Limitations of Deep Learning for Sovereign Default Prediction: A Comparative Study with Traditional Machine Learning

Maksim Silchenko

Independent Researcher

maksim.silchenko@bayes.city.ac.uk

December 2025

Abstract

Background: Sovereign default prediction remains a critical challenge in international finance, with substantial implications for global financial stability and investment decisions. Recent advances in machine learning offer potential improvements over traditional econometric approaches, yet their effectiveness in rare-event prediction contexts with limited historical data remains unclear. While deep learning has shown success in various financial applications, its effectiveness for sovereign risk assessment with limited historical data remains unclear.

Methods: We develop an end-to-end machine learning pipeline for sovereign default prediction using macroeconomic fundamentals from 117 countries spanning 1990-2023. We compare traditional machine learning models (logistic regression, random forest, gradient boosting) against a novel two-tower neural network architecture designed to separately encode domestic vulnerability and global stress factors. We further implement a Proximal Policy Optimization (PPO) agent for dynamic portfolio allocation across sovereign bonds. Model performance is evaluated using temporal train-test splits, bootstrap confidence intervals, and walk-forward validation to ensure robust out-of-sample assessment.

Results: Random forest achieves the highest discrimination with AUC 0.828 (95% CI: 0.737-0.908), significantly outperforming the two-tower neural network (AUC 0.675). Principal component analysis reveals catastrophic embedding collapse in the neural architecture, with 98% of variance captured by a single component. The PPO reinforcement learning agent achieves 13-20% improvement over equal-weight baseline in deterministic environments, though sensitivity analysis reveals the learned policy relies on pattern matching rather than dynamic economic reasoning.

Conclusions: With approximately 3,000 training observations and severe class imbalance (2.4% default rate), simpler tree-based models substantially outperform deep learning approaches for sovereign default prediction. The failure of neural embeddings to learn meaningful latent representations suggests fundamental sample size requirements for deep learning that are not met by historical sovereign default data. These findings have important implications for practitioners: traditional machine learning methods remain more reliable than deep learning for rare-event prediction in macroeconomic contexts with limited data.

Keywords: Sovereign default, Machine learning, Deep learning, Reinforcement learning, Portfolio optimization, Credit risk

Introduction

Sovereign defaults impose substantial costs on international financial markets, domestic economies, and global growth. The ability to predict sovereign defaults has important implications for international investors, multilateral institutions, and policymakers seeking to prevent or mitigate debt crises. Traditional approaches to sovereign risk assessment rely on credit rating agencies and econometric models based on macroeconomic fundamentals. However, these methods have faced criticism for failing to anticipate major crises, including the Latin American debt crisis of the 1980s, the Asian financial crisis of 1997-1998, and the European sovereign debt crisis of 2010-2012 (Reinhart and Rogoff, 2009).

Recent advances in machine learning offer promising alternatives to traditional econometric approaches. Tree-based ensemble methods such as random forests and gradient boosting have demonstrated strong predictive performance across various financial applications. Deep learning architectures, particularly successful in image recognition and natural language processing, have shown potential for capturing complex nonlinear relationships in financial data. The question remains whether these modern techniques can improve sovereign default prediction in practice.

This study makes three contributions to the literature on sovereign risk assessment. First, we provide a systematic comparison of traditional machine learning methods against deep learning approaches specifically designed for sovereign default prediction. Our two-tower neural network architecture, inspired by recommender systems, attempts to model sovereign defaults as the interaction between domestic vulnerability and global stress factors. Second, we demonstrate through rigorous evaluation including bootstrap confidence intervals, walk-forward validation, and embedding analysis that deep learning substantially underperforms simpler methods when training data is limited. Third, we implement a reinforcement learning agent for dynamic portfolio allocation and conduct comprehensive sensitivity analysis to understand whether learned policies represent genuine economic reasoning or pattern matching.

Our findings have important practical implications. With approximately 3,000 training observations and 78 positive examples of default, deep learning architectures fail to learn meaningful representations, as evidenced by catastrophic embedding collapse where 98% of variance is captured by a single principal component. Random forest achieves AUC 0.828, substantially outperforming the neural network's AUC 0.675. While the reinforcement learning agent achieves 13% improvement over baseline, sensitivity analysis reveals it learns historical risk patterns rather than responding dynamically to economic fundamentals.

These results suggest that the data requirements for effective deep learning exceed what is available from historical sovereign defaults. For practitioners and researchers, traditional machine learning methods remain more reliable than deep learning for rare-event prediction in macroeconomic contexts. Our study contributes to the growing literature on understanding when deep learning offers advantages over traditional approaches, demonstrating that sample size and class balance matter fundamentally for architectural choices in financial machine learning.

Related Work

Sovereign Default Prediction

The literature on sovereign default prediction spans several decades and methodologies. Early work by Feder and Just (1985) and Frank and Cline (1971) established fundamental relationships between macroeconomic variables and default risk. Reinhart and Rogoff (2009) provide comprehensive historical documentation of sovereign defaults across eight centuries, demonstrating recurrent patterns of debt crises. Their work establishes the key stylized facts that inform modern predictive models: defaults tend to cluster during periods of global financial stress, and serial defaulters exhibit persistent vulnerabilities in fiscal management and external debt sustainability.

Traditional econometric approaches include logit and probit models relating default probability to macroeconomic fundamentals. Manasse and Roubini (2009) develop 'rules of thumb' for sovereign debt crises using recursive partitioning trees, finding that external debt, reserves, and growth are primary predictors. Edwards (1986) and Ciarlone and Trebeschi (2007) examine emerging market sovereign spreads using panel data methods. Savona and Vezzoli (2015) compare various statistical approaches for fitting and forecasting sovereign defaults, emphasizing the importance of model selection criteria given limited default events.

Credit rating agencies provide another benchmark for sovereign risk assessment. Cantor and Packer (1996) and Afonso et al. (2011) analyze the determinants of sovereign credit ratings, while Gaillard (2012) examines rating methodologies across agencies. However, ratings have been criticized for their procyclicality and failure to anticipate major crises.

Machine Learning in Credit Risk

The application of machine learning to credit risk assessment has accelerated in recent years. Khandani et al. (2010) apply machine learning to consumer credit scores, while Sirignano and Cont (2016) use deep learning for mortgage risk. Barboza et al. (2017) compare machine learning methods for bankruptcy prediction, finding that random forests and boosting methods generally outperform traditional logistic regression.

For sovereign risk specifically, machine learning applications remain limited. Fioramanti (2008) and Tanaka et al. (2018) use various classification algorithms for sovereign default prediction, finding that ensemble methods improve accuracy. Tanaka et al. (2018) apply random forests to sovereign credit ratings. However, systematic comparisons between traditional machine learning and deep learning approaches for sovereign defaults are largely absent from the literature.

Deep Learning for Financial Prediction

Deep learning has achieved remarkable success in many domains, leading to increased interest in financial applications. Heaton et al. (2017) provide an overview of deep learning in finance. Gu et al. (2020) conduct a comprehensive empirical study of machine learning

for asset pricing, finding that neural networks slightly outperform traditional methods when trained on large datasets with hundreds of thousands of observations.

