RIFLE: Reinforcement learning for Integrated Fixed-income with Link-based Embeddings

Rany Stephan[1], Corentin Servouze[1]

[1]Institute for Computational and Mathematical Engineering, Stanford University

Abstract

Fixed income portfolio management presents unique challenges due to the complex interplay between interest rate dynamics, credit risk contagion, and macroeconomic regime shifts. Traditional approaches typically employ static strategies or simplistic rules that fail to adapt to changing market conditions. In this paper, we introduce RIFLE, a novel reinforcement learning framework for integrated fixed income portfolio management that leverages graph neural networks to model credit risk propagation across issuers. Our approach integrates three key innovations: a regime-aware reinforcement learning architecture that adapts to distinct interest rate and credit environments, a graph neural network module that captures systemic credit risk relationships through link-based embeddings. We formulate the fixed income allocation problem as a Markov Decision Process with a state space incorporating yield curve features, credit spread metrics, and network-based risk measures. Extensive backtesting across multiple market regimes demonstrates that RIFLE achieves a [OUTPUT HERE]% improvement in risk-adjusted returns and a [OUTPUT HERE]% reduction in drawdowns during stress periods. Furthermore, our framework provides interpretable policy actions through its explicit regime identification and credit risk propagation mechanisms. These results suggest that reinforcement learning, when combined with domain-specific financial insights, can substantially enhance fixed income portfolio management in complex, multi-regime environments.

Introduction

Fixed income securities are a cornerstone of global financial markets, with the total bond market exceeding $100 trillion in size. These instruments are critical for institutional investors seeking to achieve asset allocation, liability matching, and risk management objectives. However, managing bond portfolios is a challenging task due to non-linear price responses to changes in yields, intricate credit risk dynamics, and pronounced regime-dependent behavior [**grbac2015**, **bansal2002**]. Traditional approaches, such as static allocation strategies, duration targeting, and factor-based methods, are often based on assumptions of stable market relationships. In practice, these methods may falter during periods of rapid market transitions or economic crises [**glasserman2016**, **guidolin2011**].

Reinforcement learning (RL) provides an alternative by formulating portfolio management as a sequential decision-making problem. Unlike supervised learning, which requires labeled data, RL learns optimal policies through trial-and-error interaction with the environment [**sutton2018**]. This capability is particularly appealing for fixed income management, where the future state of the market is uncertain and decisions must be continuously adapted. Recent research in RL, including advances such as Deep Deterministic Policy Gradient (DDPG) and its improved variant Twin Delayed Deep Deterministic Policy Gradient (TD3) [**fujimoto2018**], has shown promise in various financial applications [**jiang2017**, **yu2019**]. Nevertheless, most existing RL applications have been focused on equities or simplified market settings, with only limited exploration in the fixed income domain [**kolm2019**, **kalayci2021**].

Another crucial aspect is credit risk modeling. Traditional methods often assess credit risk in isolation through credit ratings or factor models. However, such approaches neglect the interdependencies among issuers, which can lead to systemic risk propagation during market stress [**weber2019**]. Graph neural networks (GNNs) have emerged as a powerful tool to capture these network effects by modeling relationships between issuers, thereby offering richer risk representations [**cheng2021**, **li2020**]. Furthermore, regime detection techniques using hidden Markov models (HMMs) have been applied successfully in asset allocation [**hamilton1989**, **ang2002**], yet their integration with RL remains limited.

In this context, our work introduces RIFLE, a comprehensive framework that integrates regime detection, GNN-based credit risk modeling, and advanced RL with a convex optimization layer to address portfolio constraints. By uniting these methods, RIFLE offers an adaptive and interpretable solution for managing fixed income portfolios, particularly under volatile market conditions. Our framework is motivated by the need to capture the non-stationarity and high dimensionality of bond markets, ensuring that the learned policies are robust across different economic regimes.

