Proceedings of the 5th International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2026.38679

Why Can Positive Average Returns Coexist with Negative Compound Growth? The Arithmetic-Geometric Mean Gap in China's A-Share Market

Isac Chen

Sedbergh School, Fuzhou, China
chenisac0203@gmail.com

Abstract. A positive arithmetic mean return does not imply positive growth of wealth over multiple investment periods. This study employs monthly individual-stock returns of 4,647 A-shares listed on the Shenzhen and Shanghai Stock Exchanges in China from 2015 to 2024, and analyzes the difference between the stock-level arithmetic and geometric mean returns. The arithmetic mean return has a median of 0.78%, while the geometric mean return has a median of -0.13%. For 1,696 stocks, the arithmetic mean return is positive but the geometric mean return is negative (36.50% of the main sample; 45.79% of stocks with positive arithmetic mean returns). There is a significant and positive, but non-causal, relationship between return variance and the arithmetic-geometric mean gap. In a pooled winsorisation sensitivity check at the 1st and 99th percentiles, the variance coefficient rises from 0.147 to 0.462, consistent with the 0.5 coefficient in the small-return second-order approximation. This study demonstrates that average return should be accompanied by compound growth, volatility, and tail risk in long-horizon investment evaluation. But the evidence is not conclusive as to whether the Kelly allocation is optimal for all investors.

Keywords: arithmetic mean return, geometric mean return, compound growth, Kelly criterion

1. Introduction

In fact, the arithmetic mean returns are frequently employed to describe the average performance of an asset in investment reports, and the mean-variance framework is thought to be a standard measure for investment portfolio selection. However, over time, wealth grows multiplicatively, not additively: a large loss reduces the amount of wealth available for future growth, and equal percentage gains and losses do not cancel out. For instance, if an investor loses 50 percent of wealth and then gains 50 percent, the arithmetic average return is zero, but terminal wealth is only 75 percent of the initial amount. A positive single-period return does not imply positive compound growth over the long term.

The allocation problem was solved from the viewpoint of information theory, with maximisation of the expected logarithmic growth rate of wealth proposed as a criterion for repeated betting. Log growth embeds multi-period compounding and ruin risk into a common framework of analysis [1].

© 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: 10.54254/2754-1169/2026.38679

The growth-security trade-off was further treated as a dynamic investment problem [2]. Related work proved log-growth objectives are analysable for arbitrary bounded market sequences [3]. Bayesian betting policies were derived for scenarios with unobserved process parameters, alongside quantification of the financial value and cost of learning [4]. Modern extensions integrate risk measures, distributional uncertainty and real-world information quality into the growth-optimal framework. Weighted Value-at-Risk has been incorporated into dynamic risk management for log returns [5]. Distributionally robust optimisation frameworks for wealth growth have been constructed [6]. Empirical evidence confirms investment performance is subject to information quality [7]. Discussions revisit the relationship between geometric-mean criteria and investor preferences [8]. Analytical solutions exist for multiple-outcome settings [9]. Structural properties of the classic capital-growth model have been examined under discrete adjustment [10].

Although these studies give an insight into why cross-sectional expectations cannot replace individual investment paths in the long run, the present analysis does not disregard the relevance of utility and constraints in investment decisions. This leads to a simpler empirical question: how often is it that the arithmetic average return of a stock is not equal to the compounded growth rate? This paper offers an answer to this question with monthly returns of Chinese A-shares. It's not a new asset-pricing law; it's simply a measure of this gap. First, it takes the positive-expectation issue from a theoretical binary bet to a 10-year history of individual stocks in a range of market segments to capture the magnitude and economic scale of the mismatch. Second, the principal analysis keeps extreme returns, and a winsorized specification is used to explore deviations from the second-order approximation. Third, the paper draws a clear distinction between the empirical usefulness of a growth criterion and optimality under an investor utility criterion. The analysis therefore does not seek to compare investment methods, nor does it aim to tell investors which investment is best or worst, thereby minimising the danger of simply converting descriptive statistics into investment advice.

