Specification Sensitivity and Non-Standard Errors in the Idiosyncratic Volatility Effect: Evidence from CEE Equity Markets

Tanveer Ahmad[1], Izabela Pruchnicka-Grabias[2], Jörg Prokop[3]

Abstract

This paper examines the robustness of the idiosyncratic volatility (IVOL) effect in Central and Eastern European (CEE) equity markets through a sensitivity analysis of a systematic research design. The full design grid contains 518,400 distinct research designs, generated using IVOL estimation, portfolio specification, a weighting scheme, and sampling filters. The final sample includes daily data on more than 1,000 stocks from 2001 through November 2025. Across the design space, the spread between high and low IVOL portfolios is highly sensitive to data frequency and implementation details. Daily frequency designs generate substantially higher IVOL premium portfolio than monthly designs. Similarly, an equal-weighting scheme produces a much larger spread than value weighting or capped value weighting. Portfolio sort granularity also matters, increasing the number of portfolios from the tercile to the decile level raises the mean spread from 0.80% to 1.34%. To quantify non-standard errors (NSEs), we compare conventional time-series standard errors with cross-sectional design dispersions. NSE exceeds SE across almost all design choices. Overall, the results indicate that the magnitude and statistical significance of the IVOL effect in CEE markets depend materially on routine yet often underreported design decisions, underscoring the need for transparent reporting and robust inference across different research designs.

JEL classification: G10, G12, G14

Key words: IVOL anomaly, CEE markets, risk, return, asset pricing

1 Tanveer Ahmad, Tallinn University of Technology, Department of Economics and Finance; address: Akadeemia tee 3, 12618 Tallinn, Estonia; tanveer.ahmad@taltech.ee.

2 Izabela Pruchnicka-Grabias, SGH Warsaw School of Economics, Collegium of Socio-Economics, Al. Niepodległości 162, 02-554, Warsaw, Poland, Email: ipruch@sgh.waw.pl, ORCID 0000-0002-7385-6286

3 Jörg Prokop, Carl von Ossietzky Universität Oldenburg, Area Finance and Banking, 26111 Oldenburg, Germany, joerg.prokop@uol.de, ORCID 0000-0002-4743-1559

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly (https://app.grammarly.com/) and DeepL (https://www.deepl.com/) in order to improve language and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the manuscript.

Data availability statement

The data that support the findings of this study are available from LSEG. Restrictions apply to the availability of these data, which were used under license for this study.

Non-Standard Errors in the Idiosyncratic Volatility Effect: Evidence from CEE Equity Markets

1 Introduction

The idiosyncratic volatility (IVOL) anomaly refers to empirical findings that equities with higher unsystematic risk deliver lower subsequent returns, which runs counter to what we would expect based on classical asset pricing theory. The capital asset pricing model (CAPM; Sharpe 1964, Lintner 1965, Black 1972) and the Fama-French three-factor (FF3) model (Fama & French 1993), with Markowitz' (1952) mean-variance framework as a common starting point, suggest a positive mean-variance relationship and posit that equilibrium prices are driven by systematic risk only, while unsystematic risk is diversified and unrelated to subsequent returns. This is in line with empirical evidence by Angelidis and Tessaromatis (2008), who found equally-weighted idiosyncratic volatility to be independent of future market returns in most developed European stock markets.

However, under limited information, Merton (1987) shows that idiosyncratic volatility is positively linked to expected returns because investors cannot fully diversify it away. This dependence was later validated by empirical studies, such as Malkiel & Xu (2004), who analyze the US market and a sample of Japanese stocks and find that idiosyncratic risk positively influences stock returns. In contrast, Ang et al. (2006) use IVOL, derived based on FF3, and postulate that US stocks with high IVOLs report dramatically low average returns, which cannot be explained by the FF3 model. Simultaneously, Ang et al. (2009) confirm the same phenomenon for 23 developed countries, including the G7.

While existing empirical studies yield conflicting results and employ different empirical tools and assumptions, our goal is to assess the robustness of the idiosyncratic volatility (IVOL) effect in Central and Eastern European (CEE) equity markets through a sensitivity analysis of our systematic research design. As CEE equity markets have become an important part of the global financial markets, understanding how they function is vital for investment decisions. Following Brockmann et al. (2022), who stress the influence of market maturity on the IVOL puzzle, we also bear in mind that research conducted on developed markets may not adequately explain stock behavior in CEE markets, which may behave differently due to their lower level of development, specific characteristics, and limited liquidity. In particular, Zaremba (2015) studied the risk-return relationship in CEE markets, but with inconclusive results. He finds that CEE stocks characterized by low idiosyncratic volatility exhibit significant positive abnormal performance, a part of which he attributes to momentum-related dynamics. However, this negative relationship breaks down for microcap firms with high idiosyncratic risk, as they exhibit superior returns.

This paper contributes to the literature on IVOL anomalies in equity markets by examining the specification robustness of the IVOL effect in Central and Eastern European (CEE) equity markets. We conduct a systematic multiverse analysis to investigate the extent to which the confirmation or rejection of the IVOL effect is sensitive to reasonable research design choices. Our empirical results demonstrate that the magnitude and statistical significance of the IVOL effect in CEE markets depend to a substantial extent on the research design. This is particularly true with regard to factors such as data frequency, weighting schemes, and portfolio granularity. Specifically, we find that designs based on daily data generate high-minus-low IVOL return spreads that are on average twice as large as designs based on monthly data. Moreover, equal-weighted portfolios produce substantially higher spreads than value-weighted or capped-value-weighted portfolios. In addition, moving from a three-portfolio sort to a decile sort also raises the mean spread and increases the share of statistically significant results. Finally, with regard to statistical inference, we find that the non-standard error (NSE) exceeds the average bootstrap standard error for almost all model specifications, indicating that t-statistics based on standard errors may underestimate the true level of uncertainty in the estimates. Overall, these findings underscore the importance of transparent reporting of methodological choices and the value of rigorous multiverse-style analytical frameworks in assessing model robustness.

The rest of the paper is structured as follows. Section 2 presents details on sample construction and research design. Section 3 contains the discussion of the empirical results. Section 4 concludes.

2 Literature review

Since Ang (2006) revealed the IVOL anomaly, a substantial body of research has examined its underlying mechanisms across different asset classes, markets, and time periods, yielding mixed evidence regarding its occurrence. To better understand the equity markets, researchers have proposed a broad set of explanations, ranging from psychological factors to company size, market maturity, equity selection, and methodological issues in measuring idiosyncratic risk. Although nearly two decades of research have passed, the IVOL anomaly holds one of the most challenging puzzles in modern asset pricing.

Growing evidence from equity markets indicates that the IVOL anomaly varies with company size. This is shown, for example, by Goyal and Santa-Clara (2003), who find a positive relationship between idiosyncratic volatility and subsequent returns for some US stocks only. This rule is also noted by Ali et al. (2020a), who confirm it only for small firms listed on the Singapore Stock Market. Similarly, Ali et al. (2020b) document that, in the Indian market, the IVOL effect is statistically significant only for small firms. Also, Brown and Ferreira (2005) show that only the idiosyncratic volatilities of small companies are positively correlated with subsequent returns. On the other hand, Aboulamer and Kryzanowski (2016) admit a positive effect of IVOL for the Canadian market, but not for small companies. Also, Ali et al. (2020b) note that, in the Indian market, the IVOL anomaly is significant only for large firms.

Beyond the company itself, many other factors have been identified as influencing the IVOL puzzle. In line with Wei and Ahang (2005), who show that the IVOL anomaly is period-dependent, Fu (2009) emphasizes fluctuations in idiosyncratic volatility over time. Furthermore, Cao et al. (2021) document that the IVOL-return puzzle varies across calendar days, showing a mostly negative relationship on Fridays and a positive one on Mondays. On the other hand, Qadan and Shuval (2022) highlight its dependence on investors' risk aversion; Wang et al. (2023) seek to explain the IVOL problem in terms of speculative behavior; and Stambaugh et al. (2015) attribute the IVOL puzzle to overpricing and arbitrage asymmetry. The literature also highlights the importance of stock market development and its determinants. Brockmann et al. (2022) show, in a broad analysis of the relationship between expected idiosyncratic volatility and returns across 57 countries, including CEE markets, that this relationship diminishes with greater stock exchange development, increased availability of high-quality information, and lower transaction costsIt aligns with Park and Don-Hyun (2025), who stress that the IVOL anomaly depends, to some extent, on beta selection and portfolio implementation costs rather than on the true valuation of volatility.

In addition to the aspects discussed above, the IVOL anomaly may be shaped by the methodology employed. To prove it, Bali & Cakici (2008) question the robustness and significance of the negative relationship between IVOL and expected returns for the main US stock markets, showing that the results depend on parameters used in both IVOL and average return calculations, such as data frequency or stock selection. Similarly, Malkiel & Xu (2004) emphasize the role of the risk measure applied. Consistently, Bali et al. (2005) conclude that the IVOL puzzle is driven by the equity variance weighting scheme, the liquidity premium, and company size, and that it only holds for small caps traded on NASDAQ. By contrast, Hou and Loh (2016) develop a model that includes investors' preferences for abnormal returns, various market frictions, average variance beta, uncertainty, and earnings surprises. Authors demonstrate that, taken together, these factors explain no more than 10% of the IVOL anomaly, leaving a large unresolved component. Coqueret and Perignon (2026) show that resampling affects statistical significance because reducing correlations among methodological paths lowers the variance of the average effect estimator. In this context, the appraisal of the IVOL's significance may change.

Alternatively, Gempesaw et al. (2021) show that the IVOL anomaly can be reduced due to risk diversification. In US markets, they show that when IVOL is substituted for the omitted risk factor, the relationship between IVOL and returns is negative, but this relationship is no longer observed in well-diversified portfolios. Simultaneously, authors notice that it is not only unsystematic risk that is important here, but also systematic risk. They show that adding certain systematic risk factors to the CAPM model also reduces the anomaly. They measure IVOL in the CAPM environment and suggest that anomalies can be explained by both systematic and idiosyncratic risk.

