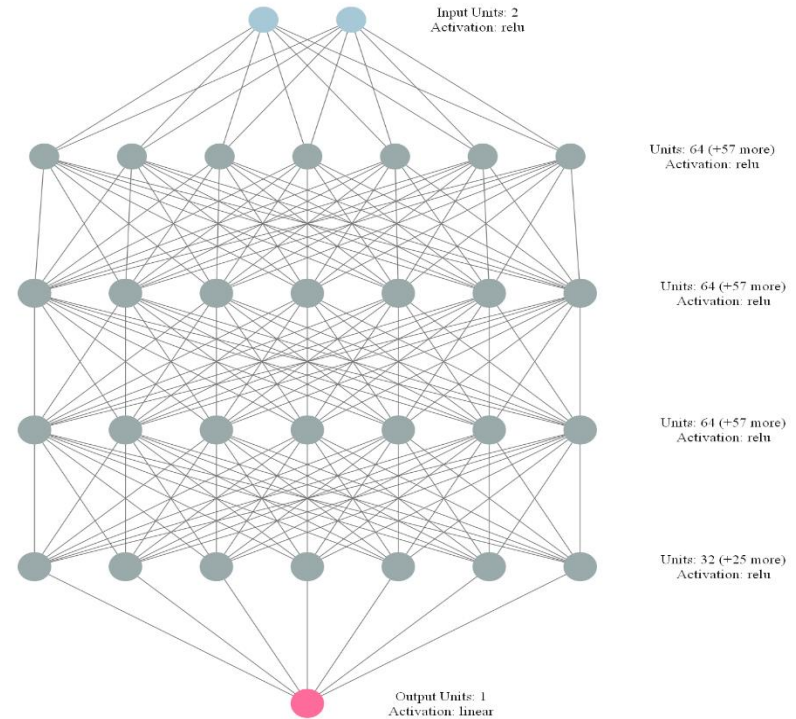


Figure. 3. Neural Network Architecture for Portfolio Strategy Models

This figure demonstrates the architectures of the Neural Network with 2 Hidden Layers (Figure. 3A.) and Neural Network with 4 Hidden Layers (Figure. 3B.) used in this study. Both models utilize two input units corresponding to the variables *chempia* (Industry-adjusted change in employees) and *consec_increase* (Number of consecutive earnings increases), with ReLU activation for all hidden layers and a linear activation function for the output layer. The NN-2 model includes two hidden layers with 64 and 32 units, respectively, while the NN-4 model incorporates four hidden layers with 64 units in the first two layers, 64 units in the third layer, and 32 units in the final layer. These configurations are used to predict stock returns based on the aforementioned input variables.

