Predicting LINK Returns: A Data-Driven Approach to Short-Term Crypto Trading

Project Class Asset Management - Ulm University

Erich Gozebina, Daria Palitzsch, Laura Gullicksen

July 8, 2025

1. Introduction

Cryptocurrency markets are known for their extreme volatility, behavioral trading dynamics, and weak-form inefficiencies. Unlike traditional financial markets, they operate 24/7, are highly speculative, and are heavily influenced by sentiment, momentum, and macro-crypto developments. This environment creates both substantial risk and potential reward, especially for strategies that can extract signals from the surrounding noise.

In this report, we design and evaluate a short-term trading strategy for Chainlink's native token LINK, a top-25 cryptocurrency that serves as a compensation token in a decentralized oracle network within the DeFi ecosystem. LINK stands out as a compelling asset due to its high liquidity, large market cap, and fundamental utility, which make it less susceptible to manipulation and more likely to exhibit predictable price dynamics. Our objective is to develop a predictive model for daily returns based on economically interpretable variables, with a focus on 7-day momentum, technical indicators, and macro-crypto factors such as Bitcoin return. We implement this model using linear regression and the LASSO shrinkage method, allowing us to balance interpretability with robustness by filtering out noisy predictors. Zaremba et al. (2021), for example, found that short-term momentum effects are more pronounced in large, liquid cryptocurrencies, supporting the suitability of LINK for a trend-based predictive strategy. Its historical trading behavior further confirms the presence of exploitable short-term return persistence. To assess the practical viability of the strategy, we apply a backtest using daily data from October 2017 to April 2024. We incorporate realistic assumptions, including transaction costs, and benchmark the strategy's performance against both buy-and-hold and naïve momentum approaches. Our contribution lies in demonstrating that even in a highly volatile and seemingly chaotic market like crypto, structured, data-driven strategies can identify persistent short-term patterns. By combining economic intuition with statistical learning, we provide a transparent and replicable framework for active management in digital asset markets.

2. Data & Descriptive Analysis

Data Aggregation and Strategy Frequency

For our project we chose a dataset that provides price data of Chainlink's native LINK token at hourly frequency. While such high-frequency data offers more granular insights, we chose to aggregate the data to daily frequency for the following reasons:

1. **Alignment with Trading Strategy**: Our core trading strategy is based on a 7-day momentum signal, which inherently reflects weekly price trends. Applying such a signal at an hourly resolution would not be consistent with the strategy's time horizon.

2. **Noise Reduction**: Hourly crypto data can be highly volatile and noisy. Aggregating to daily returns reduces microstructure noise, short-term reversals, and Whale-driven price spikes, improving the signal-to-noise ratio.

3. **Practical Execution Perspective**: A strategy that rebalances daily is more realistic to implement, considering gas fees, latency and operational constraints on decentralized exchanges or CEX APIs.

4. **Interpretability and Robustness**: Daily returns are more interpretable and robust across backtests. Most financial and technical indicators (e.g., RSI, MACD, SMA) are commonly applied on daily charts.

Our base strategy issues long/short signals based on the past 7-day log return of LINK, i.e.,

$$\text{Momentum}_t^{(7)} = \log\left(\frac{P_t}{P_{t-7}}\right)$$

This naturally assumes daily data, as each observation reflects the cumulative return over the previous seven days.

In summary, aggregating to daily frequency is a theoretically and practically sound choice. It ensures consistency between our signal construction, model estimation, and backtesting logic.

To illustrate the effects of our aggregation decision and provide initial insights into the behavior of LINK prices, we present two visualizations below:

Figure 1 shows two plots:

- The **upper panel** displays the daily closing price of LINK in USD since 2018, highlighting the asset's strong bull run during the 2020–2021 crypto boom and its subsequent volatility.

- The **lower panel** shows the corresponding daily log returns, capturing the day-to-day fluctuations in price.