2. Theory and data

Let r_{it} be the simple return of stock i during a valid month t and let T_i be the number of months for which stock i was observed. The arithmetic mean return A_i and the geometric mean return G_i are given in Model (1) and Model (2) below:

$$A_i = \frac{1}{T_i} \sum_t r_{it} \tag{1}$$

$$G_i = \exp[\frac{1}{T_i} \sum_t \ln(1 + r_{it})] - 1 \tag{2}$$

A_i is the average single period return while G_i is the fixed growth rate per valid observation that can generate cumulative wealth over the same observed return sequence. G_i is a calendar-month growth rate only when the observed months are consecutive. By the arithmetic-geometric mean inequality, $A_i \geq G_i$, provided that $1 + r_{it} > 0$. For convenience, let $D_i = A_i - G_i$, where D_i can be thought of as a sample measure of volatility drag. For small returns, a second order approximation to D_i is approximately $0.5\sigma_i^2$, where σ_i^2 is the variance of monthly returns. When the distribution of returns has large jumps, skewness or heavy tails, this expression is merely an approximation, and not an exact identity. The following empirical analysis thus aims to answer three related yet separate questions. Firstly, does D_i have a positive sign systematically across A-share stocks, and is the sign mismatch $A_i > 0$, $G_i < 0$ economically significant? Second, does the cross-sectional relationship

DOI: 10.54254/2754-1169/2026.38679

between D_i and variance follow the volatility-drag mechanism? Third, does the incidence of sign mismatch change with the variation of the minimum observation window, the use of dividend treatment, the treatment of early post-listing months, and the trading-activity screen? Since both D_i and variance are computed from the return observations, it is only as a mechanism-consistency check that these two variables are associated, and not as an independent estimate of the effect of volatility on growth.

The data are collected from the Monthly Individual Stock Return File of the China Stock Market Trading Database from January 2015 to December 2024. The raw file contains 481,639 observations for 5,696 security codes. The sample includes only stocks listed on the Shanghai and Shenzhen Main Boards, ChiNext and STAR Market (market-type codes 1, 4, 16 and 32), excluding B-shares and Beijing Stock Exchange stocks. Monthly individual-stock return (including reinvested cash dividends) (Mretwd) is the main return variable. The number of valid monthly return observations of the A-shares after cleaning is 460,274. The retained A-share sample contains no duplicate stock-month combinations and no returns less than or equal to -100%, which would render the logarithm undefined.

The main analysis uses at least 36 valid monthly returns, resulting in 4,647 stocks (the median is 113 valid months per stock). Missing returns are not imputed, and raw returns are not winsorised in the main specification so that observed tail risk is preserved. Pooled winsorisation at the 1st and 99th percentiles is used only as a sensitivity check. A_i, G_i, D_i and the variance of monthly returns are computed for each stock, as well as an indicator variable showing whether A_i is greater than 0 and G_i is less than 0. The cross-sectional relation between D_i and σ_i^2 is then estimated using HC3 heteroskedasticity-robust standard errors. Sensitivity checks use minimum counts of 24 and 60 valid months, require a complete 120-month history, exclude cash-dividend reinvestment, exclude the first 12 valid months, and retain observations for which the reported number of trading days (Ndaytrd) is missing or at least 10. The main descriptive results and the construction of the samples are summarized in Table 1.

Table 1. Sample construction and main results

Metric	Result	Explanation
Raw observations	481,639	Jan 2015-Dec 2024; 5,696 security codes
Valid A-share monthly-return observations	460,274	Market types 1/4/16/32; non-missing Mretwd
Stocks in the main sample	4,647	At least 36 valid return observations per stock
Median monthly arithmetic mean	0.78%	Cross-sectional median of stock-level A_i
Median monthly geometric mean	-0.13%	Cross-sectional median of stock-level G_i
Sign mismatch	1,696 (36.50%)	$A_i > 0$ and $G_i < 0$
Median monthly mean gap	0.93 percentage points	Stock-level $D_i = A_i - G_i$

DOI: 10.54254/2754-1169/2026.38679

3. Empirical results

3.1. Positive average returns and negative compound growth

The median stock-level monthly arithmetic mean return, geometric mean return and the arithmetic-geometric mean gap are 0.7777%, -0.1284% and 0.9297 percentage points for the main sample. None of the 4,647 stocks has a negative mean gap beyond numerical tolerance. The simple annualised median arithmetic mean is 9.33% per annum, while the geometric annualised return from monthly log growth is -1.53% per annum. These are not the same statistics; the difference cannot be interpreted as "10.86 percentage points less return". However, it is evident from the comparison that single-period averages of performance systematically overestimate compounded performance. The sign mismatch is more economically informative. Of the 3,704 stocks with $A_i > 0$, 1,696 still have $G_i < 0$. Therefore, 36.50 percent of the main sample (and 45.79 percent of stocks with positive arithmetic means) do not exhibit positive compound growth over their observed return sequences. Figure 1 shows A_i against G_i for each stock, with the points shown in red for those stocks that have the sign mismatch. However, the outcome is not a reflection of the actual performance of an investor holding the stocks, as the purchase timing, rebalancing and transaction costs as well as portfolio diversification are not modeled. It merely demonstrates that a positive historical average for the month is not enough to prove that wealth is growing over a longer period of time.