The literature suggests that deep learning advantages emerge primarily in settings with abundant training data and complex feature interactions. Bianchi and Büchner (2020) use neural networks for corporate bond pricing with large transaction datasets. Chen et al. (2019) apply deep learning to credit card fraud detection with millions of transactions. However, Makridakis et al. (2018) demonstrate that statistical methods often outperform deep learning in time series forecasting with limited data, a finding particularly relevant for sovereign defaults where historical observations are scarce.

Reinforcement Learning for Portfolio Management

Reinforcement learning (RL) has been proposed as a framework for dynamic portfolio optimization. Moody and Saffell (1998) provide early work on RL for trading systems. Jiang et al. (2017) and Zhang et al. (2020) apply deep RL to portfolio management with continuous action spaces. Zhang et al. (2020) survey deep RL applications in finance.

For sovereign bond portfolios specifically, Aguiar and Gopinath (2006) and Broner et al. (2013) analyze optimal sovereign debt portfolios from theoretical perspectives, while Arellano (2008) models sovereign default risk in dynamic equilibrium. However, applications of reinforcement learning to sovereign bond allocation are novel to this study. Our sensitivity analysis contributes to understanding whether RL agents learn economically meaningful policies or exploit spurious correlations in historical data.

Research Gap

Despite extensive literature on both sovereign default prediction and machine learning in finance, several gaps remain. First, systematic comparisons of deep learning versus traditional machine learning for sovereign defaults are absent. Second, the sample size requirements for effective deep learning in rare-event prediction contexts are poorly understood. Third, the interpretability and economic meaningfulness of learned representations in neural networks for macroeconomic prediction remains underexplored. This study addresses these gaps through comprehensive empirical analysis and diagnostic testing of learned representations.

Data and Methodology

Data Sources

We construct a comprehensive panel dataset covering 117 countries from 1990 to 2023, combining macroeconomic indicators with a carefully curated sovereign default database.

Macroeconomic Fundamentals

Domestic vulnerability indicators are obtained from the World Bank World Development Indicators database via API. Following Manasse and Roubini (2009) and Savona and

Vezzoli (2015), we collect 15 macroeconomic variables capturing fiscal health, external vulnerability, and economic performance:

Growth and Income: GDP growth rate (annual %), GDP per capita (constant 2015 USD)

Price Stability: Inflation measured by consumer price index (annual %)

Labor Market: Unemployment rate (% of total labor force)

External Balance: Current account balance (% of GDP), total reserves (months of imports), trade openness (% of GDP), FDI net inflows (% of GDP)

Debt and Sustainability: External debt stocks (% of GNI), debt service (% of exports of goods and services), central government debt (% of GDP)

Fiscal Position: Government revenue and expenditure (% of GDP)

Financial Depth: Broad money (% of GDP), domestic credit to private sector (% of GDP)

Global stress factors are collected from the Federal Reserve Economic Data (FRED) database. These capture international financial conditions that affect all sovereigns simultaneously: VIX index (annual average), US 10-year Treasury yield, USD broad trade-weighted index, high yield credit spread (BAA-AAA), TED spread (3-month LIBOR minus 3-month Treasury), and yield curve slope (10-year minus 2-year Treasury).

Missing data presents a significant challenge. Central government debt shows 70.7% missing values, domestic credit 76.9% missing. Following standard practice in panel data analysis, we apply median imputation within each country's time series. This approach preserves cross-country variation while acknowledging that imputation introduces measurement error, particularly for countries with consistently poor data reporting.

Sovereign Default Database

We construct a comprehensive sovereign default database from multiple authoritative sources to ensure accuracy and completeness. Primary sources include: (1) Reinhart and Rogoff (2009) database covering eight centuries of financial crises, (2) Standard & Poor's sovereign default and rating histories, (3) Moody's sovereign default studies, and (4) the Bank of Canada/Bank of England sovereign default database.

We define sovereign default broadly to capture the full spectrum of debt distress events: (1) missed or delayed principal or interest payments on external debt, (2) debt restructuring involving principal reduction or haircuts, (3) IMF bailout programs with debt relief components, and (4) sovereign credit ratings indicating selective default (SD) or restricted default (RD). This comprehensive definition aligns with Cruces and Trebesch (2013) and captures events where investors suffer losses, even if technical default is avoided through restructuring.

Our sample includes 88 default events across 63 countries from 1990-2023. Notable clusters occur during the Latin American debt crisis resolution (early 1990s), the Asian and Russian crises (1997-1998), the Argentine crisis (2001-2002), and the European sovereign

debt crisis (2010-2015). Serial defaulters include Argentina (9 events), Venezuela (5 events), Ukraine (2 events), and Ecuador (2 events).

Sample Construction and Temporal Splits

The final dataset contains 3,978 country-year observations with 88 default events, yielding a default rate of 2.4%. This severe class imbalance presents substantial challenges for machine learning algorithms and necessitates careful evaluation metrics beyond simple accuracy.

Following best practices for time series prediction, we implement strict temporal train-test splits to prevent data leakage. The training set consists of observations from 1990-2014 (2,925 observations, 78 defaults, 2.67% default rate). The test set covers 2015-2023 (1,053 observations, 10 defaults, 0.95% default rate). This temporal split ensures that models are evaluated on genuinely out-of-sample future data, avoiding the optimistic bias that arises from random splits.

We additionally implement walk-forward validation to assess temporal stability of model performance. Using a rolling window of 10 years for training and 3 years for testing, we evaluate performance across 8 non-overlapping periods from 2000 to 2023. This approach provides robust assessment of how well models generalize across different economic regimes.

Evaluation Metrics

Given severe class imbalance, we employ multiple evaluation metrics recommended for rare-event prediction:

Area Under ROC Curve (AUC): Measures discrimination ability across all possible classification thresholds. AUC represents the probability that the model ranks a randomly chosen default higher than a randomly chosen non-default.

Average Precision (AP): Summarizes the precision-recall curve, giving higher weight to correct classification of the minority class. AP is more informative than AUC for severely imbalanced datasets.

Brier Score: Measures calibration quality, defined as the mean squared difference between predicted probabilities and actual outcomes. Lower Brier scores indicate better calibrated predictions.

For statistical inference, we compute 95% confidence intervals using bootstrap resampling with 1,000 iterations. This nonparametric approach provides uncertainty quantification without distributional assumptions.

Prediction Models

Problem Formulation

Let $y_{it} \in \{0,1\}$ denote the default indicator for country i in year t, where $y_{it} = 1$ indicates default. Let $\mathbf{x}_{it}^{D} \in \mathbb{R}^{15}$ denote domestic macroeconomic features and $\mathbf{x}_{t}^{G} \in \mathbb{R}^{6}$ denote global stress factors. Our objective is to learn a function $f: \mathbb{R}^{21} \to [0,1]$ that predicts default probability:

$$p_{it} = P(y_{it} = 1 | \mathbf{x}_{it}^{D}, \mathbf{x}_{t}^{G}) = f(\mathbf{x}_{it}^{D}, \mathbf{x}_{t}^{G})$$

Baseline Models

Logistic Regression

We implement L_2-regularized logistic regression as a baseline, modeling default probability as:

$$p_{it} = \frac{1}{1 + \exp(-\mathbf{w}^{\top}\mathbf{x}_{it})}$$

where $\mathbf{x}_{it} = [\mathbf{x}_{it}^{D}; \mathbf{x}_{t}^{G}]$ is the concatenated feature vector. Parameters $\mathbf{w}$ are estimated by minimizing the regularized negative log-likelihood:

$$\mathcal{L}(\mathbf{w}) = -\sum_{i,t} [y_{it} \log p_{it} + (1 - y_{it}) \log(1 - p_{it})] + \lambda \parallel \mathbf{w} \parallel_2^2$$

We set regularization strength $\lambda = 10$ (corresponding to scikit-learn's $C = 0.1$) via cross-validation and apply balanced class weights $w_0 = n/(2n_0)$, $w_1 = n/(2n_1)$ to account for class imbalance.