Having a first glance on the LINK's price data, we see that there is not much movement until 2020, followed by sharp increases by a factor of ten in the consecutive nine months. The price reaches its peak of over 50\$/LINK in May 2021. In the second half of 2021 until April 2022 the data shows volatile behavior but with significant decrease on average. A trendless
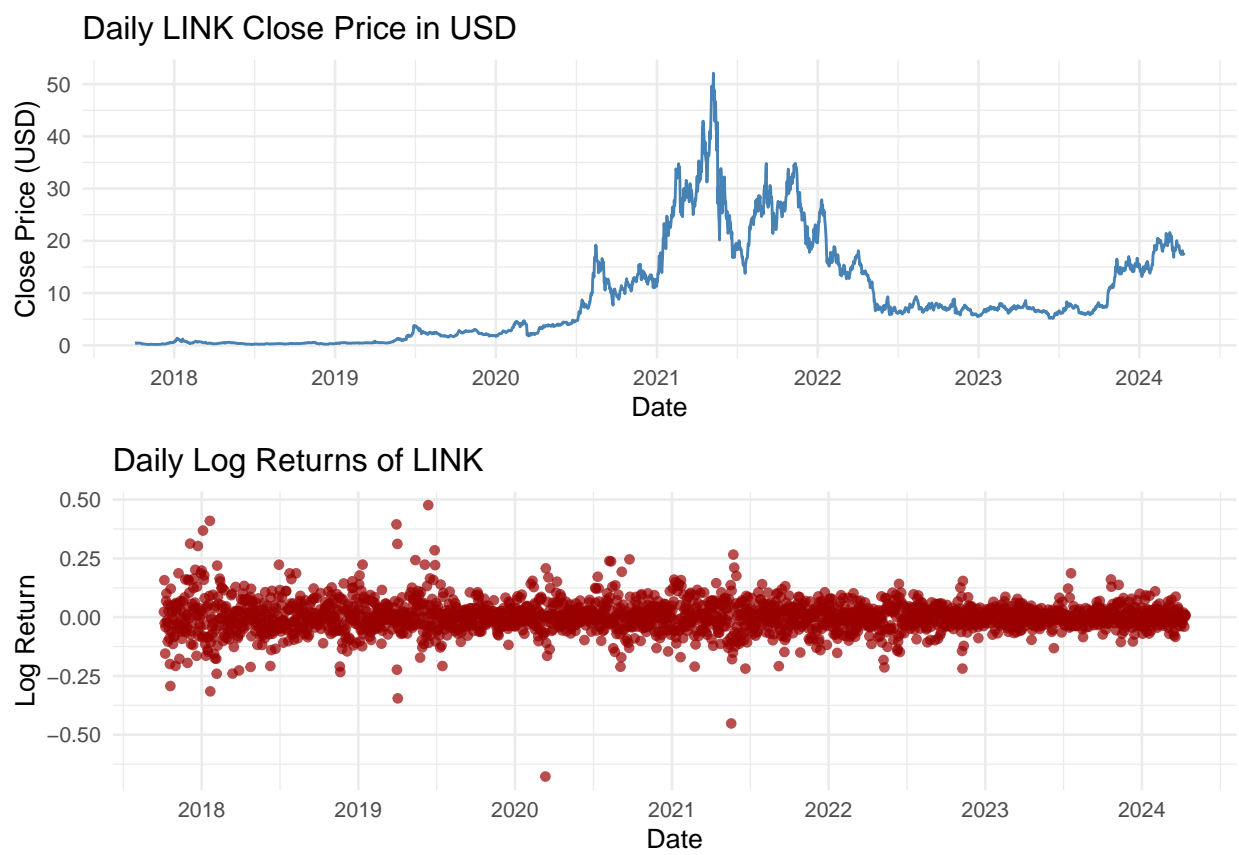
Figure 1: Close Price and Log Return of LINK

period of relatively low volatility, starting in May 2022 and ending in October 2023, shows prices between 5 and 10$/LINK. Finally, we observe another increase at the end of 2023 and beginning of 2024.

Moving away from non-stationary price data to stationary log returns, we observe a higher dispersion in the earlier years, suggesting a higher volatility then. Furthermore, most of points cluster around zero, indicating that there is no significant long term drift. Outliers, both positive and negative, imply extreme relative price movements, especially in the earlier period. Another important insight is the changing variance since the density of the points increases in the second half of the time window.

Why did we chose log returns over canonical (arithmetic) returns? Using log returns instead of canonical returns is a standard practice in financial econometrics and modeling.

$$\tilde{r}_t = \log(r_t + 1) = \log\left(\frac{p_t}{p_{t-1}}\right)$$

The underlying reason is the assumption that prices of an financial asset are log-normally distributed. This is reasonable since the log-normal distribution does not allow for negative values, which is also true for most asset prices (particularly for cryptocurrencies). Moreover, historical data provides evidence that the log-normal distribution gives a good fit for the prices of many financial assets. In reverse, since the logarithm function amplifies returns that are close to -1 more than positive returns, log-returns are distributed more symmetrically than canonical returns and indeed follow a normal distribution. Additionally, if returns are small, log returns approximate canonical returns very well. For x close to zero, it holds that

$$\log(x + 1) \approx x.$$

Daily returns are small, justifying log-return approximations. Another important property is the additivity of log returns. It allows us to aggregate returns over multiple periods by summing up the pointwise log returns - a property that canonical returns miss. These properties make log returns more suitable for linear regression models, hypothesis testing, and machine learning regressors.

To better understand the characteristics of the LINK price and return time series, we present a set of descriptive statistics based on the daily close prices and the corresponding log returns. These statistics provide a first impression of the dataset's distribution, dispersion, and extreme values, and help assess whether further preprocessing or transformation steps are necessary before applying predictive models.

The summary statistics in Table 1 reveal that our data set contains data of 2,383 days. Furthermore, we observe a mean of the daily close price of $9.1484 and a standard deviation of $9.5407. Due to the relatively high standard deviation, the computed mean loses meaningfulness regarding its predictive power. The large gap between minimum and maximum close prices undermines this hypothesis. Turning to log returns, we find that the mean daily log return of LINK is close to zero but still positive, yielding an upward trend within the

4

Table 1: Summary Statistics for LINK Price and Returns

Statistic	Value
Number of Observations	2,383.0000
Mean Close Price	9.1484
Std. Dev. Close Price	9.5407
Minimum Close Price	0.1453
Maximum Close Price	52.1000
Mean Return	0.0016
Std. Dev. Return	0.0676
Minimum Return	-0.6776
Maximum Return	0.4762

whole time frame. As for the close prices, the standard deviation of returns is relatively high, reflecting once more the well-known volatility of cryptocurrency markets. The minimum and maximum returns further highlight the presence of large price swings, showing high potential for short-term profits, but also the riskiness of the asset over the observation period.

To evaluate the temporal dependence structure of LINK's daily log returns, we analyze the autocorrelation function (ACF) shown in Figure 2. Autocorrelation aims to identify repeating patterns that may help us to explain our prediction results.