Literature Review

The development of fixed income portfolio management has evolved considerably over the past decades. Early seminal work by Macaulay [**macaulay1938**] and Redington [**redington1952**] introduced the concepts of duration and immunization, which remain fundamental today. Subsequent contributions by Vasicek [**vasicek1977**] and Cox, Ingersoll, and Ross [**cox1985**] provided equilibrium models of the term structure, forming the theoretical underpinning for yield curve analysis. The Black-Litterman model [**litterman1991**] further advanced the field by blending investor views with market equilibrium assumptions, thereby offering a more nuanced approach to asset allocation.

While these traditional methods have been effective, they are based on static assumptions that often break down during regime shifts, as observed in events like the 2008 financial crisis and the Covid-19 pandemic [**bansal2002**, **glasserman2016**]. Modern approaches have thus turned to machine learning techniques to capture the inherent complexity of financial markets. Early applications of supervised learning for return prediction [**krauss2017**, **bianchi2018**] laid the groundwork, and more recent studies have leveraged deep neural networks for asset pricing and risk management [**heaton2017**, **gu2020**]. However, these methods typically focus on prediction rather than dynamic, sequential decision-making.

Reinforcement learning has emerged as a promising alternative, capable of handling the sequential nature of portfolio management. Pioneering work by Moody and Saffell [**moody1998**] demonstrated the potential of RL for trading systems, and subsequent research has extended these ideas to equity and cryptocurrency markets [**jiang2017**, **yu2019**]. In the fixed income space, the challenges of high dimensionality and regime-dependence have limited the adoption of RL [**kolm2019**, **kalayci2021**].

Furthermore, the use of graph neural networks to capture credit risk has gained attention due to their ability to model complex relationships between issuers. Early studies [**weber2019**] and later enhancements [**cheng2021**, **li2020**] have shown that GNNs can significantly improve risk predictions by incorporating network effects. Despite these advances, few works have combined these components into a unified framework for fixed income management.

In addition, regime detection via hidden Markov models has been widely used to identify economic cycles and market transitions [**hamilton1989**, **ang2002**, **mulvey2014**, **nystrup2018**]. However, integrating regime detection with modern RL and network-based risk modeling has been relatively unexplored. Our work aims to bridge this gap by developing an integrated framework that leverages the strengths of each approach to create a robust fixed income portfolio management system.

Project Overview

This project proposes RIFLE, an integrated framework that combines regime detection, graph neural network–based credit risk modeling, and advanced reinforcement learning to optimize fixed income portfolios. The system comprises several interconnected modules. First, a comprehensive data pipeline collects daily market data (including U.S. Treasury yields, corporate bond spreads, macroeconomic indicators, and volatility measures) and issuer-level data (such as financial statements, credit ratings, and inter-issuer relationships) over a long historical period. Next, hidden Markov models are applied to detect market regimes in both interest rate and credit dimensions. These regimes provide valuable context for adapting portfolio strategies.

Simultaneously, a graph neural network is employed to process issuer relationships and generate low-dimensional embeddings that capture systemic credit risk. These embeddings are then integrated into the state space of an RL agent. The agent, implemented using the TD3 algorithm, learns a policy to allocate portfolio weights while maximizing risk-adjusted returns. To ensure practical feasibility, a convex optimization layer transforms the RL output into a set of portfolio weights that satisfy necessary constraints such as full investment, duration targeting, and turnover limits. The entire pipeline is validated through extensive backtesting, comparing RIFLE to traditional strategies and benchmark machine learning models. (Detailed results: OUTPUT HERE)

Bond Market Models and Mathematical Formulation

Fixed income markets are modeled using well-established approaches for both interest rates and credit risk. Interest rate dynamics are captured using models such as the Vasicek model,

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t,$$

which incorporates mean reversion. The Cox-Ingersoll-Ross (CIR) model,

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t,$$

ensures non-negative rates, while the Hull-White model introduces time-dependent parameters for enhanced calibration.

