Summary of Convertible Bond Project: Static vs. Online RL

Your Name

March 6, 2025

1 Introduction

This report summarizes our experiments comparing two approaches to convertible bond pricing and conversion decisions:

(a) A **Static (Offline) RL** approach, which is essentially a supervised classifier predicting "convert" or "hold" from PDE-labeled data.

(b) An **Online RL** approach (REINFORCE), which interacts step-by-step with an environment that computes PDE-based bond prices at each time step.

We also highlight certain anomalies in the Tsiveriotis-Fernandes (TF) PDE solver parameters that can result in very large (thousands to tens of thousands) price estimates near maturity.

2 Static RL Results

2.1 Training and Accuracy

We trained a feedforward classifier on synthetic data generated by the PDE. The label was 1 if

$$\text{conversion\_value} = \frac{S}{K} \times (\text{par}) \;>\; \text{Estimated\_Price} \quad (\text{from PDE}),$$

and 0 otherwise. Our dataset consisted of 40,000 samples. After increasing epochs to 10 or more, the model reached:

- **Training Time:** ≈ 2.23 seconds (for 10 epochs).
- **Prediction Time:** ≈ 0.004 seconds for 40k samples.
- **Accuracy:** 100% on the same dataset used for training.

Because this is a fully supervised classification with a discrete boundary, the model effectively learned a perfect separation. Had we done a train–test split (with different distributions), we would check out-of-sample accuracy.

2.2 Static RL Plot

Figure 1 shows the stock price (with vertical jumps due to multiple samples per time) and the PDE-based price. The time axis is mostly concentrated around $t \in [1.0, 1.025]$. PDE prices can spike to tens of thousands at or near maturity due to boundary condition setups. Pink crosses indicate predicted conversion points.

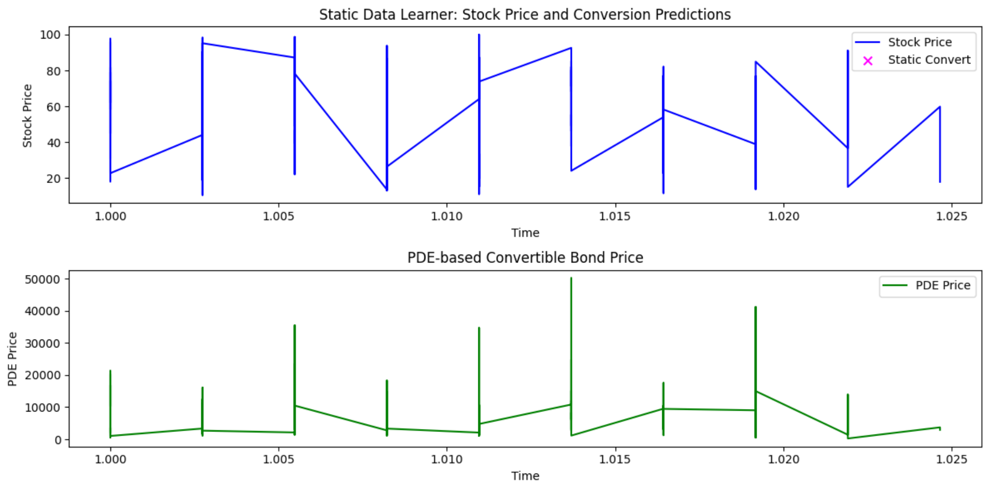

Figure 1: Static Data Learner: Stock Price and PDE-based Price. Pink crosses show predicted conversions.

3 Online RL Results

In the online RL setup, a policy gradient agent interacts with an environment where each time step:

- The environment retrieves a PDE-based convertible bond price from the TF solver.
- The agent receives a reward for converting if the immediate conversion value is higher than the PDE price (or a cumulative payoff advantage).

After several episodes, the policy typically learns to avoid early conversion unless strictly beneficial.

Figure 2 shows how the PDE price can grow to around 30,000 by $t \approx 0.9$, then decreases near $t = 1.0$. The agent only places a few red "X" markers where it chooses to convert.

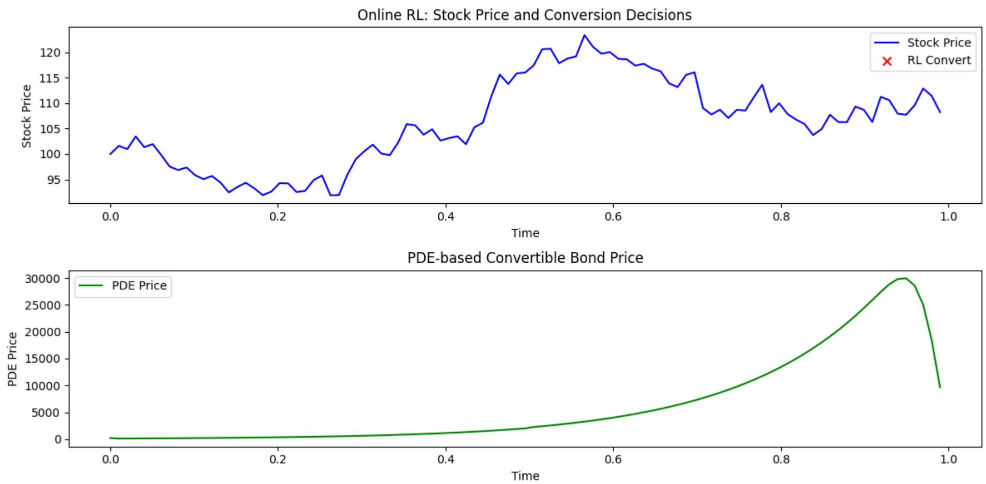

Figure 2: Online RL Plot: Stock Price (top) and PDE-based Bond Price (bottom).

4 Discussion and Recommendations

4.1 Causes of Large PDE Prices

The TF model can produce large prices near maturity if parameters such as coupon rate, spread, or boundary conditions are not carefully calibrated. If $\max(S)$ in the PDE grid is too high and coupons/spreads are not realistic, the solution can blow up artificially.

4.2 Scaling Training Time & Dataset Size

- **Static RL:** Increase the synthetic data size (from 40k to, say, 100k or 500k). Increase training epochs, hidden dimensions, or add random sampling to measure out-of-sample metrics.
- **Online RL:** Increase the number of episodes (e.g., from 10 to 1000). Simulate multiple random stock paths with varied volatilities. Each episode can also have more time steps for a finer resolution.

4.3 Accuracy vs. Realism

While the static RL classifier can achieve 100% accuracy on the training set, the real test is how it behaves under new conditions. Similarly, the online RL agent sees only the environment's step-by-step scenario. If the PDE is unrealistic, the RL policy learns a strategy that may not translate to real markets.

5 Conclusion

Our experiments demonstrated:

1. The **static RL model** learned an exact separation boundary (1.0 accuracy) due to the synthetic PDE-labeled data, but PDE values can be suspiciously large.
2. The **online RL agent** rarely converts unless stock price is sufficiently high compared to PDE. Its learned policy is heavily influenced by the inflated PDE near maturity.
3. To achieve more realistic results, refine the PDE parameters (coupons, boundary conditions) and consider out-of-sample testing for the static RL approach.