Rolling Cointegration Pairs Strategy: Theory and Empirical Analysis

Lukas Schaller

July 28, 2025

Abstract

We investigate a rolling-OLS cointegration pairs strategy that enters β-neutral long-short positions whenever the standardized spread (z-score) between two assets exceeds a pre-defined band and exits when the spread reverts. Daily data for ten major cryptocurrencies (2022-2025) and eleven economically linked ETF pairs (2018-2025) are screened and tested for cointegration. Back-testing is performed to determine the optimal parameters for entry and exit thresholds and rolling window size. ETF pairs that deliver Sharpe ratios around 1.0 and competitive maximum drawdowns are identified. Cryptocurrencies fail to yield satisfactory results. The strategy is robust to transaction costs. We conclude that rolling cointegration is a viable alpha source for fundamentally related ETFs, even though absolute returns are modest.

Contents

1 Theoretical Framework

1.1 Cointegration and the Pricing Spread

Let P_t^A and P_t^B denote the prices of two assets A and B at time t. A *cointegrating relation* exists if there is a linear combination

$$\varepsilon_t = P_t^A - (\alpha + \beta P_t^B)$$

such that ε_t is stationary. In practice we estimate the coefficients (α, β) with ordinary least squares (OLS) from the model:

$$P_t^A = \alpha + \beta P_t^B + \varepsilon_t.$$

Here no assumptions are made about the distribution of the error term ε_t.

1.2 Decision Strategy

We will assume that the pricing spread ε_t is a stationary process meaning that its mean and variance are constant over time. Hence, we expect it to revert to its mean μ over time when it deviates from it. The strategy is to enter the market as soon as a gap between the price P_t^A and the linear combination of the price P_t^B exceeds a certain threshold. We will go long on the asset that is underpriced and short the overpriced asset. As soon as the spread reverts to its mean, we close the position and take the profit. These ideas are extensively studied in the context of pairs trading [1, 2].

So, how to identify the entry and exit points? We will use the z-score z_t of the residuals ε_t to determine the entry and exit points. Recall that the z-score is defined as

$$z_t = \frac{\varepsilon_t - \mu}{\sigma},$$

where μ is the mean and σ is the standard deviation of the residuals ε_t. Our strategy S_0 will then be

$$S_0 = \begin{cases} \text{Long A, Short B} & \text{if } z_t \leq -z_{\text{entry}}, \\ \text{Long B, Short A} & \text{if } z_t \geq +z_{\text{entry}}, \\ \text{Exit} & \text{if } |z_t| < z_{\text{exit}}. \end{cases} \tag{1}$$

Here z_{entry} and z_{exit} are the defined entry and exit thresholds, respectively.

1.3 Rolling Estimation

Because in reality there is no guarantee that the cointegration relation holds forever, we cannot be sure that the residuals ε_t are stationary for very long periods of time. Hence, we will estimate the residuals as well as the corresponding

z-score in a rolling window of length w:

$$
\begin{aligned}
(\alpha_t, \beta_t) &= \mathrm{OLS}\left(P^A_{t-(w-1)}, \ldots, P^A_t \,;\, P^B_{t-(w-1)}, \ldots, P^B_t\right), \\
\varepsilon_t &= P^A_t - (\alpha_t + \beta_t P^B_t), \qquad z_t = \frac{\varepsilon_t - \mu_t}{\sigma_t}, \\
\mu_t &= \mathrm{mean}\left(\varepsilon_{t-(w-1)}, \ldots, \varepsilon_t\right), \qquad \sigma_t = \mathrm{stdev}\left(\varepsilon_{t-(w-1)}, \ldots, \varepsilon_t\right).
\end{aligned}
$$

This allows us to adapt to changing long-term trends in the prices of the considered assets.

1.4 Further Mechanisms

Recall the strategy S_0 defined in (1). We can add an additional stop-loss mechanism to it, which will close the position if the current loss exceeds a certain threshold. Another mechanism is to add an absolute threshold $\varepsilon_{\min}$ to the residuals, which needs to be exceeded before entering a position. This can be very useful for low-volatility regimes, where the z-score even after small movements might be large due to a very small standard deviation.

1.5 Testing for Stationarity

To test whether the residuals ε_t are stationary, one can use the Augmented Dickey-Fuller (ADF) test [3].

