CROSS ETF ARBITRAGE

Adrianus Christopher
Benjamin Yong
Dillon Low
Gabriel Tan
Koh Shi Ling
Low Yuan Ting
Mikhail Riyad

DUAL APPROACH STRATEGY

Simulates the balance of an **arbitrage strategy** in a highly liquid/mostly arbitrage free market and a **market maker** that provides a source of liquidity for strategy execution

Strategy
STATISTICAL ARBITRAGE

- Exploits temporary price discrepancies between cointegrated ETF pairs
- Employs mean-reversion principles based on statistical relationships

Counterparty
MARKET MAKING

- Provides liquidity to the market while capturing bid-ask spreads
- Acts as the execution venue for arbitrage trades
- Revenue stream through spread capture and inventory management
- Counter trades our algorithmic trading strategy to simulate real world market/taker situations based on market factors

To identify suitable **trading pairs**, the system employs rigorous statistical methods - beginning with correlation analysis and advancing to formal cointegration testing whereby multiple filters are implemented to ensure statistical robustness

CORRELATION FILTERING

- Initial screening applies a **minimum Pearson correlation coefficient of 0.8** between ETF price series as a preliminary filter to **identify potentially cointegrated pairs**, followed by **formal cointegration testing** to confirm long-term equilibrium relationships necessary for **mean-reversion strategies**

COINTEGRATION TESTING

- **Engle-Granger**[1] test is implemented to identify pairs with long-term equilibrium relationships
- Only pairs with **cointegration** p-values < 0.05

DATA QUALITY ASSESSMENT

- Additional sanity check filters to ensure **minimum 80% data overlap** on common dates and sufficient historical observations for reliable statistical inference

[1]Engle, R. F. & Granger, C. W. J. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55, 251–276 (1987).

TRADITIONAL APPROACH

Employs classical statistical arbitrage principles based on **price ratio analysis**

Price Ratio Analysis

For each ETF pair, the system calculates the price ratio ETF_1/ETF_2 and tracks its statistical properties over a rolling 60-day window, trading **signals are generated when the absolute z-score exceeds the threshold of 2.5 standard deviations**, indicating significant deviation from the statistical mean

Cons

- Lagging signals
- Static assumptions
- Noise sensitivity
- Inefficient timing
- Ratio contains signal and noise which are then amplified by the z-scoring process
- Alpha not orthogonal to market beta

ADVANCED APPROACH

Employs state-space modeling through **Kalman Filtering**[2] to enhance signal quality and reduce noise

State Space Model

Models spread as a **mean-reverting** process with time-varying parameters:

- State equation captures the **unobserved fair value**
- Observation equation for noisy market prices
- **Kalman Gain** automatically measures and balances:
    - Process noise (true relationship change)
    - Measurement noise (noise in observed prices)

Benefits of Kalman Filtering

- Regularisation at an **endogenous** level
- Non-parametric method
- Noise filtering and quantifies uncertainty
- Detects regime changes
- Provides smoother signals

[2]Kalman, R. E. & Buey, R. A new approach to linear filtering and prediction theory. *Trans. ASME, J. Basic Eng.* 83, 95–108 (1961).

State Space modelling through Kalman Filtering

Refers to a recursive state space model used to estimate hidden variables (e.g. true spread mean) in the presence of noise

1 — Models ETF spread as a mean reverting process

2 — Apply Kalman Filter recursion rolling windows to:

- Smooth observed price ratio ETF1/ETF2
- Adapt to changing market regimes and quantify uncertainty in signal quality

3 — Kalman Filter results:

- Filters out random price noise, avoiding false signals during volatile periods
- Allows entry only when spread deviation is statistically credible
- Adjusts dynamically to regime shifts

Results provides superior signal quality through:

- Adaptive parameter estimation
- Noise reduction and smoothing
- Uncertainty quantification
- Adaptive reversion speed estimation

