Idiosyncratic Skew Arbitrage in Equity Options

MSE 244 Final Project

Björn Thor Stefánsson Kaden Wu Sondre Rogde

March 2026

Contents

1 Introduction **3**

2 Model **3**
- 2.1 Stage 1: Skew Extraction . . . 3
  - 2.1.1 Direct Skew . . . 4
  - 2.1.2 Polynomial Skew . . . 4
  - 2.1.3 Endpoint Skew (naive) . . . 4
- 2.2 Stage 2: Pairs Trading . . . 5
- 2.3 Stage 3: Signal Construction . . . 5
- 2.4 Stage 4: Delta-Hedged Risk-Reversal Strategy . . . 6
  - 2.4.1 Leg Selection . . . 6
  - 2.4.2 P&L Decomposition . . . 7
  - 2.4.3 Portfolio Construction and Position Limits . . . 7
- 2.5 Python Implementation . . . 7

3 Methodology **8**
- 3.1 Data Sources . . . 8
- 3.2 Implied Volatility Imputation . . . 8
- 3.3 Strategy Testing Scheme . . . 9
- 3.4 Transaction Costs . . . 9

4 Results **10**
- 4.1 Pairs Trading Strategy Performance . . . 10
- 4.2 Sensitivity Analysis . . . 12

4.2.1 Delta Target (Δ_{target}) 12
4.2.2 Entry Threshold (z_{entry}) 13
4.2.3 Skew Metric 13
4.3 Regimes 14

5 Conclusion **15**

A Appendix **17**

B Appendix **18**

C Appendix **19**

D Appendix **20**

E Appendix **22**

F Appendix **23**

G Appendix **25**

H Appendix **26**
H.1 Time to Expiration Target 26
H.2 Transaction Costs 26

`https://github.com/BjornThor123/Statistical_Arbitrage_Stanford_MS-E244`

1 Introduction

This project investigates a statistical arbitrage strategy that exploits mean-reversion in the *idiosyncratic* component of equity options skew. Volatility skew is the empirical observation that out-of-the-money (OTM) puts trade at higher implied volatility than OTM calls and is a well documented feature of equity option markets. A common explanation is that investors overpay for downside protection in single stocks due to temporary liquidity demands or behavioral biases, rather than genuine shifts in the underlying risk distribution.

Previous studies utilized a range of different methods to extract skews from the implied volatility smile or market prices. Notably, in Bedendo and Hodges' work [1], a cubic model is fitted in normalized moneyness, and the linear term's coefficient was extracted as an indicator for the smile's slope and hence volatility skew. On the other hand, in Feunou, Fontaine, and Tédongap's paper [2], the skewness metric was inferred through the Homoscedastic Gamma model by matching model implied option prices to observed market prices across strikes. These papers shed light on the variety of methods that can be implemented to extract skews: by functional fit and inverting well known option pricing models. We proceed to investigate three methods of extracting the skew.

Our strategy targets the U.S. financial sector, trading liquid single-stock equity options alongside the XLF Financial Sector ETF over the period January 2015 to December 2020. Signals are constructed daily and positions are entered with a one-day lag to prevent look-ahead bias. The gross strategy achieves a Sharpe ratio of 1.05; however, execution costs reduce the net Sharpe to 0.24, illustrating the central challenge of options-based statistical arbitrage.

2 Model

While a stock's skew tends to comove with its sector (reflecting systematic fear), deviations from this relationship can signal a stock's transient mispricings. By stripping out the sector driven skew component via regression and focusing on the unexplained residual, i.e. the *idiosyncratic* skew, we identify stocks whose skew is temporarily overpriced or underpriced relative to its historical norm. We enter positions when this idiosyncratic component deviates sufficiently from zero and hold until it mean-reverts.

2.1 Stage 1: Skew Extraction

For each stock i and date t we estimate the implied volatility skew at a target maturity of $\tau^* = 15$ calendar days. Only options with at most 30 days to expiry are retained; if the exact 15-day maturity is unavailable, the skew is linearly interpolated between the nearest available maturities:

$$\widehat{\text{Skew}}_{i,t}(\tau^*) = \widehat{\text{Skew}}_{i,t}(\tau_l) + \frac{\tau^* - \tau_l}{\tau_u - \tau_l}\left[\widehat{\text{Skew}}_{i,t}(\tau_u) - \widehat{\text{Skew}}_{i,t}(\tau_l)\right]. \tag{1}$$

Where τ_l, τ_u is the closest time to maturity less than 15 and more than 15 respectively. We consider three complementary skew measures.

2.1.1 Direct Skew

The simplest and most liquid measure uses the implied volatility differential between the 25-delta OTM put and the 25-delta OTM call at the target maturity:

$$\text{Skew}^{\text{dir}}_{i,t} = \sigma^{\text{put}}_{i,t}(\Delta = -0.25) - \sigma^{\text{call}}_{i,t}(\Delta = +0.25). \tag{2}$$

A positive value indicates OTM puts are expensive relative to OTM calls, the standard signature of a negatively skewed risk-neutral distribution. Because this measure is computed at the same strikes as the traded legs of the risk reversal, it is directly consistent with the instrument used at stage 4. Note also that if the 25Δ option is not available we choose the closest one.

2.1.2 Polynomial Skew

To exploit the full cross-section of strikes, we fit a quadratic polynomial to all OTM implied volatilities (calls for $k \geq 0$, puts for $k < 0$) as a function of log-moneyness $k = \ln(K/F_t)$:

$$\hat{\sigma}(k) = \alpha + \beta k + \gamma k^2. \tag{3}$$

We define the polynomial skew as the negated linear coefficient,

$$\text{Skew}^{\text{poly}}_{i,t} = -\hat{\beta}_{i,t}, \tag{4}$$

so that a larger value again indicates that puts are expensive relative to calls. This is closely related to [1]. Negation is required because a steeper downward slope in $\sigma(k)$ (puts pricier on the left) yields a negative $\hat{\beta}$ by convention.

2.1.3 Endpoint Skew (naive)

Another alternative that avoids reliance on interpolation or regression, is the endpoint skew which compares the implied volatility at the most out-of-the-money put available and the most out-of-the-money call available on a given day:

$$\text{Skew}^{\text{end}}_{i,t} = \sigma_{i,t}(K_{\min}) - \sigma_{i,t}(K_{\max}), \tag{5}$$

where $K_{\min}$ is the lowest available strike (deepest OTM put) and $K_{\max}$ is the highest available strike (deepest OTM call). This measure is conceptually similar to the direct skew but requires no delta interpolation; it captures the full tilt of the observed smile from endpoint to endpoint. However, the downside is that the most out of the money options tend to be noisy.