Augmented Dickey–Fuller test. For a series x_t with mean 0 the ADF regression is

$$
\Delta x_t = \rho\, x_{t-1} + \sum_{i=1}^{p} \gamma_i\, \Delta x_{t-i} + u_t,
$$

where $\Delta x_t = x_t - x_{t-1}$. ρ and $(\gamma_i)_{i=1}^p$ are estimated coefficients and u_t is a white noise error term. We now test the null hypothesis $H_0 : \rho = 0$ against the alternative $H_1 : \rho < 0$. The test statistic $\tau = \hat{\rho}/\,\mathrm{SE}(\hat{\rho})$ is compared against MacKinnon critical values [4]. The more negative the value of τ, the stronger the evidence against H_0.

2 Data

Data sources.

- **Exchange-Traded Funds (ETFs).** Daily close prices were downloaded from *Yahoo Finance* [5] via the `yfinance` API in Python.
- **Cryptocurrencies.** Daily prices were downloaded from *Binance* [6] using the `ccxt` Python library [7].

All series are sampled at a fixed *1-day* interval.[1] For each asset we retain the most recent 1000 - 2000 trading days (roughly four to eight trading years), which is sufficient to calculate rolling 60-day estimates while leaving sufficient out-of-sample history for back-testing.

Data pre-processing.

- Prices of an asset pair are stored in a single `pandas DataFrame` [8] indexed by ISO date, ensuring data alignment.
- No currency conversion is required: both ETF and crypto prices are denominated in U.S. dollars.

3 Screening results

During the initial screening, 10 pairs of cryptocurrencies were tested. We calculated the R^2 of the OLS regression to filter out pairs in a first step. For the remaining pairs, we calculated the ADF test statistic to check for stationarity of the residuals. The results of the currency pairs with $R^2 \geq 0.7$ and $p_{\text{ADF}} < 0.05$ are summarized in Table 1. [2] Actual tests of the strategy performed even on

Table 1: Cryptocurrency pairs that passed the initial screening.

Pair	R^2	ADF statistic	p_{ADF}
ETH/USDT - BCH/USDT	0.731	-3.58	0.0061
ADA/USDT - XLM/USDT	0.793	-3.28	0.0158
ADA/USDT - LINK/USDT	0.722	-2.97	0.0377
XRP/USDT - XLM/USDT	0.869	-2.89	0.0468

these pairs yielded unsatisfactory results. The main reason seems to be that there is no strong enough force to push the spread back to its mean, because there is no economic reason for the prices to be cointegrated. Hence, we decided to focus on pairs of ETFs. Here it is much more likely to find cointegrated pairs,

[1] Higher-frequency data approaches were not studied in this work.

[2] The full set of cryptocurrencies of which all possible pairs were screened consists of BTC, ETH, XRP, LTC, BCH, ADA, DOT, LINK, XLM, and DOGE. All series are sampled from 01 Nov 2022 through 27 Jul 2025.

because of economic relationships or just pairs of ETFs containing very similar assets. Results of the success of the strategy on ETF pairs can be found in section 5.

4 Back-Test Design

4.1 Parameter grid

The back-test is run for different assets and parameter combinations.

- **Entry threshold (z_{entry}):** $\{2.0, 2.5\}$. This corresponds to a $> 97\%$ and $> 99\%$ confidence under normality assumptions.
- **Exit threshold (z_{exit}):** $\{0.05, 0.25\}$. Forces profit-taking once the spread has substantially decreased.
- **Rolling window (w):** $\{40, 50, 60\}$ trading days for the OLS hedge ratio and z-score moments.
- **Absolute residual filter ($|\varepsilon_t| \geq \varepsilon_{\min}$):** $\{0.0, 0.5\}$. Might be helpful in uncertain low-volatility regimes.
- **Stop-loss:** 5 % of invested capital per asset.

This yields $2^3 \cdot 3 = 24$ runs per asset pair.

4.2 Position sizing and capital at risk

All trades are β-neutral: one unit of asset A is hedged by $\beta_t \cdot P_t^A / P_t^B$ units of asset B. The gross capital invested on day t is therefore

$$G_t = \gamma \cdot \left(P_t^A + |\beta_t| P_t^A \right), \qquad \gamma \in \mathbb{R}_{>0},$$

and $\max_t G_t$ serves as capital-at-risk when computing total-return percentages. A 0.1% transaction cost is applied to each entry action.

