



Price Gap Anomaly: Empirical Study of Opening Price Gaps and Price Disparities in Chinese Stock Indices

Yuancheng Si¹ · Saralees Nadarajah²

Accepted: 16 April 2024

© The Author(s), under exclusive licence to Springer Japan KK, part of Springer Nature 2024

Abstract

In this study, we employ statistical analysis, hypothesis testing, and regression models to investigate the characteristics of opening price gap rates and price gaps in the stock market indices of Mainland China, utilizing historical data. To clarify, while both 'opening price gap rate' and 'price gaps' are central to our analysis, they represent distinct concepts. The opening price gap rate refers to the rate at which a stock's opening price differs from its previous closing price, indicating initial market sentiment and potential momentum for the trading day. In contrast, price gaps, as defined in technical analysis, are specific chart patterns formed by two adjacent candlesticks on consecutive trading days. These patterns are characterized either by one candlestick's low being higher than the following day's high, or one candlestick's high being lower than the following day's low, creating a "blank" area on the price chart. This signifies a price range with no trading activity and is a crucial indicator of market sentiment and potential directional moves. Our study tested and validated thirteen related hypotheses. The findings reveal a significant correlation between the directionality of price gaps and the fluctuations in opening price gap rates, highlighting key characteristics of the market. Notably, price gaps significantly impact daily changes in trading volume and turnover. Furthermore, we validated the efficacy of the opening price gap rate as a stock-picking factor through back-testing. This research offers a new perspective for understanding stock market behaviors and has considerable implications for investment decisions and market analysis.

Keywords Opening price gaps · Price disparities · Chinese stock indices · Stock selection strategies

✉ Yuancheng Si
siyuancheng@fudan.edu.cn

Saralees Nadarajah
saralees.nadarajah@manchester.ac.uk

¹ School of Economics, Fudan University, Shanghai 200433, China

² Department of Mathematics, University of Manchester, Manchester M13 9PL, UK

1 Introduction

In stock markets, the opening price gap, also referred to as the morning gap, is a significant phenomenon (Plastun et al., 2020; Caporale & Plastun, 2017; Plastun et al., 2019). This term describes the difference between the opening price of a trading day and the closing price of the previous trading day. This phenomenon is not only prevalent in stock markets but is also observed in other financial markets such as commodities. Price gaps can be further classified into positive and negative gaps. Positive gaps occur when the new opening price is higher relative to the previous day's closing price, often due to the market receiving positive information after-hours, leading buyers to pay higher prices at the next opening. Conversely, negative gaps occur when the opening price is lower than the previous day's closing price, which may result from negative information received after the market closes, prompting sellers to accept lower prices at opening. Overall, these price gaps reveal the dynamics of information transmission and response in the market, as well as how market participants interpret and react to information. A deeper analysis of the mechanisms and factors influencing positive and negative gaps can enhance our understanding of market behavior, thus providing valuable insights for investment decisions and market forecasting.

This study aims to explore the existence and trends of price gap anomalies in the Chinese stock market, analyzing the characteristics of this phenomenon and its effects on relevant hypotheses. In the context of stock index levels, price gaps refer to the disparities observed between consecutive trading days' index values, typically arising from the market's opening index being higher or lower than the previous trading day's closing index. On the other hand, at the individual stock level, price gaps denote the differences between the opening price and the previous trading day's closing price of individual stocks. The exclusive focus of this paper on stock index level price gaps is justified for the following reasons:

1. Stock index level price gaps offer a more comprehensive reflection of overall market behavior and characteristics, rendering them more representative.
2. Stock index level price gaps are more readily accessible and analyzable by researchers, as they can be directly obtained from market index data, whereas data at the individual stock level requires more extensive collection and processing efforts.
3. Price gaps at the stock index level hold significant reference value for investors and market participants, aiding in their understanding of market behavior and forecasting future trends.

Therefore, this paper concentrates on analyzing price gaps at the stock index level to provide a thorough exploration of the phenomenon and its implications on the market.

Our research focuses on the daily data of the Shanghai Composite Index (SH) and the Shenzhen Component Index (SZ) from 1990 to 2023. The study will employ statistical analysis, hypothesis testing, and other methods. Additionally,

simulation techniques will be used to assess whether price gap anomalies offer exploitable profit opportunities. Given the lack of comprehensive research over such an extended period in the Chinese stock market in existing literature, this study fills this gap. Therefore, this paper will include a review of related literature, discussions on data and methodology, research findings, and final conclusions.

In addressing the critical research question concerning price gap anomalies within the Efficient Market Hypothesis (EMH) framework, it is essential to delineate what constitutes price gap anomalies. These anomalies, appearing to contravene the EMH tenets, can be attributed to several factors that lead to opening price gaps. This discussion integrates various factors contributing to price gaps, providing a comprehensive understanding of the phenomena in question.

1. **Information Lag:** Despite the assumption of the Weak Form EMH that all publicly available information is fully reflected in market prices, a practical discrepancy exists due to information lag. Information such as corporate earnings reports or significant news, typically released after trading hours or during market closures, can cause opening price discrepancies when the market reopens and investors rapidly react to the newly available information, challenging the Weak Form EMH.
2. **Irrational Trading Behavior:** Contrary to the EMH assumption of investor rationality, emotional and psychological influences may lead to irrational trading decisions, such as panic selling or excessive buying at market opening. These behaviors, resulting in unreasonable price fluctuations, underscore the impact of irrational psychological patterns on market prices.
3. **Market Liquidity:** At market opening, liquidity can be relatively low, with limited trading volumes and order quantities. This insufficiency can lead to substantial price volatility, even for smaller-scale buy or sell orders, suggesting that the market might not always efficiently reflect information as posited by EMH.
4. **Trading Strategies:** Strategies such as market making or high-frequency trading that exploit price discrepancies at market opening for profit may contribute to opening price gaps. These strategies, relying on price differentials present at market opening, highlight the occurrence of short-term market inefficiencies and anomalous price movements.

Moreover, specific factors further illustrate the complexity of market dynamics contributing to price gap anomalies:

- **Publication of Macro and Microeconomic Information** (Jiang & Zhu, 2017; Tetlock, 2010): The release of significant macroeconomic information outside trading hours can alter investor expectations, leading to opening price gaps, thereby challenging the premise of Weak Form EMH.
- **Company-specific Information Leak** (Avishay et al., 2023; Tetlock, 2010): Information leaks, such as financial statements or major decisions disclosed post-market closure, can cause significant opening price differences, reflecting a deviation from EMH.

- **Changes in Market Microstructure Information** (Li et al., 2021): Alterations in market microstructure, like trading volume and frequency, influence stock prices, contributing to discrepancies between opening and closing prices.
- **Non-Economic Factors Affecting Market Sentiment** (Chi et al., 2012; Guo et al., 2017): Shifts in market sentiment due to non-economic factors, such as geopolitical events, can lead to significant opening price differences, illustrating the market's inefficiency in reflecting all available information instantaneously.
- **Liquidity Shocks**(Tetlock, 2010): Sudden changes in market liquidity at the moment of market opening can lead to significant differences between a stock's opening and closing prices, highlighting a potential flaw in the EMH assumption of market efficiency.

These factors collectively demonstrate that under certain conditions, markets may not fully adhere to the principles of EMH, challenging the hypothesis's applicability in explaining opening price gaps. The exploration of price gap anomalies, thus, provides a rich perspective for studying market efficiency and the underlying mechanisms of such anomalies, revealing the complexity of real-world market behavior beyond the idealized scenarios posited by EMH.

In technical analysis of financial markets, particularly in the analysis of stock index opening gaps, the concept of Gaps plays a crucial role. To clarify, while both 'opening price gap rate' and 'price gaps' are central to our analysis, they represent distinct concepts. The opening price gap rate refers to the rate at which a stock's opening price differs from its previous closing price, indicating initial market sentiment and potential momentum for the trading day. In contrast, price gaps, as defined in technical analysis, are specific chart patterns formed by two adjacent candlesticks on consecutive trading days. These patterns are characterized either by one candlestick's low being higher than the following day's high, or one candlestick's high being lower than the following day's low, creating a "blank" area on the price chart. This signifies a price range with no trading activity and is a crucial indicator of market sentiment and potential directional moves. Gaps can be classified into four types based on their nature and market impact:

1. **Common Gaps:** These typically manifest in a consolidating market or during periods of relative trading calm. They are generally filled promptly and do not lead to significant market consequences. Such gaps may arise due to market participants' transient reactions to non-significant information, illustrating that even in weakly efficient markets, prices can deviate from their true value in the short term due to trading noise.
2. **Continuation Gaps:** These gaps emerge within the continuation phase of a stock's upward trend, usually not getting filled swiftly, as the stock price persists in its ascent. Such gaps often indicate the release of new, materially impactful information, prompting a price adjustment to a new equilibrium. This scenario aligns with the semi-strong form of the Efficient Market Hypothesis, where market prices adjust to reflect newly public information.
3. **Breakaway Gaps:** Formed when stock prices breach existing price formations accompanied by an increase in trading volume, they typically signal the onset

of a new market trend. These gaps suggest that the market requires time to fully absorb and reflect significant information that could alter market direction. In such instances, the market may exhibit short-term inefficiencies as market participants evaluate the new situation and adjust their positions accordingly.

4. **Exhaustion Gaps:** These usually occur at the end of a price rise cycle, often with increased volume but without achieving new highs. Such gaps imply an impending market trend reversal, serving as a signal of a shift from excessive optimism to caution or bearish sentiment among investors. Exhaustion gaps reveal that even when the market appears to follow the strong form of the Efficient Market Hypothesis, irrational investor behaviors can drive market prices away from their fundamental values in the short term.

Regarding gap filling behavior, it specifically refers to the phenomenon where stock prices drop and then rise above the gap price or rise and then fall below the gap price after a decline. Although various hypotheses and speculations about gap phenomena in the Chinese stock market have garnered much attention, there is still a lack of clear academic literature to systematically verify and backtest these hypotheses.

