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

Integrating Regime Detection, Graph Neural Networks, and Modern RL Algorithms

Corentin Servouze, Rany Stephan

March 21, 2025

Outline

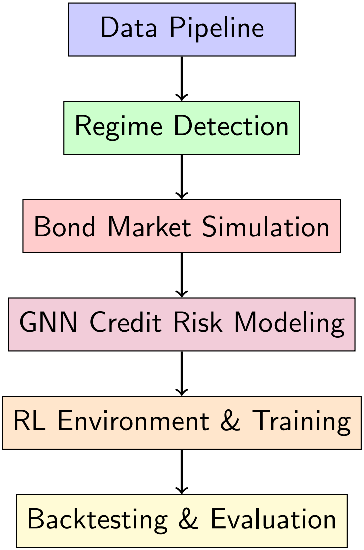
The Problem of Fixed Income Portfolio Management

- **Fixed income markets** represent over \$100 trillion globally
- Managing bond portfolios involves complex challenges:
  - Interest rate risk and credit risk management
  - Multi-dimensional features (duration, convexity, credit quality)
  - Regime-dependent dynamics (recession, expansion, crisis)
  - Limited liquidity compared to equities
- Traditional approaches:
  - Static allocation strategies (ladders, barbells)
  - Duration targeting
  - Factor-based or index-tracking methods

Project Innovation

- Comprehensive RL framework for fixed income portfolios that:
  - Integrates regime detection for regime-aware strategies
  - Incorporates credit risk through graph neural networks
  - Uses RL algorithms (TD3, DDPG)
  - Models realistic bond market dynamics
- Creates a complete simulation-to-deployment pipeline
- Addresses multi-faceted bond portfolio optimization with focus on risk-adjusted returns

Project Architecture



Prior Work in Fixed Income RL

- Limited research compared to equity RL applications
- Notable prior work:
  - Halperin & Feldshteyn (2018): Q-learning for fixed income trading
  - Kolm & Ritter (2019): Deep hedging with neural networks
  - Makinen et al. (2019): RL for corporate bond trading
- Limitations of existing approaches:
  - Often simplified market dynamics
  - Limited consideration of regime changes
  - Lack of network effects in credit risk
  - Focus on single bonds rather than portfolios

Interest Rate Models

- **Vasicek Model**:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t \tag{1}$$

  where:
  - r_t is the short rate at time t
  - κ is the mean reversion speed
  - θ is the long-term mean level
  - σ is the volatility
  - dW_t is a Wiener process increment
- Mean-reverting process capturing central tendency of rates
- Allows for negative rates (theoretical limitation)

Interest Rate Models (Continued)

- **Cox-Ingersoll-Ross (CIR) Model**:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t \tag{2}$$

  - Square root diffusion term ensures non-negative rates
  - Higher volatility when rates are higher
  - Same mean-reversion structure as Vasicek

- **Hull-White Model**:

$$dr_t = [\theta(t) - \kappa r_t]dt + \sigma dW_t \tag{3}$$

  - Time-dependent $\theta(t)$ function
  - Calibrated to match initial yield curve
  - Extension of Vasicek with greater flexibility

Credit Spread Models

- **Merton Model**:
  - Firm's asset value V follows geometric Brownian motion:

$$dV_t = rV_t dt + \sigma_V V_t dW_t \tag{4}$$

  - Default occurs if $V_T < D$ at debt maturity T
  - Credit spread:

$$s(t, T) = -\frac{1}{T - t} \ln(1 - N(-d_2)) \tag{5}$$

  where $d_2 = \frac{\ln(V_t/D) + (r - \frac{1}{2}\sigma_V^2)(T - t)}{\sigma_V \sqrt{T - t}}$

- Model parameters:
  - Asset value V_t
  - Face value of debt D
  - Asset volatility σ_V
  - Risk-free rate r

Bond Pricing

- **Zero-coupon bond price**:

$$P(t, T) = \frac{F}{(1 + y_{t,T})^{T-t}} \tag{6}$$

where F is face value, $y_{t,T}$ is yield

- **Coupon bond price**:

$$P(t, T) = \sum_{i=1}^{n} \frac{c \cdot F}{(1 + y_{t,T}/m)^{m(t_i - t)}} + \frac{F}{(1 + y_{t,T}/m)^{m(T-t)}} \tag{7}$$

where c is coupon rate, m is payments per year

- **Credit-risky bond**:

$$P(t, T) = \sum_{i=1}^{n} \frac{c \cdot F}{(1 + r_{t,t_i} + s_{t,t_i})^{t_i - t}} + \frac{F}{(1 + r_{t,T} + s_{t,T})^{T-t}} \tag{8}$$

where $r_{t,T}$ is risk-free rate, $s_{t,T}$ is credit spread

Bond Risk Metrics

- **Macaulay Duration**:

$$D = \frac{\sum_{t=1}^{T} t \cdot CF_t \cdot (1+y)^{-t}}{\sum_{t=1}^{T} CF_t \cdot (1+y)^{-t}} \quad (9)$$