Despite this, a number of concepts have been proposed, trying to explain the IVOL phenomenon by investors' psychological behavior, according to the cumulative prospect theory, leading to overestimation of lottery-like tail events (Kahneman & Tversky, 1992), which causes overestimation of the prices of stocks with outlying positive returns (Barberis and Huang 2008; Bali et al. 2011) and resulting from it their lower future returns.

Keeping in mind a broad range of empirical explanations underlying the IVOL anomaly in equity markets, we extend the above lines of research by conducting distinct research designs based on IVOL estimation, portfolio specification, weighting schemes, and sampling filters in CEE equity markets, thereby showing the role of design decisions in the process of IVOL anomaly exploration.

3 Data and methodology

3.1 Data

We use a panel of CEE common stocks with return data and firm characteristics required for portfolio construction. Our sample comprises more than 1,000 stocks from 2001 through November 2025. The sample is aligned to a monthly rebalancing frequency. We excluded financial, real estate, government-related, academic and educational services and observations with null sector classifications. At the observation level, non-positive prices and market capitalization are treated as missing values. To avoid look-ahead bias, all sorting signals are lagged so that portfolio formation at time t uses only information available at or before time $t-1$.

3.2 Methodology

3.2.1 Decision nodes

Researchers inevitably make a sequence of implementation choices when constructing portfolio sorts, and many of these choices are not uniquely determined by theory or by a single accepted convention. Following the operations research perspective adopted in the non-standard errors literature, we label each such choice a decision node. Each decision node offers two or more plausible branches, and once a branch is selected, the analysis proceeds through subsequent nodes until it reaches a terminal node. A terminal node corresponds to one fully specified empirical implementation, which we refer to as a research design. We have divided the decision nodes into four broad categories, each containing several decision nodes. The first category is IVOL construction, i.e., the model used to estimate IVOL, the estimation window, and the minimum number of observations required for estimation. The second category is sample filters, i.e., the minimum market size, minimum stock age, and minimum closing price required for inclusion in the sample. The third category is portfolio implementation, i.e., the number of portfolios, portfolio type, the minimum number of stocks required per portfolio, and the number of portfolios when using the bivariate portfolio formation methodology and dropping small-cap stocks at breakpoints. Lastly, we have two decision nodes for weighting and caps winsorization: equal weighting, value weighting, and value-weighted cap, which is further subdivided into winsorization at the 80^{th} and 90^{th} percentiles.

3.2.2 Estimating idiosyncratic volatility (IVOL)

Table 1 shows that prior IVOL studies differ in several implementation choices, including return frequency, estimation window, minimum-observation requirement, and the factor model used to estimate residual returns. Following this literature, the first IVOL construction decision node concerns the factor model used to define IVOL. We estimate IVOL from rolling time-series regressions of individual stock excess returns on common risk factors.

For the CAPM specification, we estimate:

$$r_{i,t} = \alpha_i + \beta_{i,M}(RM_t - Rf_t) + e_{i,t} \tag{1}$$

For the FF3[4] specification, we estimate:

$$r_{i,t} = \alpha_i + \beta_i^{MKT} MKT_t + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + e_{i,t} \tag{2}$$

In the above equations, $r_{i,t}$ denotes the excess return of stock i at time t, measured as the stock return minus the risk-free rate. $RM_t - Rf_t$ is the market excess return used in the CAPM specification. In the FF3 specification, MKT_t denotes the market excess return, SMB_t is the size factor, defined as the return differential between small stock and big stock portfolios, and HML_t is the value factor, defined as the return differential between high book-to-market and low book-to-market portfolios. IVOL is calculated as the standard deviation of these residuals within each rolling estimation window.

[4] SMB and HML are constructed using Fama and French (1993) methodology.

Table 1. Construction of Idiosyncratic Volatility (IVOL) Across Studies.

This table summarizes how prior papers compute IVOL, reporting the return data frequency (daily vs. monthly), the estimation window, minimum-observation required and the model used for estimation.

Paper	Data frequency	Window	Minimum obs.	Factor model
Ang et al. (2006)	Daily	Within month	17 daily obs.	FF3
Bali & Cakici (2008)	Daily and Monthly	(1) Daily: Within month; (2) Monthly: past 60 months	24 months	CAPM and FF3
Ang et al. (2009)	Daily	Within month		Regional FF3
Bhootra & Hur (2015)	Daily	Within month		FF3
Aboulamer & Kryzanowski (2016)	Daily	Within month		Carhart 4-factor (1997)
Hou & Loh (2016)	Daily	Within month	10 and 15 daily obs.	FF3
Herskovic, Kelly, Lustig & Van Nieuwerburgh (2016)	Daily	1 year daily obs	Should not have missing value in a year	Market and FF3 model
Chichernea et al. (2018)	Daily	Within month	≥ 15	FF3
Ali et al. (2020)	Daily	Within month		CAPM and FF3
Chen, Jiang, Xu & Yao (2020)	Daily	Within month	≥ 15	FF3
Cao, Chordia & Zhan (2021)	Daily	Within month	17 daily obs. and 15 obs. for robustness	FF3
Barinov (2023)	Daily and Monthly	(1) Daily: Within month; (2) Monthly: past 60 months	(1) Daily ≥15; (2) Monthly ≥ 24	FF3
Poon et al. (2022)	Daily and Monthly	(1) Daily: Within month; (2) Monthly: past 60 months	(1) Daily ≥15; (2) Monthly ≥ 36	FF3
Bali & Weigert (2024)	Monthly	36 months		Six-factor model = FF5 (2015) + UMD momentum (Carhart 1997)
Goh & Kim (2024)	Daily	252	65	CAPM

3.2.3 Window length

The second IVOL construction decision node is the length of the estimation window over which the factor model is fit, and residuals are collected. Window length governs a bias-variance tradeoff. Shorter windows are more responsive to recent changes in a firm's risk profile but can yield noisier residual volatility estimates, whereas longer windows smooth transitory variation but may be less representative if risk exposures are evolving. We implement multiple window choices that map directly to common practice in the literature and to the frequency of the underlying returns. For daily estimates, we consider the periods within 1 month, 6 months, and 12 months of the daily observations. For monthly estimations, we use a 60-month window of monthly returns, which corresponds to a long-horizon specification frequently adopted by Bali et al. (2008). These alternatives allow us to assess whether the IVOL effect and its statistical properties depend on the horizon over which idiosyncratic risk is measured.

3.2.4 Minimum Observations

The third IVOL construction decision node is the minimum number of observations required within a window to compute IVOL. Imposing a minimum-observation rule mitigates measurement error from sparsely populated windows and ensures that IVOL is based on a sufficiently informative sample of residuals, but stricter requirements can shorten the effective panel and disproportionately exclude younger or less liquid firms. In our daily implementations, we consider minimum-observation thresholds of 15, 17 days within a month, and a 50% window for 6- and 12-month daily observations. In our monthly implementations, we impose minimum thresholds of 24 and 36 months of observations. By varying these thresholds, we directly examine how sensitive IVOL estimates are to the trade-off between retaining stock coverage and enforcing estimation precision.

3.2.5 Minimum Age

Newly listed stocks often have short, discontinuous return histories, elevated volatility, and greater exposure to microstructure effects, and they may exit the sample quickly due to delisting or corporate events. We therefore treat the minimum age as a decision node and require that a stock has been listed for at least a specified minimum period before it becomes eligible for inclusion in the sample. In the design spectrum, we consider three alternative minimum-age requirements: no restriction, a minimum age of 1 year, and a minimum age of 2 years.

3.2.6 Min Market capitalization

Many cross-sectional return patterns are strongest among the smallest firms, yet these firms are often illiquid and costly to trade; hence, they do not fall within the tradable universe. In CEE equity markets, where the firm size distribution is typically highly skewed, and microcaps can be thinly traded, including very small stocks may inflate measured return spreads while reducing economic interpretability. We therefore treat minimum capitalization as a decision node and consider two alternative approaches: (i) no size filter, which maximizes coverage and reflects the full listed universe, and (ii) excluding stocks with market capitalization below a fixed minimum threshold, i.e., 0.1, 0.5, and 1 million market capitalization. This reduces the influence of microcaps and improves investability, albeit at the cost of a smaller sample size.

3.2.7 Minimum Share price

Very low-priced stocks ("penny stocks") are often illiquid, distressed, and subject to large proportional price moves and trading frictions, which can mechanically amplify anomaly returns while reducing practical tradability. In CEE markets, where thin trading and microstructure effects can be more pronounced for low-priced names, the inclusion or exclusion of penny stocks can materially affect the composition of the investable universe and the measured IVOL portfolio spreads. We therefore treat the minimum price as a decision node and consider two alternative choices, i.e., no price filter, which retains the full universe, and (ii) excluding stocks with a price below a fixed minimum threshold, i.e., \$0.1, \$0.2, \$0.5, \$1.

3.2.8 Drop Small Cap for Breaking points

The smallest stocks can exert disproportionate influence on the estimated breakpoints despite being illiquid and economically less relevant, which may distort portfolio assignments and the resulting return spreads. To address this, we introduce a "drop small cap for breakpoints" rule. Under this procedure, we first exclude the bottom fraction of the market-capitalization distribution, specifically 0%, 10%, or 20%, when computing the sorting breakpoints. Importantly, this exclusion is used only for breakpoint estimation; once the breakpoints are determined, they are applied to the full universe of stocks, including the previously excluded small-cap names, when assigning securities to portfolios.

3.2.9 Number of portfolios

Sorting stocks into IVOL-ranked portfolios is a standard test for monotonic relationships in the cross-section of returns, and the return spread between the highest- and lowest-IVOL portfolios provides a direct measure of the IVOL premium. However, the granularity of the sort is itself a research choice: using more portfolios can better capture nonlinearities and sharpen the extreme spread, but it also requires a sufficiently large cross-section and can become noisy in smaller markets such as those in CEE. We therefore treat the number of portfolios as a decision node and consider alternative one-way sort granularities, implementing terciles, quintiles, and deciles portfolios design.

