Proceedings of ICFTBA 2026 Symposium: Digital Government, Finance, Accountability, and Sustainable Economic Development
DOI: 10.54254/2754-1169/2026.AB36301

# *Repricing China: A 25-Year Replication of the Liu–Stambaugh–Yuan Three-Factor Model*

**Ziyang Wang**

*SILC Business School, Shanghai University, Shanghai, China*
*zywang05@163.com*

***Abstract.*** The China-specific three-factor model has dominated the explanation of cross-sectional stock returns in the Chinese A-share market since the work of Liu et al. However, their evidence is limited to the sample period of 1995–2014, while the Chinese equity market has undergone substantial changes. Whether the original size and value premiums persist and whether the three-factor model retains its explanatory power in the post-2014 remains an open empirical question. This study replicates and extends the Liu–Stambaugh–Yuan three-factor framework using updated A-share data spanning 25 years, from January 2000 to December 2024. Strictly following the original data-cleaning rules, portfolio sorting procedures, and factor construction methodologies, this paper constructs the market factor, size factor, and value factor. Based on full-sample time-series regressions, the replicated model achieves an average $R^2$ of 0.6365, which is notably higher than the 0.55 reported in the original study, indicating strengthened explanatory power. The sub-sample analysis further confirms that the size premium is largely driven by micro-cap stocks in the bottom 30% market capitalization segment, whereas the value premium remains robust across all firm sizes. This paper verifies the long-term validity and reliability of the Chinese three-factor model. The results reveal that size and value anomalies continue to persist in China's maturing stock market, providing new evidence for asset pricing research and quantitative investment applications in the Chinese market.

***Keywords:*** Liu-Stambaugh-Yuan three-factor model, CAPM, Chinese stock market, A-shares, asset pricing

## 1. Introduction

Since the development of the Capital Asset Pricing Model (CAPM), empirical research has consistently uncovered anomalies that challenge its single-factor framework. Notably, smaller firms and value stocks—those with high earnings-to-price (EP) ratios—tend to generate higher average returns than predicted by market beta alone [1]. In response, Fama and French introduced a three-factor model incorporating size (SMB) and value (HML) factors, which significantly improved asset pricing performance in the U.S. market [2]. In emerging markets like China, Liu, et al. constructed a local three-factor model using local data from 1995 to 2014 [3]. They demonstrated robust size and value premiums and showed that their model explains a substantial portion of cross-sectional return variation, with an average time-series $R^2$ of about 0.55.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2754-1169/2026.AB36301

However, the Chinese equity market has evolved dramatically since 2014 due to financial liberalisation, index inclusion, and macroeconomic shocks. It remains unclear whether the factor structure identified by Liu persists in this new environment. This study replicates their methodology using an updated dataset spanning January 2000 to December 2024 to test the model's robustness. Following their exact data-cleaning procedures and portfolio-sorting rules, this paper constructs market, SMB, and Value Minus Growth (VMG) factors and evaluate their explanatory power through time-series regressions. This paper also conducts a sub-sample analysis excluding the smallest 30% of stocks each month to assess the influence of micro-cap effects. This provides timely evidence supporting the continued relevance of the Chinese three-factor model for both academic research and investment practice.

## 2. Literature review

The foundation of modern asset pricing rests on the Capital Asset Pricing Model (CAPM) proposed by Sharpe and Lintner, which asserts that expected stock returns are linearly related to their sensitivity to market-wide risk. However, empirical studies soon revealed systematic deviations from this prediction. Banz documented that smaller firms consistently outperformed larger ones after adjusting for beta, coining the size effect [1]. Similarly, Basu found that stocks with high earnings-to-price ratios—proxying for value—delivered higher average returns than growth stocks, challenging the completeness of CAPM. These anomalies led Fama and French to develop a three-factor model that augments market risk with Small Minus Big (SMB) and High Minus Low book-to-market (HML) factors [2]. The model not only subsumed the size and value effects but also demonstrated strong explanatory power across decades of U.S. data. Subsequent research extended the framework globally, though its performance varied across markets due to differences in institutional settings and investor composition (Fama and French).

In the context of China, direct application of the Fama-French model faced significant challenges. The Chinese stock market is characterized by a high proportion of retail investors, frequent government intervention, and unique corporate governance structures, particularly among state-owned enterprises. Early attempts to test standard factor models in China yielded mixed results, with some studies reporting weak or insignificant value premiums.