But, the ACF reveals no statistically significant linear dependence across the first 33 lags, suggesting that past returns do not linearly predict future returns—consistent with the weak-form Efficient Market Hypothesis (EMH). We conclude that there are no hidden periodicities within a 33-days time window. Nevertheless, this result does not preclude the existence of non-linear or directional patterns. Motivated by this, we proceed with a momentum-based trading strategy that exploits the sign of multi-day past returns to generate long or short signals.
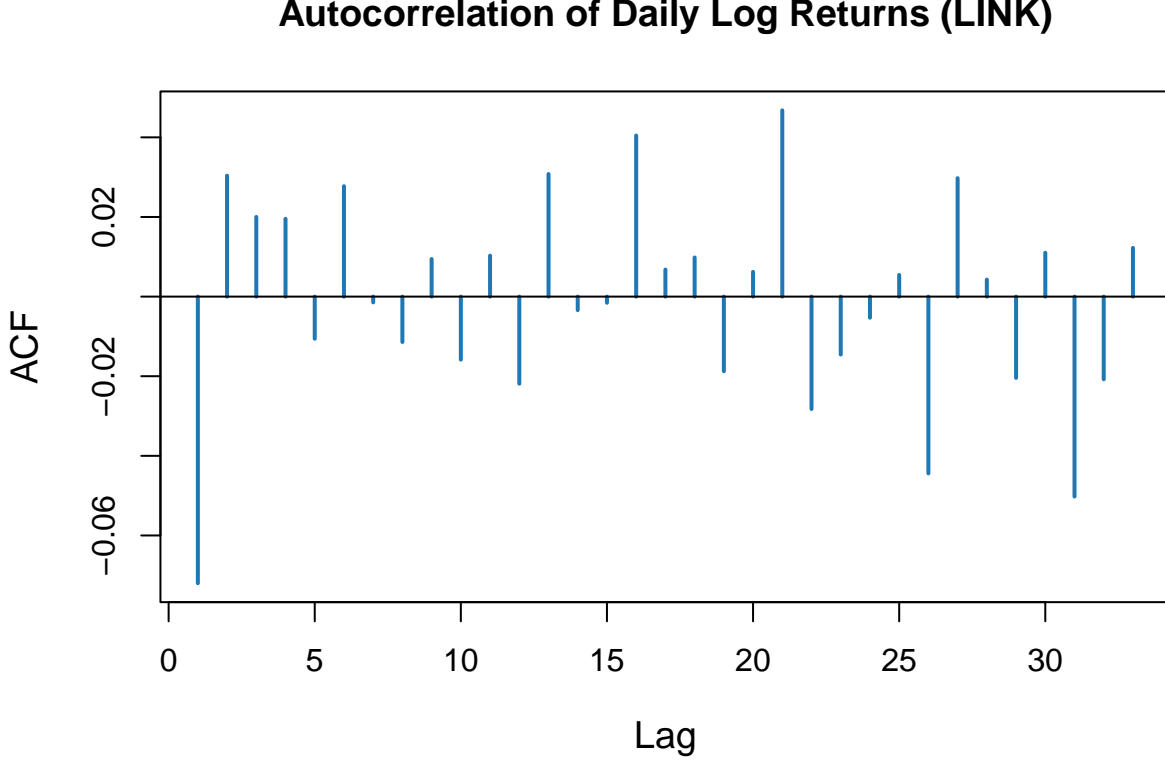
Autocorrelation of Daily Log Returns (LINK)

Figure 2: ACF of Daily Log Returns for LINK

3. Standard Model

We develop a basic model that will serve as a starting point for an extended model. Our first approach to predict future returns is a simple linear regression on one single feature - the 7-day momentum. Why linear regression? First, this technique is the underlying mechanism of many advanced models that are often generalizations of the linear case. Therefore, it is a good fit for a starting point. In general, linear regression aims to identify linear relationships between input data and the target dimension. In our case the input data is price data, trading volume, market capitalization, and every predictor that is derived from those - returns for instance. The target dimension that we are going to predict is the return of the next day. The simplicity of linear mappings make results easy to interpret, whereby the model still remains powerful since many observed relationships are indeed of linear nature. Moreover, linear regression indicates the strength of those linear ties what makes it a helpful tool for decision making.

7-Day Momentum Signal Strategy

To quantify short-term trends in LINK's price, we construct a 7-day momentum signal defined as the logarithmic return over the past seven trading days:

$$\text{Momentum}_t^{(7)} = \log\left(\frac{P_t}{P_{t-7}}\right)$$

6

This momentum measure captures the cumulative price movement over one week and serves as a central predictive feature in our empirical analysis.

In parallel, we define the **target return** as the one-day-ahead log return:

$$r_{t+1} = \log\left(\frac{P_{t+1}}{P_t}\right)$$

This simplistic setup aligns with our objective of forecasting next-day returns based on a recent trend.

Baseline Predictive Model: 7d-Momentum-Only

Table 2: Regression Results: 7-Day Momentum Strategy

	Dependent variable:
	Strategy Return
Intercept	0.0022
	(0.0088)
7-Day Momentum	0.0013
	(0.0017)
Observations	1,900
R^2	0.00003
Adjusted R^2	-0.0005
Residual Std. Error	0.0731 (df = 1898)
F Statistic	0.0605 (df = 1; 1898)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 2 presents the results of our OLS regression using 7-day momentum as a predictor of daily strategy returns. While the estimated coefficient for 7-day momentum is positive (0.0013), suggesting that higher past-week returns are associated with slightly higher next-day returns, the effect is not statistically significant at conventional levels. The intercept is also insignificant, indicating no meaningful average return when momentum is neutral. The explanatory power of the model is extremely low, meaning that it performs no better, and in fact slightly worse, than simply predicting the mean return. The F-statistic is also not significant, implying that the model lacks predictive validity in its current form. Additionally, the p-value for the slope clearly exceeds 0.05 (the threshold, which is commonly used for acceptance), we can not conclude a linear relation between the 7-day momentum and the future day return.