3.2.10 Sort Type

Double sorting extends the standard single-sort framework by examining whether the return pattern associated with the primary sorting variable persists after conditioning on an economically important secondary characteristic, most commonly firm size. Because IVOL can be correlated with size, single-sort spreads may partly reflect size-related effects rather than an independent IVOL premium. We therefore implement both independent and dependent double sorts with size as the secondary sorting variable. In dependent double sorts, we first sort stocks into size portfolios and then sort within each size portfolio into IVOL groups, so that the IVOL comparison is made conditional on firm size. In independent double sorts, we compute size breakpoints using the full cross-section and assign size groups independently of IVOL; we compute IVOL breakpoints independently in the same manner, and then intersect the size and IVOL assignments to form the bivariate portfolios.

3.2.11 Double Sorting Breaking Point

Conditional on performing double sorts, an additional design choice concerns the granularity of the secondary breakpoints. Using a small number of size groups (e.g., dividing on the basis of size into two categories based on median) provides a coarse "small vs big" conditioning set that is feasible even when the cross-section is limited, whereas using finer breakpoints (e.g., 5) allows a more detailed assessment of whether the IVOL premium is stable across the size distribution or driven by extreme small or extreme large firms. We therefore allow two alternative choices for the number of secondary-size portfolios: 2 or 5.

3.2.12 Minimum number of assets

The feasibility and reliability of fine portfolio sorts depend on having sufficient stocks available on each formation date, especially after applying eligibility screens and IVOL data requirements. If the cross-section is thin, breakpoint-based sorting can generate portfolios with very few constituents, making portfolio returns unstable and sensitive to idiosyncratic movements of individual stocks. We therefore treat the minimum number of stocks as a separate decision node and require at least a specified minimum number of constituents per portfolio. In the design spectrum, we allow alternative minimum portfolio size rules (e.g., 1, 5, 10, or 50 stocks).

3.2.13 Weighting Scheme

After assigning stocks to IVOL-sorted portfolios, we must decide how to aggregate individual stock returns into a single portfolio return. This weighting choice matters because value-weighted portfolios are typically viewed as more realistic and investable, as they concentrate on larger, more tradable firms, whereas equal-weighted portfolios place relatively greater weight on smaller, often less liquid stocks and can therefore deliver different return spreads. We therefore implement three alternatives: equal-weighted (EW) returns, standard value-weighted (VW) returns based on market capitalization, and an intermediate capped value-weighted (Cap-VW) approach that limits dominance by a few very large firms by capping the upper tail of market-cap weights each month.

3.2.14 Capped value-weighting

This decision node is only applied to the capped-value-weighted scheme. Each month, we do one-sided winsorize market capitalization in the upper tail, using caps at the 80th and 90th percentiles, so that any market cap above the chosen cutoff is replaced by the cutoff, while smaller firms remain unchanged. This yields portfolios that remain investable but is less mechanically dominated by mega-cap stocks.

Table 2 illustrates each decision node in the research design. Based on the above decision nodes, we have 518,400 distinct research designs. However, we note that decision nodes are not necessarily independent: price and size filters can remove overlapping sets of firms because low-priced stocks are frequently also small-cap, so applying both screens may partially duplicate their effects on the sample. Furthermore, we included only designs that ended up with 50% of the initial number of months in the sample. Hence, there were 423,912 feasible designs, and the following analysis is based on these. Figure 1 shows the sequential decision nodes used to construct alternative IVOL.

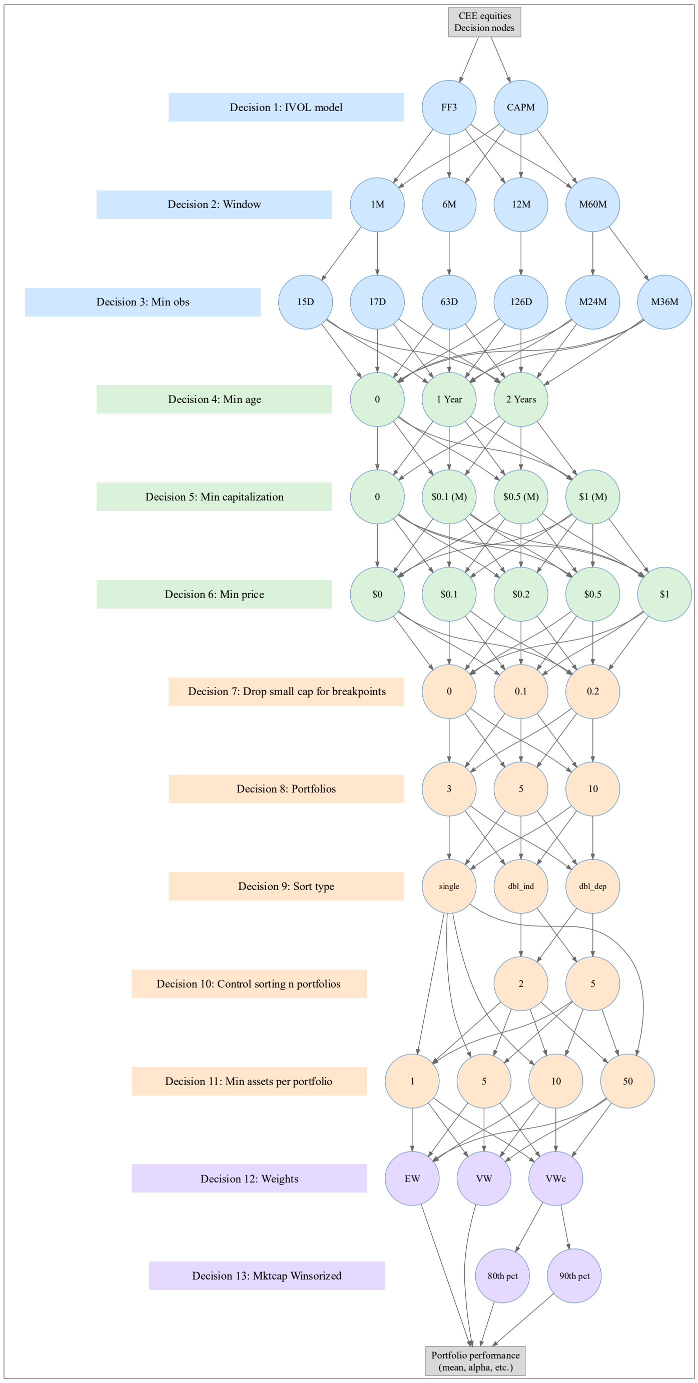
Table 2. Research-design grid and cumulative number of specifications.

This table summarizes alternative research design choices for sample preparation and portfolio construction used to assess the robustness of IVOL. Panel (a) reports the IVOL construction choices, Panel (b) reports sample filtering choices, Panel (c) reports portfolio implementation, and Panel (d) reports return weighting choices.

Design	Choices	No. of Combinations	Cumulative designs
a). IVOL Estimation Choices			
Data Input	Daily, monthly	2	2
Estimation Window	1M (1 month of daily data), 6M (daily obs. of 6 months), 12M, M60 (60 monthly observations)	6	12
Minimum Observations	15D, 17D, 63D, 126D, M24, M36		
b). Sample Filtering Choices			
Minimum age	All, 1 year, 2 years	3	36
Minimum market capitalization	0, 0.1, 0.5, 1 million	4	144
Minimum share price	0, 0.1, 0.2, 0.5, 1	5	720
c). Portfolio Implementation Choices			
Drop small cap for breaking points	0, 0.1, 0.2	3	2160
Number of portfolios	3, 5, 10	3	6480
Sort Type	Univariate, bivariate independent, bivariate dependent	5	32400
Controling sorting in portfolio	2, 5		
Minimum number of assets	1, 5, 10, 50	4	129600
d). Weighting Choices			
Weighting scheme	Equal-, value-, capped-value weighted	4	518400
Market Cap Winsorized	80^{th}, 90^{th} percentile		

Figure 1. Decision-tree representation of the research-design space.

The diagram visualizes the sequential decision nodes used to construct alternative IVOL-based portfolio designs—covering the IVOL model (CAPM vs FF3), estimation window and minimum observations, sample screens (age, size, price, breakpoint exclusions), portfolio formation choices (number of portfolios, sort type, controls, minimum assets), and weighting/winsorization rules. Arrows indicate how each choice branches into subsequent options, generating the full set of admissible specifications that map into the final portfolio-performance outcomes



4 Empirical Results

4.1 Model Estimation Choices

The first category is IVOL construction, i.e., the model used to estimate IVOL, the data frequency, the window, and the minimum observations required for estimation. The H-L return spread is presented in Figure 2 for the 12 research designs related to IVOL construction. The results reveal that IVOL spread is positive in almost all specifications, but its magnitude and dispersion vary strongly with frequency and the estimation window. The H-L spread for IVOL, calculated from daily data, is higher (i.e., a higher premium) than that calculated from monthly data. Furthermore, moving from shorter to longer daily windows tends to slightly lower the center and often changes the distribution's width, i.e., the spread is higher for IVOL generated from a longer window. The choice of model, i.e., CAPM vs. the FF3 factor model, has no impact on the H-L spread.

Figure 2. Distribution of value-weighted H–L mean returns across model and estimation-window specifications.

The figure plots violin densities of the High-minus-Low (H–L) portfolio mean return (value-weighted) for each specification on the y-axis. Labels indicate the sampling frequency (Daily vs Monthly), factor model (CAPM vs FF3), and the estimation window/minimum-observations rule (e.g., *1M_15*). Wider sections of each violin denote a higher concentration of outcomes. Markers indicate central tendency (× for the mean and • for the median).