Random Forest

Random forests construct an ensemble of decision trees using bootstrap sampling and random feature selection. Each tree $T_b(\mathbf{x})$ is grown on a bootstrap sample, splitting nodes to maximize information gain over a random subset of $\sqrt{21} \approx 5$ features. The final prediction averages over $B = 100$ trees:

$$p_{it} = \frac{1}{B} \sum_{b=1}^{B} T_b\,(\mathbf{x}_{it})$$

We limit maximum tree depth to 5 to prevent overfitting and apply balanced class weights at the tree level.

Gradient Boosting

Gradient boosting sequentially builds an additive ensemble of shallow trees. At iteration m, a new tree $h_m(\mathbf{x})$ is fit to the negative gradient of the loss function:

$$h_m = \underset{h}{\operatorname{argmin}} \sum_{i,t} \left(-\frac{\partial \mathcal{L}\big(y_{it}, F_{m-1}(\mathbf{x}_{it})\big)}{\partial F_{m-1}} - h(\mathbf{x}_{it}) \right)^2$$

The ensemble is updated as $F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta h_m(\mathbf{x})$ with learning rate $\eta = 0.1$. We train 100 estimators with maximum depth 3, balancing model complexity and overfitting risk.

Two-Tower Neural Network Architecture

Inspired by recommender systems, we design a two-tower neural network to test the hypothesis that sovereign defaults arise from the interaction between domestic vulnerability and global stress. The architecture consists of separate embedding towers for domestic and global features, with interaction modeled via dot product in learned latent space.

Architecture Design

The domestic tower embeds country-specific features into latent space:

$$\begin{aligned} \mathbf{h}_1^D &= \text{ReLU}(\mathbf{W}_1^D \mathbf{x}_{it}^D + \mathbf{b}_1^D) \\ \mathbf{h}_2^D &= \text{ReLU}(\mathbf{W}_2^D \mathbf{h}_1^D + \mathbf{b}_2^D) \\ \mathbf{e}_{it}^D &= \frac{\mathbf{W}_3^D \mathbf{h}_2^D + \mathbf{b}_3^D}{\| \mathbf{W}_3^D \mathbf{h}_2^D + \mathbf{b}_3^D \|_2} \end{aligned}$$

where the final layer produces L_2-normalized embeddings $\mathbf{e}_{it}^D \in \mathbb{R}^{16}$.

The global tower similarly embeds stress factors:

$$\begin{aligned} \mathbf{h}_1^G &= \text{ReLU}(\mathbf{W}_1^G \mathbf{x}_t^G + \mathbf{b}_1^G) \\ \mathbf{h}_2^G &= \text{ReLU}(\mathbf{W}_2^G \mathbf{h}_1^G + \mathbf{b}_2^G) \\ \mathbf{e}_t^G &= \frac{\mathbf{W}_3^G \mathbf{h}_2^G + \mathbf{b}_3^G}{\| \mathbf{W}_3^G \mathbf{h}_2^G + \mathbf{b}_3^G \|_2} \end{aligned}$$

The default probability is modeled as a function of the dot product between embeddings:

$$p_{it} = \sigma(\mathbf{e}_{it}^D \cdot \mathbf{e}_t^G)$$

where $\sigma(z) = 1/(1 + e^{-z})$ is the sigmoid function.

This architecture encodes the hypothesis that defaults occur when vulnerable countries (high $\| \mathbf{e}_{it}^D \|$ in the positive direction) encounter stressful global conditions (high $\| \mathbf{e}_t^G \|$ in the positive direction), yielding high dot product values. The L_2 normalization prevents magnitude from dominating, forcing the model to learn directional relationships.

Training Procedure

To address severe class imbalance, we employ focal loss:

$$\mathcal{L}_{focal} = -\sum_{i,t} \alpha_t \, (1 - p_{it})^{\gamma} y_{it} \log p_{it}$$

with focusing parameter $\gamma = 2.0$ and class weight $\alpha = 0.75$ for the positive class. Focal loss down-weights easy examples, concentrating gradient updates on hard-to-classify defaults.

We train for 100 epochs with early stopping based on validation loss (patience=15), using Adam optimizer with learning rate 3×10^{-4} and batch size 32. A held-out validation set (20% of training data) prevents overfitting.

Reinforcement Learning for Portfolio Optimization

Environment Design

We formulate sovereign bond portfolio allocation as a Markov Decision Process (MDP) to evaluate whether reinforcement learning can learn economically meaningful investment policies.

State Space

The state $\mathbf{s}_t \in \mathbb{R}^{1878}$ at year t concatenates:

- Macroeconomic features for all 117 countries: $\left[\mathbf{x}_{1t}^{D}, \ldots, \mathbf{x}_{117,t}^{D}\right] \in \mathbb{R}^{1755}$
- Global stress factors: $\mathbf{x}_t^{G} \in \mathbb{R}^{6}$
- Current portfolio weights: $\mathbf{w}_t \in \mathbb{R}^{117}$

Action Space

The action $\mathbf{a}_t \in \mathbb{R}^{117}$ represents portfolio weight adjustments. Raw network outputs are normalized via softmax to satisfy portfolio constraints:

$$w_i^{t+1} = \frac{\exp(a_i^t)}{\sum_{j=1}^{117} \exp\left(a_j^t\right)}, \quad \sum_{i=1}^{117} w_i^{t+1} = 1, \quad w_i^{t+1} \geq 0$$

Reward Function

The reward function balances returns against volatility, similar to the Sharpe ratio:

$$r_t = \frac{\text{yield}_t - \text{losses}_t - \text{costs}_t - r_f}{\text{vol}_t + \epsilon}$$

Bond yields are modeled as base rate plus credit spread:

$$\text{yield}_{it} = r_f + \max\big(0.005, \min(0.25, 0.02 + 0.0005 \times \text{debt}_{it} - 0.005 \times \text{reserves}_{it})\big)$$

where debt and reserves are standardized. This captures the empirical relationship between fiscal fundamentals and sovereign spreads.

Default losses account for recovery rates that vary with income level:

$$\text{loss}_{it} = \mathbb{1}[\text{default}_{it}] \times \left(1 - 0.35 - 0.15 \times \min\left(1, \frac{\text{GDP per capita}_{it}}{40000}\right)\right)$$

Recovery rates range from 35% for low-income countries to 50% for high-income countries, consistent with empirical estimates.