KALMAN FILTER FORMULAS

PREDICTION STEP

State Estimate $$\hat{x}_{k|k-1} = F_k \, \hat{x}_{k-1|k-1}$$

Covariance Estimate $$P_{k|k-1} = F_k \, P_{k-1|k-1} \, F_k^\top + Q_k$$

UPDATE STEP

Innovation $$y_k = z_k - H_k \, \hat{x}_{k|k-1}$$

Innovation Covariance $$S_k = H_k \, P_{k|k-1} \, H_k^\top + R_k$$

Kalman Gain $$K_k = P_{k|k-1} \, H_k^\top \, S_k^{-1}$$

State Update $$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \, y_k$$

Covariance Update $$P_{k|k} = (I - K_k \, H_k) \, P_{k|k-1}$$

POSITION LIMITS

30%
max capital per Individual position

30%
max net exposure

10x
max gross leverage at portfolio level

LOSS LIMITS

2%
daily value-at-risk limit

15%
max drawdown limit

real time monitoring

CONCENTRATION LIMITS

20%
max allocation to a single asset

correlation adjusted position sizing

sector and geographical diversification

```
Data obtained from Bloomberg

Ticker | ETF Name
-------|-------------------------------
IWV    | iShares Russell 3000 ETF
VTHR   | Vanguard Russell 3000 ETF
SPY    | SPDR S&P 500 ETF Trust
IVV    | iShares Core S&P 500 ETF
VOO    | Vanguard S&P 500 ETF
QQQ    | Invesco QQQ Trust (Nasdaq-100)
IWM    | iShares Russell 2000 ETF
VTWO   | Vanguard Russell 2000 ETF
EWJ    | iShares MSCI Japan ETF
FXI    | iShares China Large-Cap ETF
MCHI   | iShares MSCI China ETF

Years 2011-2025
Ask | Bid | Last | Volume
```

DATA CLEANING

- Missing value imputation
- Invalid value removal
- Timestamp alignment

OUTLIER DETECTION

- Interquartile range
- Z-score threshold

FEATURE EXTRACTION

- Price decomposition via Kalman
- Volatility: rolling window, annualised
- Correlation calculations
- Cointegration testing
- Spread calculations and half-life estimation

QUALITY REPORT

- Data shape, date range, minimum data overlap requirements
- Correlation thresholds and cointegration p-value filtering (< 0.05)

KELLY CRITERION

Is a formula that **maximises long-term geometric mean of returns** through optimal asset allocation based on signal quality[3]. It balances expected returns against variance.

FORMULA

$f^* = (bp - q) / b$,

where

b (odds ratio) = net odds received on winning bet

p = probability of winning bet

q = probability of losing bet

USAGE

Return-covariance is estimated using discrete and pair-specific historical win/loss statistics. Kelly fractions **f*** are applied to the **next period to maximise long-run geometric growth rate**. Kelly fractions are computed for each pair with **position sizings being adjusted dynamically** based on performance.

ADVANTAGES

It **avoids under/over-allocation** by **optimising geometric growth rate** rather than equal-weight allocation/risk parity allocation. Applies leverage where **empirically justified** by historical win/loss statistics.

[3]Maclean, L. C., Thorp, E. O. & Ziemba, W. T. Long-term capital growth: The good and bad properties of the kelly and fractional kelly capital growth criteria. *Quant. Finance.* **10**, 681–687 (2010).

STRUCTURED DAILY EXECUTION WORKFLOW

1	**DATA PROCESSING**	**Ingest** latest ETF price, volume and return data from market feeds Performs **cleaning**, **validation** and **formatting**
2	**SIGNAL GENERATION**	**Generates** trading signal across all active ETF pairs using both **traditional** (z-score) and **Kalman Filter** models Flags actionable opportunities
3	**RISK ASSESSMENT**	Measures **portfolio-level exposure**, **leverage**, and **factor risks**
4	**TRADE EXECUTION**	Send **qualified** trades to the integrated market maker engine and ensures **dollar neutral** execution
5	**PORTFOLIO RECONCILIATION**	**Update** all positions post-trade and **recalculates PnL**, **risks contributions**, and **position sizing** for compliance with our limits
6	**PERFORMANCE REPORTING**	Generates end-of-day reports on: **alpha contribution**, **risk-adjusted returns**, **drawdown**, and **key risk metrics**