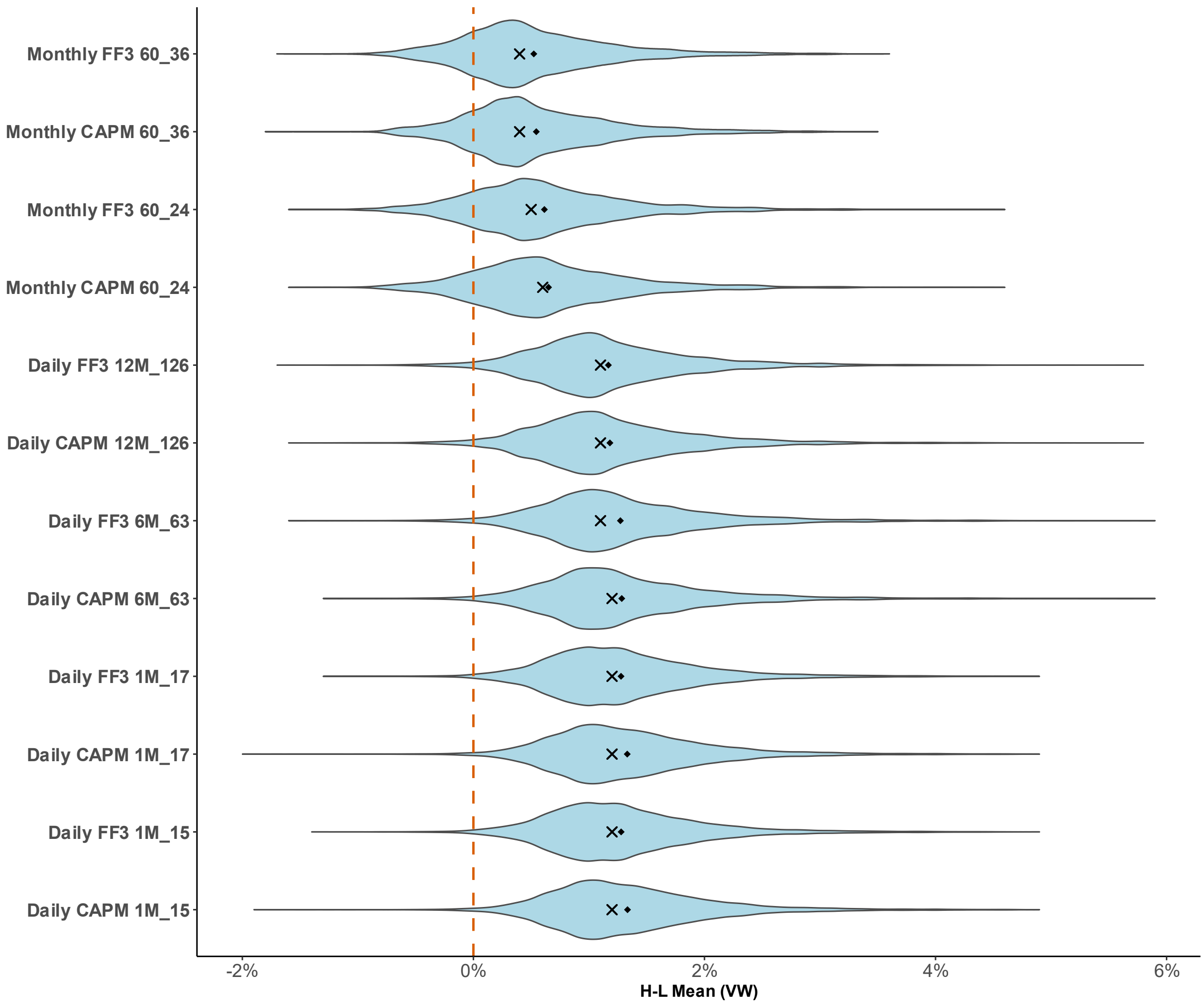


Table 3 shows that the IVOL H-L spread is economically positive across all specifications, but its strength depends strongly on the sampling frequency and the IVOL estimation window. In the daily designs, the man H-L spread is tightly clustered around 1.17-1.33 and monotonically decreases as the window length increases. The spread is positive in almost all simulations (roughly 97-99% of designs have H-L> 0). Statistical significance is also frequent for daily specifications, around two-thirds to more than four-fifths of the designs generate a significant H-L spread at the 5% significance level. By contrast, monthly specifications produce substantially smaller spread means, roughly 0.52 to 0.65, with a noticeably smaller fraction of designs (about 26% to 36%) generating positive H-L spreads at the 5% significance level.

Table 3. Summary statistics for the value-weighted H–L portfolio across specifications.

The table reports, for each model/window design, the mean and median H–L return, the fraction of simulations with H–L > 0, the share of statistically significant outcomes (at 5% level of significance), and the Sharpe ratio.

	Mean Return				
Model	Mean	Median	% H-L > 0	% of **	Sharp Ratio
Daily CAPM 1M_15	1.33	1.20	98.99	84.27	0.78
Daily FF3 1M_15	1.28	1.20	98.78	83.04	0.77
Daily CAPM 1M_17	1.33	1.20	98.95	84.15	0.77
Daily FF3 1M_17	1.28	1.20	98.77	83.00	0.77
Daily CAPM 6M_63	1.28	1.20	98.45	70.78	0.70
Daily FF3 6M_63	1.27	1.10	98.07	71.00	0.70
Daily CAPM 12M_126	1.18	1.10	97.41	66.20	0.67
Daily FF3 12M_126	1.17	1.10	97.33	66.00	0.67
Monthly CAPM 60_24	0.65	0.60	83.52	36.10	0.44
Monthly FF3 60_24	0.61	0.50	81.66	35.47	0.42
Monthly CAPM 60_36	0.54	0.40	80.45	26.57	0.36
Monthly FF3 60_36	0.52	0.40	78.34	26.74	0.35

Figure 3 shows how the estimated value-weighted H-L abnormal performance varies once returns are risk-adjusted using factor models. In panel A (CAPM), the distributions are generally towards the right of zero across both daily and monthly specifications, suggesting that the IVOL H-L spread retains a positive alpha when only the market risk factor is controlled for. In contrast, the FF3 model (panel B) is more tightly centered on zero and often overlaps it, suggesting that, after adjusting for FF3, the risk-adjusted return is close to zero. This suggests that the IVOL premium is less robust to riskier risk adjustments.

Table 4 confirms that the abnormal performance of the value H-L portfolio is highly sensitive to the risk adjustment model and the sampling design. Under the CAPM, daily specifications produce clearly positive alphas. Alpha is positive in roughly 88% to 91% of research designs. However, statistical significance is only moderate, with about 38% to 47% research designs producing significant alpha. In cases where IVOL is estimated using monthly returns, CAPM alphas are small and often near zero, with the share of positive alphas dropping to roughly 45% to 57%, generating alphas greater than zero that are significant in only 19 to 21% of the designs. Similarly, mean FF3 alphas are typically around 0.09 to 0.25; the fraction of alphas greater than 0 falls to roughly 52% to 64%; and only about one quarter of simulations are significant for IVOL construction using daily data. While FF3 alphas become negative on average, alpha is positive in only about 24% to 33% of simulations. Overall, the table implies that the apparent IVOL alpha H–L portfolios are largely a CAPM phenomenon and become weak or even negative once size and value exposures are accounted for, especially at the monthly frequency.

Figure 3. Distribution of value-weighted H–L alphas across factor models and estimation-window specifications.

The figure reports violin plots of the value-weighted High-minus-Low (H–L) portfolio alpha estimated under alternative factor models and rolling-window choices. Panel A shows alphas from the CAPM, while Panel B shows alphas from the Fama–French three-factor (FF3) model. The y-axis lists each specification by sampling frequency (Daily vs Monthly), factor model label, and the rolling estimation window/minimum-observations rule. Violin widths reflect the density of estimated alphas, highlighting dispersion and tail behavior across designs. Markers indicate central tendency (× for the mean and • for the median).

Panel A). CAPM Alpha **Panel B).** FF3 Alpha

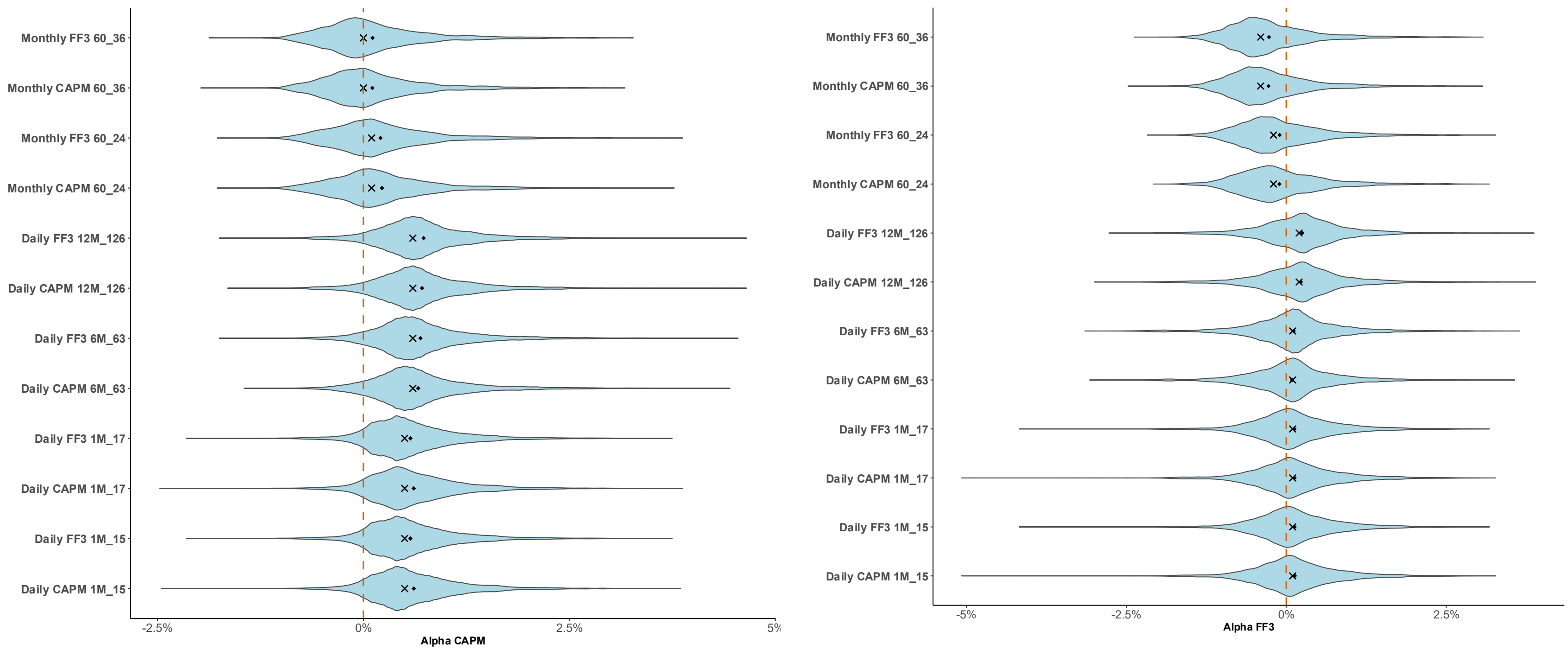


Table 4. Model Alphas H–L portfolio across specifications.
The table reports, for each model/window design, the mean and median alpha for H–L returns using the CAPM and FF3 factor models, the fraction of simulations with alpha > 0, and the share of statistically significant outcomes (at 5% level of significance).

