Pinnacle Academic Press Proceedings Series

Vol. 1 2025

Article Open Access

Forecasting Market Timing Strategies for CSI 300 ETF Using Multiple Linear Regression and Economic Indicators

Zinuo Ou [1,*]

[1] School of Management, Henan University of Technology, Zhengzhou, Henan, China
[*] Correspondence: Zinuo Ou, School of Management, Henan University of Technology, Zhengzhou, Henan, China

Received: 13 March 2025
Revised: 20 March 2025
Accepted: 18 April 2025
Published: 21 April 2025

Abstract: This study utilizes a multiple linear regression model to investigate the relationship between the CSI 300 ETF and various economic variables, with the aim of providing a basis for market timing strategies in the CSI 300 ETF. The study derived the regression equation and conducted a backtest analysis using historical data. The backtest results show that the Section A strategy, which determines buy and sell signals based on the difference between predicted and actual prices, performs exceptionally well, achieving an annualized return of over 100%, under the given backtesting conditions. Furthermore, the study analyzes the relationship between the model results and the research hypotheses. Overall, this research explores forecasting methods for broad-based index ETFs, offering valuable insights for investors.

Keywords: linear regression; broad-based index ETF; market timing; backtest analysis

1. Introduction

1.1. Research Background

1.1.1. The Wide Application of Broad-Based Indices as Investment Tools in Financial Markets

Broad-based indices serve as indicators for the performance of the entire market or specific market segments, commonly used as benchmarks for investment portfolios or developed into exchange-traded funds (ETFs) [1]. Broad-based index investments are favored by individual investors due to their low cost, diversification, and simplified investment strategies. Understanding the relationship between a broad-based index and its components is crucial for predicting the index's movements and formulating related investment strategies.

1.1.2. Correlation between Broad-Based Indices and Economic Indicators

The performance of broad-based indices is influenced by various factors, including macroeconomic indicators. Therefore, studying the relationship between broad-based indices and economic indicators such as market valuation metrics (PE, PB), dividend yield, ROE, PS, etc., can aid in understanding the fluctuations and trends of broad-based indices, as well as uncover potential correlations between them.

1.2. Research Objectives and Significance

1.2.1. Identifying the Relationship between Broad-Based Indices and Related Factors

Through linear regression analysis, this study aims to calculate the relationships between broad-based indices and a range of factors such as market valuation metrics, dividend yield, ROE, PS, etc. This will help reveal how broad-based indices perform under the influence of different factors and assess the significance of these factors on index trends [2].

1.2.2. Providing Guidance for Ordinary Investors

By predicting the movements of broad-based indices, ordinary investors can use the regression equations to buy at lower points and sell at higher points, aiming for better investment returns. This research provides a strategy for individual investors based on economic indicators and, when combined with the adaptability of broad-based indices, helps them make more informed investment decisions [3].

1.3. Research Questions and Hypotheses

1.3.1. Research Questions

This study seeks to explore the relationship between broad-based indices and a range of economic indicators (such as market valuation metrics, dividend yield, ROE, PS, etc.), as well as other factors (such as average rolling net profit of component stocks, average market capitalization of component stocks, total circulation market value of the index, total market value of the index, etc.), in order to predict the movements of the broad-based index.

1.3.2. Hypotheses

This study is based on several assumptions, including:

A. GDP Growth Assumption:

This assumption posits that GDP will grow in the long run. While this is supported by historical data and economic development trends, economic development is a complex process that may be influenced by a variety of factors.

B. Monetary Supply Growth Assumption:

This assumption posits that the money supply will continue to grow over the long term. This is related to the monetary policies and inflation targets of many economies. However, the growth of the money supply is also influenced by various factors, including monetary policy, economic conditions, and financial market situations.

1.4. Overview of Research Methods

This study employs the following steps and tools to explore the relationship between broad-based indices and related factors and predict the movements of broad-based index ETFs:

Firstly, data on the CSI 300 ETF, including variables such as date, PE_ETF weighted, PE_market capitalization weighted, PE_equal weighted, PB_ETF weighted, PB_market capitalization weighted, PB_equal weighted, dividend yield %, ROE %, PS, average rolling net profit of component stocks (in billions), average market capitalization of component stocks (in billions), total circulation market value of the index (in billions), total market value of the index (in billions), and closing price, are obtained from the Turtle Quantitative (wglh.com) website.