2.2 Stage 2: Pairs Trading

The idiosyncratic skew signal is constructed from a complete pairs trading framework. Let $\mathcal{U} = \{\text{BAC, C, BRK, JPM, SCHW, GS, XLF}\}$ be the asset universe, where XLF (the Financial Sector ETF) is included as a benchmark alongside the individual stocks. For each ordered pair (i, j) with $i \neq j$, we construct a cointegrating spread candidate using rolling OLS:

$$\varepsilon_{ij,t} = \text{Skew}_{i,t} - \hat{\alpha}_{ij,t} - \hat{\beta}_{ij,t}\,\text{Skew}_{j,t}, \tag{6}$$

where the parameters are re-estimated each day on the preceding 60 trading days:

$$\hat{\beta}_{ij,t} = \frac{\widehat{\text{Cov}}_{60}(\text{Skew}_i,\ \text{Skew}_j)}{\widehat{\text{Var}}_{60}(\text{Skew}_j)}, \qquad \hat{\alpha}_{ij,t} = \bar{S}_{i,t} - \hat{\beta}_{ij,t}\,\bar{S}_{j,t}.$$

For $|\mathcal{U}| = 7$ assets this generates $7 \times 6 = 42$ directed pairs. Including XLF naturally subsumes the "sector-based" approach as a special case: pairs of the form $(\text{stock}, \text{XLF})$ capture deviations of each stock's skew from the sector ETF, while stock-vs-stock pairs capture cross-sectional mispricings. When $\varepsilon_{ij,t}$ is abnormally large, stock i's skew is expensive relative to its historical relationship with j, and we sell the skew of i while buying the skew of j.

Before committing to any pair, we verify that the skew series are cointegrated using both the Engle-Granger and Johansen tests; cointegration is a necessary condition for the spread to be stationary and thus mean-reverting. Full results are reported in Appendix D. All 21 unique stock pairs and all stock-vs-XLF combinations are cointegrated at the 5% significance level, providing the econometric foundation for the strategy. Rolling estimation ensures all parameters use only information available at the time of trading.

2.3 Stage 3: Signal Construction

The residual $\varepsilon_{ij,t}$ from Section 2.2 is standardized into a rolling z-score:

$$z_{i,t} = \frac{\varepsilon_{i,t} - \mu_{\varepsilon_i,t}}{\sigma_{\varepsilon_i,t}}, \tag{7}$$

where $\mu_{\varepsilon_i,t}$ and $\sigma_{\varepsilon_i,t}$ are the rolling 60-day mean and standard deviation of $\varepsilon_{i,t}$. A discrete trading signal is then generated by a hard threshold:

$$s_{i,t} = \begin{cases} +1 & z_{i,t} > z_{\text{entry}}, \\ -1 & z_{i,t} < -z_{\text{entry}}, \\ 0 & \text{otherwise}, \end{cases} \tag{8}$$

The threshold z_{entry} is set to the 97.5th percentile of the expanding empirical distribution of $|z_{ij,t}|$, applied symmetrically to both tails. Using an empirical percentile rather than a fixed numerical value means we do not need to assume that the idiosyncratic residuals are normally distributed, and it provides direct control over trade frequency regardless of how the spread volatility evolves over the sample. The expanding distribution is computed without look-ahead, using only observations up to and including day t. When $s_{ij,t} = +1$, stock i's skew is abnormally high relative to j and we expect mean-reversion; hence, we sell the skew of i and buy the skew of j. Conversely, when $s_{ij,t} = -1$, the spread is abnormally negative and we take the opposite position. We also experiment with more and less selective thresholds (95th and 99th percentiles) in Section 4.

2.4 Stage 4: Delta-Hedged Risk-Reversal Strategy

Both strategies execute via the same instrument: a delta-hedged *risk-reversal* on the relevant stock. For each (ticker, date) with a non-zero lagged signal, we construct a long OTM call combined with a short OTM put (or vice versa for $s = -1$), targeting approximately $\Delta = 0.25$ and a 15-day maturity.

2.4.1 Leg Selection

For each (ticker, date) the legs are selected as follows:

1. **OTM call.** Restrict to calls with $k \geq 0$ where k is log moneyness. Among these, find the maturity bucket closest to 15 days, then select the option whose delta is closest to $+0.25$.

2. **OTM put.** Restrict to puts with $k < 0$. Apply the same maturity filter, then select the put whose absolute delta is closest to 0.25.

The risk-reversal value is:

$$\text{RR}_{i,t} = p^{\text{call}}_{i,t} - p^{\text{put}}_{i,t}, \tag{9}$$

where p denotes mid price. The net delta of a long risk reversal is $\Delta^{\text{net}}_{i,t} = \delta^{\text{call}}_{i,t} + |\delta^{\text{put}}_{i,t}| > 0$.

2.4.2 P&L Decomposition

The daily P&L decomposes into an option leg component and a delta-hedge component.

Option leg return (normalized by spot price):

$$r_{i,t}^{\text{opt}} = s_{i,t-1} \times \frac{\text{RR}_{i,t} - \text{RR}_{i,t-1}}{S_{i,t-1}}. \tag{10}$$

Delta-hedge return (from a short position in the underlying that neutralizes the risk reversal's net delta at the start of each day):

$$r_{i,t}^{\text{hedge}} = -\, s_{i,t-1} \times \Delta_{i,t-1}^{\text{net}} \times \frac{S_{i,t} - S_{i,t-1}}{S_{i,t-1}}. \tag{11}$$

The total daily position return is $r_{i,t}^{\text{pos}} = r_{i,t}^{\text{opt}} + r_{i,t}^{\text{hedge}}$.

2.4.3 Portfolio Construction and Position Limits

Capital is allocated equally across all active positions on each day. To prevent over-concentration in any single name, a per-stock weight cap of $w_{\max} = 20\%$ is enforced:

$$w_{i,t} = \min\left(\frac{1}{N_t^{\text{active}}},\ w_{\max}\right), \tag{12}$$

where $N_t^{\text{active}} = \sum_i \mathbf{1}[s_{i,t-1} \neq 0]$ is the number of active names. The gross portfolio return is:

$$R_t^{\text{gross}} = \sum_i w_{i,t}\, r_{i,t}^{\text{pos}}. \tag{13}$$

2.5 Python Implementation

The strategy is implemented in Python. Raw data is ingested and stored in a local DuckDB database via a custom `DataLoader` class that handles zip extraction, schema inference, and type casting. Skew extraction and imputation are implemented in vectorized `pandas`/`NumPy` operations. The pairs trading engine iterates over all directed pairs each day, updating rolling OLS coefficients using `numpy.linalg.lstsq` on a sliding window. Signal generation, option leg selection, P&L computation, and portfolio aggregation are each encapsulated in dedicated modules (`pairs_trading_skew.py`, `sector_pairs_trading_skew.py`, and `run.py`).