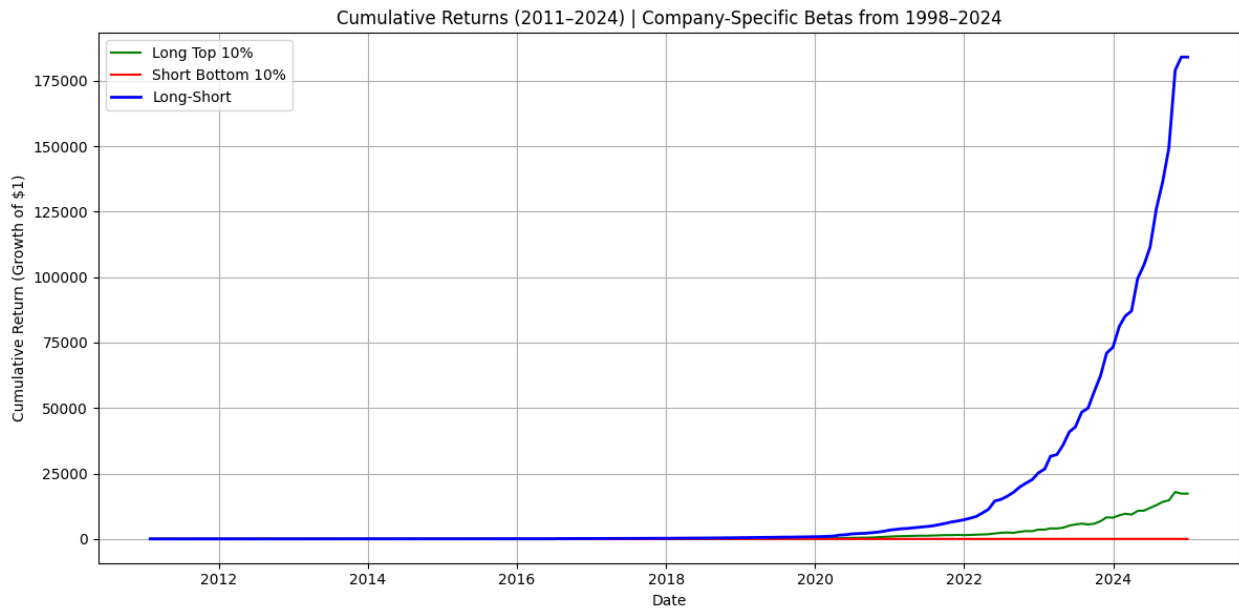


Figure. 4. Cumulative Returns (2011–2024) | Full-Sample OLS-Based Long-Short Strategy Using Company-Specific Betas

This figure shows the cumulative returns from 2011 to 2024 for the Long Top 10% (green line), Short Bottom 10% (red line), and Long-Short (blue line) strategies. These strategies are based on company-specific betas, utilizing the *chempia* (Industry-adjusted change in employees) and *consec_increase* (Number of consecutive earnings increases) variables, and are evaluated using the full-sample OLS method. Consistent with Green, Hand, and Zhang (2017), who validate the effectiveness of such characteristics in predicting stock returns, the Long-Short strategy outperforms the others. This aligns with typical findings in finance journals and mirrors the outcomes of cross-sectional analysis, where the long-short strategy yields the best performance, reinforcing the predictive power of these variables in identifying profitable stock portfolios.