2 Literature Review

From a theoretical perspective, according to the Efficient Market Hypothesis (EMH) (Fama, 1965, 1970), in an efficient market, asset prices will fully and promptly reflect all available information. Thus, there is a close relationship between the Efficient Market Hypothesis and stock index price differences. In a perfectly efficient market, investors cannot obtain excess returns through any means, as all information is fully absorbed by the market and reflected in prices. However, the occurrence of opening price differences suggests a certain degree of market inefficiency. Specifically, if there is a difference between the closing price of the previous trading day and the opening price of the current day, it may indicate that the market did not fully incorporate all information into the closing price of the previous trading day or that there is a delay in the market's reaction to new information (Jiang & Zhu, 2017; Tetlock, 2010). Although stock index price differences imply possible market inefficiency, it does not mean that investors can easily obtain excess returns. In practice, investors need to consider many other factors, including transaction costs, market volatility, and other factors that may affect investment decisions. Moreover, even if there is a certain degree of market inefficiency, the Efficient Market Hypothesis reminds us that this inefficiency may quickly disappear as market participants react. From a practical perspective, Plastun et al. (2019) study rigorously examines the Ukrainian stock market, specifically analyzing the UX index from 2009 to 2018, to discern patterns and anomalies related to price gaps. Through comprehensive statistical and regression analyses, it finds no significant evidence of seasonality or abnormal behavior post-gaps, aligning with the Efficient Market Hypothesis, except for a momentum effect on days with negative gaps, suggesting a profitable trading strategy that contradicts market efficiency. Similarly, Caporale and Plastun (2017) undertakes an extensive analysis of price gaps across stock, FOREX, and commodity

markets from 2000 to 2015, employing various statistical tests to explore six hypotheses regarding market efficiency. It concludes that while most market behaviors align with efficiency, FOREX markets exhibit an anomaly that allows for the generation of abnormal profits through a specific trading strategy, highlighting a distinct deviation from market efficiency in the FOREX sector. Adding to the complexity of market efficiency, Si et al. (2024) developed a hybrid statistical model designed to accurately capture the dynamic patterns of opening price gaps in Chinese stock markets. Recent studies have delved into the nuanced dynamics of stock market (fang et al., 2022; Ho et al., 2023; Zhang et al., 2023), revealing intriguing exceptions to the momentum effect. A novel examination into this market demonstrates that intraday and overnight returns significantly influence future stock returns in differing manners. Investors show a tendency to underreact to intraday information, while overreacting to overnight information, leading to the formulation of intraday momentum and overnight momentum strategies. This dichotomy not only challenges the traditional understanding of market reactions but also illustrates the persistence of profitability, showcasing its resilience against momentum crashes. Furthermore, the relationship between overnight returns and investor sentiment on the Taiwan Stock Exchange (TWSE) has been reassessed (Zhang et al., 2023), corroborating the findings by Aboody et al. that overnight returns reflect investor sentiment. This study extends the understanding by highlighting how trading activities by different investor types amplify the patterns of overnight returns, with a significant role played by retail trading volume. It elucidates that overnight returns contribute to both short-term persistence and long-term return reversals, driven by investor sentiment. These insights not only validate the use of overnight returns as a measure of investor sentiment in the TWSE but also suggest the influence of market structure and investor behaviors as critical determinants in non-US markets (Aboody et al., 2018). These recent findings enrich the discourse on market efficiency by illustrating how specific market mechanisms and investor behaviors can lead to anomalies that both challenge and complement the Efficient Market Hypothesis. They underscore the importance of considering intraday and overnight information separately in analyzing market dynamics and formulating trading strategies. To bridge the insights from specific market behaviors and anomalies highlighted in Plastun et al. (2019); Caporale and Plastun (2017) with the broader considerations of market dynamics, it's imperative to understand the underlying factors contributing to price differences.

3 Data and Preprocessing

3.1 Data Description

In this study, we first establish the relationship between price gaps (hereafter referred to as "gaps") and the difference rate (hereafter referred to as "diffrate"). In the context of financial market analysis, diffrate can be understood as a broad indicator of price movement, while price gaps represent a subset of diffrate under specific conditions. Specifically, a price gap refers to a sudden jump in trading price due to a temporary interruption in order flow (commonly known as the "liquidity vacuum" phenomenon),

often manifesting as a significant difference between the opening price of one trading day and the closing price of the previous day. Thus, we consider diffrate as a comprehensive measure that captures both micro-adjustments in market prices and macro-fluctuations driven by various factors. In contrast, a price gap is a special case within diffrate, specifically denoting those significant and sudden price movements caused by an "order flow vacuum." In this study, we will explore a series of theoretical hypotheses regarding these two phenomena, particularly the unique characteristics of price gaps in the operation of stock indices.

This research utilizes daily trading data of the Shanghai Stock Exchange Index (SH) and the Shenzhen Component Index (SZ) downloaded from sources like Sina Finance and Yahoo Finance. The sample period covers all daily index data since the inception of the SH index (1990–2023) and the SZ index (1991–2023), including 8028 SH index and 7936 SZ index trading data points. Each data point includes eight attributes: opening price, closing price, highest price, lowest price, trading volume, trading amount, price increase/decrease, and price change percentage. Notably, the dataset contains no missing values, ensuring the completeness of the analysis. For strategy testing and gap analysis, particular attention is paid to historical data from 2021 to 2023, aiming to analyze the profitability of related gap strategies.

In the dataset under consideration, gap types are defined in the following manner:

- An **Upward Gap** is observed when the low price on a given day, denoted as L_t , exceeds the high price from the previous day, H_{t-1} . This phenomenon is mathematically represented as:

$$L_t > H_{t-1}. \quad (1)$$

- Conversely, a **Downward Gap** is identified when the high price on a given day, denoted as H_t , falls below the low price from the prior day, L_{t-1} . This condition is mathematically expressed as:

$$H_t < L_{t-1}. \quad (2)$$

These gap types indicate a range in prices where no trades were executed between two successive trading days, serving as critical indicators of market sentiment and potential shifts in market direction. Furthermore, the dataset includes a variable termed "diffrate," which captures the daily opening price difference rate, with "shdiffrate" for the Shanghai index and "szdiffrate" for the Shenzhen index. The "diffrate" is defined by the equation:

$$\text{diffrate}_t = \left(\frac{\text{open}_t - \text{close}_{t-1}}{\text{close}_{t-1}} \right) \times 100\% \quad (3)$$

where diffrate_t represents the opening price difference rate on day t , open_t is the opening price on day t , and close_{t-1} is the closing price on day $t - 1$. It is noteworthy that this phenomenon has also been referred to as "overnight return" in previous studies (Ho et al., 2023; Zhang et al., 2023). The significance of diffrate lies in providing the daily opening price's relative movement compared to the previous

day's closing price. Positive values indicate an opening price higher than the previous day's closing price, while negative values indicate a lower opening price. This metric is crucial for understanding market volatility and trends, helping investors, traders, and analysts gauge stock market directions and price movement magnitudes. Diffrate not only reflects market sentiment but also reveals supply and demand dynamics in the market. For instance, positive values might signify positive market expectations for a stock or index, while negative values could reflect market concerns or cautious sentiment. Additionally, when the absolute value of diffrate is large, it may signal significant market fluctuations or extraordinary events, significantly impacting the stock market. Descriptive statistical analysis of diffrate for the SH index (*shdiffrate*) and SZ index (*szdiffrate*) yields a series of key statistical indicators reflecting the performance characteristics of the two indices during the sample period, as shown in Table 1. The SH index has an average diffrate of 0.0089, while the SZ index has an average diffrate of -0.0217 . The median diffrate for the SH index is -0.0076 , and for the SZ index, it is -0.0049 . The negative median values indicate that in both indices, more than half of the trading days have opening prices lower than the previous day's closing prices, with specific distribution ranges detailed in Table 2.

The maximum diffrate value for the SH index was 104.2691, occurring on 1992-05-21; the minimum value was -21.82161 , occurring on 1992-08-12. For the SZ index, the maximum diffrate value was 18.28696, occurring on 1995-05-18, and the minimum value was -21.16085 , occurring on 1991-08-19. These extreme values likely reflect the impact of specific

events or market fluctuations, such as the deregulation of listed stock prices on 1992-05-21. The standard deviation of the SH index diffrate is 1.5637, and for the SZ index, it is 0.86301, indicating greater volatility in the SH index diffrate. The skewness of the SH index diffrate is 37.73352, and for the SZ index, it is -0.20888 . The high positive skewness of the SH index diffrate indicates a long tail in the positive direction, while the slight negative skewness of the SZ index diffrate suggests a slight leftward tilt in its distribution. The kurtosis of the SH index diffrate is 2485.598, and for the SZ index, it is 112.5812, both exhibiting high kurtosis characteristics, especially for the SH index diffrate, showing an unusually high peak. The Lilliefors (Kolmogorov-Smirnov) normality tests were conducted to assess the distribution of the variables *shdiffrate* and *szdiffrate*. The test results are as follows: For *shdiffrate*, the test statistic was $D = 0.26284$, with a p -value $< 2.2 \times 10^{-16}$, indicating a significant deviation from the normal distribution. Similarly, for *szdiffrate*, the test statistic was $D = 0.17053$, with a p -value $< 2.2 \times 10^{-16}$, also suggesting a considerable departure from normality. Given the extremely low p -values in both tests, there is strong evidence to reject the null hypothesis of normality for both variables. These findings imply that the distributions of *shdiffrate* and *szdiffrate* do not conform to a normal distribution, which also could be found in Fig 1 visually. To explore the specific distribution characteristics of *shdiffrate* and *szdiffrate* further, it would be beneficial to consult the methodology outlined in the article by Si and Nadarajah (2022), which provides a comprehensive framework for analyzing non-normal financial data distributions.

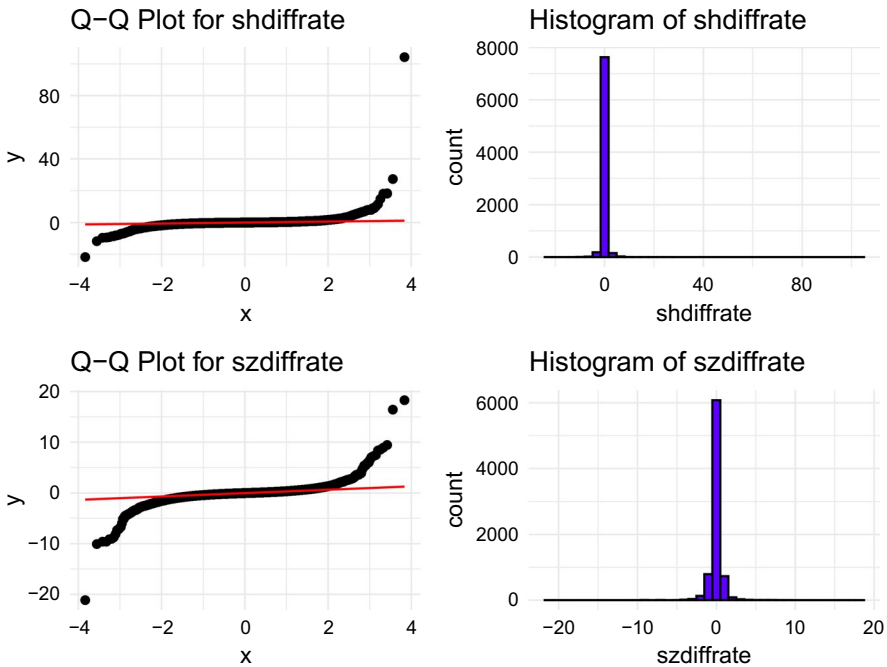


Fig. 1 Normality Plot

Overall, these statistical results reveal the behavior characteristics of the SH and SZ indices' opening price difftrate under different market conditions. The SH index's opening price difftrate demonstrates greater volatility and extreme price behavior, while the SZ index's opening price difftrate is relatively more stable. These findings may reflect the differences between the two markets in responding to external information and internal dynamics.