4.3 Performance metrics

For every run we record

- total Sharpe ratio,
- Sharpe ratio during invested days,
- percentage of days invested,
- maximum draw-down,
- total return over capital-at-risk, and
- the same metrics for a 50-50 buy-and-hold benchmark.

5 Empirical results

5.1 Global grid sweep

Table 2 shows the top five parameter runs during back-testing on the parameter grid for different pairs of ETFs[3]. We observe that the best-performing pairs are TLT/IEF and EEM/VWO. TLT/IEF dominates in terms of Sharpe ratio, but is invested for a larger percentage of time and suffers from a larger maximum draw-down than EEM/VWO. The parameters chosen are similar, even though EEM/VWO seems to prefer a slightly higher entry threshold which also explains the difference in the percentage of days invested. The detailed metrics of the best-performing run of TLT/IEF can be found in Table 3.

Table 2: Top five parameter runs sorted by total Sharpe ratio.

Pair	Sharpe	Time in [%]	Max DD	z_{entry}	z_{exit}	w	$\varepsilon_{\min}$
TLT/IEF	1.04	56.8	−4.33	2.0	0.05	50	0.0
TLT/IEF	0.97	34.1	−3.57	2.0	0.25	50	0.0
TLT/IEF	0.95	55.1	−4.33	2.0	0.05	50	0.5
EEM/VWO	0.87	17.5	−0.38	2.5	0.05	50	0.0
EEM/VWO	0.87	22.1	−0.50	2.5	0.05	60	0.0

Table 3: Key performance metrics for the first run of TLT/IEF from Table 2.

Metric	**Value**
Sharpe ratio (total)	1.04
Sharpe ratio (invested)	1.37
Percentage of days invested	56.76 %
Maximum drawdown (strategy)	–4.33 %
Total return (strategy)	5.52 %
Total return (50-50 benchmark)	–4.62 %
Sharpe ratio (50-50 benchmark)	0.00
Maximum drawdown (50-50 bench.)	–0.50 %

Figure 1 shows the Sharpe ratio of the best-performing run for each pair of ETFs. The left panel varies the entry and exit thresholds, while the right panel varies the rolling window size and exit threshold. One might recognize some of the best runs from the results in Table 2 as some of cells in the heatmaps.

[3]The full set of ETF pairs screened consists of EEM / VWO, XLF / XLU, IWM / QQQ, XLK / XLY, TLT / IEF, EWJ / EZU, XLE / XLK, SHY / TLT, GDX / GLD, MTUM / SPLV, USO / IEF. All series are sampled from 01 Jan 2018 through 27 Jul 2025.

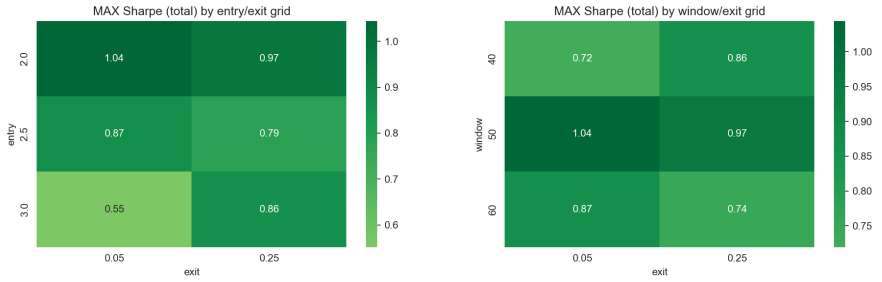

Figure 1: Sharpe ratio of the best-performing run for each pair of ETFs. Left: entry vs. exit thresholds. Right: window size vs. exit threshold.

5.2 Case study: `TLT/IEF`

We will now analyse the pair of ETFs `TLT` and `IEF`. `TLT` is the *iShares 20+ Year Treasury Bond ETF* and `IEF` is the *iShares 7-10 Year Treasury Bond ETF*. The back-test was run with a rolling window of $w = 50$ days, an entry threshold of $z_{\text{entry}} = 2.0$, and an exit threshold of $z_{\text{exit}} = 0.05$. The absolute residual filter was set to $\varepsilon_{\min} = 0$, and a stop-loss of 5% was applied. The data ranges from 01 Jan 2018 to 27 Jul 2025 in 1 day intervals.
Figure 2 shows the trades executed by the strategy. In the top panel the prices of the two ETFs are displayed and in the bottom panel z-score of the residuals is shown. The horizontal lines indicate the entry and exit thresholds. The strategy enters a position when the z-score crosses the entry threshold and exits when it crosses the exit threshold or the stop-loss is triggered.