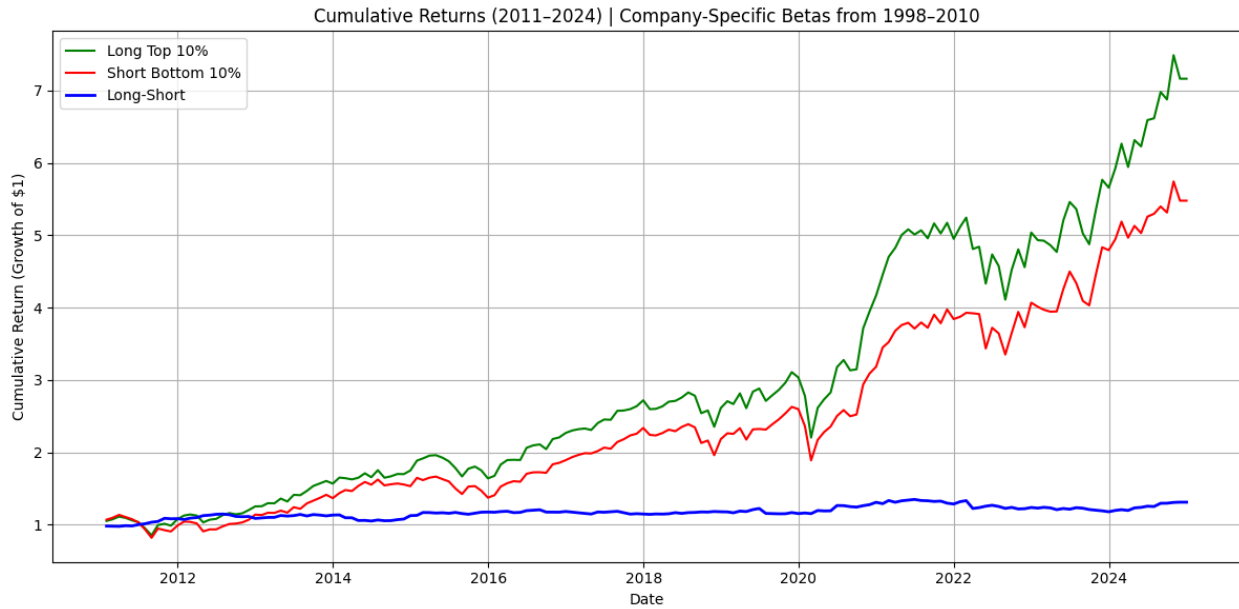


Figure. 5. Cumulative Returns (2011–2024) | Long-Short Strategy Based on OLS Predicted Monthly Returns Using Company-Specific Betas Trained on 1998–2010 Data

This figure displays the cumulative returns from 2011 to 2024 for the Long Top 10% (green line), Short Bottom 10% (red line), and Long-Short (blue line) strategies. The strategies are based on company-specific betas, utilizing the *chempia* and *consec_increase* variables. The OLS model is trained on data from 1998 to 2010, with predicted returns used for cross-sectional analysis starting from 2011. The figure shows that the Long Only strategy yields the best performance, while the Short Bottom 10% strategy does not generate negative returns as expected, resulting in the Long-Short strategy producing the worst performance. These results differ from those observed in the full-sample OLS model, highlighting the unique dynamics of the strategy when trained on data from 1998 to 2010 and predicting returns from 2011 onward.

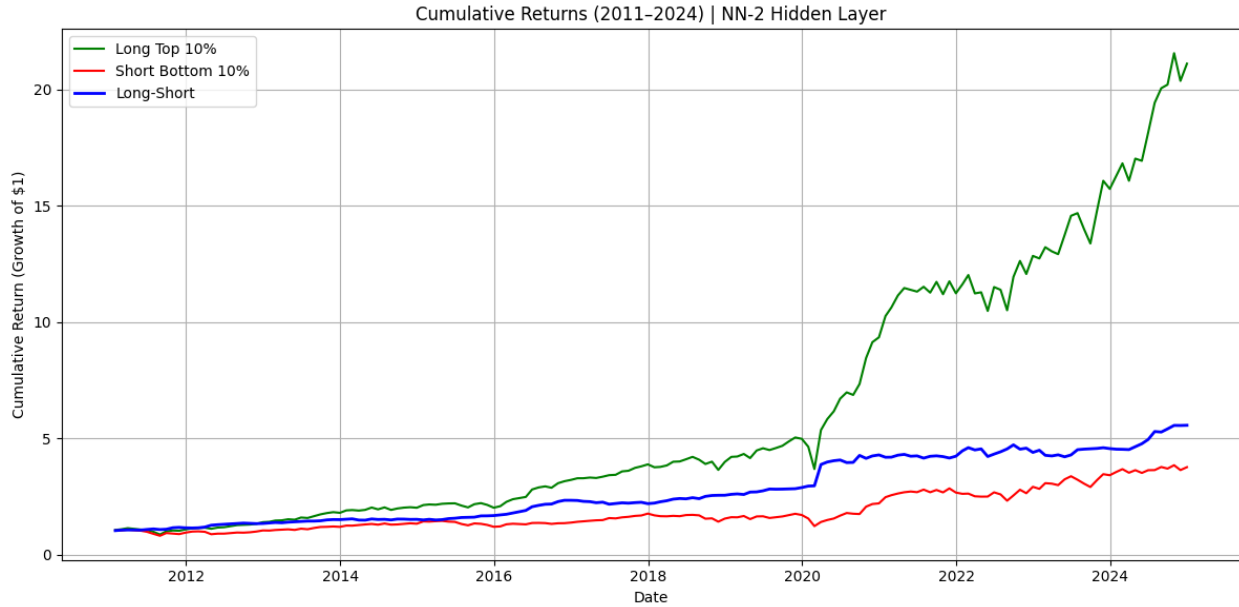


Figure. 6. Cumulative Returns (2011–2024) | Neural Network Model with 2 Hidden Layers Using Company-Specific Features