VECTORISED KALMAN FILTERING — 1

- Implemented using **linear algebra optimisation** to improve runtime
- Kalman Filter written in matrix form for prediction and update steps
- State-space modelling allows **adaptive reversion estimation** in real time

PARALLEL COMPUTE ARCHITECTURE — 2

- Arbitrage and market making components run on a single thread **asynchronously**
- Enables asynchronous quote updates and real-time execution logic

REAL-TIME PERFORMANCE — 3

- Efficient runtime achieved through **vectorisation**

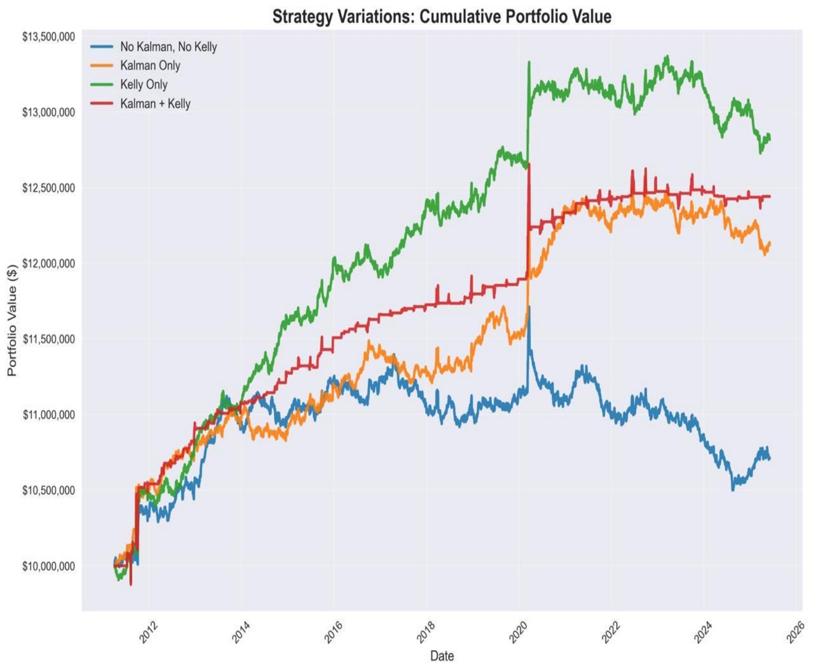
ROLLING WINDOW FRAMEWORK — 4

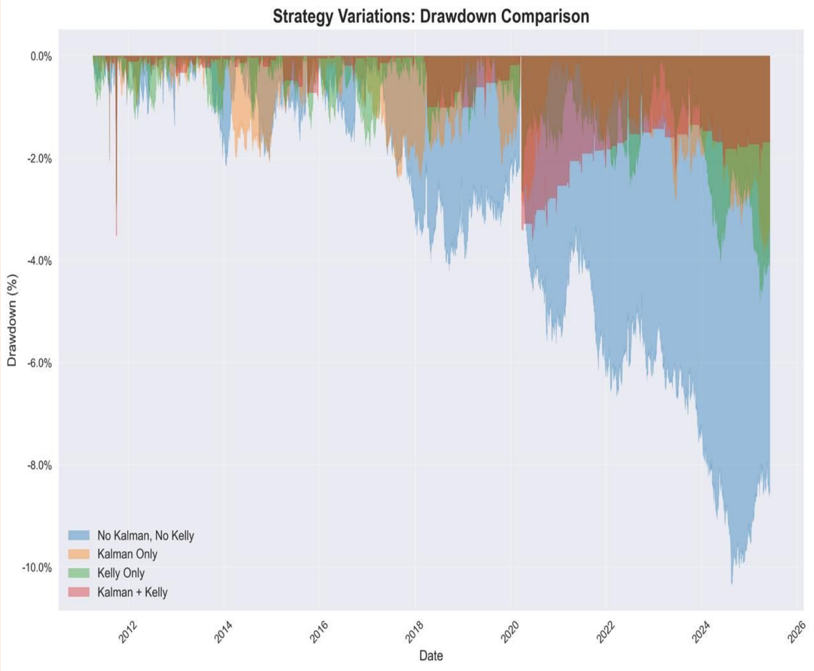
- **Kalman**: trained on 6-month window, walk forward cross-validation, validated on next-month data
- **Kelly**: estimates return-covariance matrix on rolling basis