3 Methodology

3.1 Data Sources

The strategy requires three data inputs:

1. **Equity options data.** Daily option chains (calls and puts) across a range of strikes and maturities for each stock and the sector ETF. Fields used include: date, expiration date, option type, strike price, best bid, best offer, implied volatility, delta, gamma, and contract size.

2. **Underlying equity prices.** Daily closing prices from CRSP for each stock in the universe. These are used to compute returns for the delta-hedge positions.

3. **Risk-free interest rates.** Zero-coupon yield curves matched to option maturities, used to compute forward prices and to invert the Black-Scholes formula where implied volatility is missing.

The sample period runs from January 2015 to December 2020. We chose this period for several reasons. First, it does not contain too many missing values. Second, it is a short enough period so the data is small enough to handle without extra compute. Third, it contains a diverse set of market regimes. The sector proxy is the XLF Financial Sector ETF, and the stock universe consists of liquid financial sector equities for which option chains are available across multiple maturities throughout the sample. Only options with at most 30 calendar days to expiry are retained, consistent with a near-term skew signal targeting a 15-day tenor. To properly handle the data we have included a detailed description of data ingestion and structuring in Appendix A

3.2 Implied Volatility Imputation

The dataset contains a meaningful fraction of options where implied volatility is missing. A short overview of the missing values in the dataset can be found in Appendix B. Rather than discarding these observations, we impute missing values by numerically inverting the Black-Scholes formula.

For a European call the Black-Scholes price is:

$$C^{\mathrm{BS}}(S, K, r, \tau, \sigma) = S\,\Phi(d_1) - Ke^{-r\tau}\,\Phi(d_2), \tag{14}$$

where

$$d_1 = \frac{\ln(S/K) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \qquad d_2 = d_1 - \sigma\sqrt{\tau},$$

and Φ is the standard normal CDF. For each option with a missing implied volatility but a valid mid price, spot price, strike, and time to expiry, we solve

$$C^{\mathrm{BS}}(S_t, K, r, \tau, \hat{\sigma}) = p_{\mathrm{mid}}$$

for $\hat{\sigma}$ using Brent's root-finding algorithm on $[10^{-4}, 20]$. Options whose mid price violates the no-arbitrage intrinsic-value lower bound are left as `NaN` and excluded downstream. Only OTM options are used in the skew regression, consistent with market convention for smile fitting. After imputation we also notice that options with short time to maturity sometimes exhibit very high implied volatilities (upwards of 800%). This is because options near expiry are sensitive to gamma and hedging from market participants. This introduces noise into our model so we therefore remove all options with time to maturity less than 4 days. After doing this, some values of implied volatilities are still extremely high (upwards of 800%) we therefore clip values at the 90th percentile. We do this on a daily basis to avoid lookahead bias. Another approach would be to clip it at an absolute predetermined level.

3.3 Strategy Testing Scheme

The backtest follows a strict out-of-sample discipline. The first 60 trading days of the sample serve as an estimation warm-up period during which all rolling OLS coefficients and z-score statistics are initialized; no positions are taken during this period. After warm-up, the model operates in a fully online fashion: all rolling statistics (OLS coefficients, z-score mean and standard deviation, expanding empirical percentile) are updated daily using only observations available at the time of signal generation.

Signals formed at the close of day t are executed at day $t + 1$ to prevent look-ahead bias; this one-day signal lag is applied consistently throughout. Hence the strategy does not hedge continuously and that is why we see PnL from hedging diluting our results. The instrument universe is fixed at the six financial-sector stocks plus XLF for the entire sample period. No survivorship bias correction is applied, as all stocks remained listed throughout the sample. Assumptions for missing data are described in Section 3: options with missing IV after imputation are excluded, and any stock-day with insufficient option data to compute the skew measure is simply assigned no signal on that day.

3.4 Transaction Costs

Two components of transaction cost are applied whenever a position opens, closes, or reverses.

In principle the option leg cost should be the full bid-ask spread. However, as Appendix E shows, the bid-ask spread for near-dated OTM options in our universe averages between

100% and 200% of the option mid-price for most stocks. Far-OTM options are particularly illiquid and attract the widest spreads due to low volume and high adverse selection. Charging the full spread would trivially render the strategy unprofitable regardless of signal quality. Instead, we model both option legs and the equity delta-hedge as a flat b basis points of spot price per trade, where $b = 20$ at the baseline. This represents a stylized but more tractable execution cost assumption, applicable when trading with institutional counterparties at tighter-than-quoted rates. We conduct a sensitivity analysis on b in Section 4.

Option leg cost. For each risk-reversal leg (call and put) traded:

$$c_{i,t}^{\text{opt}} = \frac{b}{10{,}000}(p_{i,t}^{call} + p_{i,t}^{put}), \tag{15}$$

Delta-hedge slippage. Trading the underlying stock to neutralize net delta:

$$c_{i,t}^{\text{hedge}} = |\Delta_{i,t-1}^{\text{net}}| \times \frac{b}{10{,}000}. \tag{16}$$

Net returns are obtained by subtracting the weighted sum of both cost components from the gross return:

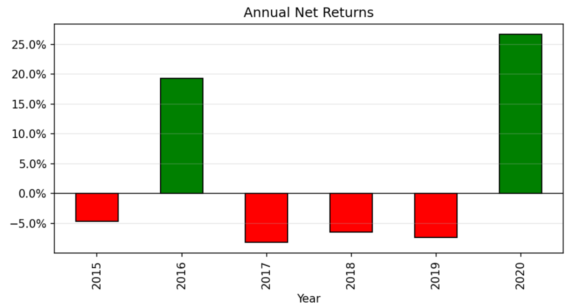
$$R_t^{\text{net}} = R_t^{\text{gross}} - \sum_i w_{i,t}\, \mathbf{1}[\text{trade}_{i,t}]\big(c_{i,t}^{\text{opt}} + c_{i,t}^{\text{hedge}}\big). \tag{17}$$

4 Results

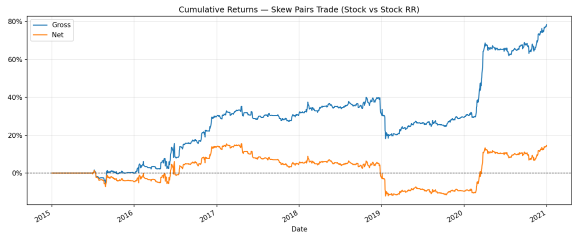
4.1 Pairs Trading Strategy Performance

Figures 1a and 1b give an overview of the return profile. The strategy generates positive returns in total (Figure 1a); losses, where they occur, are modest and largely attributable to transaction costs rather than adverse signal realizations. The cumulative return path (Figure 1b) is broadly stable, with two notable episodes. A sharp decline occurs in late 2018 and early 2019, coinciding with the broad S&P 500 sell-off of December 2018 and the subsequent recovery in January 2019: a rapid market-wide move followed by a reversal is the type of regime most likely to confound a mean-reversion strategy whose rolling OLS parameters are calibrated on a prior 60 days of more normal conditions. A sharp increase occurs in early 2020, driven primarily by the COVID-19 volatility spike. Importantly, this gain is attributable largely to the delta-hedge component of the position (see Figure 11 in Appendix G) rather than skew mean-reversion. This is undesirable as we would prefer P&L to come from the signal, not from directional equity exposure. Normally, a hedge is not directionally exposed, but we do daily hedging and not continuous hedging so there is some directional exposure. For a full month-by-month breakdown see Appendix F.