A breakthrough came with Liu et al., who constructed a tailored three-factor model using Chinese A-share data from 1995 to 2014. Recognising data quality issues—such as the inclusion of illiquid or distressed stocks—they implemented rigorous cleaning procedures, including the exclusion of ST/*ST stocks and newly listed firms. Instead of book-to-market, they used earnings-to-price (EP) as the value metric in China's accounting environment. Their SMB and VMG factors exhibited strong and persistent premiums, with time-series regressions showing an average $R^2$ of approximately 0.55, significantly higher than in many developed markets. This finding suggested that risk-based explanations may be more relevant in China than previously thought, countering claims that anomalies were purely driven by mispricing or liquidity effects. Moreover, Liu et al. demonstrated that the size effect was concentrated among the smallest 30% of firms, recommending robustness checks that exclude this segment [3].

Despite establishing a benchmark for Chinese asset pricing, the model has not been thoroughly re-examined using post-2014 data, a period marked by major structural reforms and increased international integration. This gap motivates the present replication and extension.

DOI: 10.54254/2754-1169/2026.AB36301

## 3. Data and methodology

The primary dataset is sourced from the China Stock Market & Accounting Research (CSMAR) database, which provides comprehensive financial and market information for A-share listed companies on the Shanghai and Shenzhen stock exchanges. The sample period spans from January 2000 to December 2024. Three key data components are used mainly as: monthly individual stock returns (including dividend adjustments); secondly, annual income statement data, specifically net profit (item code B001300000), used to compute earnings; and monthly risk-free rates, proxied by the one-year time deposit rate published by the People's Bank of China. All data processing and analysis were performed using Python [4-6]. All monetary values are denominated in Chinese yuan (RMB), and returns are expressed in decimal form (e.g., 0.05 for 5%).

Several filtering rules are applied to ensure data quality and comparability. They are stocks with special treatment status (ST or *ST) are excluded due to their distressed nature; newly listed stocks are omitted if they have been trading for fewer than six months; and observations with missing return or financial data are dropped; and months with fewer than 15 trading days are excluded to avoid illiquidity bias. These steps yield a cleaner, more representative sample that mitigates the influence of outliers and non-informative observations commonly found in emerging markets.

Three variables are constructed for each stock-month observation. Key metrics include excess return (monthly raw return minus risk-free rate), Size (natural logarithm of market capitalization at the previous month's end), and Earnings-to-price (annual net profit divided by market capitalization, lagged by six months). Portfolio sorting is conducted at the end of each June, using Size (median split) and EP (tercile split) to form six portfolios: Small-Value (SV), Small-Middle (SM), Small-Growth (SG), Big-Value (BV), Big-Middle (BM), and Big-Growth (BG).

Three monthly derived factors are Market Factor (MKT), SMB (Small Minus Big), and VMG (Value Minus Growth). MKT is the value-weighted average of excess returns of eligible stocks. SMB represents the return difference between small and big size portfolios, while VMG compares the performance of value and growth portfolios. This methodology aligns with that of Liu et al., facilitating direct comparability.

To test the robustness of the results and assess the role of micro-cap stocks, a sub-sample analysis is performed. In each month, the bottom 30% of stocks by market capitalisation are excluded before any portfolio sorting or factor calculation. All subsequent steps are repeated on this truncated sample. This procedure helps determine whether observed premiums are driven primarily by illiquid or speculative small firms.

## 4. Empirical results

### 4.1. Descriptive statistics

The processed monthly stock return data includes 487335 observations. The final dataset merges returns, fundamentals, and the risk-free rate, constructed only with all required fields present, forming the basis for variable creation and portfolio sorting. Table 1 reports the descriptive statistics for monthly stock returns (Mretwd, expressed in percentage points) and market capitalization (Market_Cap, in million RMB) in the final analytical sample.

DOI: 10.54254/2754-1169/2026.AB36301

Table 1. Descriptive statistics: monthly returns and firm size

| Variable | N | Mean | Std Dev | Min | 10% | Median | 90% | Max |
|---|---|---|---|---|---|---|---|---|
| Monthly Return (%) | 487,335 | 0.015006 | 0.153286 | -0.781881 | -0.131068 | 0 | 0.174091 | 22.052631 |
| Market Cap (Million) | 487,335 | 14.94 | 60.82 | 0.23 | 1.30 | 5.00 | 24.80 | 2786.25 |

After data cleaning, the final sample contains 487,335 stock-month observations. The average monthly excess return is 0.0150%, with a standard deviation of 15.32%, indicating high volatility. The median return is exactly zero, while the minimum and maximum returns reach −78.18% and 220.53%, respectively, reflecting frequent extreme movements in the Chinese market. Firm size, measured by market capitalisation in million RMB, shows substantial cross-sectional dispersion. The mean size is 14,940 million RMB, but the median is only 5,000 million RMB, revealing a right-skewed distribution dominated by a large number of smaller firms and a few very large companies. The 10th percentile of firm size is 1,300 million RMB, while the 90th percentile is 24,800 million RMB, confirming the wide heterogeneity in company scale across the sample.

