Project Report: Helformer Replication & Analysis of Trading Signal Robustness

Yong Cui

December 31, 2025

1 Executive Summary

This project replicates the Helformer model of Kehinde et al. (2025) to evaluate its efficacy in forecasting Bitcoin (BTC) price movements. The original study highlights the potential of combining Holt-Winters smoothing layer with Transformer and replacing the convention FFN layers with LSTM layers, reporting significant profitability.

I implemented the model to test the strategy under a rigorous, no-lookahead backtest environment (Jan 2023 – June 2024).

Key Finding: This replication observed a divergence between statistical fit and realizable trading performance. While the model replicated the high R^2 fit observed in the literature, the volatility-adjusted trading strategy (inclding the long short strategy) was found to be highly sensitive to transaction costs, yielding negative net returns in this specific backtest implementation.

2 Methodology guided by the Source Paper Kehinde et al. (2025)

2.1 Model Formulation

Data Acquisition & Preparation:

- Daily historical Close for Bitcoin (BTC-USD) was downloaded using the yfinance library, targeting a period from 2017-01-01 to 2024-06-30 as specified in the paper.
- No missing values found.
- The data was split chronologically into 80% for training and 20% for testing.
- Fix a lookback window $L = 30$. Two inputs to the model were made:
  1. **Price stream:** a sequence of scaled raw prices of shape $(L, 1)$.
  2. **Seasonal phase stream:** an integer phase index of shape $(L,)$ used to encode calendar seasonality.

  Phase is computed from an absolute-day count relative to an epoch date and reduced modulo $m = 365$:

  $$\text{phase}(t) \;=\; \big(\text{days}(t) - \text{days}(\text{epoch})\big) \bmod 365,$$

  so each day is assigned a stable seasonal coordinate aligned across the full sample.

- Instead of MinMax scaling, we use *constant scaling* to avoid additive shifts that can interact poorly with multiplicative seasonal components:

$$x_t = \frac{P_t}{S}$$

where

$$P_t \text{ is the closing price of BTC on day t and } S = \text{median}\{P_t : t \in \text{TRAIN}\}.$$

The scale constant S is saved to *price_scale_const.joblib* so that inference outputs can be deterministically inverted:

$$\widehat{P}_{t+1} = S\,\widehat{x}_{t+1}.$$

Training examples are created by sliding windows: each sample uses a window ending at day t and targets day $t+1$, producing tensors

$$X \in \mathbb{R}^{n \times L \times 1}, \qquad \Phi \in \mathbb{Z}^{n \times L}, \qquad y \in \mathbb{R}^{n \times 1}.$$

Helformer Model Implementation:

- The Helformer architecture was implemented using the TensorFlow2/Keras deep learning framework in Python.
- The model structure consists of one Helformer block. Inside this block, a Holt-Winters smoothing layer is connected to a MHA layer which is connected to an LSTM layer. We added another LSTM readout after this block.
- Key hyperparameters were found using Optuna according to the paper's specifications: 50 trials, the Optuna Pruner feature enabled and MSE as the primary objective to minimize.

Training& Evaluation:

- The model was trained on the first 80% of the downloaded data.
- After training, the model's R-squared score and MAE were evaluated on the test set (last 20% of the downloaded data).
- Visualizations were generated comparing the predicted Close prices against the actual Close prices between 2023-01-01 and 2024-06-30.

2.2 Trading Strategy (Volatility-Adjusted)

To translate forecasts into positions, I utilized a signal-to-noise ratio approach. The trading signal is derived from the z-score of the predicted return relative to realized volatility:

$$z_t = \frac{\hat{R}_{t+1}}{\sigma_t}$$

- **Long:** if $z_t > \tau$
- **Short:** if $z_t < -\tau$
- *(For this evaluation, $\tau = 0$ for directional testing as in the paper)*

3 Performance Analysis

3.1 Quantitative Results

Models	R^2	Tot. Return	Ann. Ret	Ann. Vol	MDD	log SR	SR
Replicated model	0.9836	-18.84%	-1.64%	39.88%	-45.76%	-0.24	-0.04
Original Paper	1	**+925.29%**	–	**1.78%**	**≈0%**	**18.06**	–

Table 1: Replication performance (0.1% fees included) vs. Reported metrics using the same long-short strategy