While the 7-day momentum coefficient is directionally consistent with momentum theory (positive trend continuation), the statistical insignificance and poor model fit highlight a

key limitation: momentum alone is not enough to predict returns in such a volatile and noisy market. This result is aligned with the autocorrelation analysis (Figure 2), which showed no significant linear dependence between the log return and its 7-day lag. The underperformance of the current model reinforces the need for extending the model to include additional explanatory features and employing a Lasso regression to avoid overfitting and improve predictive power. Although the positive sign of the momentum coefficient aligns with the underlying economic intuition, the weak performance highlights the need for a more comprehensive approach.

4. Extension

Extension of our Linear Regression Model

To enhance the predictive power of the benchmark model, we extend it by incorporating a broader set of explanatory variables that capture not only short- and medium-term price dynamics, but also market sentiment, technical indicators, and inter-asset relationships. These include:

- Momentum indicators over 3, 7, and 14 days,

- Lagged daily returns (1-day and 2-day),

- A 7-day rolling volatility measure,

- Technical indicators such as the 14-day Relative Strength Index (RSI), MACD value and histogram, Simple Moving Average difference, and Average True Range (ATR),

- Day-of-week dummy variables to capture potential calendar effects,

- BTC-based predictors: daily BTC return, 7-day BTC momentum, and 7-day BTC volatility,

- ETH-based predictors: daily ETH return, 7-day ETH momentum, and 7-day ETH volatility,

- ETH trading volume: daily ETH volume return, 7-day ETH volume momentum, and 7-day ETH volume volatility,

- ETH market capitalization: daily ETH market capitalization return, 7-day ETH market capitalization momentum, and 7-day ETH market capitalization volatility,

- Ethereum gas fees: daily gas return, 7-day gas momentum, and 7-day gas volatility.

For the calculation of the technical indicators we reference Appendix A. The extended predictive regression model is specified as:

$$
r_{t+1} = \alpha + \sum_{h \in \{3,7,14\}} \beta_h \cdot \text{Momentum}_t^{(h)} + \gamma_1 \cdot r_t + \gamma_2 \cdot r_{t-1} + \delta \cdot \text{Volatility}_t^{(7)} + \sum_j \theta_j \cdot X_t^{(j)} + \varepsilon_{t+1}
$$

where $X_t^{(j)}$ represents the set of technical indicators (RSI, MACD, ATR, SMA), weekday dummies, ETH-based and BTC-based predictors.

$$r_{t+1} := \log\left(\frac{P_{t+1}}{P_t}\right) \quad \text{(one-day-ahead LINK return)}$$

$$\text{Momentum}_t^{(h)} := \log\left(\frac{P_t}{P_{t-h}}\right) \quad \text{for } h \in \{3, 7, 14\}$$

$$\text{Volatility}_t^{(7)} := \text{std}\left(r_{t-6}, \ldots, r_t\right)$$

$$\text{BTC return}_t := \log\left(\frac{P_t^{\text{BTC}}}{P_{t-1}^{\text{BTC}}}\right)$$

$$\text{BTC Momentum}_t^{(7)} := \log\left(\frac{P_t^{\text{BTC}}}{P_{t-7}^{\text{BTC}}}\right)$$

$$\text{BTC Volatility}_t^{(7)} := \text{std}\left(r_{t-6}^{\text{BTC}}, \ldots, r_t^{\text{BTC}}\right)$$

The ETH-based predictors are constructed analogously to the BTC-based predictors. In a next step we perform a linear regression on all of these features. Therefore, the parametrization of this model is estimated via Ordinary Least Squares (OLS) on the in-sample period. By incorporating this rich feature set, we aim to capture a range of return drivers including price trends, market overreaction, volatility clustering, inter-market dependencies, and behavioral biases tied to trading weekdays. Note that the model is trained on a training set, the in-sample period, which includes the first 80% of available LINK data. This allows us to backtest the model in an out-of-sample manner in a later step.

Table 3 presents the output of our extended OLS regression model, where the dependent variable is the daily log return of LINK. This model includes not only LINK-specific momentum indicators but also technical indicators, calendar dummies, and a broad set of macro-crypto variables derived from Bitcoin (BTC) and Ethereum (ETH). The purpose of this model is to evaluate whether combining crypto-specific and cross-asset features can improve return predictability over the basic 7-day momentum model (as seen in Table 2). Several predictors emerge as statistically significant. The strongest within-asset signals are the one-day and two-day lagged LINK returns, both of which have positive and significant coefficients (0.0796 and 0.0920), indicating short-term positive autocorrelation. Interestingly, the coefficient for 3-day momentum is significantly negative, suggesting some evidence of short-term reversal over that horizon. In contrast, longer momentum terms (7-day and 14-day) are negative but not statistically significant.

From a cross-asset perspective, the results reveal that BTC returns are negatively associated with LINK returns (-1.0111), pointing to potential substitution effects or temporary decoupling between the two assets. This negative correlation seems economically reasonable, since major bull or bear runs in the dominant cryptocurreny BTC appeal traders to readjust portfolio composition, heavily impacting the prices of altcoins. Similarly, BTC volatility has a negative and significant impact, implying that increased uncertainty in BTC reduces returns for LINK. A similar pattern is observed for ETH returns, which also carry a negative and marginally significant coefficient (-1.0707). Notably, the interaction between ETH momentum and market cap is positive and significant, suggesting that when ETH trends upward in conjunction with growing market cap, it positively influences LINK performance.