5.3 ADF as predictor of Sharpe ratio

From the theory described in section 1.5 one could expect that the ADF test statistic is a good predictor of the Sharpe ratio of the strategy. Figure 3 shows the ADF test statistic against the total Sharpe ratio for each pair of ETFs on their best set of parameters (by total Sharpe ratio). But it is apparent that there is no reason to believe a more negative ADF statistic implies a higher Sharpe ratio. The correlation coefficient is approximately 0.544 which would suggest a moderate positive relationship between the two variables. A similar picture emerges when looking at the Sharpe ratio during invested days or the difference of Sharpe ratio of the strategy and the 50-50 benchmark. A reason for this could be that the ADF test statistic is not a good predictor of the stationarity of the residuals, but rather just a measure of how far the residuals are from being stationary.

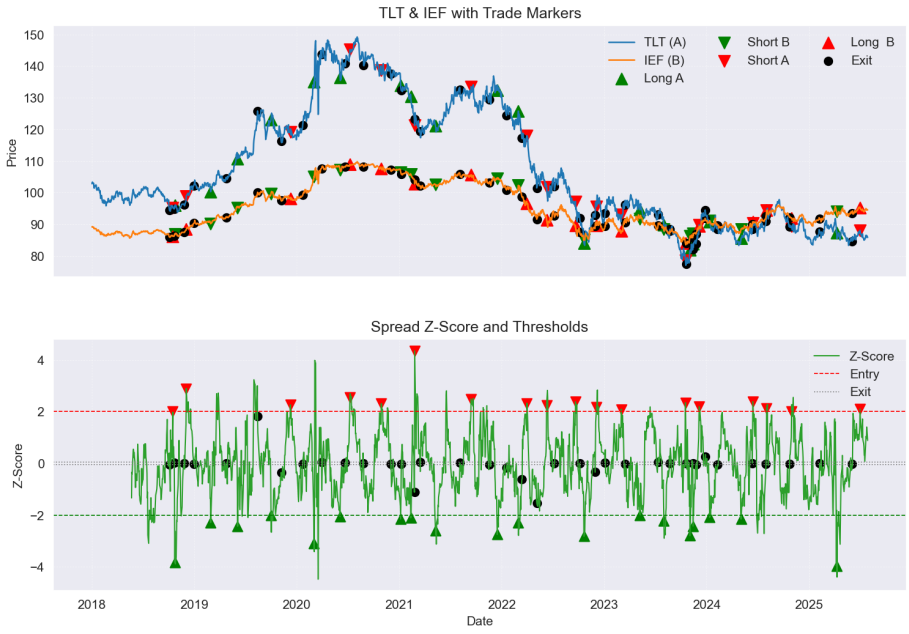

Figure 2: Trades executed by the strategy on the pair of ETFs TLT/IEF. Top: prices of the two ETFs. Bottom: z-score of the residuals with entry and exit thresholds.

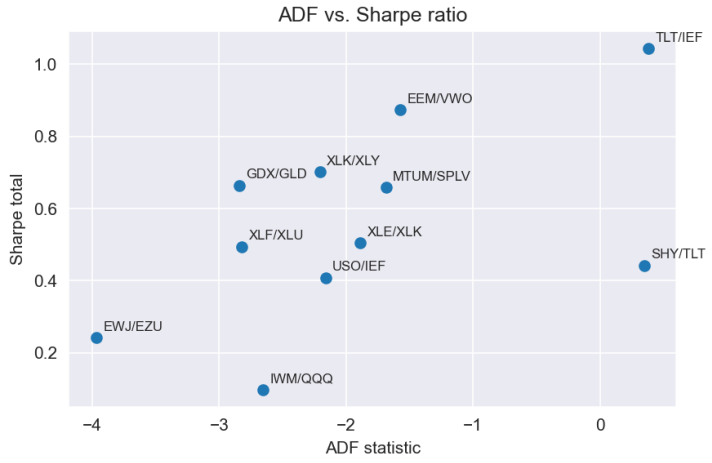

Figure 3: ADF test statistic vs. total Sharpe ratio for each pair of ETFs on their best set of parameters.