	Capm Alpha				FF3 Alpha			
Model	Mean	Median	% alpha > 0	% of **	Mean	Median	% alpha > 0	% of **
Daily CAPM 1M_15	0.61	0.50	88.59	46.75	0.13	0.10	53.38	25.37
Daily FF3 1M_15	0.57	0.50	87.73	44.19	0.13	0.10	51.87	25.39
Daily CAPM 1M_17	0.61	0.50	88.56	46.77	0.13	0.10	53.11	25.29
Daily FF3 1M_17	0.57	0.50	87.69	44.37	0.13	0.10	51.88	25.50
Daily CAPM 6M_63	0.67	0.60	89.04	38.17	0.09	0.10	52.84	25.59
Daily FF3 6M_63	0.69	0.60	90.16	41.1	0.12	0.10	55.28	25.33
Daily CAPM 12M_126	0.71	0.60	90.55	42.51	0.23	0.20	63.16	25.34
Daily FF3 12M_126	0.73	0.60	91.11	45.34	0.25	0.20	64.05	25.66
Monthly CAPM 60_24	0.22	0.10	57.22	21.52	-0.11	-0.20	32.90	37.85
Monthly FF3 60_24	0.21	0.10	55.48	22.22	-0.10	-0.20	33.34	38.64
Monthly CAPM 60_36	0.11	0.00	45.66	19.72	-0.28	-0.40	23.90	45.40
Monthly FF3 60_36	0.11	0.00	45.18	21.35	-0.27	-0.40	24.31	46.30

Table 5 summarizes the sensitivity of the H-L return to individual research design decisions. The most consequential choice is the data frequency: using daily data yields a much larger, more reliably positive spread, i.e., a mean of 1.27 with a 98% design, producing an H-L spread of greater than zero. In the monthly specification, a much weaker H-L spread is observed, i.e., a mean of 0.58, with 81% of designs producing an H-L spread greater than zero. Within the daily framework, shortening the window size tends to increase the return and significance of the outcome. Screening and portfolio construction choices materially shape the magnitude of the premium. Dropping small stocks earlier, before breaking points, and calculating the spread more aggressively reduces the H-L spread from 1.22 to 0.89, indicating that the spread is partially concentrated among small stocks. Weighing schemes generate especially large differences: equal weighing delivers the strongest performance, while value weighing produces a much smaller spread, and the capped value-weighted sits in between. Additional robustness checks show that winsorizing market capitalization, increasing the number of portfolios, and imposing stricter size/price/assets filters generally compress the spread and reduce statistical significance, suggesting that the IVOL H–L premium is strongest when the sample includes more marginal, smaller, and lower-priced firms. Finally, moving from univariate sorting to bivariate dependent sorting and adding controls tends to improve performance metrics and significance rates, while restrictions based on firm age have comparatively minor effects.

Table 5. Mean-return robustness across research-design choices.

The table summarizes the H–L portfolio mean and median returns, the fraction of cases with H–L > 0, the share of statistically significant results (at 5% level of significance), and the Sharpe ratio for alternative design settings. Each row reports the outcome when varying one design dimension while holding the remaining setup fixed.

Research Desings	Mean	Median	% H-L > 0	% of **	Sharp Ratio
Data Input: Daily	1.27	1.20	98.34	76.06	0.73
Data Input: Monthly	0.58	0.50	81.02	31.30	0.39
Window: 1M	1.31	1.20	98.87	83.62	0.77
Window: 6M	1.28	1.20	98.26	70.89	0.70
Window: 12M	1.17	1.10	97.37	66.10	0.67
Window: M60	0.58	0.50	81.02	31.30	0.39
Min obs: 15	1.31	1.20	98.88	83.66	0.77
Min obs: 17	1.30	1.20	98.86	83.58	0.77
Min obs: 63	1.28	1.20	98.26	70.89	0.70
Min obs: 126	1.17	1.10	97.37	66.10	0.67
Min obs: M24	0.63	0.50	82.59	35.79	0.43
Min obs: M36	0.53	0.40	79.40	26.65	0.35
Drop small AT: None	1.22	1.20	94.49	67.24	0.69
Drop small AT: 0.1	1.02	1.00	92.29	60.34	0.61
Drop small AT: 0.2	0.89	0.90	91.24	56.62	0.56
Weights: Equal-weighted	1.69	1.60	99.50	88.79	1.01
Weights: Value-weighted	0.79	0.80	88.49	47.52	0.45
Weights: VW-Cap	0.85	0.90	91.35	54.65	0.50
Mktcap winsorize: None	1.24	1.10	94.00	68.15	0.73
Mktcap winsorize: 80%	0.89	0.90	92.40	57.21	0.53
Mktcap winsorize: 90%	0.81	0.80	90.29	52.09	0.48
Portfolios: 3	0.80	0.80	92.46	62.18	0.64
Portfolios: 5	1.04	1.00	92.39	61.77	0.62
Portfolios: 10	1.34	1.30	93.27	59.98	0.58
Cap: 0	1.16	1.10	93.66	65.49	0.67
Cap: >$0.1m	1.12	1.00	93.46	64.11	0.66
Cap: >$0.5m	1.01	1.00	92.61	60.60	0.61
Cap: >$1m	0.88	0.80	90.91	55.23	0.53
Price: >$0	0.56	0.50	78.75	32.81	0.34
Price: >$0.10	1.05	1.00	94.16	58.83	0.61
Price: >$0.20	1.05	1.00	94.50	65.33	0.63
Price: >$0.50	1.29	1.20	97.98	77.20	0.77

Price: >$1	1.30	1.20	99.16	75.73	0.77
Assets: >1	1.15	1.10	94.81	66.93	0.65
Assets: >10	1.09	1.00	94.35	64.39	0.63
Assets: >20	0.99	0.90	92.40	59.08	0.61
Assets: >50	0.73	0.60	82.42	43.08	0.52
Sorting: Univariate	1.01	0.80	82.41	46.10	0.55
Sorting: Bivariate independent	0.99	0.90	94.03	58.99	0.55
Sorting: Bivariate dependent	1.12	1.10	97.54	73.09	0.73
Controls: None	1.01	0.80	82.41	46.10	0.55
Controls: 2	1.05	1.00	95.09	64.37	0.63
Controls: 5	1.05	1.00	96.65	68.11	0.65
Age: All	1.06	1.00	92.74	61.84	0.62
Age: >1 year	1.06	1.00	92.70	62.12	0.63
Age: >2 years	1.01	1.00	92.58	60.23	0.61

Table 6 examines alpha robustness as one research design dimension is varied at a time and reports results separately for the CAPM and the FF3 factor model. The result reveals that CAPM alphas are generally positive, especially in daily implementation. In the monthly implementation, the return weakens. Once the FF3 model is introduced, the picture changes substantially. The FF3 alpha is much smaller and often near zero or negative. For daily data, the mean FF3 alpha is positive, and roughly 50% of cases are significant, whereas for monthly data, the return is negative and only 28% of cases are significant. Across design dimensions, several choices materially shift the alpha distribution. In contrast to the returns in Table 5, the alpha increases with window size in daily data, and the % H-L greater than zero also increases in both the CAPM and FF3 alpha. Sample screening also matters, such as dropping small stocks. The construction of returns is especially influential, i.e., equal-weighted portfolios produce the strongest alphas, whereas value-weighted and capped value-weighted implementations yield much weaker CAPM and FF3 alphas. Similarly, market-cap winsorization tends to compress CAPM alpha and makes FF3 alpha more negative. Overall, this table suggests that abnormal performance is more robust under CAPM than under FF3 and that robustness is highly sensitive to portfolio construction and sample inclusion decisions, with the strongest alphas arising in settings that place more weight on small stocks.

Table 6. Mean-return robustness across research-design choices.

The table summarizes the H–L portfolio mean and median returns, the fraction of cases with H–L > 0, the share of statistically significant results (at 5% level of significance), and the Sharpe ratio for alternative design settings. Each row reports the outcome when varying one design dimension while holding the remaining setup fixed.