Table 3: Extended Regression Model: Predicting LINK Returns with Crypto Features

	Target Return
Intercept	-0.0292\(0.0180)
Momentum (3d)	-0.0816*\(0.0339)
Momentum (7d)	-0.0156\(0.0194)
Momentum (14d)	-0.0192\(0.0175)
Lagged Return (1d)	0.0796*\(0.0331)
Lagged Return (2d)	0.0920*\(0.0360)
Volatility (7d)	-0.0710\(0.0548)
RSI (14)	0.0004\(0.0003)
SMA Diff	-0.0010\(0.0018)
MACD Value	0.0004\(0.0005)
MACD Histogram	0.0011\(0.0017)
ATR (14)	-0.0020\(0.0014)
Monday	0.0063\(0.0048)
Tuesday	-0.0010\(0.0045)
Wednesday	0.0014\(0.0045)
Thursday	0.0041\(0.0045)
Friday	0.0079.\(0.0045)
BTC Return	-1.0111***\(0.0374)
BTC Momentum (7d)	-0.0040\(0.0143)
BTC Volatility (7d)	-0.1576*\(0.0744)
ETH Return	-1.0700\(7.7903)
ETH Momentum (7d)	-6.6440\(4.1274)
ETH Volatility (7d)	23.1704.\(13.8990)
ETH Volume	0.0016\(0.0071)
ETH Momentum (7d) × Volume	0.0044\(0.0050)
ETH Volatility (7d) × Volume	0.0262.\(0.0157)
ETH Market Cap	1.0881\(7.7917)
ETH Momentum (7d) × Market Cap	6.6656\(4.1276)
ETH Volatility (7d) × Market Cap	-22.9841.\(13.8968)
ETH Gas	0.0030\(0.0053)
ETH Momentum (7d) × Gas	-0.0016\(0.0036)
ETH Volatility (7d) × Gas	-0.0031\(0.0067)

Significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Model statistics: $R^2 = 0.342$, $R^2_{adj} = 0.331$, F statistic = 30.97.

The overall model fit improves substantially compared to the single momentum specification: the R^2 increases to 0.342, and the adjusted R^2 to 0.331, indicating that over 33% of the

variation of LINK returns is now explained. The model's F-statistic (30.97) confirms the joint significance of the regressors.

However, the extended OLS regression model considers all features equivalently, yielding poor interpretability. Furthermore, having equal weights is a drawback, because there might exist highly correlated features that sum up to an overweighted influence on the prediction. Similarly, many non-significant features are taken into account, leading to undesired high variance. A solution to these downsides are variable selection and regularization techniques, which we consider in the next section.

LASSO Model

To prevent overfitting and perform automatic variable selection, we extend our linear modeling approach using the Least Absolute Shrinkage and Selection Operator (LASSO). The LASSO adds a penalty term to the standard OLS loss function, shrinking some coefficient estimates toward zero. This results in a sparse model that may improve predictive performance, particularly when dealing with multiple correlated predictors. Furthermore, reducing the number of relevant features allows for a better interpretation of simulation results.

The LASSO estimator is defined as the solution to the following optimization problem:

$$\hat{\beta}^{\text{LASSO}} = \arg\min_{\beta_0,\beta} \left\{ \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

where:

- y_i is the target variable (e.g., one-day-ahead return),
- x_{ij} are the predictor variables,
- β_j are the coefficients,
- $\lambda \geq 0$ is the tuning parameter controlling the strength of the penalty.

As λ increases, more coefficients are shrunk toward zero. For $\lambda = 0$, the solution coincides with OLS.

We use 10-fold cross-validation to select the optimal λ that minimizes the mean squared prediction error on held-out data.

Optimal Lambda from Cross-Validation: $\lambda^* = 0.002780$

Given this optimal lambda, we now run the LASSO regression, including all variables from the previous linear regression:

The LASSO regression identified two non-zero predictors for explaining LINK returns:

1. Bitcoin Daily Return (Coefficient: -0.9552): This variable has a large and negative coefficient, indicating that when BTC's daily return increases by 1 unit (in our scaled units), the predicted return of our LINK-based trading strategy decreases by approximately 0.9552 units, all else equal. This suggests a strong inverse relationship between

Table 4: Non-Zero Coefficients from LASSO Regression

Predictor	Coefficient
Intercept	0.001414
BTC Daily Return	-0.955208
ETH 7-Day Volume Momentum	0.001887

BTC movements and our strategy, potentially due to hedging behavior or negative spillovers, as we have already detailed in the discussion of Table 3.

2. ETH 7-Day Volume Momentum (Coefficient: 0.0019): This predictor captures short-term trends in ETH's trading volume. The positive but small coefficient implies that higher recent momentum in ETH trading volume is weakly associated with increased LINK returns, possibly due to spillover effects from rising market activity in related tokens.

3. Intercept (Coefficient: 0.0014): The intercept represents the model's baseline prediction when all predictors are zero. Here, it suggests a small positive base return, though in practice this often has less interpretive value than the covariates.

Remarkably, the ETH 7-day volume momentum was not statistically significant in the extended OLS model. This points out a key strength of the LASSO technique. While p-values from the OLS model reflect only marginal significance, LASSO aims to minimize prediction error with a penalty that encourages sparsity, taking combined predictive contribution into account. Therefore, LASSO can identify relevant explanatory variables that are non-significant in their marginal influence and drop significant ones as well. We investigate whether reducing features is fruitful by providing a comprehensive backtesting analysis.

5. Forecasting & Backtesting

In-Sample testing

To evaluate the performance of our predictive models, we begin by conducting in-sample (IS) testing. This involves fitting each model on a fixed training sample and evaluating how well the model explains historical variation in the data.

We assess in-sample performance using the following criteria:

- **Mean Squared Error (MSE)**: Measures the average squared difference between predicted and actual returns.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

- **Coefficient of Determination - Adjusted R^2**: Indicates the proportion of variance explained by the model, adjusted for the number of predictors. It estimates the percentage of predictable variations in the dependent variable with respect to the explanatory variables.

$$R^2_{\text{adj}} = 1 - \frac{\text{RSS}/(n - p - 1)}{\text{TSS}/(n - 1)}$$

- **Directional Accuracy**: The fraction of times the predicted direction matches the actual direction of returns.

$$\text{Accuracy} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\left(\text{sign}(\hat{y}_i) = \text{sign}(y_i)\right)$$

Thereby, n denotes the respective sample size, p the number of predictors, RSS the residual sum of squares, TSS the total sum of squares, and $\mathbb{1}$ the indicator function. Hence, RSS is the sum of squared prediction errors, while TSS is the sum of squared differences between mean and predictions, which is proportional to the variance.

