# Fama-French ML Project — Daily Progress Report

Date: 2025-10-30

## 1. Dataset Finalization

Confirmed that the modeling dataset is clean and ready for factor or ML modeling. The final dataset includes 106 European financial-sector stocks (IBKR-tradable), covering 2013–2025 and updated automatically via the daily price updater.

## 2. Automated Data Pipeline

Created and validated the daily auto-updater script (update\_prices\_daily.py). It fetches new prices from Yahoo Finance, appends them to the dataset, and updates returns automatically. The dataset now refreshes manually or can be scheduled to run daily.

## 3. Hypothesis Testing Implementation

Implemented and debugged hypothesis\_tests.py to analyze econometric properties:  
- Newey–West HAC robust t-statistics  
- 12-month momentum (skip 1 month) strategy  
- Fama–MacBeth regressions for 20-day momentum and volatility  
- Daily Spearman Information Coefficient (IC) tests

Fixed multiple issues: KeyErrors due to non-trading days, deprecated resampling, and constant input warnings. Standardized output to a clean table with t-stats, SEs, and notes.

## 4. Hypothesis Test Results & Interpretation

Key results from reports/hypothesis\_tests\_summary.csv:

|  |  |  |
| --- | --- | --- |
| Test | Result | Interpretation |
| Zero-mean across assets (daily) | t ≈ 0 | Returns centered around zero; stationary and stable |
| Momentum L–S (12m, skip1m) | +0.27%/month, t=1.17 | Weak, insignificant positive momentum |
| Fama–MacBeth α | t=-0.27 | No pricing bias; neutral intercept |
| Fama–MacBeth β\_mom20d | t≈0.06 | No short-term momentum predictability |
| Fama–MacBeth β\_vol20d | t≈-0.94 | Slightly negative, insignificant relationship |
| Spearman IC (mom20d → next-day) | t≈-0.42 | No daily alpha evidence |

Conclusion: Dataset is statistically clean and economically stable, ready for modeling with no strong biases.

## 5. README Automation Upgrade

Integrated automatic README update (update\_readme.py) to include hypothesis testing summary tables. Every time the script runs, it refreshes the README with data cleaning, assumption tests, and hypothesis results.

## 6. Codebase Structure Review

Cleaned the repository and moved legacy scripts and artifacts to /bin. Main operational scripts: build\_modeling\_universe.py, update\_prices\_daily.py, hypothesis\_tests.py, update\_readme.py.

## 7. Next Steps

1. Begin modeling (Fama–French regressions or ML).  
2. Optionally include factor updates in auto-updater.  
3. Schedule daily automation for dataset updates.  
4. Maintain daily README logs for reproducibility.

## End Status

✅ Clean modeling dataset (2013–2025)  
✅ Auto-updater operational  
✅ Hypothesis testing completed  
✅ README automation integrated  
All systems functional and documented.