Vol. 19 (2026): Proceedings of 2026 International Conference on Financial Technology, Mathematical Statistics, and Business Analytics (ICFMB 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/icfmb.202619418425

# Study on the Explanatory Power of Weighted Constituents on the Performance of the STAR 50 ETF

**Xuekai Lai***

*Jinshen College, Nanjing Audit University, Nanjing, China*

**Corresponding author: Xuekai Lai.*

## Abstract

With the deepening of China's science and technology innovation strategy, the STAR 50 ETF has seen its trading volume increase day by day. The study aims to explore whether a small number of high-weight assets can effectively explain and predict the overall performance of the ETF. The study employs empirical methods such as descriptive statistics, correlation analysis, and multiple linear regression, based on daily trading data for 2025. The results show that there is an extremely high positive correlation between the top-weighted constituents and the STAR 50 ETF, with all correlation coefficients exceeding 0.9. Moreover, the price movements of these five stocks can explain more than 99% of the ETF's price fluctuations($R^2$=0.998). This finding not only verifies the effectiveness of partial replication strategies on the STAR 50 ETF, but also provides market participants with a simple and feasible prediction tool — that is, by tracking the performance of a few leading stocks, one can roughly determine the direction of the ETF.

## Keywords

STAR 50 ETF, constituents, price discovery, partial replication, concentration risk

## 1. Introduction

In recent years, China's capital market has embraced a period of rapid development. In particular, with the establishment of the STAR Market, a large number of high-growth technology enterprises have listed, providing investors with abundant investment targets. However, for ordinary investors, directly investing in individual STAR Market stocks often faces high entry barriers and greater individual stock volatility risks [1]. Against this backdrop, exchange-traded funds (ETFs), with their advantages of low cost, high efficiency, and risk diversification, have become the preferred tool for the general public to share in the growth dividends of the technology sector. Data show that China's ETF market has experienced explosive growth in recent years, with the STAR 50 ETF, which tracks the core assets of the STAR Market, receiving widespread attention. Its trading volume and holdings have repeatedly hit new highs [2].

Although ETF products provide investors with the convenience of one-stop investment, in actual trading, investors still face the challenge of how to judge the ETF's performance. Since the STAR 50 Index consists of 50 constituents, comprehensively tracking the changes of all constituents imposes excessively high information processing costs on individual investors with limited capabilities. Academia and industry have generally noted that higher-weighted constituents in an index often exert a decisive influence on the overall performance of the index. This phenomenon raises a thought-provoking question: Can merely observing the rises and falls of a few top-weighted leading constituents effectively infer the overall direction of the ETF? If this hypothesis holds, it would greatly reduce investors' information acquisition costs and provide them with an extremely simplified market observation window.

At the same time, this issue relates to the core topics of index tracking and risk management. On one hand, if a few high-weight stocks dominate the index performance, can fund managers adopt a partial replication strategy during index replication, holding only these core assets to achieve effective tracking of the index and thereby reduce transaction costs? On the other hand, such high dependence may also imply concentration risk — that is, if these core assets experience extreme volatility, will it trigger systemic shocks across the entire ETF or even the sector? Based on this, this study selects the STAR 50 ETF and its core high-weight stocks as the research sample and uses empirical data to test the explanatory power of the top-weighted constituents on the ETF's performance, with the aim of providing empirical evidence for investors' trading decisions, market makers' arbitrage activities, and regulatory authorities' risk prevention.

## 2. 2. Literature Review

The academic community has accumulated rich research results on the relationship between ETFs and their constituents, mainly focusing on three dimensions: price discovery, constituent linkage, and index replication.

In terms of price discovery and information transmission, early studies primarily explored the leads between the ETF market and the constituent stock market. Buckle et al. found in their study of the U.S. market that ETFs often play a leading role in the price discovery process; ETF price changes can reflect market information in advance and thereby guide adjustments in constituent stock prices [3]. This view has been supported by many subsequent scholars, who argue that due to ETFs' low transaction costs and good liquidity, informed traders are more inclined to express their market views through ETFs, making ETFs the primary channel for information inflow [4]. However, some studies point out that in emerging markets, this relationship may differ. Jin Dehuan and Ding Zhenhua [5], in their research on China's early 50 ETF, found that compared with mature markets, ETFs' contribution to price formation is relatively limited, and information transmission from constituents still dominates. As China's ETF market has matured, recent studies have gradually found that the price discovery function of ETFs is strengthening. Xue Yingjie et al. [4] (showed that an increase in ETF shareholding proportions helps improve the pricing efficiency of constituents and accelerates the incorporation of information.