These metrics are computed for all three models:

1. Benchmark (7-day momentum only),

2. Extended linear model with multiple features,

3. LASSO-regularized regression with automatic feature selection.

Table 5 shows that the Extended and LASSO models have superior in-sample performance compared to the Benchmark model. The Benchmark model has the highest mean squared error (MSE = 0.0053) and the lowest directional accuracy (0.5051). This indicates poor predictive precision and a limited ability to correctly predict the direction of change. Furthermore, its negative adjusted R^2 (-0.05%) suggests that it fails to explain the variation in the dependent variable and performs worse than a simple mean model.

Table 5: In-Sample Performance of Benchmark, Extended, and LASSO Models

Model	MSE	Directional Accuracy	R^2_{adj}
Benchmark	0.0053	0.5051	-0.05%
Extended	0.0035	0.7110	33.10%
LASSO	0.0036	0.7142	32.29%

Compared with the Benchmark model, the Extended and Lasso models demonstrate substantially improved performance with notably lower MSEs (0.0035 and 0.0036, respectively) and significantly higher directional accuracy (0.7110 and 0.7142, respectively). Their adjusted R^2 values (33.10% for the Extended model and 32.29% for the Lasso model) suggest that both models meaningfully explain the variance in the data, though the Extended model has a slight advantage. These results confirm that the Extended and Lasso models offer stronger in-sample predictive capabilities than the Benchmark model.

Out-of-sample testing:

After performing the in-sample analysis, we now test our models out-of-sample (OOS) on the test data set:

Table 6: Out-of-Sample Model Evaluation

Model	MSE	Directional Accuracy	R^2_{OOS}
Benchmark	0.0014	0.5191	-0.00%
Extended	0.0012	0.7277	18.81%
Lasso	0.0010	0.7362	33.63%

The benchmark model fails to generalize out-of-sample ($R^2_{OOS} = 0$), confirming the weak predictive power of raw 7-day momentum. In contrast, the LASSO model improves out-of-sample accuracy substantially ($R^2_{OOS} = 0.34$), likely due to its ability to regularize noise and select relevant predictors.

To evaluate the trading performance of our predictive models beyond return forecasting, we implement a simple In-and-Out (I/O) strategy. This rule-based strategy enters the market ("IN") only when the model predicts strong positive returns — i.e., when the predicted value exceeds the 75th percentile of the respective model's prediction distribution. Otherwise, the model takes the money out of the market ("OUT").

Signal Construction:

We construct separate trading signals for:

- the Benchmark model (7-day momentum),

- the Extended Linear Regression model,

- the LASSO model, and

- the actual target returns (for reference).

The procedure:

- 1 (enter position) if the predicted return exceeds the threshold

- 0 (stay out) otherwise

This setup computes the daily strategy returns by multiplying the lagged signal with the actual log return. Additionally, we estimate the trading costs by applying a fixed fee per position change to simulate realistic trading costs. For the trading costs we consider the average Ethereum gas price per day and multiply it by 60,000, which roughly coincides with the factor for ERC-20 token transfer on the Ethereum blockchain.

This approach emphasizes selective participation in the market based on the model's confidence and allows us to assess whether models can not only predict returns but also generate economically meaningful signals.
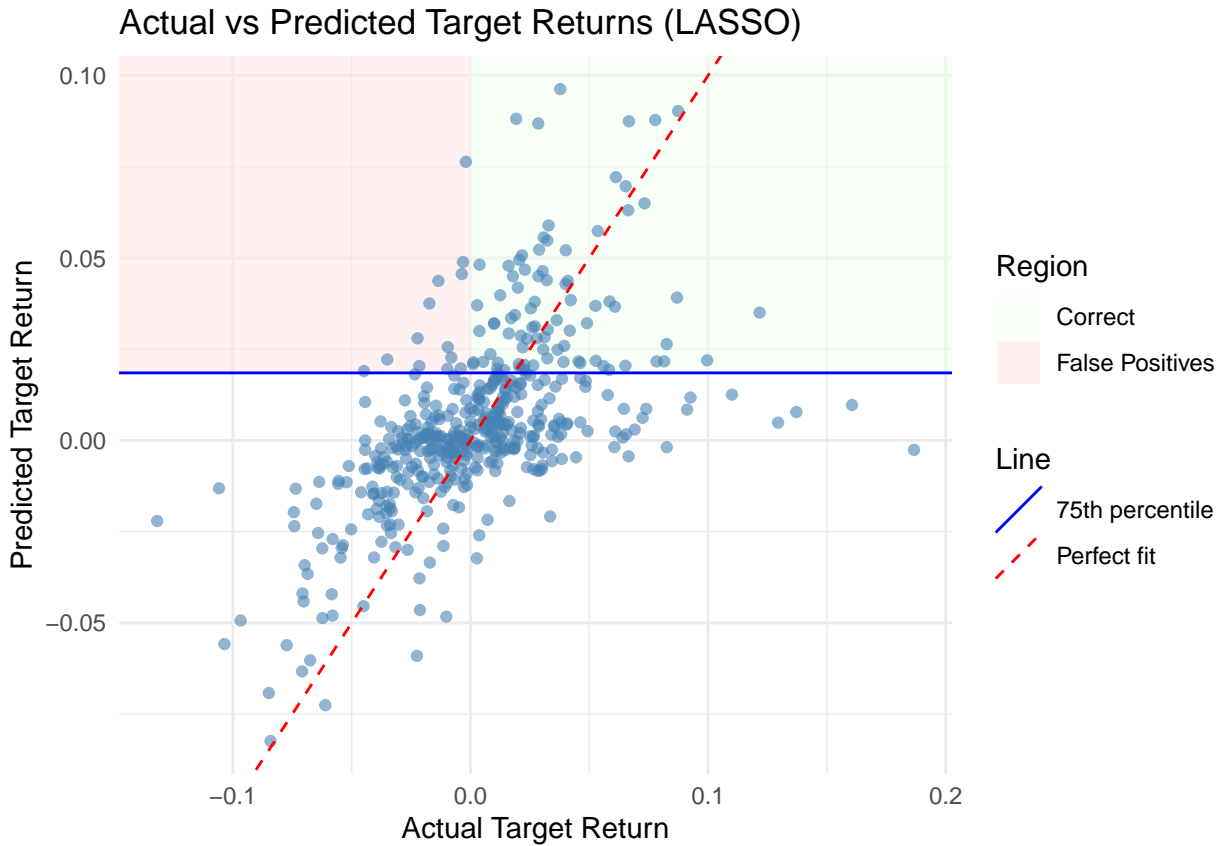