	Capm Alpha				FF3 Alpha			
	Mean	Median	% H-L > 0	% of **	Mean	Median	% H-L > 0	% of **
Data Input: Daily	0.65	0.60	89.18	43.65	0.15	0.10	55.70	25.43
Data Input: Monthly	0.16	0.10	50.97	21.21	-0.19	-0.30	28.69	41.98
Window: 1M	0.59	0.50	88.14	45.52	0.13	0.10	52.56	25.39
Window: 6M	0.68	0.60	89.60	39.63	0.10	0.10	54.06	25.46
Window: 12M	0.72	0.60	90.83	43.92	0.24	0.20	63.60	25.50
Window: M60	0.16	0.10	50.97	21.21	-0.19	-0.30	28.69	41.98
Min obs: 15	0.59	0.50	88.16	45.47	0.13	0.10	52.63	25.38
Min obs: 17	0.59	0.50	88.12	45.57	0.13	0.10	52.49	25.39
Min obs: 63	0.68	0.60	89.60	39.63	0.10	0.10	54.06	25.46
Min obs: 126	0.72	0.60	90.83	43.92	0.24	0.20	63.60	25.50
Min obs: M24	0.22	0.10	56.35	21.87	-0.11	-0.20	33.12	38.24
Min obs: M36	0.11	0.00	45.42	20.53	-0.27	-0.40	24.10	45.85
Drop small AT: None	0.65	0.60	83.78	44.40	0.19	0.20	57.07	33.31
Drop small AT: 0.1	0.47	0.40	76.28	36.07	0.02	0.00	47.78	29.00
Drop small AT: 0.2	0.34	0.30	69.96	28.44	-0.10	-0.10	35.71	30.25
Weights: Equal-weighted	1.07	1.00	94.03	72.88	0.66	0.60	79.56	57.02
Weights: Value-weighted	0.33	0.40	73.23	21.07	-0.12	-0.10	39.51	16.86
Weights: VW-Cap	0.27	0.30	69.71	25.63	-0.19	-0.10	34.17	24.77
Mktcap winsorize: None	0.70	0.60	83.63	46.98	0.27	0.20	59.54	36.94
Mktcap winsorize: 80%	0.30	0.30	72.06	27.62	-0.18	-0.10	35.55	25.34
Mktcap winsorize: 90%	0.24	0.30	67.36	23.64	-0.21	-0.20	32.79	24.19
Portfolios: 3	0.37	0.30	76.27	38.36	-0.01	0.00	42.58	34.46
Portfolios: 5	0.48	0.50	76.07	36.65	0.04	0.00	47.49	30.87
Portfolios: 10	0.65	0.60	77.89	33.30	0.10	0.10	51.48	26.28
Cap: 0	0.59	0.50	80.57	41.19	0.17	0.10	53.92	32.93
Cap: >$0.1m	0.57	0.50	79.57	39.68	0.13	0.10	51.94	32.05
Cap: >$0.5m	0.46	0.40	76.09	34.70	0.00	0.00	45.14	29.33
Cap: >$1m	0.33	0.30	70.28	29.45	-0.15	-0.10	36.11	29.04
Price: >$0	0.02	0.00	47.75	17.63	-0.57	-0.60	13.27	42.94
Price: >$0.10	0.50	0.40	78.97	31.06	0.06	0.00	43.56	25.43
Price: >$0.20	0.50	0.40	80.29	34.07	0.08	0.00	45.67	26.91
Price: >$0.50	0.72	0.60	89.66	51.59	0.31	0.20	64.78	30.26
Price: >$1	0.73	0.70	89.25	49.58	0.37	0.30	70.70	28.13
Assets: >1	0.56	0.50	81.12	40.42	0.08	0.00	49.68	29.97
Assets: >10	0.52	0.50	79.65	37.55	0.08	0.00	49.57	29.52
Assets: >20	0.46	0.40	75.56	33.78	0.03	0.00	46.78	30.68
Assets: >50	0.26	0.10	58.27	27.59	-0.16	-0.20	31.11	37.68
Sorting: Univariate	0.58	0.40	68.78	33.92	0.04	-0.20	40.30	39.40

Sorting: Bivariate independent	0.38	0.40	74.01	28.03	-0.08	-0.10	40.74	23.77
Sorting: Bivariate dependent	0.54	0.50	84.11	46.03	0.16	0.10	56.94	32.75
Controls: None	0.58	0.40	68.78	33.92	0.04	-0.20	40.30	39.40
Controls: 2	0.48	0.40	79.37	36.76	0.03	0.00	47.42	27.86
Controls: 5	0.44	0.40	78.69	37.36	0.05	0.10	50.60	28.75
Age: All	0.50	0.40	77.07	36.49	0.07	0.00	48.93	31.29
Age: >1 year	0.51	0.50	77.29	38.82	0.04	0.00	47.55	31.17
Age: >2 years	0.45	0.40	75.64	33.57	0.00	0.00	44.04	30.08

4.2 Standard Error VS Non-Standard Error

Once we have summarized the baseline results, we now turn to quantify the uncertainty in a way that separates the sampling noise within a given design, i.e., the standard error (within-design variation), from variation introduced by different research designs, i.e., the non-standard error (across-design variation). Our methodology is closely related to Menkveld et al. (2021) and Soebhag et al. (2022). Firstly, we compute the standard error as a within-design measure: for each design, we generate 1000 block bootstrap samples and take the standard deviation of these bootstrap samples. The reported SE is the average of these bootstrapped standard deviations across all designs. Whereas the non-standardized error (NSE) is intended to capture design uncertainty and is measured as the standard deviation of the statistic across the entire grid of research design choices. To gauge the relative importance of these to sources of uncertainties, we report the N/S ratio, i.e., NSE divided by SE. The larger the value of N/S, the more it indicates that differences across design choices contribute more to overall uncertainty than within-design variations.

Figure 4 compares the average uncertainty of the H-L estimates using SE and NSE. A clear, consistent pattern is that NSE is larger than SE across all IVOL estimation methods. The gap between NSE and SE is especially pronounced in the daily specification, where both the error measures are higher overall than in the monthly design, suggesting greater noise in daily implementation. Even within daily estimation, the longer-horizon estimation is prone to greater SE and NSE errors. The difference between CAPM and FF3 under the same frequency and window specifications is negligible, suggesting that model choice has less impact than the window and frequency specifications.

Figure 4. Standard errors by model and window specification.

The plot compares average standard errors (SE) and non-standard errors (NSE) for the value-weighted H–L estimate across the listed monthly and daily CAPM/FF3 rolling-window designs.

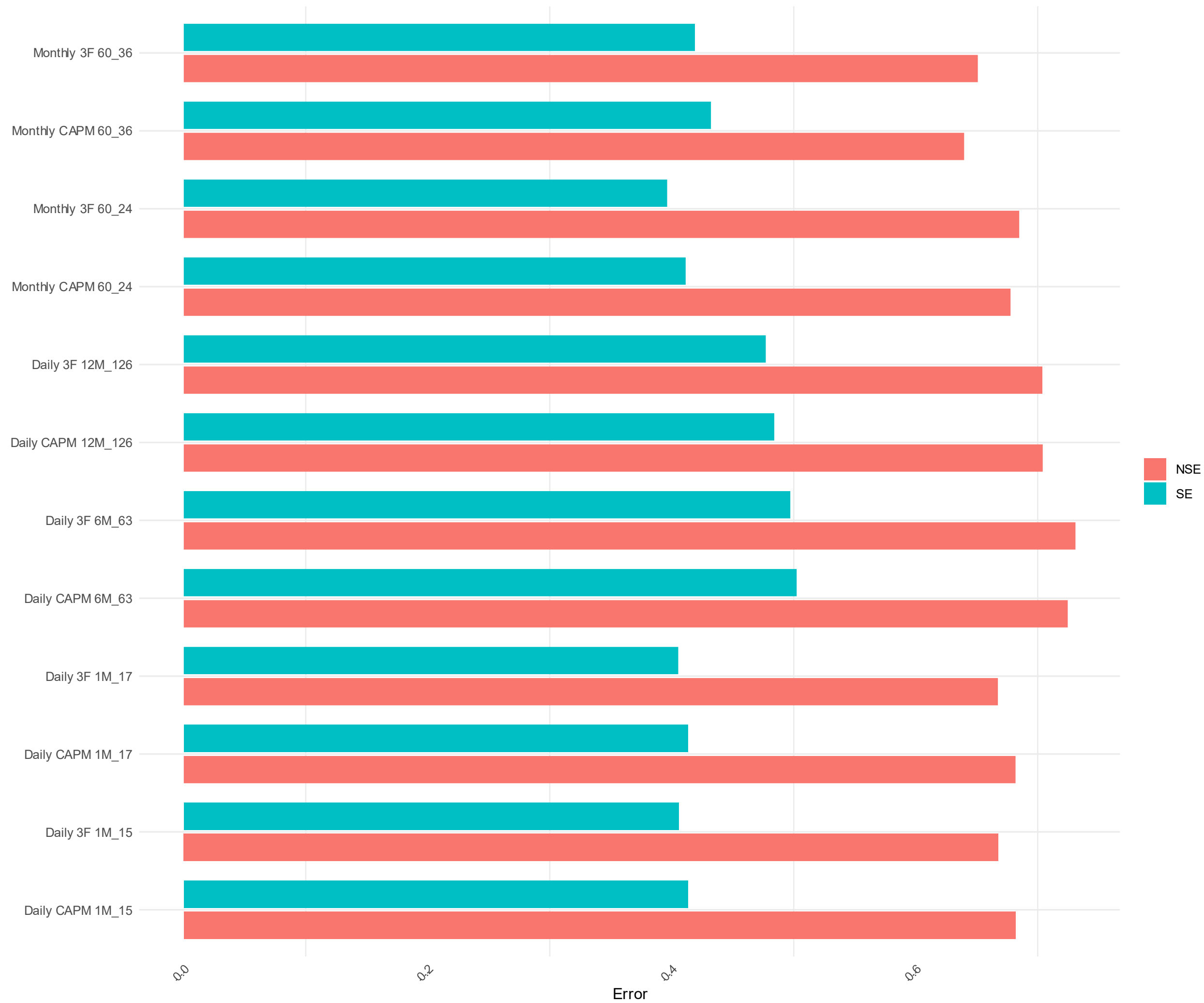


Table 5 also shows that NSE exceeds SE across all IVOL estimation specifications. For returns, SE is fairly stable and only slightly higher in longer daily estimation windows, and a similar pattern holds for the Sharpe ratio. However, if we look at the ratio of NSE vs SE, i.e., N/S, we observe that the N/S is lower for the longer daily specification, suggesting that once the SE has been catered for, the remaining dispersion (NSE) is not as large compared to within design uncertainty, i.e., results are more stable across research designs in a longer window. Overall, this table implies that relying on conventional SE can materially underestimate the uncertainty, and that inference becomes substantially more conservative once design variations are incorporated via NSE.

Table 5. Standard errors versus non-standard errors across models.
For each model/window specification, the table reports the conventional standard error (SE) and the non-standard error (NSE) for the H–L mean return and Sharpe ratio, along with the inflation factor NSE/SE.