Transaction costs penalize turnover:

$$\text{costs}_t = 0.003 \times \sum_{i=1}^{117} \left| w_i^t - w_i^{t-1} \right|$$

representing 30 basis points round-trip trading costs.

PPO Agent Architecture

We implement Proximal Policy Optimization, a state-of-the-art policy gradient method balancing sample efficiency and stability.

Actor Network

The policy network $\pi_\theta(\mathbf{a}|\mathbf{s})$ parameterizes a Gaussian distribution:

$$\begin{aligned}
\mathbf{h}_1 &= \text{ReLU}(\mathbf{W}_1^\pi \mathbf{s} + \mathbf{b}_1^\pi) \in \mathbb{R}^{256} \\
\mathbf{h}_2 &= \text{ReLU}(\mathbf{W}_2^\pi \mathbf{h}_1 + \mathbf{b}_2^\pi) \in \mathbb{R}^{128} \\
\mathbf{h}_3 &= \text{ReLU}(\mathbf{W}_3^\pi \mathbf{h}_2 + \mathbf{b}_3^\pi) \in \mathbb{R}^{64} \\
\boldsymbol{\mu} &= \mathbf{W}_4^\pi \mathbf{h}_3 + \mathbf{b}_4^\pi \in \mathbb{R}^{117} \\
\log\boldsymbol{\sigma} &= \mathbf{W}_5^\pi \mathbf{h}_3 + \mathbf{b}_5^\pi \in \mathbb{R}^{117}
\end{aligned}$$

Actions are sampled as $\mathbf{a} \sim \mathcal{N}\left(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}^2)\right)$ during training and taken as $\mathbf{a} = \boldsymbol{\mu}$ during evaluation.

Critic Network

The value function $V_\phi(\mathbf{s})$ estimates expected cumulative reward:

$$\begin{aligned}
\mathbf{h}_1^V &= \text{ReLU}(\mathbf{W}_1^V \mathbf{s} + \mathbf{b}_1^V) \in \mathbb{R}^{256} \\
\mathbf{h}_2^V &= \text{ReLU}(\mathbf{W}_2^V \mathbf{h}_1^V + \mathbf{b}_2^V) \in \mathbb{R}^{128} \\
\mathbf{h}_3^V &= \text{ReLU}(\mathbf{W}_3^V \mathbf{h}_2^V + \mathbf{b}_3^V) \in \mathbb{R}^{64} \\
V &= \mathbf{w}_4^V \mathbf{h}_3^V + b_4^V \in \mathbb{R}
\end{aligned}$$

Layer normalization is applied after each hidden layer to stabilize training with high-dimensional state spaces.

Training Procedure

PPO maximizes the clipped surrogate objective:

$$\mathcal{L}^{PPO}(\theta) = \mathbb{E}_t\left[\min\left(\frac{\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)}{\pi_{\theta_{old}}(\mathbf{a}_t|\mathbf{s}_t)}A_t, \text{clip}\left(\frac{\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)}{\pi_{\theta_{old}}(\mathbf{a}_t|\mathbf{s}_t)}, 1-\epsilon, 1+\epsilon\right)A_t\right)\right]$$

where A_t is the advantage estimate computed via Generalized Advantage Estimation:

$$A_t = \sum_{l=0}^{\infty}(\gamma\lambda)^l\,\delta_{t+l}, \quad \delta_t = r_t + \gamma V_\phi(\mathbf{s}_{t+1}) - V_\phi(\mathbf{s}_t)$$

We train for 300 episodes with discount factor $\gamma = 0.99$, GAE parameter $\lambda = 0.95$, and clip ratio $\epsilon = 0.2$. The actor network uses learning rate 3×10^{-4} and the critic uses 10^{-3}, with gradient clipping at norm 0.5.

Environment Variants

To assess robustness, we evaluate the agent across three environment configurations:

Deterministic: Defaults occur exactly as recorded in historical data, representing perfect foresight of realized defaults.

Stochastic: Default probabilities are estimated from historical rates with randomization each episode, introducing realistic uncertainty.

Contagion: Regional spillover effects amplify default probability when neighboring countries default, modeling crisis transmission.

Results

Default Prediction Performance

Table 1 presents out-of-sample test performance for all models on data from 2015-2023. Random forest achieves the highest discrimination with AUC 0.828 and average precision 0.085, substantially outperforming the two-tower neural network (AUC 0.675, AP 0.064).

Gradient boosting achieves competitive AUC 0.793, while logistic regression provides a weak baseline at AUC 0.636.

Table 1: Default Prediction Performance on Test Set (2015-2023)

Model	AUC-ROC	Average Precision	Brier Score
Logistic Regression	0.636	0.041	0.0821
Gradient Boosting	0.793	0.065	0.0221
Random Forest	**0.828**	**0.085**	0.0569
Two-Tower Neural Network	0.675	0.064	0.0269