The annual Sharpe ratio (Figure 2a) fluctuates substantially from year to year. This is not unusual for statistical arbitrage strategies whose signal strength varies with market regimes, but the variability is larger than we would ideally want and reflects the sensitivity of the strategy to the COVID distortion period.

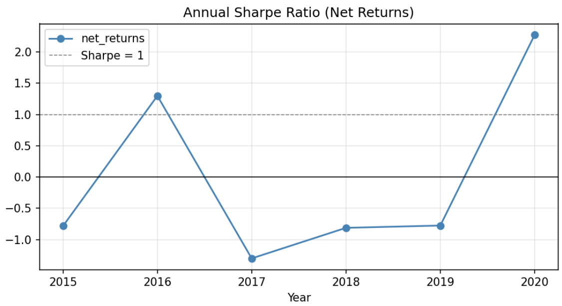


(a) Annual returns over the trading period.

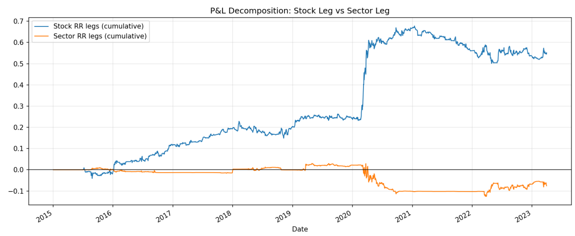


(b) Cumulative returns over the trading period.

Figure 1: Trading period returns.



(a) Annual Sharpe over the trading period.



(b) Breakdown risk reversal leg return over the trading period.

Figure 2: Trading period Sharpe and PnL decomposition.

Table 1 quantifies the performance (see Appendix C for a definition of the metrics). The gross Sharpe ratio of 1.05 is attractive, but costs reduce it to 0.24 on a net basis. The strategy retains a positive signal but barely covers its execution costs. The gross and net annualized volatility are nearly identical (9.6% vs 9.5%), confirming that transaction costs affect the mean return rather than the return variance, as expected. The win rate of 44.4% gross (41.2% net) is below 50% and should not be interpreted as the strategy being wrong more often than right: with a fat-tailed return distribution (excess kurtosis ≈ 16.6), the strategy profits through a small number of large-magnitude reversion events rather than frequent small gains. The skewness turns from slightly positive (0.03 gross) to slightly negative (-0.16 net), indicating that costs trim the small wins while the large losses remain. With an average of 1.3 risk-reversal legs traded per day, the annualized transaction cost of 7.4% consumes about 73% of the gross return, a stark illustration of how sensitive options-based strategies are to execution costs. Figure 2b shows that the majority of gross P&L derives from the stock risk-reversal legs (rather than XLF legs), confirming that individual stock skews tend to revert toward the sector, not the other way around.

Metric	**Gross**	**Net**
Total Return	78.32%	14.62%
Ann. Return	10.13%	2.30%
Ann. Volatility	0.096	0.095
Sharpe Ratio	1.053	0.241
Max Drawdown	-15.66%	-23.91%
Calmar Ratio	0.647	0.096
Win Rate	0.444	0.412
Skewness	0.029	-0.160
Kurtosis	16.603	16.751
Avg Active Pairs		4.214
Avg Daily RR Legs Traded		1.304
Ann. Transaction Cost		7.38%

Table 1: Strategy Performance Metrics

4.2 Sensitivity Analysis

To assess how robust the results are to modeling choices, we vary some key parameters one at a time while holding all others at their baseline values.

4.2.1 Delta Target (Δ_{target})

The baseline risk reversal targets the 25-delta strike on each leg. We re-run the backtest for $\Delta_{\text{target}} \in \{0.15, 0.25, 0.50\}$. The delta choice is also a cost lever: higher delta means a larger net delta to hedge, increasing equity execution cost. We want a strike far enough OTM to capture skew information, but not so far OTM that liquidity deteriorates and option prices become noisy.

Table 2: Sensitivity to delta target.

Value	0.15	0.25	0.5
Gross Sharpe Ratio	0.371	1.053	0.133
Gross Ann. Return	0.021	0.101	0.023
Gross Max Drawdown	-0.116	-0.157	-0.404
Net Sharpe Ratio	-0.393	0.241	-0.604
Net Ann. Return	-0.023	0.023	-0.104
Net Max Drawdown	-0.248	-0.239	-0.633
Total Transaction Cost	0.265	0.442	0.797

The 25-delta target performs best by a wide margin. Performance changes drastically across delta values, partly because we construct the skew signal using the 25-delta measure, switching to a different execution delta therefore introduces a mismatch between the signal and the traded instrument. Whether or not that mismatch explains all the variation, the sensitivity highlights that the results are not robust to this parameter choice, and that the 25-delta is the natural and internally consistent option.

4.2.2 Entry Threshold (z_{entry})

The entry threshold determines how extreme the idiosyncratic skew signal must be before a position is initiated. We sweep $z_{\text{entry}} \in \{0.95, 0.975, 0.99\}$ (percentiles). A higher threshold admits fewer but higher conviction trades, reducing transaction costs at the expense of lower capital utilization. A lower threshold increases trade frequency and diversification but may let in noisier signals.

Table 3: Sensitivity to entry threshold (percentile of expanding $|z_{ij,t}|$ distribution).

Value	0.95	0.975	0.99
Gross Sharpe Ratio	1.264	1.053	1.102
Gross Ann. Return	0.122	0.101	0.093
Gross Max Drawdown	-0.138	-0.157	-0.066
Net Sharpe Ratio	0.155	0.241	0.625
Net Ann. Return	0.015	0.023	0.052
Net Max Drawdown	-0.253	-0.239	-0.085
Total Transaction Cost	0.601	0.442	0.227

The gross Sharpe ratio does not vary dramatically across thresholds (1.26, 1.05, and 1.10), suggesting some robustness in the underlying signal quality. The net Sharpe, however, improves substantially with a tighter threshold: from 0.16 at the 95th percentile to 0.63 at the 99th. This improvement comes partly from simply trading less as total transaction costs fall from 0.60 to 0.23, but also from more selectively entering only the strongest-signal trades. The 99th-percentile variant also achieves a much more contained drawdown (8.5% vs. 25.3%), making it the most favourable risk-adjusted configuration.