- **Modified Duration**:

$$D_{mod} = \frac{D}{1+y} \quad (10)$$

- **Convexity**:

$$C = \frac{\sum_{t=1}^{T} t(t+1) \cdot CF_t \cdot (1+y)^{-t}}{P \cdot (1+y)^2} \quad (11)$$

- **Price sensitivity to yield changes**:

$$\frac{\Delta P}{P} \approx -D_{mod} \cdot \Delta y + \frac{1}{2} \cdot C \cdot (\Delta y)^2 \quad (12)$$

Metrics capture different aspects of interest rate risk

Economic Regimes in Fixed Income

- Different market regimes significantly impact fixed income:
  - **Normal regime**: Steady growth, stable rates
  - **Expansion regime**: Rising rates, flattening yield curve
  - **Stress regime**: Flight to quality, widening credit spreads
  - **Crisis regime**: Rate cuts, extreme volatility, liquidity issues
- **Importance for bond investors**:
  - Risk factors behave differently across regimes
  - Optimal allocations vary by regime
  - Regime shifts create both risks and opportunities
  - Traditional mean-variance assumptions break down during transitions

Hidden Markov Models for Regime Detection

- **Hidden Markov Model (HMM)** framework:
  - Observable features $\mathbf{X}_t$ (yields, spreads, etc.)
  - Hidden states (regimes) $z_t \in \{1, 2, ..., K\}$
  - Emission distributions $p(\mathbf{X}_t | z_t)$
  - Transition matrix A where $A_{ij} = p(z_t = j | z_{t-1} = i)$
- **Model specification**:

$$p(\mathbf{X}_t | z_t = k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{13}$$

- **Parameter estimation** via Expectation-Maximization:
  - E-step: Calculate posterior $p(z_t | X_{1:T})$ using forward-backward algorithm
  - M-step: Update $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, and A

Regime Characterization and Transitions

- **Statistical regime characteristics**:
  - Expected returns, volatilities, correlations
  - Yield curve shapes (normal, flat, inverted)
  - Average spread levels by rating
  - Typical duration of each regime (persistence)
- **Transition matrix visualization**:

	Normal	Expansion	Stress	Crisis
Normal	0.983	0.010	0.005	0.002
Expansion	0.015	0.975	0.008	0.002
Stress	0.008	0.005	0.967	0.020
Crisis	0.010	0.000	0.025	0.965

- **Regime forecasting**:
  - Short-term prediction via Markov property
  - Confidence metrics for regime identification

Integration with Market Simulator and RL

- **Market simulator integration**:
  - Regime-specific parameters for interest rate models
  - Regime-dependent credit spread dynamics
  - Transition probabilities for regime switching
- **RL environment integration**:
  - Current regime as part of state representation
  - Regime-specific reward scaling to account for different risk environments
  - Enables learning of regime-appropriate strategies
- **Benefits**:
  - More realistic simulation of market dynamics
  - Allows RL agent to learn regime-specific policies
  - Better generalization to different market conditions

Credit Risk and Network Effects

- **Traditional limitations**:
  - Credit ratings provide point-in-time assessments
  - Standard models treat issuers independently
  - Interconnections between companies often ignored
- **Network effects in credit markets**:
  - Supply chain dependencies
  - Counterparty relationships
  - Common exposures to risk factors
  - Contagion effects during crises
- **GNN advantage**: Can explicitly model these relationships

Graph Neural Network Formulation

- **Graph representation**:
  - Nodes: Bond issuers with features $\mathbf{X}_i$
  - Edges: Relationships between issuers
  - Target: Credit spreads or default probabilities
- **Message passing framework**:

$$\mathbf{h}_i^{(l+1)} = \text{UPDATE}\left(\mathbf{h}_i^{(l)}, \text{AGGREGATE}\left(\{\mathbf{h}_j^{(l)} : j \in \mathcal{N}(i)\}\right)\right) \quad (14)$$

where:

- $\mathbf{h}_i^{(l)}$ is the node representation at layer l
- $\mathcal{N}(i)$ is the neighborhood of node i
- AGGREGATE combines information from neighbors
- UPDATE incorporates aggregated information

GNN Model Architectures

- **Graph Convolutional Network (GCN)**:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{15}$$

  where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ and $\tilde{\mathbf{D}}$ is degree matrix