This figure depicts the cumulative returns from 2011 to 2024 for the Long Top 10% (green line), Short Bottom 10% (red line), and Long-Short (blue line) strategies, utilizing a Neural Network Model with 2 Hidden Layers. The model is trained using *chempia* and *consec_increase* variables, with a training set from 1998 to 2010 and a test set from 2011 to 2024. As shown, the Long Only strategy outperforms the others, achieving the highest cumulative return among the three strategies, and surpassing the performance of the OLS-based strategy. Additionally, the Short Bottom 10% strategy also performs better than its OLS counterpart. These results highlight the improved predictive power of the Neural Network Model with 2 Hidden Layers compared to the traditional OLS approach.

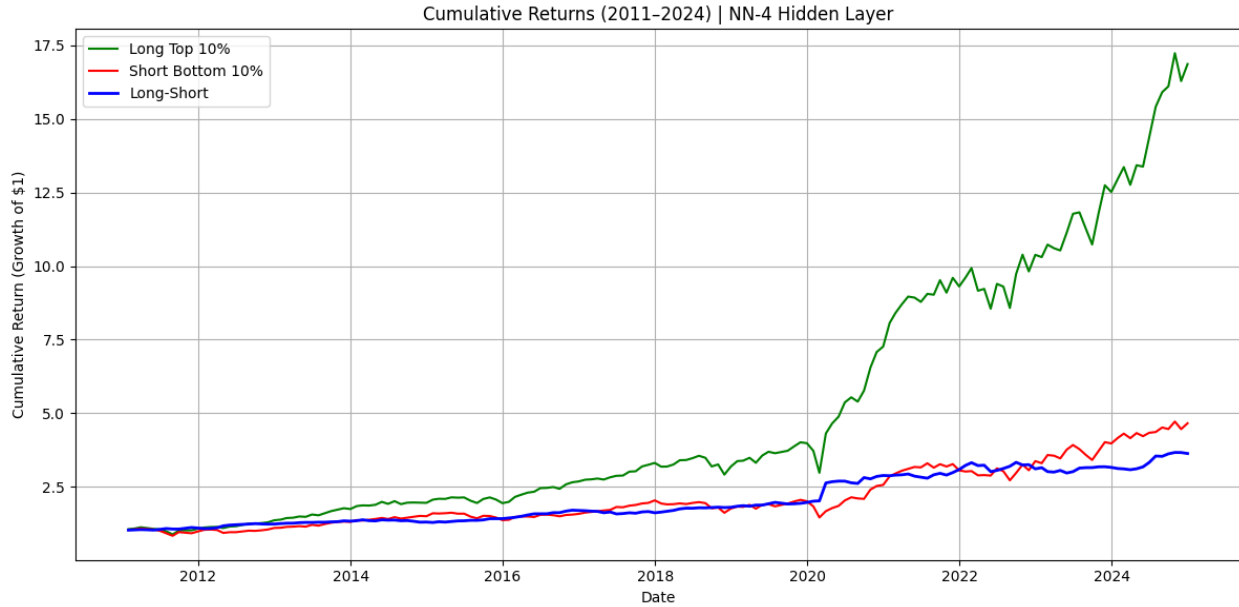


Figure. 7. Cumulative Returns (2011–2024) | Neural Network Model with 4 Hidden Layers

This figure illustrates the cumulative returns from 2011 to 2024 for the Long Top 10% (green line), Short Bottom 10% (red line), and Long-Short (blue line) strategies, using a Neural Network Model with 4 Hidden Layers. The model utilizes `chempia` and `consec_increase` as input variables. Trained on data from 1998 to 2010, the model predicts returns for the test set from 2011 to 2024. The Long Only strategy yields the highest cumulative return among the three, with a slight difference compared to the 2-hidden layer model. While the 4 Hidden Layer model outperforms the OLS model in terms of overall returns, the performance of the Long Only, Short Only, and Long-Short strategies is slightly weaker compared to the 2 Hidden Layer model. This suggests that additional hidden layers may lead to some degree of overfitting, which impacts the model's ability to generalize across all strategies.