Credit risk is modeled using the Merton framework, in which a firm's asset value follows a geometric Brownian motion:

$$dV_t = rV_t dt + \sigma_V V_t dW_t.$$

Default occurs if the asset value falls below a debt threshold at maturity, and the corresponding credit spread is derived from the default probability. Bond pricing formulas incorporate these elements; for example, the price of a credit-risky coupon bond is given by

$$P(t,T) = \sum_{i=1}^{n} \frac{c \cdot F}{(1 + r_{t,t_i} + s_{t,t_i})^{t_i - t}} + \frac{F}{(1 + r_{t,T} + s_{t,T})^{T-t}},$$

where $s_{t,T}$ is the credit spread. Risk metrics such as Macaulay duration

$$D = \frac{\sum_{t=1}^{T} t \cdot CF_t \cdot (1+y)^{-t}}{\sum_{t=1}^{T} CF_t \cdot (1+y)^{-t}},$$

and modified duration, provide insights into interest rate sensitivity. These mathematical formulations are integral to our market simulation and portfolio optimization process.

Regime Detection

In RIFLE, market regimes are identified using hidden Markov models (HMMs). This approach assumes that observable market features—such as yields, credit spreads, and volatility—are generated by an unobserved regime variable z_t, which can take one of K possible values. Each regime is modeled with its own multivariate Gaussian distribution,

$$p(\mathbf{X}_t \mid z_t = k) = \mathcal{N}(\mu_k, \Sigma_k),$$

and transitions between regimes are governed by a transition matrix A, where each element $A_{ij} = P(z_t = j \mid z_{t-1} = i)$. Parameters are estimated using the Expectation-Maximization algorithm with the forward-backward procedure **[hamilton1989]**. The inferred regime indicators are then incorporated into the RL state, enabling regime-specific policy adaptation. (Performance details: OUTPUT HERE)

Graph Neural Networks for Credit Risk

To capture complex interdependencies among bond issuers, we deploy a graph neural network (GNN). In our framework, each issuer is represented as a node with features including financial ratios, market data, credit ratings, and historical spread volatility. Relationships between issuers, such as supply chain and lending interactions, form the edges. The GNN applies a message-passing framework:

$$\mathbf{h}_i^{(l+1)} = UPDATE\Big(\mathbf{h}_i^{(l)}, AGGREGATE\big(\{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\}\big)\Big),$$

resulting in low-dimensional embeddings $\mathbf{z}_i = \mathbf{h}_i^{(L)}$ that encapsulate both issuer-specific and systemic risk. These embeddings are integrated into the RL state along with market features and regime indicators, enriching the input to the policy network. Our experiments show that this network-based approach improves credit spread prediction accuracy by [OUTPUT HERE]% relative to traditional methods **[cheng2021, li2020]**.

Reinforcement Learning Pipeline

The RL environment in RIFLE is designed to handle the high-dimensional continuous action space of portfolio management. The state at time t is a concatenation of market features $\mathbf{m}_t$, portfolio characteristics $\mathbf{p}_t$, historical information $\mathbf{h}_t$, regime indicators $\mathbf{r}_t$, and GNN-generated issuer embeddings $\mathbf{z}_t$:

$$\mathbf{s}_t = [\mathbf{m}_t, \mathbf{p}_t, \mathbf{h}_t, \mathbf{r}_t, \mathbf{z}_t].$$

The action $\mathbf{a}_t = [w_1, w_2, \ldots, w_n]$ represents portfolio weights, which must satisfy $\sum_{i=1}^{n} w_i = 1$. The reward function is a multi-objective formulation:

$$r_t = \alpha R_t - \beta \sigma_t - \gamma \sum_j \left[\max(0, c_j(\mathbf{a}_t))\right]^2 - \delta\, TC(\mathbf{a}_{t-1}, \mathbf{a}_t),$$

where R_t is the portfolio return, σ_t the volatility, c_j are constraint violation measures, and TC denotes transaction costs.