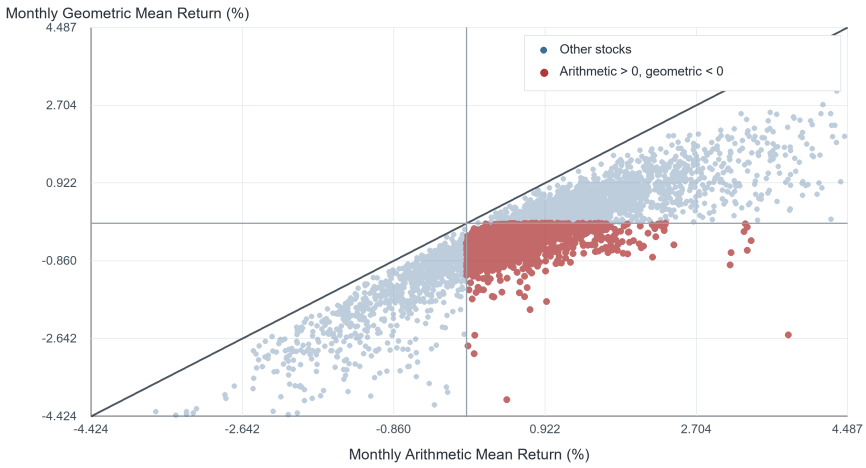

Figure 1. Arithmetic and geometric mean returns of individual stocks (picture credit: original)

Notes: The sample contains 4,647 Chinese A-share stocks with at least 36 valid monthly return observations during 2015-2024. Axes are limited to the 1st and 99th percentile bounds of A_i and G_i for display; observations outside these ranges are not shown. The 45-degree line denotes $A_i = G_i$, and red points satisfy $A_i > 0$ and $G_i < 0$. Source: Author's calculations based on CSMAR data.

3.2. Variance, tail returns, and the second-order approximation

Monthly return variance is highly positively correlated with the mean gap: Pearson's correlation coefficient is 0.802 and Spearman's rank correlation coefficient is 0.989. In the raw-sample regression, the fitted intercept is 0.00682 (HC3 standard error = 0.00110), and the variance coefficient is 0.147 (HC3 standard error = 0.044, $p < 0.001$), with $R^2 = 0.643$. P-values are for the hypothesis of a zero slope, and not equality with 0.5. This coefficient is quite small compared to the theoretical value of 0.5, but this does not rule out the approximation. The second-order expansion might be less accurate for A-share returns, as the sample contains returns affected by price-limit

DOI: 10.54254/2754-1169/2026.38679

adjustments, volatile returns around IPOs, and other unusual observations. If the monthly returns are winsorized at the pooled 1st and 99th percentiles (-31.24% and 48.57%), then the fitted intercept falls to 0.000307 (HC3 standard error = 0.000012), the variance coefficient increases to 0.462 (HC3 standard error = 0.000745, $p < 0.001$), and R^2 rises to 0.993. The intercepts represent 0.682 and 0.0307 percentage points and D_i and σ_i^2 are based on decimal returns. In Figure 2 the fitted line from the raw-sample regression is plotted along with the line of slope 0.5. This evidence is in line with volatility drag as the foremost mechanical source of the mean gap, and reveals that the approximation is more quantitatively illuminating when controlling for tail observations. However, the results shown are the realized returns as per the sample before winsorization is applied, and thus a better statistical fit is not a substitute for the observed tail of the return distribution.