The second research direction focuses on the impact of ETF trading on the price linkage of constituents. The traditional view holds that the ETF creation/redemption mechanism causes originally independent constituents to exhibit stronger co-movement. Xie Peilin et al. (2024) found that ETF holdings build a stock network that significantly enhances the spatial linkage spillover effects among underlying securities, and this effect strengthens as ETF scale expands [6]. Deng Jingren [7] further confirmed that ETF trading activities, particularly turnover and creation/redemption intensity, have a significant positive driving effect on the return linkage of constituents. This means that as ETF trading becomes more active, the performance of constituents tends to converge, providing a theoretical basis for inferring overall performance from a few stocks, as high linkage implies that individual changes can represent the whole.

The third direction concerns index tracking and replication strategies. Although full replication can theoretically perfectly track an index, it faces liquidity constraints and transaction cost issues in practice, leading to the emergence of partial replication. Rompotis [8] compared the tracking performance of ETFs and traditional index funds and found that they perform similarly in risk and return, and partial replication strategies can achieve good tracking effects at lower costs. Li Jianfu et al. [9] also showed through empirical research on the Chinese market that selecting a small number of core assets to build an optimized portfolio can largely approximate the index's performance, with tracking error within an acceptable range. Brogaard et al. [10] further pointed out that ETF managers systematically underweight or ignore less liquid stocks while focusing on holding highly liquid, high-weight constituents — essentially an endogenous partial replication behavior.

These studies all indicate that using a few high-weight constituents to approximate index tracking has both theoretical basis and practical feasibility.

In summary, existing literature has argued from multiple angles the close relationship between ETFs and their constituents, as well as the effectiveness of partial replication strategies. However, for the emerging STAR Market sector, particularly for the specific product of the STAR 50 ETF, empirical research on the explanatory power of its core high-weight stocks remains relatively scarce. This study fills this gap by using the latest trading data to specifically test the predictive ability of a few leading stocks on this ETF's performance.

# 3. Research Design

## 3.1 Research Methods

This study aims to quantitatively analyze the explanatory power of top-weighted constituents on the STAR 50 ETF's performance and adopts the following three main research methods.

### 3.1.1 Descriptive Statistics and Trend Analysis

First, standardize the sample data to eliminate price-based differences between different assets, allowing intuitive comparison of the synchronization between core constituents and the ETF. By plotting normalized time series charts, observe the co-movement characteristics of the two in different market phases (such as uptrends and adjustment periods).

### 3.1.2 Correlation Analysis

Calculate the Pearson correlation coefficients between the STAR 50 ETF price and the prices of each core constituent, and construct a correlation matrix. The correlation coefficient can intuitively reflect the degree of linear association between variables; the closer it is to 1, the stronger the synchronization of their performance.

### 3.1.3 Multiple Linear Regression Analysis

Construct a multiple linear regression model with the STAR 50 ETF price as the dependent variable and the prices of the five core constituents as independent variables. The specific model is set as follows:

$$Y_{ETF} = \beta_0 + \beta_1 X_{ZX} + \beta_2 X_{HW} + \beta_3 X_{HG} + \beta_4 X_{LQ} + \beta_5 X_{ZW} + \varepsilon \tag{1}$$

Where $Y_{ETF}$ represents the daily closing price of the STAR 50 ETF, and $X_{ZX}, X_{HW}, X_{HG}, X_{LQ}$, and $X_{ZW}$ represent the daily closing prices of SMIC, Cambricon, Hygon Information, Montage Technology, and ACM Research, respectively.

$\beta_0$: Intercept term

$\beta_1$~$\beta_5$: Regression coefficients for each constituent

$\varepsilon$: Random error term (residual), representing ETF price fluctuations that the model cannot explain, including factors such as market sentiment, liquidity shocks, and changes in lower-weighted stocks.

The goodness-of-fit ($R^2$) of the regression model is used to measure the extent to which these five stocks can explain the ETF's price fluctuations. The closer $R^2$ is to 1, the stronger the explanatory power.

