**🐞 RiskPipeline Debug Log**

**📅 Date: 26 May 2025**

**🧠 Context: Running python run\_pipeline.py --quick on RiskPipeline (Volatility Forecasting Framework)**

**🧪 Goal: Run a quick test on AAPL, CBA.AX, and ^VIX using the pipeline**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Bug ID | Description | Root Cause | Fix | Date Fixed | Verified |
| BUG-001 | KeyError: 'Adj Close' | yfinance now sets auto\_adjust=True by default, which removes 'Adj Close' | Added auto\_adjust=False in yf.download() | 26-May | ✅ |
| BUG-002 | KeyError: 'Adj Close' in correlation step | 'Adj Close' was hardcoded; fallback logic was missing | Used price\_col = 'Adj Close' if ... else 'Close' | 26-May | ✅ |
| BUG-003 | ValueError: If using all scalar values... | Empty or invalid returns were added to correlation dictionary | Skipped invalid Series with .dropna().empty checks | 26-May | ✅ |
| BUG-004 | No valid return series found. Correlation skipped. | All returns were invalid due to missing columns or empty price data | Logged asset-level issues, improved validations | 26-May | ✅ |
| BUG-005 | KeyError: [('A', 'A', 'P', 'L')] not in index | MultiIndex columns caused by multi-ticker downloads | Started flattening columns using get\_level\_values() | 26-May | ✅ |
| BUG-006 | Columns looked like ['AAPL', 'AAPL', ...] | Incorrect column level flattened — took ticker name, not OHLCV | Changed to df.columns.get\_level\_values(1) to get actual data fields | 26-May | ✅ |
| BUG-007 | KeyError: 'Close' persisted | After flattening, columns were still just ticker names due to wrong level being dropped | Fixed flattening logic to drop correct level (data field names retained) | 26-May | ✅ |
| BUG-008 | Columns still named ['AAPL', 'AAPL', ...] after flattening | Downloading multiple tickers together causes yfinance to return MultiIndex with ticker as outer level; flattening dropped the wrong level | Changed to download each symbol individually, and drop level 0 (ticker) to retain proper OHLCV column names | 26-May | Pending ✅ |
| BUG-009 | KeyError: 'Close' during feature engineering | Wrong level of MultiIndex was being dropped (droplevel(1)), flattening to tickers only (e.g., 'AAPL') instead of expected price labels like 'Close'. | Corrected logic in download\_data() to drop level 0 instead (droplevel(0)). Added explicit validation that 'Close' or 'Adj Close' exists. This allows features to be computed and prevents crash at model training. | 26-May | ✅ Yes |
| BUG-010 | Pipeline crashes at summary report with KeyError: 'Task' | results\_df passed to \_generate\_summary\_report() does not contain a 'Task' column | Validate that model results were created, and that 'Task' column is added before summary | 26-May | ✅ Yes |
| BUG-011 | KeyError: 'Asset' during regression plotting | pipeline.results is empty or missing expected columns because no models were trained | Add checks before plotting to skip if results\_df.empty or 'Asset' not in columns | 26-May |  |
| BUG-012 | Visualization crashes with KeyError: 'Asset' | results\_df was empty, so pivoting on 'Asset' column failed | Added results.empty check and skipped plotting if true | 26-May | ✅ Yes |
| BUG-013 | Empty results dataframe | Model training silently failed; no predictions or metrics returned | Fixed data loader, feature engineering, and added logging + fallback checks | 26-May | ✅ Yes |
| BUG-014 | DL model evaluation crashes with ValueError: Found array with 0 sample(s) | Walk-forward splits failed due to small dataset; sequences not generated | Rewrote \_evaluate\_dl\_model() with shape checks and safe error handling | 26-May | ✅ Yes |
| BUG-015 | Walk-forward splits inappropriate for small datasets | test\_size=252 too large for ~300-row quick test data | Rewrote WalkForwardValidator.split() to handle small datasets better | 26-May | ✅ Yes |
| BUG-016 | SHAP analysis crashes with KeyError: ['VIX', 'VIX\_change'] | SHAP tried to use global VIX features not present in per-asset features | Added missing column check in \_generate\_shap\_analysis() and improved SHAP + result saving logic | 26-May | ✅ Yes |
| BUG-017 | Matplotlib seaborn style deprecation warning | seaborn-v0\_8-darkgrid deprecated in newer matplotlib | Replaced with manual rcParams configuration | 26-May | ✅ |
| BUG-018 | Radar chart creates new figure inside subplot | \_plot\_radar\_comparison() called plt.figure() inside subplot | Fixed with ax.remove() and add\_subplot() | 26-May | ✅ |
| BUG-019 | Pivot operation fails with duplicate entries | data.pivot() crashes when duplicate Asset-Model combos exist | Added duplicate detection & aggregation before pivot | 26-May | ✅ |
| BUG-020 | Hardcoded market classification causes KeyError | US/AU asset lists don’t cover all possible symbols | Implemented fallback using asset naming conventions | 26-May | ✅ |
| BUG-021 | Unsafe DataFrame access in market comparison | Used market\_avg.loc['US'] without checking existence | Added existence validation + single-market fallback | 26-May | ✅ |
| BUG-022 | SHAP crashes with mismatched input/output dimensions | SHAP values vs feature names misaligned | Added full input validation and dimensional checks before SHAP runs | 26-May | ✅ |
| BUG-023 | Walk-forward validation generates 0 splits | test\_size=252 was too large for small datasets (~346 rows total) | Rewrote validator with adaptive sizing and expanding window | 26-May | ✅ |
| BUG-024 | KeyError: 'Precision' in visualization | Plots accessed metrics that didn't exist when no models were trained | Added column existence checks and empty result fallback | 26-May | ✅ |
| BUG-025 | Model evaluation crashes with empty arrays | No error handling for empty prediction arrays in evaluation routines | Wrapped in try-catch blocks and added fallback metrics | 26-May | ✅ |
| BUG-026 | SHAP feature mismatch | SHAP expects correlation features, but SHAP data reconstruction skips correlation step | Added proper reconstruction pipeline using cached data + feature list filtering | 26-May | 🔧 Fix applied |
| BUG-027 | Radar chart subplot bug | ax.get\_figure() returns None in some matplotlib contexts | Added subplot-safe fallback, converts radar to bar plot if needed | 26-May | 🔧 Fix applied |
| BUG-028 | Font warning spam: findfont: serif not found | Times New Roman hardcoded in visualization.py; missing on Linux/WSL systems | ✅ Replaced with cross-platform font fallback: Times New Roman → Times → Liberation Serif → DejaVu Serif → serif | 26-MAY | 🔧 FIX APPLIED |
| BUG-029 | SHAP feature mismatch between training and analysis | Models trained without VIX features; SHAP tries to include them, causing dimension mismatch | ✅ Store training feature list from scaler.feature\_names\_in\_ | 26-MAY | 🔧 FIX APPLIED |
| BUG-030 | NameError: name 'mplstyle' is not defined | mplstyle.use('fast') called at the top of visualization.py before import matplotlib.style | ✅ Move import matplotlib.style as mplstyle to top of file |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**🛠️ Key Fixes Summary**

|  |  |
| --- | --- |
| Area Fixed | Fix Description |
| Data Download | Added auto\_adjust=False and handled MultiIndex correctly with .droplevel(0) |
| Column Fallback | Used 'Adj Close' if exists else 'Close' logic for all price accesses |
| Log Returns Validation | Skipped assets if log returns were invalid or empty |
| Correlation Engine | Now robust against missing or corrupt data |
| Debugging | Added structured logging for: data shape, column names, valid/invalid returns, and feature creation summary |