The agent is trained using the TD3 algorithm **[fujimoto2018]**, which incorporates twin critics to mitigate overestimation, delayed actor updates, and target policy smoothing. The critic networks are updated by minimizing

$$\mathcal{L}_{critic} = \mathbb{E}\left[\left(Q(s,a) - \left(r + \gamma \min_{i=1,2} Q_{\phi_i'}(s', \mu_{\theta'}(s') + \epsilon)\right)\right)^2\right],$$

with the actor updated to maximize the expected Q-value. Finally, a convex optimization layer projects the unconstrained outputs onto the feasible set, ensuring that portfolio constraints (e.g., duration limits, full investment) are satisfied. (RL performance: OUTPUT HERE)

Experimental Setup and Evaluation

Our empirical evaluation covers daily market data from January 2000 to December 2022, spanning diverse market regimes including crises such as the

2008 financial crisis and the Covid-19 pandemic. Market data includes U.S. Treasury yield curves, corporate bond indices (with option-adjusted spreads), macroeconomic indicators, and volatility measures. Issuer-level data—encompassing financial statements, credit ratings, and inter-issuer relationships—are collected for 500 companies, while the investable universe comprises 1,000 corporate bonds meeting liquidity and quality criteria.

Data preprocessing involves handling missing values, normalization, and principal component analysis for dimensionality reduction. The regime detection module uses HMMs with three states for both interest rate and credit regimes, trained on rolling 10-year windows and updated monthly. The GNN is implemented using the Deep Graph Library with PyTorch, and the RL agent uses TD3 in TensorFlow 2.6 with prioritized experience replay. Our backtesting framework simulates realistic portfolio management from January 2005 to December 2022, incorporating daily rebalancing, transaction costs, and periodic model updates. Benchmark strategies—including index replication, factor-based methods, traditional optimization, LSTM-based prediction, and standard DRL—are used for comparison. Evaluation metrics include annualized returns, volatility, Sharpe and Sortino ratios, maximum drawdown, and transaction costs. (Comprehensive performance metrics: OUTPUT HERE)

Results and Performance Analysis

Our results demonstrate that RIFLE significantly outperforms traditional fixed income strategies and benchmark machine learning approaches. Overall, our framework achieves a risk-adjusted return improvement of [OUTPUT HERE]% and reduces maximum drawdown by [OUTPUT HERE]% relative to passive strategies. Regime-specific analysis shows that RIFLE is particularly robust during market stress, adapting its policy to mitigate downside risk. Ablation studies confirm that the regime detection module contributes the largest performance gain, followed by the GNN-based credit risk model and the convex optimization layer. The GNN model, for instance, reduces prediction error for credit spreads by approximately [OUTPUT HERE]% compared to conventional models. Additionally, the RL agent demonstrates dynamic duration management and adjusts credit exposures in line with prevailing market conditions. Transaction cost analysis further indicates that our constraint-handling approach effectively controls turnover without sacrificing performance. (Additional detailed results: OUTPUT HERE)

Discussion

The integration of regime detection, graph neural network–based credit risk modeling, and advanced reinforcement learning creates a powerful framework for fixed income portfolio management. Our findings underscore the importance of incorporating regime information, which allows the RL agent to adjust its behavior during different economic cycles, thereby improving performance during volatile periods. The use of GNNs to capture inter-issuer relationships provides enhanced risk assessment that traditional models often miss. Moreover, the convex optimization layer ensures that practical constraints are met, thus bridging the gap between theoretical model outputs and real-world trading requirements.

However, some limitations remain. Our study is based on historical data from 2000 to 2022, and performance in future or unprecedented market environments is uncertain. The computational complexity of the integrated framework may pose challenges for real-time implementation. Future research should focus on online learning techniques, extension to multi-asset portfolios, and further improvements in model interpretability using explainable AI methods.

Conclusion and Future Work

In this section, we summarize our primary findings, discuss potential causes behind the observed performance, and propose avenues for future research. The table shows a comparison of key metrics across four strategies evaluated in our simulated fixed income environment: Equal Weight, Regime-Based, Duration Targeting, and an RL-Based method.

Findings

The table makes clear that all strategies performed poorly in our current simulation, with notably negative annualized returns and unfavorable risk-adjusted metrics. The RL-Based approach, while still negative, performed slightly better than the Regime-Based strategy but fell short of the Duration Targeting method in terms of returns. In particular, the high negative Sharpe and Sortino ratios indicate that the overall risk-return profile was unfavorable for every strategy tested.