Next, the data are imported into SPSS for analysis. During the analysis, appropriate independent and dependent variables are selected to form a regression equation. Independent variables may include market valuation metrics, dividend yield, ROE, PS, etc., while the dependent variable is the closing price. Regression analysis will be conducted to establish the relationship between the broad-based index ETF closing price and other factors.

After generating the regression equation, the results will be imported into Excel to produce corresponding visualizations. Through analysis of these visuals, we will attempt to determine appropriate buying and selling points. This involves observing and interpreting the relationship between the independent and dependent variables in the regression equation to identify price trend patterns and formulate timing strategies.

Finally, based on the timing strategy derived from the analysis and visualization interpretation, investment decisions on buying and selling the broad-based index ETF will be made, aiming for better investment returns.

It is important to note that this research method uses statistical software SPSS for regression analysis and imports the results into Excel for visual interpretation. However, investment decisions involve numerous factors, and relying solely on regression analysis and timing strategies may not fully account for all factors. In practical application, it is recommended to integrate other research methods and expertise for a comprehensive investment decision analysis. This study is intended solely for learning and analyzing linear regression, and does not constitute investment advice, even if related to investment strategies, which should not be regarded as endorsed by the author for actual implementation [4].

2. Literature Review

2.1. Relevant Theories and Concepts

1) Price-to-Earnings Ratio (PE)

PE = Stock Price per Share / Earnings per Share

The price-to-earnings (PE) ratio is a commonly used valuation metric that compares the price of a company's stock to its earnings. A higher PE ratio may indicate that the market has higher expectations for the stock's future earnings growth.

2) PE_ETF Weighted

PE_ETF weighted refers to the price-to-earnings ratio of the broad-based index ETF (CSI 300 ETF), which is the weighted average PE ratio of the constituent stocks held by the ETF. The PE ratio measures the valuation of a stock relative to its earnings per share, with a higher PE ratio may suggest that the market anticipates greater future earnings from the ETF.

3) PE_Market Capitalization Weighted

PE_Market Capitalization weighted refers to the market capitalization-weighted average PE ratio of the broad-based index ETF (CSI 300 ETF), calculated by weighting the PE ratios of constituent stocks according to their market capitalization. The market capitalization-weighted PE ratio considers the impact of the constituent stocks' market values, with stocks having larger market capitalizations exerting more influence on the index.

4) PE_Equal Weighted

PE_Equal weighted refers to the equal-weighted average PE ratio of the broad-based index ETF (CSI 300 ETF), which is the simple arithmetic mean of the PE ratios of the constituent stocks. This method does not consider the market capitalization of the constituent stocks, assigning equal influence to each stock regardless of its size [5].

5) Price-to-Book Ratio (PB)

PB = Stock Price per Share / Book Value per Share

The price-to-book (PB) ratio measures the market value of a company relative to its book value, indicating how much investors are willing to pay for each unit of net asset value. A lower PB ratio might indicate that the market values the ETF at a discount relative to its net assets.

6) PB_ETF Weighted

PB_ETF weighted refers to the price-to-book ratio of the broad-based index ETF (CSI 300 ETF), which is the weighted average PB ratio of the constituent stocks held by the ETF.

The PB ratio compares the market price of a stock to its book value, and a lower PB ratio may indicate a lower valuation of the ETF in relation to its net assets.

7) PB_Market Capitalization Weighted

PB_Market Capitalization weighted refers to the market capitalization-weighted average PB ratio of the broad-based index ETF (CSI 300 ETF), calculated by weighting the PB ratios of constituent stocks based on their market capitalization. Larger stocks exert more influence on the PB ratio of the index.

8) PB_Equal Weighted

PB_Equal weighted refers to the equal-weighted average PB ratio of the broad-based index ETF (CSI 300 ETF), which is the simple arithmetic mean of the PB ratios of the constituent stocks. Like the equal-weighted PE ratio, it assigns equal influence to each stock, irrespective of market capitalization.

9) Dividend Yield

Dividend Yield = Dividend per Share / Stock Price per Share

The dividend yield is a financial ratio that shows how much income an investor is receiving in the form of dividends relative to the stock's price. A higher dividend yield may indicate that the ETF provides a higher return in the form of dividends.