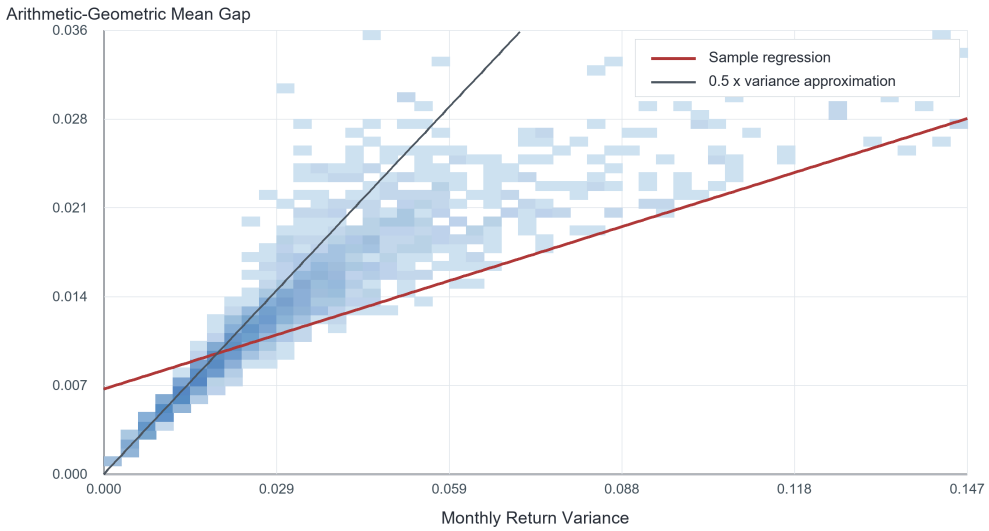

Figure 2. Monthly return variance and the arithmetic-geometric mean gap (picture credit: original)

Notes: The sample contains 4,647 Chinese A-share stocks with at least 36 valid monthly return observations during 2015-2024. Shading shows the log-transformed number of stocks in each two-dimensional bin. Both axes are limited at their 99th percentiles for display; the regression line uses all 4,647 stocks. The gray line denotes $D_i \approx 0.5\sigma_i^2$. D_i is shown in decimal-return units, and σ_i^2 is the variance of returns expressed as decimals. Source: Author's calculations based on CSMAR data.

3.3. Market-segment differences and robustness

Descriptive differences between market segments are presented in Table 2. The sign-mismatch rate is highest among all market segments for the Shenzhen Main Board, at 40.00%. The rates of the Shanghai main board, ChiNext and STAR market are 36.36%, 34.85% and 27.93%, respectively. ChiNext's monthly mean gap has a median of 1.14 percentage points, higher than those of the two major exchanges, consistent with its higher return volatility. The median geometric mean for the STAR Market is -1.06%, but just 376 stocks are listed on the STAR Market and the median number of valid months is only 50, with many of the stocks listed in recent years. These data do not support a claim of an independent "segment effect." A comparison of segments is therefore descriptive only.

DOI: 10.54254/2754-1169/2026.38679

Table 2. Stock-level results by market segment

Market segment	Stocks	Median valid observations	Median A_i	Median G_i	Sign mismatch
Shanghai Main Board A shares	1,675	115	0.68%	-0.10%	36.36%
Shenzhen Main Board A shares	1,500	116	0.87%	-0.06%	40.00%
ChiNext	1,096	95.5	1.04%	-0.12%	34.85%
STAR Market	376	50	0.00%	-1.06%	27.93%

Table 3 demonstrates that, for a minimum observation window of 24 months, the sign-mismatch rate is 35.17%, while for 60 months it is 38.53%. Requiring a complete 120-month history reduces the rate to 34.00% but it also removes stocks from the sample due to "survivorship" and "listing date" selection effects. When excluding cash dividend reinvestment, the rate is 36.39%; after excluding the first 12 valid months post-listing, the rate is 36.63%; and under the implemented trading-activity rule, which retains months for which Ndaytrd is missing or at least 10, the rate is 36.30%. The estimates vary only slightly around the base estimate of 36.50%, suggesting that the central result is not sensitive to the dividend treatment, unusual volatility just after the listing, or low trading activity.

Table 3. Sensitivity analysis of the sign-mismatch rate

Specification	Stocks	Mismatch rate	Relative to baseline
At least 24 valid return observations	5,001	35.17%	-1.32 percentage points
Baseline: at least 36 valid return observations	4,647	36.50%	Baseline
At least 60 valid return observations	3,743	38.53%	+2.03 percentage points
Complete 120-month history	1,097	34.00%	-2.49 percentage points
Excluding cash-dividend reinvestment	4,647	36.39%	-0.11 percentage points
Excluding the first 12 valid return observations	4,152	36.63%	+0.14 percentage points
Reported Ndaytrd >= 10 or missing	4,645	36.30%	-0.20 percentage points

4. Kelly implications, model risk, and the scope of the study

First, the empirical results give a practical reason to use a log-growth perspective. However, this is not a test of whether a Kelly strategy is better. The analysis does not provide quantitative estimates of subjective probabilities, cross-asset covariances, risk-free interest rates, or optimal portfolio weights, and it does not construct a tradable portfolio. The conclusion is confined to the analysis of multi-period wealth (or compound growth): the arithmetic mean return is a poor measure of compound growth. This study does not recommend following the Kelly criterion.