4.2.3 Skew Metric

We compare all three skew measures defined in Section 2.1 on otherwise identical strategy configurations.

Performance is highly sensitive to the skew metric. The direct 25-delta measure is the only one that produces a positive net return. The polynomial skew yields a *negative* gross Sharpe of -0.53, i.e. the signal is directionally wrong. This is the expected outcome: since we trade risk reversals at the 25-delta level, the direct skew measure is the most

Table 4: Sensitivity to skew estimation method.

Value	direct	polynomial	naive
Gross Sharpe Ratio	1.091	-0.529	0.220
Gross Ann. Return	0.085	-0.036	0.016
Gross Max Drawdown	-0.084	-0.211	-0.173
Net Sharpe Ratio	0.081	-1.368	-0.791
Net Ann. Return	0.006	-0.094	-0.056
Net Max Drawdown	-0.162	-0.448	-0.329
Total Transaction Cost	0.451	0.373	0.438

internally consistent choice. The polynomial coefficient is estimated from all available strikes and may be distorted by the noisy far-OTM options, which are precisely the strikes we exclude from the instrument. The endpoint ("naive") measure performs better than the polynomial but is still unprofitable on a net basis, likely for similar reasons. Together, these results underscore that signal construction and instrument selection must be aligned.

We also did sensitivity analysis on the time to expiration target and the transaction cost. They are both included in Appendix H.

4.3 Regimes

To see what really drives our returns, we looked at how the strategy performs in different market conditions. As Figure 11 in Appendix G shows, regular trading costs constantly eat away at our profits over time. Breaking this down further reveals some patterns.

First, as seen in Figure 13, the strategy generates almost all of its net returns during down markets. In falling markets, there is a sudden rush to buy downside protection; this panic stretches the skew beyond fair value, creating mispricings large enough to easily cover our trading costs.

Second, we see a distinct "U shaped" outcome across volatility regimes (Figure 12), where we make money at the extremes but lose in the middle. The success in low volatility markets might simply come from stability: the market is quiet, so when a rare pricing anomaly does happen, it reliably reverts back, and our hedging and trading costs stay low. The "mid vol" environment, however, might just be the worst of both worlds. It could be too noisy, triggering false signals and forcing us to pay continuous trading fees but lacks the fast and predictable reversion needed to actually turn a profit.

5 Conclusion

This paper asked whether idiosyncratic equity option skew is mean reverting and whether that mean reversion is large enough to trade profitably after costs. The honest answer is: the signal is real, but the results are not strong.

Econometrically, the foundation is solid. All 21 unique skew pairs in our financial-sector universe are cointegrated at the 5% significance level under both Engle-Granger and Johansen tests, establishing that idiosyncratic skew spreads are stationary. The gross strategy performance bears this out: a Sharpe ratio of 1.05 and a cumulative return of 78.3% over 2015-2020 at the 25-delta, 15-day baseline. However, once we account for transaction costs, even at the modest 20 bps stylized assumption rather than the actual bid-ask spread, which averages 100-200% of option mid-prices, the net Sharpe collapses to 0.24 and the net return to 2.3%. The strategy breaks even around 25 bps.

The sensitivity analysis reinforces the fragility of the results. Sharpe ratios and drawdowns change drastically with the delta target, time-to-expiry, and skew metric. Only the direct 25-delta measure produces a positive net return; the polynomial skew produces a *negative* gross Sharpe, demonstrating that poor signal construction can actively destroy value. The one robust finding across all sensitivity dimensions is that tighter entry thresholds consistently improve net performance by reducing unnecessary turnover: the 99th-percentile threshold achieves a net Sharpe of 0.63 with a drawdown of only 8.5%. The strategy also benefits disproportionately from the COVID volatility episode of 2020, whose gains come largely from the delta hedge rather than from skew mean-reversion, an undesirable source of return.

Extensions. The most impactful improvements would directly address the cost problem. Weighting options by open interest when constructing the skew measure would concentrate signal in the most liquid strikes, reducing noise and improving signal consistency. SVI (stochastic volatility inspired) parametrization [3] provides an arbitrage-free smile interpolation and a principled way to extract skew without fitting to illiquid endpoints. A multi-factor decomposition of the systematic component of skew, separating market-wide fear, sector demand, and firm-specific components, would sharpen idiosyncratic isolation beyond the single-factor OLS regression. More ambitiously, constructing a vega-neutral rather than notional-equal-weighted portfolio would control overall volatility exposure and reduce unintended exposure to realized variance, particularly during stress periods. Furthermore, expanding the asset universe allows the strategy to enforce stricter entry thresholds without depressing trade frequency. Given that the sensitivity analysis linked higher thresholds to significant increases in net performance, a larger pool of assets is most likely essential for isolating high conviction trades at scale that are worth the cost.

References

[1] Mascia Bedendo and Stewart D. Hodges. The dynamics of the volatility skew: A kalman filter approach. *Journal of Banking & Finance*, 33(6):1156–1165, 2009.

[2] Bruno Feunou, Jean-Sébastien Fontaine, and Roméo Tédongap. Implied volatility and skewness surface. *Review of Derivatives Research*, 20(2):167–202, 2017.

[3] Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces, 2013.

A Appendix

Data Ingestion and Storage

We ingest raw data files (per-ticker option CSV archives, CRSP equity files, and a wide-format risk-free rate panel) are ingested into a local DuckDB database via a custom `DataLoader` class. The loader handles zip extraction, schema inference from accompanying metadata dictionaries, and type casting via `TRY_CAST` to handle mixed or malformed fields without aborting the load.

Feature Engineering and Enrichment

After ingestion, an `options_enriched` table is constructed by joining the raw options table with equity prices and risk-free rates, and computing the following derived fields:

- **Strike.** Raw strike prices are stored in tenths of cents in the source data; they are divided by 1,000 to obtain dollar-denominated strikes.
- **Mid price.** $\text{mid} = (\text{bid} + \text{ask})/2$.
- **Time to expiry.** Calendar days τ_{days} and annualized fraction $\tau = \tau_{\text{days}}/365$.
- **Forward price.** Using the cost-of-carry formula with the nearest-maturity risk-free rate r:

$$F_t = S_t\, e^{r\tau}.$$

- **Log-moneyness.** The forward log-moneyness:

$$k = \ln\left(\frac{K}{F_t}\right).$$

  Using the forward price as the normalising base ensures that at-the-money options correspond to $k = 0$ independent of the interest rate level.

B Appendix

Figure 3 visualizes the missing values per day after applying the filters discussed in Section 3.1. From Figure 3 we see that the number of missing values are not distributed equally over the dataset but are clustered on a few days. From Figure 4 we see that the number of missing values are quite low for most of the different stocks except BRK which has a lot of missing values. A solution to this could be to drop BRK from the dataset, but we decided to keep it.