Table. 1. Descriptive Statistics of Annual Firm-Level Data (1998–2024)

This table presents the annual summary statistics for key firm-level variables from 1998 to 2024. The data includes `firm_months`, representing the total number of months of valid data for each distinct firm within a given year; `distinct_firms`, indicating the number of firms with available data in each year; `mean_ret`, the average monthly return for firms in a given year; and `median_ln_mktcap`, the median of the log-transformed market capitalization for firms in each year. This table describes the distribution of market capitalization, mean monthly returns, and the number of firms across years, providing insights into firm characteristics and trends over time.

year	firm_months	distinct_firms	mean_ret	median_ln_mktcap
1998	24836	3205	0.010102	20.7964
1999	35667	3760	0.027562	20.7052
2000	39034	3998	0.018832	20.6749
2001	36585	3545	0.02049	20.7809
2002	35263	3309	-0.00106	20.8565
2003	36712	3316	0.064508	20.9049
2004	37433	3504	0.017396	21.1538
2005	38947	3684	0.009777	21.1826
2006	33360	3115	0.018316	21.587
2007	32666	3121	0.00703	21.7525
2008	33094	3135	-0.02776	21.4525
2009	34299	3136	0.049637	21.1665
2010	33748	3027	0.030219	21.5045
2011	33703	2994	-0.00059	21.6976
2012	34236	2897	0.015948	21.6917
2013	34015	2906	0.024735	21.9017
2014	35333	3067	0.006866	22.0145
2015	36534	3155	-0.00078	21.9293
2016	36345	3092	0.019828	21.8673
2017	35599	3021	0.024328	22.0612
2018	35603	3064	0.00234	22.1167
2019	36154	3073	0.024639	22.0783
2020	37455	3260	0.029853	21.9633
2021	37229	3305	0.02353	22.3378
2022	37909	3242	-0.00349	22.1611
2023	37896	3159	0.025004	22.1485
2024	34221	3048	0.023389	22.2869

Table. 2. Descriptive Statistics of Key Variables (1998–2024) for 953,876 US Stocks

This table presents the summary statistics for the key variables used to predict return predictability as outlined in Green, Hand, and Zhang (2017). The variables include chempia (Industry-adjusted change in employees), consec_increase (Number of consecutive earnings increases), prccm (Monthly closing price), and ln_mktcap (Logarithm of monthly closing price multiplied by shares outstanding). The data spans from 1998 to 2024, covering a sample of North American stocks. As emphasized by Green et al. (2017), since 2003, chempia and consec_increase have become independent predictors of average stock returns, significantly contributing to the understanding of stock return behavior.

	chempia	consec_increase	prccm	ln_mktcap
count	953876	953876	953876	953876
mean	0.2467	2.0084	120.4425	21.9372
std	7.3676	2.5641	4811.833	1.5971
min	-409.851	0	0.05	18.9607
10%	-1.2134	0	8.81	20.1582
25%	-0.2972	0	16.34	20.7762
50%	-0.0184	1	29.09	21.6496
75%	0.3469	3	49.8	22.8639
90%	1.7696	7	84.61	24.1416
max	491.6679	8	724040	32.4178

Table. 3. Architecture Summary of Neural Network Models

This table describes the architectures of the Neural Network–2 Hidden Layer and Neural Network–4 Hidden Layer models. Each model uses fully connected *dense* layers, where every neuron connects to all neurons in the previous layer. The output shape indicates the dimensionality of each layer, with *None* representing the unspecified batch size during training. The parameter count (Param #) reflects the total number of trainable weights and biases within each dense layer, indicating the model’s capacity and complexity. All models are trained using the Adam optimizer with a learning rate of 0.001 and mean squared error (MSE) as the loss function, and their performance is evaluated based on *loss* and *compile_metrics*.