4 Data Visualization and Trend Analysis

4.1 Time Series Analysis of Opening Price Difference Rate

Through plotting a time series graph of the 'Opening Price Difference Rate' (Fig. 2), we further explore its trends and volatility. The chart provides a detailed representation of the evolution of the opening price difference rate (shdifftrate and szdifftrate) over time for the Shanghai Stock Exchange Index (SH, depicted in blue) and the Shenzhen Component Index (SZ, depicted in red).

In the early 1990 s, the difftrate of the SH index exhibited notable high volatility, especially in the initial stages of the time series, with several significant peaks reflecting substantial price fluctuations in the market's early days.

As we moved into the late 1990s and early 2000s, the volatility of the SH index's difftrate decreased. Although the extreme fluctuations seen at the beginning of the

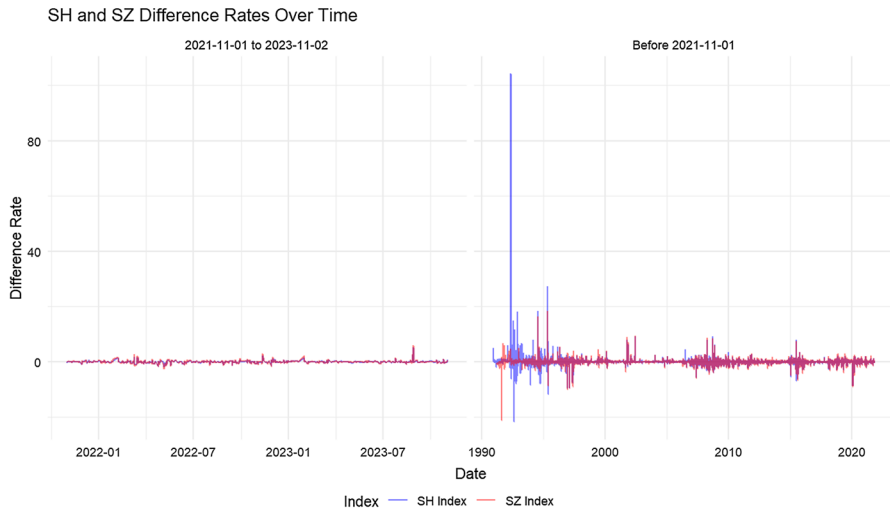


Fig. 2 Time series plots of diffrate

Table 1 Descriptive Statistics of shdiffrate and szdiffrate from period shdiffrate, 1990–2023; szdiffrate, 1991–2023

Statistic	shdiffrate	szdiffrate
Mean	0.0089	-0.0217
Median	-0.0076	-0.0049
Maximum	104.2691	18.28696
Minimum	-21.82161	-21.16085
Standard Deviation	1.5637	0.86301
Skewness	37.73352	-0.20888
Kurtosis	2485.598	112.5812
Number of Observations	8028	7936

series weakened, there were still several intense movements. During this period, the SZ index's diffrate began to appear in the chart with smaller fluctuations, but the continuous movement indicated that the market had not yet fully stabilized.

From the mid-2000s to the early 2010s, the opening price difference rate of both indices was relatively stable, suggesting more mature market behavior during this phase. Despite this, occasional spikes in the data pointed to the market's sensitive response to specific information.

In the more recent period from the 2010s to the early 2020s, the diffrate of both indices remained within a lower range of volatility, suggesting a quicker market reaction to information and a reduced difference between the opening price and the previous day's closing price.

Overall, the SH index's diffrate demonstrated greater volatility in the earlier part of the time series, gradually stabilizing over time. The SZ index's diffrate, on the other hand, consistently showed more stable opening price difference rate behavior. The

Table 2 Count distribution for SH and SZ difference rates

Interval	SH Count	Interval	SH Count	Interval	SZ Count	Interval	SZ Count
$(-\infty, -10]$	2	$[-10, -9)$	3	$(-\infty, -10]$	2	$[-10, -9)$	4
$[-9, -8)$	4	$[-8, -7)$	4	$[-9, -8)$	2	$[-8, -7)$	2
$[-7, -6)$	6	$[-6, -5)$	6	$[-7, -6)$	2	$[-6, -5)$	3
$[-5, -4)$	14	$[-4, -3)$	34	$[-5, -4)$	8	$[-4, -3)$	19
$[-3, -2)$	69	$[-2, -1)$	279	$[-3, -2)$	70	$[-2, -1)$	275
$[-1, -0.8)$	120	$[-0.8, -0.6)$	234	$[-1, -0.8)$	161	$[-0.8, -0.6)$	264
$[-0.6, -0.4)$	403	$[-0.4, -0.2)$	939	$[-0.6, -0.4)$	470	$[-0.4, -0.2)$	918
$[-0.2, 0)$	1998	$[0, 0.2)$	1920	$[-0.2, 0)$	1833	$[0, 0.2)$	1814
$[0.2, 0.4)$	825	$[0.4, 0.6)$	389	$[0.2, 0.4)$	948	$[0.4, 0.6)$	432
$[0.6, 0.8)$	207	$[0.8, 1)$	208	$[0.6, 0.8)$	258	$[0.8, 1)$	140
$[1, 2)$	234	$[2, 3)$	61	$[1, 2)$	219	$[2, 3)$	55
$[3, 4)$	23	$[4, 5)$	12	$[3, 4)$	17	$[4, 5)$	3
$[5, 6)$	9	$[6, 7)$	9	$[5, 6)$	6	$[6, 7)$	2
$[7, 8)$	6	$[8, 9)$	1	$[7, 8)$	3	$[8, 9)$	3
$[9, 10)$	2	$(10, +\infty)$	7	$[9, 10)$	1	$(10, +\infty)$	2

adjusted transparency of the SZ index in the chart allows for a clearer comparison of the volatility between the two indices. Although the SZ index’s diff rate is relatively less volatile, the fluctuation pattern of the SH index’s diff rate remains distinctly visible.

5 Methods and Validation

In our study on the data outlined in , we will perform hypothesis testing for both the Shanghai and Shenzhen stock indices using the following methods:

- **Welch’s t-test:** Welch’s t-test formula is given by:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4}$$

where \bar{X}_1 and \bar{X}_2 are the means of the two samples, s_1^2 and s_2^2 are the variances of the samples, and n_1 and n_2 are the sample sizes. The degrees of freedom (df) are approximated by:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} \tag{5}$$

Welch's t-test is used when two independent samples have unequal variances. It provides more reliable results compared to the standard t-test (Student's t-test) in cases of unequal sample variances.

- **Wilcoxon Signed Rank Test:**

- Test statistic: Involves the signs and ranks of the differences.
- Null Hypothesis (H_0): No difference between the sample median and a specified value.
- Alternative Hypothesis (H_1): Difference exists between the sample median and the specified value.

- **Chi-Squared Test:**

- Test statistic: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, where O_i is the observed frequency, and E_i is the expected frequency.
- Null Hypothesis (H_0): No significant difference between observed and expected frequencies.
- Alternative Hypothesis (H_1): Significant difference exists between observed and expected frequencies.

- **Proportion Test:**

- Test statistic: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, where \hat{p} is the sample proportion, p_0 is the hypothesized population proportion, and n is the sample size.
- Null Hypothesis (H_0): No difference between the sample proportion and the hypothesized population proportion.
- Alternative Hypothesis (H_1): A difference exists between the sample proportion and the hypothesized population proportion.

- **Dummy Variable Regression:** The process of regression with dummy variables involves the following steps:

1. Mark each trading day as either an upward gap, a downward gap, or a non-gap day in the dataset.
2. Use a linear regression model to estimate the impact of gap type on price fluctuations.
3. Re-run the regression with the non-gap day as the reference category to adjust coefficients.

The structure of the linear regression model is as follows:

$$Y_t = \beta_0 + \beta_1 D_{\text{up},t} + \beta_2 D_{\text{down},t} + \epsilon_t$$

where:

- Y_t represents the price fluctuation on day t .

- $D_{up,t}$ is the dummy variable for upward gaps.
- $D_{down,t}$ is the dummy variable for downward gaps.
- β_0 is the intercept, representing the average fluctuation on non-gap days.
- β_1 and β_2 respectively represent the difference in price fluctuations on upward and downward gap days compared to non-gap days.
- ϵ_t is the random error term for period t .

The size, sign, and statistical significance of the dummy coefficients provide insights into potential market patterns. Upon identifying potential market patterns, this study uses Wind Information Platform’s stock screening and quantitative back-testing tool (EQBT feature) to construct and optimize specific stock strategies. Centered around the core variable of opening price difference rate, we select stocks from a predetermined pool and retrospectively simulate the strategy’s performance using historical data. This step involves adhering to established trading rules and patterns, aiming to validate if strategies based on opening price difference rate and price gaps can yield consistent profits under real market conditions. This phase is crucial in testing market efficiency and exploring viable investment strategies, potentially revealing and leveraging systematic behaviors in the market.

- **Welch’s t-test:** Welch’s t-test formula is given by:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{6}$$

where \bar{X}_1 and \bar{X}_2 are the means of the two samples, s_1^2 and s_2^2 are the variances of the samples, and n_1 and n_2 are the sample sizes. The degrees of freedom (df) are approximated by:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}} \tag{7}$$

Welch’s t-test is used when two independent samples have unequal variances and provides more reliable results compared to the standard t-test (Student’s t-test) in cases of unequal sample variances.

- **Wilcoxon Signed Rank Test:**
 - Test statistic: Involves the signs and ranks of the differences.
 - Null Hypothesis (H_0): No difference between the sample median and a specified value.
 - Alternative Hypothesis (H_1): Difference exists between the sample median and the specified value.

- **Chi-Squared Test:**

- Test statistic: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, where O_i is the observed frequency, and E_i is the expected frequency.
- Null Hypothesis (H_0): No significant difference between observed and expected frequencies.
- Alternative Hypothesis (H_1): Significant difference exists between observed and expected frequencies.