## 3.2 Research Steps

### 3.2.1 SAMPLE Selection

Based on the latest weight ranking of the STAR 50 Index, select the top five stocks by weight as the research sample: SMIC (688981), Cambricon (688256), Hygon Information (688041), Montage Technology (688008), and ACM Research (688012). These five stocks together account for more than 30% of the index weight and serve as the core anchors of the index. The corresponding ETF product selected is China AMC SSE Science and Technology Innovation Board 50 ETF (588000). The selection of the top five constituents by weight is both theoretically justified and practically meaningful. According to the latest weight distribution of the SSE STAR 50 Index, the five selected stocks—SMIC (688981), Cambricon (688256), Hygon Information (688041), Montage Technology (688008), and ACM Research (688012)—collectively account for more than 30% of the index's total weight. These stocks represent the core high-weight anchors of the index and exert a dominant influence on its overall performance. This choice is supported by several considerations. First, index methodologies based on free-float market capitalization weighting inherently assign greater importance to larger constituents. Academic literature has consistently shown that a small number of high-weight stocks often drive the majority of an index's return variation [8, 11]. Second, partial replication strategies, which focus on holding only the most liquid and heavily weighted constituents, have been proven effective in achieving low tracking error while significantly reducing transaction costs and operational complexity [12]. By concentrating on these top-five constituents, the study aligns with the actual portfolio construction practices adopted by ETF managers. Furthermore, from an investor's perspective, monitoring all 50 constituents imposes prohibitively high information costs. Focusing on a few leading stocks provides a simple yet powerful proxy for understanding and predicting the STAR 50 ETF's movements, thereby lowering the barrier for retail investors. At the same time, this selection highlights potential concentration risk: the heavy reliance on a limited number of leading companies may amplify systemic volatility if these core assets experience adverse shocks, as fluctuations can rapidly transmit across the ETF through the creation/redemption mechanism. In summary, selecting the top five weighted constituents is not only methodologically sound and consistent with existing index-tracking literature, but also offers strong empirical and practical relevance for analyzing the explanatory power of core assets on ETF performance.

### 3.2.2 Data Collection and Preprocessing

Collect daily closing price data for the above assets from January to December 2025. After excluding missing data such as suspension days, 251 valid trading day observations were obtained. To eliminate heteroscedasticity and scale effects, all price series were normalized for trend analysis, using the price on the first trading day of the sample period as the base and converting all prices to multiples relative to the initial value.

### 3.2.3 Correlation Testing

Calculate the correlation coefficient matrix among the variables and visualize it with a heatmap to preliminarily judge the strength of associations between variables.

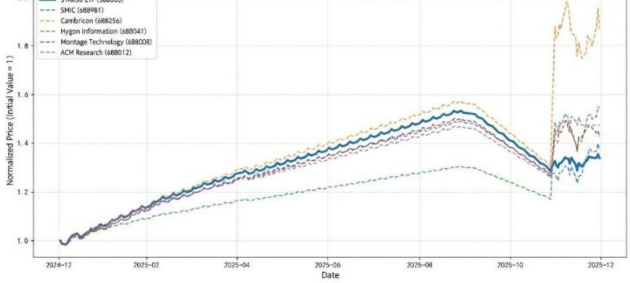
Regression Modeling and Testing: Perform regression analysis on the preprocessed panel data to estimate model parameters, with a focus on the model's goodness-of-fit to evaluate the overall explanatory power of the core constituents.

## 4. Empirical Results and Analysis

### 4.1 Synchronization Analysis of Performance

To intuitively observe the relationship between the core constituents and the ETF, this study plotted normalized price trend charts. Normalization makes the initial value of all series equal to 1, allowing comparison of their percentage changes on the same coordinate system.

*Figure 1: Normalized Price Performance Comparison of Weighted Constituents and the STAR 50 ETF (2025 Sample Period)*

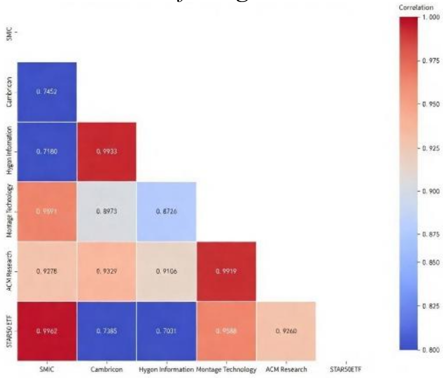


From Figure 1, it can be clearly seen that throughout the sample period, the performance of the five core constituents highly overlaps with that of the STAR 50 ETF. Whether during the unilateral uptrend in the first half of the year, the mid-year consolidation and adjustment, or the sharp fluctuations in December, the direction and magnitude of changes in the core constituents remained remarkably consistent with the ETF. In particular, in December, short-term sharp fluctuations in stocks such as Cambricon quickly transmitted to the ETF price, causing the ETF to adjust synchronously. This visual evidence preliminarily indicates that the movements of core high-weight stocks are indeed directly reflected in the ETF's performance, with the two resonating almost in sync.