DATA VALIDATION FILTERING — 5

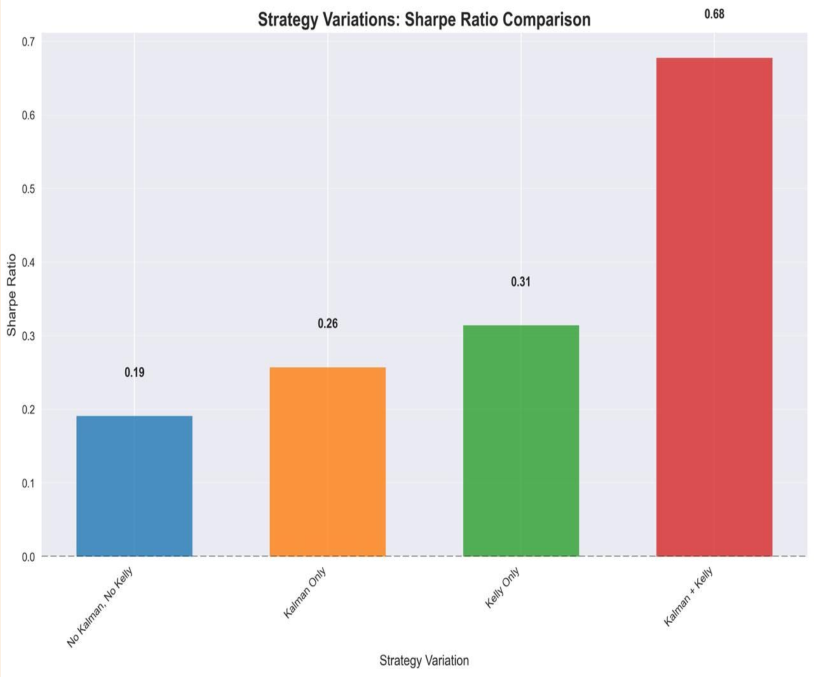
Pairs screened by :
- **Correlation** > 0.8
- **Engle-Granger** cointegration p-value < 0.05
- 80% minimum historical data overlap

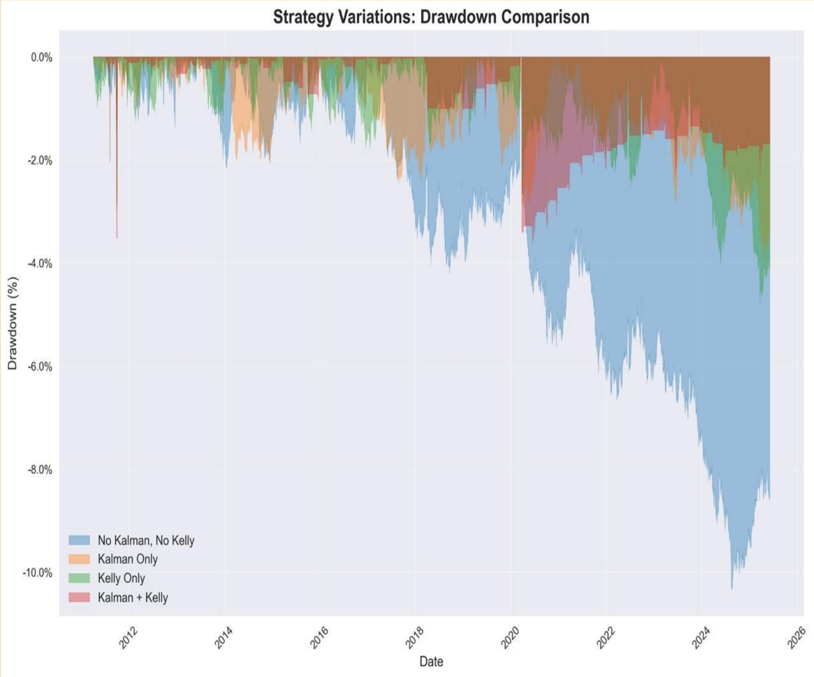
Kalman Filtering alongside Kelly Criterion outperform baseline Z-score strategy