Neural Network - 2 Hidden Layer		
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	192
dense_1 (Dense)	(None, 32)	2,080
dense_2 (Dense)	(None, 1)	33
Neural Network - 4 Hidden Layer		
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	192
dense_1 (Dense)	(None, 64)	4,160
dense_2 (Dense)	(None, 64)	4,160
dense_3 (Dense)	(None, 32)	2,080
dense_4 (Dense)	(None, 1)	33

Table. 4. OLS Estimation Results for Long, Short, and Long-Short Portfolios

This table reports the OLS estimation results for the Long Only, Short Only, and Long-Short portfolios using the variables *chempia* (Industry-adjusted change in employees) and *consec_increase* (Number of consecutive earnings increases). The upper panel presents the results based on the full-sample OLS model (1998–2024) applied to the portfolio formation period from 2011 to 2024. The lower panel shows the results for the OLS model trained on the 1998–2010 period, with coefficients used to predict and evaluate portfolio performance during 2011–2024. Reported statistics include the mean, median, and dispersion measures for the estimated coefficients, as well as t-statistics and p-values for model significance.

1998–2024 Long	β_0	β_1 : CHEMPIA	β_2 : CONSEC INCREASE	R ²	Adj. R ²
# of Firms	2489	2489	2489	2489	2489
Min	-3.8492	-9.5806	-1.2103	0.0000	-0.0938
Max	2.5379	73.0288	6.8386	0.4695	0.4383
Mean	0.0267	0.0573	0.0046	0.0368	0.0067
Median	0.0224	-0.0013	-0.0001	0.0209	-0.0003
25%	0.0056	-0.0306	-0.0069	0.0088	-0.0114
75%	0.0421	0.0319	0.0066	0.0470	0.0152
t-stat	9.8837***	1.3473	1.5004	40.2387***	8.1155***
p-value	<0.001	0.178	0.134	<0.001	<0.001

1998–2024 Short	β_0	β_1 : CHEMPIA	β_2 : CONSEC INCREASE	R ²	Adj. R ²
# of Firms	2951	2951	2951	2951	2951
Min	-3.8492	-15.1974	-1.2103	0.0000	-0.0945
Max	2.5379	73.0288	6.8386	0.4695	0.4383
Mean	0.0143	0.0403	0.0024	0.0325	0.0061
Median	0.0095	-0.0008	-0.0009	0.0178	-0.0002
25%	-0.0012	-0.0205	-0.0061	0.0074	-0.0081
75%	0.0249	0.0144	0.0041	0.0399	0.0128
t-stat	6.5919***	1.1225	0.9477	41.5322***	8.7961***
p-value	<0.001	0.262	0.343	<0.001	<0.001

Table. 4.
Continued

1998–2024 Long-Short	β_0	β_1 : CHEMPIA	β_2 : CONSEC INCREASE	R^2	Adj. R^2
# of Firms	7198	7198	7198	7198	7198
Min	-3.8492	-15.1974	-1.2103	0.0000	-0.0947
Max	3.6887	73.0288	6.8386	0.6478	0.6143
Mean	0.0183	0.0145	0.0006	0.0283	-0.0012
Median	0.0116	-0.0004	-0.0002	0.0127	-0.0050
25%	0.0020	-0.0169	-0.0047	0.0042	-0.0164
75%	0.0271	0.0141	0.0035	0.0342	0.0060
t-stat	15.0925***	0.9694	0.5854	56.9950***	-2.6865***
p-value	<0.001	0.332	0.558	<0.001	0.007

1998–2010 Long	β_0	β_1 : CHEMPIA	β_2 : CONSEC INCREASE	R^2	Adj. R^2
# of Firms	1111	1111	1111	1111	1111
Min	-0.4701	-3.2561	-0.5022	0.0002	-0.0878
Max	1.7979	9.7058	0.3204	0.3816	0.3227
Mean	0.0366	0.0080	-0.0009	0.0421	0.0083
Median	0.0238	-0.0029	-0.0009	0.0263	0.0000
25%	0.0061	-0.0397	-0.0074	0.0120	-0.0134
75%	0.0502	0.0322	0.0062	0.0553	0.0201
t-stat	13.0977***	0.6911	-0.9285	29.3622***	6.0768***
p-value	<0.001	0.49	0.353	<0.001	<0.001

