



# Three-Factor OLS Model for Prediction

This study explores a **new approach** to predict cryptocurrency returns using an OLS model.

The findings indicate a strong correlation between factors and **cryptocurrency price movements** in the market.

## Three-Factor OLS Model for Cryptocurrency Return Prediction

Author: Alan Zhao  
Email: azhao13@chicagoboth.edu  
Affiliations: University of Chicago

### Introduction

Cryptocurrency markets exhibit complex dynamics driven by volatility, momentum, and liquidity. This poster presents a cross-sectional Ordinary Least Squares (OLS) model to predict 24-hour ahead returns using three factors:

1. Volatility: 7-day rolling standard deviation of returns
2. Momentum: 7-day price change
3. Liquidity: 7-day rolling mean of VWAP (volume-weighted average price)

### Data & Preprocessing

- Dataset: 355 crypto trading pairs, from March 1, 2021 onward
- Resolution: 15-minute intervals
- Paths:
- Training data: ``/tmp/train_data``
- Submission IDs: ``/tmp/submission_id.csv``

#### Feature Matrix Construction

- Compute VWAP per symbol
- Calculate returns and rolling statistics
- Align target: returns 96 intervals ahead

### Methodology

1. Load & Batch VWAP
  - Parquet → DataFrame → per-symbol VWAP
2. Factor Calculation
  - ``df_vol = returns.rolling(672).std()``
  - ``df_mom = vwap.pct_change(672)``
  - ``df_liq = vwap.rolling(672).mean()``
3. OLS Regression
  - Stack panel data → drop NaNs
  - Fit ``Return = α + β1·Vol + β2·Mom + β3·Liq``
  - Evaluate: Weighted Spearman correlation

### Implementation Highlights

- Path Autodetection for ``/kaggle/input``
- Robust VWAP: Inf/NaN handling, forward/backfill
- Deprecation Fix: ``pct_change(fill_method=None)``
- Validation: ID format, numeric checks, no missing predictions

```
```python
model = LinearRegression().fit(X, y)
corr = weighted_spearmanr(y.values, model.predict(X))
```
```

### Results & Diagnostics

- Training Samples: 21,812,837
- Spearman Correlation: 0.0044
- Submission Rows: 21,633,318

#### Training Data Stats

- Target  $\sigma$ : 0.0123
- Non-zero targets: 65%

#### Feature Correlations

- Volatility: -0.0152
- Momentum: +0.0125
- Liquidity: +0.0034

#### Prediction Diagnostics

- Pred  $\sigma$ : 0.0056
- Non-zero predictions: 72%

### Submission & Reproducibility

- UID: 129F0U
- Files:
  - ``129F0U.py`` (code)
  - ``Description_129F0U.md`` (strategy + team info)
- Output: ``final_submission_129F0U.csv``

### Conclusions & Next Steps

- OLS yields modest positive rank correlation in cross-sectional crypto returns.
- Future work: nonlinear models, alternative data (sentiment, on-chain), regime-adaptive factors.

### Acknowledgments

Kaggle Community  
@reallygreatsite