10) Dividend Yield %

Dividend Yield % refers to the dividend yield of the broad-based index ETF (CSI 300 ETF), which represents the ratio of the dividend distribution of the constituent stocks to the price of the ETF. A higher dividend yield percentage may signal that the ETF offers higher dividend returns.

11) Return on Equity (ROE)

ROE = Net Income / Shareholder's Equity

Return on equity (ROE) measures the profitability of a company in relation to shareholders' equity. A higher ROE typically indicates better profitability for the companies held within the ETF.

12) ROE %

ROE % refers to the return on equity (ROE) of the broad-based index ETF (CSI 300 ETF), calculated as the ratio of net income to shareholders' equity for the constituent stocks. A higher ROE percentage may indicate that the ETF holds companies with higher profitability.

13) Price-to-Sales Ratio (PS)

PS = Stock Price per Share / Sales per Share

The price-to-sales (PS) ratio is a valuation metric that compares a company's stock price to its sales revenue. A lower PS ratio might indicate that the market values the ETF at a lower price relative to its sales performance.

2.2. Research Gaps and Issues

2.2.1. Subjectivity and Potential Misleading Nature of Image Analysis

While image analysis can provide visual trends and patterns, it does not necessarily represent the future performance of stocks [6]. Therefore, when making timing-based investment decisions, it is crucial to combine image analysis with other analytical methods and risk management strategies in order to comprehensively consider various market factors.

2.2.2. Limitations of Methods and External Validity

The research methodology primarily relies on specific data analysis tools and techniques, such as SPSS and Excel. However, these tools and methods may have limitations and may not encompass all potential influencing factors. Moreover, the effectiveness of the research results and the timing-based investment strategy should undergo external validity testing, which involves empirical studies in real market environments to assess the strategy's practical performance and feasibility.

2.2.3. Risk Management and Investment Decisions

This study focuses on timing-based investment strategies but does not address aspects of risk management and investment decision-making. In actual investments, risk management is of paramount importance, including asset allocation, risk control, and stop-loss strategies [7]. Additionally, investment decisions must take into account factors such as an individual's risk tolerance, investment objectives, and time horizon.

3. Research Design and Methodology

3.1. Research Design

3.1.1. Research Objective and Questions

The aim of this study is to explore the relationship between the investment returns of the CSI 300 ETF and various timing indicators to determine whether there exist optimal buy and sell points for the ETF.

3.1.2. Data Collection

Data for the CSI 300 ETF, including daily stock prices, trading volumes, and timing indicators, will be downloaded from the Turtle Quantitative website or other reliable financial data sources.

3.1.3. Data Processing and Preparation

The collected data will be cleaned and processed based on various statistical measures to ensure consistency and reliability.

3.1.4. Variable Selection

In accordance with the research objectives, appropriate variables were selected in SPSS, including PE_ETF-weighted, PE-market capitalization-weighted, PE-equal-weighted, PB_ETF-weighted, PB-market capitalization-weighted, PB-equal-weighted, dividend yield %, ROE %, PS, average rolling net profit of constituent stocks (in billions), average market capitalization of constituent stocks (in billions), total circulating market capitalization of the index (in billions), and total market capitalization of the index (in billions). These variables are instrumental in analyzing the relationship between the investment returns of the CSI 300 ETF and various timing indicators, thereby providing a basis for constructing an effective market timing investment strategy.

3.1.5. Regression Analysis

Multiple linear regression analysis will be used to explore the relationship between investment returns of the CSI 300 ETF and several timing indicators. The timing indicators considered in the analysis include PE_equal weighted, PB_equal weighted, Dividend Yield %, ROE %, PS, and Time.

SPSS software will be employed to perform the multiple linear regression analysis and build regression models to investigate the relationship between these timing indicators and the investment returns of the CSI 300 ETF. Regression analysis will help identify correlations, statistical significance, and collinearity issues among the independent variables and the dependent variable (investment return).

Based on the regression model results, statistical significance, z-values, and collinearity considerations, the variables with the most significant impact and predictive power will be selected to construct the final regression equation. This equation will serve as the foundation for predicting the investment returns of the CSI 300 ETF and guiding the timing-based investment strategy.