- **Proportion Test:**

- Test statistic: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, where \hat{p} is the sample proportion, p_0 is the hypothesized population proportion, and n is the sample size.
- Null Hypothesis (H_0): No difference between the sample proportion and the hypothesized population proportion.
- Alternative Hypothesis (H_1): A difference exists between the sample proportion and the hypothesized population proportion.

- **Dummy Variable Regression:** The process of regression with dummy variables involves marking each trading day as either an upward gap, a downward gap, or a non-gap day in the dataset, using a linear regression model to estimate the impact of gap type on price fluctuations, and re-running the regression with the non-gap day as the reference category to adjust coefficients. The model is structured as:

$$Y_t = \beta_0 + \beta_1 D_{\text{up},t} + \beta_2 D_{\text{down},t} + \epsilon_t$$

where Y_t represents the price fluctuation on day t , $D_{\text{up},t}$ and $D_{\text{down},t}$ are dummy variables for upward and downward gaps respectively, β_0 is the intercept, and ϵ_t is the error term.

- **Granger Causality Test:** Within our analytical framework, we employ Vector Autoregression (VAR) models as a foundational tool for analyzing the dynamic interrelationships between multiple time series. The VAR model's comprehensive nature allows us to capture the essence of these interactions comprehensively. A critical extension of our VAR analysis is the application of the Granger Causality Test, which serves to further elucidate the directional influences among the variables under study. The Granger Causality Test is a statistical procedure designed to ascertain whether one time series can be utilized to forecast another. The premise of this test is rooted in the concept of temporal precedence, wherein the causality is determined by the informational value of past values of a predictor series in forecasting a target series. Specifically, a time series X is said to Granger-cause another time series Y if past values of X contain unique information that is beneficial in predicting Y , beyond what is already available from past values of Y itself. The empirical application of the Granger Causality Test involves the estimation of two distinct models:

$$Y_t = \alpha + \sum_{i=1}^n \beta_i Y_{t-i} + \sum_{j=1}^m \gamma_j X_{t-j} + \epsilon_t \quad (8)$$

$$Y_t = \alpha + \sum_{i=1}^n \beta_i Y_{t-i} + \epsilon_t \quad (9)$$

Here, Y_t represents the current value of the target time series, while X_{t-j} denotes past values of the predictor series. The coefficients α , β_i , and γ_j embody the model parameters, with ϵ_t signifying the error term. The null hypothesis for the Granger Causality Test (H_0) asserts that $\gamma_j = 0$ for all j , suggesting that the predictor series X does not Granger-cause the target series Y . The integration of Granger Causality Tests within the VAR framework enhances our analysis by allowing us to identify the directionality of relationships between time series. The VAR model provides a multi-dimensional view of the data dynamics, setting the stage for a detailed Granger causality analysis. By assessing the significance of the γ_j coefficients in the context of the VAR model, we can deduce the presence and direction of causal relationships between the variables, thus providing a richer understanding of the underlying economic phenomena. In the following contexts, Table 3 summarizes the specific methodologies we employed for different hypothesis, while Tables 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 present the detailed results of hypothesis testing based on historical data.

The following table outlines the specific hypotheses tested, the corresponding statistical methods utilized, and the key equations or models underpinning our analysis:

6 Empirical Results

6.1 Hypothesis Development and Classification

The investigation into price gap anomalies within financial markets necessitates a rigorous and systematic approach to hypothesis formulation. This section is dedicated to elucidating the theoretical underpinnings and empirical expectations concerning the impact of opening price differences on daily returns, market reactions to news, distribution patterns of price gaps, and their effects on trading volume and turnover rate. By categorizing hypotheses into distinct themes, this study aims to dissect the multifaceted nature of price gap anomalies, providing a structured framework for subsequent empirical analysis. This classification not only enhances the clarity of the research objectives but also aligns with the Efficient Market Hypothesis (EMH) to critically assess market behavior in response to new information, trading patterns, and investor sentiment. Through this meticulous approach, the research endeavors to contribute valuable insights into the dynamics of price gaps, addressing gaps in the literature and offering implications for both market participants and regulatory bodies.

Table 3 Overview of Hypotheses, Statistical Tests, and Models

Hypothesis	Statistical Method	Key Equation(Statistics)/Model
Impact of Opening Price Gaps on Daily Returns	Welch's t-test	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
Market Reaction to News	Dummy Variable Regression	$Y_t = \beta_0 + \beta_1 D_{up,t} + \beta_2 D_{down,t} + \epsilon_t$
Distribution Patterns of Price Gaps	Chi-Squared Test and Proportion Test	$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}; \quad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Effects of Price Gaps on Trading Volume and Turnover Rate	Wilcoxon Signed-Rank Test	Median test for changes in trading volume and turnover rate
Dynamic Relationships between Opening Price Gaps and Market Returns	VAR Model and Granger Causality Test	VAR Model: $Y_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + \epsilon_t$ Granger Causality: Testing if X Granger-causes Y

Each method is selected based on its suitability to address the specific research question at hand, ensuring rigorous and robust empirical investigation. The Welch's t-test is applied to compare mean differences between two groups under the assumption of unequal variances. The Dummy Variable Regression is used to quantify the impact of categorical events on continuous outcome variables. The Chi-Squared Test examines the distribution of categorical data, while the Wilcoxon Signed-Rank Test is employed for non-parametric comparison of paired samples. Lastly, the VAR Model and Granger Causality Test are instrumental in analyzing the temporal dynamics and causal relationships between time series variables

Table 4 Test Results for the Impact of Positive Opening Price Difference Rate on Daily Returns

Index	t-value	Degrees of Freedom	p-value
SH Index	14.028	3890	$< 2.2e - 16$
SZ Index	16.077	3897	$< 2.2e - 16$

Table 5 Test Results for the Impact of Negative Opening Price Difference Rate on Daily Returns

Index	t-value	Degrees of Freedom	p-value
SH Index	-15.671	4114	$< 2.2e - 16$
SZ Index	-13.944	4032	$< 2.2e - 16$

Table 6 Summary of Granger Causality Tests for Opening Price Difference Rates and Daily Return Percentages in H_A and H_B

Variable relationship	F-test	p-value	Significance
<i>H_A: Positive Opening Price Difference Rate (SH Index)</i>			
shdiffrate → sh rate	3.1347	4.219e-15	***
sh rate → shdiffrate	1.0422	0.3865	
<i>H_A: Positive Opening Price Difference Rate (SZ Index)</i>			
szdiffrate → sz rate	4.9613	2.472e-10	***
sz rate → szdiffrate	1.8118	0.0242	*
<i>H_B: Negative Opening Price Difference Rate (SH Index)</i>			
shdiffrate → sh rate	3.5556	5.947e-13	***
sh rate → shdiffrate	3.2938	2.306e-11	***
<i>H_B: Negative Opening Price Difference Rate (SZ Index)</i>			
szdiffrate → sz rate	2.0999	0.005104	**
sz rate → szdiffrate	3.0019	3.04e-05	***

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 7 Regression Analysis and F-test Results for Impact of Gaps on Returns in Shanghai and Shenzhen Stock Markets

Measure	Shanghai		Shenzhen	
	Estimate	p-value	Estimate	p-value
Intercept	0.04540	.0828	0.04241	.0607
Gap Day Down	-2.48820	<.0001	-2.62272	<.0001
Gap Day Up	2.30332	<.0001	2.70468	<.0001
F-Statistic	448.15		677.25	
F-Test p-value	$< 2.2e - 16$		$< 2.2e - 16$	

1. Impact of Opening Price Difference Rate on Daily Returns

- a. H_A : Investigates the effect of a positive opening price difference rate ($diffrate > 0$) on daily returns, positing an expected discrepancy of zero in line with the

Table 8 Proportional Test Results for the Impact of Previous Day's Price Change on Next Day's diffrate Direction

Test item	Index	Proportion	Degrees of freedom	p -value
H_C Previous Day Increase Affecting Today's diffrate	SH	0.615	3890	$< 2.2e - 16$
H_C Previous Day Increase Affecting Today's diffrate	SZ	0.628	3897	$< 2.2e - 16$
H_D Previous Day Decrease Affecting Today's diffrate	SH	0.657	4114	$< 2.2e - 16$
H_D Previous Day Decrease Affecting Today's diffrate	SZ	0.645	4032	$< 2.2e - 16$

Table 9 Mean Value T-Test Results for the Impact of Previous Day's Price Change on Next Day's diffrate Direction

Test Item	t-Value	Degrees of Freedom	p -Value	Mean of Next Day diffrate
Positive Return Day (SH Index)	5.6841	4213	$7.019e - 09$	0.16676
Negative Return Day (SH Index)	-9.917	3811	$< 2.2e - 16$	-0.16671
Positive Return Day (SZ Index)	8.4251	3966	$< 2.2e - 16$	0.122132
Negative Return Day (SZ Index)	-13.269	3963	$< 2.2e - 16$	-0.16532

Table 10 Summary of Granger Causality Tests for Impact of Previous Day's Returns on the Next Day's Opening Price Difference Rate

Variable Relationship	F-Test	p -Value	Significance
<i>H_C: Impact of Previous Day's Rise (SH Index)</i>			
shdiffrate \rightarrow sh rate	2.9432	$< 2.2e-16$	***
sh rate \rightarrow shdiffrate	1.17	0.1228	
<i>H_C: Impact of Previous Day's Rise (SZ Index)</i>			
szdiffrate \rightarrow sz rate	2.8574	0.00148	**
sz rate \rightarrow szdiffrate	1.8986	0.04061	*
<i>H_D: Impact of Previous Day's Decline (SH Index)</i>			
shdiffrate \rightarrow sh rate	2.5086	$1.206e-06$	***
sh rate \rightarrow shdiffrate	2.8575	$1.835e-08$	***
<i>H_D: Impact of Previous Day's Decline (SZ Index)</i>			
szdiffrate \rightarrow sz rate	1.8465	0.02371	*
sz rate \rightarrow szdiffrate	4.2011	$8.218e-08$	***

*** $p < .001$, ** $p < .01$, * $p < .05$

Efficient Market Hypothesis (EMH). According to EMH, all available information is already reflected in stock prices, making it difficult for investors to consistently profit from trading based on past price movements. Therefore, under EMH, the expected difference between the opening price difference rate and daily returns should be zero, as any deviations would imply the presence of systematic patterns or anomalies.