1998–2010 Short	β_0	β_1 : CHEMPIA	β_2 : CONSEC INCREASE	R^2	Adj. R^2
# of Firms	1212	1212	1212	1212	1212
Min	-0.4822	-3.2561	-0.5022	0.0002	-0.0797
Max	1.2845	3.2375	0.2573	0.3816	0.3227
Mean	0.0216	0.0004	-0.0035	0.0383	0.0089
Median	0.0137	-0.0015	-0.0019	0.0234	0.0008
25%	-0.0005	-0.0277	-0.0081	0.0110	-0.0098
75%	0.0313	0.0177	0.0038	0.0500	0.0186
t-stat	10.3881***	0.0569	-4.2916***	30.7134***	7.8890***
p-value	<0.001	0.955	<0.001	<0.001	<0.001

Table. 4.
Continued

1998–2010 Long-Short	β_0	β_1 : CHEMPIA	β_2 : CONSEC INCREASE	R ²	Adj. R ²
# of Firms	4834	4834	4834	4834	4834
Min	-1.1233	-7.8961	-0.5022	0.0000	-0.0947
Max	3.6887	9.7058	0.3204	0.6478	0.6143
Mean	0.0212	-0.0005	-0.0017	0.0327	-0.0010
Median	0.0123	-0.0004	-0.0006	0.0167	-0.0078
25%	0.0007	-0.0190	-0.0057	0.0059	-0.0186
75%	0.0300	0.0174	0.0036	0.0412	0.0085
t-stat	16.1792***	-0.1261	-4.2410***	50.7997***	-1.6164
p-value	<0.001	0.9	<0.001	<0.001	0.106

Table 5. Performance Evaluation of Portfolio Strategies Across Models

This table reports the performance metrics for different models and trading strategies, including the Full Sample OLS, Test Set OLS, Neural Network with 2 Hidden Layers (NN-2), and Neural Network with 4 Hidden Layers (NN-4). The strategies evaluated are Long Only, Short Only, and Long–Short portfolios. Performance is assessed using five key indicators: Mean Return, Volatility, Sharpe Ratio, Maximum Drawdown, and Compound Annual Growth Rate (CAGR).

Model	Strategy	Mean Return	Volatility	Sharpe Ratio	Max Drawdown	CAGR
Full Sample OLS	Long Only	0.0645*** (7.2920)	0.1143	0.5643	-0.1944	0.9983
	Short Only	-0.014*** (-3.2860)	0.0551	-0.2543	-0.9275	-0.1707
	Long - Short	0.0785*** (9.9123)	0.1023	0.7670	-0.0074	1.371
Test Set OLS	Long Only	0.0131*** (3.1114)	0.0541	0.2415	-0.2971	0.1478
	Short Only	0.0109*** (2.7416)	0.0513	0.2128	-0.283	0.1207
	Long - Short	0.0022* (1.6765)	0.0166	0.1301	-0.1229	0.0244
NN-2 Hidden Layer	Long Only	0.0197*** (4.3121)	0.0591	0.3337	-0.2699	0.2388
	Short Only	0.0091** (2.3019)	0.0511	0.1781	-0.3067	0.0971
	Long - Short	0.0106*** (4.2682)	0.0322	0.3303	-0.1091	0.1282
NN-4 Hidden Layer	Long Only	0.0183*** (4.0674)	0.0582	0.3147	-0.259	0.2192
	Short Only	0.0103*** (2.6804)	0.0496	0.2074	-0.2928	0.1135
	Long - Short	0.008*** (3.4328)	0.0302	0.2656	-0.109	0.0948