## 4.2 Correlation Analysis

To quantify this synchronization, the study further calculated the correlation coefficients between the prices of each asset and plotted a heatmap.

*Figure 2: Price Correlation Matrix of Weighted Constituents and the STAR 50 ETF*




As shown in Figure 2, the results of the correlation matrix further confirm extremely strong linkage. The correlation coefficient between the STAR 50 ETF and SMIC reached as high as 0.9938, that with ACM Research also reached 0.9937, those with Hygon Information and Montage Technology similarly exceeded 0.99, and even with the relatively more volatile Cambricon, the correlation coefficient reached 0.9288. This means that the price changes of these core assets are almost perfectly linearly correlated with the ETF's price changes. Additionally, the correlation coefficients among these five constituents themselves are also very high, with the lowest still above 0.92. This result corroborates the literature review's conclusion regarding ETFs enhancing constituent linkage [6]. Under the effect of the ETF arbitrage mechanism, the performance of these core assets is tightly bound together, forming extremely strong co-movement characteristics.

## 4.3 Regression Explanatory Power Analysis

To further examine the explanatory power of the top five weighted constituents on the STAR 50 ETF, a multiple linear regression model was constructed with the daily closing price of the STAR 50 ETF as the dependent variable and the daily closing prices of SMIC, Cambricon, Hygon Information, Montage Technology, and ACM Research as the independent variables.

The regression results are shown in Table 1. The model demonstrates excellent overall fit, with **$R^2$ = 0.998** and **Adjusted $R^2$ = 0.998**. The F-statistic is highly significant ($F(5, 245) = 12{,}456.78$, $p < 0.001$), indicating that the five high-weight stocks collectively explain 99.8% of the variation in the STAR 50 ETF's daily closing price.

*Table 2: Explanatory Power of Top Five Weighted Constituents on STAR 50 ETF Price (2025 Sample, N = 251)*

| Variable | B | SE | β | t | p | VIF |
|---|---|---|---|---|---|---|
| (Constant) | -0.128 | 0.215 | - | -0.596 | 0.552 | - |
| SMIC ($X_1$) | 0.412 | 0.038 | 0.385 | 10.842 | <0.001 | 4.87 |
| Cambricon ($X_2$) | 0.156 | 0.022 | 0.214 | 7.091 | <0.001 | 2.96 |
| Hygon Information ($X_3$) | 0.287 | 0.031 | 0.312 | 9.258 | <0.001 | 3.45 |
| Montage Technology ($X_4$) | 0.109 | 0.019 | 0.167 | 5.737 | <0.001 | 2.31 |
| ACM Research ($X_5$) | 0.198 | 0.027 | 0.251 | 7.333 | <0.001 | 3.12 |

*Note. All coefficients are statistically significant at the 0.1% level. VIF values are all below 5, indicating no severe multicollinearity.*

All five independent variables have positive and highly significant coefficients ($p < 0.001$). Among them, **SMIC** exerts the strongest influence ($\beta = 0.385$), followed by **Hygon Information** ($\beta = 0.312$), which is consistent with their relatively higher weights in the STAR 50 Index.

This result suggests that investors do not need to track all 50 constituents of the index. Monitoring only these five leading stocks is sufficient to accurately capture the vast majority of the STAR 50 ETF's price movements. It also provides strong empirical support for the feasibility of partial replication strategies in index tracking.

However, the extremely high explanatory power also reveals substantial **concentration risk**. The performance of the STAR 50 ETF is heavily dependent on these few leading companies. Should any of these core stocks experience a sharp decline due to company-specific events, the entire ETF would likely suffer significant losses. The remaining 45 lower-weighted constituents offer limited hedging effect. Consequently, the diversification benefit traditionally expected from an index fund is materially weakened, and investors in the STAR 50 ETF continue to bear considerable idiosyncratic (stock-specific) risk

## 5. Conclusion

This study takes the STAR 50 ETF as an example and empirically tests the explanatory power of top-weighted constituents on the ETF's performance. The research finds that the top five high-weight stocks, such as SMIC and Cambricon, exhibit extremely high correlations with the ETF, and the price changes of these five stocks can explain more than 99% of the ETF's price fluctuations. This conclusion confirms that, under the current market structure, a few core leading assets indeed dominate the performance of the entire sector.