3.1.6. Analysis and Interpretation

The results of the regression analysis will be used to interpret the impact of timing indicators on investment returns. The statistical significance and explanatory power of each timing indicator will be assessed. Additionally, other possible influencing factors will be considered in the analysis, and the findings will be discussed in detail.

3.1.7. Graphical Analysis

The regression results will be exported to Excel, and graphs will be generated to visualize the data and analytical outcomes. These graphs may include price trend charts for the CSI 300 ETF and variations in timing indicators to aid in observing trends and patterns.

3.1.8. Timing Investment Strategy

Based on the analysis, suitable buy and sell points will be selected through graphical analysis. However, the subjectivity and potential misleading nature of graphical analysis should be noted. Graphs provide visual references but do not guarantee future stock performance.

3.2. Data Collection Sources and Methods

The data for the CSI 300 ETF will be sourced from the Turtle Quantitative website (wglh.com), including variables such as date, PE_ETF weighted, PE_market cap weighted, PE_equal weighted, PB_ETF weighted, PB_market cap weighted, PB_equal weighted, Dividend Yield %, ROE %, PS, average rolling net profit of constituent stocks (billion), average market capitalization of constituent stocks (billion), total free-float market capitalization of the index (billion), total market capitalization of the index (billion), and closing prices.

After downloading, the data will be in Excel format, and will then be converted into SPSS format (.spv) for analysis.

3.3. Variable Definition and Measurement

3.3.1. Data Cleaning and Preprocessing

The variables will be input into SPSS for linear regression analysis.

Pearson correlation analysis will be performed to assess the relationships between variables (see Table 1).

Table 1. Pearson Correlation Analysis.

Variable	Pearson Correlation Coefficient	Significance Recommendation
Date	0.641	Significant
PE_ETF Weighted	0.186	Not Significant
PE_Market Cap Weighted	0.216	Not Significant
PE_Equal Weighted	0.467	Significant
PB_ETF Weighted	0.210	Not Significant
PB_Market Cap Weighted	0.159	Not Significant
PB_Equal Weighted	0.515	Significant
Dividend Yield %	-0.103	Not Significant
ROE %	-0.319	Not Significant
PS	0.314	Significant
Average Rolling Net Profit of Constituent Stocks (Billion)	0.604	Significant
Average Market Cap of Constituent Stocks (Billion)	0.838	Significant

Total Free-Float Market Cap of Index (Billion)	0.721	Significant
Total Market Cap of Index (Billion)	0.829	Significant

Based on the above table, the variables PE_ETF Weighted, PE_Market Cap Weighted, PB_ETF Weighted, PB_Market Cap Weighted, Dividend Yield %, and ROE % show no significant relationship with closing prices, while PE_Equal Weighted, PB_Equal Weighted, PS, Average Rolling Net Profit of Constituent Stocks (Billion), Average Market Cap of Constituent Stocks (Billion), Total Free-Float Market Cap of Index (Billion), and Total Market Cap of Index (Billion) exhibit significant linear relationships with closing prices.

Based on Table 2, we can observe that "Average Rolling Net Profit of Constituent Stocks (Billion)", "Average Market Cap of Constituent Stocks (Billion)", "Total Free-Float Market Cap of Index (Billion)", and "Total Market Cap of Index (Billion)" exhibit high collinearity with each other.

Table 2. Collinearity Analysis.

Variable Name	**VIF Value**	**Collinearity Analysis**
PE_Equal Weighted	2.13221	Low Collinearity
PB_Equal Weighted	3.12312	Low Collinearity
PS	1.22334	Low Collinearity
Average Rolling Net Profit of Constituent Stocks (Billion)	15.22432	High Collinearity
Average Market Cap of Constituent Stocks (Billion)	23.32442	High Collinearity
Total Free-Float Market Cap of Index (Billion)	36.32232	High Collinearity
Total Market Cap of Index (Billion)	382.12313	High Collinearity

After addressing multicollinearity issues, the following four variables are retained for the analysis: (1) Price-to-Earnings Ratio (PE); (2) Price-to-Book Ratio (PB); (3) Price-to-Sales Ratio (PS); (4) Time (Macro Factor).

