Replication Curve – A Black Box Hurdle

Author: Syed Bashir Hydari

The Core Problem

Our attempt to replicate Carta & Conversano (2020) revealed a fundamental truth about computational finance research: exact numerical replication is nearly impossible without access to the original source code and data. While we successfully replicated the paper's Monte Carlo simulations with over 99% accuracy (thanks to their transparent, deterministic parameters) the empirical portfolio analyses proved irreproducible at a precise quantitative level despite our best efforts.

This reflects a structural challenge in computational finance: academic papers cannot possibly document every implementation decision, financial data is messy and proprietary, and portfolio optimization is numerically unstable. The result is a "black box" where we can see the inputs (stock returns) and outputs (optimal portfolios) but cannot replicate the machinery connecting them.

Data: More Complex Than It Appears

The paper states it uses "42 equities from the EuroStoxx 50 index" from 2000-2018, converting "daily adjusted close prices" to "monthly returns." This sounds clear but conceals numerous ambiguities.

Ticker Identification: The paper lists company names (Adidas, Total, Schneider Electric) but not ticker symbols. European companies trade on multiple exchanges—Total trades as FP.PA in Paris, TTE in New York, and TOT elsewhere. Each has slightly different price histories. We made educated guesses, but even a handful of wrong ticker choices cascade through the entire analysis since covariance matrices depend on which specific assets are included.

Data Source: The paper doesn't specify Bloomberg, Refinitiv, Yahoo Finance, or another vendor. This matters enormously. Professional data vendors make thousands of proprietary decisions about stock splits, dividend adjustments, mergers, and delistings. Studies show return calculations can differ by 1-5% annually across vendors for the same security. We used Yahoo Finance for accessibility, but this guarantees our data differs from theirs.

Temporal Details: Does "January 2000" mean returns calculated using December 1999 prices or February 2000 prices? Are "monthly" returns calculated from the last trading day of each month? The 15th? An average? These seemingly pedantic details matter because portfolio optimization is exquisitely sensitive to inputs—a single month's difference changes the covariance matrix, which changes optimal portfolios.

Financial Crises Omissions: several major financial crises impacted Italian and European markets between 2017-2018, which have been noticeably omitted in the data selection process of Carta (given their exceptionally low volatility for their equity curves). We do not know what this omission profile looks like and thus cannot match the data selection blocks Carte performed.

Corporate Actions: Between 2000-2018, these 42 companies underwent countless splits, dividends, spinoffs, and restructurings. How should these be handled? Different adjustment methodologies yield different return histories. The paper provides no guidance, likely because the authors used vendor-adjusted data without questioning the specifics.

Optimization: The Numerical Minefield

Even with identical data, portfolio optimization itself is irreproducible without implementation details. The Kelly optimization problem appears straightforward mathematically but executing it requires dozens of unstated choices.

Solver Selection: Which quadratic programming solver did the authors use? R's quadprog? MATLAB's quadprog? Python's CVXOPT? Commercial solvers like Gurobi? Each algorithm converges differently and produces subtly different solutions. The paper provides no indication of solver choice, convergence tolerances, maximum iterations, or other numerical parameters.

Covariance Matrix Conditioning: With 42 stocks and ~228 months of data, the sample covariance matrix is almost certainly ill-conditioned (condition number exceeding 1,000). This creates numerical instability where tiny input changes cause enormous output changes. The standard solution is regularization (adding a diagonal term or shrinking toward a simpler structure). But the paper never mentions regularization. Did they use it? If so, what method and parameter? We applied ridge regularization ($\lambda = 0.01\%$ of max eigenvalue) to ensure stability, but this is our choice.

Constraint Interpretation: The paper specifies $\Sigma f_i \leq 1$ but reports results where $\Sigma f_i = 1$ exactly. Did they enforce equality or inequality? This affects the solution. We reverse-engineered equality constraints to match their reported outputs, but that's guessing.