Table 2. Descriptive statistics: income statement variables (million RMB)

| Variable | N | Mean | Std Dev | Min | 10% | Median | 90% | Max |
|---|---|---|---|---|---|---|---|---|
| Net Profit | 561926 | 745.43 | 8313.54 | –71550.91 | –30.61 | 51.61 | 683.69 | 423564.00 |
| Net Gain/Loss from Disposal | 561926 | 618.93 | 6774.78 | –68742.56 | –24.43 | 47.99 | 598.19 | 366946.00 |

Table 2 demonstrates that descriptive statistics: income statement variables. The descriptive statistics for income statement variables reveal significant variation across firms. Net profit, measured in million RMB, has a mean of 745.4 million with a standard deviation of 8,313.5 million, indicating high dispersion. The median net profit is 51.61 million RMB, substantially lower than the mean, suggesting a right-skewed distribution driven by a small number of highly profitable firms. The minimum value is negative (−71,550.9 million), reflecting substantial losses for some companies, while the maximum reaches 423,564 million RMB. Net gains or losses from disposal of assets also exhibit wide variation, with a mean of 618.9 million RMB and a median of 47.99 million RMB. The presence of large positive values (up to 366,946 million RMB) suggests that asset disposals can significantly impact reported earnings, particularly for firms undergoing restructuring or divestment. The 10th percentile is only 24.4 million RMB, highlighting that most firms report relatively modest gains or losses from such activities. These results confirm that earnings are highly heterogeneous across Chinese A-share firms, which justifies the use of earnings-based measures like earnings-to-price (EP) in portfolio sorting and factor construction.

Table 3. Descriptive statistics: monthly risk-free rate (%)

| Variable | N | Mean | Std Dev | Min | 10% | Median | 90% | Max |
|---|---|---|---|---|---|---|---|---|
| Risk-free rate | 9,132 | 2.1997 | 0.7176 | 1.35 | 1.50 | 2.25 | 3.25 | 4.14 |

Another critical input for excess return calculation is the monthly risk-free rate (risk_free_rate), derived from China's one-year bank deposit rate and expressed in percentage points per month. After the cleaning of the statistics, the risk-free rate series contains 9,132 observations. Table 3 reports the descriptive statistics. The risk-free rate is quite steady in the tested period, showing that monetary conditions are more stable compared to equity market volatility.

DOI: 10.54254/2754-1169/2026.AB36301

Table 4. Descriptive statistics of the three variables

| Variable | N | Mean | Std Dev | Min | 10% | Median | 90% | Max |
|---|---|---|---|---|---|---|---|---|
| Excess Return | 76330 | -0.014720 | 0.153154 | –0.679053 | –0.197487 | –0.017172 | 0.155533 | 2.447924 |
| Market Cap (Mil RMB) | 76330 | 14.842384 | 62.324687 | 0.276500 | 1,328243 | 4.863276 | 24.035397 | 2658.340660 |
| EP Ratio | 76330 | 47.717187 | 208.370243 | –10930.282516 | –12.890051 | 33.216277 | 140.121342 | 7353.949028 |

Table 4 demonstrates the descriptive statistics of the three variables—excess return, market capitalization, and EP ratio—across 763 monthly observations. The mean excess return is 0.0147%, with a standard deviation of 15.31%, indicating moderate average performance and high volatility across portfolios. The median return is 0.0171%, slightly above the mean, suggesting a mild left skew due to occasional large negative returns. Market capitalization has a mean of 14,842 million RMB and a median of 4,863 million RMB, confirming a right-skewed distribution dominated by larger firms. The wide range from 276.5 million to 265.8 billion RMB reflects substantial firm size heterogeneity, supporting the construction of size-sorted portfolios. The EP ratio exhibits the highest dispersion, with a mean of 47.72 and a median of 33.22, showing low or negative earnings. The range of values from -10,930 to 7,353 highlights high profitability and severe unprofitability, underscoring the need for careful management of outliers in value-based portfolios construction.