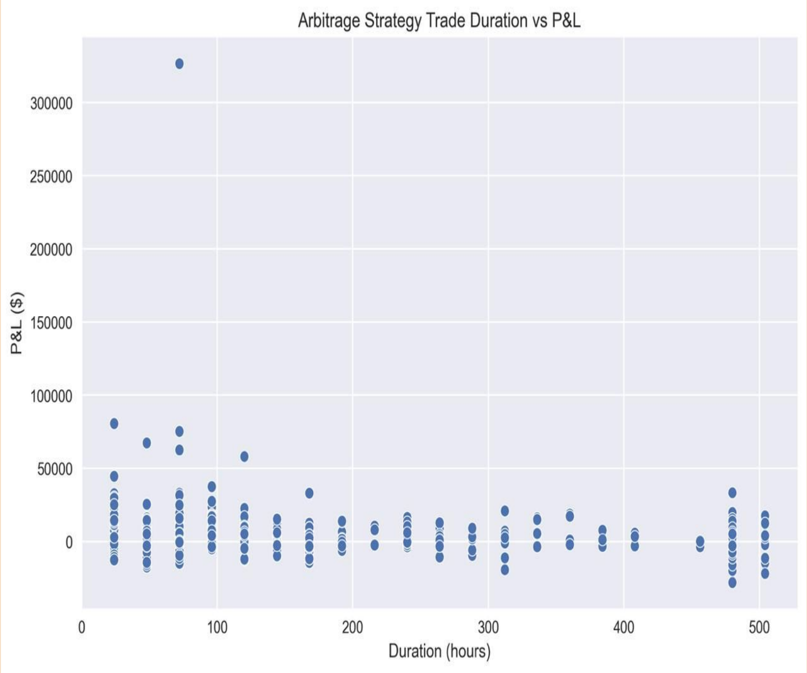


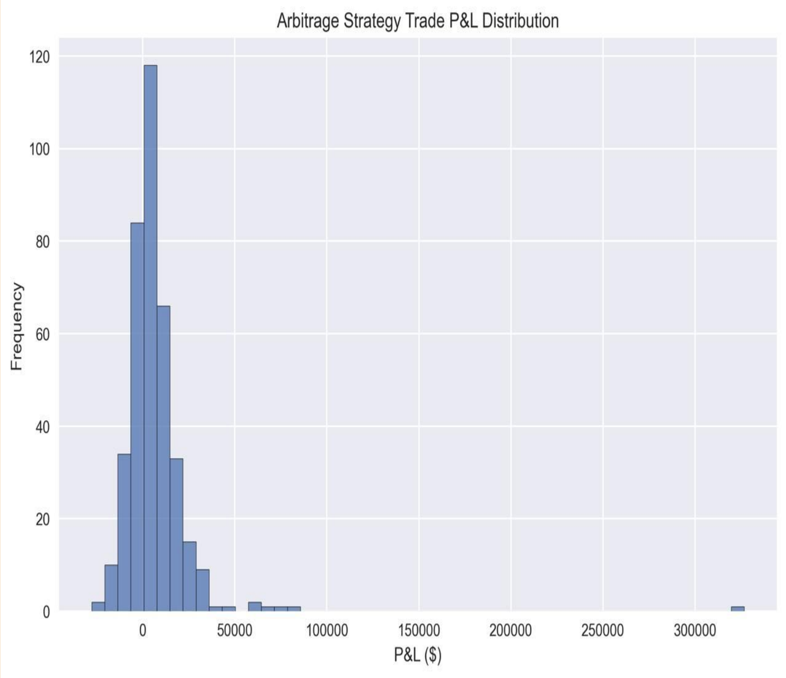
Kalman Filtering alongside Kelly Criterion gives a higher Sharpe with lower max drawdown