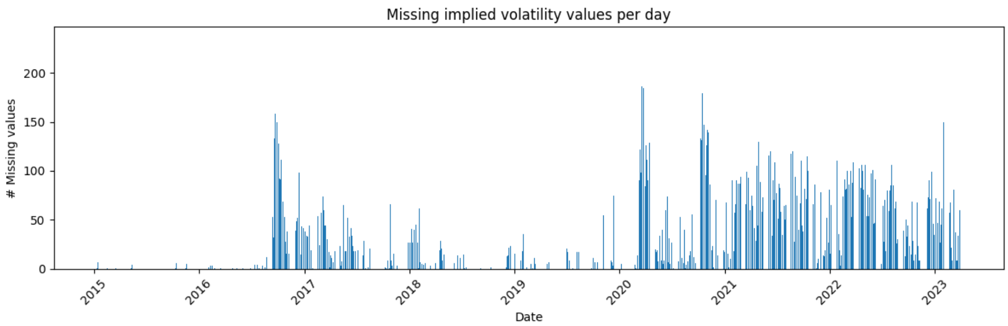


Figure 3: Number of missing values for implied volatility per day for the whole dataset.

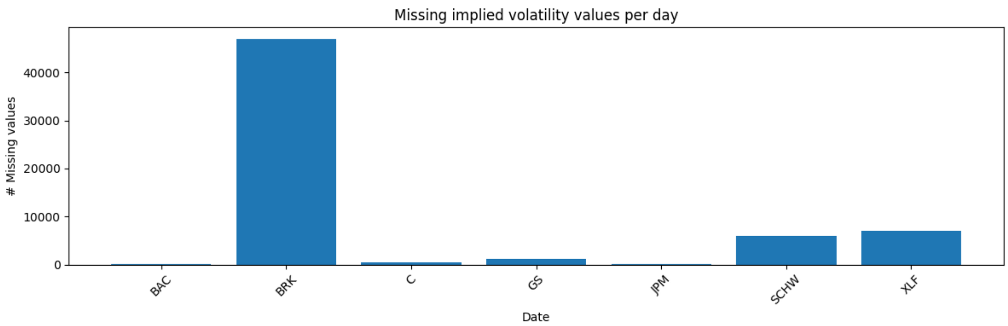


Figure 4: Number of missing values for implied volatility per ticker for the whole dataset.

C Appendix

- **Total and annualized return:** $\mathrm{TR} = \prod_t (1 + R_t) - 1; \quad \mu = (1 + \mathrm{TR})^{252/T} - 1.$
- **Annualized volatility:** $\sigma = \hat{\sigma}_R \times \sqrt{252}.$
- **Sharpe ratio:** $\mathrm{SR} = (\mu - r_f)/\sigma.$
- **Maximum drawdown:** $\mathrm{MDD} = \min_t \big(V_t / \max_{s \le t} V_s - 1 \big).$
- **Calmar ratio:** $\mu / |\mathrm{MDD}|.$
- **Win rate, skewness, and excess kurtosis** of daily returns.

D Appendix

Below is an overview of tests of cointegration. Table 5 shows metrics from the Johansen test and Table 6 shows metrics from the Engle-Granger.

Table 5: Johansen trace cointegration test: all unique ticker pairs of IV skew. H_0: no cointegration. Cointegrated when trace stat > 5% critical value.

Ticker 1	Ticker 2	Trace stat	p-value	Crit. 1%	Crit. 5%	N
SCHW	XLF	2250.506	—	19.935	15.494	4005
C	XLF	2064.541	—	19.935	15.494	4465
BRK	XLF	2059.845	—	19.935	15.494	3311
BRK	SCHW	2016.867	—	19.935	15.494	3238
BAC	BRK	1909.971	—	19.935	15.494	3311
BAC	XLF	1803.336	—	19.935	15.494	4485
GS	XLF	1728.543	—	19.935	15.494	4472
BAC	SCHW	1650.681	—	19.935	15.494	4003
BRK	GS	1613.443	—	19.935	15.494	3311
C	SCHW	1603.954	—	19.935	15.494	3983
JPM	XLF	1590.404	—	19.935	15.494	4467
BRK	C	1461.936	—	19.935	15.494	3310
JPM	SCHW	1366.628	—	19.935	15.494	4005
BAC	C	1355.933	—	19.935	15.494	4563
BRK	JPM	1308.270	—	19.935	15.494	3311
GS	SCHW	1303.963	—	19.935	15.494	4005
C	GS	1153.107	—	19.935	15.494	4548
C	JPM	954.842	—	19.935	15.494	4543
BAC	GS	906.297	—	19.935	15.494	4568
BAC	JPM	881.147	—	19.935	15.494	4563
GS	JPM	698.736	—	19.935	15.494	4565

Note: Cointegrated at 5%: 21/21 pairs.

Table 6: Engle-Granger cointegration test: all unique ticker pairs of IV skew. H_0: no cointegration. Stars: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Ticker 1	Ticker 2	τ-stat	p-value	Crit. 1%	Crit. 5%	N
BRK	GS	-29.315	0.0000***	-3.900	-3.338	3311
BRK	JPM	-13.631	0.0000***	-3.900	-3.338	3311
C	GS	-12.831	0.0000***	-3.899	-3.337	4548
BAC	XLF	-12.726	0.0000***	-3.899	-3.337	4485
C	JPM	-12.113	0.0000***	-3.899	-3.337	4543
BRK	C	-11.941	0.0000***	-3.900	-3.338	3310
BAC	GS	-11.814	0.0000***	-3.899	-3.337	4568
C	XLF	-11.658	0.0000***	-3.899	-3.337	4465
BAC	C	-11.553	0.0000***	-3.899	-3.337	4563
BAC	JPM	-11.505	0.0000***	-3.899	-3.337	4563
BAC	SCHW	-11.381	0.0000***	-3.899	-3.338	4003
BRK	XLF	-11.252	0.0000***	-3.900	-3.338	3311
C	SCHW	-11.107	0.0000***	-3.899	-3.338	3983
BRK	SCHW	-10.468	0.0000***	-3.900	-3.338	3238
JPM	XLF	-8.698	0.0000***	-3.899	-3.337	4467
SCHW	XLF	-8.534	0.0000***	-3.899	-3.338	4005
GS	JPM	-6.551	0.0000***	-3.899	-3.337	4565
GS	XLF	-6.328	0.0000***	-3.899	-3.337	4472
BAC	BRK	-5.492	0.0000***	-3.900	-3.338	3311
GS	SCHW	-5.343	0.0000***	-3.899	-3.338	4005
JPM	SCHW	-5.228	0.0001***	-3.899	-3.338	4005

Note: Cointegrated at 5%: 21/21 pairs.

.

E Appendix

Figure 5 and Table 7 shows that the spread as a percent of midprice for the option can be as big as 200%. Crossing this spread, results in extreme cost and is detrimental for trading strategies.