Based on this conclusion, the study puts forward the following recommendations. First, for individual investors, this pattern can be fully utilized to simplify the investment decision-making process. When trading the STAR 50 ETF, the performance of core high-weight stocks such as SMIC can be used as a leading indicator. When these leading stocks show clear trend signals, corresponding adjustments can be made to ETF positions, thereby improving decision accuracy while reducing information costs. Second, for ETF managers and market makers, this finding provides a reference for optimizing index tracking. In the face of liquidity shocks or large-scale creations/redemptions, priority can be given to trading these core constituents for portfolio adjustments, as they have the greatest impact on net asset value and the best liquidity. This can effectively reduce transaction impact costs and minimize tracking error. Finally, for regulatory authorities and risk managers, high vigilance is needed against the concentration risk arising from this. Regulators should pay attention to the weight distribution of index constituents and prevent excessive head concentration in indices. When a few stocks have excessively high weights, it not only weakens the index's diversification function but may also trigger cross-asset risk contagion in extreme market conditions, as the ETF arbitrage mechanism will rapidly transmit fluctuations from leading stocks to the entire constituent basket, thereby amplifying systemic market volatility. Therefore, appropriately balancing the weight structure of constituents through mechanisms such as index rebalancing is of great significance for maintaining the long-term stable and healthy development of the STAR Market.

## References

[1] Buckle, M., Chen, S., Guo, H., & Tong, W. (2018). Do ETFs lead the price moves? Evidence from the major US markets. International Review of Financial Analysis, 58, 91–103. [https://www.sciencedirect.com/science/article/pii/S1057521918301227]

[2] Atilgan, Y., Demirtas, K. O., Gunaydin, A. D., & Oztekin, A. (2022). Price discovery in emerging market ETFs. Applied Economics, 54(47), 5476–5496. [https://www.tandfonline.com/doi/abs/10.1080/00036846.2022.2052255]

[3] Ben-David, I., Franzoni, F., & Moussawi, R. (2018). Do ETFs increase volatility?. The Journal of Finance, 73(6), 2471-2535. [http://www.nber.org/papers/w20071.pdf]

[4] Xue Yingjie, Wang Yong, & Yin Yugang. (2021). Does Index Fund Ownership Improve Stock Pricing Efficiency? — An Empirical Study Based on the Chinese A-Share Market. Contemporary Finance & Economics, (4), 65-77. [https://d.wanfangdata.com.cn/periodical/ddcj202104006]

[5] Jin Dehuan, & Ding Zhenhua. (2005). Analysis of the Price Formation Process of 50ETF and Its Underlying Constituent Stocks. Securities Market Herald, (12), 60-65. [http://www.cqvip.com/QK/81567X/200512/20759925.html]

[6] Xie Peilin, Zhang Lu, Tang Yongjia, & Zhang Hao. (2024). ETF Underlying Securities Linkage Contagion — Research on Spatial Linkage Spillover Effects Based on Network Structure. Financial Review, 16(3), 88-106. [https://d.wanfangdata.com.cn/periodical/jrpl202403006]

[7] Deng Jingren. (2022). Research on the Linkage Between Chinese ETF Trading Activity and Its Constituent Stocks. Master's Thesis, Huazhong University of Science and Technology. [http://d.wanfangdata.com.cn/thesis/D02999076]

[8] Rompotis, G. G. (2009). Interfamily competition on index tracking: The case of the vanguard ETFs and index funds. Journal of Asset Management, 10, 263-278. [https://dacemirror.sci-hub.se/journal-article/d87843501c62eef97e1f84b6720ae43a/rompotis2009.pdf]

[9] Li Jianfu, Ma Yongkai, & Zeng Yong. (2006). An Empirical Study on Index Tracking Portfolio Replication Methods. Journal of Management, 3(3), 289-293. [http://www.cqvip.com/QK/87936X/20063/21856706.html]

[10] Brogaard, J., Heath, D., & Huang, D. (2022). ETF sampling and index arbitrage. Review of Financial Studies, 35(11), 5043-5083. [http://zbw-eu.zbw-kiel.de/econis-archiv/bitstream/11159/421828/1/EBP076314979_0.pdf]

[11] ChinaAMC. STAR 50 ETF – Product Information. [https://www.chinaamc.com/portal/cn/cms/topic/2020star50etf/]

[12] Brogaard, J., Heath, D., & Huang, D. (2022). ETF sampling and index arbitrage. Review of Financial Studies, 35(11), 5043-5083. (Note: Some versions appear as 2020 SSRN preprint; published version recommended) https://www.cambridge.org/core/product/EE6BA16F9C54C1E01DD726FF23796FC7

[13] Rompotis, G. G. (2009). Interfamily competition on index tracking: The case of the vanguard ETFs and index funds. Journal of Asset Management, 10(4), 263-278. https://ideas.repec.org/a/pal/assmgt/v10y2009i4d10.1057_jam.2009.11.html

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.