

AlphaMark: A Benchmarking Pipeline for Financial Forecast Evaluation

Authors: Dhruv Patel, Haoran Li

January 26, 2026

1 Introduction

The development of predictive signals for financial markets is a central objective in quantitative finance. While the construction of such signals has been extensively studied, their empirical evaluation remains fragmented and often inconsistent across studies. Differences in data preprocessing, portfolio construction, and performance metrics make it difficult to compare results or reproduce findings in a systematic way.

To address this issue, we introduce *AlphaMark*, a benchmarking pipeline for evaluating financial forecasts using profit-and-loss (PnL)-based performance metrics. AlphaMark is designed to take financial time series data and user-defined predictive signals as input and produce a standardized set of evaluation statistics and visual summaries in the form of an automated PDF report.

This document serves as a brief overview of the AlphaMark pipeline. The goal is to enable researchers to easily apply the framework to their own studies, while ensuring consistency, transparency, and comparability across different models, datasets, and research projects in quantitative finance.

2 AlphaMark Pipeline

2.1 Overview

AlphaMark implements a standardized pipeline for evaluating predictive signals on financial time series using profit-and-loss (PnL)-based metrics. The pipeline comes in the form of a Python-based framework in which users can input financial time series data and generate a PDF report containing various summary metrics and charts. [1].

2.2 Data Inputs

AlphaMark consumes daily cross-sectional datasets containing (at minimum) tickers, one or more signal columns, one or more target forward-return columns, and one or more bet-size columns. Each trading day is processed independently, and all metrics are later aggregated across the evaluation window.

Each daily input file is made of:

- **Identifier:** ticker or instrument identifier
- **Signals:** one or more predictive signal columns
- **Targets:** one or more forward-return target columns
- **Bet sizes:** one or more position-sizing columns

The pipeline itself is highly customizable so users can select which signals, targets, bet sizes, quantile thresholds for portfolio construction, and evaluation metrics are included in the output PDF.

2.3 Output PDF

The pipeline produces a consolidated PDF report summarizing performance and behavior across multiple dimensions.

The following link contains a sample output PDF that uses CRSP data over a 20 day interval:
<https://drive.google.com/file/d/1t3efiVhbger00mhVmKy5y7qg-YJwkI-u/view?usp=sharing>

The PDF report contains:

- **Barplot pages** comparing metrics across signals, targets, quantile buckets, and bet-size choices.
- **Heatmaps and correlation panels** summarizing similarity between signals at the signal level and at the realized-PnL level.
- **Temporal pages** showing how cumulative performance and activity evolve through time.
- **Cross-correlation analysis** quantifying lead-lag correlation between each individual stock and a user selected benchmark.
- **Outlier tables** highlighting extreme metric values across days and evaluation configurations.

2.4 Conclusion

In the next week, we will release the pipeline code as well as a more thorough guide. This new guide will dive into how to use and customize the pipeline and will also explain the mathematical foundations of relevant metrics and financial concepts.

References

- [1] M. Cucuringu. *A Pipeline for Benchmarking Forecasts in Financial Time Series Data (PnL Markouts)*. 2025.