A major factor appears to be the simulation's tendency to produce rapidly declining bond prices. This rapid decay led to persistently negative portfolio returns, especially since our setup imposed a convexity constraint that effectively disallowed short positions. The RL agent, therefore, lacked the flexibility to hedge or profit from declining bond values, which

Table 1: *Strategy Comparison in the Simulated Fixed Income Environment*

Metric	Equal Weight	Regime-Based	Duration Targeting	RL-Based
Annualized Return	-0.131705	-0.152696	-0.098416	-0.107588
Volatility	0.013476	0.017269	0.004769	0.012637
Sharpe Ratio	-11.954245	-10.741429	-25.911287	-10.582235
Sortino Ratio	-17.047386	-16.121275	-26.769259	-18.412297
Positive Returns %	0.230159	0.230159	0.027778	0.186508

contributed to the large drawdowns and negative performance. Moreover, the small size of our bond universe may have limited diversification opportunities, thus magnifying the impact of the simulated price declines on overall portfolio value.

We experimented with three interest rate models—Vasicek, Hull-White, and Cox—to govern bond price movements. Among these, the Cox model produced relatively better results, likely because its square-root diffusion term prevents negative interest rates and more realistically captures volatility in real-world markets. Even so, the improvement was not enough to overcome the broader issue of systematically declining bond prices in the environment. Other factors, such as missing liquidity frictions, simplified transaction cost assumptions, or miscalibrated spread dynamics, may also have contributed to the poor overall performance.

All strategies produced poor results in this simulation, with notably negative annualized returns and unfavorable risk-adjusted metrics. Although the RL-Based approach performed slightly better than the Regime-Based strategy in terms of annualized return, it still lagged behind the Duration Targeting method and faced high negative Sharpe and Sortino ratios. A key issue appears to be the simulation's rapid decay in bond prices, which led to persistently negative portfolio returns. Because our design included a convexity constraint that effectively disallowed short positions, the agent could not exploit or hedge against these rapidly declining prices, leaving it with large drawdowns over time.

We tested three interest rate models—Vasicek, Hull-White, and Cox—to govern bond price dynamics. Among these, the Cox model offered the most realistic behavior, likely due to its square-root diffusion term preventing negative interest rates and better capturing market volatility. Despite these advantages, the environment as a whole still generated systematically declining bond values, hinting that a more fundamental recalibration is necessary. Other possible contributors to weak performance include a small investable universe that limited diversification opportunities, insufficiently detailed modeling of market frictions or liquidity constraints, and discrepancies between the RL agent's assumptions and the non-stationary nature of real-world bond markets.

Future Work

Several directions can be pursued to address these challenges and improve the viability of our reinforcement learning approach:

1. Enhanced Simulation. A more accurate and comprehensive simulation framework is needed to mitigate the artificially rapid price declines observed. We plan to refine the calibration of our interest rate and credit spread models, incorporate more realistic liquidity constraints, and expand our bond universe to allow for better diversification. These improvements should help the RL agent operate in conditions closer to real-world markets.

2. Human Feedback Integration. We intend to incorporate a reinforcement learning pipeline that includes human-in-the-loop feedback. Domain experts can provide real-time input on strategy decisions, spot anomalies in the environment, and guide the model toward more stable allocations. This approach will help us identify and correct issues in the code or modeling assumptions before they lead to large portfolio losses.

3. Larger Bond Universe and Market Frictions. By including a broader array of bonds—varying in maturities, credit ratings, and sectors—we aim to offer the RL agent richer opportunities for diversification and hedging. Additionally, introducing market frictions such as bid-ask spreads, market impact, and partial fill constraints can bring our environment closer to real-world trading conditions.

4. Extended Regime Analysis. Although we used a regime-based approach in one of our baselines, future work could integrate regime detection more tightly with the RL agent. This might involve real-time updates to regime probabilities and dynamic policy shifts, offering the potential for more effective responses to market transitions.

Overall, while the results here were disappointing, they offer valuable lessons for improving the fidelity

of our simulation, refining the RL architecture, and exploring new strategies for risk management. We remain optimistic that, with a more realistic environment and the addition of expert feedback, reinforcement learning can become a powerful tool for managing fixed income portfolios.