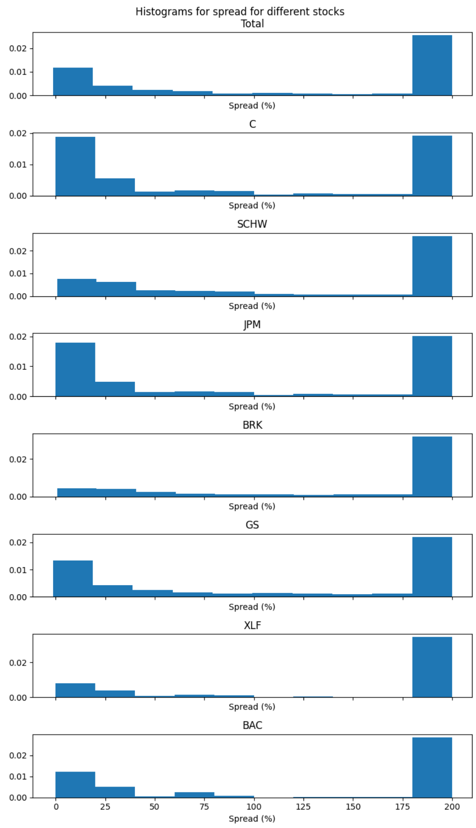


Figure 5: Spread in percent of option midprice for different stocks.

Table 7: Spread Percentage Statistics by Ticker

ticker	count	mean	std	min	25%	50%	75%	max
BAC	326220.00	126.12	87.39	0.00	22.22	200.00	200.00	200.00
BRK	389737.00	150.44	72.13	0.69	82.76	200.00	200.00	200.00
C	396493.00	95.44	87.91	0.00	9.52	50.00	200.00	200.00
GS	475516.00	112.83	85.42	-1.47	16.51	128.89	200.00	200.00
JPM	438228.00	99.45	88.08	0.00	9.84	66.67	200.00	200.00
SCHW	294685.00	129.34	80.54	0.63	36.73	200.00	200.00	200.00
XLF	336958.00	148.60	79.40	0.00	66.67	200.00	200.00	200.00

F Appendix

Below are plots from the experiment not included in the main text.

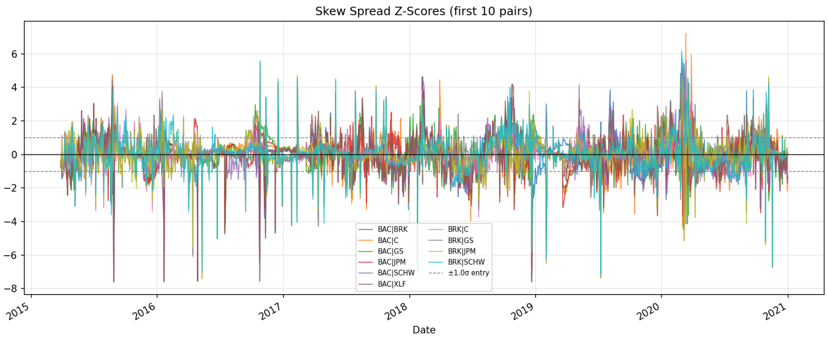


Figure 6: Z-scores for the regressions. The z-scores are quite heavy-tailed.

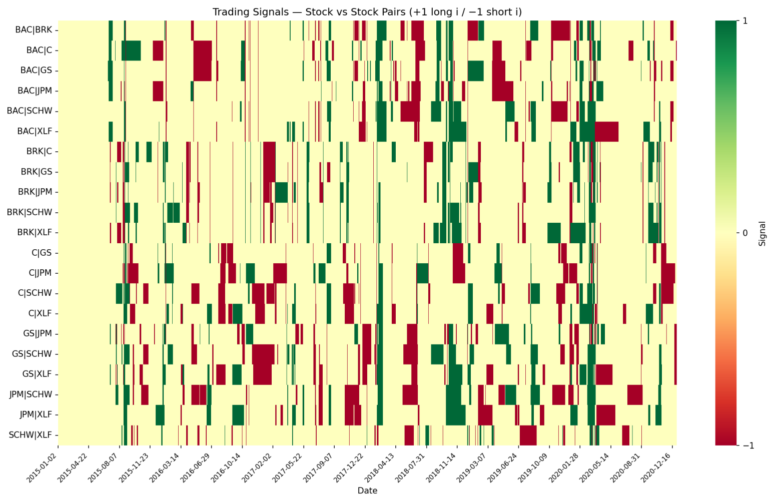


Figure 7: The plot of signals shows that we trade in almost all pairs. The column of empty trades is the estimation period.

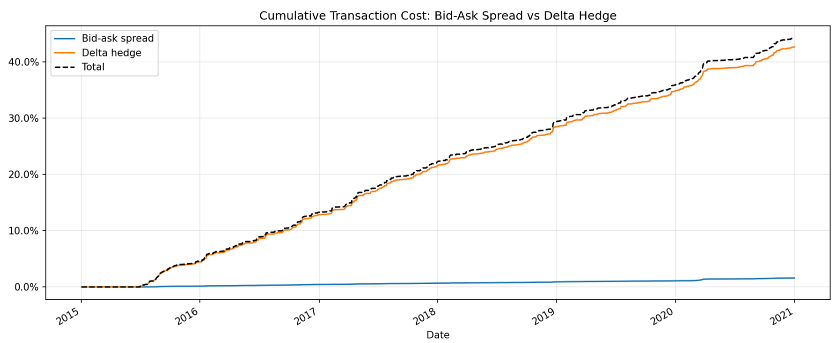


Figure 8: The transaction breakdown shows that most of the trading cost come from the hedging. The transaction cost for options would have been higher if we modeled it as the bid-ask difference.

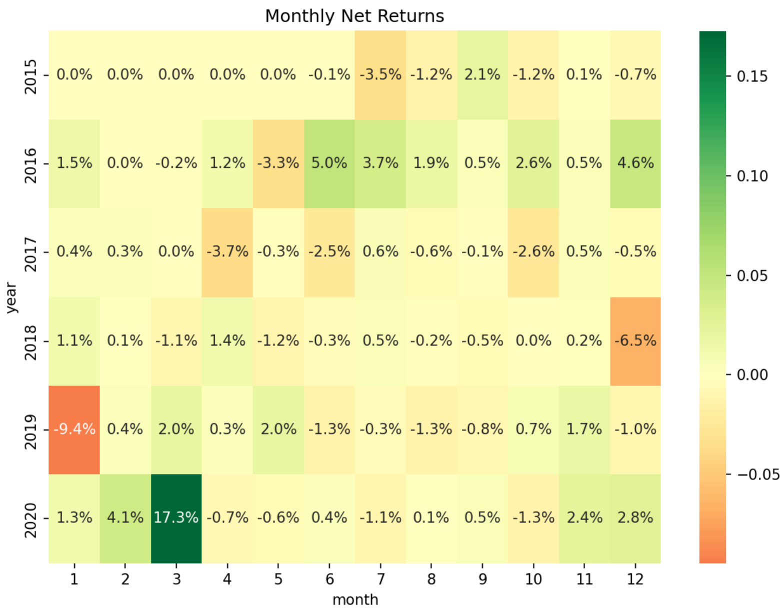


Figure 9: Monthly returns over the trading period. Most of the negative returns comes from December 2018 and January 2019 while most of the upside comes from March 2020.