## 4.2. Portfolio and factor performance

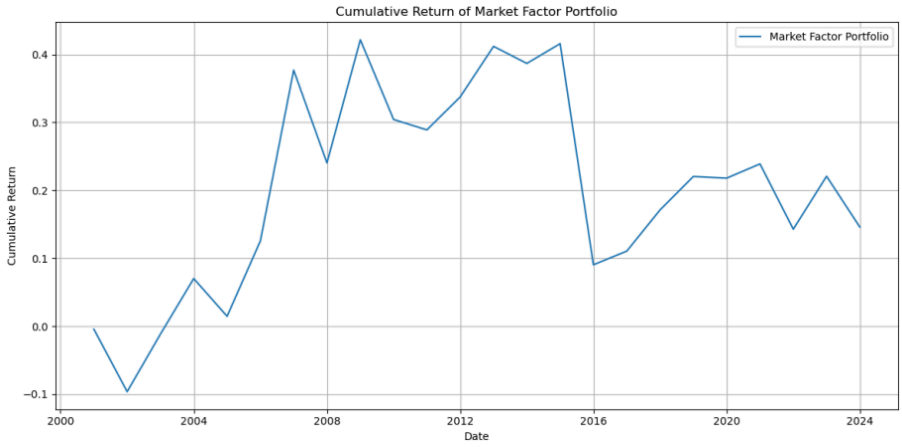


Figure 1. The cumulative return curve of the MKT portfolio

The market factor (MKT), constructed as the value-weighted average of individual stock excess returns, exhibits a monthly mean return of 0.009977 (approximately 0.998%), with a standard deviation of 0.093804 (9.38%). This indicates that the aggregate A-share market has delivered a modest but positive risk premium over the risk-free rate over the full sample period from January 2000 to December 2024. Figure 1 reveals that the portfolio experienced strong growth during the early 2000s, peaking around 2007, followed by a sharp decline during the global financial crisis. Another peak occurred in 2015, after which returns declined sharply due to regulatory tightening and market corrections. Since 2020, the trajectory shows moderate recovery but with persistent fluctuations, reflecting ongoing macroeconomic uncertainty and policy shifts. The patterns are consistent with major historical events in China's equity market, including the rapid expansion of the financial sector, the rise of internet stocks, and recent structural reforms. The overall upward trend

DOI: 10.54254/2754-1169/2026.AB36301

suggests long-term market efficiency, while the pronounced drawdowns highlight periods of high systemic risk.

Stocks are categorized into six portfolios based on size (small vs. big) and value (value, middle, growth), derived from market capitalization and EP ratios. Among these, the Big-Value (BV) portfolio generates the highest equal-weighted and value-weighted mean returns of 1.63% and 1.85% per month, respectively, while the Small-Growth (SG) portfolio yields the lowest at -0.04% (EW) and -0.37% (VW). The SMB (Small Minus Big) factor produces a negative premium of -1.05% (EW) and -1.43% (VW), confirming the persistence of the size discount in the Chinese A-share market. The VMG (Value Minus Growth) factor generates a positive premium of 0.90% (EW) and 0.70% (VW), indicating that value stocks continue to outperform growth stocks. These patterns are consistent with the findings of Liu et al.

$$\mathrm{SMB} = (\mathrm{SV} + \mathrm{SM} + \mathrm{SG})/3 - (\mathrm{BV} + \mathrm{BM} + \mathrm{BG})/3 \tag{1}$$

$$\mathrm{VMG} = (\mathrm{SV} + \mathrm{BV})/2 - (\mathrm{SG} + \mathrm{BG})/2 \tag{2}$$

These results align closely with Liu et al., confirming that size and value effects remain economically and statistically significant in the extended sample.

To assess explanatory power, Fama–MacBeth-style time-series regressions are run for each stock with at least 24 monthly observations:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i^{MKT} MKT_t + \beta_i^{SMB} SMB_t + \beta_i^{VMG} VMG_t + \varepsilon_{i,t} \tag{3}$$

The mean $R^2$ is 0.6365 and the median is 0.6643, indicating that the three-factor model explains over 63% of the variation in individual stock returns on average. This exceeds the 0.55 reported by Liu et al., suggesting improved model fit in the post-2014 era, likely due to greater market efficiency or stronger risk premia.