Table 11 Results of Hypothesis Tests on the Distribution of Gaps Across Trading Days

Hypothesis Test Content	Index	Chi-squared	Degrees of Freedom	<i>p</i> -Value
H_E : Uniform Distribution of Upward Gaps on Weekdays (SH Index)	SH	273.39	6	$< 2.2e - 16$
H_E : Uniform Distribution of Upward Gaps on Weekdays (SZ Index)	SZ	264.23	6	$< 2.2e - 16$
H_F : Uniform Distribution of Downward Gaps on Weekdays (SH Index)	SH	206.91	6	$< 2.2e - 16$
H_F : Uniform Distribution of Downward Gaps on Weekdays (SZ Index)	SZ	205.71	6	$< 2.2e - 16$

- b. H_B : Explores the impact of a negative opening price difference rate ($diff_{rate} < 0$) on daily returns, also anticipating a zero discrepancy under EMH assumptions. Similarly to H_A , under the Efficient Market Hypothesis, the expected difference between a negative opening price difference rate and daily returns should be zero. Any significant deviations from zero would suggest potential inefficiencies or anomalies in the market that contradict the assumptions of EMH.

2. Market Reaction to News

- a. H_C : Considers the market's underreaction to positive news released before trading hours, affecting the next day's opening price difference rate. This hypothesis stems from the idea that investors may not fully incorporate positive news into stock prices during pre-market hours, leading to a subsequent adjustment in prices during regular trading hours. Factors such as limited trading volume and reduced liquidity in pre-market sessions may contribute to this underreaction phenomenon.
- b. H_D : Addresses the market's overreaction to negative news disclosed before trading hours, influencing the subsequent day's opening price difference rate. This hypothesis suggests that investors may overreact to negative news released outside of regular trading hours, causing exaggerated price movements at the opening of the next trading day. Behavioral biases, such as panic selling or herding behavior, could amplify the impact of negative news announcements during pre-market periods.

3. Distribution Patterns of Price Gaps

- a. H_E : Tests for a uniform distribution of upward price gaps ($gap=1$) across weekdays, seeking patterns influenced by systematic weekly behaviors. In financial markets, investor behavior may vary depending on specific days of the week, such as the Monday effect or Friday effect, thus this hypothesis seeks to verify if there are statistically significant differences in the distribution of opening price gaps across different weekdays.

Table 12 Distribution of Gaps by Weekday for SH (1990–2023) and SZ (1991–2023) Indexes

Weekday	SH Index		SZ Index	
	Gap -1	Gap 1	Gap -1	Gap 1
Monday	119	158	120	142
Tuesday	57	66	57	58
Wednesday	44	74	50	53
Thursday	57	70	55	51
Friday	63	74	64	55

Table 13 Results of Hypothesis Tests on the Distribution of Upward and Downward Gaps Across Months

Hypothesis Test	Index	Chi-Squared Value	<i>p</i> -Value
H_G Upward Gaps (SH Index)	SH	6.0724	0.8685
H_G Upward Gaps (SZ Index)	SZ	7.0501	0.795
H_H Downward Gaps (SH Index)	SH	12.871	0.3019
H_H Downward Gaps (SZ Index)	SZ	26.832	0.004871

Table 14 Filled and Unfilled Gaps Distribution by Month for SH(1990–2023) and SZ(1991–2023) Indexes

Month	SH Index		SZ Index	
	Gap -1	Gap 1	Gap -1	Gap 1
JAN	26	37	26	27
FEB	22	38	19	28
MAR	30	39	26	33
APR	27	42	42	34
MAY	26	40	39	34
JUE	32	33	40	33
JUL	41	32	33	26
AUG	37	39	29	25
SEP	23	28	31	21
OCT	27	40	24	33
NOV	24	42	18	34
DEC	25	32	19	31

- b. H_F : Examines the uniform distribution of downward price gaps ($gap=-1$) across weekdays, to identify weekday-specific patterns. Different weekdays in financial markets may be influenced by specific events or factors, such as interest rate announcement days or economic data release days, which may lead to deviations in the occurrence of price gaps on certain weekdays.
- c. H_G : Investigates the uniformity in the distribution of upward price gaps across different trading months, aiming to uncover monthly effects. Various months in financial markets may be influenced by seasonal factors, end-of-quarter

financial reporting, or other market events, resulting in differences in the distribution of price gaps across different months.

- d. H_H : Tests the uniform distribution of downward price gaps across trading months, determining the influence of seasonal or monthly phenomena. Different months in financial markets may be affected by seasonal factors, market sentiment, or other macroeconomic factors, potentially leading to uneven distributions of price gaps across different months.

4. Effects of Price Gaps on Trading Volume and Turnover Rate

- a. H_J : Explores the effect of upward price gaps on the trading volume and turnover rate of the subsequent trading day. In financial markets, significant price gaps may attract increased investor attention and trading activity, leading to higher trading volumes and turnover rates. This hypothesis aims to investigate whether upward price gaps have a statistically significant impact on trading volume and turnover rate.
- b. H_K : Examines the impact of downward price gaps on trading volume and turnover rate, considering potential increased selling pressure or trading activity. Downward price gaps may indicate negative market sentiment or unexpected events, potentially leading to higher trading volumes as investors react to the gap. This hypothesis seeks to determine whether downward price gaps have a significant effect on trading volume and turnover rate.

5. Methodology for Gap Analysis

- a. H_I : Defines the time required to fill gaps, detailing the methodology for calculating the average time to close a gap, supported by descriptive statistics. In financial markets, the time it takes for price gaps to be filled can provide valuable insights into market dynamics and investor behavior. This hypothesis aims to establish a systematic approach for measuring the duration from the occurrence of a price gap to its subsequent closure. Understanding the average time required to fill gaps can help investors formulate trading strategies and manage risk more effectively.

6.2 Tests Related to Opening Price Difference Rate (Difftrate)

Focusing on the historical data of the Shanghai and Shenzhen stock indices, this study delves into several aspects.

6.2.1 H_A : Impact of Positive Opening Price Difference Rate (Difftrate > 0) on Daily Returns

Null Hypothesis (H_0): The mean of daily return following a positive opening price difference rate (difftrate) equals zero.

Alternative Hypothesis (H_1): The mean of daily return following a positive opening price difference rate (diffrate) is greater than zero.

6.2.2 Analysis of H_A Hypothesis Through VAR Model and Granger Causality Tests

This study investigates the dynamic relationships between opening price difference rates and daily return percentages in stock markets, using vector autoregression (VAR) models and Granger causality tests on two distinct datasets. The first dataset (*data1*) captures the dynamics in the context of *shdiffrate* and daily return percentages for positive opening price difference rates, while the second dataset (*data2*) explores similar dynamics with *szdiffrate* and daily return percentages under the same condition. For both datasets, optimal lag lengths for the VAR models were determined based on the Akaike Information Criterion (AIC), with *data1* for *SH Index* implementing a lag of 61 and *data2* for *SZ Index* a lag of 16, reflecting adjustments to capture the dynamics more accurately in the context of positive diffrate scenarios.

VAR Model Summary: The VAR models provided a comprehensive view of the interdependencies within each stock market under the specific condition of positive opening price difference rates. The estimation results highlighted significant coefficients across various lags, indicating the predictive power of past values on current market conditions. Specifically, the model for *data1* demonstrated an intricate relationship between *shdiffrate* and daily return percentages, while the model for *data2* revealed similar insights for *szdiffrate* and its corresponding daily return percentages.

Granger Causality Tests: The Granger causality tests offered compelling evidence of the impact of positive opening price difference rates on daily return percentages. For *data1*, the causality test confirmed that *shdiffrate* Granger-causes daily return percentages with an *F*-Test value of 3.1347 and a *p*-value of 4.219×10^{-15} . However, the reciprocal causality from daily return percentages to *shdiffrate* was not found significant (*F*-Test = 1.0422, *p*-value = 0.3865). Similarly, for *data2*, *szdiffrate* was found to Granger-cause daily return percentages (*F*-Test = 4.9613, *p*-value = 2.472×10^{-10}), with the reciprocal relationship also being significant (*F*-Test = 1.8118, *p*-value = 0.0242). Instantaneous causality tests indicated significant within-period interactions between the variables in both markets, with *p*-values less than 2.2×10^{-16} .

Conclusion: The findings underscore the interconnectedness of positive opening price difference rates and daily return percentages in financial markets. The Granger causality analysis reveals a nuanced interaction where past positive diffrate values can significantly forecast future values of daily returns, particularly in the context of *data1*. These results have important implications for understanding market dynamics, informing investment strategies, and enhancing predictive models under specific market conditions. The evidence of instantaneous causality further suggests that market movements are not only influenced by past trends but also by concurrent market conditions, highlighting the multifaceted nature of stock market dynamics in response to positive opening price differences.

6.2.3 H_B : Impact of Negative Opening Price Difference Rate (Difftrate < 0) on Daily Returns

Null Hypothesis (H_0): The mean daily return following a negative opening price difference rate (difftrate) equals zero.

Alternative Hypothesis (H_1): The mean daily return following a negative opening price difference rate (difftrate) is less than zero.

6.2.4 Analysis of H_B Hypothesis Stock Market Dynamics Through VAR Model and Granger Causality Tests

Optimal lag lengths for the VAR models were meticulously selected based on the Akaike Information Criterion (AIC), with *data1* and *data2* adopting tailored lags to best capture the underpinning temporal structures indicative of how past values influence current market behavior in negative difftrate scenarios.

Granger Causality Tests: Granger causality tests reveal significant findings regarding the influence of negative difftrates on daily return percentages. For *data1*, the test indicated that *shdifftrate* Granger-causes daily return percentages with an *F*-Test value of 3.5556 and a *p*-value of 5.947×10^{-13} . Conversely, the impact of daily return percentages on *shdifftrate* was also significant (*F*-Test = 3.2938, *p*-value = 2.306×10^{-11}). Similarly, in *data2*, *szdifftrate* Granger-causes daily return percentages (*F*-Test = 2.0999, *p*-value = 0.005104), with the reciprocal relationship holding significance (*F*-Test = 3.0019, *p*-value = 3.04×10^{-05}). Instantaneous causality tests further corroborated significant interactions within the same period between the variables, with *p*-values less than 2.2×10^{-16} across both markets.

Conclusion: The analysis highlights the intricate dynamics between negative opening price difference rates and daily return percentages in financial markets. The evidence from Granger causality tests points to a significant bidirectional influence, suggesting that past negative difftrate values possess substantial predictive power over future daily returns, and vice versa. These insights are pivotal for comprehending market dynamics under specific conditions, guiding investment strategies, and refining predictive models. The pronounced instantaneous causality underscores the complexity of market movements, revealing that they are influenced by both past trends and concurrent market conditions, thus illuminating the multifaceted nature of stock market dynamics in response to negative opening price differences.

In addition to the VAR Model and Granger Causality Tests, regression analyses were conducted to further validate the impact of opening price gaps (both positive and negative) on daily returns in the Shanghai and Shenzhen stock markets. The regression models were designed to quantify the relationship between the type of gap day (upward or downward) and the consequent daily return percentages, with the aim of providing empirical evidence to support the hypotheses H_A and H_B . The results, as summarized in the table below, clearly indicate significant effects of both upward and downward gap days on daily returns across both markets. Specifically, in the Shanghai market, downward gap days were associated with a mean decrease in daily returns of 2.48820% (*p*-value <.0001), while upward gap days corresponded to a mean increase of 2.30332% (*p*-value <.0001). Similarly, in the Shenzhen market,

downward gap days led to a mean decrease of 2.62272% in daily returns (p -value $<.0001$), and upward gap days resulted in a mean increase of 2.70468% (p -value $<.0001$). These findings robustly confirm that the direction of the opening price gap significantly influences the daily return percentage, thereby providing strong empirical support for both H_A and H_B hypotheses.

This comprehensive analysis, encompassing both statistical tests and regression models, underscores the significant impact of opening price gaps on daily returns. The results not only affirm the predictive value of gap days on market behavior but also contribute valuable insights for market participants, enhancing the understanding of market dynamics and informing investment strategies under specific market conditions.

6.2.5 H_C : Impact of Previous Day's Rise on the Next Day's Opening Price Difference Rate (Difftrate)

Null Hypothesis (H_0): On days following a positive return, the proportion of days with the opening price difference rate (difftrate) moving in the same direction equals 50%.

Alternative Hypothesis (H_1): On days following a positive return, the proportion of days with the opening price difference rate (difftrate) moving in the same direction is not equal to 50%.

6.2.6 H_D : Impact of Previous Day's Decline on the Next Day's Opening Price Difference Rate (Difftrate)

Null Hypothesis (H_0): On days following a negative return, the proportion of days with the opening price difference rate (difftrate) moving in the same direction equals 50%.

Alternative Hypothesis (H_1): On days following a negative return, the proportion of days with the opening price difference rate (difftrate) moving in the same direction is not equal to 50%.

6.2.7 H_C with Mean Value Test

Null Hypothesis (H_0): For days with a positive return, the mean of the next day's opening price difference rate (difftrate) equals zero.

Alternative Hypothesis (H_1): For days with a positive return, the mean of the next day's opening price difference rate (difftrate) is greater than zero.

6.2.8 H_D with Mean Value Test

Null Hypothesis (H_0): For days with a negative return, the mean of the next day's opening price difference rate (difftrate) equals zero.

Alternative Hypothesis (H_1): For days with a negative return, the mean of the next day's opening price difference rate (diffrate) is less than zero.

6.2.9 H_C and H_D with Granger Causality Tests

Through proportional, mean value tests and Granger Causality Tests, we analyzed the relationship between the opening price difference rate (diffrate) and the previous day's price change for both the Shanghai and Shenzhen stock indices. The results indicate a significant positive correlation between the previous day's price change and the next day's diffrate direction for both indices, verified in both positive gaps following an increase and negative gaps following a decrease. Specifically, when the previous day's stock index change was positive, the mean diffrate of the next day was significantly greater than zero; conversely, when the previous day's change was negative, the mean diffrate of the next day was significantly less than zero.

6.3 Gap Nature Test

6.3.1 H_E : Test for Uniform Distribution of Upward Price Gaps Across Weekdays (Testing if Gap=1 is Uniformly Distributed Across Weekdays)

Null Hypothesis (H_0): Upward price gaps (Gap=1) are uniformly distributed across the weekdays of a week.

Alternative Hypothesis (H_1): Upward price gaps (Gap=1) are not uniformly distributed across the weekdays of a week.

6.3.2 H_F : Test for Uniform Distribution of Downward Price Gaps Across Weekdays (Testing if Gap=-1 is Uniformly Distributed Across Weekdays)

Null Hypothesis (H_0): Downward price gaps (Gap=-1) are uniformly distributed across the weekdays of a week.

Alternative Hypothesis (H_1): Downward price gaps (Gap=-1) are not uniformly distributed across the weekdays of a week.

Details of the distribution of gaps across weekdays are presented in Table 12.

6.3.3 H_G : Test for Uniform Distribution of Upward Gaps Across Trading Months (Testing if gap=1 is uniformly distributed across trading months)

Null Hypothesis (H_0): Upward price gaps (Gap=1) are uniformly distributed across the trading months (Table 13).

Alternative Hypothesis (H_1): Upward price gaps (Gap=1) are not uniformly distributed across the trading months.

Table 15 Unfilled Gaps in SH and SZ Indices (Up to November 2, 2023)

Year	SH Gaps	SZ Gaps
1991	23	3
1992	3	–
1994	3	1
1995	3	1
1996	9	7
1997	1	–
2000	–	1
2002	1	1
2005	3	–
2006	2	3
2007	1	–
2008	2	3
2014	5	1
2015	3	3
2018	1	–
2019	2	7
2020	3	1
2021	2	–
2022	–	1
2023	4	1

6.3.4 H_H : Test for Uniform Distribution of Downward Gaps Across Trading Months (Testing if $\text{gap}=-1$ is uniformly distributed across trading months)

Null Hypothesis (H_0): Downward price gaps ($\text{Gap}=-1$) are uniformly distributed across the trading months.

Alternative Hypothesis (H_1): Downward price gaps ($\text{Gap}=-1$) are not uniformly distributed across the trading months.

Details of gap distribution across months are presented in Table 14, and statistics for unfilled gaps are available in Table 15. To address the weekday and month effects observed in the price gap (H_E, H_F, H_H, H_G), we can provide explanations grounded in behavioral finance theories. For hypotheses H_E and H_F , which examine the uniform distribution of upward and downward price gaps across weekdays, respectively, the observed deviations from uniformity can be attributed to investor sentiment and trading behavior patterns. Behavioral finance suggests that investors may exhibit certain biases or preferences based on the day of the week. For example, the "Monday effect" posits that returns on Mondays are often lower compared to other days of the week, possibly due to negative sentiment lingering from the weekend. Similarly, the "Friday effect" suggests that investors may adopt a more risk-averse stance on Fridays, leading to lower volatility or smaller price gaps. These behavioral tendencies can result in non-uniform distributions of price gaps across weekdays, as observed in our findings. Regarding hypotheses H_H and H_G , which test the uniform distribution of upward and downward price

Table 16 Welch's t-test Results for Average Gap Filling Time

Metric	t-value	Degrees of Freedom	p-value	Group Mean Values
Upward Gaps vs Downward Gaps (SH Index)	-0.32257	650.61	0.7471	(Average Days to Fill Upward Gaps) 69.82025 (Average Days to Fill Downward Gaps) 75.56329
Upward Gaps vs Downward Gaps (SZ Index)	-2.3241	627.68	0.02044	(Average Days to Fill Upward Gaps) 47.71893 (Average Days to Fill Downward Gaps) 78.46847

Table 17 Summary of Gap Fill Dates by Type

Gap Type	Index	Average Fill Time (Days)	Median Fill Time (Days)	Mode Fill Time (Days)
Upward Gap (SH)	SH	69.8	6	1
Downward Gap (SH)	SH	75.6	6	1
Upward Gap (SZ)	SZ	47.7	6	1
Downward Gap (SZ)	SZ	78.5	8	1

gaps across trading months, respectively, the deviations from uniformity may stem from seasonal or calendar-related factors. Behavioral finance literature highlights the influence of investor mood and sentiment on market outcomes, which can vary across different months of the year. For example, the "January effect" suggests that stock prices tend to rise more in January compared to other months, potentially impacting the occurrence of price gaps. Additionally, seasonal factors such as holidays, earnings seasons, or macroeconomic events may influence investor behavior and trading volumes, leading to non-uniform distributions of price gaps across trading months. In summary, the observed weekday and month effects in price gaps can be explained by behavioral biases, sentiment-driven trading behavior, and seasonal factors. These explanations align with the principles of behavioral finance, which complement traditional financial theories by incorporating psychological and emotional aspects of investor decision-making.

The hypothesis test results demonstrate that both upward and downward gaps in trading days are non-uniformly distributed, indicating the presence of a calendar effect. Notably, there is a significantly higher frequency of upward gaps on Mondays. Additionally, the monthly distribution tests reveal that while upward gaps appear to be uniformly distributed across months, the downward gaps in the Shenzhen Index show significant non-uniformity, particularly with a higher occurrence in November. This might reflect market behavior and investor psychology at specific time points. These findings suggest that calendar effects

Table 18 Formation Count, Filled Count, and Fill Rates of Different Types of Gaps

Type	Formation Count	Filled Count	Upward Gap Fill Rate	Downward Gap Fill Rate
Upward Gap (SH Index)	442	316	89.37%	–
Downward Gap (SH Index)	340	316	–	92.94%
Upward Gap (SZ Index)	359	338	94.15%	–
Downward Gap (SZ Index)	346	333	–	96.24%

associated with gap phenomena should be considered in trading strategy formulation and risk management.

6.3.5 H_1 : Average Time to Fill Gaps

Null Hypothesis (H_0): There is no significant difference in the average time to fill upward and downward gaps (Tables 16).

Alternative Hypothesis (H_1): There is a significant difference in the average time to fill upward and downward gaps.

Detailed statistics on the types of gaps and their filling scenarios are presented in Tables 17 and 18, covering the operational period from 1990 to 2023 for the SH Index and 1991 to 2023 for the SZ Index.

The proposed algorithm for identifying and calculating the time required to fill price gaps in stock market data is structured as follows:

1. **Data Preprocessing:** The initial step involves augmenting the dataset with calculated columns for the high and low prices of the previous and following days. This preparation aids in the identification of upward and downward price gaps, characterized by the absence of trading activity within a specific range between consecutive trading days.
2. **Initialization of Gap Fill Metrics:** The process includes initializing two metrics within the dataset to track the process of gap closure: the number of days taken to fill the gap and the date on which the gap was filled, both set initially to indicate unfilled gaps.
3. **Iterative Gap Fill Detection:** The dataset undergoes an iterative examination to detect the presence of gaps for each trading day. For identified gaps, the algorithm further evaluates subsequent trading data to determine whether the gap has been filled.
4. **Upward Gap Closure Criteria:** A gap is considered filled if, for an upward gap, on any day following its occurrence, the day's highest price exceeds the gap's upper limit while its lowest price drops below the lower limit of the gap.
5. **Downward Gap Closure Criteria:** For a downward gap, closure is determined if, on any subsequent day, the highest price surpasses the gap's lower limit, and the lowest price falls below its upper limit.

Table 19 Descriptive Statistics of Gap-Filled After Days for All Closed Gaps, Including Both Upward and Downward Gaps for SSE Composite Index and SZSE Component Index

Statistic	SSE Composite Index	SZSE Component Index
Minimum	1.00	1.00
1st Quartile	2.00	2.00
Median	6.00	7.00
Mean	72.27	62.98
3rd Quartile	33.00	42.50
Maximum	1876.00	1749.00

- 6. Time To Fill Gaps:** The time required to fill a gap is quantified by the number of days from the gap’s appearance to its eventual closure. This measurement is recorded once a gap is identified as filled, noting the elapsed days and the specific date of closure.
- 7. Average Time Taken To Fill Gaps:** To ascertain the average time taken to fill gaps, an average calculation is performed based on the days recorded for all filled gaps. Descriptive statistics such as the minimum, maximum, mean, and standard deviation are employed to detail the distribution of the time required for gap closure, providing insights into market sentiment and price adjustment dynamics.

Additionally, it’s worth noting that the definition of gap closure provided here is tailored to the conventions and computational convenience within the Chinese mainland stock market. Different regions and markets may have varying definitions of gap closure. Nonetheless, the core principles remain consistent across different market contexts.

In our research on the average time to fill gaps in the Shanghai Composite Index (1990–2023) and the Shenzhen Component Index (1991–2023), we employed Welch’s t-test to compare the average time taken to fill upward and downward gaps.

Table 20 Median Changes in Trading Volume and Turnover for Different Gap Types in the SH and SZ Indices

Gap Type	Median Change in Trading Volume (%)		Median Change in Turnover (%)	
	SH Index	SZ Index	SH Index	SZ Index
No Gap	-2.18	-1.51	-1.96	-1.37
Downward Gap	-0.57	-0.36	-2.06	-3.43
Upward Gap	28.1	19.4	27.9	22.2

Table 21 Median Test Results for Trading Volume and Turnover Changes for Different Gap Types in the SH and SZ Indices

Gap Type	Upward Gap Trading Volume Change		Downward Gap Turnover Change	
	SH Index	SZ Index	SH Index	SZ Index
W Value	673,893	556,078	1,361,969	1,401,622
p Value	$< 2.2e - 16$	$< 2.2e - 16$	0.1814	0.02588

Table 22 T Test Results for Trading Volume and Turnover Changes for Upward Gaps in the SH and SZ Indices

Indicator	Index	<i>t</i> Value	DF	<i>p</i> Value	Difference Interval for Non-Upward vs. Upward Gaps
Trading Volume Change	SH	-4.8694	446.3	1.556e-06	[-65.6688, -27.9033]
Turnover Change	SH	-4.2341	443.57	2.791e-05	[-78.57837, -28.75721]
Trading Volume Change	SZ	-3.6366	368.3	0.0003156	[-60.03844, -17.89668]
Turnover Change	SZ	-3.1722	360.66	0.001643	[-75.89384, -17.80543]

Table 23 T Test Results for Trading Volume and Turnover Changes for Downward Gaps in the SH and SZ Indices

Indicator	Index	<i>t</i> Value	DF	<i>p</i> Value	Difference Interval for Non-Downward vs. Downward Gaps
Trading Volume Change	SH	-0.67361	359.48	0.501	[-13.900561, 6.807443]
Turnover Change	SH	0.29119	368.51	0.7711	[-8.113101, 10.933572]
Trading Volume Change	SZ	-0.49945	615.28	0.6176	[-6.678734, 3.970392]
Turnover Change	SZ	0.043135	464.1	0.9656	[-5.860116, 6.123157]

Table 24 Quantitative Trading Strategy Based on Opening Price Difference Rate (Difftrate)

Screening Criteria	Description
Stock Pool	All A-shares listed on the Shanghai Stock Exchange
Backtest Period	From November 2, 2021, to November 2, 2023
Rebalancing Frequency	Weekly
Rebalancing Weight	Based on the circulating market value proportion of stocks meeting S1 and S2 criteria
Criterion 1: S1	Stocks ranked in the 18%-30% range of the difftrate from high to low on the rebalancing day
Criterion 2: S2	Stocks ranked in the 30%-55% range of the difftrate from high to low on the previous trading day

The findings indicated that for the Shenzhen Component Index, the average time to fill upward gaps was significantly shorter than downward gaps, while for the Shanghai Composite Index, there was no significant difference between the two. This discovery suggests a divergence in market responses to gaps of different directions, particularly evident in the speed of gap filling. Additionally, we meticulously calculated the formation count, filled count, and fill rates of different types of gaps, revealing that upward gaps generally have a lower fill rate than downward gaps, especially in the Shenzhen Component Index.

6.3.6 H_J : Impact of Upward Gaps on Daily Trading Volume and Turnover Rate

Null Hypothesis (H_0): Upward gaps in the index do not significantly influence the daily trading volume and turnover rate.

Alternative Hypothesis (H_1): Upward gaps in the index significantly influence the daily trading volume and turnover rate.

6.3.7 H_K : Impact of Downward Gaps on Daily Trading Volume and Turnover Rate

Null Hypothesis (H_0): Downward gaps in the index do not significantly influence the daily trading volume and turnover rate.

Alternative Hypothesis (H_1): Downward gaps in the index significantly influence the daily trading volume and turnover rate. In our study of H_J and H_K , we chose to explore the impact of gaps on the daily trading volume and turnover rates, rather than directly examining the impact on the day's trading volume and turnover. This decision was based on the following considerations:

- **Standardized Comparison:** The change rates of trading volume and turnover provide a standardized comparison method that accurately reflects the intensity and direction of market reactions without being affected by the scale of original values.
- **Removing Trends and Seasonality from Time Series:** Change rates help mitigate the effects of long-term trends and seasonality, focusing the study on the immediate impact of events on the market.
- **Stability and Stationarity:** Change rates are generally closer to a stationary process than original values, making statistical testing and modeling more reliable.
- **Market Efficiency:** Financial markets are often considered efficient, where information is quickly absorbed and reflected in prices and trading volumes. Change rates better capture the market's rapid response to new information.
- **Risk Management:** Change rates are closely related to market volatility and risk. Analyzing change rates helps understand the market's risk response to specific events, which is crucial for risk management and investment decisions.
- **Comparative Analysis:** Change rates offer a method to compare various securities or market reactions without being influenced by the size or liquidity of individual securities.

The relevant data statistics in the H_J and H_K hypothesis tests are as follows in Tables 20 and 21, with categorized statistics for groups with and without gaps:

We explored the impact of stock index gaps on the daily change rates of trading volume and turnover. The null hypothesis posited that gaps do not significantly affect these change rates, while the alternative hypothesis suggested that gaps indeed have a significant impact. The choice to analyze change rates rather than absolute values is due to the fact that change rates offer a standardized benchmark for comparison, more accurately reflecting the market's response to gap phenomena. Additionally, change rates help eliminate trends and seasonality factors in time series data, allowing us to focus more on the immediate effects of gap events on the market.

Our analysis revealed that the average change rates in trading volume and turnover for upward gaps were significantly higher than those in situations without gaps, indicating a more pronounced market reaction to upward gaps. While the impact of downward gaps was not as statistically significant as upward gaps, it still showed a certain market response. The results of median tests and mean tests further corroborated this finding. These findings underscore the importance of gap phenomena in market microstructure and provide empirical grounding for gap-based trading strategies.

7 Strategy Simulation

Building on the established impact of positive price gaps in opening price differences, a **quantitative stock selection strategy** was formulated, detailed information can be found in Table 24. This strategy hinged on the opening price difference rate (**diffrate**), aimed at pinpointing prospective investments within the Shanghai Stock Exchange A-share market. A **two-step screening process** was implemented, consisting of screener **S1**, which shortlisted stocks within the top 18% to 30% based on **diffrate** rankings on rebalancing days, and screener **S2**, which selected stocks within the top 30% to 55% range from the preceding trading day's **diffrate** rankings. The investment domain encompassed the entirety of A-shares traded on the Shanghai Stock Exchange. The **backtesting methodology** employed was comprehensive, rigorously assessing the efficacy of the trading strategy over the period from November 2, 2021, to November 2, 2023, with a **weekly rebalancing** protocol. Portfolio allocations were determined by the circulating market capitalization of each stock. The strategy's return was meticulously computed for each rebalancing interval throughout the backtesting timeline and aggregated to reflect the overall performance through the compounded product of the periodic returns. Emphasis was placed on the exclusion of stocks under trading suspension during the selection process to uphold the **backtest accuracy**. The **cumulative return graph** delineated the comparative performance of the strategy, the benchmark, and the excess return across any selected temporal interval. Detailed records of daily profit and loss offered insights into the strategy's operational dynamics. Performance metrics, as illustrated in Fig. 3, included a total return of 31.70% and an annualized return of 14.67%, which notably surpassed the benchmark. The strategy's alpha of 0.255 signified its superior risk-adjusted return, as shown in Table 25, while a beta of 0.741 suggested a lower volatility and systemic risk profile relative to the benchmark. Risk-adjusted return measures, including a Sharpe Ratio of 0.722 and a Sortino Ratio of 1.016, favored the strategy considering overall volatility and downside risk, respectively. The Information Ratio stood at an impressive 67.788, underlining the excess return per unit of risk compared to the benchmark. The strategy's volatility was measured at 18.22%, with a maximum drawdown of -27.77%, delineating the potential loss spectrum. Additional risk assessments were encapsulated by a tracking error of 0.87% and a downside risk of 12.96%.

- **Annualized Return:** The theoretical return of the strategy over a standard one-year period. It is calculated as:

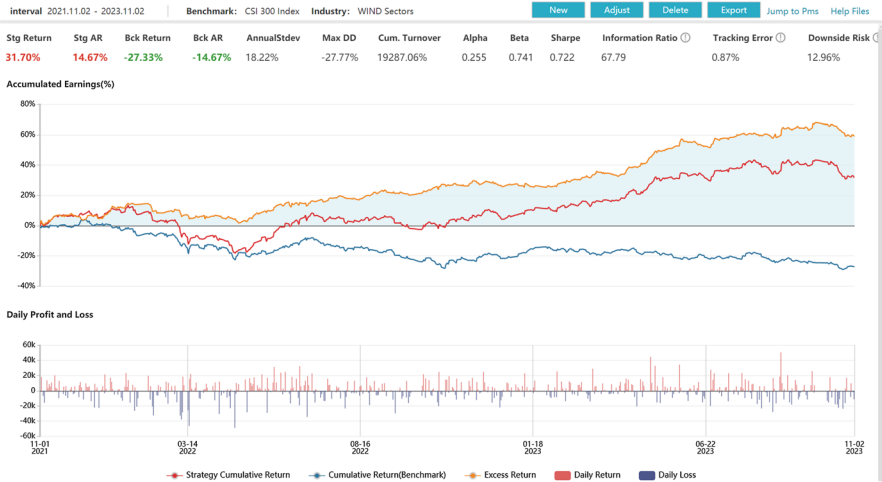


Fig. 3 Performance of Strategy

Table 25 Performance Metrics of the Quantitative Trading Strategy

Metric	Value
Strategy Return	31.70%
Annualized Strategy Return	14.67%
Benchmark Return	-27.33%
Annualized Benchmark Return	-14.67%
Alpha	0.255
Beta	0.741
Sharpe Ratio	0.722
Sortino Ratio	1.016
Information Ratio	67.788
Volatility	18.22%
Maximum Drawdown	-27.77%
Tracking Error	0.87%
Downside Risk	12.96%

$$\text{Annualized Return} = (\text{Total Return over the Period} + 1)^{\frac{260}{\text{Number of Trading Days in the Period}}} - 1 \quad (10)$$

- **Relative Return:** The difference between the total return of the strategy and the total return of the benchmark.
- **Alpha:** Reflects the part of the strategy’s return that exceeds the expected risk-adjusted return.

$$\text{Alpha} = \text{Annualized Return of the Strategy} - \text{Beta} \times \text{Annualized Return of the Benchmark} \quad (11)$$

- **Beta:** Indicates the sensitivity of the strategy's return to market fluctuations.

$$\text{Beta} = \frac{\text{Cov}(r_a, r_m)}{\sigma_m^2} \quad (12)$$

Here, r_a represents the return series of the strategy, and r_m is the market return series.

