AMProject\_Clean

Laura Gullicksen, Erich Gozebina, Daria Palitzsch

23/05/2025

## 1. Introduction

#TODO: describe data choice -> Laura

## 2. Data & Descriptive Analysis

**Data Aggregation and Strategy Frequency**

The raw dataset provides CHAINLINK price data at . While such high-frequency data offers more granular insights, we chose to for the following reasons:

The extended predictive regression model is specified as:

where represents the set of technical indicators (RSI, MACD, ATR, SMA), weekday dummies, and BTC-based predictors.

The ETH-based predictors are constructed analogously to the BTC-based predictors. 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.

#DONE: add ethereum data -> Erich

#TODO: generate nicer latex table output of the regression results -> Daria

#TODO: Description and interpretation of output -> Laura

**Lasso Model**

To prevent overfitting and perform automatic variable selection, we extend our linear modeling approach using the Lasso (Least Absolute Shrinkage and Selection Operator). 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.

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

where:

As increases, more coefficients are shrunk toward zero. For , the solution coincides with OLS.

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

#TODO: generate nicer output

#TODO: explain the results

## 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:

These metrics are computed for all three models:

#TODO: interpret results

**Out-of-sample testing:**

#TODO: review code, does not work at the moment

#TODO: Evaluate: o Sharpe ratio o Cumulative return o OOSR2 o Hit rate (how often you correctly predict direction)

#TODO: include Transaction fees as extra path

## 6. Conclusion