- **Graph Attention Network (GAT)**:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{ij}\mathbf{W}^{(l)}\mathbf{h}_j^{(l)}\right) \tag{16}$$

  with attention coefficients α_{ij} learned from data

- **Message Passing Neural Network (MPNN)**:

$$\mathbf{m}_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \mathrm{MSG}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{e}_{ij}) \tag{17}$$

$$\mathbf{h}_i^{(l+1)} = \mathrm{UPDATE}^{(l)}(\mathbf{h}_i^{(l)}, \mathbf{m}_i^{(l+1)}) \tag{18}$$

Credit Spread Prediction with GNN

- **Node features $\mathbf{X}_i$** for issuer i:
  - Financial ratios (debt/equity, interest coverage)
  - Market capitalization and volatility
  - Industry and sector indicators
  - Current credit rating
  - Historical spread volatility
- **Edge features $\mathbf{e}_{ij}$** between issuers i and j:
  - Strength of relationship
  - Type of connection (supply chain, competitor, etc.)
  - Correlation of historical spreads
- **Prediction target**:

$$\hat{s}_i = f_\theta(\mathbf{X}_i, \{\mathbf{X}_j, \mathbf{e}_{ij} : j \in \mathcal{N}(i)\}) \tag{19}$$

where $\hat{s}_i$ is predicted credit spread, f_θ is GNN

Node Embeddings and Integration with RL

- **Node embeddings** capture rich credit risk information:

$$\mathbf{z}_i = \mathbf{h}_i^{(L)} \in \mathbb{R}^d \tag{20}$$

  - $\mathbf{z}_i$ is low-dimensional embedding of issuer i
  - Encodes both issuer-specific and network information
  - Dimensionality d typically 32-128

- **Integration into RL state space**:

$$\mathbf{s}_t = [\mathbf{m}_t, \mathbf{r}_t, \mathbf{z}_{i_1}, \mathbf{z}_{i_2}, ..., \mathbf{z}_{i_n}] \tag{21}$$

  where:
  - $\mathbf{m}_t$ is market state (rates, economic indicators)
  - $\mathbf{r}_t$ is current regime
  - $\mathbf{z}_{i_1}, ..., \mathbf{z}_{i_n}$ are embeddings of bonds in investable universe

- Enriches state representation with network-aware credit risk information

RL Environment Design for Fixed Income

- **State space** $\mathcal{S}$:

$$\mathbf{s}_t = [\mathbf{m}_t, \mathbf{p}_t, \mathbf{h}_t, \mathbf{r}_t, \mathbf{z}_t] \tag{22}$$

  where:
  - $\mathbf{m}_t$: Market features (rates, spreads, volatility)
  - $\mathbf{p}_t$: Portfolio features (current weights, durations)
  - $\mathbf{h}_t$: Historical returns and features (lookback window)
  - $\mathbf{r}_t$: Regime indicator (one-hot encoded)
  - $\mathbf{z}_t$: GNN embeddings of issuers in universe
- **Dimensions**: Typically 500-700 features in total

RL Environment Design (Continued)

- **Action space** $\mathcal{A}$:

$$\mathbf{a}_t = [w_1, w_2, ..., w_n] \quad \text{s.t.} \quad \sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0 \tag{23}$$

  - w_i is portfolio weight for bond i
  - Continuous action space with simplex constraint
  - Dimensionality = number of bonds in universe
- **Transition dynamics**:

$$\mathbf{s}_{t+1} = f(\mathbf{s}_t, \mathbf{a}_t) \tag{24}$$

  - Based on bond market simulation
  - Incorporates regime transitions
  - Updates portfolio based on new weights and market movements

Reward Function Design

- **Multi-objective reward function**:

$$r_t = \alpha \cdot r_{\text{return}} + \beta \cdot r_{\text{risk}} + \gamma \cdot r_{\text{constraint}} - \delta \cdot r_{\text{cost}} \tag{25}$$

- **Component rewards**:

$$r_{\text{return}} = R_t \tag{26}$$

$$r_{\text{risk}} = -\sigma_t \tag{27}$$

$$r_{\text{constraint}} = -\sum_j \max(0, c_j(\mathbf{a}_t))^2 \tag{28}$$

$$r_{\text{cost}} = TC(\mathbf{a}_{t-1}, \mathbf{a}_t) \tag{29}$$

where:

- R_t is portfolio return
- σ_t is portfolio volatility
- c_j are constraint functions (e.g., duration limits)
- TC is transaction cost function

Deep Deterministic Policy Gradient (DDPG)