Negative Weight Handling: The tangent portfolio often produces negative weights (short positions). The paper constrains weights to be non-negative, but how? Solve unconstrained then iteratively remove negatives? Use constrained optimization from the start? Zero out negatives and renormalize? Each approach yields different results with different numbers of non-zero positions.

Rolling Window: Additional Ambiguities

The out-of-sample analysis uses a "24-month lookback window" rebalanced monthly. More ambiguities arise:

Window Type: Does "24-month lookback" mean a sliding window (always 24 months) or expanding window (grows over time)? The phrase suggests sliding but isn't stated explicitly.

Rebalancing Timing: When exactly does rebalancing occur? End of month t using month t prices? Beginning of month t+1? The timing creates one-day execution lags that can shift annual returns by 0.5-1%.

Failure Handling: Over 132 months, optimization occasionally fails—singular matrices, convergence issues, data glitches. How should code respond? Use previous weights? Fall back to equal-weight? Skip the period? We chose to maintain previous weights, but the paper provides no guidance.

Transaction Costs: The paper never mentions transaction costs, implying zero costs. This is standard in academic research but highly unrealistic. Real implementation would see returns eroded by 0.5-2% annually depending on turnover.

Successes & Deviations

We replicated the paper's Monte Carlo simulations (Section 3.1) with >99% accuracy while struggling with empirical sections. Why? Monte Carlo is deterministic given stated parameters: $\mu=0.05$, $\sigma=0.20$, $r=0.01$, $f^*=0.20$. No data vendors, no ticker ambiguities, no solver choices, no covariance regularization—just clean mathematics. We implemented the formulas and matched their results within rounding error.

Empirical sections depend on unspecified external inputs (proprietary data, corporate action adjustments, vendor-specific calculations) and unstated implementation choices (solvers, tolerances, regularization methods, constraint handling). This is the difference between theoretical and empirical science: theory replicates from first principles, but empirical work requires replicating the messy reality of data collection and computational implementation.

The Reproducibility Crisis in Finance

Our struggles reflect a broader problem. Hou, Xue, and Zhang (2020) attempted to replicate 452 published stock market anomalies—36% completely failed even with original code. Harvey, Liu, and Zhu (2016) found that over half of published "factors" lose statistical significance out-of-sample. Chen and Zimmermann (2022) created an open-source library of 400+ return predictors and found substantial variation across seemingly equivalent implementations.

The root cause: financial data is historical and constantly revised retroactively. Unlike physics experiments that can be repeated with identical conditions, or mathematical proofs that are self-contained, empirical finance depends on proprietary data that cannot be freely shared and computational workflows with hundreds of implementation branches.

Our Replication Caliber

Given these obstacles, exact numerical replication is impossible without source code and original data. We use realistic standards:

Tier 1 - Qualitative Replication (Minimum Standard):

- Same patterns: Kelly portfolio more concentrated than tangent ✓
- Same risk-return relationships: Kelly higher return and higher volatility ✓
- Same performance ranking: Kelly > Tangent > Equal > Min-Variance ✓
- Same position on efficient frontier ✓

Tier 2 - Semi-Quantitative Replication (Strong Success):

- Portfolio weights within ±20% (top holding: 51% vs 40-55%)
- Performance metrics within ±30% (CAGR: 3.27% vs 2.5-4.0%)
- Same order of magnitude for all statistics

Tier 3 - Exact Quantitative Replication (Impossible Without Source):

- Identical portfolio weights to two decimals
- Identical CAGR to two decimals
- Identical time series of returns

By this framework, our replication achieved Tier 1 and Tier 2.

Conclusion

Replicating empirical portfolio optimization research without source code is very difficult. You can identify the main ingredients (data, optimization method, constraints), but you'll never know the exact measurements, cooking temperatures, timing, or techniques that produced the original.

There is an implicit message too: researchers should share code, data, and computational environments. Until that becomes standard practice, we must accept that "successful replication" in computational finance means qualitative agreement in many cases.