3.2 Graphic Results

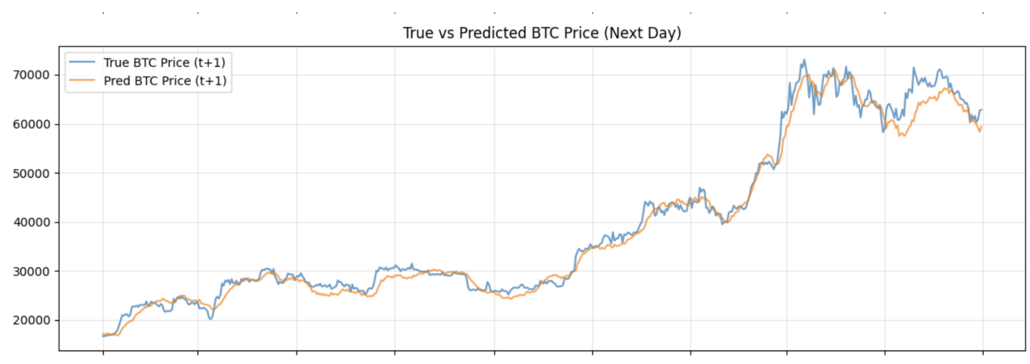

Figure 1: True vs. Predicted BTC Price by log return model

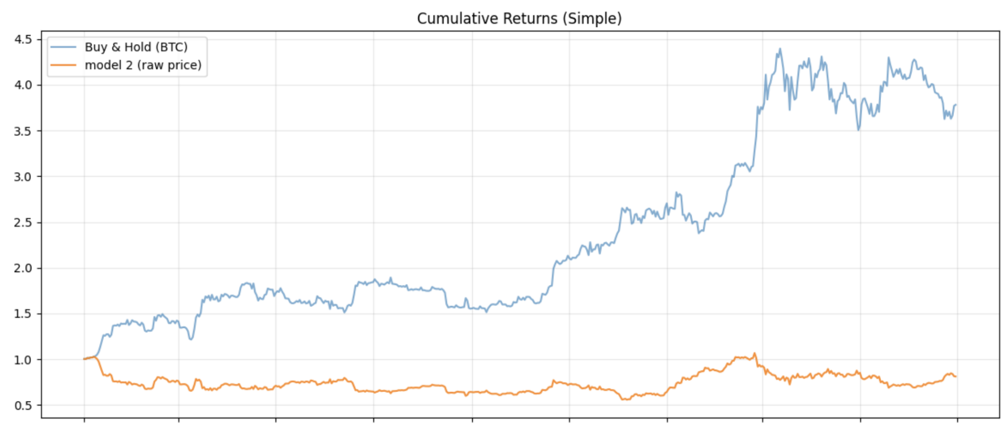

Figure 2: Cumulative returns of long-short strategy vs. Buy & Hold benchmark (day trading)

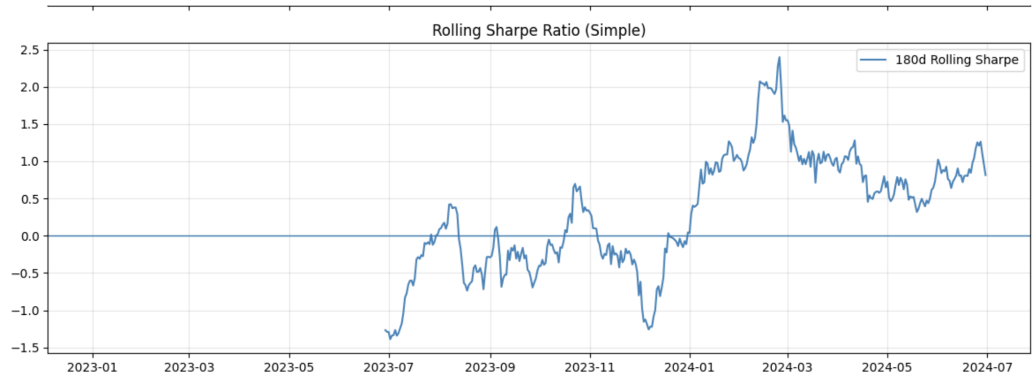

Figure 3: Rolling SR

3.3 Analysis of the long-short strategy using the replicated model

- The direction hit-rate of the strategy is 51.7%. This hit-rate is not enough to outweigh the transaction fees.
- The Information Coefficient 0.0597 is weak for cryptos like BTC.
- The high R^2 was largely derived from tracking the trend rather than predicting forward returns
- Decomposing the prediction error reveals severe overconfidence. A regression of realized vs. predicted returns yields a slope of $b \approx 0.037$, indicating the model predicts price moves $\approx 25\times$ larger than reality. This magnitude mismatch—rather than directional error—is the primary reason the model's MSE (0.0024) significantly underperforms the random walk baseline (0.0006).

3.4 Future research direction

- **Reduce Trading Frequency:** The divergence between the model's high R^2 (trend capture) and poor daily return forecasting (MSE) suggests the signal is effective only at lower frequencies. Future work should shift from daily rebalancing to a weekly or regime-based holding period. This would filter out the daily noise where the model currently struggles and align execution with the trend-following strengths of the Helformer architecture.
- **Calibrate magnitude:** To correct the overconfidence ($b \approx 0.037$), future work should implement a scaling layer or volatility-constrained loss function. Forcing predicted volatility to match realized market volatility would drastically reduce MSE and likely restore the model's advantage over the baseline.

4 Conclusion

This replication confirms that the Helformer architecture can effectively model the statistical properties of BTC price levels, achieving high R^2 values consistent with Kehinde et al. (2025). However, transforming these predictions into a profitable trading strategy proved sensitive to transaction costs and signal definition in the 2023–2024 window.

References

Kehinde, T. O., Adedokun, O. J., Joseph, A., Kabirat, K. M., Akano, H. A., and Olanrewaju, O. A. (2025). Helformer: an attention-based deep learning model for cryptocurrency price forecasting. *Journal of Big Data*, 12(81).

Contact: cuiyong@umd.edu