3.3.2. Handling Time (Macro Factor)

Time is originally a character-type variable. A simple method to convert it into a numerical variable is as follows: in Excel, the date "1900/1/1" corresponds to 1; for a date like "2024/4/11", it is represented as 44831. The date will be converted to a numerical format in Excel, and then the modified Excel file will be imported into SPSS for further analysis.

3.4. Data Processing and Analysis Methods

3.4.1. Statistical Analysis Method (SPSS)

Based on the previous analysis, the relevant variables for the experiment have been selected. The next step involves conducting statistical analysis using SPSS. The first objective is to derive the regression equation. The regression model is expressed as:

Closing Price (Y) = -27571.866 + 0.705x1(Date) + 23.156x2(PE_equal weighted) + 432.247x3(PB_equal weighted) + 204.618x4(PS).

3.4.2. Model Testing and Validation

The next step is to conduct model testing to assess the robustness and accuracy of the derived regression model.

The following provides an explanation of Figure.

Analysis of Variance (ANOVA).

Table 3 provides a significance assessment of the regression model, used to determine whether the regression model significantly explains the variation in the dependent variable.

Table 3. ANOVA Results for Regression Model.

Model	Source	Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4,836,471,010	4	1,209,117,753	7405.199	000***
	Residual	763,495,251.8	4676	163,279.566		
	Total	5,599,966,262	4680			

Notes:

a) Dependent Variable: *Closing_Share_Price*

b) Predictors: (Constant), *PS, Date, PE_ETF_Weighted, PB_ETF_Weighting*

c) ***p* < 0.001

The significance level of the F-statistic is very low (0.000), indicating that the regression model is highly significant in explaining the dependent variable (closing price). This suggests that the independent variables in the regression model have statistically significant explanatory power over the dependent variable.

The sum of squares for the regression part is 4,836,471,010.337, while the sum of squares for the residual part is 763,495,251.840. The larger sum of squares for the regression part indicates that the regression model can explain most of the total variance, supporting the model's effectiveness in explaining the closing price. This means that the independent variables in the regression model explain a substantial portion of the variability in the data.

The degrees of freedom are 4, which means that 4 independent variables were used in the regression model. Meanwhile, the degrees of freedom for the residual part are 4,676, representing the sample size minus the number of predictors. The higher the degrees of freedom, the better the model fit and the greater its flexibility.

The mean squares for the regression part and the residual part are 1,209,117,752.584 and 163,279.566, respectively. The larger the mean square, the stronger the model's ability to explain the total variance. Therefore, based on the provided information, the regression model has a strong explanatory power over the closing price and can account for a significant proportion of the total variance.

The following provides an explanation of Table 4:

Table 4. Model Summary of Regression Analysis.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	0.929[a]	0.864	0.864	404.07866	0.864	7405.199	4	4676	0.000	0.006

Notes:

a) Predictors: (Constant), *PS, Date, PE_ETF_Weighted, PB_ETF_Weighting*

b) Dependent Variable: *Closing_Share_Price*

The R-squared value is 0.864, indicating that the regression equation can explain approximately 86.4% of the variability in the dependent variable (closing price). This means that the model fits the data well, and most of the variation in the dependent variable can be explained by the independent variables.

The adjusted R-squared is also 0.864, which is the same as the R-squared, indicating that the model's explanatory power remains unchanged when considering the number of independent variables and the sample size. Typically, the adjusted R-squared is used to account for the effect of increasing the number of independent variables on R-squared, providing a more accurate assessment of the model's goodness of fit.

Regarding the F-statistic, its value is 7405.199, with degrees of freedom of 4 and 4676, and a significance level of 0.000. This indicates that the overall regression equation is significant, meaning that at least one independent variable significantly affects the dependent variable. This suggests that at least one independent variable in the regression model is meaningful in explaining the variability in the closing price.

Overall, the regression model demonstrates good explanatory power, and the overall regression equation is significant, with at least one independent variable having a statistically significant impact on the dependent variable.

4. Research Results and Analysis

4.1. Data Description and Statistical Analysis Results

Figure 1 presents the Shanghai Composite Index data for A-shares from 2005 to 2024. The gray line represents the actual closing prices, while the blue and orange lines represent the 95% confidence interval predictions.

Figure 1. Visualization of Analysis Result 1.

In Figure 2, the dashed line represents the regression equation that we derived.