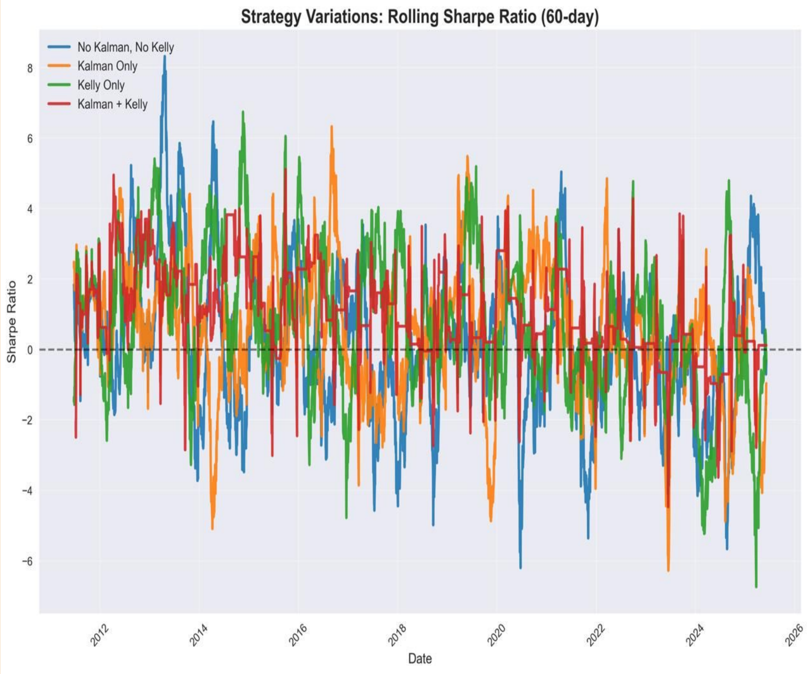


P&L Distribution for Arbitrage Strategy





RAS Penalised Sharpe and Rolling Sharpe



Market Making component employs the **Multi-Factor Spread Model** to adapt to market conditions
ie. it combines multiple factors to determine optimal bid-ask spreads

1. **Volatility Component**
Adjusts the spread in response to market volatility

Volatility multiplier calculated by
$1 + (\text{Current Vol} / \text{Median Vol} - 1) \times 0.5$

2. **Time-of-Day Component**
Applies a premium to the base spread depending on the time of day

Market open/close: +30% spread premium
lunch period: +10% spread premium
normal hours: base spread

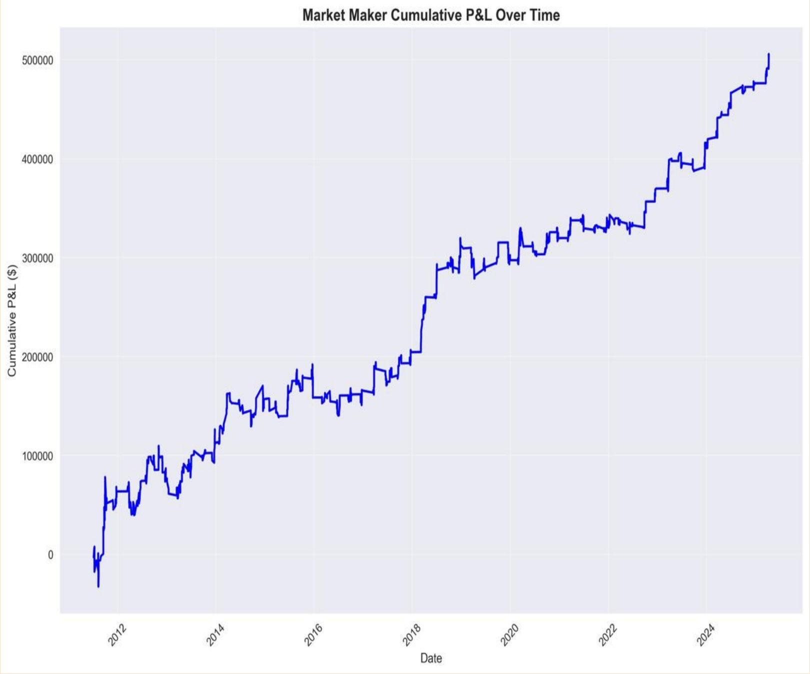
3. **Inventory Penalty**
Is applied to manage inventory risk and maintain liquidity discipline

