Dynamic Portfolio Optimization - Analytics

Anuj Panwar
Ashoka University

April 30, 2025

Abstract

We convert a legacy weekly loan-tape into a dynamic portfolio-optimization framework, engineering three financial ratios and using ideas taught in Reinforcement Learning to make decisions. Across four phases—vanilla REINFORCE, baseline variance reduction, safety-penalty shaping, and tail-risk metrics—we convert a static data and transform it as a continuous Gaussian policy to allocate capital to loans.

1 Introduction

Bond portfolio managers typically rely on static loan-level data to assess risk and return, missing the evolving interplay of interest costs, borrower leverage, and payment behavior. Drawing on prior work analyzing a weekly-refreshed securitization tape—originally indexed to Libor and later SONIA—we repurposed it into a Markov decision process.

2 Dataset Source and Description

Legacy Tape from Fixed-Income Operations

Our dataset is a historical weekly snapshot of securitized loans, originally used by a fixed-income desk to monitor portfolio performance.

We utilize an anonymized securitization dataset consisting of weekly loan-tape snapshots. Each record includes:

- **Interest Rate:** loan coupon, recorded as a percentage.
- **Debt-to-Income (DTI):** borrower leverage ratio.
- **Days Past Due (DPD):** measure of payment delinquency.
- **Loan Age:** months since loan origination.
- **Principal Balances:** original and current outstanding amounts.

This dataset was originally updated each week to reflect market rate changes byh extracting SONIA rates from Bloomberg terminal and borrower performance, but here serves purely as a static archive for RL experimentation.

3 Data Exploration and Important Features

We began with over 1,000 loan records. Our exploratory steps included:

Key Cleaning & Feature Engineering

1. *Header realignment & NA removal*: Skipped the first five metadata rows so that row six became the true header; dropped any rows with missing values in our four core features.

2. *Percentage parsing*: Used regular expressions to extract trailing "%" from `Loan Product` into `InterestRatePct` and stripped "%" from `Qualifying DTI`.

3. *Column selection*: Retained only `InterestRatePct`, `DPD`, `Loan Age`, and `Qualifying DTI`, discarding all other fields.

Descriptive Statistics

Table 1: Descriptive Statistics for Core State Variables

Feature	**Min**	**25th%ile**	**Median**	**75th%ile**	**Max**
Interest Rate (%)	2.0	4.2	5.4	6.7	10.0
Debt-to-Income (%)	0.0	20.5	30.1	42.8	60.0
Days Past Due (DPD)	0	0	0	15	90
Loan Age (months)	0	12	24	36	60

Tail-Risk Metrics

To quantify extreme-loss behavior, we computed:

- **$\text{VaR}_{5\%}$:** the 5% empirical quantile of the 200-episode return distribution.
- **$\text{ES}_{5\%}$:** the average of the worst 5% of returns.

These quantitative metrics validate that phase 4's penalty shaping significantly compresses the tail of the loss distribution.

4 Methods

1. **Baseline Variance Reduction (Phase 3):** Introduced a moving-average baseline ($\beta = 0.2$), so updates used advantage $G(t) - b$. This cut gradient variance by a significant margin, producing a much smoother learning curve.

2. **REINFORCE + Penalty + Baseline Comparison (Phase 4):** Combined the Phase 3 baseline with a proportional DTI penalty $r_t = a_t \cdot \mathrm{r}_t - 0.01(\mathrm{DTI}_t - 0.20)\mathbf{1}_{\mathrm{DTI}>0.20}$. This hybrid method drove DTI violations to zero by episode 60 while preserving positive returns.

3. **Tail-Risk Analysis & Frontier Visualization (Phase 5):** Computed 5% VaR and ES on the episode-return distribution, and rendered a 3D "yield–risk frontier" of $\mu_\theta(x)$ over interest rate and DPD. This produced an actionable decision surface and demonstrated significant tail-risk improvements.

5 Experimentation

Conducted Experiments

1. **Variance Comparison:** Measured standard deviation of Phase 2 vs. Phase 3 returns to confirm baseline efficacy.

2. **Penalty Sensitivity:** Swept DTI thresholds (15%, 20%, 25%) and penalty scales (0.005, 0.01, 0.02), recording yield vs. violation trade-offs.

3. **Frontier Region Identification:** Utilized the Phase 5 3D contour plot to pinpoint "hot" allocation regions—high-interest (7%) and low-DPD (5 days).

6 Final Results

Quantitative Summary

- **Mean Quarterly Return:** +1.23%±0.05% (95% CI, on-policy MC).
- **5% VaR / ES:** –0.82%/–1.15%.
- **Violation Rate:** Phases 1–2 averaged 2.8 violations/episode; Phase 4 converged to 0 by episode 60.
- **Yield–Risk Frontier:** Capital allocation peaks above 0.8 when rate8% DPD=0, declines smoothly as risk increases.

7 Conclusion

We have demonstrated a fully reproducible reinforcement-learning framework for dynamic bond-pool allocation under safety constraints. By progressively adding a variance-reducing baseline and penalty shaping, the policy achieves stable, positive returns while enforcing DTI regulations. Tail-risk metrics and the yield–risk frontier furnish both quantitative and visual decision tools for portfolio managers. Future work includes on-policy TD(λ) refinements, multi-pool coordination, and real-time deployment on live collateral feeds. This study demonstrates how a four-phase policy-gradient pipeline—augmented by a moving-average baseline and safety-penalty shaping—can learn continuous loan allocations that maximize yield while enforcing DTI constraints. Tail-risk metrics (VaR, ES) and the 3D frontier provide transparent, practitioner-ready decision tools. Future work includes TD(λ) extensions and real-time adaptation on live collateral streams.