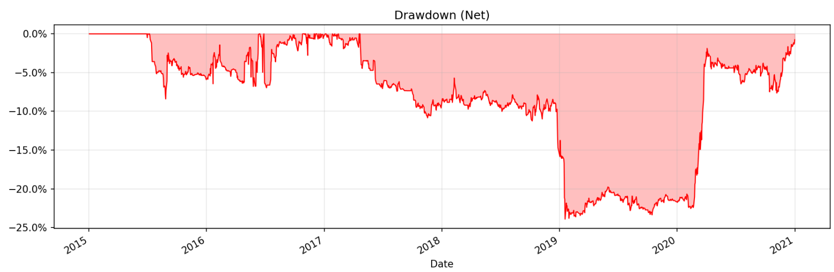


Figure 10: Drawdown during trading period. Again we see that the trading strategy experienced a sharp decline in late 2018 early 2019 and a sharp increase in 2020.

.

G Appendix

This section provides a granular breakdown of the strategy's return drivers. It includes a longitudinal view of cumulative P&L components, isolating option gains from hedging activity and transaction costs, alongside performance attribution across varying volatility regimes and market directions. These visualizations serve to highlight the strategy's convexity profile and its sensitivity to execution friction.

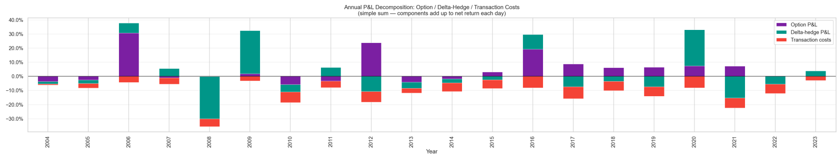


Figure 11: Cumulative P&L decomposition showing the performance of the option position, the offsetting delta-hedge, and the steady erosion caused by transaction costs from 2004 to 2024.

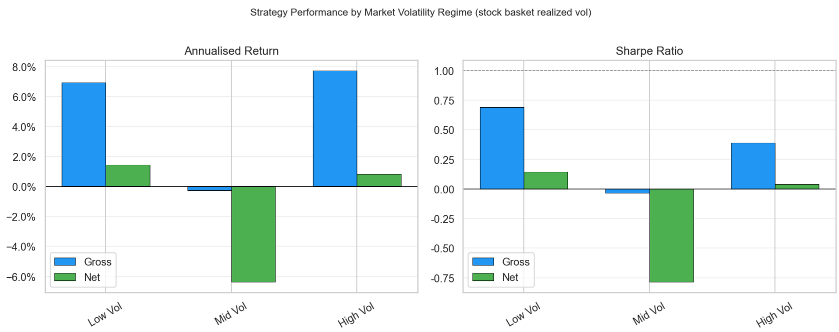


Figure 12: Strategy performance (Annualised Return and Sharpe Ratio) segmented by market volatility regimes.

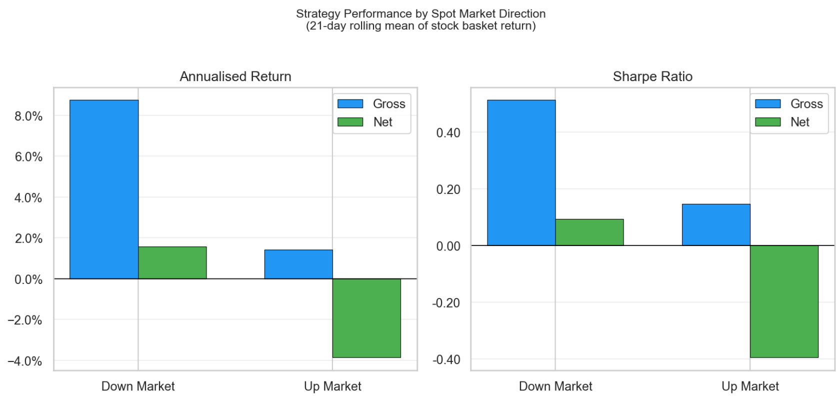


Figure 13: Analysis of strategy returns and risk-adjusted performance across different spot market directions, demonstrating a performance bias toward "Down Market" environments on both a gross and net basis.

H Appendix

H.1 Time to Expiration Target

Table 8: Sensitivity to time-to-expiry target (calendar days).

Value	15	20	25
Gross Sharpe Ratio	1.053	0.584	0.423
Gross Ann. Return	0.101	0.056	0.039
Gross Max Drawdown	-0.157	-0.201	-0.204
Net Sharpe Ratio	0.241	-0.210	-0.393
Net Ann. Return	0.023	-0.020	-0.036
Net Max Drawdown	-0.239	-0.322	-0.350
Total Transaction Cost	0.442	0.445	0.446

Performance degrades monotonically as the target maturity increases. Transaction costs are roughly constant across maturities (the bps model does not depend on tenor), so the worsening net performance directly tracks the weakening gross signal. Near-dated options are more sensitive to short-lived demand imbalances and revert faster; longer-dated skew incorporates more structural, slow-moving information that does not revert on the timescale our strategy assumes. The 15-day tenor is the only configuration that survives costs, and even then only marginally.

H.2 Transaction Costs

As described in Section 2.4, we model all execution costs (both option legs and equity hedge) as a flat b basis points rather than the actual bid-ask spread, which is prohibitively wide (Appendix E).

Table 9: Sensitivity to execution cost assumption (bps applied to both option legs and equity hedge).

Value	0	20	40
Gross Sharpe Ratio	1.053	1.053	1.053
Gross Ann. Return	0.101	0.101	0.101
Gross Max Drawdown	-0.157	-0.157	-0.157
Net Sharpe Ratio	1.053	0.241	-0.522
Net Ann. Return	0.101	0.023	-0.050
Net Max Drawdown	-0.157	-0.239	-0.388
Total Transaction Cost	0.000	0.442	0.885

Gross performance is identical regardless of the cost assumption, as it must be. At zero cost the strategy is strong (Sharpe 1.05). At 20 bps the net Sharpe falls to 0.24. At 40 bps the strategy is clearly unprofitable. The strategy is not cost-efficient: it trades frequently, and higher costs hurt it severely. See Figure 7 in Appendix F for how much the base strategy trades. This is the central limitation of the approach as currently designed.