6 Conclusion

We observed that the cointegration pairs trading strategy can be successfully applied to pairs of ETFs, yielding Sharpe ratios significantly larger than the naive 50-50 benchmark. Also, the maximum draw-down and the percentage of time invested could be reduced compared to the benchmark. The best performing parameters vary slightly between pairs, but a rolling window of 50 days, an entry threshold of 2.0, and an exit threshold of 0.05 seem to be a good choice. The ADF test statistic does not seem to be a good predictor of the strategy's Sharpe ratio. In contrast to the ETFs, the strategy did not yield satisfactory results on pairs of cryptocurrencies, which is likely due to the lack of a strong economic relationship between their prices, as the most successful pairs of ETFs were very similar in their underlying assets. Transaction costs were considered in the back-test.

Even though the Sharpe ratios are significantly better than the naive benchmark, the absolute returns are modest given the long time span of several years. But a strength of this strategy is that profits can be made independent of overall market developments, based only on the relative developments of the given assets.

Future work could focus on (i) exploring higher-frequency data, (ii) using more sophisticated entry and exit mechanisms as well as softer windowing, (iii) testing the strategy on other asset classes and (iv) investigating different tests for stationarity of the residuals as predictor for strategy success.

A Algorithm pseudo-code

Algorithm 1 Cointegration Pairs Trading Loop

Require: price series (P_t^A, P_t^B), window w, thresholds $z_{\text{entry}}, z_{\text{exit}}$, stop-loss s_{SL}

for $t = w, w+1, \ldots, T$ **do**
  Estimate hedge ratio ▷ OLS on the last w observations

$$(\alpha_t, \beta_t) \leftarrow \text{OLS}\big(P^A_{t-(w-1):t} \text{ on } P^B_{t-(w-1):t}\big)$$

  Compute spread and z-score

$$\varepsilon_t = P_t^A - (\alpha_t + \beta_t P_t^B), \quad z_t = (\varepsilon_t - \mu_t)/\sigma_t$$

  if position $= 0$ **then**
    if $z_t < -z_{\text{entry}}$ **then**
      long A, short $\beta_t B$
    else if $z_t > z_{\text{entry}}$ **then**
      short A, long $\beta_t B$
    end if
  else
    if $|z_t| < z_{\text{exit}}$ **or** loss $> s_{\text{SL}}$ **then**
      close position
    end if
  end if
  record PnL and update current position
end for

B Additional figures

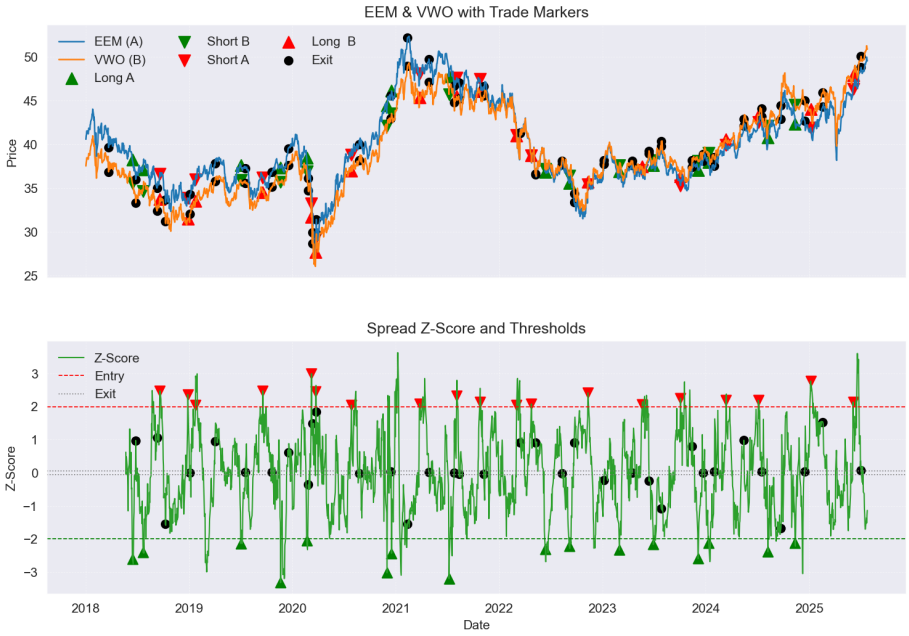

Figure 4: Trades executed by the strategy on the pair of ETFs EEM/VWO. Top: prices of the two ETFs. Bottom: z-score of the residuals with entry and exit thresholds.