Secondly, model risk should not be left only as a disclaimer at the end of the paper. Forecast quality, parameter uncertainty, distributional misspecification, and risk controls can significantly influence Kelly-type allocations. The historical sample selection of the study needs to be considered

DOI: 10.54254/2754-1169/2026.38679

as well. Another key time-series point is that bull and bear markets occur at different times, while stocks enter and leave the sample in different periods. The 36-month minimum does not account for securities with shorter histories, and does not completely mitigate delisting and survivorship biases. Extreme observations may be indicative of true risks or may be due to conventions for price adjustment, special treatment, or trading rules. This sensitivity is revealed by the winsorization and sample-definition checks.

Thirdly, individual-stock growth does not necessarily translate into portfolio growth. Transaction costs, taxes, and liquidity constraints will limit realized growth, while diversification and rebalancing will lower the portfolio variance and therefore the arithmetic-geometric mean difference. For further research, equal-weighted, value-weighted, constant-proportion and fractional Kelly strategies should be compared using investable A-share portfolios, with rolling windows, out-of-sample estimation and adjustment of delisting returns. The relative merits of these strategies can only be judged after the investor's utility function, constraints, and rebalancing rules are defined.

5. Conclusion

This study, based on monthly individual-stock returns of Chinese A-share stocks compiled by CSMAR for the period 2015-2024, reveals the existence of an economically significant difference between the arithmetic and geometric mean returns of Chinese A-share stocks. For the main sample of 4,647 stocks having a minimum of 36 valid monthly observations, the median value of the arithmetic-geometric monthly mean difference is 0.93 percentage points; the sign mismatch of the arithmetic mean being positive while the geometric mean is negative is present in 36.50% of the stocks. The percentage remains in a similar range across the robustness analyses, with different observation thresholds and dividend treatment and/or different screens for trading activity, ranging between 34% and 39%. The empirical results are consistent with the volatility-drag mechanism, but the approximation that the average gap equals one-half of the return variance exhibits high quantitative explanatory power only in the winsorised sensitivity check. It is therefore important that long-horizon investment analysis also shows geometric mean returns, return volatility and tail risk, as well as arithmetic mean returns, and that the period of the sample, dividend treatment and missing-value rules are also transparently disclosed. The average single period return is not necessarily a good indicator of the long-run compound growth of an asset.

The main contribution of the paper is to consider the arithmetic-geometric mean gap based on actual returns of A-shares from China; to measure the frequency and economic significance of the gap between average return and compound growth; and to separate descriptive empirical facts from the normative question of optimal investment under the Kelly criterion. This separation is a way to avoid drawing a trading strategy from statistical patterns. This evidence can be used to evaluate assets and performance, and it highlights the risk involved in compounding. Several limitations remain. The sample covers specific bull and bear market periods and is vulnerable to survivorship bias; no strategies are tested, and no out-of-sample portfolio analysis is performed. In future, more detailed research can explore the feasibility of growth-optimal strategies under various constraints using rolling-window analysis and adjustments for delisting returns. Different portfolio rules, such as equal weighting, value weighting, and fractional Kelly, could be compared.

References

[1] Kelly, J. L., Jr. (1956). A new interpretation of information rate. *Bell System Technical Journal, 35*(4), 917–926.

DOI: 10.54254/2754-1169/2026.38679

[2] MacLean, L. C., Ziemba, W. T., & Blazenko, G. (1992). Growth versus security in dynamic investment analysis. *Management Science, 38*(11), 1562–1585.
[3] Cover, T. M. (1991). Universal portfolios. *Mathematical Finance, 1*(1), 1–29.
[4] Browne, S., & Whitt, W. (1996). Portfolio choice and the Bayesian Kelly criterion. *Advances in Applied Probability, 28*(4), 1145–1176.
[5] Wei, P., & Xu, Z. Q. (2025). Dynamic growth-optimal portfolio choice under risk control. *European Journal of Operational Research, 322*(1), 325–340.
[6] Sun, Y. M., & Zou, Z. F. (2025). Data-driven distributionally robust Kelly portfolio optimization based on coherent Wasserstein metrics. *JUSTC, 55*(8), Article 0805.
[7] Heikkinen, R., Karvanen, J., & Miettinen, K. (2025). A Bayesian model for portfolio decisions based on debiased and regularized expert predictions. *Journal of Business Economics, 95*(5), 669–706.
[8] Levy, H. (2025). The maximum geometric mean criterion: Revisiting the Markowitz-Samuelson debate: Survey and analysis. *Annals of Operations Research, 346*(1), 263–284.
[9] Vecer, J. (2026). Analytical solution for Kelly's criterion for multiple outcomes. *Quantitative Finance, 26*(5), 671–684.
[10] Watt, R. (2026). When Harry met Kelly: An overlooked result in the classical theory of optimal capital growth. *Annals of Finance, 22*(1), Article 8.