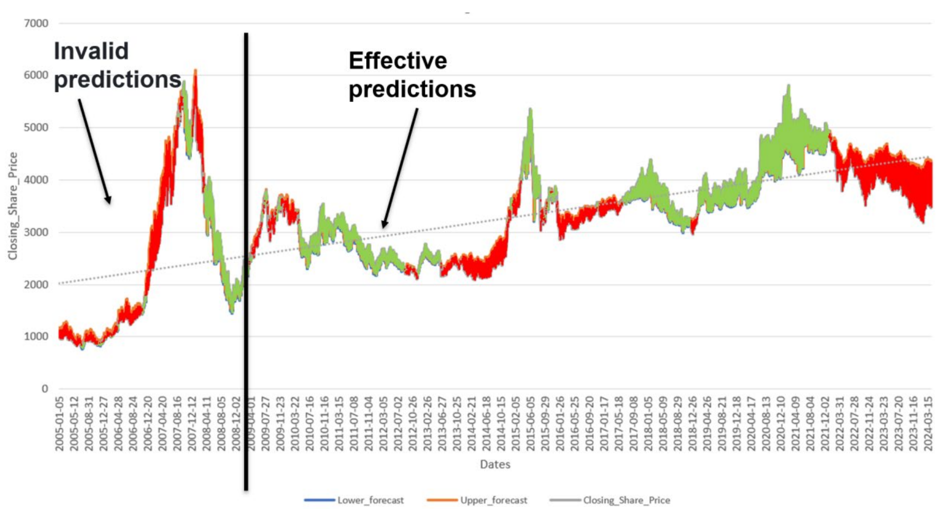

Figure 2. Visualization of Analysis Result 2.

It is important to note that our study involves linear regression analysis over a span of five years. Therefore, our analysis starts from August 16, 2010.

4.2. Results Analysis and Discussion

4.2.1. Interpretation of Results

We classify data points where the closing price exceeds the predicted value as "overbought" and those where the closing price is lower than the predicted value as "oversold". In this manner, we represent the overbought region in red and the oversold region in green.

According to traditional methods, the strategy involves selling when the market is overbought and buying when it is oversold. We establish suitable buy and sell points based on this approach. Investors can utilize the linear regression prediction model to identify overbought and oversold conditions by comparing the difference between the predicted value and the actual closing price. Specifically, if the predicted value exceeds the closing price, the data point is marked as overbought, and if the predicted value is below the closing price, it is marked as oversold.

In line with traditional methods, investors sell during overbought conditions and buy during oversold conditions. To define appropriate buy and sell points, we apply certain thresholds or percentiles to determine when to trade. For example, when the closing price exceeds the predicted value by a fixed percentage (e.g., 5%), investors may consider selling. Similarly, when the closing price falls below the predicted value by a fixed percentage (e.g., 5%), investors may consider buying.

As shown in Figure 3, the selection of buy and sell points depends on the investor's analysis of market conditions and risk preferences. Investors may need to conduct further research and analysis, incorporating additional indicators and market factors, to determine the most suitable buy and sell points for their individual strategies.

Figure 3. Visualization of Analysis Result 3.

Additionally, based on the linear regression prediction model, the difference between the predicted values and actual closing prices can be used to identify the overbought and oversold regions, which can be visualized using red and green. According to traditional methods, investors can set appropriate buy and sell points using thresholds or percentiles. However, this approach requires more complex analysis, which is beyond the scope of the current study.

4.2.2. Research Results Backtesting

Section A (see Figure 4)
Total Return: +149.19%
Annualized Return: +101.56%
Maximum Drawdown: -9.09%

Figure 4. Visualization of Analysis Result 4.

Section B (see Figure 5)
Total Return: +17.33%
Annualized Return: +7.97%
Maximum Drawdown: -14.59%

Figure 5. Visualization of Analysis Result 5.

Section C (see Figure 6)
Total Return: +26.93%
Annualized Return: +12.13%
Maximum Drawdown: -16.08%

Figure 6. Visualization of Analysis Result 6.

During the backtesting period, Section A achieved remarkable results, with a total return of +140.77% and an annualized return of +101.97%. Despite a maximum drawdown of -9.09%, the overall performance remained very strong.