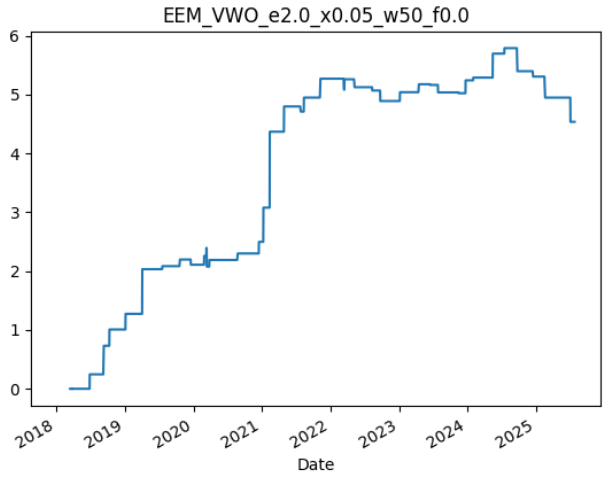

Figure 5: Absolute returns (USD) of the strategy on the pair of ETFs EEM/VWO with parameters $z_{\text{entry}} = 2.0$, $z_{\text{exit}} = 0.05$, $w = 50$, and $\varepsilon_{\min} = 0$. Capital-at-risk is 108 USD.

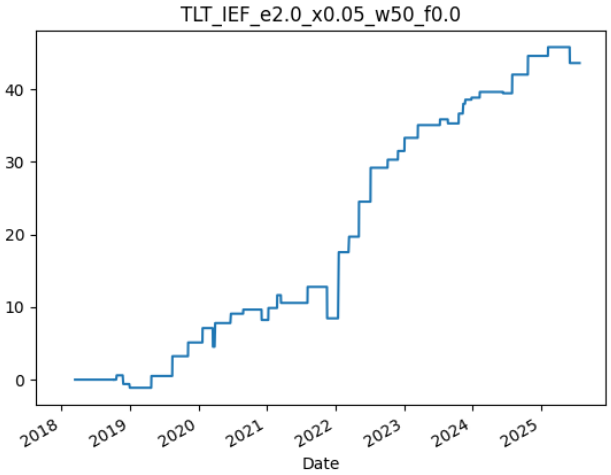

Figure 6: Absolute returns (USD) of the strategy on the pair of ETFs TLT/IEF with parameters $z_{\text{entry}} = 2.0$, $z_{\text{exit}} = 0.05$, $w = 50$, and $\varepsilon_{\min} = 0$. Capital-at-risk is 706 USD.

References

[1] Ganapathy Vidyamurthy. *Pairs Trading: Quantitative Methods and Analysis*. John Wiley & Sons, 2004. ISBN: 978-0471460671. URL: https://www.wiley.com/en-us/Pairs%2BTrading%3A%2BQuantitative%2BMethods%2Band%2BAnalysis-p-9780471460671.

[2] Evan Gatev, William N. Goetzmann, and K. Geert Rouwenhorst. "Pairs Trading: Performance of a Relative-Value Arbitrage Rule". In: *The Review of Financial Studies* 19.3 (2006). DOI: 10.1093/rfs/hhj020. URL: https://doi.org/10.1093/rfs/hhj020.

[3] David A. Dickey and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root". In: *Journal of the American Statistical Association* 74.366 (1979), pp. 427–431. ISSN: 01621459, 1537274X. URL: http://www.jstor.org/stable/2286348 (visited on 07/28/2025).

[4] James G. MacKinnon. *Critical Values for Cointegration Tests*. Queen's Economics Department Working Paper 1227. Last accessed 2025-07-28. Kingston, Ontario: Queen's University, Department of Economics, 2010. URL: https://econpapers.repec.org/paper/qedwpaper/1227.htm.

[5] *Yahoo Finance API*. URL: https://finance.yahoo.com/ (accessed 2025-07-28).

[6] *Binance Exchange*. URL: https://www.binance.com/ (accessed 2025-07-28).

[7] Igor Kroitor and contributors. *CCXT: CryptoCurrency eXchange Trading Library*. https://github.com/ccxt/ccxt. Version 4.x, accessed 28 July 2025. 2025.

[8] The pandas development team. *pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.