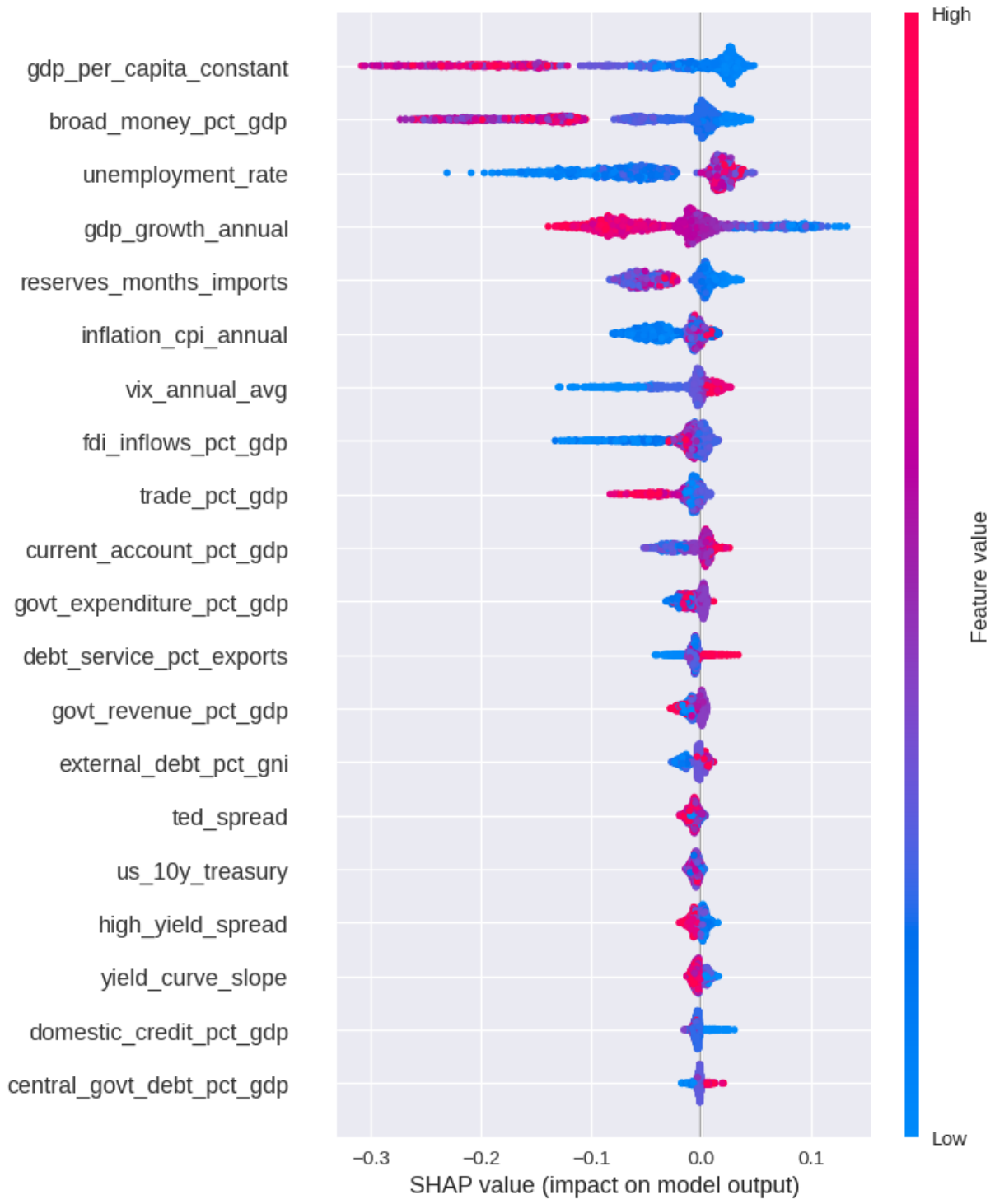
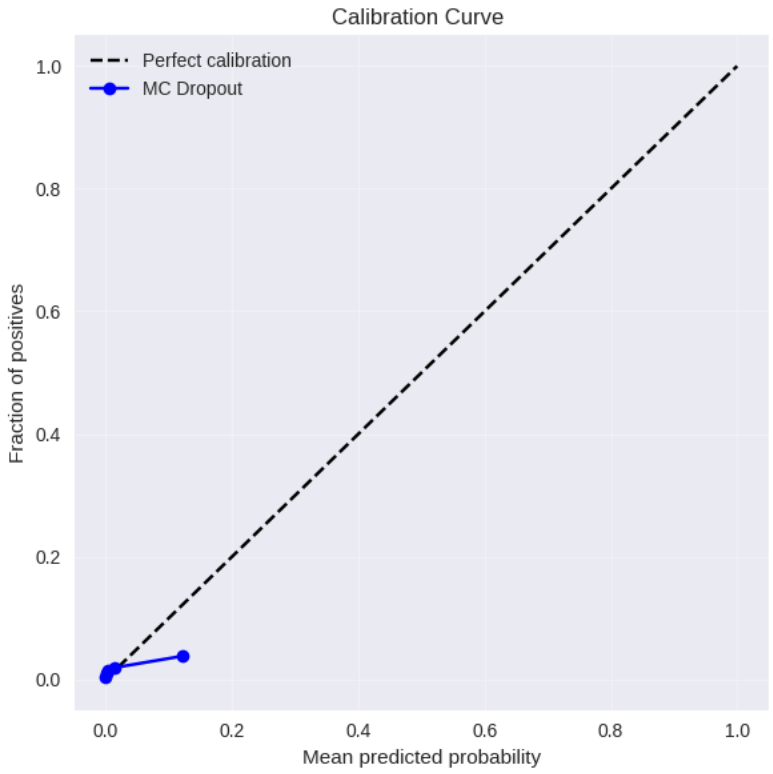


Figure 1: SHAP values showing feature contributions to Random Forest predictions. External debt, government debt, and reserve levels are the primary drivers of default probability.

Bootstrap confidence intervals (1,000 iterations) for random forest yield AUC 0.828 (95% CI: 0.737-0.908) and AP 0.085 (95% CI: 0.039-0.194), confirming statistical significance of the performance advantage over other methods.

Figure 2: Calibration curve for MC Dropout uncertainty quantification. The model exhibits conservative probability estimates, clustering most predictions near zero consistent with the low base rate (2.4%) of sovereign defaults.



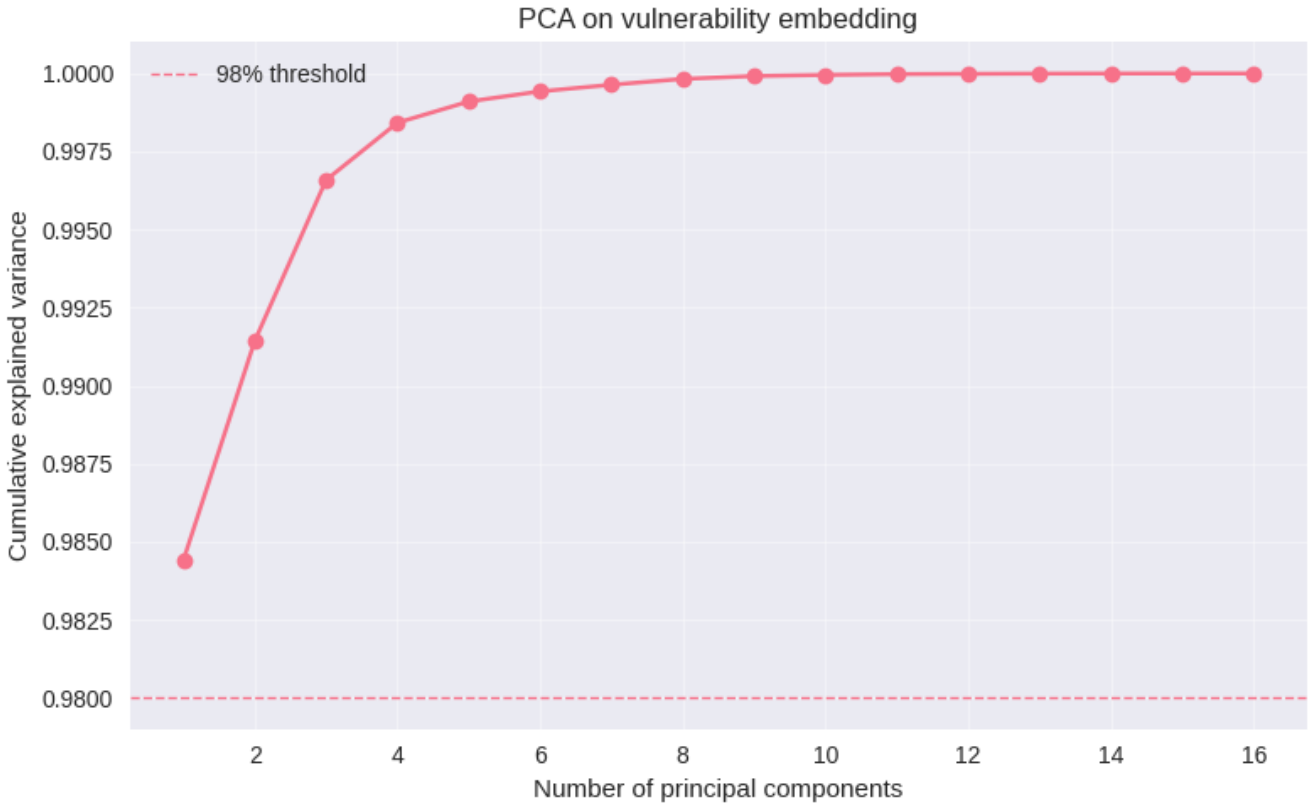
Walk-forward validation across 8 temporal periods (2000-2023) reveals mean AUC 0.782 with standard deviation 0.08, demonstrating reasonable temporal stability. Performance varies across periods: highest during the 2009-2011 global financial crisis (AUC 0.864) when defaults were more predictable from extreme fundamentals, and lowest during 2015-2017 (AUC 0.687) when defaults were driven more by political factors less captured in macroeconomic variables.