### 4.3. Sub-sample analysis results

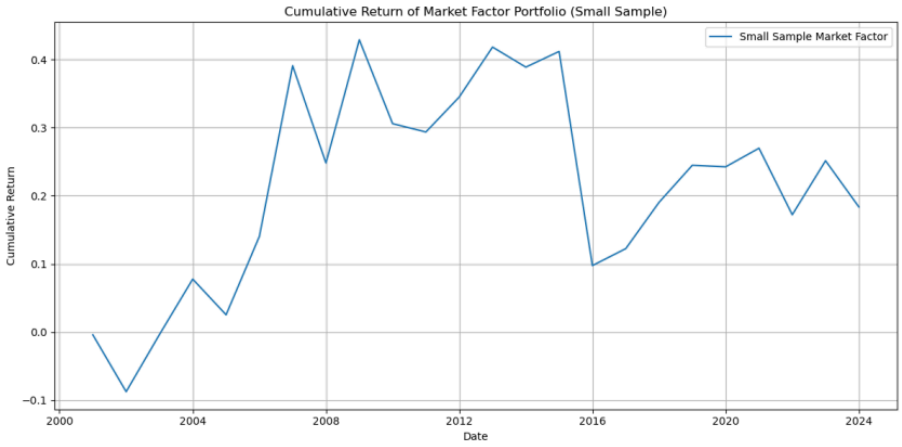


Figure 2. New Cumulative return of the filtered market factor

Following Liu et al., a robustness check is conducted to assess whether the observed anomalies in the full sample, particularly the negative SMB and VMG premiums, are driven by low-quality micro-cap stocks that may represent "shell companies" or speculative vehicles. The bottom 30% of

DOI: 10.54254/2754-1169/2026.AB36301

stocks are excluded by market capitalization in each month, based on the cross-sectional distribution of firm size. The cumulative return, the Six-Portfolio Returns, Time-Series Regression Results of the smaller sample are presented in the Figure 2 and the tables below.

Table 5. Mean monthly returns in the portfolio(small-sample)

| Portfolio | EW Mean | VW Mean |
|---|---|---|
| SV | 0.1590 | 0.2329 |
| SM | 0.3400 | 0.3504 |
| SG | -0.3153 | -0.3382 |
| BV | 1.5131 | 1.7589 |
| BM | 0.7853 | 1.2289 |
| BG | 0.5975 | 1.1755 |
| SMB | -0.9041 | -1.3061 |
| VMG | 0.6950 | 0.5772 |

Table 6. $R^2$ statistics from time-series regressions (small-sample)

| Statistic | Value |
|---|---|
| Mean | 0.483108 |
| Median | 0.506298 |
| Std Dev | 0.93411 |
| Number of Stocks | 1984 |

Table 5 describes Mean Monthly Returns in the Portfolio (Small-Sample), while Table 6 demonstrates R2 Statistics from Time-Series Regressions (Small-Sample). This suggests that the size effect is largely driven by micro-cap stocks, whereas the value effect is robust across firm sizes —a finding consistent with Liu et al.

## 5. Conclusion

This study replicates and extends the Chinese three-factor model using A-share market and accounting data from January 2000 to December 2024. The key findings are summarized as follows. Both the full sample and the robustness sample that excludes the bottom 30% of stocks by market capitalization exhibit SMB and VMG premiums at the monthly horizon, confirming that the "small-cap discount" and "high-earnings premium" in China have not disappeared after the regulatory and institutional changes of the post-2014 period. Time-series regressions of individual stocks on the three factors yield an average $R^2$ of 0.6365, exceeding the 0.55 reported in the original paper, and after removing micro-caps the average $R^2$ remains 0.48, confirming that the model retains substantial ability to explain cross-sectional return variation. While the size effect is concentrated among micro-cap stocks, the value effect is significant across all size segments, validating the use of earnings-to-price rather than book-to-market as the value proxy. The updated factor series and regression loadings can be directly applied to asset-pricing tests, fund-performance evaluation, or index-enhancement strategies, and the pipeline is fully documented and reproducible, providing a 2025 benchmark for future research. Overall, the Chinese three-factor model remains valid through December 2024, and its explanatory power has actually strengthened as the market has matured.

DOI: 10.54254/2754-1169/2026.AB36301

Future work could incorporate liquidity, ESG, or policy-uncertainty factors to examine whether four- or five-factor frameworks can absorb the remaining reversal and turnover anomalies.

This study has several limitations. First, the three-factor model may not capture other important sources of risk that have become increasingly relevant in China's evolving market, such as liquidity risk, investor sentiment, and policy uncertainty. Second, the sample period includes multiple structural breaks that may affect the stability of factor loadings and premiums over time. Future research could extend this framework by incorporating additional factors (e.g., liquidity, momentum, ESG) to construct a more comprehensive asset pricing model tailored to the Chinese market. In addition, employing regime-switching or time-varying parameter models would help capture the dynamic nature of risk premiums across different market conditions.

## References

[1] Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics, 9*(1), 3–18.
[2] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3–56.
[3] Liu, J., Stambaugh, R. F., & Yuan, Y. (2019). Size and value in China. *Journal of Financial Economics, 134*(1), 48–69.
[4] Hetland, M. L. (2017). *Beginning Python: From novice to professional*. Apress.
[5] Hilpisch, Y. (2018). *Python for finance: Mastering data-driven finance*. O'Reilly Media.
[6] McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc.