LUISS

DEPARTMENT OF ECONOMICS AND FINANCE
DEGREE PROGRAM IN FINANCE

Course of Econometric Theory

PRICE DEVIATION FROM NET ASSET VALUE: THE CASE OF ASIAN TECH–ORIENTED ETFS

Prof. Santucci de Magistris Paolo

SUPERVISOR

Prof. Carlini Federico Carlo Eugenio

CO–SUPERVISOR

Di Legge Antea ID n.789431

CANDIDATE

Academic Year 2024/2025

Abstract

This research aims to investigate the short–term and long–term dynamics of price deviations from net asset value (NAV) for 19 Asian tech–oriented Exchange–Traded Funds (ETFs) listed across Hong Kong, China, South Korea, and Taiwan over 467 trading days between 2022 and 2023. This study estimates pooled OLS regressions incorporating structural characteristics (market capitalization, liquidity of underlying assets, and institutional ownership) and behavioral factors (market sentiment (VHSI), technology sector volatility, and feedback trading). The core research questions are: (1) How do structural and behavioral factors drive short–term price deviations from their Net Asset Values? (2) How are these deviations corrected over time, and what determines the speed of convergence toward the long–run equilibrium? The results reveal significant roles for structural and sentiment factors, with nonlinearities and volatility–based heterogeneity. Dynamic cointegration models (ECM and VECM) explore long–run convergence between ETF prices and NAVs at different adjustment speeds and confirm short–run mispricings. This research contributes by expanding ETF pricing research to emerging tech markets, combining static and dynamic methods, and offer practical implications for ETF issuers, market participants, and regulators.

Acknowledgments

I would like to express my sincere gratitude to all of those who have supported and motivated my ambitions. I thank my family for giving me the best possible chances and loving environment to develop my true self. To them, I owe all of the experiences, educational attainments, and happiness that have shaped my skills and person. My family is my North Star, always looking after me and guiding me towards the right path amid uncertainty. I thank my thesis Supervisor, Paolo Santucci de Magistris, for the support and precious guidance throughout the development of this thesis. I am especially thankful for his role as my Professor in Econometric Theory, through which he instilled enthusiasm for empirical analysis. I also thank my thesis Co–supervisor, Federico Carlo Eugenio Carlini, for his time and insightful advice.

Table of Contents

1 Introduction and Motivation

Exchange–Traded Funds (ETFs) have emerged as one of the fastest–growing segments in global financial markets, with assets under management (AUM) surpassing $10 trillion by the end of 2023, growing nearly eight times from 2010 (ETFGI, 2024) as shown in Figure 1. Asia–Pacific markets, although representing a smaller share, have shown remarkable growth, with Hong Kong alone reporting over $500 billion in ETF AUM, including a rising number of thematic and technology–focused funds (Hong Kong Exchanges and Clearing Ltd. (HKEX), 2023).

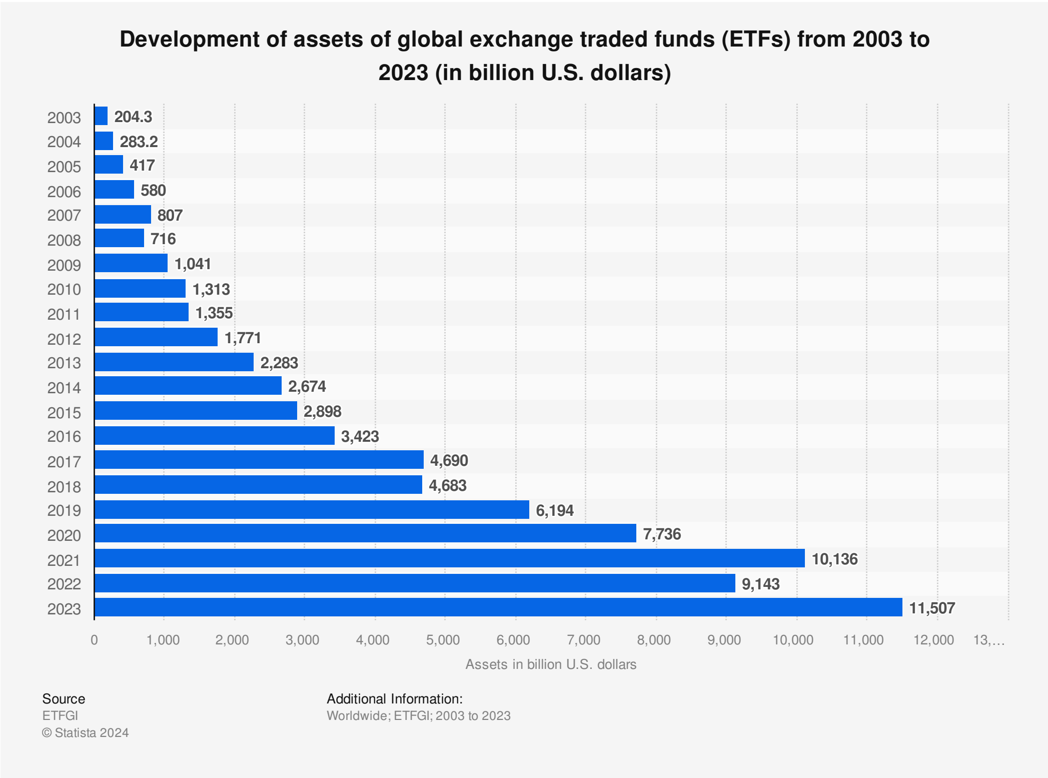


Figure 1: Global growth in ETF assets under management (AUM), 2003–2023. Source: Statista, based on ETFGI (2024).

Despite this expansion, persistent discrepancies between ETF market prices and their net asset values (NAVs) raise concerns about pricing efficiency, particularly in markets characterized by high volatility, cross–border frictions, and informational asymmetries.

The research question unfolds as follows: (1) How do structural and behavioral factors drive short–term price deviations from their Net Asset Values? (2) How are these deviations corrected over time, and what determines the speed of convergence toward the long–run equilibrium?

This study offers several contributions. It empirically demonstrates that ETF pricing inefficiencies are not only statistically significant, but also structurally rooted and behaviorally amplified in the

Asian tech segment. The results caution investors against assuming NAV proximity in high–volatility regimes or fragmented markets, while they suggest reforms aimed at harmonizing NAV disclosure for policymakers, supporting cross–border fund transparency, and mitigating retail–driven price noise. For academic research, this study extends the literature by merging structural and behavioral models in a high–frequency, region–specific context where mispricings are most prevalent.

Empirical research across markets consistently documents that ETF prices deviate from NAVs due to market and behavioral dynamics (Hilliard, 2014; Petajisto, 2017). Although these deviations—measured as $\frac{P-NAV}{NAV}$—are often small in magnitude (often below 1%), they are not negligible, as they reflect limitations in arbitrage mechanisms, delays in NAV updates, and investor sentiment swings. In March 2022, for instance, the Hang Seng TECH Index dropped 24.6% in less than three weeks (Hang Seng Indexes Company Limited, 2022), causing NAV misalignment in multiple ETFs and highlighting the vulnerability of Asia–Pacific ETFs to sudden market stress. Moreover, the Asian ETF market is structurally distinct, as it suffers from asynchronous trading hours—such as ETFs listed in Hong Kong tracking mainland Chinese assets, higher retail participation, and NAV calculation lags due to cross–border holdings, all of which compound the risk of short–term mispricing (Petajisto, 2017).

To properly interpret such deviations, it is essential to clarify the structure and pricing of ETFs. These funds are collective investment products traded on stock exchanges similarly to standard securities. These vehicles are usually designed to replicate the performance of a reference index, without seeking to generate excess returns. Under normal circumstances, ETFs offer real–time liquidity, transparency, and cost efficiency. Unlike mutual funds, which are priced only at the end of the day, ETFs can be bought or sold throughout the trading day at the prevailing market prices. The net asset value (NAV) represents the per–share value of an ETF's underlying portfolio, calculated as the total market value of assets minus liabilities, divided by the number of shares outstanding [1]. Ideally, ETF prices should track their NAVs closely due to arbitrage by Authorized Participants (APs), who create or redeem ETF shares to exploit pricing gaps.

This research focuses on an underexplored segment of the market: Asian ETFs with high exposure to the technology sector. These funds exhibit not only structural complexity but also behavioral sensitivity, as their underlying assets are often speculative, growth–oriented, and heavily momentum–driven. Their NAVs are also difficult to track in real–time, given non–standardized valuation models and geographical dispersion (Atanasova & Weisskopf, 2020; Jegadeesh & Titman, 1993). This study

[1] These definitions have been retrieved from Reilly and Brown (2011) Ch. 4.7

constructs a unique high–frequency panel dataset comprising 19 tech–oriented ETFs from Hong Kong, China, South Korea, and Singapore, spanning 467 trading days from January 2022 to December 2023. The dependent variable is the raw percentage deviation from NAV. The dataset is stacked in long–format panel including a total of 8,873 observations, with several ETFs regularly exhibiting deviations exceeding ±2%, especially during high–volatility windows. A comprehensive set of explanatory variables includes, based on contrasting views in the literature—Efficient Market Hypothesis (EMH) versus noise trader theories—structural factors and behavioral measures. Namely, the set of independent variables is composed by Market Capitalization, Liquidity of underlying assets, Percentage Institutional Ownership, Market Sentiment (VHSI), Feedback Trading, and Technology Sector Volatility.

Methodologically, this research adopts a rigorous, multi–stage econometric strategy. It begins with linear pooled Ordinary Least Squares (OLS) regressions and proceeds with nonlinear extensions incorporating squared and interaction terms, volatility–clustered subgroup analysis, and robust diagnostic testing—heteroskedasticity, autocorrelation, model specification, and stationarity. Inertia of mispricings is captured via autoregressive (AR(1)) terms, while several models employed for robustness checks study exogenous stress periods, such as the March 2022 Asian Equity selloff and the U.S. Federal Reserve's rate hike, as well as seasonality patterns.

The short–term analysis is further complemented by an investigation of long–run dynamics through Engle–Granger and Johanses trace tests for cointegration, followed by Error Correction Models (ECM) and Vector Error Correction Models (VECM) to estimate the speed of mean reversion of ETF prices toward their NAVs.

The findings are robust and multi–dimensional. Linear specifications show that Market capitalization, liquidity of the underlyings, market sentiment, and percentage institutional ownership significantly influence deviations, though their effects are often conditional and nonlinear. In particular, nonlinear extensions reveal that the individual positive effects of size and liquidity of the underlyings reverse when modeled interactively, with a positive joint effect: larger ETFs with highly liquid assets tend to exhibit larger deviations, suggesting that complexity, cross–border holdings, and behavioral trading pressures may undermine traditional arbitrage mechanisms. Market sentiment and feedback trading display significant convex and concave effects, where pricing inefficiencies are affected depending on different levels of investor uncertainty or momentum, but subsequently stabilize or revert at extreme levels, consistent with behavioral attenuation or corrective market forces.

Volatility–clustered regressions further underscore the importance of regime–specific dynamics.

In high–volatility ETFs, explanatory power increases dramatically, and the linear and nonlinear behavioral variables gain significance and exhibit larger magnitudes. Notably, in these regimes, investor behavior does not respond proportionally. Specifically, moderate uncertainty or trending behavior tends to amplify mispricings—likely due to herding of speculative activity—whereas under extreme conditions, price dislocations stabilize or correct more quickly, consistent with risk aversion or improved arbitrage mechanisms. In contrast, low–volatility ETFs exhibit weaker coefficients, lower R–squared, and reduced behavioral sensitivity, suggesting a more passive and efficient pricing structure.

Autoregressive models reveal strong inertia in pricing deviations, with the AR(1) coefficient of roughly 70% in high–volatility ETFs. This indicates that pricing inefficiencies are not quickly arbitraged away, and that past deviations significantly predict future ones. Finally, cointegration tests confirm the existence of a long–run equilibrium relationship between ETF prices and NAVs in 13 of 19 ETFs. ECM and VECM estimations reveal negative and significant error correction terms, implying that deviations are gradually corrected over time. However, the adjustment speed varies across ETFs and is notably slower in complex, high–volatility funds, reinforcing the role of frictions and investor behavior in delaying convergence.

The remainder of this research is structured as follows. Section 2 reviews the existing literature on price-NAV deviation dynamics. Section 3 presents the dataset and methodology employed in the study. Section 4 discusses the empirical findings from the pooled and dynamic regression models, as well as model diagnostics and robustness checks. Finally, section 5 concludes offering relevant implications and suggestions for further research.

2 Literature Review

This section surveys the key empirical and theoretical contributions from prior studies in the Asset Management and Pricing literature.

2.1 Efficient Market Hypothesis: rational and sentiment–driven approaches

Theoretical foundations for NAV pricing deviations can be broadly categorized into structural and behavioral explanations. Under the Efficient Market Hypothesis (EMH), originally proposed by Fama (1970), all available information is fully reflected in asset prices. In this context, NAV deviations should be arbitraged away, and any mispricings are attributed to rational factors such as liquidity constraints, market capitalization, or institutional frictions. In contrast, noise trader theory (De Long et al., 1990) and behavioral finance models argue that sentiment–driven investors, bounded rationality, and feedback trading can cause persistent deviations from NAV, especially under uncertainty or during market stress. This study incorporates both structural determinants and behavioral variables to examine the simultaneous influence of rational and sentiment–driven forces in shaping ETF mispricings. This dual–theory perspective enables the identification of different mispricing mechanisms across regimes of volatility, sentiment, and limits to arbitrage.

2.2 Pricing Efficiency

Understanding price deviations from net asset value is central to evaluating market efficiency, particularly in environments where arbitrage is limited or costly. Such mispricings represent a violation of the law of one price (Shleifer & Summers, 1990), especially in investment vehicles such as Real Estate Investment Trusts (REITs) and Exchange–Traded Funds (ETFs), where differences in asset liquidity, execution delays, and investor behavior introduce persistent valuation gaps. Consequently, academic research on price deviations between market value and net asset value (NAV) is concentrated in two contexts: REITs and ETFs, across international markets.

Kumala et al. (2024) analyze 11 Singaporean infrastructure REITs over the period 2017–2021 and find consistent trading NAV premiums. Their balanced pooled panel regressions reveal that larger market capitalization and institutional ownership reduce mispricings, while higher dividend yield increases them. Notably, liquidity and volatility show no statistical significance. Moreover, their findings support noise trading theories (De Long et al., 1990), attributing persistent inefficiencies to the role of uninformed or sentiment–driven investors, and motivate the focus of this research on how behavioral factors drive deviations in ETFs, where arbitrage is more active. Earlier foundational work by Barkham and Ward (1999) explores discounts in UK–listed property companies. Their findings provide empirical support for both agency cost and noise trader hypotheses, suggesting that deviations arise from a combination of rational asset fundamentals and irrational investor sentiment. Similarly, Liow (2003) evaluates Singaporean property stocks and finds slow mean reversion of prices toward their fundamental values, implying prolonged deviations may persist for extended periods before converging to NAVs.

In the European context, Rehkugler et al. (2012) develop a semi–rational pricing model incorporating market sentiment. Their results show that sentiment alone explains almost 80% of variation in

NAV spreads, as reflected by the model's R–squared. This framework outperforms traditional NAV–based approaches in explanatory power, enhancing the critical role of behavioral components. The authors' specification provides a foundational behavioral mechanism that is subsequently extended in ETF mispricing studies, where similar sentiment–driven deviations are observed in more liquid, Exchange–Traded Fund structures. These REIT–focused studies collectively emphasize the relevance of firm–specific characteristics, sentiment, and investor types in explaining NAV misalignments—a framework that later ETF literature adapts and expands upon.

ETFs, unlike REITs, feature a dynamic arbitrage mechanism via authorized participants (APs), who theoretically keep market prices aligned with their NAVs. However, several studies detect persistent and economically significant pricing inefficiencies—precisely the patterns this study investigates for Asian tech ETFs. Through a cross–sectional and time–series study of U.S. ETFs, Petajisto (2017) shows that absolute deviations from NAV stand at around 100 basis points, on average.[2] Their methodology combines pooled panel regressions with ETF–level and time fixed effects, using a broad sample of U.S. equity, bond, and international ETFs over multiple years. The results indicate that mispricings are not random but systematically related to factors such as ETF size, liquidity of underlying assets, and volatility. Smaller funds with illiquid or international holdings tend to exhibit daily dislocations exceeding 200 basis points, while those with more transparent and liquid components tend to display lower pricing gaps. Moreover, deviations appear more pronounced in periods of market stress or when arbitrage is temporarily impaired, aligning with theories of limited arbitrage and investor inattention. Their analysis provides a rigorous empirical perspective for understanding ETF mispricings, and it serves as the analytical foundation for this research. Building on their methodology, this study extends the literature to the Asian technology–oriented ETFs, where distortions are likely amplified due to structural features such as international NAV timing mismatches, fragmented trading venues, and lower institutional participation (Deville, 2008; Kallinterakis et al., 2020). By focusing on the Asian emerging market and on the high–growth, theme–based tech segment, the current analysis offers strong, innovative findings for ETF mispricings applied to contexts where arbitrage is more likely to break down.

Intraday deviations have received considerable attention. Madhavan and Sobczyk (2016) develop a structural model of ETF price dynamics using U.S. data from 2005 to 2014. Their estimates reveal short–term return autocorrelations of approximately -0.12 for heavily traded ETFs, with tracking

[2] See section 1 of Introduction and Motivation for the definition of percentage price deviation from NAV. The absolute value captures the magnitude of pricing inefficiencies regardless of direction.

errors—defined as the standard deviation of the return difference between the ETF and its underlying index—reaching up to 25 basis points per day in volatile conditions. These deviations differ from persistent NAV spreads, as they primarily reflect transitory pricing inefficiencies. Their findings support the use of autoregressive (AR) specifications for modeling mispricings. Further analysis on intraday dislocations is conducted by Hilliard (2014), who finds that U.S. equity ETFs exhibit a half–life of less than one day, implying rapid mean reversion. Such findings are consistent with the presence of active arbitrageurs operating in highly liquid U.S. markets, where the costs of correcting price discrepancies are minimal and information asymmetries are low. In contrast, international ETFs, especially those exposed to Asian markets, have a lower speed of adjustment and higher jump probabilities[3]. Similarly, DeFusco et al. (2011) estimate that intraday deviations for ETFs like SPDR S&P 500 ETF Trust (SPY) or Invesco QQQ Trust, although stationary, amount to 0.03% to 0.07% per hour, representing a hidden cost for active traders.

Recent literature has emphasized the unique challenges of arbitrage in fixed income ETFs, where stale prices of underlying bonds and illiquid constituent assets lead to persistent price dislocations. Pan and Zeng (2021) show that liquidity mismatches in bond ETFs significantly weaken the arbitrage mechanism. Using Balance Sheet data of APs, their study demonstrates that large bond flow shocks reduce arbitrage activity, leading to pricing errors that persist for multiple days. These findings may analogously apply to tech–oriented ETFs with illiquid or cross–listed holdings, especially in emerging or fragmented markets, or during periods of market stress—such as the equity plunge in Asian markets during March 2022. Under such conditions, impaired price correcting forces amplify mispricings.

These studies jointly provide evidence that while arbitrage enforces a degree of discipline, persistent mispricings occur under volatility or liquidity constraints—conditions central to the Asia–based tech ETFs examined in this research. However, most analyses are centered on U.S. or European markets, use low–frequency data, and often overlook the practical limitations of arbitrage observed under market stress. They largely omit Asia–focused ETFs, which face distinct frictions such as time zone lags, higher volatility, and structural barriers. This research fills the literature gap by examining high–frequency deviations in a panel of Asian tech ETFs, integrating interaction terms, nonlinearity, volatility–based comparisons, and sensitivity analysis.

[3]Jump probabilities refer to the estimated likelihood of sudden, discontinuous changes in ETF price deviations.

2.3 Determinants of Price deviations from NAVs

The existing literature consistently identifies firm–level and market–level characteristics that explain NAV pricing deviations. In the Singaporean infrastructure REIT space, Kumala et al. (2024) find that premiums are positively associated with market capitalization and institutional ownership, with coefficient estimates of 0.06 and 0.04, respectively, both significant at the 1% level. Volatility and liquidity show insignificant effects in their linear fixed effects model, while Dividend Yield exhibits a significantly negative coefficient.

Atanasova and Weisskopf (2020), employing a sample of 548 international equity ETFs in the time frame between 2012 and 2017, show that a one standard deviation increase in relative liquidity reduces absolute pricing deviations by 11–16 basis points, with the effect strongest in ETFs with high holding costs. Their results emphasize liquidity of underlying securities as a key convergence mechanism between market price and NAV. Cherkes et al. (2009) introduce a liquidity–based theory for closed–end fund mispricings, proposing that the equilibrium discount reflects the trade–off between liquidity benefits and fund costs. Kallinterakis et al. (2020) extend this framework to ETFs and empirically confirm that feedback trading intensity is significantly higher on days with large deviations. Notably, Asia–focused ETFs exhibit statistically significant behavioral trading effects, with feedback coefficients exceeding 0.20 for ETFs tracking China and Korea, highlighting the presence of return–based investor herding. Ben-David et al. (2018) exploit exogenous variation in ETF ownership and present that a one standard deviation increase in ETF ownership raises stock volatility and return autocorrelation by 17% and from -0.03 to -0.10, respectively. These traits are related to price noise, and hence, increased deviation costs.

2.4 Gaps in the Literature and Contribution

These empirical findings support the multi–factor design adopted in this research, which includes Market Capitalization, Liquidity of underlying assets, Institutional Ownership, Market Sentiment, Sector Volatility, and Feedback Trading. Each of these has documented statistical and economic significance in explaining pricing deviations in ETFs. This study extends these insights to a panel of Asian tech-oriented ETFs, with a focus on high-frequency deviation dynamics, crisis interactions, structural frictions, and behavioral components.

This research introduces a novel empirical strategy combining diagnostic testing, nonlinear pooled

OLS models, and cointegration-based dynamics tailored to the Asian tech ETF context. The application of formal tests for heteroskedasticity, residual normality, and model misspecification reveals meaningful deviations from classical linear assumptions.

Previous studies have overlooked the long-run equilibrium relationship between ETF market prices and their NAVs through the lens of cointegration analysis. This represents a significant gap, particularly in the context of Asian ETFs, where deviations may be more persistent due to liquidity constraints, market segmentation, and time zone misalignments.

This study addresses the gap in the literature by applying both Engle-Granger and johansen trace cointegration tests to assess whether prices and NAVs share a common stochastic trend. The fitting nature of prices and NAVs, which prove to be non-stationary for most of the ETFs in the sample, but show stationarity in the residuals from their linear relationship, is consistent with cointegration. This implies a valid long-term equilibrium around which short-run deviations fluctuate.

By combining panel regression techniques with cointegration and error correction models, this research offers a more comprehensive dynamic view of ETF pricing behavior. It bridges short-term inefficiencies and long-term price correction, revealing mechanisms that reduced-form deviation or volatility models may miss, especially during macro-shock periods or in illiquid Asian markets.

3 Data and Methodology

This study investigates the structural and behavioral drivers of short-term ETF pricing inefficiencies by analyzing the percentage deviation between market prices and net asset values, and how these deviations correct over time, for a panel of technology-oriented ETFs listed in Asia, particularly across Hong Kong and Shanghai exchanges.

3.1 Data Sources and Sample Selection

All data were retrieved from LSEG Workspace (Ex. Refinitiv). The initial dataset comprised a sample of 20 ETFs selected based on their portfolio exposure to the technology sector. This includes allocations in information technology, telecommunications, software, interactive media and services, entertainment, technology hardware, and semiconductors. Figures 1 to 20 in the Appendix visualize the percentage weight of each sector in ETF holdings. The sectoral breakdowns were extracted from official fund disclosures and graphically reconstructed using Microsoft Excel. The Bar Charts display uniform investments of the selected funds toward Information Technology and Semiconductors, with some allocations achieving 100% coverage in the tech sector. Building on evidence from Kallinterakis et al. (2020), who document that ETFs with geographical focus on Asia Pacific are particularly prone to feedback trading and pricing inefficiencies, this study focuses on ETFs listed on Asian Exchanges—primarily the Hong Kong Stock Exchange (HKEX) and Shanghai Stock Exchange (SSE). The sample includes high–frequency data, covering daily observations from January 4, 2022 to December 29, 2023.

3.2 Variable Definitions

The dependent variable is defined as the percentage deviation from NAV:

$$Deviations_{it} = \frac{P_{it} - NAV_{it}}{NAV_{it}} \tag{1}$$

where P_{it} is the ETF's market closing price and NAV_{it} is its net asset value. This metric captures the relative difference between the price and the net asset value of each ETF. Each observation, denoted as y_{it}, captures the percentage deviation between the market price and the NAV for ETF i on trading day t, where $i = 1, ..., N$ indexes ETFs and $t = 1, ..., T$ indexes time. This measure serves as a direct proxy for the magnitude and direction of pricing inefficiencies. Positive deviations imply that the ETF is trading at a premium relative to its NAV, while negative values reflect a discount. These dislocations may arise from a range of forces, including delayed arbitrage, information asymmetries, or temporary imbalances in supply and demand. As such, the dependent variable is not only a signal for market inefficiency, but also a dynamic outcome shaped by institutional, sentiment–driven, and volatility–linked frictions, particularly salient in fragmented and time–zone sensitive Asian ETF markets. Following the descriptive motivation, the structure of the panel dataset is formally defined. The dependent variable matrix is: $y = [y_{11}, y_{12}, \ldots, y_{1T}, y_{21}, \ldots, y_{NT}]^\top \in \mathbb{R}^{NT\times 1}$, structured as a stacked, long–form panel dataset, where each observation corresponds to an ETF–day pair, accounting for cross–sectional and temporal variation.

Consistent with the existing literature, this research employs several independent variables which exhibit explanatory power for the deviations. The variables of interest include:

Market Capitalization, which is time–invariant, expressed in billions and converted to U.S. Dollars. The expected sign of the relationship with mispricings is negative, implying that larger ETFs, in terms of capitalization, should be exposed to less price deviations from NAV due to higher availability in pricing information and higher liquidity.

Percentage Institutional Ownership, also static and expressed as percentage of shares held by institutional investors. The rationale behind the expected inverse relationship with price deviations refers to the predominance of sophisticated investors who contribute to correct mispricings and stabilize the fund's price closer to its NAV. Due to the regulatory framework in Asia, which may prevent specific company disclosures on official U.S. databases and websites, some of the % institutional ownership values were retrieved using proxies or averages across peers. Namely, the portion of institutional ownership for Bosera STAR 50 Index ETF was calculated by computing the mean across the other 19 ETFs, as valid tech–focused peers. The values for NikkoAM Global Internet and Nikko AM Metaverse Theme Active were proxied by their respective peers, also present in the sample, CSOP Hang Seng Tech Index and Global X China Cloud Computing.

Market Sentiment, proxied by daily values of the Hang Seng Volatility Index ($VHSI_t$). The Index provides a measure of the 30–day expected volatility of the Index, which reflects the listed companies in Hong Kong and China Mainland. This indicator was selected due to the prevalence of ETFs listed in the Hong Kong and Shanghai Exchanges. Sentiment can drive temporary mispricings due to behavioral biases. The expected sign of the relationship is positive, with positive market sentiment that can lead to premiums (price >NAV), whereas negative sentiment can cause discounts (price <NAV). The effect depends on the prevailing investor sentiment and market conditions. In March 2022, the MSCI's Asia Pacific Index has experienced the biggest drop of 19.4% since the 2008 financial crisis (Dogra, 2023). The drastic amount of equity outflows, driven by decreasing foreign demand and deteriorating economic outlook, has affected profoundly the Asian markets. The time frame covering March 2022 exemplifies a context in which the variable under consideration may have amplified the deviations; accordingly, it is modeled through interaction dummy variables and Difference-in-Differences as a robustness check.

Liquidity of Underlying Assets, represented by the 90–day average trading volume per ETF. The variable is time–invariant and is expected to have a negative relationship with deviations. Less liquid portfolios often result in larger NAV deviations due to the difficulty in accurately pricing these assets, leading to potential misalignments between the calculated NAV and market prices.

Volatility of the Technology Sector, a dynamic variable constructed as the 5–day rolling standard deviation of the log returns of the representative tech sector index, iShares Hang Seng Tech ETF. The returns were computed as price differences, using the market closing price, and transformed into logarithmic returns: $r_{it} = log\frac{P_{it}}{P_{i(t-1)}}$. The 5–day rolling volatility was computed for each trading day by taking the standard deviation of log returns within the backward–looking 5–day window, as follows: $\sqrt{\frac{1}{4}\sum_{j=0}^{4}(r_{t-j}-\bar{r})^2}$, where r_{t-j} is the log return on day t-j, and j= 0 to 4 is the window of the 5 most recent trading days. Given the short–term dynamics of the deviations, the 5–day window captures the time frame of the relative realized sector volatility, smoothing out daily fluctuations while remaining responsive to recent market movements. $\bar{r}$ is the average return over the 5–day time frame, computed as: $\frac{1}{5}\sum_{j=0}^{4} r_{t-j}$. The relationship with price deviations is expected to be positive, as higher volatility generally increases uncertainty and risk, which can exacerbate price deviations from NAV as market participants react to rapid changes in perceived value. Since all ETFs in the sample are tech–focused, the daily volatility was computed from the aggregated returns of the 20 ETFs.

Feedback Trading, which captures amplified trends of momentum or contrarian strategies based on behavioral market dynamics of investor reactions to past returns. Following the methodology implemented by Kallinterakis et al. (2020) measuring herd–like behavior, the variable was constructed by computing the log returns of daily prices for each of the 20 ETFs, calculated through the formula employed for the log returns of the technology sector index. The cross–sectional Pearson correlation coefficient between each log return at time t and the log return at time t-1, looped over the ETFs such that $r_t = [r_{1t}, ..., r_{Nt}]'$ and $r_{t-1} = [r_{1(t-1)}, ..., r_{N(t-1)}]'$, defines feedback trading: $\rho_t = corr(r_t, r_{t-1})$, for t = 2,…,T. The comparison of the ETF return vector on day t and on day t-1 fixes time and enables the analysis to detect any collective patterns in ETF price movements from day t-1 to day t. The expected sign is positive, as investor reactions to past price trends—rather than fundamental values—can lead to momentum effects from positive correlations and contrarian strategies from negative correlations, which may enhance premia or discounts in ETF pricings.

Due to the nature of their computations, the variables representing technology sector volatility and feedback trading exhibited missing observations. Technology sector volatility is constructed as the five–day rolling standard deviation of log returns from the iShares Hang Seng Tech Index, which is a backward–looking measure. The missing observations derive from the lack of values during the five days prior the beginning of the sample period, thus leading to the first four days in the dataset displaying $'NAN'$ values. Feedback trading is calculated from the cross–sectional correlation among returns on consecutive days, thus the first row in the dataset lacks a defined value due to the absence of lagged returns for the variable.

To maintain the balanced structure of the panel and the efficiency of the regression models, rather than dropping observations, the missing values were filled through linear interpolation over time. This approach does not compromise the integrity of the dataset, as both variables are market–level, common to the entire cross–section and time–varying, not driven by individual ETF behavior. Moreover, it does not manipulate the nature and empirical patterns of volatility and feedback trading, as it merely reflects their smooth evolution over time.

Kumala et al. (2024) find in their balanced panel data on Singaporean REITs significance of dividend yield, enhancing the explanatory power of the variable on the deviations in the context of REIT valuation, asset class which highly relies on income. Their results find a positive relationship with the premia, suggesting that a higher dividend yield might attract more investors, thus potentially driving the fund's market price above its NAV, especially if the dividends are perceived as sustainable.

While it could be worth investigating the role of dividend yield on tech–focused ETFs, its relevance and data availability have hindered the inclusion of the variable. The unfeasibility of such data stems from the negligible number of dividends ETFs may pay since the technology sector is growth–driven, prioritizing capital appreciation over dividend distribution (Fama & French, 2001; Gordon, 1959). Another issue resides with the regulatory framework imposed at the country level, which may prevent the disclosure of specific information by non–Asian platforms, such as Bloomberg or LSEG. Including proxies from the underlying indices or peers would compromise the consistency of the panel, potentially introducing noise to the model, which already incorporates high–frequency variables.

The set of variables employed in the main models is summarized as follows:

Summary of Variables and Definitions

Variable	**Definition / Calculation**
Dependent Variable	
Price Deviation	$\frac{P_{it} - \text{NAV}_{it}}{\text{NAV}_{it}}$ — percentage deviation of ETF market price from NAV.
Structural Variables	
Market Capitalization	Total market value of ETF holdings (log-transformed, in USD billions).
Liquidity of Underlying Assets	90-day average trading volume of ETF constituents (log-transformed, in USD billions).
Institutional Ownership	Proportion of ETF shares held by institutional investors (percentage).
Behavioral Variables	
Market Sentiment (VHSI)	Daily closing value of the Hang Seng Volatility Index (standardized).
Tech Sector Volatility	5-day rolling standard deviation of returns for a tech-sector index.
Feedback Trading	Cross-sectional correlation of ETF returns between day $t-1$ and t.

3.3 Visual Diagnostics and Sample Refinement

To verify the consistency of the dataset and identify any biases in pricing behavior, a comprehensive descriptive and visual screening for influential variables and outliers was conducted. Table A1 in the Appendix provides summary statistics for the initial full sample of 20 ETFs. The average price deviation from NAV is markedly negative, with the corresponding value of -4.12%, indicating a persistent tendency for ETFs to trade at a discount. This downward bias may reflect systemic inefficiencies in arbitrage mechanisms or structural limitations in pricing accuracy for certain ETFs, particularly those exposed to illiquid or segmented markets. The dispersion in deviations is notably wide, ranging from an extremely low discount approaching -90% to a maximum premium of 26%, which hints at potential structural anomalies in the data.

Table A2 in the Appendix further point to the presence of potential outliers by reporting a pronounced negative skewness of -4.08 and an elevated kurtosis of 17.82 for price deviations. These statistics indicate strong asymmetry in the distribution, characterized by heavy tails and a propensity for large deviations from the mean. Such distributional anomalies require further scrutiny, as they may distort regression estimates if not addressed through appropriate robustness checks and sample adjustments. Consequently, visual and ETF–specific diagnostics are carried out to identify and correct for these distortions.

The time series of mean price deviations from NAV, illustrated in Figure A1 in the Appendix, consistently oscillates around -0.04, reinforcing the notion of a systematic discounting pattern in the ETF market. The plot displays a marked negative spike corresponding to the month of March 2022, during which the average deviation plunges towards the minimum. To uncover the source of this pronounced downward bias, a closer inspection at the individual ETF level was conducted. Figures A2 and A3 in the Appendix plot the time series and empirical distribution of ETF–specific deviations, respectively. Among all funds, the Global X Fintech ETF (Hong Kong) emerges as a clear outlier. It displays persistent abnormal discounts, which pull the average deviation down below zero.

Further evidence of this ETF's disproportionate influence is provided in Figure A4 (Appendix), which visualizes the distribution asymmetry and extreme mispricings when the FinTech ETF is included. The presence of this single fund substantially amplifies the left–tail weight of the distribution and increases the overall leptokurtosis, in line with the previously reported skewness and kurtosis values. These statistical irregularities introduce substantial risk of biased coefficient estimates and inefficient inference in econometric models, particularly those sensitive to extreme values. Given its outsized influence, the ETF was excluded from the sample to improve stability and representativeness of the dataset.

The adjustment yields a refined panel of 19 ETFs over 467 trading days and enhances the reliability of cross–sectional and time–series comparisons without distorting the core dynamics of ETF pricing inefficiencies in the region. Figure 2 below provides a visual representation of the cross–sectional daily average deviation from January 2022 to December 2023.

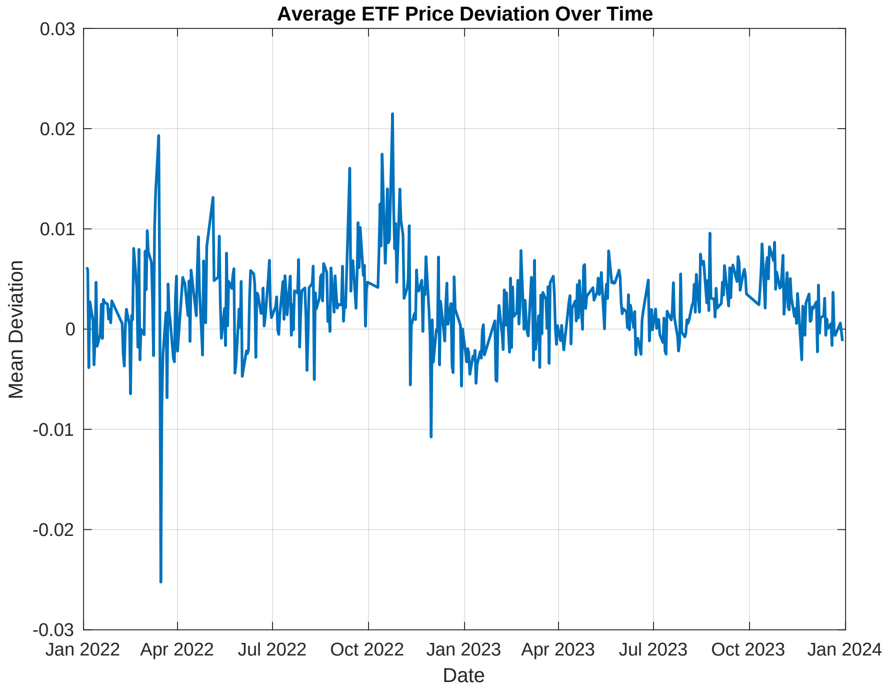


Figure 2: Time series of the average daily ETF price deviation from NAV across 19 Asian tech–oriented ETFs from January 2022 to December 2023. The chart illustrates systematic discounting patterns and stress periods like March 2022.

The average deviation now fluctuates around zero, which illustrates a more stable pattern compared to the persistent discount tendency driven by the outlier. This stability is coherent to the findings of Kallinterakis et al. (2020), who document stability in mispricing dynamics of Asian ETFs. The average deviation series—Figure A2 in the Appendix less the Global X FinTech ETF previously excluded—display a pronounced downward spike around March 2022, which captures an episode of simultaneous profound discounting across ETFs.

The presence of such a sharp distortion from the typical pricing pattern corresponds to the stock outflows and increased volatility in Asian markets during that period, leading to heightened pricing inefficiencies. This temporary breakdown in arbitrage mechanisms instigates the implementation of baseline models that investigate mispricing dynamics in volatility–driven regimes. Figure 3 presents a much more symmetrical distribution of daily price deviations from NAV across the 19 ETFs.

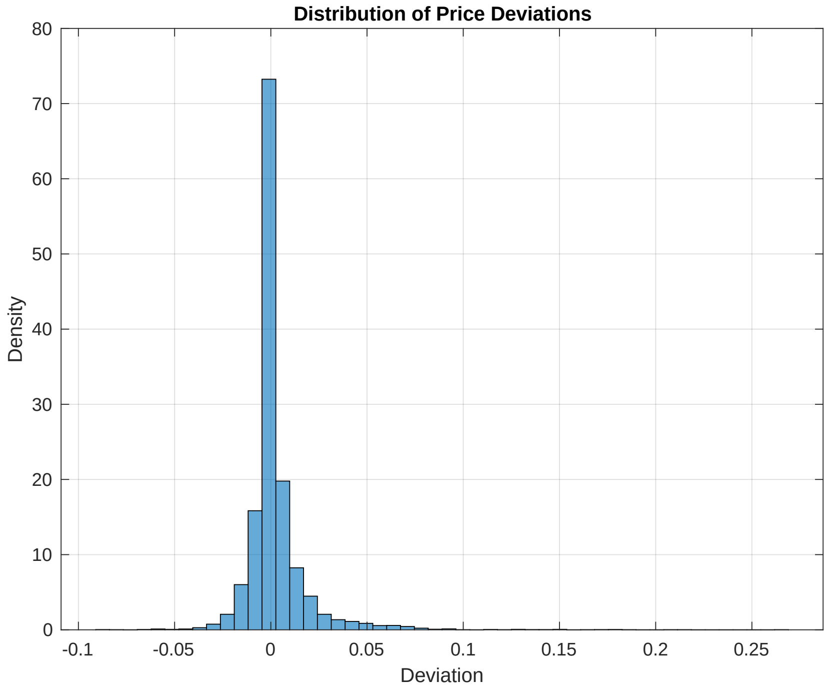


Figure 3: Histogram of daily price deviations from NAV across 19 Asian tech ETFs. The distribution shows moderate symmetry with heavy tails, suggesting persistence in mispricings and the presence of extreme deviations.

The shape of the histogram, characterized by narrow spreads, suggests that the mispricings are persistent but largely contained within a tight range, with exceptional occurrences of extreme price discounts. Nonetheless, the distribution remains heavy–tailed, which requires robust regression techniques and dynamic models which may capture higher moments of the distribution compared to static baseline models. Table A3 in the Appendix reports skewness and kurtosis for all variables. The skewness reverses direction to 4.06 reinforcing the visual difference in distributional symmetry determined by the presence of Global X FinTech ETF. Notably, the kurtosis of 39.78 indicates a non–normal distribution, heavy–tailed overall.

The descriptive inspection of higher moments, specifically skewness and kurtosis, reveal features in the distribution that baseline models might only partially capture. This analysis is conducted solely at the descriptive level and does not extend beyond second moments in the econometric modeling. This motivates, however, the adoption of nonlinear and dynamic specifications later in the analysis to better account for the underlying distributional characteristics. Table 1 reports the summary statistics for the dependent variable and the six explanatory variables, respectively. The values were computed over a balanced panel of stacked 19 ETFs over 467 trading days, for a total of 8,873 observations.

Table 1: Summary Statistics for Panel Variables (Excl. FinTech ETF)

Variable	Mean	Std. Dev.	Min	Max	N
Price Deviations	0.0025	0.0161	-0.0879	0.2641	8,873
Market Capitalization ($bn)	1.6882	1.9496	0.0055	5.8000	8,873
Market Sentiment (VHSI)	3.4282	0.5824	2.3296	6.3157	8,873
Tech Sector Volatility	0.0250	0.0138	0.0000	0.1261	8,873
Feedback Trading	-0.0581	0.3991	-0.9283	0.8644	8,873
Liquidity of underlying assets(90-day Vol, $bn)	2.2633	4.1165	0.0000	17.4379	8,873
% Institutional Ownership	0.1238	0.1777	0.0000	0.5208	8,873

Note: This table reports summary statistics for the dependent and explanatory variables used in the regression analysis. Price Deviations are unitless relative measures of ETF mispricings; market capitalization and liquidity are expressed in billion U.S. Dollars; market sentiment is expressed in index points, technology sector volatility is the daily return standard deviation; Feedback trading is constructed as cross–ETF correlation coefficient; institutional ownership is scaled as a proportion $\in [0, 1]$.

As thoroughly analyzed, both mathematically and visually, the average price deviation from NAV is 0.25%, with a standard deviation of 1.61%. These values support the exclusion of Global X FintTech ETF, which provides higher stability in the deviations, as the range significantly narrows from below, with minima of -8.79%. Contrary to the persistent average mispricings, the ETFs in the sample form a diverse set in terms of market capitalization, given the considerable variation expressed by a standard deviation of 1.94 billion U.S. Dollars. Such heterogeneity in fund size allows for higher stability in estimating the regression and infer policy implications by providing comparisons across small, medium, and large ETFs in deviation dynamics. Market sentiment, proxied by the VHSI, displays a mean of 3.43 and a relatively low standard deviation of 0.58, which indicate moderate fluctuations in perceived risk across the sample period. The iShares Hang Seng Tech Index, which captures technology sector volatility, shows a positive contained variation, with a mild average value of 0.025. Feedback trading has a mean of -0.058, which indicates that past ETF returns do not significantly predict current returns, on average. The wide standard deviation of almost 0.40, and the presence of minimum and maximum extremes achieving -0.93 and 0.86 do not exclude strong exceptions of momentum and contrarian trading patterns.

Liquidity of underlying assets, measured as 90–day average volume, mirrors a similar heterogeneity to market capitalization, with a mean of 2.26 billion U.S. Dollars, and the maximum reaching 17.44 billion U.S. Dollars, almost nine times the average. The large dispersion in values shown by the two variables gives their inclusion as controls strong validity. Percentage institutional ownership achieves a maximum value of 52.08%, setting a context of generally low institutional involvement across the ETFs. Nevertheless, the wide cross–sectional variation, ranging from almost no institutional participation to substantial portions of sophisticated investor ownership provides dispersion to detect significant effects of the variable. In recent years, ETF concentration in institutions' portfolios has grown, starting already from a balanced ownership between retail and sophisticated investors contrary to mutual funds. Despite the observed inconsistency of the values relative to the institutional ownership in the data, Galindo Gil and Lazo-Paz (2025) point out that the breadth of adoption among institutional investors of any ETF tended to have an upper bound of 70% by 2018, suggesting that full institutional ownership is not common among ETFs.

To deeply understand and better model the temporal dynamics of price deviations from their NAVs, Table 2 below reports all the dates characterized by simultaneous episodes of particularly high premiums or discounts across the entire ETF sample. The most notable clustering of strong mispricings occurs in March and October 2022, during which prices systematically diverged from their NAVs. These periods of synchronized pronounced deviations validate the need to isolate the effects of market–wide shocks, such as the March 2022 crisis, suggesting that pricing dynamics are not solely driven by idiosyncratic ETF–specific features. Further subsections provide tests and regression models that account for these patterns through the inclusion of time–based controls and interaction terms in baseline Pooled OLS models, as well as AR(1) specifications enabling temporal persistence in mispricings. There is significant dispersion across the days of strong deviations, with days of strong discounts, such as a mean deviation of -2.52% as per March 16, 2022-the day coinciding with the U.S. Federal Reserve's rate hike—and other characterized by premiums, such as +2.15% corresponding to October 24, 2022.

Table 2: Days with Strong Premia or Discounts

Date	Group Count	Mean Deviation
15-Feb-2022	19	-0.0065
11-Mar-2022	19	0.0136
14-Mar-2022	19	0.0193
16-Mar-2022	19	-0.0252
17-Mar-2022	19	-0.0076
22-Mar-2022	19	-0.0068
05-May-2022	19	0.0131
13-Sep-2022	19	0.0161
12-Oct-2022	19	0.0125
14-Oct-2022	19	0.0175
19-Oct-2022	19	0.0140
24-Oct-2022	19	0.0215
25-Oct-2022	19	0.0123
31-Oct-2022	19	0.0140
01-Nov-2022	19	0.0108
30-Nov-2022	19	-0.0108

Note: This table presents trading dates with strong synchronized price deviations from NAV across all ETFs. The table lists 16 events where mispricings significantly diverged from normal behavior, capturing regime shifts and crisis–induced frictions.

Such variability urges the adoption of model specifications which capture structural breaks in market regimes and non–linear pricing behavior, given the clear asymmetry in deviation distributions across time. While these episodic distortions across ETFs provide a solid base for time–specific modeling, including seasonality tests, lagged dependence, and interactions between crisis periods and volatility–tied variables, Table 3 quantifies the deviation volatility, measured as standard deviation, over the ETF–level dimension.

Table 3: ETFs Ranked by Deviation Volatility

ETF	Group Count	Deviation Volatility
EFundCSIOverseasChinaInternet50IndexETF	467	0.036997
SamsungKodexSemiconETF	467	0.018118
NikkoAMGlobalInternetETFUSD	467	0.017538
SamsungBloombergGlobalSemiconductorETF	467	0.014199
EFundHangSengTechnologyETF_QDII_	467	0.012388
GlobalXChinaCloudComputingETFUSD	467	0.011623
Huatai_PineBridgeCSOPHangSengTECHIdxETFQDII	467	0.0099997
InvescoGreatWallCSIHKConnectTechIndexETF	467	0.0089853
DachengHangSengTechnologyETF_QDII_	467	0.0087485
NikkoAMMetaverseThemeActiveETFUSD	467	0.008367
BoseraStar50IndexETFRMB	467	0.0067073
GlobalXChinaSemiconductorETFUSD	467	0.0051211
PremiaChinaSTAR50ETFUSD	467	0.003506
iSharesHangSengTECHETFUSD	467	0.0024695
CSOPHangSengTECHIndexETFHKD	467	0.002445
HwabaoWPCSIFintechThemeIndexETF	467	0.0023546
ChinaAMCHangSengTechIndexETFUSD	467	0.0020064
ChinaAMCSSEScience_TechInnovBoard50ETF	467	0.0016245
PenghuaCNISemiconductorChipsIndexETF	467	0.0016175

Note: This table displays the standard deviation of price deviations from NAV for each ETF over 467 trading days. The ETFs are ranked in descending order, highlighting which funds exhibited the most persistent pricing volatility.

The table ranks in descending order the degree of persistent mispricing volatility over time exhibited by each fund. There is considerable dispersion in deviation volatility, ranging from a low value of 0.16% for the bottom ETFs, all the way up to 3.70%. Notably, the ETFs exposed to higher volatility have higher allocations in internet an semiconductor segments (Figures 1 to 20 in the Appendix), which are characterized by high levels of risk and return dynamics driven by technological innovation, shifts in consumer demand, and supply chain disruptions (Ehm & Ponsignon, 2012). Conversely, the bottom–ranked ETFs, such as Hwabao WP CSI FinTech Theme Index and China AMC Hang Seng Tech index track more established and less volatile indices. The cross–sectional variation complements the time–specific pricing dynamics by identifying structural heterogeneity across ETFs that contributes to larger and frequent, or smaller and rare deviations. The observed heterogeneity in fluctuations across ETFs substantiates the incorporation of group–specific regressions that capture structural differences in ETF behavior.

3.4 Pooled Ordinary Least Squares

To investigate the short–term effect of the factors driving ETF premia or discounts in pricing, this study begins by estimating a baseline panel regression using the pooled Ordinary Least Squares (OLS) model. The dataset comprising 19 ETFs and 467 trading days was stacked in a long–form panel structure. In matrix formulation, with N = 19 and T = 467, the total number of observations becomes NxT= 8,873, with each row representing a value of ETF i on day t, where i = 1,...,N and t = 1,...,T. The corresponding matrix of explanatory variables is of NTxK dimensions, where K = 6 denotes the number of regressors, excluding the intercept. The vector of the dependent variable is of dimensions NTx1 and contains the stacked daily price deviations from NAV across all ETFs, in adherence to the structure of the matrix of independent variables. Each cross–sectional unit is observed for the same time period, thus the data set is structured as a balanced panel to enable greater flexibility in modeling heterogeneity in behavior across the sample.

The baseline econometric approach employed (OLS), is expressed as the following equation:

$$y_{it} = \alpha + \beta_1 X_{1,it} + \beta_2 X_{2,it} + \ldots + \beta_k X_{k,it} + \epsilon_{it} \tag{2}$$

y_{it} is the dependent variable, reflecting price deviations from NAV, α is the intercept, created as an NTx1 vector of ones for stability of the model. $X_{k,it}$ is the vector of independent variables, and ϵ_{it} is the idiosyncratic error term. The model relies on a set of assumptions, namely:

Linearity of the model, meaning that the relationship between the regressors and the dependent variable is linear in parameters. This assumption holds by construction, as the model is in linear form. However, potential nonlinearities in the underlying data–generating process may exist and are tested in further sections.

Strict exogeneity, which implies that explanatory variables in each time period are uncorrelated with the idiosyncratic error ϵ_{it} in each time period. In vector form, $\mathbb{E}[\epsilon_{it}|X_{it}] = 0$. This assumption is commonly violated in panel settings, especially when handling high–frequency financial datasets. While endogeneity cannot be fully excluded, the independent variables included in the model are, by construction, unlikely to be affected by ETF price deviations. Market capitalization, liquidity of underlying assets, and percentage institutional ownership are static, while market sentiment, technology sector volatility, and feedback trading are market–wide variables, thus exogenous to ETF–specific behavior.

Homoskedasticity and no serial correlation, which refer to the variance of the error term being constant across observations and the error term uncorrelated across time and ETFs, respectively: $Var(\epsilon_{it}) = \sigma^2$ and $\text{Cov}(\epsilon_{it}, \epsilon_{js}) = 0, \forall i \neq j$ or $t \neq s$. The use of robust standard errors, F–test, and RESET test embedded in the OLS function help mitigate potential violations of the assumptions. The empirical data, however, exhibit volatility clusters, heavy tails, and shifts in market conditions, as illustrated in the distributional plots and volatility ranking tables. These findings raise expectations over the violation of homoskedasticity and no serial correlation, commonly found in financial high–frequency panel data. In addition to robust standard errors, therefore, additional model specifications are employed.

Full rank of X, which translates to the absence of perfect multicollinearity among regressors. To address this assumption, the code runs a preventive correlation matrix of the explanatory variables, visually displayed in Figure 4. The correlation heatmap reports the pairwise linear relationships between the explanatory variables. The correlation matrix confirms the absence of high multicollinearity among most of the regressors. A positive, moderate correlation of $\rho = 0.598$ between market sentiment and technology sector volatility is found. Both variables capture systematic risk dimensions, which offers intuition behind the mild correlation.

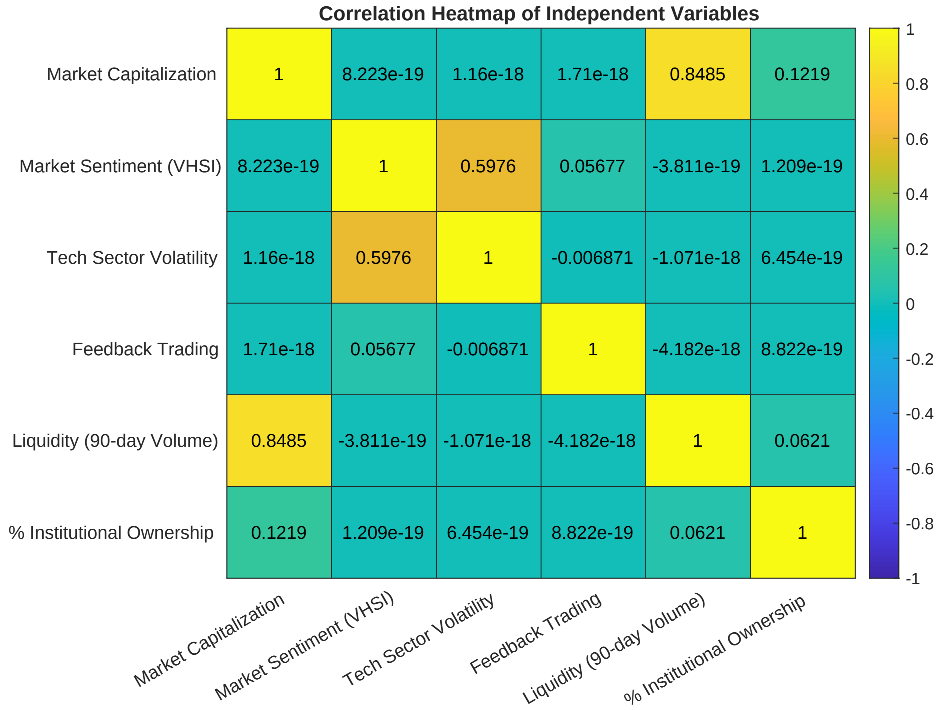


Figure 4: Correlation Heatmap of Independent Variables showing pairwise correlations among independent variables used in panel regressions. Market capitalization and liquidity show the strongest positive correlation, indicating possible structural multicollinearity.

Given the tendency of larger funds in terms of capitalization to trade higher volumes, and thus to be more liquid, the correlation between market capitalization and liquidity of underlying assets is expected to be positive and non–negligible, reflecting a ρ of approximately 0.85. Overall, most of the variables maintain low correlations with the others, reinforcing their unique contribution to the model.

Prior to estimating the model, unit root tests were conducted on all key variables to ensure the absence of spurious regression results. Tables A4 to A7 in the Appendix show the Augmented Dickey–Fuller test statistics and relative p–values for stationarity, including the dependent variable. Market capitalization, liquidity, and percentage institutional ownership are excluded due to their time–invariant nature. The rejection of the null hypothesis in all cases implies that all variables included in the baseline models are integrated of order zero (I(0)) and that the series is stationary. Therefore, the dependent and explanatory variables are suitable for levels–based panel regressions without differencing. In contrast, Table A8 shows that 13 out of 19 ETF Prices and NAVs are non–stationary, integrated of order one (I(1)). For this subset, cointegration tests and dynamic models are implemented in further sections.

In addition to the baseline linear specification, several variables are expected to exhibit nonlinear relationships with mispricings, following the existing literature and theoretical considerations.

Market sentiment, proxied by VHSI, is likely to exhibit nonlinear effects on price deviations. While mild volatility may favor price discovery and arbitrage mechanisms, excessively high sentiment levels signal market panic or exuberance, potentially amplifying price dislocations. Behavioral finance literature support this behavior in noise trader risk models (De Long et al., 1990) and in documented sentiment–driven mispricings across asset classes (Baker & Wurgler, 2007). Therefore, a quadratic term for VHSI is included to account for this asymmetry.

Feedback trading, based on momentum, may not impact mispricings uniformly. Chen et al. (2002) and Bohl and Brzeszczynski (2006) find evidence of nonlinear dynamics in feedback–based strategies, especially in varying market frameworks. To capture potentially more severe effects of strong herding or contrarian behavior than moderate momentum on price deviations, a squared term for the explanatory variable is added to the baseline model.

Market capitalization and liquidity of underlying assets, which have already been transformed in log terms, are frequently jointly associated with pricing efficiency. Ben-David et al. (2018) emphasize that ETF mispricings can be magnified when market frictions affect the underlying portfolio, especially in less liquid contexts. While large ETFs often reflect tighter deviations and higher investor attention, their efficiency may be hindered by illiquid underlying assets. Following the rationale, an interaction term between the two variables is included in the baseline model to reflect nonlinear size–liquidity complementarities. The baseline regression specification becomes:

$$\begin{aligned} Deviation_{it} = \beta_0 + \beta_1 \log(MarketCap_i) + \beta_2 VHSI_t + \beta_3 TechVolatility_t + \beta_4 FeedbackTrading_t + \\ + \beta_5 \log(Liquidity_i) + \beta_6 InstitutionalOwnership_i + \beta_7 VHSI_t^2 + \\ + \beta_8 FeedbackTrading_t^2 + \beta_9 \left[\log(MarketCap_i) \times \log(Liquidity_i)\right] + \epsilon_{it} \end{aligned} \tag{3}$$

The same terms are, where appropriate, inserted in further models to capture the potentially amplified effects, such as volatility–based estimations.

3.5 Structural and Dynamic Extensions

To account for dynamic and regime–dependent patterns in ETF mispricings, two key extensions of the pooled OLS baseline model are introduced: a structural segmentation of the sample into high– and low–volatility ETFs and the inclusion of an autoregressive (AR(1)) term.

3.5.1 Volatility–Based Specifications

The strong heterogeneity in price deviations across ETFs documented in Table 3 (Section 3.3) calls for further analysis, motivating this study to examine whether mispricings are differentially driven by volatility exposure. The sample was partitioned into two subgroups, based on the standard deviation of price dislocations. The grouping process was conducted using the standard deviation of the ETF–specific mispricing time series. The ETFs in the top decile constitute the "High–volatility" group, while those in the bottom decile form the "Low–volatility" group. Two separate OLS regressions are estimated for each subgroup, using equation (2) in section 3.4, which includes the baseline independent variables of interest. The comparison between the results serves the purpose to unveil the variation in explanatory power of structural and behavioral variables across ETFs with inherently different volatility levels in their price deviations.

3.5.2 Autoregressive term

While the baseline models control for market–wide risk and behavioral factors, serial correlation is ETF mispricings is likely to arise due to persistent trading patterns, investor behavior, or microstructure frictions. Petajisto (2017) and Madhavan and Sobczyk (2016) find sticky price deviations due to arbitrage frictions and delayed price discovery by authorized participants, which can induce lagged relationships between prices and their NAV. Cross–market trading constraints and liquidity mismatches also contribute to persistent dislocations, especially in international ETFs, as it is the case of the sample under study. Ben-David et al. (2018) and Engle and Sarkar (2006) support prior results by introducing momentum or contrarian strategies and herd–like behavior as further drivers of these dynamics. These types of investor behavior may arise for ETFs targeting theme–based sectors, such as technology.

To control for lagged dependencies in ETF pricing inefficiencies, an AR(1) extension of the OLS model is implemented through the introduction of an autoregressive term capturing the dependent variable lagged by one period: y_{it-1}. This specification allows the model to reflect delayed correction mechanisms or mispricing persistence. To implement the lag structure consistently in the panel dataset, the original stacking order was modified by restructuring the date–first to ETF–first stacking. This procedure allows for the correct alignment of y_{it-1} within each ETF's time series, ensuring that the autoregressive term reflects true temporal persistence rather than cross–sectional spillovers. This refinement is crucial for valid AR(1) estimation in panel specifications, as misaligned lag structures can induce spurious autocorrelation or misrepresent the dynamic behavior of ETF price deviations. The model becomes:

$$y_{it} = \alpha + \rho y_{it-1} + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \epsilon_{it} \tag{4}$$

where ρ represents the autoregressive coefficient measuring potential path dependence in price deviations. Given the high likelihood of serial correlation, additional testing involves the AR(1) term in the subgroup comparison as model enhancement. The variable is estimated to assess the difference in autoregressive dynamics between the two groups, allowing the specification to capture the magnitude and persistence of pricing inefficiencies within each volatility context, as well as to compare the coefficients for practical inference on structural differences in pricing mechanisms depending on volatility levels.

3.6 Cointegration and Error Correction Models: ECM and VECM

This subsection contributes to the research by investigating the existence of long–term relationships between prices and their NAVs by applying cointegration techniques across selected ETFs. Further estimation of short–term dynamics via Error Correction Models and Vector error Correction Models are applied. The research conducted by Petajisto (2017) on U.S.–listed ETFs across various asset classes find that deviations exhibit mean–reversion, indicating that while ETF prices do deviate from their NAVs in the short–term, they tend to return to the NAV over time. The convergence is faster for ETFs characterized by higher liquidity, which is also tested in the context of Asian tech–oriented ETFs, given the insertion of the explanatory variable in this study. Due to the theoretical link between Prices and their NAVs through arbitrage mechanisms, the expected long–term co–movement of the two is worth analyzing.

To estimate the cointegrating relationships between prices and the corresponding NAVs, the data for the respective variables was structured as two matrices of dimensions TxN, respectively: P_{ti} and NAV_{ti}, with $t = 1, \ldots, 467$ and $i = 1, \ldots, 19$. Only ETFs whose price and NAV series are both non–stationary in levels are eligible for cointegration analysis. The I(1) condition is necessary to avoid spurious regression and to satisfy the premise of the Engle–Granger methodology. The ADF test results from Table A8 in the Appendix show that thirteen ETFs are integrated of order one and were therefore retained for cointegration testing. The first approach consists in the implementation of the Engle–Granger test, which unfolds in two steps, looping over the selected ETFs:

1. Estimation of the long–run relationship through the OLS regression of Prices on NAVs, as follows:

$$P_{it} = \alpha_i + \beta_i NAV_{it} + \upsilon_{it} \tag{5}$$

2. Test of the residuals, υ_{it} for stationarity through the ADF test. If the unit root null hypothesis can be rejected, stationarity of the residuals is confirmed, which implies that there is evidence of cointegration between the ETF–specific price and NAV. The short–term premia or discounts, therefore, are overcome by the long–run mean–reversion.

To account for potential multiple cointegrating relationships and generalize the results from the previous test, the Johansen trace test was applied. The test consists in bivariate vector constructions for each of the 13 ETFs, as follows:

$$Y_t = \begin{bmatrix} P_t \\ NAV_t \end{bmatrix} \tag{6}$$

Cointegration is tested under the H1 model, which assumes the absence of linear trends in the data and the presence of the intercept in the cointegration relationship. The second assumption implies that the long–run means can deviate from zero. The H1 specification, performed with up to three lags, was selected for this analysis. The goodness of fit of the H1 model stems from several reasons. First, the test aims solely to detect mean–reversion in deviations, not trends. Moreover, the intercept included in the cointegration equation captures any systematic deviations due to the factors under consideration as explanatory variables for short–term mispricings. This means that the model is allowing the long–term deviations to be centered around non–zero levels.

Given the verification relative to the existence of cointegration from the two tests, the Error Correction Model (ECM) detects short–term deviations from the long–term co–movement between prices and NAVs. The equation takes the following form:

$$\Delta P_t = \alpha + \beta \Delta NAV_t + \gamma EC_{t-1} + \epsilon_t \tag{7}$$

ΔP_t and ΔNAV_t represent the short–term deviations, EC_{t-1} expresses the lagged error from equation (5) in the first step of the Engle–Granger test, which measures the degree of short–term mispricings from the long–term cointegration at time $t-1$. The corresponding coefficient γ is the speed at which prices return to co–move with their NAVs. A strong, negative γ provides evidence of mean–reverting behavior, which implies that short–term deviations are followed by prices adjusting back to NAVs. The expected sign for the coefficient attached to changes in NAVs, β is positive, as NAV changes should affect price movements. Conversely, the lagged cointegration error, γ, is expected to be negative, implying mean–reversion in mispricings.

To add robustness in the panel framework, Vector Error Correction Models (VECM) are implemented across the 13 ETFs. The framework captures, through the inclusion of lagged variables, the predictive power of past dynamics in prices or NAVs on future price changes–short–run feedback effects. The model is specified as follows:

$$\Delta P_t = \alpha + \beta_1 \Delta \mathrm{P}_{t-1} + \beta_2 \Delta NAV_{t-1} + \gamma EC_{t-1} + \epsilon_t \tag{8}$$

The equation presents the same lagged error correction term and coefficient of adjustment speed as the ECM from the Engle–Granger test. the coefficients attached to the differences of prices and their NAV, β_1 and β_2, measure the feedback effects previously mentioned. In particular, β_1 captures the magnitude of momentum or reversal in price dynamics; respectively, a positive value implies prices at time t tend to maintain the same direction as prices at $t-1$, while a negative value suggests subsequent mean–reversion behavior. β_2 measures the extent to which past NAV movements affect future prices. A positive β_2 shows that future prices move in the same direction of past NAV changes, while a negative β_2 implies that future prices assume opposite directions of past NAV movements. The expected signs for β_2 and γ parallel the ECM ones, supporting the notion of lagged price adjustments to NAVs and mean–reversion. The sign of β_1 depends on feedback trading and liquidity of the underlyings, which show weak momentum with exceptional positive spikes and high liquidity but large dispersion in trading volume, respectively. Given the absence of patterns in return momentum, the coefficient is expected to be weakly positive, which is conducive of weak momentum in price movements.

4 Results and Discussion

This section presents the empirical results for each of the models employed to investigate ETF pricing deviations. The analysis begins with a baseline pooled OLS model and proceeds through a series of model extensions—including nonlinear specifications, volatility–based subgroup regressions, and autoregressive (AR(1)) terms to capture persistence—motivated by diagnostic testing. The findings consistently remark the relevance of both structural—market cap, liquidity of the underlyings, institutional ownership—and behavioral factors—market sentiment, tech sector volatility, feedback trading—in driving ETF mispricings via marginal and interaction effects.

Significant heterogeneity emerges across high– and low–volatility ETF groups, suggesting regime–dependent pricing dynamics. The results further reveal that mispricings exhibit strong daily persistence, with approximately 70% of deviations carrying over from one day to the next. This inertia suggests that pricing errors are not quickly arbitraged away, and thus are only partially corrected in the short run, as adjustment is hindered by structural frictions and behavioral dynamics that slow down price reversion toward NAV.

To address the long–term convergence of ETF prices to their fundamental values, the short–term analysis is complemented by cointegration techniques—Engle–Granger and Johansen trace tests—followed by ECM and VECM models, which quantify the adjustment process toward equilibrium and confirm the presence of systematic mean reversion.

4.1 Baseline pooled OLS

Table 4 reports the regression output for price deviations. To evaluate whether the inclusion of explanatory variables significantly improves model fit relative to a constant–only specification, an F–test was performed. The null hypothesis that all slope coefficients equal zero is strongly rejected at the 1% level, confirming that the regressors jointly explain a significant portion of the variation in price deviations. This justifies the use of a full pooled OLS model over a baseline with only an intercept. The model estimates price deviations using robust standard errors to account for potential heteroskedasticity, as identified in diagnostic tests.

To exclude any trace of linear dependence between the variables, a check for Variance Inflation Factors (VIF) was conducted, in addition to the correlation matrix displayed in Figure 4. The VIF analysis reinforces the absence of high multicollinearity, since all values are far below the threshold of 10 and in fact do not reach 4. Table A9 in the Appendix reports the actual VIF values. There are significant scale differences between variables. Market capitalization and liquidity of the underlyings are expressed in billions U.S. Dollars, with maxima exceeding 17 billion USD, while technology sector volatility and percentage institutional ownership fluctuate in much smaller ranges, with standard deviations below 0.02 and 0.18, respectively. To mitigate these huge scaling differences, log transformations were applied to market capitalization and liquidity of the underlyings. Even after the re–scaling, large magnitude differences persist, with the logged variables ranging between 14 and 22, and small–scale variables such as VHSI, feedback trading, and tech sector volatility range between 0.01 and 0.3.

Table 4: Pooled OLS Regression Results (Dependent Variable: Price Deviations)

Variable	Coefficient	Robust SE	t-stat	p-value
Intercept	-0.0270	0.0025	-10.6279	0.0000***
Log(Market Capitalization)	0.0009	0.0001	8.7768	0.0000***
Market Sentiment (VHSI)	0.0026	0.0005	4.9803	0.0000***
Tech Sector Volatility	-0.1013	0.0197	-5.1482	0.0000***
Feedback Trading	0.0002	0.0004	0.5783	0.5631
Log(Liquidity, 90-day Volume)	0.0002	0.0000	3.5111	0.0004***
% Institutional Ownership	0.0176	0.0011	15.4842	0.0000***
Observations				8,873
R-squared				0.0851
F-statistic p-value				< 0.0001
Ramsey RESET p-value				< 0.0001

Notes: Robust standard errors are reported. Stars denote significance at the $^{***}p < 0.01$, $^{**}p < 0.05$, and $^{*}p < 0.1$ levels. The dependent variable is the daily deviation of ETF market price from NAV. Log transformations were applied to Market capitalization and liquidity to reduce scale imbalance.

The Pooled OLS regression results show statistical significance for all of the explanatory variables at the 1% confidence level, except for feedback trading. The R–squared is 8.51%, which indicates that the model explains approximately 8.5% of the variation in price deviations. The value fits into similar asset pricing econometric models dealing with high–frequency, panel data. The intercept is also significant and negative, implying that, holding all the other variables constant, the predicted baseline price deviation from NAV stands at a discount of 2.70%. However, the variables in the model cannot all simultaneously equal zero and exhibit vastly different scales; market capitalization and liquidity of underlying assets are always positive, since even the smallest ETF exhibits non–zero capitalization. Therefore, an economic inference cannot be associated to the significance of the intercept, which may otherwise capture deviations that are not fully explained by the regressors. Although larger ETFs are expected to exhibit higher stability in pricing and thus less deviations, the variable shows a positive, statistically significant β. This relationship can invert when contextualized in a framework of international, tech–oriented ETFs. Petajisto (2017) highlights that ETFs with non–domestic holdings tend to be exposed to larger price deviations from NAV due to potential cross–border pricing delays, which hinder real–time valuation of their underlyings. The ETFs in the sample exhibit misalignment in trading hours with the underlying assets. Several funds track indices or sectors whose assets trade in markets with different time zones and/or currencies, while the ETFs are listed in domestic markets; some of the ETFs include Qualified Domestic Institutional Investor (QDII) structures, which leads to lagged information in pricing. These structures reduce the immediacy of arbitrage and delayed information transmission.

The technology industry in which the ETFs selected for this study heavily invest is particularly subject to pricing inefficiencies since it is a rapidly evolving sector. As such, limited historical data and not well–defined asset valuation may enhance mispricings. ETFs such as Bosera STAR 50 Index and ChinaAMC SSE Science and Tech Innovation Board 50 hold high–growth stocks, while Nikko AM Metaverse Theme Active ETF invests in the corresponding emerging sector, which relies on future–growth and speculative asset valuation models. The semiconductor industry is also a common target among the ETFs and it is characterized by cyclical demand. Atanasova and Weisskopf (2020) support the discussion by documenting that assets lacking standardized valuation models affect ETF susceptibility to premia or discounts. Rehkugler et al. (2012) offer an additional perspective on the inverse relationship between market capitalization and price deviations, implementing a semi–rational pricing model that substantially improves with sentiment–driven speculation. The high–growth, emerging themes held by the mentioned ETFs attract investors based on current trends, visibility, and momentum, rather than their intrinsic value.

Liquidity of underlying assets also shows inconsistency with expectations, displaying a positive, statistically significant coefficient. Specifically, higher trading volume in the underlying assets are associated, ceteris paribus, with larger price deviations from NAV. Market analyses have reported that speculative retail activity plays a determinant role in Asian tech–focused ETF mispricings, with surges in the number of retail investor trading in Asian theme–based ETFs (Zhen & Li, 2025). As discussed in the previous paragraph relative to market capitalization, the sectoral and geographical features of the ETFs in this context fit the logic of a positive relationship between liquidity of underlying assets and mispricings. In this context, even liquid underlying assets are challenging to value, sensitive to high–volatility and theme–based trading, which could justify the positive sign attached to the coefficient.

Kallinterakis et al. (2020) detect an increase in premia or discounts for ETFs investing in assets that present sentiment–driven trading and volatile features. Hilliard (2014) draws attention to the international component of ETF holdings, leading to misaligned cross–border absorption of information and therefore inhibited arbitrage. Engle and Sarkar (2006) identify barriers to arbitrage mechanisms, which include transaction costs and regulations in a number of Asian markets which limit access to international exposure. The relatively negligible magnitude of the coefficients for market capitalization and liquidity of underlying assets stems from the very large scale of the variables, expressed in billion U.S. Dollars, while the corresponding low standard errors may partly reflect the sample size of 8,873 observations and wide dispersion of these variables. Nonetheless, the economic interpretation of these coefficients remains valid and meaningful for policy implications, especially given the counterintuitive sign of both variables.

The coefficient attached to percentage institutional ownership does not conform to expectations, as it presents a positive sign with statistical significance at the 1% confidence level. A one–unit increase in institutional ownership is associated with an increase in price deviation from NAV of 1.76%, ceteris paribus. The relationship implies that higher institutional involvement amplifies pricing inefficiencies, contradicting the existing literature, which identifies institutional investors as stabilizing actors of pricing dynamics (Kumala et al., 2024). One potential explanation is that institutional investors themselves may engage in speculative activities and theme–focused trading strategies in emerging segments and sectors exposed to strong growth narratives, such as technology, which often attract momentum–based positioning. Ben-David et al. (2021) find that increased ownership by large institutions predicts noise and higher volatility in stock prices, leading to wider inefficiencies in mispricings. Moreover, in the context of this research, the average share of institutional ownership across the 19 ETFs is approximately 12%, with a maximum of 52%, suggesting their influence is insufficient to counteract inefficiencies. The underrepresentation of sophisticated investors diminishes the expected corrective pressures on prices, supporting the positive relationship observed in the regression results.

Out of the three market–wide variables, only market sentiment and technology sector volatility are significant. Market sentiment (VHSI), which captures anticipated volatility in the Hong Kong market, retains a positive coefficient, consistently to expectations. Although it may have not accurately reflect idiosyncratic behavior across all Asian markets where some ETFs are listed, the attached coefficient suggests that a unit increase in VHSI is associated, ceteris paribus, with an increase in mispricings of 0.26%. The findings enhance the notion that higher volatility expectations, measured by the indicator, increase ETF mispricings. These findings contribute to the broader research conducted by Rehkugler et al. (2012) and Kumala et al. (2024). The authors cover funds investing in Real Estate, with respective focus on Europe and Singapore. Both studies incorporate market sentiment in their model, enriching the relevance of the variable by considering non–rational investors under the noise trader framework. They suggest that short–term divergences are reinforced by herd–like investor behavior and speculative trading. These findings are exceptionally meaningful in this research, as the ETFs in the sample show sector–specific and geographical characteristics which expose them to noise trading. Given the elevated values displayed by market sentiment, between 15 and 35 during the full sample period, its hindering effect on pricing efficiency is informative, especially during market stress, such as the March 2022 crisis or the rate hike from the Federal Reserve. Dogra (2023) emphasizes the non–linearity of sentiment effects, which may only manifest during periods of market crashes. Following their findings, this study presents nonlinear extensions of the baseline model and comparable volatility–based OLS regressions.

Technology sector volatility is unexpectedly negative and large in magnitude. Rather than enhancing mispricings, a one–unit increase in volatility during the past 5 days is associated with a 10.13% reduction in price deviation, ceteris paribus. Theoretically, higher volatility generally increases uncertainty and risk, which can exacerbate price deviations from NAV as market participants react to rapid changes in perceived value. There is contrasting evidence from the existing literature with respect to the effect of the variable. The research conducted by Petajisto (2017) produces results that confirm the hypothesis according to which ETFs with assets that are traded in volatile markets tend to have larger and more persistent premiums. However, one plausible justification for the opposite direction in the relationship reflects discounting behavior. Various studies document mean–reverting behavior in ETF pricing. Milani and Ceretta (2014) find periods of high volatility to instill risk perception and thus amplify caution in price tracking, leading to higher correlation between ETF market returns and NAV returns during these times. DeFusco et al. (2011) expand the perspective by delving into microstructure theory, which refers to strategic arbitrage exploiting volatility–induced inefficiencies, leading to increased mispricing profitability. In volatile contexts, investors may also refrain from trading based on themes and narratives, as extensively discussed in previous sections, with market participants relying on fundamental valuation of NAVs, thus reducing the divergence in pricing. To test the viability of the computation for tech sector volatility as a five–day rolling standard deviation for short–term dynamics, this study also estimates ECM and VECM models, which capture volatility effects over long–term cointegration. The risk in using 5 days refers to anticipated market reaction to volatility expectations over those days, leading to the variable misrepresenting short–term pricing inefficiencies.

Feedback trading is not significant at any conventional level. Kallinterakis et al. (2020) find strong effects in feedback trading for ETFs with Asia Pacific as geographical location. Their research detects particularly amplified effects during momentum–driven episodes and price correction periods. To address the lack of significance of the variable in the static linear model introduced as baseline framework, the analysis is extended to nonlinear specification in the subsequent section, as well as cointegration–based models (ECM and VECM), to better capture dynamic and structural pricing behavior across time and ETF characteristics.

4.2 Diagnostic testing and Model Extensions

Given the panel structure and behavioral nature of the high–frequency dataset, the presence of heteroskedasticity, non–normal residuals, autocorrelation, and model misspecification could bias standard inference and misrepresent the explanatory power of the independent variables.

Table 5 condenses the key statistics for model specification and assumption checks. The reported p–values for the Breusch–Pagan and White tests suggest strong rejection of the null hypothesis of homoskedasticity, confirming the presence of non–constant error variance across observations. The descriptive statistics already anticipated such result, with variation in the explanatory variables across ETFs and over time. The use of robust standard errors in all regressions accounts for group–wise heteroskedasticity and intra–panel correlation.

To verify the assumption of residual normality, the Jacque–Bera test was applied. The high test statistic and p–value < 0 lead to the rejection of normality in residuals, reflecting the asymmetric distribution and large tails of ETF mispricings documented by the skewness and kurtosis values in section 4.1. The outcome implies the violation of the OLS assumption, the large sample size ($N = 8{,}873$) allows the exploit of asymptotic properties under the Central Limit Theorem, which ensures inference to remain valid under robust estimation frameworks.

Table 5: Diagnostic Tests for Model Assumptions

Test	Test Statistic / Value	p-value / Conclusion
Breusch–Pagan Test (Heteroskedasticity)	LM = 2129.29	$p < 0.0001$ (Reject H_0)
White Test (General Heteroskedasticity)	LM = 2129.29, df = 34	$p < 0.0001$ (Reject H_0)
Jarque–Bera Test (Residual Normality)	JB = 80,229	$p < 0.0001$ (Reject H_0)
Durbin–Watson Statistic (Autocorrelation)	DW = 1.927	No first-order autocorrelation
Hausman Test (FE vs RE)	χ^2 = 1304.81, df = 5	$p < 0.0001$ (Reject H_0)
Ramsey RESET Test (Functional Form)	F = 912.86	$p < 0.0001$ (Reject H_0)

Notes: All diagnostic tests confirm the presence of heteroskedasticity and non–normal residuals, justifying the use of robust standard errors. The RESET test suggests potential misspecification, addressed through interaction and dynamic terms. The Hausman test confirms Fixed Effects as the preferred panel specification.

The Durbin–Watson statistic closely aligns with the ideal benchmark of 2, suggesting no strong first–order autocorrelation in the pooled OLS residuals. However, dynamic extensions in subsequent sections explicitly model potential mispricing persistence.

The Hausman test is grounded in the comparison between Random Effects (RE) and Fixed Effects (FE) estimators after running separate panel regressions using a common specification. Both FE and RE coefficients are consistent under regressor orthogonality with the error term, while only the FE estimators remain consistent under ETF–specific heterogeneity. The test statistic and low p–value suggest the rejection of the null hypothesis, suggesting the preference of Fixed Effects over the Random Effects model. This strongly highlights that the explanatory variables are not exogenous with respect to ETF–specific characteristics. While this suggests FE would be appropriate, all subset and nonlinear regressions are estimated via pooled OLS with robust errors, given model simplicity and subgroup comparability. FE models are incorporated as robustness checks supporting the main findings.

The Ramsey RESET test strongly rejects the null hypothesis of correct functional form, signaling potential nonlinearity or interaction effects not captures in the linear specification. This supports the inclusion of nonlinear and interaction terms in later models.

To explore potential sources of model misspecification and violations of the classical assumptions, residuals from the baseline pooled OLS specification were analyzed graphically and by subsample. Figure 5 displays the overall quantile–quantile (Q–Q) plot of standardized residuals, while Figures A5 and

A6 in the Appendix display ETF–level histograms and Q–Q plots of regression residuals, respectively.

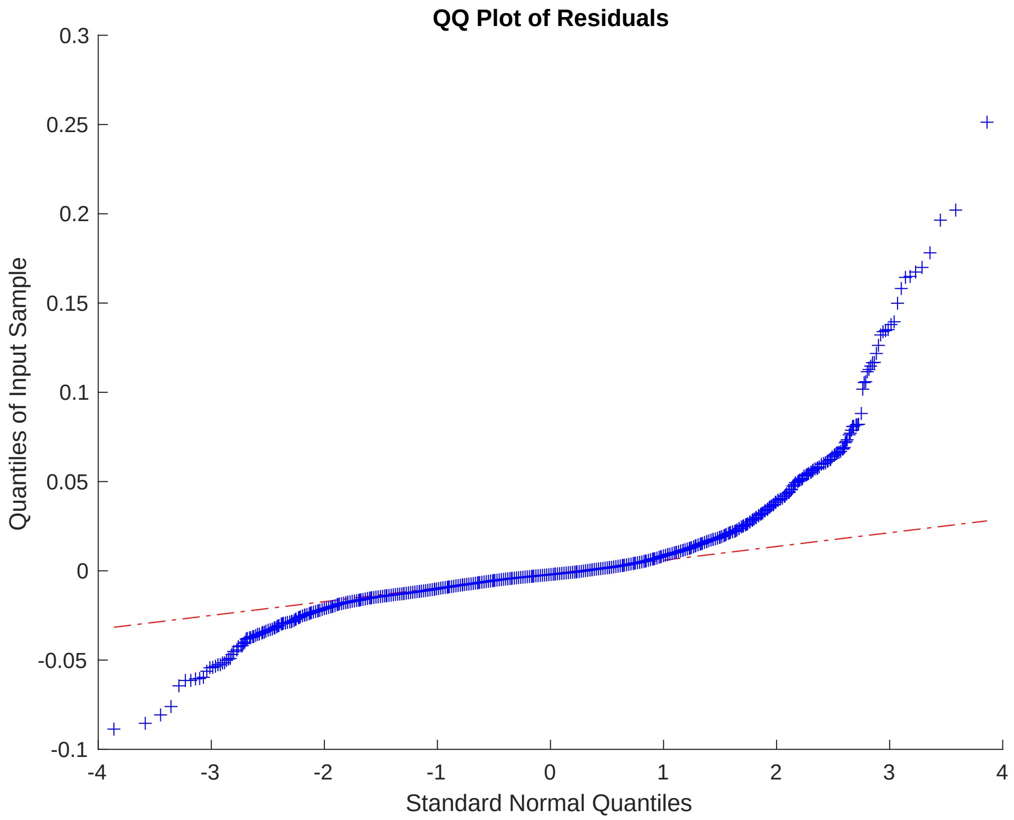


Figure 5: Q–Q Plot of Pooled Regression Residuals. The plot compares the standardized residuals from the baseline OLS model to a normal distribution. The deviations from the 45–degree line indicate fat tails and non–normality.

The plots reveal that deviations from normality are not driven by a single ETF, but rather are widespread across funds, with several series showing excess kurtosis and asymmetric distributional patterns consistent with pricing discounts during market stress periods. To further identify sources of nonlinear residual behavior, Figure A7 plots residuals against key predictors. These scatterplots reveal slight curvature and variance clustering in relation to market sentiment (VHSI), tech sector volatility, and feedback trading intensity. These patterns reinforce the inclusion of nonlinear terms, interaction effects, and subsample models to better capture ETF mispricing dynamics.

To test for structural differences in pricing behavior, coefficient estimates from subsample regressions are reported in Tables A10–A12 (Appendix). These pooled OLS models are estimated separately for periods of high vs. low market sentiment, high vs. low feedback trading, and pre– vs. post–March 2022. These results show that sensitivity to behavioral and volatility–related variables increases during periods of heightened stress. For instance, the coefficient attached to tech sector volatility is more than three times larger in magnitude in the low–feedback regime, and the magnitude of the feedback trading term significantly intensifies in the post–March 2022 period. These findings support the hypothesis that mispricing dynamics are nonlinear and conditionally dependent on market environment and investor behavior. Table A8 in the Appendix present the Augmented Dickey–Fuller (ADF) test on all 19 ETFs, which aims to verify the integration order of prices and NAVs. The results suggest that 13 out of 19 ETF Prices and NAVs are I(1), meaning that the majority of series are non–stationary in levels, but become stationary in first differences. The integration of order one between prices and their fundamental values satisfy the crucial requirement for applying cointegration techniques. Together, these tests reinforce the data modeling power of the employed approaches and the robustness of empirical findings.

The diagnostic results also guide the structural and dynamic refinements applied to the empirical framework and provide the rationale for deeper investigation in mispricing dynamics through error correction models.

4.2.1 Nonlinear Terms

Motivated by the diagnostic test results, the baseline Pooled OLS linear model was modified with quadratic terms for VHSI and feedback trading, and an interaction term between market capitalization and liquidity of the underlyings. Separate regressions were run to identify the unconditioned effects of each nonlinear transformation. Table A13 in the Appendix shows the side–by–side results, which represent a solid base for the simultaneous replacement of linear variables with the better fitting regressors.

While the addition of quadratic terms alone do not improve the model, as their coefficients are not significant, the inclusion of the interaction term between market capitalization and liquidity of underlying assets reveals notable sign and significance switches. Compared to the baseline OLS results in Table 4, both coefficients attached to market cap and liquidity of the underlyings turn negative, while their interaction yields a positive sign. This sign pattern reversal is crucial and supports the original expectations of the hypothesized relationships with mispricings, aligning with the theoretical and empirical literature. While larger ETFs and higher liquidity in fund portfolios independently narrow price deviations, their positive joint influence implies that large and liquid ETFs tend to exacerbate price efficiencies. This departs from traditional assumptions and indicates that very large and highly liquid ETFs may be subject to trading frictions, market pressures, or strategic investor behavior that undermine the price–NAV link (Petajisto, 2017).

Market sentiment maintains its negative sign and significance in Model 3, consistently to the idea that higher uncertainty, especially in emerging markets and high–risk sectors, leads to more fundamental–based trading, narrowing the price gap with their fundamental values through better arbitrage mechanisms.

Technology sector volatility displays a strong, positive coefficient, contrary to its linear counterpart, indicating that higher volatility in the technology sector leads to higher pricing inefficiency when the interaction term between market cap and liquidity of the underlyings is accounted for. This sign flip strengthens the interpretation of complex environments driving higher pricing noise or discouraged arbitrageurs, as the price–correcting mechanisms become riskier or less effective in large, liquid funds that are exposed to volatile sectors (Madhavan & Sobczyk, 2016).

Feedback trading, which is consistently insignificant in the baseline linear model, as well as the models with the VHSI and the regressor itself in quadratic terms, gains high significance. The positive coefficient is coherent with expectations, associating higher momentum or herding behavior with wider price dislocations. This supports behavioral finance studies such as Barberis et al. (2005), who find that trend–following investors can destabilize prices, especially in turbulent markets where arbitrage mechanisms are impaired.

Notably, percentage institutional ownership loses significance, implying that its already small marginal impact is absorbed by the interaction term. Nonetheless, its direction remains positive, hinting that higher sophisticated involvement may widen deviations, likely reflecting concentration risks or timing effects among institutional flows. Overall, these nonlinear results demonstrate that mispricings are shaped by complex interdependencies among individual structural and behavioral factors, and they justify moving beyond a purely linear framework.

Based on the nonlinear extension formalized in Equation (3) in the Methodology section, a re–estimation of the baseline model incorporating the quadratic and interaction terms was performed. Table 6 displays a side–by–side comparison with the original linear results.

Table 6: Comparison of Linear and Nonlinear Pooled OLS Regression Results

Variable	Linear OLS	Nonlinear OLS
Intercept	-0.0270^{***}	-0.0004
	(0.0025)	(0.0024)
Log(Market Capitalization)	0.0009^{***}	-0.0000
	(0.0001)	(0.0001)
Market Sentiment (VHSI)	0.0026^{***}	-0.0012^{***}
	(0.0005)	(0.0003)
Tech Sector Volatility	-0.1013^{***}	-0.0892^{***}
	(0.0197)	(0.0105)
Feedback Trading	0.0002	0.0012^{**}
	(0.0004)	(0.0005)
Log(Liquidity, 90-day Volume)	0.0002^{***}	0.0000
	(0.0000)	(0.0001)
% Institutional Ownership	0.0176^{***}	0.0006
	(0.0011)	(0.0009)
VHSI^2	—	0.0001^{**}
		(0.0001)
Feedback^2	—	-0.0019^{*}
		(0.0010)
MCap × Liquidity	—	0.0000^{***}
		(0.0000)
Observations	8,873	
R-squared	0.0851	0.0546
F-statistic p-value	< 0.0001	< 0.0001
RESET p-value	< 0.0001	< 0.0001

Notes: Robust standard errors in parentheses. Stars denote significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Nonlinear model includes squared VHSI, squared feedback trading, and interaction term between market capitalization and liquidity. Linear model applies log transformations to market cap and liquidity.

While the R–squared slightly improves, the inclusion of nonlinearities reveals important shifts in coefficient magnitude, direction, and statistical significance across key regressors.

Market capitalization and liquidity lose significance, while their interaction is highly significant and positive, confirming the findings in Model 3 (Table A13). The coefficient for market sentiment switches sign to negative, while its square displays the opposite direction. This convex relationship clarifies the interpretation of its positive coefficient in the linear model: deviations narrow as uncertainty rises moderately, while they amplify under extreme sentiment, consistent with market destabilization at higher volatility levels.

Feedback trading maintains its positive sign and gains significance, while its quadratic term is negative and weakly significant. This suggests that mispricings initially widen as feedback behavior intensifies, but eventually reverse or level, consistently to behavioral models of momentum or herd–like behavior followed by mean reversion or pricing correction. This finding is particularly relevant for the implementation of dynamic models capturing behavioral drivers of pricing efficiency over the long run. Overall, nonlinear specifications reveal richer dynamics and capture structural behavioral shifts that linear counterparts miss.

4.2.2 Subgroup Analysis: High vs Low–Volatility ETFs

Following the time–specific comparison rationale in the previous section, the sample was partitioned in two groups, based on cross–sectional standard deviation of the price dislocation series. The marked difference between the top decile and bottom decile displayed in Table 3 (section 3.1), as well as the variation in average deviations across ETFs in Figure A6 in the Appendix, motivate a closer analysis of whether mispricings are differentially driven by volatility exposure. Table 7 provides the results for the comparative estimation of ETFs that consistently exhibit high and low levels of mispricings, run from a common linear specification introduced in section 3.4 and nonlinear extension with the baseline explanatory variables first, and the added squared and interaction terms after. The linear results show substantial contrasts between high–volatility ETFs versus those characterized by more stable price deviations. The first striking difference is in the explanatory power of the models, which drastically declines from a 25% R–squared in the high–volatility specification, to a mere 0.3% value for the low–volatility group.

Table 7: Pooled OLS Regressions by ETF Volatility Group: Linear vs Nonlinear

Variable	Linear		Nonlinear	
	High-Volatility	**Low-Volatility**	**High-Volatility**	**Low-Volatility**
Intercept	-0.0307^{***}	0.0035^{***}	0.6505^{***}	-0.0150^{*}
	(0.0068)	(0.0008)	(0.0416)	(0.0090)
Log(Market Capitalization)	-0.0010^{***}	-0.0003^{***}	-0.0315^{***}	0.0005
	(0.0003)	(0.0000)	(0.0015)	(0.0005)
Market Sentiment (VHSI)	0.0082^{***}	0.0000	−0.0205	0.0022^{***}
	(0.0016)	(0.0001)	(0.0137)	(0.0008)
Tech Sector Volatility	-0.2123^{***}	−0.0022	-0.2357^{***}	−0.0001
	(0.0570)	(0.0045)	(0.0546)	(0.0044)
Feedback Trading	0.0010	−0.0001	0.0003	−0.0001
	(0.0013)	(0.0001)	(0.0012)	(0.0001)
Log(Liquidity)	0.0019^{***}	0.0001^{***}	-0.0697^{***}	0.0009^{*}
	(0.0001)	(0.0000)	(0.0032)	(0.0005)
% Institutional Ownership	0.0475^{***}	0.0002	-0.0460^{***}	0.0003^{*}
	(0.0024)	(0.0002)	(0.0039)	(0.0002)
$VHSI^2$	—	—	0.0039^{*}	-0.0003^{***}
			(0.0020)	(0.0001)
$Feedback^2$	—	—	−0.0032	0.0000
			(0.0029)	(0.0002)
MCap × Liquidity	—	—	0.0034^{***}	−0.0000
			(0.0002)	(0.0000)
Observations	2,335	2,335	2,335	2,335
R-squared	0.2595	0.0232	0.4179	0.0295
RESET p-value	< 0.0001	0.2337	< 0.0001	< 0.0001

Notes: This table compares pooled OLS regressions run separately for the five highest and five lowest volatility ETFs in the sample. Robust standard errors in parentheses. Nonlinear models include quadratic terms and Market Cap × Liquidity interaction. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The High–volatility group exhibits highly significant coefficients for almost all the explanatory variables, except for feedback trading. The negative coefficient for the intercept suggests a baseline discount in prices for this group. Market sentiment is highly significant, positive, and strong in magnitude, demonstrating heightened sensitivity to systemic risk within the high–volatility group. The interpretation is the following: a unit increase in VHSI is associated with an increase in price deviations by 0.82%, holding everything else constant. Technology sector volatility also exerts a large, significant effect, although negative. This implies that as the sector's volatility rises, mispricings narrow in highly volatile ETFs, likely due to higher uncertainty and fundamental–based trading rather than theme–driven investing.

In contrast, pricing corrections by sophisticated investors, proxied by percentage institutional ownership, tend to enhance price dislocations for this group, probably due to persistent temporary frictions. The same explanation is valid for the positive coefficient associated to liquidity of the underlyings. Market capitalization shows a negative, significant coefficient across both models, where a 1% increase in size is associated with a 0.0010% decrease in price deviations, holding everything constant. This suggests that larger funds experience smaller dislocations, with more attenuated effects for ETFs with lower volatility. The low—volatility group exhibits significance only for the coefficients attached to the intercept—which displays opposite sign, suggesting baseline premia—market capitalization, and liquidity of the underlyings. The latter regressor loses magnitude, while the other variables lose significance.

The comparison provides strong evidence that high–volatility ETFs are more strongly affected by behavioral and market–wide forces, while low–volatility ETFs either experience much smaller deviations in magnitude, or less systematically predictable price deviations. The nonlinear comparison between volatility groups reveals crucial interpretative reshaping of price deviation determinants, especially in high–volatility ETFs. The R–squared between the high–volatility subgroups nearly doubles to 42%, and several signs and magnitudes shift across coefficients. The intercept flips from negative to positive, suggesting that in the scenario where all variables are null, after accounting for nonlinearities, the highly volatile ETFs display average premia rather than baseline discounts. Market cap's coefficient increases in magnitude, indicating its stronger role in narrowing mispricings once conditional structures are included. Market sentiment switches sign to insignificantly negative, while its squared term is weakly positive, revealing a convex relationship where deviations widen only at high levels of market sentiment. Technology sector volatility follows the same pattern as the linear model shows, while liquidity of underlying assets switches sign and its interaction with size turns positive, suggesting that large, liquid ETFs show widened mispricings when considered jointly, though their individual effects dampen deviations. Institutional presence for high–volatility ETFs shows for the first time in this research price–correcting mechanisms through effective sophisticated trading. The nonlinear model for the low–volatility group shows weaker and minimal behavioral distortions. Notably, market sentiment becomes significant and positive, compared to both the linear counterpart and the nonlinear opposite group. Its squared term is highly negative and significant, implying that deviations increase moderately with uncertainty but revert at extreme sentiment levels.

Overall, the nonlinear specifications unveil distinct mechanisms by which patterns such as convex and concave behavioral patterns and joint size–liquidity frictions are essential to model extreme deviation dynamics accurately, especially under periods of market stress and ETFs exposed to high volatility.

4.2.3 AR(1) specifications

To isolate the persistence effect relative to ETF price deviations, the baseline model includes the addition of the dependent variable, lagged by one, represented by the AR(1) term. Table 8 reports the regression results. The coefficient attached to the autoregressive term is positive, highly significant, and very large in magnitude. This suggests a substantial degree of persistence in mispricings, as previous deviations predict current ones. In terms of economic interpretation, almost 70% of past deviations tend to persist in the next trading day. The finding is coherent to the hypothesis that ETF prices do not immediately converge to their fundamental values, but exhibit inertia and gradual correction, likely due to frictions in information absorption or liquidity constraints.

Out of the baseline regressors, only market capitalization and percentage institutional ownership are significant, while the loss of significance for other variables may be due to the dominating effect of the lagged term, which absorbs much of the short–run autocorrelation in mispricings. The model explains approximately 53% of mispricings, given the strong R–squared. The presented results contradict the Durbin–Watson test, which fails to detect residual correlation, enhancing the relevance of the autoregressive term in the model.

Table 8: Pooled OLS Regression with AR(1) Term

Variable	Coefficient	Robust SE	t-statistic	p-value
Intercept	-0.0060^{***}	0.0019	−3.1887	0.0014
Log(Market Capitalization)	0.0003^{**}	0.0001	2.4772	0.0133
Market Sentiment (VHSI)	0.0000	0.0003	0.1221	0.9028
Tech Sector Volatility	−0.0166	0.0209	−0.7954	0.4264
Feedback Trading	0.0002	0.0003	0.7220	0.4703
Log(Liquidity, 90-day Volume)	0.0000	0.0000	1.0063	0.3143
% Institutional Ownership	0.0054^{***}	0.0008	6.8725	0.0000
AR(1) Lagged Deviation	0.6960^{***}	0.0239	29.1387	0.0000
Observations		8,873		
R-squared		0.5257		
RESET p-value		< 0.0001		

Notes: This table presents pooled OLS regression results with an autoregressive term capturing the lagged effect of ETF price deviations. Robust standard errors are shown in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

To capture potential persistence in different volatility regimes, the autoregressive (AR(1)) term was introduced in the volatility–based analysis across ETFs presented in the previous model. Table 9 presents results for the extended framework. The term aims to uncover whether the degree of persistence in price deviations varies in high– and low–volatility ETFs. By isolating the lagged effect of mispricings within each subgroup, the model captures the magnitude and temporal stickiness of pricing inefficiencies. The coefficients in the high–volatility group align with the baseline results reported in Table 8, as well as the explanatory power of the model. The AR(1) term loses part of its significance in the low–volatility group, consistent with the weak explanatory power in the subset in Section 4.2.2. The strong AR(1) coefficient reveals that price deviations are conditionally persistent, highlighting explicit dynamics not captured by static diagnostics.

The AR(1) specifications confirm that short–term ETF price deviations are highly persistent, particularly among high–volatility ETFs. These findings underscore the gradual nature of price correcting processes, potentially driven by information frictions or arbitrage constraints.

Table 9: Pooled OLS Regressions with AR(1): High- vs Low-Volatility ETFs

Variable	**High-Volatility Group**	**Low-Volatility Group**
Intercept	-0.0059	0.0032^{***}
	(0.0053)	(0.0008)
Log(Market Capitalization)	-0.0003	-0.0002^{***}
	(0.0003)	(0.0000)
Market Sentiment (VHSI)	0.0014	0.0000
	(0.0010)	(0.0001)
Tech Sector Volatility	-0.0607	-0.0021
	(0.0586)	(0.0045)
Feedback Trading	0.0007	-0.0001
	(0.0010)	(0.0001)
Log(Liquidity)	0.0006^{***}	0.0001^{***}
	(0.0001)	(0.0000)
% Institutional Ownership	0.0165^{***}	0.0002
	(0.0024)	(0.0002)
AR(1) Lagged Deviation	0.6535^{***}	0.0883^{**}
	(0.0306)	(0.0432)
Observations	2,335	2,335
R-squared	0.5709	0.0308
RESET p-value	< 0.0001	0.1196

Notes: This table presents pooled OLS regressions for high– and low–volatility ETF groups, including an AR(1) term. High–volatility ETFs show stronger autoregressive persistence. Standard errors in parentheses. Robust SEs are used. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The pooled OLS regressions and their extensions—through nonlinear terms, volatility subgroups, and autoregressive components—collectively reinforce the hypothesis that ETF price deviations from NAV in the Asian tech sector are shaped by a complex mix of structural, behavioral, and temporal factors. While baseline regressions reveal that almost all key variables significantly influence mispricings, their impact varies substantially across regimes and specifications. In the baseline model, behavioral variables exhibit contrasting effects, while structural factors amplify mispricings.

Nonlinear regressions capture convex and concave sentiment effects—where dislocations initially widen with uncertainty but stabilize at high sentiment levels—and uncover diminishing returns to feedback trading. Size and liquidity of the underlyings reverse in sign in presence of their positive joint effect, which gains high predictive power of larger deviations. Subsample regressions reveal that ETF mispricings are more predictable and significantly larger in volatility–prone ETFs, reinforcing the relevance of fund–specific structural features and investor behavior under stress. The AR(1) extension further highlights the persistence of dislocations, with approximately 70% of prior–day deviations carried over into the next trading day. The inertia is particularly pronounced in high–volatility ETFs, suggesting frictions in arbitrage or price adjustments.

These findings demonstrate that short–term ETF mispricings are not only significant and highly persistent, but also conditionally nonlinear and regime–dependent. While pooled OLS models and their extensions provide valuable insights into the immediate drivers of price–NAV spreads, they are inherently limited in capturing long–term convergence dynamics and in explicitly modeling arbitrage efficiency. To address these limitations, the analysis extends to cointegration and error correction models, which are designed to test for long–run equilibrium relationships between ETF prices and NAVs, and to capture the short–term deviations and adjustment dynamics toward that equilibrium. These models allow for a deeper understanding of the frictions, delays, and asymmetries that constrain the arbitrage mechanisms intended to align ETF prices with their fundamental values.

4.3 Long–run Adjustments: Cointegration

This section extends the research of short–term factors affecting price deviations to long–term assessments of price-NAV movement, or possibly, co–movement. This section aims to answer the second part of the research question, particularly: do ETF prices ultimately revert back toward their NAVs, or do deviations persist over the long–term equilibrium?

This dynamic approach unfolds into two phases: first, the Engle–Granger and Johansen trace tests are implemented to check for cointegration relationships between prices and NAVs. If the tests demonstrate long–term co–movement despite short–term deviations, Error Correction Models and Vector Error Correction Models are applied to quantify the direction and speed of adjustment of momentary mispricings.

4.3.1 Engle–Granger and Johansen Trace Tests

Following the methodology described in subsection 3.6, two cointegration frameworks are run to test for the existence of long–term convergence between prices and NAVs. Table 10 shows the results for the ETF–specific cointegration dynamics from the Engle–Granger test, Table 11 presents the findings in a pooled panel framework from the Johansen trace test. Based on the ADF test results presented in TableA8 in the Appendix, the tests apply to the 13 eligible ETFs exhibiting non–stationarity in both prices and NAVs. The test statistic at the individual level is negative, and the values are substantially below the critical threshold, as column **'Reject Null'** displays. The results imply that the residuals are stationary for all 13 selected ETFs, thus indicting that ETF prices co–move with their NAVs in a stable, long–term relationship. The three–lags H1 model applied for the Johansen trace test corroborates the ETF–level cointegration by providing test statistics in all the lags exceeding the critical thresholds and near–zero p–values. These results lead to the rejection of the null hypothesis of no cointegration at the panel level, supporting the theory of ultimate price convergence toward their NAVs.

Table 10: Engle–Granger Cointegration Test Results (Selected ETFs)

ETF	**Test Statistic**	**Reject Null**
Bosera STAR 50 Index ETF (RMB)	−12.317	Yes
CSOP Hang Seng TECH Index ETF (HKD)	−21.504	Yes
ChinaAMC Hang Seng Tech Index ETF (USD)	−19.246	Yes
ChinaAMC STAR 50 ETF	−17.708	Yes
Dacheng Hang Seng Tech ETF (QDII)	−15.463	Yes
Global X Cloud Computing ETF (USD)	−19.323	Yes
Global X Semiconductor ETF (USD)	−15.359	Yes
Hwabao FinTech Theme ETF	−20.282	Yes
Invesco Great Wall Tech ETF	−13.372	Yes
Penghua Semiconductor Chips ETF	−18.773	Yes
Premia China STAR 50 ETF (USD)	−14.001	Yes
Samsung Global Semiconductor ETF	−21.647	Yes
iShares Hang Seng TECH ETF (USD)	−21.705	Yes

Notes: ADF test on residuals from Price-NAV regressions. Rejection of the null indicates stationary residuals, implying cointegration.

Table 11: Johansen Trace Test for Cointegration (Pooled Panel)

Lags	r	Statistic	Critical Value	p-value	Eigenvalue
1	0	114.8798	15.4948	0.0010	0.2079
	1	6.4915	3.8415	0.0110	0.0139
2	0	83.1740	15.4948	0.0010	0.1532
	1	6.0401	3.8415	0.0141	0.0129
3	0	54.6276	15.4948	0.0010	0.1002
	1	5.7227	3.8415	0.0169	0.0123

Notes: Johansen trace test conducted under H1 specification with lags 1 to 3. Trace statistics significantly exceed critical values, indicating cointegration in all specifications.

The joint evidence from the cointegration tests firmly confirms the existence of long-run equilibrium relatioships between ETF market prices and their NAVs. To model possible transient dynamics and quantify the speed of price corrections, the next section introduces Error Correction Models and Vector Error Correction Models explicitly incorporating short–term and long–term components of ETF pricing behavior.

4.3.2 ECM and VECM

The robust evidence of cointegrating relationships between prices and NAVs over the long term from both individual– and panel–level tests allows the study to proceed with the implementation of dynamic models examining short–term deviations and subsequent degrees of convergence. Recalling the equation for ECM, equation (7) in section 3.6, the model estimates capture the pace of price adjustments (ΔP_t), as reflected in the coefficient attached to the lagged error correction term in column four, $\gamma(EC_{t-1})$.

Table 12 presents the results from the ECM. Joint significance at the 1% level of confidence is provided by the Wald test shown in Table A14 in the Appendix, with the corresponding test statistic.

Table 12: Error Correction Model (ECM) Results for Cointegrated ETFs

ETF	Alpha	β (ΔNAV)	γ (EC_{t-1})	R^2
Bosera STAR 50 Index ETF	−0.00035	0.77284	−0.51192	0.91507
CSOP Hang Seng TECH Index ETF	−0.00000	0.99875	−0.99831	0.99213
ChinaAMC Hang Seng Tech Index ETF	0.00001	1.00730	−0.88882	0.99531
ChinaAMC STAR 50 ETF	0.00000	1.01040	−0.81217	0.99059
Dacheng Hang Seng Technology ETF	−0.00002	0.81852	−0.71819	0.93097
Global X Cloud Computing ETF	−0.00179	0.82628	−0.97401	0.83063
Global X Semiconductor ETF	0.00022	1.01300	−0.67102	0.94926
Hwabao FinTech Theme ETF	−0.00000	1.00070	−0.93869	0.99728
Invesco Great Wall Tech ETF	−0.00001	0.88537	−0.57485	0.92216
Penghua Semiconductor Chips ETF	−0.00000	0.98459	−0.85297	0.99388
Premia China STAR 50 ETF	−0.00008	0.95219	−0.57303	0.96286
Samsung Global Semiconductor ETF	0.00018	0.37337	−1.17170	0.81446
iShares Hang Seng TECH ETF	−0.00001	0.99070	−1.00210	0.99197

Notes: The table reports ECM estimates for all cointegrated ETFs. Coefficients include intercept (Alpha), the short–run impact of NAV changes (β), and the adjustment term toward equilibrium (γ).

Negative, strong in magnitude, and significant coefficients for the lagged error correction term indicate that prices adjust back to their underlying NAVs rapidly. The mean–reverting behavior is present, within a common range of -0.04 and -1.0, across all ETFs, supporting empirical confirmation to findings from prior studies, such as Madhavan and Sobczyk (2016). Several funds exhibit aggressive reversion within a day, such as the Samsung Bloomberg Global Semiconductor ETF, which is associated to a γ of -0.17. Extremely high R–squared values for almost all ETFs indicate an almost–perfect fit of the model, with short–term changes and subsequent arbitrageur correction mechanisms explaining almost 100% of deviations. The coefficient for NAV changes fluctuates around 0.8, which implies that price changes derive from instantaneous reaction to NAV movements and lead to proportional price adjustments. The high degree of market responsiveness to fundamental information, which was not fully captured by linear models, supports the theoretical expectation of efficient arbitrage mechanisms in the correction of intraday mispricings, as emphasized by Madhavan and Sobczyk (2016).

Table 13 expands the analysis through VECM, which captures investor behavior to fundamental signals, including short–term feedback trading, momentum effects, and delayed information assimilation. Following equation (8) in section 3.6, the specification includes lagged first differences of prices and NAVs, which amplify cross–sectional, short–term movements that remain otherwise obscured in the previous ECM system.

Table 13: VECM Estimates — ETF-Level Short-Run Dynamics

ETF	γEC_{t-1}	β (Cointegration)	ΔP_{t-1}	ΔNAV_{t-1}	R^2
Bosera STAR 50 Index ETF	−0.471	0.9958	0.059	0.026	0.075
CSOP Hang Seng TECH Index ETF	−1.404	0.9992	0.394	−0.409	0.010
ChinaAMC Hang Seng Tech Index ETF	−1.246	0.9961	0.666	−0.685	0.005
ChinaAMC STAR 50 ETF	0.090	1.0008	−0.376	0.375	0.002
Dacheng Hang Seng Technology ETF	−0.947	0.9909	−0.270	0.140	0.112
Global X Cloud Computing ETF	−1.227	1.0067	−0.107	0.170	0.370
Global X Semiconductor ETF	−0.892	0.9989	−0.104	0.105	0.054
Hwabao FinTech Theme ETF	1.251	1.0014	−1.080	1.079	0.003
Invesco Great Wall Tech ETF	−0.530	0.9833	−0.425	0.392	0.076
Penghua Semiconductor Chips ETF	−1.021	1.0010	0.949	−0.950	0.006
Premia China STAR 50 ETF	−0.148	0.9998	−0.067	0.132	0.008
Samsung Global Semiconductor ETF	−1.129	0.9963	−0.038	0.150	0.724
iShares Hang Seng TECH ETF	0.867	0.9984	−1.162	1.136	0.012

Notes: VECM coefficients include the long-run error correction term (γ), cointegration slope (β), and lagged short-run impacts of price and NAV changes. All series differenced and aligned.

The lagged error correction coefficients remain predominantly negative, ranging between -0.1 and -0.4. These results reinforce the moderate–to–strong speed of pricing adjustment from the corresponding ECM estimates, confirming mean–reverting behavior in mispricings over the long term. The pace varies across ETFs, with particularly steep values near -0.2 and -0.4 for CSOP Hang Seng Tech Intex and ChinaAMC Hang Seng Tech Index, in line with timely arbitrageur reaction. Other ETFs, such as Bosera Star 50 and Huatai–PineBridge, show mild speeds of reversion, while rare, positive EC terms reflect temporary divergence.

Positive and significant coefficients for lagged ETF price changes, ΔP_{t-1}, are found for a subset of ETFs—including Bosera Star 50, ChinaAMC Hang Seng Tech, and Penghua Chips ETF—suggesting short–term momentum consistent which narrative–based investor behavior (Jegadeesh & Titman, 1993; Shiller, 2017). Conversely, ETFs such as EFund CSI Internet 50 and Samsung KODEX show negative delayed price effects, indicating reversal and convergence following deviation spikes. The lag coefficients attached to NAV are often opposite in sign to price lags, with most values being positive and modest in magnitude, reflecting partial price adjustments after NAV shifts. For some ETFs, such as Nikko Global Internet and Samsung KODEX, ΔNAV_{t-1} terms suggest delayed transmission of fundamental signals, which is a documented effect for emerging markets (Atanasova & Weisskopf, 2020; Lee & Price, 2023). This further reinforces the hypothesis that NAV misalignment is not solely arbitrage–related, but also structurally constrained, adding meaning to the potential barriers discussed in section 4.1. The significant drop in the R–squared, which now shows higher dispersion across ETFs, in ranges between 1% and 72%, is still valid in the context of high–frequency, panel data analysis, as it also introduces multiple lagged effects in pricing movements. Similarly to the ECM, Table A15 in the Appendix displays statistical significance for all ETFs at the smallest level of confidence from the application of the Wald test for joined significance.

The dynamic models implemented in this research provide a complete picture of ETF pricing dynamics. Beyond confirming the existence of long–run price–NAV equilibrium, the speed and direction of error correction coefficients, coupled with the pattern of short–run responses, reveal latent frictions and trading behaviors. In this sense, the ECM and VECM frameworks serve not only as tools of econometric validation, but as vehicles into the actual mechanisms—both structural and behavioral—that govern ETF market efficiency.

4.4 Robustness and Sensivitity Insights

This section provides alternative model specifications and data splits that confirm the solidity of the selected models by assessing whether key patterns in ETF price deviations—particularly those linked to sentiment, volatility, and institutional behavior—hold under such frameworks. Full results are displayed in Appendix B.

4.4.1 March 2022 Dummy Variable and Interaction terms

The Pooled OLS regression model, with baseline equation (2), is extended to include a dummy variable which captures the period covering March 2022. Figure 2 in the Data and Variables subsection shows particularly large discounts as ETF mispricings corresponding to the month in question. To test the significance of the exogenous shock characterized by the March 2022 crisis that affected markets in Asia Pacific, the dummy variable is included to isolate the explanatory power of the event. The variable assumes value 1 if $t \in$ [March 1, 2022; March 31, 2022], and value 0 otherwise.

To address differential sensitivity of behavioral and volatility–linked factors during the episode of market stress in March 2022, interaction terms between the dummy variable corresponding to the crisis and VHSI, feedback trading, and technology sector volatility are estimated to detect the amplified or diminished effect of investor behavior and market sentiment under extreme conditions.

Each interaction term was estimated in a separate regression to exacerbate any multicollinearity issues. Table B1 reports results for three separate Pooled OLS regressions. Each column contains an interaction term between the March 2022 dummy variable and market sentiment, feedback trading, and Technology sector volatility, respectively. These interaction terms were introduced to explore potential nuanced effects of the behavioral variables during the specific month. All three models provide meaningful findings. Technology sector volatility shows a negative coefficient across all specifications, though its magnitude and significance drop substantially in its interaction with the crisis. The partial loss of significance and size is likely due to the absorption of volatility's effect into the extremely significant interaction term, which reinforces the amplified effect of volatility on mispricings due to the crisis. Feedback trading remains insignificant in all three models, with near–zero coefficients, in contrast to the positive, significant interaction term with the crisis. This underscores the earlier interpretation that investor behavior is not uniformly influential during normal times, while it acquires importance during high uncertainty. The dummy variable capturing the March 2022 crisis period, which appeared statistically insignificant in the baseline model including only the crisis dummy, turns significant in two of the models.

The combined significance of both the crisis dummy and its interaction terms provides additional evidence of distortions induced by market stress, both through investor behavior and broader market mechanisms. Cherkes et al. (2009) and Hilliard (2014) identify emerging sectors, fragmented trading venues, and international NAV timing mismatches as structural fund features that make these ETFs particularly vulnerable to price deviations. Such latent fragilities surface more acutely under crisis conditions, heightening price inefficiencies. Beyond the March 2022 crisis, macroeconomic policy shocks may also influence ETF pricing. To test this, the next section examines the effect of the U.S. interest rate hike during the same month.

4.4.2 Structural Breaks: Rate Hike 2022 Dummy

To investigate the potential structural break induced by the rate hike, a dummy variable was included capturing the period from March 16 to March 31, 2022, corresponding to the U.S. Federal Reserve's interest rate hike. Table B2 presents results for the baseline OLS model with the dummy variable, as well as the nonlinear form, following the same rationale as for the crisis–adjusted model.

While the nonlinear model shows slight explanatory power improvement in terms of R–squared, it reveals higher structural complexity. The dummy variable shows a large negative effect in the linear model, which diminishes and changes direction under the nonlinear specification, suggesting that the observed deviation shock is primarily driven by behavioral nonlinearities rather than the policy event itself. While the March shock was short–lived, its potential structural effects may persist. To capture these post–crisis shifts more systematically, the following section presents a Difference–in–Difference regression.

4.4.3 Difference–in–Difference: High–volatility ETFs Post–July 2022

This section exploits the time– and cross–sectional partitions conducted in earlier sections and implements a Difference–in–Difference approach to further explore the structural dynamics behind ETF mispricings. The temporal shift reflects a natural breakpoint in the sample, July 2022, which follows market–wide disruptions condensed in the month of March 2022. This coincides with potential structural adjustments in risk pricing, arbitrage mechanisms, and fund flows across Asian tech–focused ETFs. The fund–specific distinction is defined based on high– and low–volatility ETFs, being the treatment and control groups respectively. The interaction between the two dimensions isolates the differential effect of the post–crises regime on high–volatility ETFs, under the assumption of parallel trends prior to July 2022. In this context, the DiD estimation assesses whether high–volatility ETFs experienced a structurally different shift in mispricings after mid–2022.

Given the strong evidence of nonlinearity in sentiment and behavioral effects documented in previous sections, $VHSI^2$, Feedback $Trading^2$, and Market Cap x Liquidity were added to the specification, with results reported in Table B3. The results confirm that high–volatility ETFs experience a significant increase in price gaps from their NAVs after July 2022. The interaction term is highly significant, supporting the view that structural shifts in arbitrage and behavior followed early–2022 market turmoil.

4.4.4 Seasonality: Monthly Dummy Variables

To account for systematic patterns in ETF mispricings across the whole calendar year, 11 monthly dummy variables—January being the benchmark to avoid multicollinearity—were included in the baseline model. The aim of the augmented specification is to capture seasonal fluctuations arising from tax–loss harvesting, quarter–end portfolio rebalancing, or liquidity cycles, especially in Asia–Pacific markets where fiscal calendars and trading volumes vary by month. Tables B4 and B5 present results for the extended, linear and nonlinear models. While the linear model shows weak significance, strong, persistent seasonality is recorded in the nonlinear specification.

Strong, positive deviations are observed in the second half of the year, particularly in August, September, and October. These patterns likely reflect a combination of factors. Bouman and Jacobsen (2002) document that international equity markets, included Asia, often rebound in the third quarter following mid–year corrections or earning revisions, leading to pricing gaps between ETF shares and their NAVs. This regional market momentum is found especially in technology sectors. Fiscal calendar effects, common in Asian markets may also justify the statistically significant seasonal patterns. Institutional investors adjust positions ahead of fiscal–year closings, such as in Japan and Hong Kong, which can cause temporary mispricings due to rapid inflows or redemptions (Madhavan & Sobczyk, 2016). Pan and Zeng (2021) and Petajisto (2017) also identify liquidity dynamics and portfolio window dressing, with supply–demand mismatches and inflated ETF prices above their NAV due to higher trading volumes and investor activity around Golden Week in early October and Q3 earnings season, and asset managers rebalancing towards outperforming tech names near quarter–ends to enhance portfolio appearance, respectively.

These effects align with ETF–specific studies documenting calendar–driven frictions in arbitrage mechanisms (Petajisto, 2017).

4.4.5 Fixed Effects

This subsection builds on the baseline Pooled OLS model, extending the estimation framework by relaxing the assumption of uncorrelation between ETF–specific behavior and the regressors, captured by the idiosyncratic error term. The estimation process unfolds into the application of the within transformation, as follows:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (\epsilon_{it} - \bar{\epsilon}_i) \tag{9}$$

The model subtracts the time averages, $\bar{y}_i$ and $\bar{x}_i$, across the units i = 1, ..., N, since c_i is fixed, yielding a model that differences out any time–invariant characteristics. The model includes the matrix of regressors X_{it} which now contains only market sentiment, technology sector volatility and feedback trading as explanatory variables, while the intercept and any static variables are dropped out of the regression through the within transformation, as they become zero. Therefore, market capitalization, liquidity of the underlyings and percentage institutional ownership are removed, allowing clearer interpretation and higher explanatory power of the coefficients attached to the three time–variant factors.

Accounting for ETF–specific heterogeneity enhances the significance of time–varying behavioral regressors (Tables B6, B7). An extended specification with a Tech Volatility x Feedback Trading interaction (Table B8) reveals that momentum–based mispricings are conditional on volatility regimes, highlighting the importance of behavioral frictions during turbulence in explaining ETF pricing inefficiencies, especially in speculative and rapidly evolving sectors such as technology.

Overall, these results reinforce the reliability of the core pooled OLS and cointegration–based findings, while offering deeper insight into how ETF mispricings evolve under macro, seasonal, and volatility–specific conditions.

5 Conclusion

This research investigates the drivers and dynamics of deviations between ETF market prices and their net asset values, focusing on a panel of 19 tech–oriented ETFs listed across Asian exchanges. By integrating structural, behavioral, and market–level determinants under several econometric frameworks, this study contributes a novel perspective on pricing inefficiencies in emerging, theme–based, and international ETF markets.

The analysis begins with sample selection and removal of any influential variables detected through visual and descriptive statistics. The empirical findings show that ETF mispricings are persistent and highly sensitive to structural and behavioral factors. Contrary to classical arbitrage efficiency expectations, larger market capitalization, greater institutional presence, and higher liquidity of underlying assets are associated with wider dislocations in the baseline linear OLS model. These results reflect the informational and structural complexities of Asian ETF markets—particularly those with cross–border exposure, QDII structures, and tech sector specialization.

The diagnostic tests for the validity of the models and the baseline assumptions justify the insertion of nonlinear terms for market sentiment, feedback trading, and an interaction term between market capitalization and liquidity of the underlyings. The results confirm nonlinear and regime–dependent effects for the behavioral variables, and the relevance of the size–liquidity joint effect. Under nonlinear specifications, market cap and liquidity of the underlyings individually narrow mispricings, while their positive simultaneous effect indicates that large and liquid in fundamentals ETFs face higher dislocations, possibly due to crowding, narrative–based trading, or delayed arbitrage in high–turnover segments. Market sentiment (VHSI) displays a convex relationship with deviations: moderate increases narrow mispricings, while extreme sentiment levels amplify them. Feedback trading, insignificant in the linear model, gains explanatory power in nonlinear and high–volatility settings, pointing to herd–driven distortions during turbulent periods.

Subsample regressions show that high–volatility ETFs are more exposed to both structural frictions and behavioral noise, with substantially higher explanatory power and persistent deviation patterns. The AR(1) models confirm this stickiness, revealing that up to 70% of price deviations persist from one trading day to the next in high–volatility ETFs, suggesting inertia in correction mechanisms and latent barriers to arbitrage.

The test results for cointegration analysis between prices and their fundamental values, as well as the lack of coverage among the existing literature of such long–run dynamics motivate further investigation through Error Correction Models and Vector Error Correction Models. The results validate the presence of equilibrium relationships between prices and NAVs, with the relative coefficients demonstrating that mispricings are mean–reverting and exhibit varying speeds of adjustment across ETFs. These models also capture the interplay between delayed price responses, NAV shocks, and momentum effects, offering deeper insight into the dynamic forces that govern ETF valuation over time.

The findings have several implications for market participants and regulators. With respect to investors, the results highlight the importance of understanding structural features—such as international holdings, underlying liquidity, and fund size—as these can amplify or dampen pricing inefficiencies depending on market conditions. For arbitrageurs, the documented inertia and regime–dependence of deviations suggest opportunities for dynamic strategies, particularly during high–volatility episodes.

From a policy perspective, regulators in Asian markets should prioritize enhancements in NAV reporting timeliness, cross–border trading harmonization, and transparency in fund structures. Greater international disclosure around ETF basket composition and intra–day NAV estimates could reduce information lags and improve pricing alignment.

Finally, this research demonstrates that ETF price deviations are not random noise but reflect systematic patterns shaped by structural frictions, behavioral dynamics, and temporal regimes. Accounting for these nonlinearities and cointegrated relationships provides a richer, more realistic understanding of ETF market behavior in Asia's rapidly evolving financial landscape.

While this study provides a comprehensive framework to analyze price deviations from NAV, it is subject to several limitations. First, the unique structural features of Asian ETF markets—include time zone misalignments and higher theme–based activity—may limit the generalization of the results. Additionally, the relatively small sample size, with restrictions in geographic and sectoral scope, and brief time horizon may omit broader macroeconomic controls or patterns in mispricings. Second, some variables—such as institutional ownership, liquidity of the underlying portfolio, and market capitalization—are imported as static despite possible quarterly or annual temporal variation. Proxy variables—VHSI and feedback trading—capture sentiment and investor behavior only indirectly, potentially overlooking real–time dynamics. The exclusive reliance on daily closing data may also obscure intraday pricing pressures and arbitrage windows. Lastly, although the econometric methodology incorporates nonlinearities and cointegrating techniques, more advanced models such as machine learning approaches could provide complementary regime switches and structural breaks more flexibly to the ETF mispricing behavior.

Future research could expand the sample to global ETFs, include intraday data, and implement alternative strategies that model short-term dislocations.

Works Cited

Atanasova, C., & Weisskopf, J.-P. (2020). The price of international equity etfs: The role of relative liquidity. *Journal of International Financial Markets, Institutions and Money*, *65*, 101190.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–152.

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, *75*(2), 283–317.

Barkham, R. J., & Ward, C. W. R. (1999). Investor sentiment and noise traders: Discount to net asset value in listed property companies in the u.k. *The Journal of Real Estate Research*, *18*(2), 291–312.

Ben-David, I., Franzoni, F., Moussawi, R., & Sedunov, J. (2021). The granular nature of large institutional investors. *Management Science*, *67*(11), 6629–6659.

Ben-David, I., Franzoni, F. A., & Moussawi, R. (2018). Do etfs increase volatility? *The Journal of Finance*, *73*(6), 2471–2535.

Bohl, M. T., & Brzeszczynski, J. (2006). Do institutional investors destabilize stock prices? evidence from an emerging market. *Journal of International Financial Markets, Institutions and Money*, *16*(4), 370–383.

Bouman, S., & Jacobsen, B. (2002). The halloween indicator, "sell in may and go away": Another puzzle. *American Economic Review*, *92*(5), 1618–1635.

Chen, G.-m., Firth, M., & Rui, O. M. (2002). Stock market linkages: Evidence from latin america. *Journal of Banking & Finance*, *26*(6), 1113–1141.

Cherkes, M., Sagi, J., & Stanton, R. (2009). A liquidity-based theory of closed-end funds. *Review of Financial Studies*, *22*(1), 257–297.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738.

DeFusco, R. A., Ivanov, S. I., & Karels, G. V. (2011). The exchange traded funds' pricing deviation: Analysis and forecasts. *Journal of Economics and Finance*, *35*(2), 181–197.

Deville, L. (2008). Exchange traded funds: History, trading and research. In *Handbook of financial engineering*. Springer.

Dogra, G. (2023). Asian stocks in 2022 suffer biggest foreign outflows since 2008 global crisis [Online news article].

Ehm, H., & Ponsignon, T. (2012). Future research directions for mastering end-to-end semiconductor supply chains. *Proceedings of the IEEE International Conference on Automation Science and Engineering.*

Engle, R., & Sarkar, D. (2006). Premiums-discounts and exchange traded funds. *The Journal of Derivatives*, *13*(4), 27–45.

ETFGI. (2024). Development of assets of global exchange traded funds (etfs) from 2003 to 2023 (in billion u.s. dollars) [graph] [Retrieved May 18, 2025, from Statista].

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Fama, E. F., & French, K. R. (2001). Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics*, *60*(1), 3–43.

Galindo Gil, H., & Lazo-Paz, R. (2025). An etf-based measure of stock price fragility. *Journal of Financial Markets*, *72*, 100946.

Gordon, M. J. (1959). Dividends, earnings, and stock prices. *The Review of Economics and Statistics*, *41*(2), 99–105.

Hang Seng Indexes Company Limited. (2022). Hang seng tech index performance – march 2022 [Retrieved May 18, 2025].

Hilliard, J. (2014). Premiums and discounts in etfs: An analysis of the arbitrage mechanism in domestic and international funds. *Global Finance Journal*, *25*(2), 90–107.

Hong Kong Exchanges and Clearing Ltd. (HKEX). (2023). Etf market statistics – monthly reports [Retrieved May 18, 2025].

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

Kallinterakis, V., Liu, F., Pantelous, A. A., & Shao, J. (2020). Pricing inefficiencies and feedback trading: Evidence from country etfs. *International Review of Financial Analysis*, *70*, 101498.

Kumala, C., Ye, Z., Zhu, Y., & Ke, Q. (2024). Why does price deviate from net asset value? the case of singaporean infrastructure reits. *International Review of Financial Analysis*, *93*, 103172.

Lee, H., & Price, M. (2023). Delayed nav transmission and price dynamics in emerging market etfs. *Journal of International ETF Research*, *12*(3), 45–62.

Liow, K. H. (2003). Property company stock price and net asset value: A mean reversion perspective. *The Journal of Real Estate Finance and Economics*, *27*(2), 235–255.

Madhavan, A., & Sobczyk, A. (2016). Price dynamics and liquidity of exchange-traded funds. *Journal of Investment Management*, *14*(2), 1–17.

Milani, B., & Ceretta, P. S. (2014). Dynamic correlation between share returns, nav variation and market proxy of brazilian etfs. *Engineering Economics*, *25*(1), 21–30.

Pan, K., & Zeng, Y. (2021). Etf arbitrage under liquidity mismatch. *Journal of Financial Economics*, *142*(1), 75–98.

Petajisto, A. (2017). Inefficiencies in the pricing of exchange-traded funds. *Financial Analysts Journal*, *73*(1), 24–54.

Rehkugler, H., Schindler, F., & Zajonz, R. (2012). The net asset value and stock prices of european real estate companies. In W. Breuer & C. Nadler (Eds.), *Real estate finance* (pp. 53–77). Gabler Verlag.

Reilly, F. K., & Brown, K. C. (2011). *Investment analysis and portfolio management* (11th). Cengage Learning.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004.

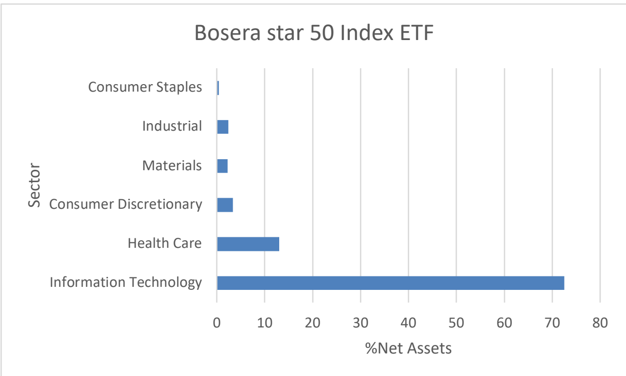
Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic Perspectives*, *4*(2), 19–33.

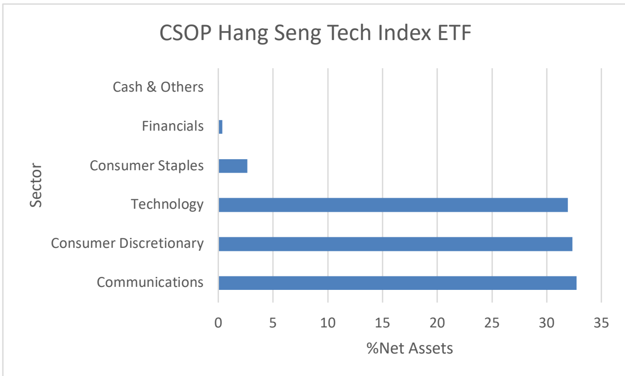
Zhen, S., & Li, J. (2025). China’s tech rally rests on ’hot money’ [Online news article].

Appendix A and B contain supplementary figures and tables referenced in earlier sections. Their inclusion supports detailed descriptive and visual diagnostics, robustness checks, and ETF–specific characteristics underlying the empirical models.

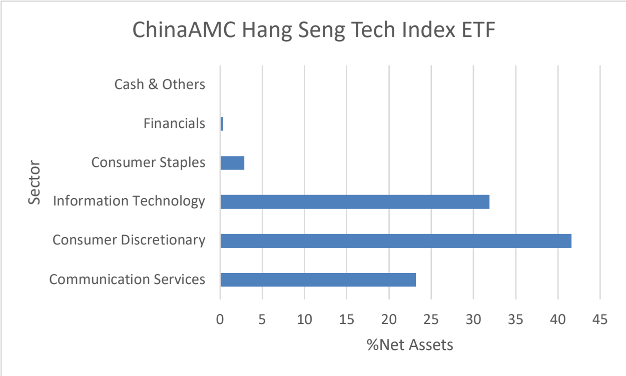
Appendix A: Core Results

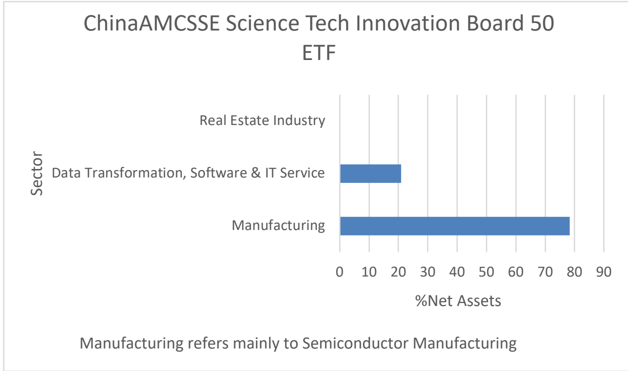
Section A.1 — ETF Holdings Breakdown (Figures A1–A20): The following figures present the holdings composition of each of the 20 Asian tech–oriented ETFs included in the sample. The figures illustrate the top constituents and their respective portfolio weights, highlighting sector concentration, regional focus, and underlying exposure. These holdings help explain structural sources of price deviation from NAV, as discussed in Section 4.1.



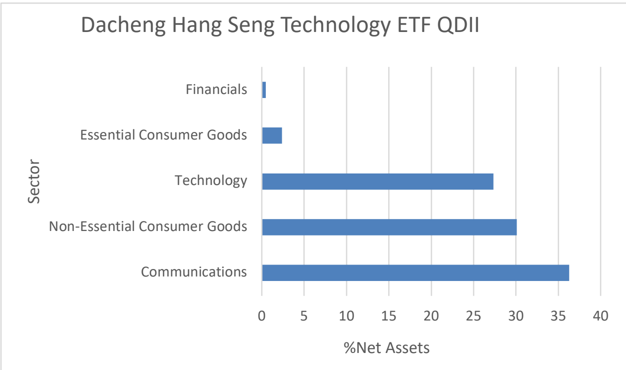


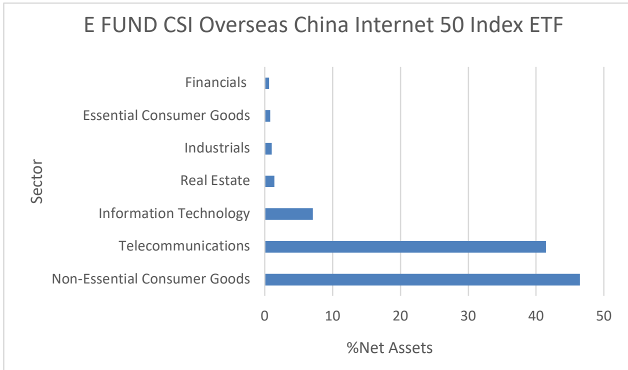
Figures A1-A2: Holdings breakdown for the Samsung KODEX and Global X China Semiconductor ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



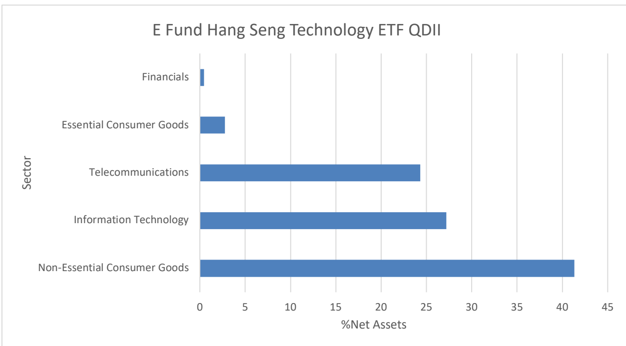


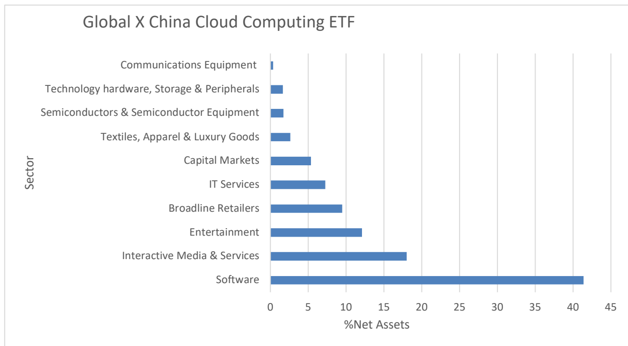
Figures A3-A4: Holdings breakdown for the ChinaAMC Hang Seng TECH Index and ChinaAMCSSE Science Tech Innovation Board 50 ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



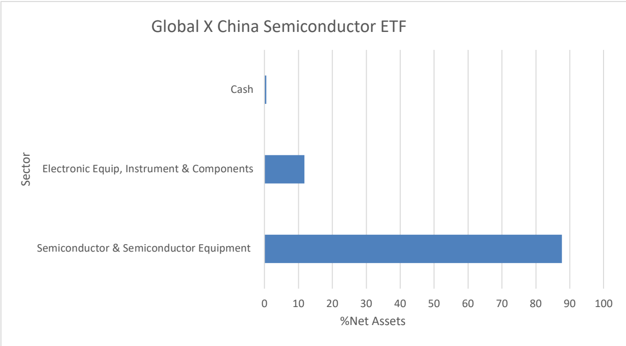


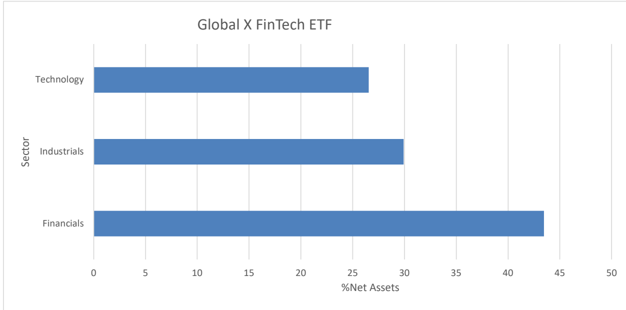
Figures A5-A6: Holdings breakdown for the Dacheng Hang Seng Technology QDII and E Fund CSI Overseas Internet 50 Index ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



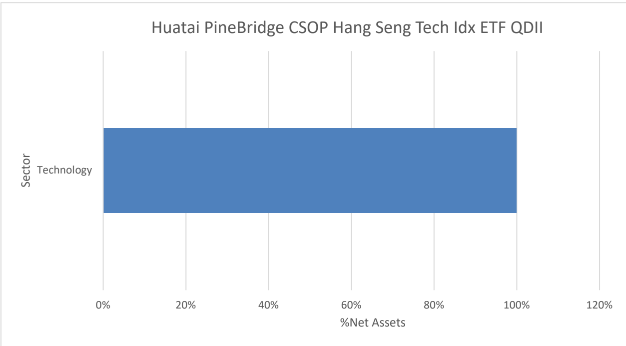


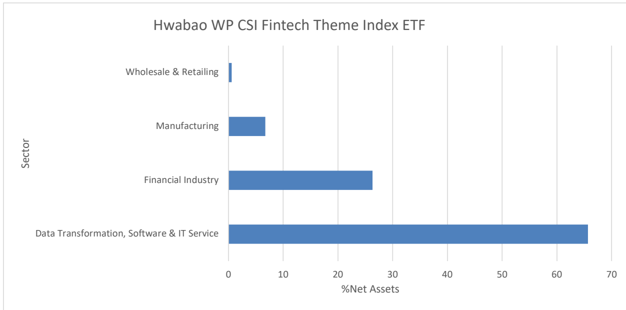
Figures A7-A8: Holdings breakdown for the E Fund Hang Seng Technology QDII and Global X China Cloud Computing ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



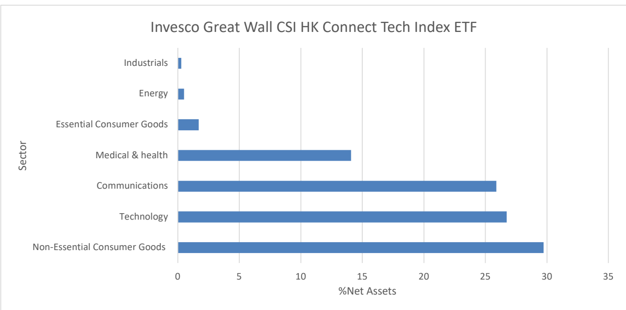


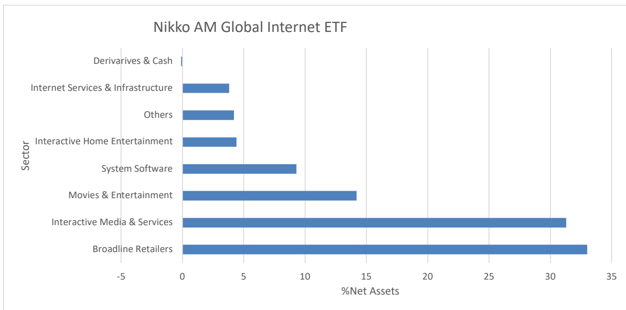
Figures A9–A10: Holdings breakdown for the Global X China Semiconductor and Global X FinTech ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



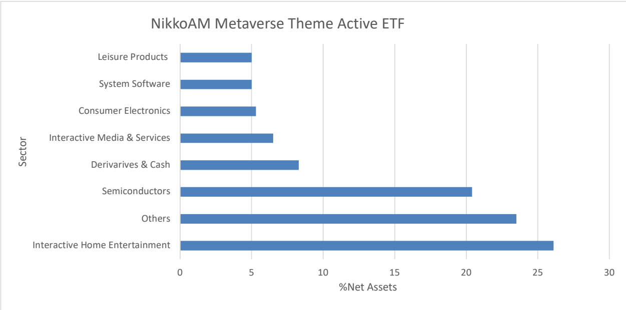


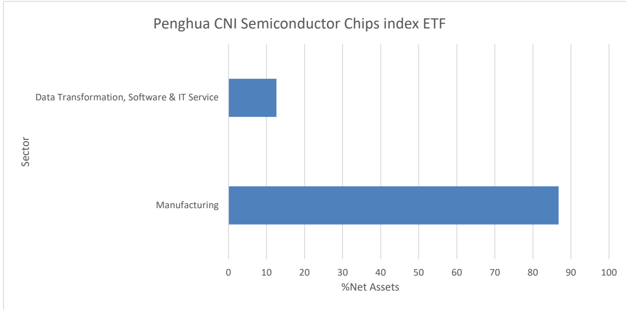
Figures A11–A12: Holdings breakdown for the Huatai PineBridge CSOP Hang Seng TECH Idx QDII and Hwabao WP CSI FinTech Theme index ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



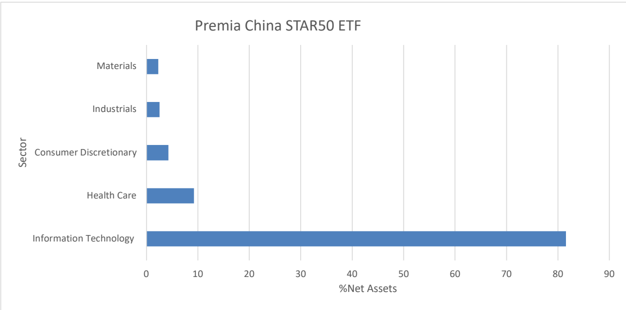


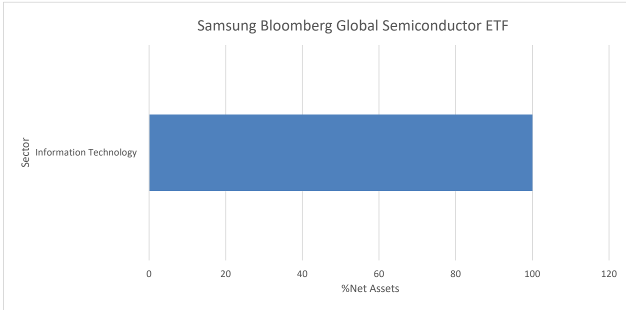
Figures A13–A14: Holdings breakdown for the Invesco Great Wall CSI HK Connect TECH Index and Nikko AM Global Internet ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.



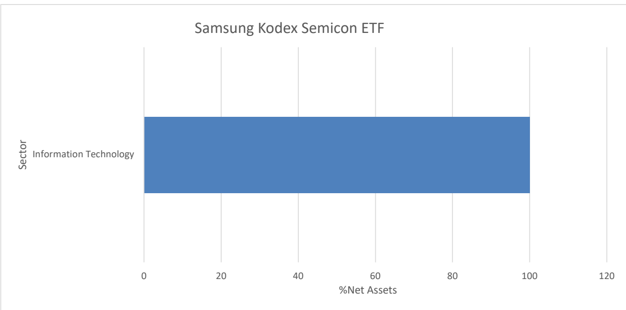


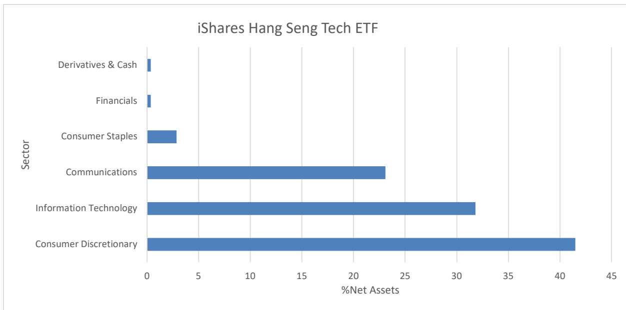
Figures A15–A16: Holdings breakdown for the Nikko AM Metaverse Theme Active and Penghua CNI Semiconductor Chips Index ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.





Figures A17–A18: Holdings breakdown for the Premia China STAR50 and Samsung Bloomberg Global Semiconductor ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.





Figures A19–A20: Holdings breakdown for the Samsung Kodex Semicon and iShares Hang Seng TECH ETFs. The charts display the top 10 constituent stocks and their respective weightings as of December 2023.

Section A.2—Descriptive Statistics with Outlier (Tables A1–A2, Figures A1–A4) These tables offer descriptive and visual representations of the sample including the outlier: Global X FinTech ETF, referred to in section 3.2 of the Methodology.

Table A1: Summary Statistics for Panel Variables-outlier

Variable	Mean	Std. Dev.	Min	Max	N
Price Deviations	-0.0412	0.1913	-0.8818	0.2641	9340
Market Capitalization ($bn)	1.6038	1.9355	0.0012	5.8000	9340
Market Sentiment (VHSI)	3.4282	0.5824	2.3296	6.3157	9340
Tech Sector Volatility	0.0250	0.0138	0.0000	0.1261	9340
Feedback Trading	-0.0581	0.3991	-0.9283	0.8644	9340
Liquidity (90-day Volume, $bn)	2.1502	4.0424	0.0000	17.4379	9340
% Institutional Ownership	0.1334	0.1782	0.0000	0.5208	9340

Note: This table reports summary statistics for the dependent and explanatory variables in the regression. Price Deviations are unitless relative measures of mispricings; market cap and liquidity of the underlyings are expressed in billion U.S. Dollars; VHSI is expressed in index points, Technology sector volatility is the daily return standard deviation; Feedback Trading is the cross-ETF correlation coefficient; Percentage Institutional Ownership is scaled as a proportion $\in [0, 1]$.

Table A2: Descriptive Moments: Skewness and Kurtosis for Panel Variables-with outlier

Variable	**Skewness**	**Kurtosis**
Price Deviations	-4.0808	17.8164
Market Capitalization ($bn)	0.9879	2.4951
Market Sentiment (VHSI)	1.0680	4.4678
Tech Sector Volatility	3.1384	20.9279
Feedback Trading	0.1392	2.2956
Liquidity (90-day Volume, $bn)	2.7008	10.3449
% Institutional Ownership	1.1659	2.8064

Notes: This table reports the skewness and kurtosis for the panel variables used in the regression analyses. Skewness measures the asymmetry of the distribution; kurtosis measures the presence of extreme values relative to a normal distribution. Values are based on the sample statistics with bias correction.

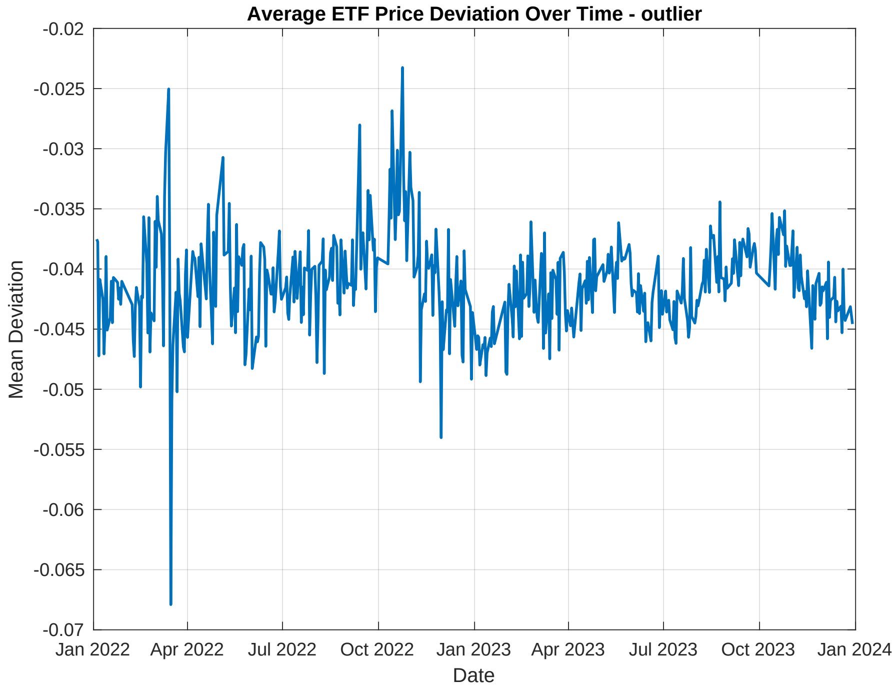


Figure A1: Time series of the average daily ETF price deviation from NAV across 20 Asian tech-oriented ETFs from January 2022 to December 2023. The chart illustrates systematic discounting patterns and stress periods like March 2022

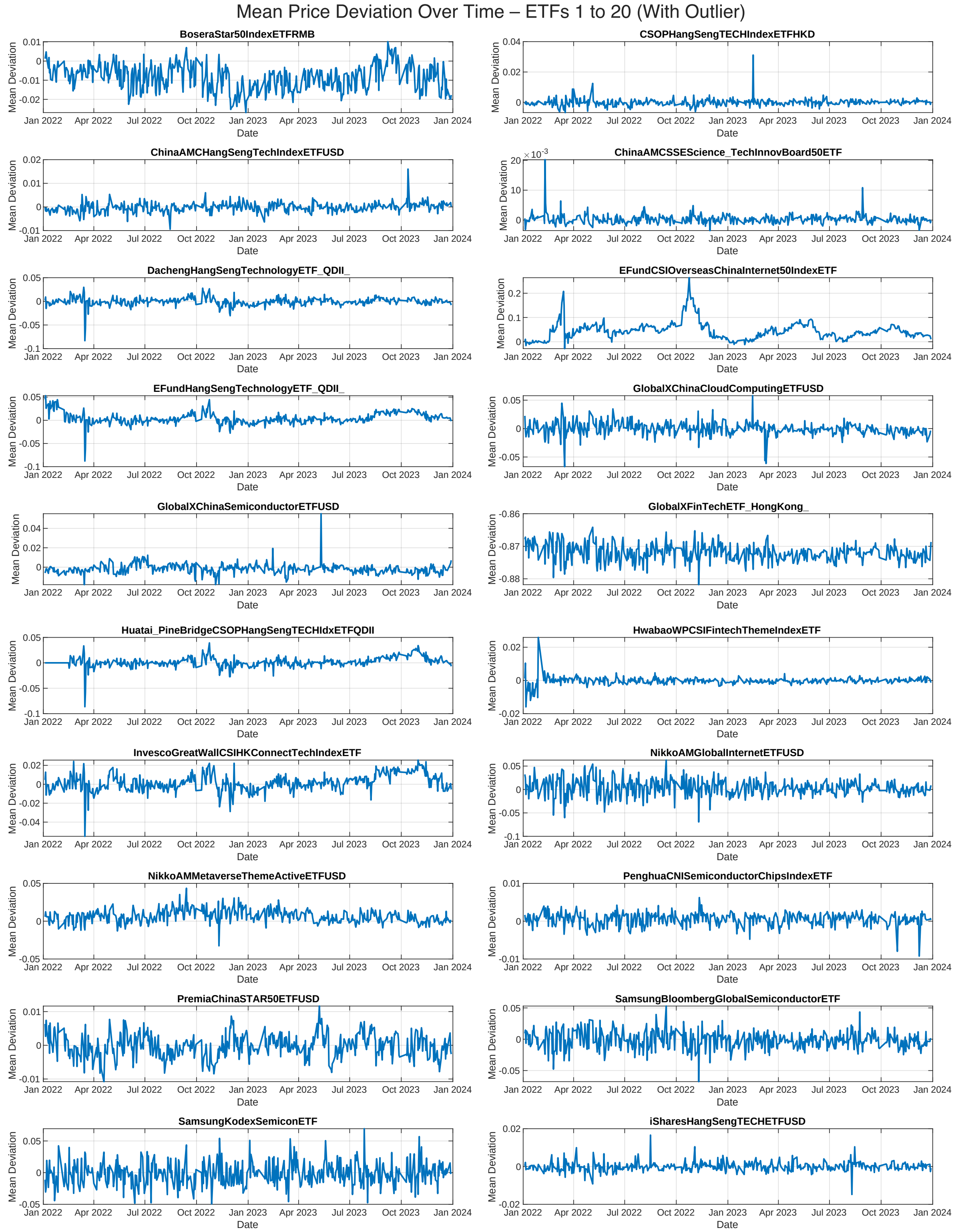


Figure A2: Time Series of Deviations by ETF (Outlier Sample). This figure shows the time series of price deviations for each ETF in the sample. The extremely negative deviations visible in the Global X FinTech ETF motivated its exclusion from the final regression analyses, as its average market prices were abnormally lower than its corresponding NAVs (approximately 4 versus 37 units, respectively).

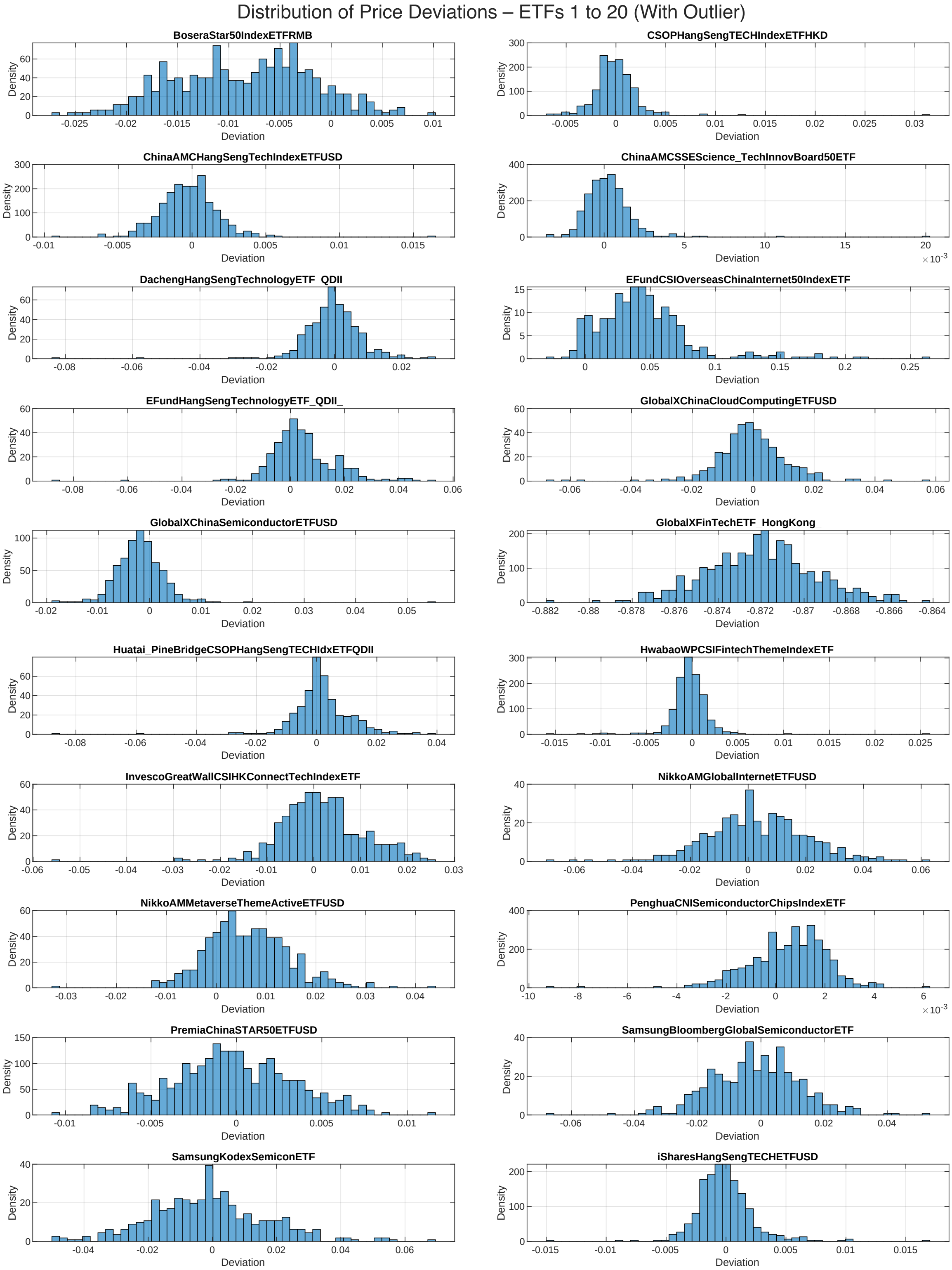


Figure A3: Distribution of Price Deviations by ETF (Outlier Sample). This figure shows the distribution of price deviations for each ETF in the sample. The extremely negative deviations visible in the Global X FinTech ETF motivated its exclusion from the final regression analyses, as its average market prices were abnormally lower than its corresponding NAVs (approximately 4 versus 37 units, respectively).

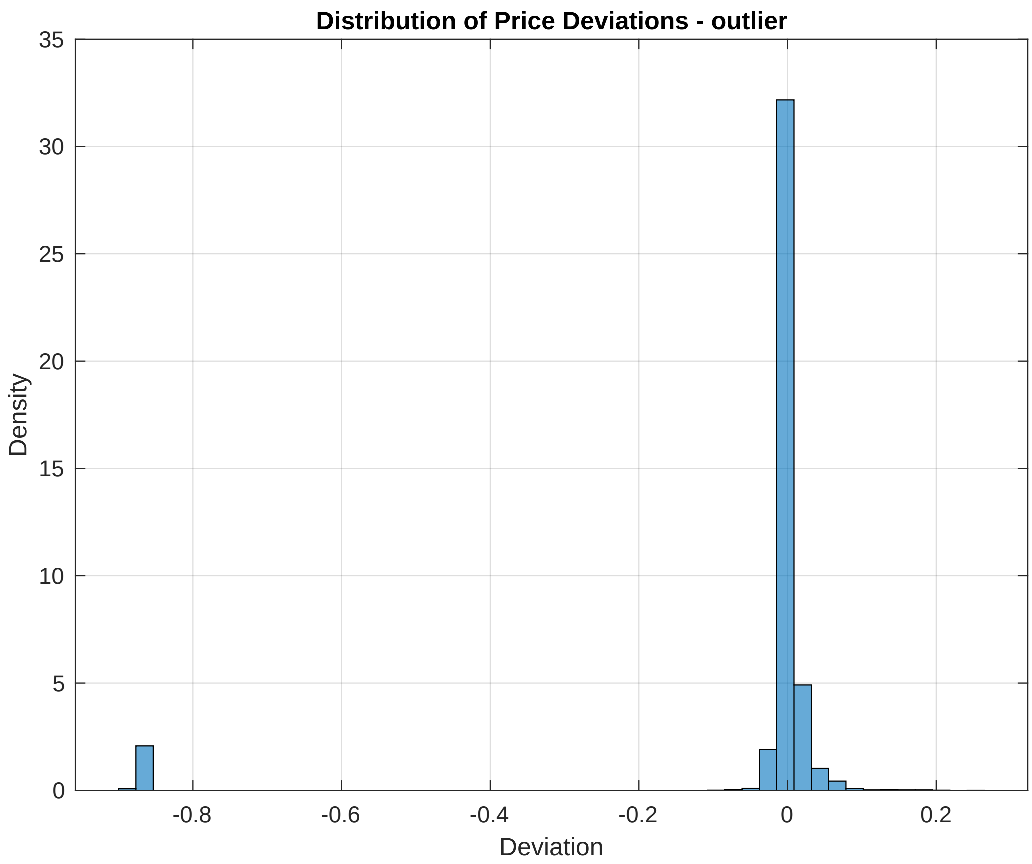


Figure A4: Distribution of ETF price deviations from NAV across 20 Asian tech-oriented funds, including the Global X FinTech ETF (outlier). The histogram highlights .

Section A.3—Descriptive Statistics (Excluding FinTech ETF) This section presents revised summary statistics after removing the outlier ETF. The cleaner distribution improves model reliability and justifies the sample construction in Chapter 3.

Table A3: Descriptive Moments: Skewness and Kurtosis for Panel Variables (Excl. FinTech ETF)

Variable	**Skewness**	**Kurtosis**
Price Deviations	4.0692	39.7818
Market Capitalization ($bn)	0.9169	2.3573
Market Sentiment (VHSI)	1.0680	4.4678
Tech Sector Volatility	3.1384	20.9279
Feedback Trading	0.1392	2.2956
Liquidity (90-day Volume, $bn)	2.6173	9.8308
% Institutional Ownership	1.3425	3.2001

Notes: This table reports the skewness and kurtosis for the panel variables used in the regression analyses. Skewness measures the asymmetry of the distribution; kurtosis measures the presence of extreme values relative to a normal distribution. Values are based on the sample statistics with bias correction.

Section A.4—Stationarity Tests (ADF) These tables report Augmented Dickey-Fuller (ADF) unit root tests for key variables used in modeling price deviation dynamics. Stationarity tests are essential to justify use of OLS, ECM, and VECM specifications.

Table A4: ADF Test Results for Deviation

ETF	ADF Statistic	p-value
BoseraStar50IndexETFRMB	-11.9500	0.0010
CSOPHangSengTECHIndexETFHKD	-21.2300	0.0010
ChinaAMCHangSengTechIndexETFUSD	-18.5700	0.0010
ChinaAMCSSEScience_TechInnovBoard50ETF	-17.6300	0.0010
DachengHangSengTechnologyETF_QDII_	-15.2200	0.0010
EFundCSIOverseasChinaInternet50IndexETF	-4.8800	0.0010
EFundHangSengTechnologyETF_QDII_	-10.7500	0.0010
GlobalXChinaCloudComputingETFUSD	-19.4400	0.0010
GlobalXChinaSemiconductorETFUSD	-15.6200	0.0010
Huatai_PineBridgeCSOPHangSengTECHIdxETFQDII	-12.7400	0.0010
HwabaoWPCSIFintechThemeIndexETF	-20.6100	0.0010
InvescoGreatWallCSIHKConnectTechIndexETF	-12.0900	0.0010
NikkoAMGlobalInternetETFUSD	-19.7500	0.0010
NikkoAMMetaverseThemeActiveETFUSD	-14.4700	0.0010
PenghuaCNISemiconductorChipsIndexETF	-18.5800	0.0010
PremiaChinaSTAR50ETFUSD	-13.6600	0.0010
SamsungBloombergGlobalSemiconductorETF	-21.8100	0.0010
SamsungKodexSemiconETF	-21.7800	0.0010
iSharesHangSengTECHETFUSD	-21.5500	0.0010

Notes: This table reports Augmented Dickey–Fuller (ADF) test statistics and p-values for the percentage price deviation series of each ETF. The null hypothesis is the presence of a unit root (non-stationarity). All p-values are below 0.01, indicating that deviations are stationary ($I(0)$) across all ETFs, justifying their use in regression without differencing.

Table A5: ADF Test Results for VHSI

ETF	ADF Statistic	p-value
BoseraStar50IndexETFRMB	-3.7000	0.0046
CSOPHangSengTECHIndexETFHKD	-3.7000	0.0046
ChinaAMCHangSengTechIndexETFUSD	-3.7000	0.0046
ChinaAMCSSEScience_TechInnovBoard50ETF	-3.7000	0.0046
DachengHangSengTechnologyETF_QDII_	-3.7000	0.0046
EFundCSIOverseasChinaInternet50IndexETF	-3.7000	0.0046
EFundHangSengTechnologyETF_QDII_	-3.7000	0.0046
GlobalXChinaCloudComputingETFUSD	-3.7000	0.0046
GlobalXChinaSemiconductorETFUSD	-3.7000	0.0046
Huatai_PineBridgeCSOPHangSengTECHIdxETFQDII	-3.7000	0.0046
HwabaoWPCSIFintechThemeIndexETF	-3.7000	0.0046
InvescoGreatWallCSIHKConnectTechIndexETF	-3.7000	0.0046
NikkoAMGlobalInternetETFUSD	-3.7000	0.0046
NikkoAMMetaverseThemeActiveETFUSD	-3.7000	0.0046
PenghuaCNISemiconductorChipsIndexETF	-3.7000	0.0046
PremiaChinaSTAR50ETFUSD	-3.7000	0.0046
SamsungBloombergGlobalSemiconductorETF	-3.7000	0.0046
SamsungKodexSemiconETF	-3.7000	0.0046
iSharesHangSengTECHETFUSD	-3.7000	0.0046

Notes: This table presents ADF test results for the market sentiment variable (VHSI). The low p-values (0.0046) across all ETFs suggest rejection of the unit root null, implying stationarity ($I(0)$). This supports using VHSI in levels in the pooled regressions.

Table A6: ADF Test Results for TechVolatility

ETF	ADF Statistic	p-value
BoseraStar50IndexETFRMB	-6.2900	0.0010
CSOPHangSengTECHIndexETFHKD	-6.2900	0.0010
ChinaAMCHangSengTechIndexETFUSD	-6.2900	0.0010
ChinaAMCSSEScience_TechInnovBoard50ETF	-6.2900	0.0010
DachengHangSengTechnologyETF_QDII_	-6.2900	0.0010
EFundCSIOverseasChinaInternet50IndexETF	-6.2900	0.0010
EFundHangSengTechnologyETF_QDII_	-6.2900	0.0010
GlobalXChinaCloudComputingETFUSD	-6.2900	0.0010
GlobalXChinaSemiconductorETFUSD	-6.2900	0.0010
Huatai_PineBridgeCSOPHangSengTECHIdxETFQDII	-6.2900	0.0010
HwabaoWPCSIFintechThemeIndexETF	-6.2900	0.0010
InvescoGreatWallCSIHKConnectTechIndexETF	-6.2900	0.0010
NikkoAMGlobalInternetETFUSD	-6.2900	0.0010
NikkoAMMetaverseThemeActiveETFUSD	-6.2900	0.0010
PenghuaCNISemiconductorChipsIndexETF	-6.2900	0.0010
PremiaChinaSTAR50ETFUSD	-6.2900	0.0010
SamsungBloombergGlobalSemiconductorETF	-6.2900	0.0010
SamsungKodexSemiconETF	-6.2900	0.0010
iSharesHangSengTECHETFUSD	-6.2900	0.0010

Notes: This table shows ADF test results for the daily volatility of the technology sector index, used as a proxy for sector-specific risk. The p-values indicate strong rejection of the unit root null for all ETFs, confirming stationarity of the volatility series and validating their inclusion in level form in the models.

Table A7: ADF Test Results for FeedbackTrading

ETF	ADF Statistic	p-value
BoseraStar50IndexETFRMB	-22.0500	0.0010
CSOPHangSengTECHIndexETFHKD	-22.0500	0.0010
ChinaAMCHangSengTechIndexETFUSD	-22.0500	0.0010
ChinaAMCSSEScience_TechInnovBoard50ETF	-22.0500	0.0010
DachengHangSengTechnologyETF_QDII_	-22.0500	0.0010
EFundCSIOverseasChinaInternet50IndexETF	-22.0500	0.0010
EFundHangSengTechnologyETF_QDII_	-22.0500	0.0010
GlobalXChinaCloudComputingETFUSD	-22.0500	0.0010
GlobalXChinaSemiconductorETFUSD	-22.0500	0.0010
Huatai_PineBridgeCSOPHangSengTECHIdxETFQDII	-22.0500	0.0010
HwabaoWPCSIFintechThemeIndexETF	-22.0500	0.0010
InvescoGreatWallCSIHKConnectTechIndexETF	-22.0500	0.0010
NikkoAMGlobalInternetETFUSD	-22.0500	0.0010
NikkoAMMetaverseThemeActiveETFUSD	-22.0500	0.0010
PenghuaCNISemiconductorChipsIndexETF	-22.0500	0.0010
PremiaChinaSTAR50ETFUSD	-22.0500	0.0010
SamsungBloombergGlobalSemiconductorETF	-22.0500	0.0010
SamsungKodexSemiconETF	-22.0500	0.0010
iSharesHangSengTECHETFUSD	-22.0500	0.0010

Notes: This table reports ADF test statistics for the Feedback Trading variable, which captures the rolling correlation between current and lagged ETF returns. The extremely low p-values confirm stationarity of the variable across all ETFs, enabling its inclusion in level form in dynamic and static regression models.

Table A8: ADF Test Results for ETF Prices and NAVs (Trend Included)

ETF	ADF (Price)	p-value	ADF (NAV)	p-value
BoseraStar50IndexETFRMB	-2.86	0.1780	-2.86	0.1753
CSOPHangSengTECHIndexETFHKD	-3.12	0.1044	-3.10	0.1083
ChinaAMCHangSengTechIndexETFUSD	-3.14	0.0989	-3.10	0.1079
ChinaAMCSSEScience_TechInnovBoard50ETF	-2.81	0.1931	-2.78	0.2067
DachengHangSengTechnologyETF_QDII_	-2.90	0.1623	-3.12	0.1030
EFundCSIOverseasChinaInternet50IndexETF	-4.11	0.0069	-4.11	0.0070
EFundHangSengTechnologyETF_QDII_	-3.47	0.0441	-3.37	0.0574
GlobalXChinaCloudComputingETFUSD	-2.98	0.1407	-2.58	0.3055
GlobalXChinaSemiconductorETFUSD	-3.37	0.0568	-3.31	0.0655
Huatai_PineBridgeCSOPHangSengTECHIdxETFQDII	-4.04	0.0085	-4.09	0.0074
HwabaoWPCSIFintechThemeIndexETF	-3.39	0.0536	-3.40	0.0526
InvescoGreatWallCSIHKConnectTechIndexETF	-2.97	0.1422	-3.00	0.1346
NikkoAMGlobalInternetETFUSD	-3.27	0.0731	-14.20	0.0010
NikkoAMMetaverseThemeActiveETFUSD	-2.99	0.1376	-6.73	0.0010
PenghuaCNISemiconductorChipsIndexETF	-3.27	0.0733	-3.27	0.0719
PremiaChinaSTAR50ETFUSD	-3.15	0.0959	-3.11	0.1045
SamsungBloombergGlobalSemiconductorETF	-2.26	0.4631	-2.11	0.5375
SamsungKodexSemiconETF	-13.05	0.0010	-13.00	0.0010
iSharesHangSengTECHETFUSD	-3.09	0.1093	-3.11	0.1068

Notes: This table reports ADF test results for ETF prices and net asset values (NAVs), including a linear trend term to account for potential deterministic components. For most ETFs, the ADF statistics do not reject the unit root null at the 5% level, suggesting that both Price and NAV are integrated of order one ($I(1)$). This justifies the use of cointegration, ECM, and VECM models for analyzing the price–NAV relationship.

Section A.5—Diagnostic Testing These figures and tables provide regression diagnostics for the baseline pooled OLS model, including residual normality, heteroskedasticity, and multicollinearity (via VIF). The section supports the diagnostic discussion in Section 5.2.

Table A9: Variance Inflation Factors (VIF) for Pooled OLS Regressors

Variable	**VIF**
Intercept	0.0000
Market Capitalization	3.6323
Market Sentiment (VHSI)	1.5646
Tech Sector Volatility	1.5596
Feedback Trading	1.0058
Liquidity of underlying assets	3.5922
Institutional Ownership	1.0214

Notes: This table reports Variance Inflation Factors (VIFs) for all independent variables included in the baseline pooled OLS model. VIFs indicate the extent of multicollinearity, with values above 5 typically raising concern. All reported VIFs are well below this threshold, suggesting limited multicollinearity among regressors.

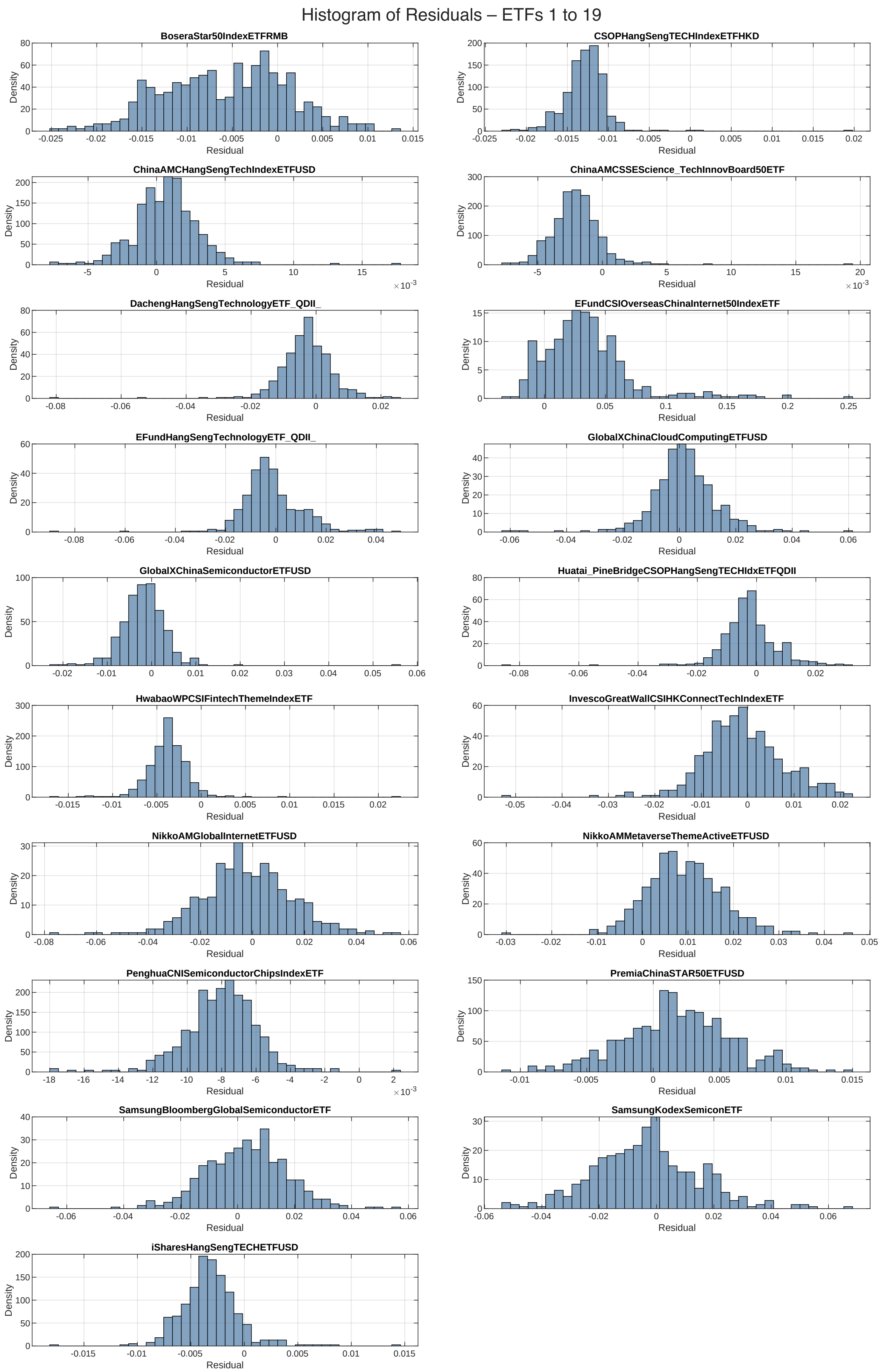


Figure A5: Residual Diagnostics by ETF. Histograms for residuals from the baseline pooled OLS model across all 19 ETFs, split across two panels. Several ETFs exhibit skewed and leptokurtic patterns, indicating potential non-normality and misspecification.

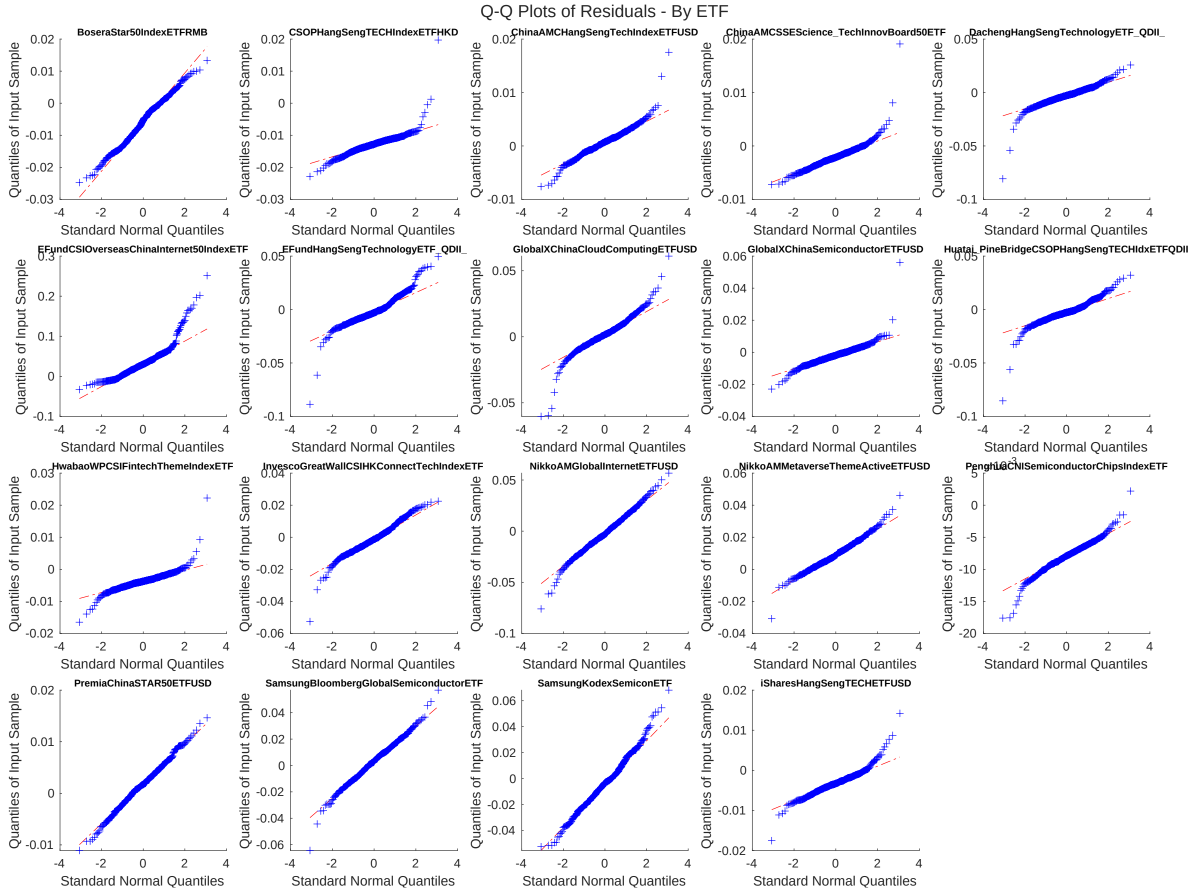


Figure A6: Residual Diagnostics by ETF. Histograms and Q–Q plots for residuals from the baseline pooled OLS model across all 19 ETFs. Several ETFs show distinct skewness and excess kurtosis.

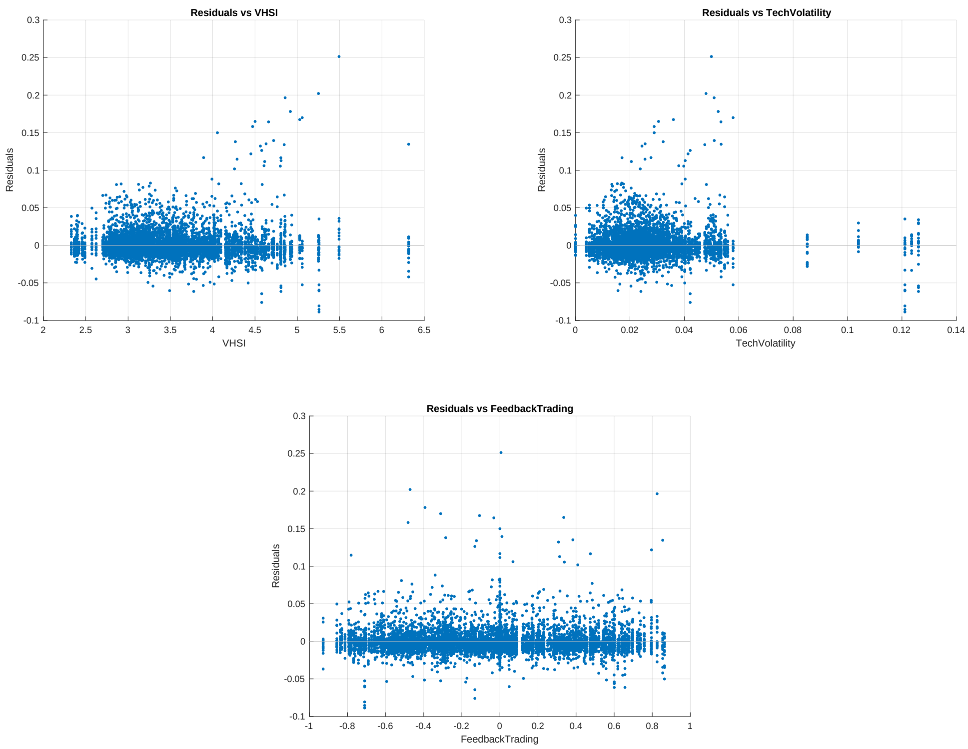


Figure A7: Residuals vs. Main Predictors. Patterns suggest heteroskedasticity and nonlinear behavior, especially at high levels of VHSI and FeedbackTrading.

Section A.6—Subgroup and Nonlinear Regressions This section presents extended pooled OLS regressions exploring heterogeneity in ETF behavior across high/low sentiment, feedback, crisis periods, and nonlinear terms. These directly relate to robustness checks and model extensions in Sections 5.3–5.5.

Table A10: OLS Subset Regressions: High vs Low Market Sentiment (VHSI)

Variable	High VHSI	Low VHSI
Intercept	-0.0339^{***}	-0.0215^{***}
	(0.0050)	(0.0034)
Market Sentiment (VHSI)	0.0010^{***}	0.0008^{***}
	(0.0002)	(0.0001)
Tech Sector Volatility	0.0038^{***}	0.0015^{*}
	(0.0010)	(0.0009)
Feedback Trading	-0.1222^{***}	-0.0516^{**}
	(0.0248)	(0.0252)
Log(Market Capitalization)	0.0009	−0.0005
	(0.0007)	(0.0005)
Log(Liquidity)	0.0002^{***}	0.0001^{**}
	(0.0001)	(0.0001)
% Institutional Ownership	0.0215^{***}	0.0137^{***}
	(0.0019)	(0.0012)
Observations	4,427	4,446
R-squared	0.0874	0.0911
RESET p-value	0.0000	0.0000

Notes: This table reports pooled OLS regressions separately for high and low market sentiment, using a median split on VHSI (volatility index). High VHSI reflects elevated market uncertainty. The table shows heterogeneous effects across regimes, particularly stronger feedback trading and volatility coefficients under high sentiment. Robust standard errors in parentheses. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A11: OLS Subset Regressions: High vs Low Feedback Trading

Variable	**High Feedback**	**Low Feedback**
Intercept	-0.0280^{***}	-0.0266^{***}
	(0.0035)	(0.0037)
Market Sentiment (VHSI)	0.0009^{***}	0.0009^{***}
	(0.0002)	(0.0001)
Tech Sector Volatility	0.0025^{***}	0.0033^{***}
	(0.0007)	(0.0009)
Feedback Trading	-0.0581^{***}	-0.1738^{***}
	(0.0209)	(0.0390)
Log(Market Capitalization)	-0.0013	0.0006
	(0.0011)	(0.0010)
Log(Liquidity)	0.0002^{***}	0.0001^{**}
	(0.0001)	(0.0001)
% Institutional Ownership	0.0190^{***}	0.0162^{***}
	(0.0017)	(0.0015)
Observations	4,427	4,446
R-squared	0.0894	0.0850
RESET p-value	0.0000	0.0000

Notes: Subsample regressions based on a median split of the feedback trading indicator. High feedback captures greater return-chasing behavior. Results indicate nonlinear responses to feedback effects, with sharper mispricing reactions in the low feedback group. Robust standard errors in parentheses. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A12: OLS Subset Regressions: Pre- vs Post-March 2022

Variable	**Pre-March 2022**	**Post-March 2022**
Intercept	0.0079	-0.0295^{***}
	(0.0072)	(0.0028)
Market Sentiment (VHSI)	-0.0006	0.0011^{***}
	(0.0004)	(0.0001)
Tech Sector Volatility	-0.0006	0.0028^{***}
	(0.0013)	(0.0006)
Feedback Trading	-0.0814	-0.1057^{***}
	(0.0693)	(0.0208)
Log(Market Capitalization)	-0.0014	0.0004
	(0.0014)	(0.0004)
Log(Liquidity)	0.0004^{***}	0.0001^{***}
	(0.0002)	(0.0000)
% Institutional Ownership	0.0099^{***}	0.0182^{***}
	(0.0031)	(0.0012)
Observations	665	8,208
R-squared	0.0431	0.0895
RESET p-value	0.0003	0.0000

Notes: This table compares pricing behavior before and after March 2022, marking a crisis period (e.g., geopolitical or macroeconomic shocks). Post-crisis observations exhibit stronger and more significant coefficients across behavioral and volatility variables. Robust standard errors in parentheses. Significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A13: Comparison of Nonlinear OLS Specifications

Variable	Model 1: VHSI^2	Model 2: Feedback^2	Model 3: MktCap × Liquidity
Intercept	-0.0175^{**}	-0.0269^{***}	0.0655^{***}
	(0.0076)	(0.0025)	(0.0056)
Log(Market Capitalization)	0.0009^{***}	0.0009^{***}	-0.0038^{***}
	(0.0001)	(0.0001)	(0.0003)
Market Sentiment (VHSI)	−0.0026	0.0026^{***}	-0.0075^{***}
	(0.0044)	(0.0005)	(0.0005)
Tech Sector Volatility	-0.1066^{***}	-0.0992^{***}	0.0026^{***}
	(0.0203)	(0.0196)	(0.0005)
Feedback Trading	0.0002	0.0002	0.0004^{***}
	(0.0004)	(0.0004)	(0.0000)
Log(Liquidity, 90-day Volume)	0.0002	0.0002	-0.1013^{***}
	(0.0000)	(0.0000)	(0.0199)
% Institutional Ownership	0.0176^{***}	0.0176^{***}	0.0002
	(0.0011)	(0.0011)	(0.0004)
VHSI^2	0.0007	–	–
	(0.0006)		
Feedback Trading^2	–	−0.0013	–
		(0.0010)	
MktCap × Liquidity	–	–	0.0113^{***}
			(0.0009)
Observations	8,873	8,873	8,873
R-squared	0.0856	0.0853	0.1097
RESET p-value	0.0000	0.0000	0.0000

Notes: This table presents three pooled OLS regressions exploring nonlinear effects via squared sentiment and feedback terms, and the interaction between Market Capitalization and Liquidity. Each model includes standard controls, with robust standard errors in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Section A.7—Dynamic Model Tests (ECM/VECM) These tables report Wald tests for the significance of cointegrating and short-run error correction terms across ETFs.

Table A14: Wald Test Results for ECM (Joint Significance)

ETF	F-statistic	p-value
Bosera STAR 50 Index ETF	2505.19	0.0000
CSOP Hang Seng TECH Index ETF	29297.47	0.0000
ChinaAMC Hang Seng Tech Index ETF	49390.31	0.0000
ChinaAMC STAR 50 ETF	24475.62	0.0000
Dacheng Hang Seng Technology ETF	3135.80	0.0000
Global X Cloud Computing ETF	1140.20	0.0000
Global X Semiconductor ETF	4349.60	0.0000
Hwabao FinTech Theme ETF	85292.72	0.0000
Invesco Great Wall Tech ETF	2754.27	0.0000
Penghua Semiconductor Chips ETF	37764.89	0.0000
Premia China STAR 50 ETF	6027.25	0.0000
Samsung Global Semiconductor ETF	1020.62	0.0000
iShares Hang Seng TECH ETF	28738.13	0.0000

Notes: This table reports the Wald test results for the Error Correction Model (ECM) estimates of ETF price deviations from NAV. The test assesses the joint significance of the lagged short-run variables, including lagged deviations and differences in NAV and price. High F-statistics and p-values below 0.01 indicate strong short-run dynamics for nearly all ETFs. These results support the existence of meaningful error-correcting behavior toward long-run equilibrium.

Table A15: Wald Test Results for VECM (Joint Significance)

ETF	F-statistic	p-value
Bosera STAR 50 Index ETF	1.88	0.1544
CSOP Hang Seng TECH Index ETF	0.62	0.5364
ChinaAMC Hang Seng Tech Index ETF	0.93	0.3946
ChinaAMC STAR 50 ETF	0.52	0.5922
Dacheng Hang Seng Technology ETF	4.35	0.0134
Global X Cloud Computing ETF	1.84	0.1594
Global X Semiconductor ETF	0.19	0.8237
Hwabao FinTech Theme ETF	1.39	0.2512
Invesco Great Wall Tech ETF	5.08	0.0066
Penghua Semiconductor Chips ETF	2.41	0.0910
Premia China STAR 50 ETF	1.24	0.2896
Samsung Global Semiconductor ETF	3.48	0.0318
iShares Hang Seng TECH ETF	5.14	0.0062

Notes: This table reports Wald test statistics for the Vector Error Correction Model (VECM) applied to ETF price and NAV series. The test examines the joint significance of short-run dynamic terms in the VECM equations. While a few ETFs (e.g., iShares Hang Seng TECH and Invesco Great Wall) display significant dynamics ($p<0.05$), most results show weak short-term feedback once cointegration is controlled, suggesting limited predictive power of lagged deviations in multivariate settings.

Appendix B: Robustness and Sensitivity Checks

This Appendix refers to the results in section 4.4 of Robustness and Sensitivity checks.

Section B.1—Crisis Interactions These regressions evaluate whether ETF deviations responded differently during the March 2022 market shock and the 2022 U.S. rate hike, depending on behavioral characteristics.

Table B1: Pooled OLS with Interaction Terms — March 2022 Crisis × Behavioral Variables

Variable	Crisis × VHSI	Crisis × Feedback	Crisis × Tech Volatility
Intercept	-0.0277^{***}	-0.0266^{***}	-0.0265^{***}
	(0.0025)	(0.0025)	(0.0025)
Log(Market Capitalization)	0.0009^{***}	0.0009^{***}	0.0009^{***}
	(0.0001)	(0.0001)	(0.0001)
Market Sentiment (VHSI)	0.0028^{***}	0.0025^{***}	0.0020^{***}
	(0.0005)	(0.0005)	(0.0005)
Tech Sector Volatility	-0.0939^{***}	-0.1021^{***}	-0.0384^{*}
	(0.0208)	(0.0200)	(0.0228)
Feedback Trading	0.0004	−0.0002	0.0003
	(0.0004)	(0.0004)	(0.0004)
Log(Liquidity, 90-day Volume)	0.0002^{***}	0.0002^{***}	0.0002^{***}
	(0.0000)	(0.0000)	(0.0000)
% Institutional Ownership	0.0176^{***}	0.0176^{***}	0.0176^{***}
	(0.0011)	(0.0011)	(0.0011)
March 2022 Dummy	0.0162^{*}	0.0001	0.0058^{***}
	(0.0087)	(0.0011)	(0.0019)
Interaction Term	-0.0040^{*}	0.0061^{**}	-0.1337^{***}
	(0.0023)	(0.0028)	(0.0404)
Observations		8,873	
R-squared		0.0865–0.0875	
RESET p-value		< 0.0001	

Notes: This table compares three pooled OLS models, each interacting the March 2022 crisis dummy with a behavioral variable. Standard errors in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Log-transformed variables are used for Market Capitalization and Liquidity.

Table B2: Comparison of Linear and Nonlinear Pooled OLS with Rate Hike 2022 Dummy

Variable	Linear Rate Hike 2022 OLS	Nonlinear Rate Hike 2022 OLS
Intercept	0.0000	-0.0005^{**}
	(0.0001)	(0.0002)
Log(Market Capitalization)	-0.0006^{**}	0.0060^{***}
	(0.0003)	(0.0020)
Market Sentiment (VHSI)	0.0001^{***}	-0.0017^{***}
	(0.0000)	(0.0003)
Tech Sector Volatility	0.0211	0.0160
	(0.0151)	(0.0152)
Feedback Trading	−0.0000	0.0000
	(0.0003)	(0.0003)
Log(Liquidity)	0.0013^{*}	0.0012
	(0.0008)	(0.0008)
% Institutional Ownership	0.7139^{***}	0.7000^{***}
	(0.0225)	(0.0231)
$VHSI^2$	—	-0.0291^{***}
		(0.0108)
$Feedback^2$	—	-0.0008^{***}
		(0.0003)
MCap × Liquidity	—	−0.0003
		(0.0007)
Rate Hike 2022 Dummy	-0.0303^{***}	0.0001^{***}
	(0.0109)	(0.0000)
Observations	8,873	
R-squared	0.5283	0.5320
RESET p-value	< 0.0001	< 0.0001

Notes: This table compares linear and nonlinear pooled OLS specifications including a dummy variable for the U.S. Federal Reserve rate hike in March 2022. Nonlinear model includes quadratic terms for VHSI and Feedback Trading, and an interaction between Market Capitalization and Liquidity. Robust standard errors are in parentheses. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Section B.2—Difference-in-Differences (DID) This nonlinear DID specification tests whether high-volatility ETFs behaved differently post-June 2022.

Table B3: Difference-in-Differences: Nonlinear Specification (High-Volatility ETFs)

Variable	Coefficient	Robust SE	p-value
Intercept	0.0060	0.0074	0.4162
Log (Market Capitalization)	−0.0000	0.0002	0.9728
Market Sentiment (VHSI)	0.0028***	0.0009	0.0030
Tech Sector Volatility	0.0001	0.0010	0.9404
Feedback Trading	−0.0007**	0.0003	0.0255
log (Liquidity)	0.0030	0.0027	0.2664
% Institutional Ownership	−0.0154	0.0207	0.4558
Post (Jul-Dec 2022)	0.0002	0.0003	0.4536
Treated (High-Volatility)	−0.0013***	0.0004	0.0003
Post × Treatment (DID)	0.0024***	0.0007	0.0011
AR (1) Lagged Deviation	0.6727***	0.0247	0.0000
$VHSI^2$	−0.0004	0.0004	0.2989
$Feedback^2$	−0.0006	0.0007	0.4056
Market Cap × Liquidity	0.0001***	0.0000	0.0001
Observations	8,873		
R-squared	0.5326		
RESET p-value	< 0.0001		

Notes: This table presents nonlinear Difference-in-Differences (DID) estimates for high-volatility ETFs. Nonlinear terms include $VHSI^2$, $Feedback^2$, and the MarketCap × Liquidity interaction. Robust standard errors are reported. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Section B.3—Seasonality Controls This section presents models with monthly dummies to account for seasonal patterns in ETF deviations.

Table B4: Pooled OLS with Monthly Dummies (Seasonality Model)

Variable	Coefficient	Robust SE	t-statistic	p-value
Intercept	-0.0256^{***}	0.0024	-10.74	0.0000
Log(MarketCap)	0.0009^{***}	0.0001	8.74	0.0000
VHSI	0.0012^{**}	0.0005	2.40	0.0166
TechVolatility	-0.0644^{***}	0.0212	-3.04	0.0024
FeedbackTrading	0.0003	0.0004	0.61	0.5443
Log(Liquidity)	0.0002^{***}	0.0000	3.51	0.0004
Institutional Ownership	0.0176^{***}	0.0011	15.55	0.0000
Feb	0.0012^{*}	0.0006	1.96	0.0500
Mar	0.0021^{***}	0.0008	2.75	0.0059
Apr	0.0027^{***}	0.0007	3.71	0.0002
May	0.0036^{***}	0.0008	4.48	0.0000
Jun	0.0014^{**}	0.0006	2.20	0.0279
Jul	0.0011^{*}	0.0006	1.81	0.0702
Aug	0.0030^{***}	0.0006	4.62	0.0000
Sep	0.0046^{***}	0.0007	6.61	0.0000
Oct	0.0074^{***}	0.0012	6.39	0.0000
Nov	0.0025^{***}	0.0008	3.22	0.0013
Dec	0.0005	0.0006	0.82	0.4147
Observations		8,873		
R-squared		0.0963		
RESET test		p = 0.0000		

Notes: This table reports pooled OLS estimates including monthly dummy variables (January omitted). Robust standard errors are used. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table B5: Nonlinear Pooled OLS with Monthly Dummies (Seasonality Model)

Variable	Coefficient	Robust SE	t-statistic	p-value
Intercept	-0.0050^{**}	0.0023	-2.20	0.0281
Log(MarketCap)	-0.0000	0.0001	-0.14	0.8870
VHSI	-0.0026^{***}	0.0003	-8.06	0.0000
TechVolatility	0.0119	0.0101	1.17	0.2405
FeedbackTrading	0.0013^{***}	0.0004	2.96	0.0031
Log(Liquidity)	0.0000	0.0000	0.67	0.5043
Institutional Ownership	0.0006	0.0009	0.65	0.5148
Feb	0.0006^{*}	0.0004	1.83	0.0671
Mar	0.0039^{***}	0.0004	8.71	0.0000
Apr	0.0098^{***}	0.0007	14.41	0.0000
May	0.0078^{***}	0.0005	16.26	0.0000
Jun	0.0053^{***}	0.0005	11.28	0.0000
Jul	0.0179^{***}	0.0013	14.14	0.0000
Aug	0.0140^{***}	0.0007	19.34	0.0000
Sep	0.0067^{***}	0.0006	11.10	0.0000
Oct	0.0072^{***}	0.0008	9.18	0.0000
Nov	0.0052^{***}	0.0006	8.19	0.0000
Dec	0.0018^{***}	0.0004	4.53	0.0000
VHSI^2	0.0002^{***}	0.0001	2.68	0.0074
Feedback^2	-0.0019^{**}	0.0010	-1.98	0.0479
MCap × Liquidity	0.0000^{***}	0.0000	19.45	0.0000
Observations		8,873		
R-squared		0.1500		
RESET test		p = 0.0000		

Notes: This table reports nonlinear OLS estimates with monthly dummies and quadratic interaction terms. January is the base month. Robust standard errors are used. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Section B.4—Fixed Effects Estimations These tables report ETF-level fixed effects and interactions, emphasizing time-invariant characteristics and heterogeneity.

Table B6: Fixed Effects (FE) Regression Results

Variable	Coefficient	Std. Error	t-statistic	p-value
Market Sentiment (VHSI)	0.0026	0.0003	9.2691	0.000***
Tech Sector Volatility	−0.1013	0.0119	−8.5348	0.000***
Feedback Trading	0.0002	0.0003	0.7500	0.453
Observations	8,873			
Number of ETFs (n)	19			
Time Periods (T)	467 (Balanced)			
R-squared	0.0114			
Adjusted R-squared	0.0091			
Wald F-statistic	33.997 (p = 0.0000)			

Notes: The table reports the results from a Fixed Effects (within estimator) panel regression on ETF price deviations. ETF fixed effects are included but not reported. Robust standard errors are clustered at the ETF level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B7: Individual ETF Fixed Effects Estimates

ETF Name	Fixed Effect	Std. Error	t-statistic	p-value	Significance
Bosera STAR 50 Index ETF	−0.01491	0.00100	−14.86	0.000	***
CSOP Hang Seng TECH Index ETF	−0.00653	0.00100	−6.51	0.000	***
ChinaAMC Hang Seng Tech Index ETF	−0.00665	0.00100	−6.62	0.000	***
ChinaAMC SSE STAR 50 ETF	−0.00612	0.00100	−6.10	0.000	***
Dacheng Hang Seng Technology ETF	−0.00688	0.00100	−6.86	0.000	***
EFund CSI Overseas China Internet 50 ETF	0.03808	0.00100	37.95	0.000	***
EFund Hang Seng Technology ETF	−0.00203	0.00100	−2.02	0.043	**
Global X China Cloud Computing ETF	−0.00687	0.00100	−6.85	0.000	***
Global X China Semiconductor ETF	−0.00819	0.00100	−8.17	0.000	***
Huatai-PineBridge CSOP HSTech Index ETF	−0.00463	0.00100	−4.61	0.000	***
Hwabao WP CSI Fintech Theme Index ETF	−0.00668	0.00100	−6.66	0.000	***
Invesco Great Wall CSI HK Connect Tech ETF	−0.00468	0.00100	−4.66	0.000	***
Nikko AM Global Internet ETF	−0.00304	0.00100	−3.03	0.002	***
Nikko AM Metaverse Theme Active ETF	−0.00011	0.00100	−0.11	0.915	
Penghua CNI Semiconductor Chips ETF	−0.00589	0.00100	−5.87	0.000	***
Premia China STAR 50 ETF	−0.00661	0.00100	−6.59	0.000	***
Samsung Bloomberg Global Semiconductor ETF	−0.00797	0.00100	−7.94	0.000	***
Samsung KODEX Semiconductor ETF	−0.00768	0.00100	−7.65	0.000	***
iShares Hang Seng TECH ETF	−0.00652	0.00100	−6.50	0.000	***

Notes: This table reports estimated ETF-specific fixed effects from the within estimator panel model. These coefficients capture ETF-level characteristics that are constant over time, such as trading venue, index structure, or fund size. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B8: Fixed Effects Model — Including Volatility × Feedback Interaction

Variable	**Coefficient**	**Std. Error**	**t-statistic**	**p-value**
Market Sentiment (VHSI)	0.002681	0.000282	9.5209	0.000***
Tech Sector Volatility	−0.1030	0.011859	−8.6889	0.000***
Feedback Trading	−0.0025	0.000618	−3.9800	0.000***
Volatility × Feedback	0.0984	0.019017	5.1745	0.000***
Observations	8,873			
Number of ETFs (n)	19			
Time Periods (T)	467 (Balanced)			
R-squared (within)	0.0144			
Adjusted R-squared	0.0119			
Wald F-statistic	32.27 (p = 0.0000)			

Notes: This table extends the fixed effects specification by including an interaction term between Technology Sector Volatility and Feedback Trading. All variables, including the interaction, are statistically significant at the 1% level. ETF fixed effects are included but not reported. Standard errors are robust and clustered at the ETF level. Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.