- **Actor-critic architecture** for continuous action spaces:
  - Actor $\mu_\theta(s)$: Deterministic policy mapping states to actions
  - Critic $Q_\phi(s, a)$: Action-value function estimator
- **Learning algorithm**:

$$\mathcal{L}_{\text{critic}} = \mathbb{E}_{(s,a,r,s')\sim\mathcal{D}}\left[(Q_\phi(s, a) - y)^2\right] \tag{30}$$

$$y = r + \gamma Q_{\phi'}(s', \mu_{\theta'}(s')) \tag{31}$$

$$\mathcal{L}_{\text{actor}} = -\mathbb{E}_{s\sim\mathcal{D}}\left[Q_\phi(s, \mu_\theta(s))\right] \tag{32}$$

  where ϕ' and θ' are parameters of target networks
- **Exploration** with Ornstein-Uhlenbeck process:

$$a_t = \mu_\theta(s_t) + \mathcal{N}_t \tag{33}$$

Twin Delayed Deep Deterministic Policy Gradient (TD3)

- **Improvements over DDPG**:
  - Twin critics to reduce overestimation bias
  - Delayed policy updates
  - Target policy smoothing
  - Clipped double Q-learning
- **Twin critics update**:

$$y = r + \gamma \min_{i=1,2} Q_{\phi'_i}(s', \mu_{\theta'}(s') + \epsilon) \tag{34}$$

$$\mathcal{L}_{\text{critic}_i} = \mathbb{E}_{(s,a,r,s')\sim\mathcal{D}} \left[\left(Q_{\phi_i}(s,a) - y \right)^2 \right] \tag{35}$$

where $\epsilon \sim \text{clip}(\mathcal{N}(0,\sigma), -c, c)$

- **Delayed policy updates**:

$$\nabla_\theta \mathcal{L}_{\text{actor}} = -\mathbb{E}_{s\sim\mathcal{D}} \left[\nabla_a Q_{\phi_1}(s,a)|_{a=\mu_\theta(s)} \nabla_\theta \mu_\theta(s) \right] \tag{36}$$

Updated every d critic updates (typically $d = 2$)

RL Training and Hyperparameter Tuning

- **Training procedure**:
  - Episode length: 252 steps (1 trading year)
  - Batch size: 64-128 transitions
  - Replay buffer size: 100,000 transitions
  - Learning rates: 1e-4 (actor), 1e-3 (critic)
  - Discount factor: 0.99
  - Target network update: $\tau = 0.001$ (soft updates)
- **Data efficiency techniques**:
  - Experience replay with prioritization
  - Random start points within simulation
  - Data augmentation through regime resampling
  - Curriculum learning (gradually increasing difficulty)
- **Evaluation metrics** during training:
  - Average return
  - Sharpe ratio
  - Constraint violation frequency
  - Portfolio turnover

Backtest Evaluation Framework

- **Benchmark strategies**:
  - Equal weight (naive diversification)
  - Market value weight (passive approach)
  - Duration targeting (fixed income standard)
  - Regime-based rule strategies (manually defined)
- **Performance metrics**:
  - Total return and volatility
  - Sharpe and Sortino ratios
  - Maximum drawdown
  - Regime-conditional performance
  - Turnover and transaction costs
- **Statistical significance tests**:
  - Bootstrap resampling
  - Spanning tests
  - Out-of-sample robustness checks

Key Takeaways

1. **Fixed income markets** benefit significantly from RL approaches due to their complex, regime-dependent dynamics and asymmetric risk profiles
2. **Regime detection** provides crucial context that improves both simulation realism and strategy performance
3. **Graph neural networks** capture issuer relationships and network effects in credit risk that traditional models miss
4. **Advanced RL algorithms** like TD3 handle the high-dimensional continuous action space effectively while managing constraints
5. **Integrated approach** combining multiple modeling techniques yields superior performance to any single method alone

References

Halperin, I., & Feldshteyn, I. (2018). Market self-learning of signals, impact and optimal trading: Invisible hand inference with free energy. arXiv preprint arXiv:1805.06126.

Kolm, P. N., & Ritter, G. (2019). Dynamic replication and hedging: A reinforcement learning approach. The Journal of Financial Data Science, 1(1), 159-171.

Makinen, Y., Kanniainen, J., Gabbouj, M., & Iosifidis, A. (2019). Forecasting jump arrivals in stock prices: new attention-based network architecture using limit order book data. Quantitative Finance, 19(12), 2033-2050.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. The Journal of finance, 29(2), 449-470.

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. ICLR.

Fujimoto, S., Hoof, H., & Meger, D. (2018). Addressing function approximation error in actor-critic methods. ICML.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.