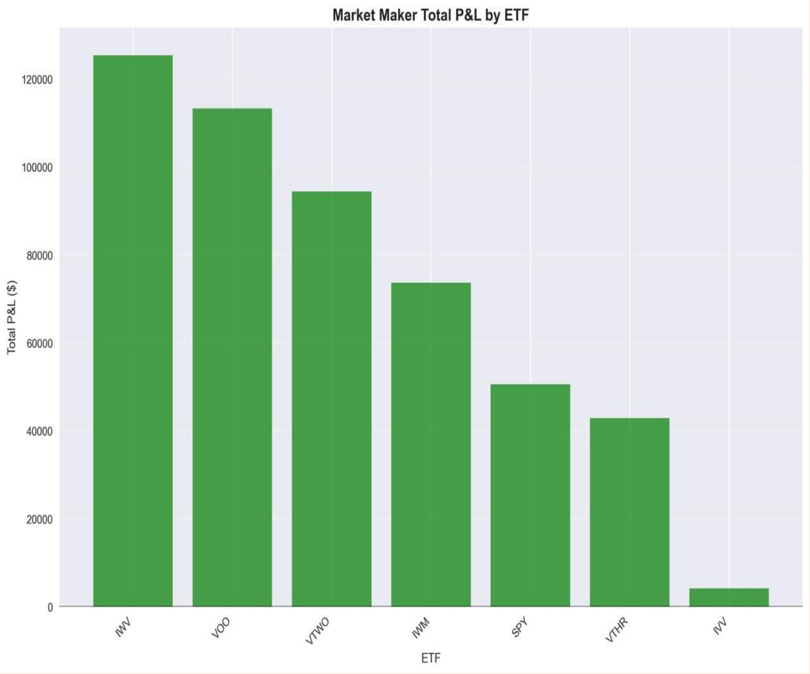
Calculated by
$(\text{Current Position} / \text{Max Position})^2 \times 0.001$

When arbitrage trades are executed, the Market Maker:
- takes the opposite side of each trade
- captures the bid-ask spread
- manages inventory exposure
- applies appropriate slippage models for large orders

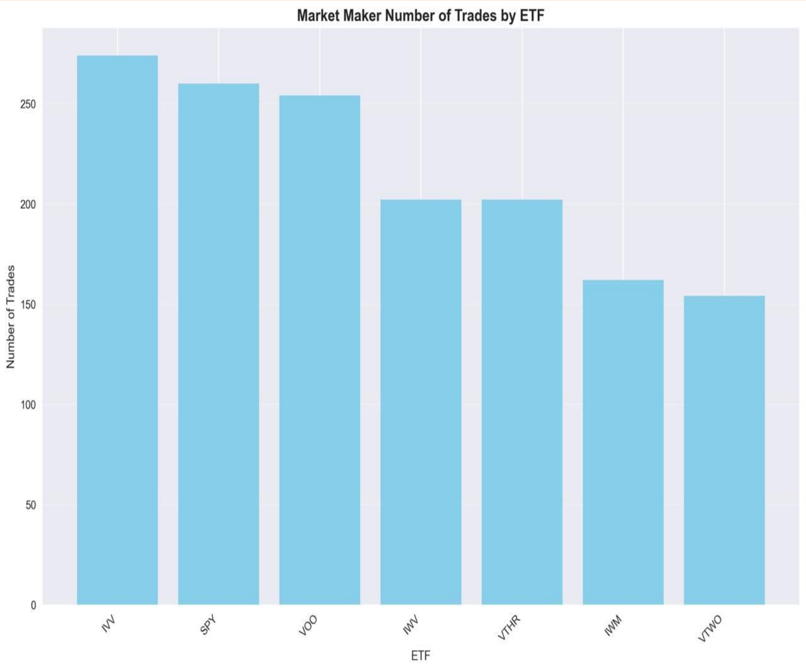
MM P&L calculated by taking **(Execution Price - Mid Price) × Quantity**