Model	Return			Sharp Ratio		
	SE	NSE	NSE/SE	SE	NSE	NSE/SE
Daily CAPM 1M_15	0.41	0.68	1.65	0.21	0.33	1.63
Daily FF3 1M_15	0.41	0.67	1.65	0.21	0.34	1.61
Daily CAPM 1M_17	0.41	0.68	1.65	0.20	0.33	1.63
Daily FF3 1M_17	0.41	0.67	1.65	0.21	0.34	1.62
Daily CAPM 6M_63	0.50	0.72	1.44	0.22	0.37	1.68
Daily FF3 6M_63	0.50	0.73	1.47	0.22	0.37	1.69
Daily CAPM 12M_126	0.48	0.70	1.45	0.22	0.38	1.74
Daily FF3 12M_126	0.48	0.70	1.48	0.22	0.39	1.75
Monthly CAPM 60_24	0.41	0.68	1.65	0.25	0.43	1.69
Monthly FF3 60_24	0.40	0.68	1.73	0.25	0.44	1.75
Monthly CAPM 60_36	0.43	0.64	1.48	0.26	0.40	1.53
Monthly FF3 60_36	0.42	0.65	1.55	0.26	0.41	1.57

Figure 5 decomposes the uncertainty in H-L estimates across individual research designs dimension by plotting average SE and NSE for each choice. NSE exceeds SE for every design node, indicating that designing uncertainty adds more dispersion than the conventional within-design sampling variation. NSE is highest when no control variable is used and when single sorting is used, suggesting that the results are particularly sensitive to these choices. In contrast, nodes such as window and minimum observations show a smaller NSE–SE gap. This suggests that inferences based solely on standard errors can be overly optimistic and that the credibility of the H-L effect depends more on research design than on sampling variation.

Figure 5. Standard errors under alternative research-design choices

Horizontal bars report average standard errors (SE) and non-standard errors (NSE) for the H–L estimate across design dimensions.

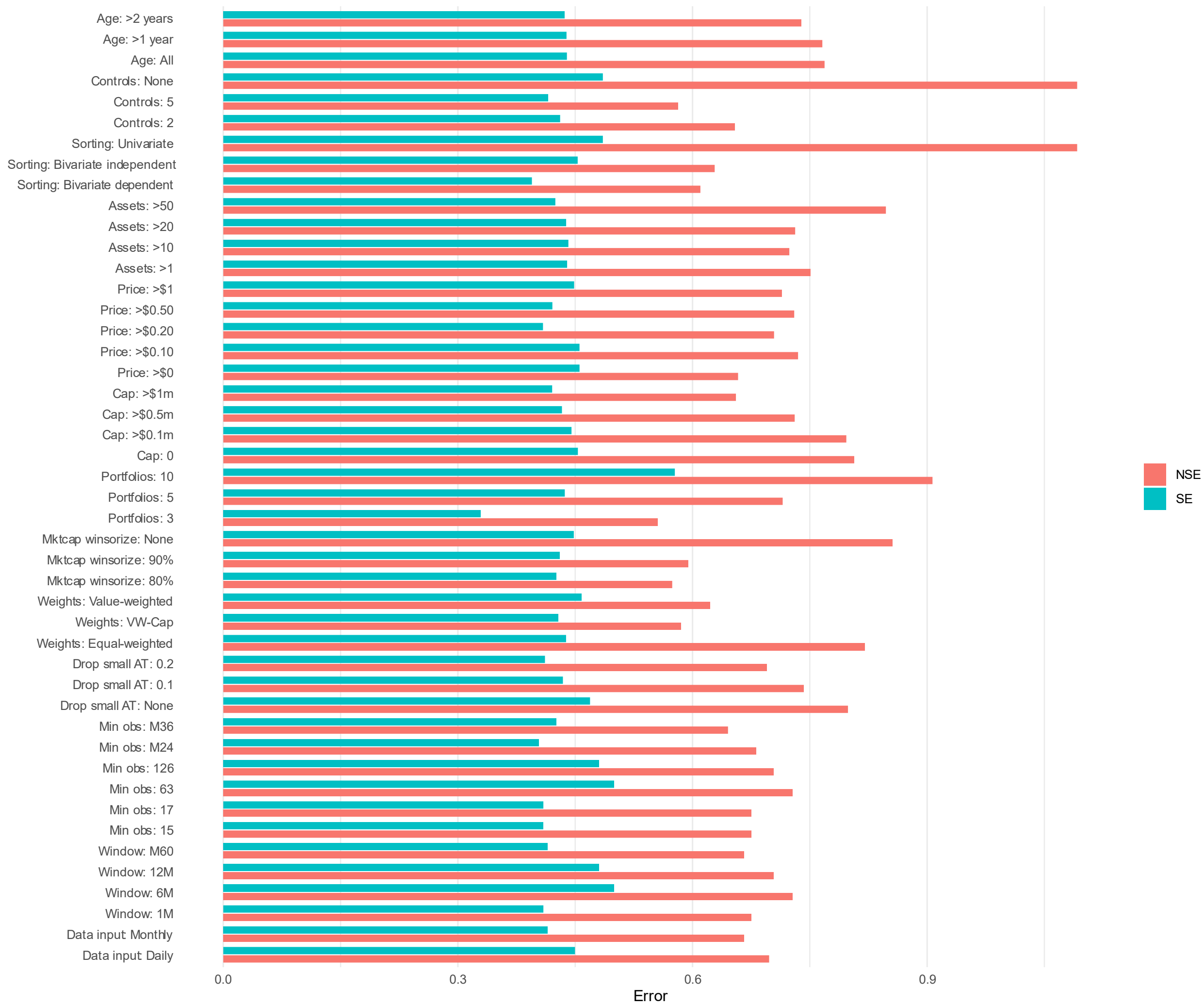


Table 6 supports the conclusion that specification uncertainty is substantial across essentially all design dimensions. NSE is consistently larger than SE, and the N/S ratio is always well above one, suggesting that NSE matters more than SE. Overall, the table implies that treating conventional SE as the sole uncertainty measure substantially understates the true fragility of the inference and that a small set of design decisions, especially those related to sorting and size-related treatments, are the primary drivers of NSE.

Table 6. Standard errors versus non-standard errors across research designs.

The table reports conventional SE, NSE, and the inflation ratio NSE/SE for the H–L mean return and the Sharpe ratio as key design dimensions vary.

Research Design	Return			Sharp Ratio		
	SE	NSE	NSE/SE	SE	NSE	NSE/SE
Data Input: Daily	0.45	0.70	1.55	0.21	0.36	1.69
Data Input: Monthly	0.41	0.67	1.61	0.26	0.42	1.64
Window: 1M	0.41	0.68	1.65	0.21	0.34	1.62
Window: 6M	0.50	0.73	1.46	0.22	0.37	1.69
Window: 12M	0.48	0.70	1.46	0.22	0.39	1.75
Window: M60	0.41	0.67	1.61	0.26	0.42	1.64
Min obs: 15	0.41	0.68	1.65	0.21	0.34	1.62
Min obs: 17	0.41	0.68	1.65	0.21	0.34	1.62
Min obs: 63	0.50	0.73	1.46	0.22	0.37	1.69
Min obs: 126	0.48	0.70	1.46	0.22	0.39	1.75
Min obs: M24	0.40	0.68	1.69	0.25	0.44	1.72
Min obs: M36	0.43	0.65	1.52	0.26	0.41	1.55
Drop small AT: None	0.47	0.80	1.70	0.23	0.42	1.84
Drop small AT: 0.1	0.43	0.74	1.71	0.23	0.41	1.81
Drop small AT: 0.2	0.41	0.69	1.69	0.23	0.40	1.75
Weights: Equal-weighted	0.44	0.82	1.87	0.22	0.39	1.81
Weights: Value-weighted	0.46	0.62	1.36	0.24	0.34	1.39
Weights: VW-Cap	0.43	0.59	1.37	0.23	0.32	1.42
Mktcap winsorize: None	0.45	0.86	1.91	0.23	0.46	2.02
Mktcap winsorize: 80%	0.43	0.57	1.35	0.23	0.31	1.38
Mktcap winsorize: 90%	0.43	0.59	1.38	0.23	0.33	1.44
Portfolios: 3	0.33	0.56	1.69	0.23	0.42	1.81
Portfolios: 5	0.44	0.72	1.64	0.23	0.42	1.81
Portfolios: 10	0.58	0.91	1.57	0.22	0.39	1.81
Cap: 0	0.45	0.81	1.78	0.23	0.43	1.89
Cap: >$0.1m	0.45	0.80	1.79	0.23	0.43	1.88
Cap: >$0.5m	0.43	0.73	1.69	0.23	0.40	1.77
Cap: >$1m	0.42	0.66	1.56	0.23	0.37	1.63
Price: >$0	0.46	0.66	1.45	0.25	0.38	1.53
Price: >$0.10	0.46	0.73	1.61	0.23	0.39	1.69
Price: >$0.20	0.41	0.70	1.72	0.22	0.40	1.80
Price: >$0.50	0.42	0.73	1.73	0.21	0.38	1.79
Price: >$1	0.45	0.71	1.59	0.22	0.36	1.59
Assets: >1	0.44	0.75	1.71	0.22	0.38	1.76
Assets: >10	0.44	0.72	1.64	0.22	0.39	1.73
Assets: >20	0.44	0.73	1.67	0.23	0.41	1.76
Assets: >50	0.42	0.85	2.00	0.26	0.54	2.11

Sorting: Univariate	0.49	1.09	2.25	0.23	0.59	2.53
Sorting: Bivariate independent	0.45	0.63	1.39	0.23	0.31	1.36
Sorting: Bivariate dependent	0.39	0.61	1.55	0.22	0.34	1.53
Controls: None	0.49	1.09	2.25	0.23	0.59	2.53
Controls: 2	0.43	0.65	1.52	0.23	0.35	1.56
Controls: 5	0.42	0.58	1.40	0.23	0.32	1.43
Age: All	0.44	0.77	1.75	0.23	0.41	1.83
Age: >1 year	0.44	0.77	1.75	0.23	0.42	1.83
Age: >2 years	0.44	0.74	1.69	0.23	0.41	1.78

5 Conclusions

In this paper, we contributed to the literature on the relationship between idiosyncratic volatility and future returns. Specifically, we analyzed the robustness of the IVOL anomaly in CEE stock markets by conducting a multiverse-style sensitivity analysis of plausible empirical model design choices. Our analysis yielded two main insights: First, we showed that the magnitude and statistical significance of the IVOL premium are highly sensitive to routine design decisions. In particular, working with daily instead of monthly data, using equal-weighted instead of value-weighted portfolios and using more granular portfolio sorts produced the largest high-minus-low IVOL return spreads. Second, we showed that conventional t-statistics tend to overstate the precision of model specifications, while using non-standard errors yields substantially more conservative estimates of model parameter uncertainty.