Figure 3: Actual vs Predicted Target Returns (LASSO)

Figure 3 visualizes the relationship between actual and predicted one-day-ahead returns from the LASSO regression model. Each point represents an observation in the test set, with the x-axis showing the realized return and the y-axis the model's predicted return. The dashed

red line indicates the 45° line of perfect predictions, while the blue horizontal line marks the 75th percentile of predicted returns.

We observe a moderate clustering of points around the diagonal, suggesting that the LASSO model captures some predictive structure in the return dynamics, particularly for returns near zero. However, deviations from the red line highlight prediction errors, especially in the tails.

The green region denotes instances where the predicted and actual returns have the same sign — i.e., the model correctly forecasts the return direction. These cases reflect successful directional predictions, which are critical for long/short trading strategies. In contrast, the red-shaded area highlights false positives, where the model incorrectly predicts a positive return, but the actual return is negative.

While the spread around the diagonal indicates that exact return magnitudes are not always accurately predicted, the high density in the green region supports the model's strong directional accuracy (see Table 6). This suggests that the LASSO model is well-suited for sign-based trading rules, even if its point predictions are noisy.
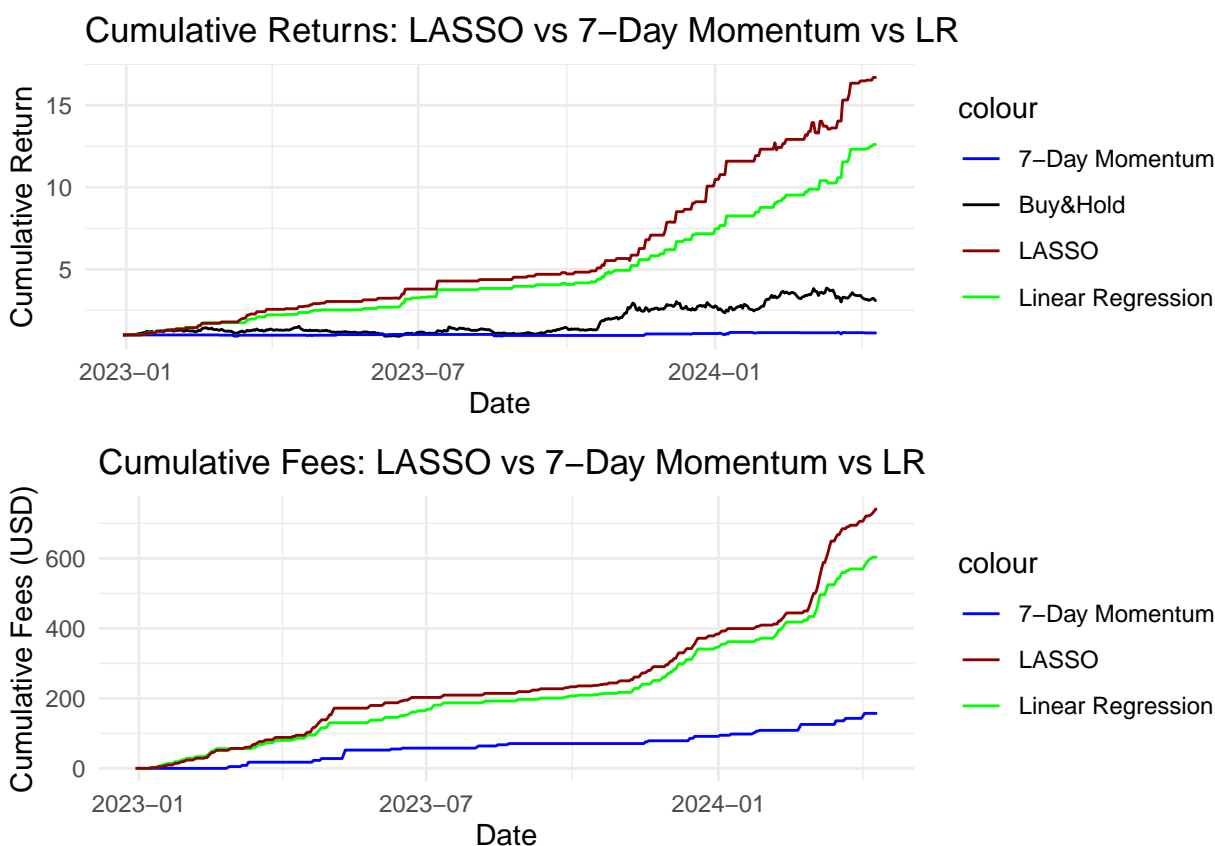

Figure 4: Cumulative Performance of Trading Strategies (Log Returns and Fees)

The **top panel** of Figure 4 compares the cumulative returns of four strategies:

- LASSO model (dark red) achieves the highest return, significantly outperforming all others.