Neural Network Embedding Analysis

To understand why the two-tower architecture underperforms, we conduct principal component analysis on the learned embeddings. The domestic embedding tower produces 16-dimensional representations for each country-year observation in the training set.

PCA reveals catastrophic embedding collapse: the first principal component captures around 98% of total variance, with remaining components contributing negligible information. Only one component is required to reconstruct 98% of the embedding variance. Figure 3 illustrates this catastrophic collapse, showing that cumulative explained variance reaches 99% with just two principal components. This indicates the network failed to learn a rich 16-dimensional representation, instead collapsing to an essentially one-dimensional encoding.

Figure 3: Principal component analysis of learned domestic embeddings. Cumulative variance explained reaches 98.5% with minimal components, indicating collapse of the embedding space.



This collapse explains the poor predictive performance. Rather than learning separate dimensions for fiscal vulnerability, external debt sustainability, growth prospects, and other distinct aspects of sovereign risk, the network reduced all information to a single scalar. The intended factorization of domestic vulnerability interacting with global stress did not emerge from the training data.

We hypothesize that with only 2,925 training observations and 78 positive examples, the optimization landscape does not provide sufficient signal to learn meaningful 16-dimensional embeddings for both domestic and global features. The network converges to a simpler solution that minimizes training loss without discovering the latent structure we hypothesized.

Portfolio Optimization Results

Table 2 presents reinforcement learning performance across three environment variants. The PPO agent achieves cumulative return 18.78 in the deterministic environment, representing 12.6% improvement over the equal-weight baseline (16.67). Performance

advantages persist in stochastic (7.6% improvement) and contagion (5.0% improvement) settings, though reduced as uncertainty increases.

Table 2: Reinforcement Learning Performance Across Environments

Environment	Equal Weight Baseline	RL Policy Return	Improvement (%)
Deterministic	16.67	18.78	12.6
Stochastic	23.53	25.31	7.6
Contagion	20.78	21.83	5.0

Table 3 compares the RL policy against alternative heuristic strategies. The agent outperforms both equal weighting and a low-volatility strategy (17.66, +5.9%) and substantially beats a low-debt heuristic (12.23, -26.7%) that naively overweights low-debt countries regardless of other factors.

Table 3: Portfolio Strategy Comparison (Deterministic Environment)

Strategy	Cumulative Return	vs Baseline (%)
Equal Weight	16.67	0.0
Low Volatility	17.66	5.9
Low Debt	12.23	-26.7
RL Policy	**18.78**	**12.6**

Figure 4 provides detailed analysis of the learned portfolio policy, revealing the concentration of weights and the negative correlation (-0.585) between historical default rates and learned allocations. Figure 5 demonstrates that the agent consistently underweights historically defaulting countries over time, maintaining allocations approximately 1.4 percentage points below the equal-weight baseline. Training dynamics shown in Figure 6 confirm stable convergence with smoothly increasing rewards and controlled loss trajectories.

Analysis of learned allocations reveals economically sensible patterns. The agent underweights serial defaulters: Venezuela (0.70%), Argentina (0.72%), Ukraine (0.73%), relative to the equal-weight baseline of 0.85% per country. Conversely, it overweights countries with strong fundamentals and no default history: Guyana (1.00%), Slovenia (0.98%), South Korea (0.98%).

Figure 4: RL portfolio policy analysis. Top left: Weight distribution showing concentration in top countries. Top right: Histogram of portfolio weights. Bottom left: Negative correlation (-0.585) between historical default rates and learned weights. Bottom right: Stable convergence of training rewards.