Our results are interesting for policymakers, supervisors, investors and other market participants alike. Overall, they suggest that implied volatility in CEE stock markets should be treated as a single risk factor yielding a clear trading signal. Rather, it seems to reflect a mixture of genuine risk differences and empirical model design choices. On the one hand, this underscores the need for further research into the distinct determinants of IVOL, for instance involving alternative risk-adjustment models (such as the Fama & French (2015) five-factor model), and respective comparisons with other developed and emerging markets. On the other hand, our study highlights the necessity of transparent reporting of modeling choices and of conducting multiverse-style robustness tests that account for alternative plausible model designs.

Reference list

Ali, S. R. M., Hasan, M. N., & Östermark, R. (2020). Are idiosyncratic risk and extreme positive return priced in the Indian equity market?. *International Review of Economics & Finance*, *70* (C), 530-545. DOI: 10.1016/j.iref.2020.08.008

Ali, S. R. M., Rahman, M. A., Hasan, M. N., & Östermark, R. (2020a). Positive IVOL-MAX effect: A study on the Singapore Stock Market. *The North American Journal of Economics and Finance*, *54* (C), 101245. DOI: 10.1016/j.najef.2020.101245

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, *61*(1), 259-299. DOI: 10.1111/j.1540-6261.2006.00836.x

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further US evidence. *Journal of Financial Economics*, *91*(1), 1-23. DOI: 10.1016/j.jfineco.2007.12.005

Angelidis, T., & Tessaromatis, N. (2008). Does idiosyncratic risk matter? Evidence from European stock markets. *Applied Financial Economics*, *18*(2), 125-137. DOI: 10.1080/09603100601118276

Bali G., T., & Weigert, F. (2024). Hedge funds and the positive idiosyncratic volatility effect. *Review of Finance*, *28*(5), 1611-1661. DOI:10.1093/rof/rfae022

Bali, T. G., & Cakici, N. (2008). Idiosyncratic volatility and the cross section of expected returns. *Journal of Financial and Quantitative Analysis*, *43*(1), 29-58. DOI: 10.1093/rof/rfae022

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Ffinancial Eeconomics*, *99*(2), 427-446. DOI: 10.1016/j.jfineco.2010.08.014

Bali, T. G., Cakici, N., Yan, X., & Zhang, Z. (2005). Does idiosyncratic risk really matter?. *The Journal of Finance*, *60*(2), 905-929. DOI: 10.1111/j.1540-6261.2005.00750.x

Barberis, N., & Huang, M. (2008). Stocks as lotteries: The implications of probability weighting for security prices. *American Economic Review*, *98*(5), 2066-2100. DOI: 10.1257/aer.98.5.2066

Barinov, A., & Chabakauri, G. (2023). Idiosyncratic volatility, growth options, and the cross-section of returns. *The Review of Asset Pricing Studies*, *13*(4), 653-690.

Bhootra, A., & Hur, J. (2015). High idiosyncratic volatility and low returns: A prospect theory explanation. *Financial Management*, *44*(2), 295-322. DOI: 10.1111/fima.12057

Black F. (1972), Capital market equilibrium with restricted borrowing, *Journal of Business* , 45(3), 444-455. DOI: 10.1086/295472

Brockman P., Tao Guo T., Maria Gabriela Vivero M.G., Wayne Yu W. (2022),. Is idiosyncratic risk priced? The international evidence, *Journal of Empirical Finance*, Volume 66, 2022, 121-136, DOI:10.1016/j.jempfin.2022.01.004.

Brown, D.avid P., and Ferreira M.A., Miguel Almeida (2025),. Idiosyncratic Volatility of Small Public Firms and Entrepreneurial Risk (Decemberr 2005). Available at SSRN: https://ssrn.com/abstract=892145 or http://dx.doi.org/10.2139/ssrn.892145

Cao J., & Tarun Chordia T., & Xintong Zhan X. (2021), 2021. "The Calendar Effects of the Idiosyncratic Volatility Puzzle: A Tale of Two Days?," Management Science, INFORMS, vol. 67(12), pages 7866-7887, December. DOI: 10.1287/mnsc.2020.3803

Cao, J., Chordia, T., & Zhan, X. (2021). The calendar effects of the idiosyncratic volatility puzzle: A tale of two days?. *Management Science*, *67*(12), 7866-7887. DOI: 10.1287/mnsc.2020.3803

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of fFinance*, *52*(1), 57-82. DOI:10.1111/j.1540-6261.1997.tb03808.x

Chen, L. H., Jiang, G. J., Xu, D. D., & Yao, T. (2020). Dissecting the idiosyncratic volatility anomaly. *Journal of Empirical Finance*, *59(C)*, 193-209. DOI: 10.1016/j.jempfin.2020.10.004

Chichernea, D. C., Kassa, H., & Slezak, S. L. (2019). Lottery preferences and the idiosyncratic volatility puzzle. *European Financial Management*, *25*(3), 655-683. DOI: 10.1111/eufm.12178

Coqueret G., Pérignon Ch. (2025). Persistent Anomalies and Nonstandard Errors, HEC Research Papers Series 1578, HEC Paris. DOI: 10.2139/ssrn.5276723

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56. DOI:10.1016/0304-405X(93)90023-5

Fu, Fangjian, 2009. "Idiosyncratic risk and the cross-section of expected stock returns," *Journal of Financial Economics*, Elsevier, vol. 91(1), 24-37., January. DOI:10.1016/j.jfineco.2008.02.003

Gempesaw D., Haimanot Kassa, Blerina Bela Zykaj, Does idiosyncratic volatility proxy for a missing risk factor? Evidence using portfolios as test assets, *European Financial Management*, 28(3) pp., 693-721, https://doi.org/10.1111/eufm.12317

Goh, J., & Kim, D. (2024). The role of arbitrage risk in the MAX effect: evidence from the Korean stock market. *Journal of Derivatives and Quantitative Studies*, *32*(2), 159-180. DOI: 10.1108/JDQS-09-2023-0031

Goyal, A. and Santa-Clara, P. (2003) Idiosyncratic risk matters!, *Journal of Finance*, 58, 975–1008. DOI: 10.1111/1540-6261.00555

Herskovic, B., Kelly, B., Lustig, H., & Van Nieuwerburgh, S. (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics*, 119(2), 249-283. DOI: 10.1016/j.jfineco.2015.09.010

Hou, K., & Loh, R. K. (2016). Have we solved the idiosyncratic volatility puzzle?. *Journal of Financial Economics*, 121(1), 167-194. DOI: 10.1016/j.jfineco.2016.02.013

Lintner, J. (1965), The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics*, 47(1), 13-37. DOI:10.2307/1924119

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. DOI:10.2307/2975974

McMillan, D., Kambouroudis, D., & Khasawneh, M. (2022). Lottery Stocks in the UK: Evidence, Characteristics and Cause. *International Journal of Banking, Accounting and Finance*, 14(1), 58-96.

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., ... & Khomyn, M. K. (2024). Nonstandard errors. *The Journal of Finance*, *79*(3), 2339-2390. DOI: 10.1111/jofi.13337

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, *42*(3), 483-510. DOI:10.1111/j.1540-6261.1987.tb04565.x

Park H., Dong-Hyun A. (2025), The Anatomy of Idiosyncratic Volatility. Joint Conference with the Allied Korea Finance Association, May 2025, Available at SSRN: https://ssrn.com/abstract=6533361 or http://dx.doi.org/10.2139/ssrn.6533361

Poon, S.-H., Yao, J., & Zhang, Y. (2022). The alphas of beta and idiosyncratic volatility. *Journal of Financial Markets*, 61(C). DOI: 10.1016/j.finmar.2022.100720

Qadan M. & Shuval K. (2022), Variance risk and the idiosyncratic volatility puzzle, *Finance Research Letters*, 45(C),102176. DOI:10.1016/j.frl.2021.102176.

Sharpe, W. F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19(3), 425-442. DOI: 10.1111/j.1540-6261.1964.tb02865.x

Soebhag, A., Van Vliet, B., & Verwijmeren, P. (2024). Non-standard errors in asset pricing: Mind your sorts. *Journal of Empirical Finance*, 78(C), 101517. DOI: 10.1016/j.jempfin.2024.101517

Stambaugh R.F. & Jianfeng Yu & Yu Yuan, 2015. Arbitrage Asymmetry and the Idiosyncratic Volatility Puzzle, *Journal of Finance*, American Finance Association, vol. 70(5), pages 1903-1948, October. DOI: 10.1111/jofi.12286

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297-323. DOI: 10.1007/BF00122574

Wang, Jianqiu & Wu, Ke & Pan, Jiening & Jiang, Ying (2023). Disagreement, speculation, and the idiosyncratic volatility, *Journal of Empirical Finance*, Elsevier, vol. 72(C), 232-250. DOI: 10.1016/j.jempfin.2023.03.011

Wei, S. X., & Zhang, C. (2005). Idiosyncratic risk does not matter: A re-examination of the relationship between average returns and average volatilities. *Journal of Banking & Finance*, 29(3), 603-621. DOI:10.1016/j.jbankfin.2004.05.021

Xu, Yexiao and Malkiel, Burton G., Idiosyncratic Risk and Security Returns (May 2004). Available at SSRN: https://ssrn.com/abstract=255303 or DOI: 10.2139/ssrn.255303

Zaremba A., 2015. Low Risk Anomaly In in The the Cee CEE Stock Markets, *Journal for Economic Forecasting*, Institute for Economic Forecasting, 0(3), 81-102, September.