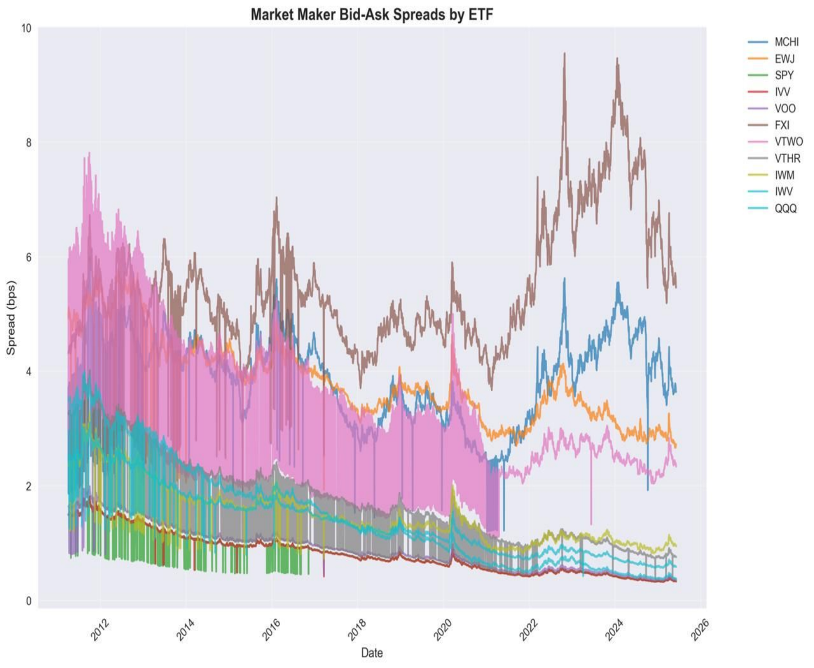
Market Maker P&L and Profit Breakdown by ETF name





Market Maker Trades and Spread





Rademacher Complexity helps **control data snooping bias** and **false discoveries** by **penalising the empirical Sharpe Ratio** based on the **complexity of the strategy space** where

penalised Sharpe = empirical Sharpe - 2 * Rademacher Complexity

This penalty reflects the **models capacity to overfit** by accounting for the size and complexity of the strategy space, serving as a **pre-emptive adjustment before evaluating overfitting risk**. This helps enable a more **robust** and **statistically valid performance comparison** across strategies.

Rademacher Anti-Serum for Sharpe:

$$\theta_n \geq \widehat{\theta}_n - 2\widehat{R} - 3\sqrt{\frac{2\log(2/\delta)}{T}} - \sqrt{\frac{2\log(2N/\delta)}{T}}$$

* Rademacher Anti-serum

IN SUMMARY

- **Kalman and Kelly combination** provides **robust**, **adaptive signal generation** complemented with statistical validation, risk constraints and penalty theory
- Market Maker demonstrates **real life market making problems**, included bid/ask spread penalty system based on time of day liquidity, inventory etc. to model how market makers behave in real life
- **Negative Rademacher penalty** indicates **strategy robustness**, Kalman and Kelly combination naturally creates anti-noise properties, which signifies orthogonality to market beta

POSSIBLE FUTURE INTEGRATIONS

- **Hidden Markov Models** with multiple states, alongside ensemble methods for **regime classification**
- **Structure Break Detection** (CUSUM)
- **Volatility Clustering** for Spread (GARCH, EGARCH)
- **Threshold Cointegration** models
- **Monte Carlo** Permutation
- **TWAP/VWAP** execution algorithms for **smooth execution/offloading**

END