- **Downside Risk:** Measures the potential risk of the strategy's return being lower than a target level.

$$\text{Downside Risk} = \sqrt{260 \times \frac{\sum_{i=1}^T (R_i - R_f)^2}{T - 1}}, \quad R_i < R_f \quad (13)$$

Here, R_i is the daily return rate of the strategy, R_f is the annual deposit rate/365, and T is the number of trading days in the calculation period.

- **Information Ratio:** The excess return brought by each unit of active risk.

Information Ratio

$$= \frac{\text{Daily Returns of the Strategy} - \text{Daily Returns of the Benchmark}}{\text{Annualized Standard Deviation}}$$

- **Jensen's Alpha:** Represents the excess return of the strategy over the benchmark.

$$\text{Jensen} = (\mathbf{R}_p - \mathbf{R}_f) - \text{Beta} \times (\mathbf{R}_m - \mathbf{R}_f) \quad (14)$$

Here, \mathbf{R}_p is the annualized return rate of the strategy, \mathbf{R}_m is the annualized return rate of the benchmark, \mathbf{R}_f is the annual deposit rate, and **Beta** is the Beta value of the strategy over the period.

- **Maximum Drawdown:** Describes the worst-case scenario for the strategy, being the largest drop in value within a period.
- R^2 (**Coefficient of Determination**): Indicates the impact of the benchmark's performance variation on the strategy's return.

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

Here, \hat{y}_i is the predicted value of the strategy's return series, y_i is the actual return series, and \bar{y} is the average of the actual return series.

- **Sharpe Ratio:** Represents the excess return of the strategy per unit of risk.

$$\text{Sharpe} = \frac{(\mathbf{R}_p - \mathbf{R}_f)}{\sigma_p}$$

Here, R_p is the expected return rate of the strategy, R_f is the risk-free rate, and σ_p is the annualized volatility of the strategy.

- **Sortino Ratio:** Represents the excess return of the strategy per unit of downside risk.

$$\text{Sortino} = \frac{(\text{Annualized Return of the Strategy} - \text{Risk-Free Rate})}{\text{Downside Risk}}$$

- **Tracking Error:** The standard deviation of the difference between the return of the strategy and the return of the benchmark.
- **Treynor Ratio:** The excess return of the strategy per unit of systematic risk.

$$\text{Treynor} = \frac{(R_p - R_f)}{\text{Beta}} \tag{15}$$

- **Annualized Volatility:** Measures the riskiness of the strategy.

$$\text{Volatility} = \sqrt{52} \times \sqrt{\frac{\sum (R_i - \bar{R})^2}{N - 1}}$$

Here, N is the number of trading periods (weekly frequency) within the interval, R_i is the return for the corresponding period, and \bar{R} is the average return over the interval.

- **Correlation Coefficient:** Indicates the correlation between the return rates of the strategy and the benchmark.

$$\text{Correlation Coefficient} = \frac{\text{Cov}(R_a, R_b)}{\sigma_a \times \sigma_b}$$

Here, R_a and R_b are the return rate series of the strategy and the benchmark, respectively, **Cov** is the covariance, and σ is the standard deviation.

8 Conclusion

In this study, we conducted extensive hypothesis testing on the opening price difference rate (diffrate) and price gap (gap) of the Shanghai Composite Index and Shenzhen Component Index in China's mainland stock market using statistical methods. We delved into the dynamic relationships between these market variables and stock market behaviors—specifically, price, trading volume, and turnover. By testing series of hypotheses, we discovered significant correlations between the directionality of price gaps, the variability of the opening price difference rate, and certain key characteristics of the market.

Our findings offer fresh insights into unique behavioral patterns of the Chinese stock market, which have rarely been reported in international market research literature. In our comprehensive analysis, we delved into the distinctions between the Shanghai Stock Market and the Shenzhen Stock Market, recognizing that these two markets, while both pivotal to China's economic landscape, exhibit divergent characteristics that influence their respective market dynamics. The Shanghai Stock Market, often viewed as more conservative, tends to attract established, state-owned enterprises, resulting in a market composition that might exhibit different volatility patterns and investor behaviors compared to the Shenzhen Stock Market. The latter, known for its innovative and entrepreneurial spirit, hosts a significant number of high-tech and small-to-medium enterprises, potentially leading to different gap behaviors and price movements. To elucidate the variations in results observed between these two markets, we formulated several hypotheses. For instance, we hypothesized that the Shenzhen Stock Market, with its concentration of high-growth companies, would exhibit a higher frequency of price gaps due to more dynamic news flows and investor sentiments. Conversely, the Shanghai Stock Market might show a pattern of more moderate price gaps, reflective of its more stable and mature market participants. Additionally, we conducted an in-depth analysis based on historical data on the regularity of stock index gaps appearing on different weekdays (calendar effects) and their distribution across months (monthly effects), as well as the gap filling cycle, leading to a series of new conclusions and discoveries. This nuanced approach allowed us to capture the unique temporal and behavioral aspects of gap phenomena in both markets. Based on these results, we developed a stock selection strategy and validated its profit potential through historical back-testing, further confirming the reliability of our research hypotheses. This analysis not only enhances our understanding of the distinct market mechanisms at play in China's dual stock market system but also provides investors with actionable insights for navigating these nuances effectively. We are particularly interested in whether these patterns and hypotheses are equally effective in markets outside of China and the underlying economic dynamics behind them. Future work will focus on deeper analysis of the causal relationships of these new findings and exploring the universality of these patterns in global stock markets. These efforts will contribute to enriching the current understanding of the workings of global financial markets and provide more precise market strategy recommendations for investors.

References

- Aboudy, D., Even-Tov, O., Lehavy, R., & Trueman, B. (2018). Overnight returns and firmspecific investor sentiment. *Journal of Financial and Quantitative Analysis*, 53(2), 485–505. <https://doi.org/10.1017/S0022109017000989>
- Avishay, A., Gil, C., & Vladimir, G. (2023). Stocks opening price gaps and adjustments to new information. *Computational Economics*. <https://doi.org/10.1007/s10614-023-10363-w>
- Caporale, G. M., & Plastun, A. (2017). Price gaps: Another market anomaly? *Investment Analysts Journal*, 46, 279–293. <https://doi.org/10.1080/10293523.2017.1333563>
- Chi, L., Zhuang, X., & Song, D. (2012). Investor sentiment in the Chinese stock market: An empirical analysis. *Applied Economics Letters*, 19, 345–348. <https://doi.org/10.1080/13504851.2011.577003>

- Fama, E. F. (1965). The behavior of stock-market prices (Vol. 38). Retrieved from <https://www.jstor.org/stable/2350752>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work (Vol. 25).
- Fang Su, Z., Bao, H., Li, Q., Xu, B., & Cui, X. (2022). The prediction of price gap anomaly in Chinese stock market: Evidence from the dependent functional logit model. *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2022.102702>
- Guo, K., Sun, Y., & Qian, X. (2017). Can investor sentiment be used to predict the stock price? Dynamic analysis based on china stock market. *Physica A: Statistical Mechanics and its Applications*, 469, 390–396. <https://doi.org/10.1016/j.physa.2016.11.114>
- Ho, H.-W., Hsiao, Y.-J., Lo, W.-C., & Yang, N.-T. (2023). Momentum investing and a tale of intraday and overnight returns: Evidence from Taiwan. *Pacific-Basin Finance Journal*, 82, 102151. <https://doi.org/10.1016/j.pacfin.2023.102151>
- Jiang, G. J., & Zhu, K. X. (2017). Information shocks and short-term market underreaction. *Journal of Financial Economics*, 124, 43–64. <https://doi.org/10.1016/j.jfineco.2016.06.006>
- Li, C., Huang, W., Wang, W.-S., & Mun Chia, W. (2021). Price change and trading volume: Behavioral heterogeneity in stock market. *Computational Economics*, 61, 677–713. <https://doi.org/10.1007/s10614-021-10224-4>
- Plastun, A., Kozmenko, S., Plastun, V., & Filatova, H. (2019). Market anomalies and data persistence: The case of the day-of-the-week effect. *Journal of International Studies*, 12, 122–130. <https://doi.org/10.14254/2071-8330.2019/12-3/10>
- Plastun, A., Sibande, X., Gupta, R., & Wohar, M. E. (2020). Price gap anomaly in the us stock market: The whole story. *North American Journal of Economics and Finance*. <https://doi.org/10.1016/j.najef.2020.101177>
- Si, Y., & Nadarajah, S. (2022). A statistical analysis of Chinese stock indices returns from approach of parametric distributions fitting. *Annals of Data Science*, 10, 73–88.
- Si, Y., Nadarajah, S., Zhang, Z., & Xu, C. (2024). Modeling opening price spread of shanghai composite index based on arima-gru/lstm hybrid model. *PLOS ONE*, 19(3), 1–19. <https://doi.org/10.1371/journal.pone.0299164>
- Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *Review of Financial Studies*, 23, 3520–3557. <https://doi.org/10.1093/rfs/hhq052>
- Zhang, H., Tsai, W.-C., Weng, P.-S., & Tsai, P.-C. (2023). Overnight returns and investor sentiment: Further evidence from the Taiwan stock market. *Pacific-Basin Finance Journal*, 80, 102093. <https://doi.org/10.1016/j.pacfin.2023.102093>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.