- Linear Regression (green) performs closely behind LASSO.

- The 7-Day Momentum (blue) lags far behind, reflecting its weak signal quality.

- Buy-and-Hold (black) delivers minimal returns with no active decision-making.

The **bottom panel** shows cumulative trading fees incurred due to signal changes:

- Both LASSO and Linear Regression generate more trades and hence higher fees.

- Despite these higher transaction costs, their net performance (gross return - fees) still can be superior if the initial investment is chosen high enough.

- The 7-Day Momentum strategy results in fewer position switches and lower fees but offers little economic value due to poor returns.

Overall, the LASSO model demonstrates superior performance, not only in terms of return predictability but also as a trading signal generator. In practice, one must account for transaction costs as well.

Limitations

However, the model does face certain limitations when transitioning to real-world application. One key challenge lies in the assumption of perfect trade execution at the daily closing price. In highly volatile markets, the actual execution price often deviates from the expected one due to slippage, which can impact strategy performance. This effect may be further intensified in periods of low liquidity, where execution latency and order book depth become critical factors.

Additionally, the LASSO-based approach, by design, captures only linear relationships between LINK returns and the selected features. While this promotes interpretability and robustness, it also limits the model's ability to reflect the nonlinear dynamics that frequently characterize cryptocurrency markets. As a result, certain predictive patterns may remain unexplored.

These limitations, however, open promising avenues for further enhancement. Future iterations of the strategy could integrate execution-aware trading algorithms and explore nonlinear modeling techniques, such as deep neural networks or tree-based ensembles, to better capture complex market behavior. By addressing these challenges, the strategy could evolve into a more powerful and realistic tool for active crypto asset management.

6. Conclusion

Our results demonstrate that short-term predictability in the crypto space is possible using a structured, data-driven approach. We developed and backtested a daily trading strategy for LINK that issued "in" or "out" signals based on predicted next-day returns. The strategy enters the market only when the forecasted return exceeds the model-specific 75th percentile, thereby focusing on high-confidence signals. While a simple 7-day momentum signal proved insufficient, we enhanced performance by incorporating a broad feature set: momentum signals, volatility indicators, cross-asset spillovers, technical indicators, and calendar effects. Our final model uses Lasso regression to select only the most predictive variables and prevent overfitting. The Lasso model achieves the highest out-of-sample R^2 (33.6%) and directional accuracy (73.6%), outperforming both buy-and-hold and naïve momentum strategies. However, we made simplifying assumptions, such as perfect trade execution at the daily close price, which limits the model's applicability in real-world settings. In practice, slippage and latency can significantly impact returns, especially in volatile or illiquid markets. Additionally, the model captures only linear relationships, potentially missing complex market dynamics. These limitations emphasize opportunities for further enhancement, such as incorporating deep neural networks to better model nonlinear behavior.

A Appendix

14-Day Relative Strength Index (RSI)

The RSI is an index that takes values between 0 and 100. Its computation is derived from the relative strength

$$RS = \frac{\text{SMMA}(U, n)}{\text{SMMA}(D, n)}$$

$$\Rightarrow RSI = 100 - \frac{100}{1 - \text{RS}}$$

SMMA denotes the smoothed or modified moving average, which is an exponentially weighted average with weighting factor $1/n$. Moreover,

$$U = \text{close}_{t-n} - \text{close}_t$$

denoting an upward change and

$$D = \text{close}_t - \text{close}_{t-n}$$

a downward change respectively.

In our model n is set to 14. RSI is often associated with overbought and oversold conditions.

Moving Average Convergence/Divergence (MACD)

The MACD value is computed under the inclusion of exponentially weighted moving averages (EWMA). EWMA assigns exponentially decreasing weights to all past values. The speed of decline depends on the time constant (which we denote in the index) that indicates the number of time steps until the weighting factor reaches 63.2% of the initial value. MACD then is constructed by:

$$\text{MACD} = \text{EWMA}_9(\text{EWMA}_{12} - \text{EWMA}_{26}),$$

where $\text{EWMA}_n(x)$ denotes the moving average of x with a time constant of n.

From an economic perspective, the MACD value can reveal changes in the trend of a financial asset, since it captures differences between averages of long- and short-term periods. This makes it a strong predictor, especially for assets with infrequent trend changes.

Simple Moving Average (SMA) Difference

By the SMA Difference we denote the difference between a simple moving average of the close prices of the last 7 days and the last 14 days. Let SMA_n be the average close price of the last n days. Then

$$\text{SMA}_{Diff} = \text{SMA}_7 - \text{SMA}_{14}$$

The SMA Difference indicates short-term momentum relative to the longer-term trend. A positive value suggests recent prices are rising faster than the broader trend, potentially signaling bullish sentiment, while a negative value may indicate emerging bearish pressure.

Average True Range (ATR)

ATR measures market volatility by averaging the true range over a specified period. Let $\text{TR}_t = \max\{\text{High}_t - \text{Low}_t, |\text{High}_t - \text{Close}_{t-1}|, |\text{Low}_t - \text{Close}_{t-1}|\}$ be the true range on day t. Then the 14-day ATR is defined as:

$$\text{ATR}_{14} = \frac{1}{14} \sum_{i=0}^{13} \text{TR}_{t-i}$$

The ATR reflects how much an asset typically moves in a day, with higher values indicating greater volatility. It does not indicate price direction but helps assessing risk and adjusting position sizes or stop-loss levels accordingly.