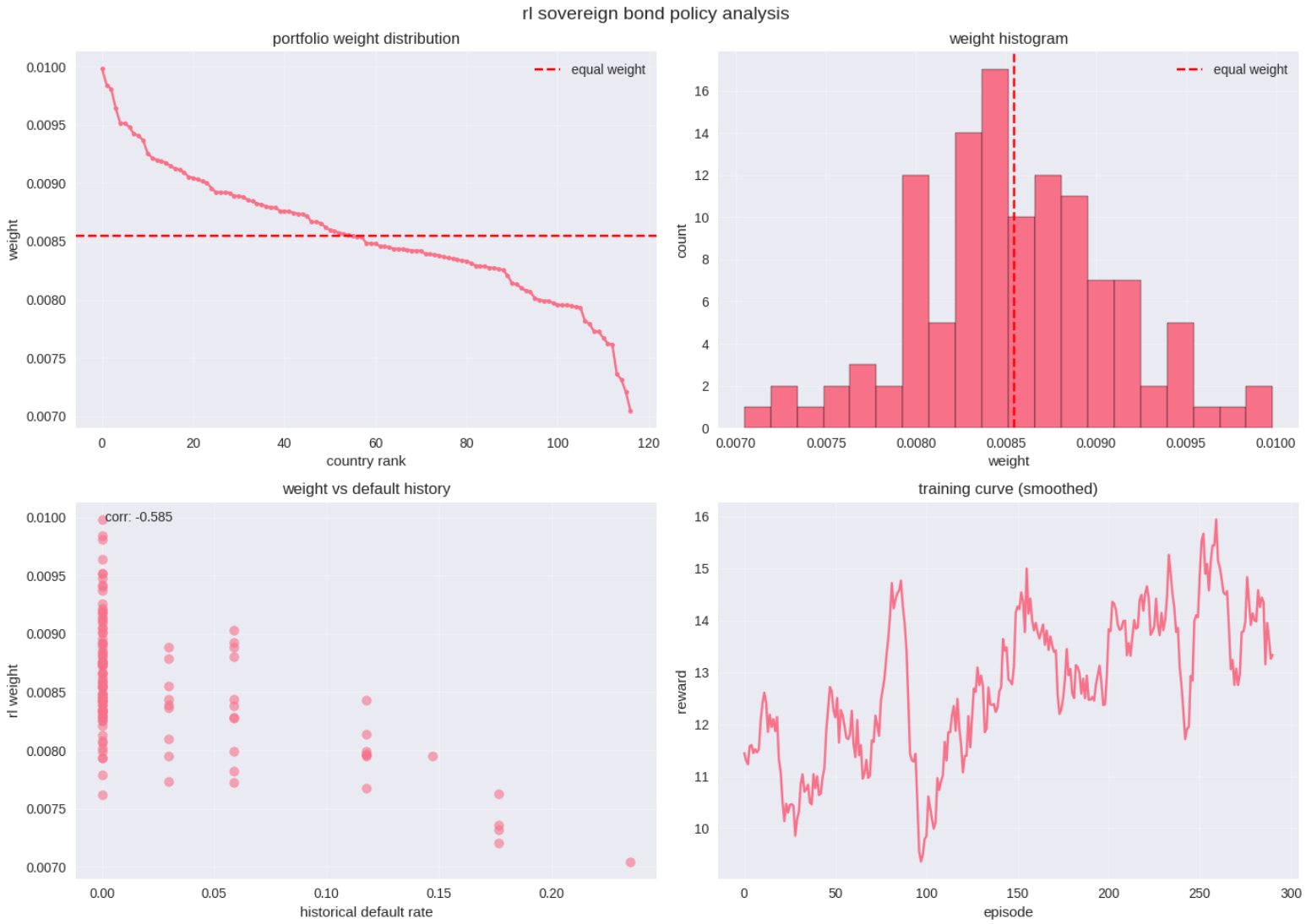
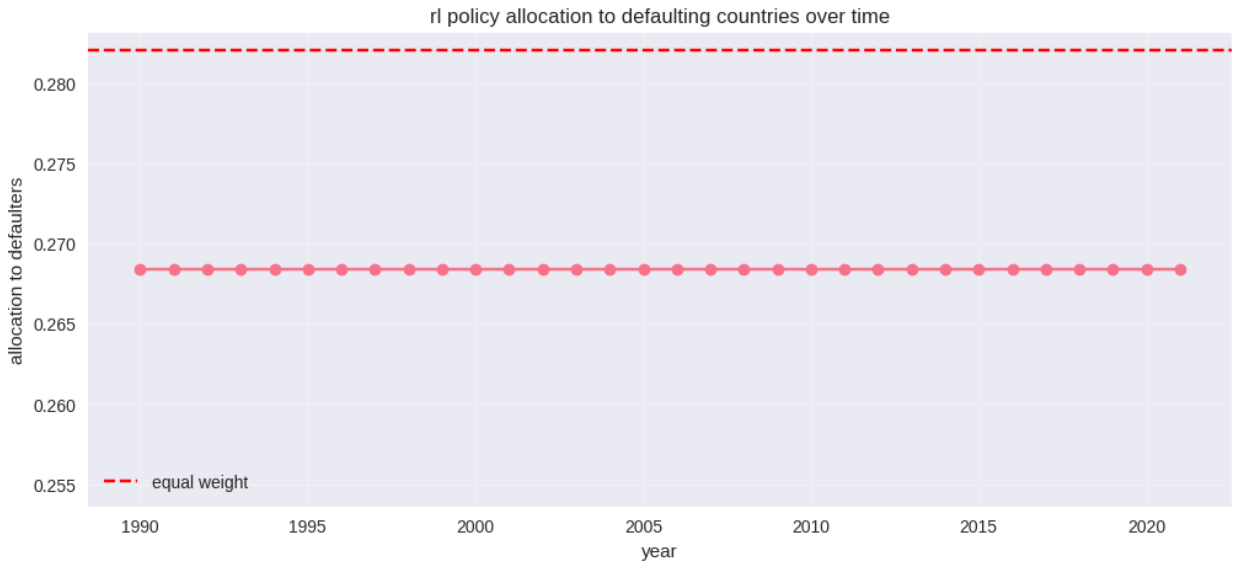
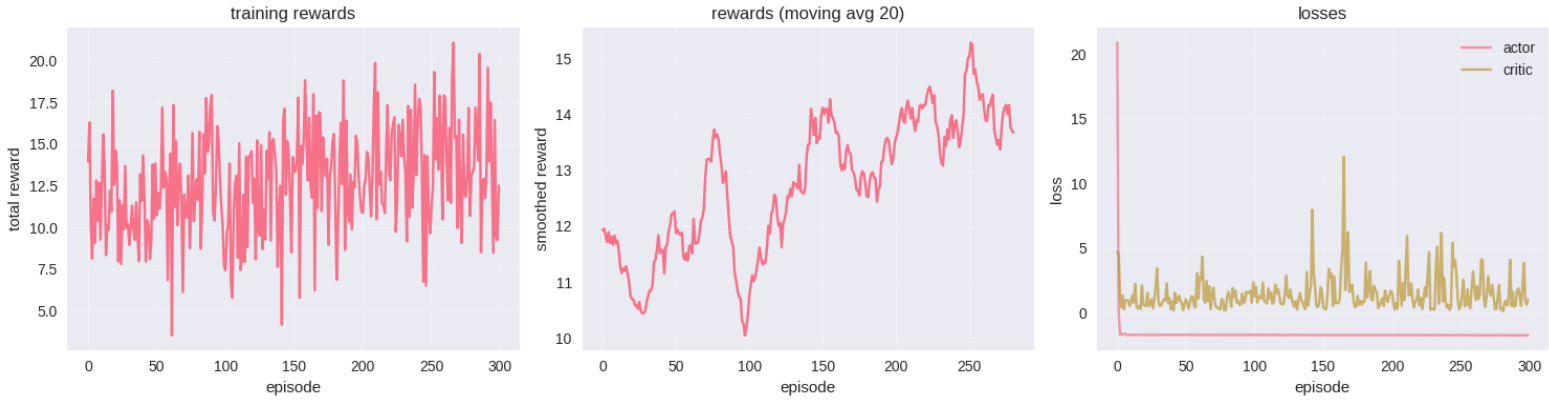


Figure 5: RL policy allocation to historically defaulting countries over time. The agent consistently underweights defaulters relative to equal-weight baseline (red dashed line), demonstrating learned risk aversion.



However, the weight spread is narrow (0.70%-1.00%), suggesting conservative risk-taking. The agent allocates 26.8% to countries that defaulted historically, only 1.4 percentage points below the 28.2% equal-weight allocation would imply. This limited differentiation raises questions about whether the policy captures sophisticated risk assessment or simply pattern-matches historical defaults.

Figure 6: PPO training dynamics. Left: Episode rewards showing learning progress. Center: Smoothed rewards demonstrating convergence. Right: Actor and critic loss trajectories.



Sensitivity Analysis

To probe whether the RL agent performs genuine economic reasoning versus pattern matching, we conduct two sensitivity experiments:

Feature Perturbation: We increase individual macroeconomic variables by 10% and measure portfolio weight changes. For GDP growth, inflation, external debt, government debt, and reserves, perturbations produce zero change in portfolio weights. This indicates the agent does not respond dynamically to marginal changes in fundamentals.

Random State: We replace the current state with a completely different randomly generated state vector. This produces total weight difference of 53.49 across all countries, confirming the agent is state-dependent at the macro level.

These findings reveal a critical limitation: the PPO agent learned to recognize overall state patterns and map them to a relatively fixed allocation but does not perform the economic reasoning we intended. It identified that certain countries historically default more often and learned to underweight them slightly but does not adjust weights in response to changes in specific economic indicators. The policy represents sophisticated pattern matching rather than dynamic risk assessment based on causal understanding of default drivers.

Discussion

Why Deep Learning Underperformed

The substantial performance gap between random forest (AUC 0.828) and the two-tower neural network (AUC 0.675) contradicts the narrative that deep learning universally outperforms traditional methods. Our results demonstrate that sample size fundamentally constrains deep learning effectiveness.