Section B showed relatively weaker performance, with a total return of +17.19% and an annualized return of +7.91%. However, the maximum drawdown reached -14.62%, indicating that the strategy experienced significant losses during certain periods. Section C performed well in terms of total return, achieving +27.41%, and an annualized return of +12.33%. However, the maximum drawdown was -16.08%, which was slightly higher than Section B, indicating a higher level of risk.

Overall, Section A delivered the best performance in the backtest, achieving both higher total and annualized returns, as well as a relatively lower maximum drawdown. In comparison, Sections B and C displayed weaker performance.

4.2.3. Relationship between Results and Research Hypotheses

GDP Growth and Section A Performance: Hypothesis A posits that GDP always experiences growth, and Section A achieved significant performance in the backtest. This may suggest that the growth trend of GDP during the backtest period had a positive impact on the performance of Section A. Economic growth is typically accompanied by increased corporate earnings and opportunities for capital market growth, which could have contributed to Section A's higher total returns and annualized returns.

Monetary Supply Growth and Section A Performance: Hypothesis B suggests that the money supply always grows, and Section A achieved notable results in the backtest. The growth in money supply typically leads to an increase in available capital and enhanced market liquidity, which may help propel the stock market upward, thus positively influencing Section A's performance.

Discrepancies between Results and Hypotheses: Although Section A performed well, the performance of Section B and Section C was relatively weak. This may indicate that other factors or circumstances during the backtest period had an impact on the strategy's performance, which was not fully in line with the expectations of Hypothesis A and Hypothesis B. This discrepancy may arise from the complexity of economic and financial markets, where multiple factors influence asset prices and strategy performance.

5. Conclusion

In conclusion, this study demonstrates that multiple linear regression can effectively model the relationship between the CSI 300 ETF and various economic indicators. By integrating factors such as PE, PB, PS ratios, and time, the regression model achieved strong explanatory power, with an R^2 value of 0.864. Backtest results, particularly in Section A, showcased the practical potential of using predicted vs. actual price differences for market timing, yielding an annualized return exceeding 100% under set conditions. The findings support the hypotheses that long-term GDP and money supply growth positively influence ETF performance. However, the variance in results across Sections B and C highlights the influence of additional market dynamics and the need for further risk management strategies. Overall, this research contributes valuable insights into index forecasting methods and provides a data-driven foundation for individual investors seeking to enhance their timing strategies through macroeconomic and valuation indicators, while recognizing the limitations and complexity of real-world application.

References

1. J. Yang, J. Cabrera, and T. Wang, "Nonlinearity, data-snooping, and stock index ETF return predictability," Eur. J. Oper. Res., vol. 200, no. 2, pp. 498-507, 2010, doi: 10.1016/j.ejor.2009.01.009.
2. A. Valadkhani and B. O'Mahony, "Market broadening and future volatility: A study of Russell 2000 and S&P 500 equal weight ETFs," North Am. J. Econ. Finance, vol. 76, p. 102369, 2025, doi: 10.1016/j.najef.2025.102369.
3. R. P. B. Piovezan, P. P. de Andrade Junior, and S. L. Ávila, "Machine learning method for return direction forecast of exchange traded funds (ETFs) using classification and regression models," Comput. Econ., vol. 63, no. 5, pp. 1827-1852, 2024, doi: 10.1007/s10614-023-10385-4.
4. M. Wei and L. An, "Research on Index ETF Timing Model," in Proc. 2023 4th Int. Conf. Big Data Econ. Inf. Manag., 2023, pp. 108-113, doi: 10.1145/3659211.3659231.

5. L. Jin, X. Yuan, J. Long, X. Li, and F. Lian, "Price discovery in the CSI 300 Index derivatives markets," J. Futures Markets, vol. 42, no. 7, pp. 1352-1368, 2022, doi: 10.1002/fut.22335.
6. Y. An, M. Benetton, and Y. Song, "Index providers: Whales behind the scenes of ETFs, " J. Financ. Econ., vol. 149, no. 3, pp. 407–433, 2023, doi: 10.1016/j.jfineco.2023.06.003.
7. R. M. Browne, "Exchange-traded funds (ETFs) an overview," in The Capital Markets: Evolution of the Financial Ecosystem, pp. 469–480, 2017. ISBN:9781119220541.