With approximately 3,000 training observations and severe class imbalance (78 positive examples), the optimization problem for learning 16-dimensional embeddings across two towers is severely underdetermined. The network has thousands of parameters to estimate from dozens of positive examples. In contrast, random forest with balanced class weights and regularization through tree depth limits naturally handles this regime.

The embedding collapse to one dimension provides direct evidence of insufficient training signal. The network converged to a minimal solution that reduces training loss without discovering the intended latent structure. This failure mode is characteristic of deep learning in small-sample regimes.

Our findings align with Makridakis et al. (2018), who demonstrate that statistical methods often outperform deep learning in time series forecasting with limited data. However, our study goes beyond prior work by providing direct diagnostic evidence for why deep learning fails in this regime. The embedding collapse revealed through PCA directly demonstrates that the optimization problem is underdetermined, the network cannot learn the hypothesized domestic-global interaction structure from available data. This diagnostic approach offers actionable guidance: practitioners can use embedding analysis to detect when deep learning architectures are inappropriate, rather than relying solely on validation performance. For sovereign defaults, the historical record simply does not provide enough examples for deep architectures to learn meaningful representations.

Interpretability and Feature Importance

Beyond predictive performance, the random forest model provides interpretable insights into default drivers through SHAP value analysis (Figure 1). External debt stocks, central

government debt, and reserve adequacy emerge as the primary predictors, consistent with theoretical frameworks emphasizing debt sustainability and liquidity buffers.

The model's feature importance rankings align with established empirical findings in the sovereign risk literature. External debt (% of GNI) shows the strongest positive association with default probability, reflecting vulnerability to external shocks and foreign currency obligations. Reserve coverage (months of imports) exhibits protective effects, consistent with its role as a liquidity buffer during stress periods. Central government debt captures fiscal sustainability concerns, while GDP growth and per capita income reflect debt servicing capacity.

Notably, global stress factors (VIX, credit spreads, yield curve) contribute meaningfully but secondarily to country-specific fundamentals. This suggests that while financial contagion matters during crisis periods, idiosyncratic fiscal and external vulnerabilities remain the primary default drivers. The random forest naturally captures these nonlinear interactions and threshold effects (for instance, high debt becomes particularly dangerous when combined with low reserves and negative growth).

This interpretability advantage represents a key strength of tree-based methods over neural networks in policy-relevant contexts. Policymakers and investors can understand which specific indicators drive risk assessments, facilitating targeted interventions and due diligence.

Reinforcement Learning Limitations

While the PPO agent achieves 12.6% improvement over baseline, our sensitivity analysis reveals it does not perform the dynamic economic reasoning we hypothesized. The agent learned historical risk patterns (serial defaulters receive lower weights) but does not adjust allocations in response to changes in specific macroeconomic indicators.

This limitation likely stems from the environment design. Our reward function evaluates terminal portfolio performance rather than incentivizing interpretable decision rules. The agent optimized what we measured (cumulative returns accounting for historical defaults) rather than what we intended (causal understanding of default drivers).

This highlights a broader challenge in financial reinforcement learning: agents may discover policies that perform well in-sample through pattern matching without learning economically meaningful strategies that generalize to new regimes. Our deterministic environment with fixed historical defaults is particularly susceptible to this issue, as the agent can simply memorize which countries defaulted.

Practical Implications

For practitioners and researchers in sovereign risk assessment, our results offer several actionable insights:

Traditional ML remains reliable: Random forests and gradient boosting provide robust predictive performance with limited data and severe class imbalance. These methods should remain the default choice for sovereign default prediction until substantially larger datasets become available.

Deep learning requires scale: Neural architectures require orders of magnitude more training examples than are available from historical sovereign defaults. While we cannot specify the exact threshold, our results suggest that for rare-event prediction with 2-3% base rates, deep learning likely requires at least 10,000+ observations with 200+ positive examples to avoid embedding collapse. This implies a fundamental mismatch between deep learning's data requirements and the inherent scarcity of sovereign default events. Claims of deep learning superiority should be scrutinized in rare-event prediction contexts.

Interpretability matters: Our sensitivity analysis proved essential for understanding that the RL agent learned pattern matching rather than economic reasoning. Without such diagnostic testing, we might have misinterpreted the performance improvements as evidence of genuine risk assessment capability.

Class imbalance demands attention: Standard machine learning practices (random train-test splits, accuracy metrics, unweighted loss functions) produce misleading results for sovereign defaults. Temporal splits, AUC/AP metrics, focal loss, and class weights are essential.

For multilateral institutions (IMF, World Bank): Our findings suggest that traditional econometric early warning systems augmented with ensemble machine learning methods offer the best balance of performance and interpretability for sovereign risk monitoring. The transparency of random forest feature importance enables clearer communication with member countries about vulnerability indicators.

For credit rating agencies: The substantial AUC advantage of random forests (0.828 vs. 0.636 for logistic regression) suggests potential for improving rating models through modern machine learning while maintaining interpretability. However, the failure of deep learning in this context cautions against uncritical adoption of complex architectures.

For researchers: Sample size requirements should guide methodological choices. When historical data is limited and rare events are the outcome of interest, devoting resources to careful feature engineering and domain-informed model selection yields better returns than experimenting with increasingly complex architectures.

Methodological Contributions

Beyond the specific findings on sovereign default prediction, this study contributes methodological insights for machine learning in finance:

Architecture design does not substitute for data: We designed the two-tower architecture based on economic intuition about vulnerability-stress interactions. This conceptual

motivation proved insufficient when training data was limited. Clever architectures cannot overcome fundamental sample size constraints.

Embedding analysis provides diagnostics: PCA on learned embeddings revealed the collapse that explained underperformance. This diagnostic approach should be standard practice when applying deep learning to small datasets.

Sensitivity analysis is essential for RL: Performance improvements alone do not validate that an RL agent learned the intended strategy. Our perturbation experiments revealed pattern matching that would have been missed by examining returns alone.

Limitations

Several limitations qualify our findings and suggest directions for future work.

Data Limitations

Sample size: With 88 default events across 34 years, the historical record provides limited training data. This is a fundamental constraint that cannot be easily addressed, as sovereign defaults are genuinely rare events.

Missing data: Central government debt (70.7% missing) and domestic credit (76.9% missing) forced us to rely on median imputation. This introduces measurement error, particularly for countries with poor data reporting that may also have higher default risk.

Temporal clustering: Many defaults in 1990 represent carryover from the 1980s debt crisis rather than events our models could have predicted from 1990 data. This conflates crisis resolution with crisis prediction.

Survivorship bias: Our sample includes only countries that existed throughout 1990-2023. Countries that ceased to exist (Soviet Union, Yugoslavia, East Germany) had different risk profiles not captured in our data.

Methodological Limitations

No yield curve data: Real sovereign bond analysis requires modeling the full-term structure. Our simplified yield function based on debt ratios and reserves, while reasonable, does not capture the rich information in actual bond prices.

No political risk: Defaults often stem from political decisions (Argentina 2001, Greece 2015) not fully captured by macroeconomic fundamentals. Incorporating political risk indicators or text data from news and IMF reports could improve predictions.

Static features: We use levels of macroeconomic variables without lag features or momentum indicators. Models with temporal dynamics (LSTM, temporal convolutional networks) might capture deterioration patterns.

Simplified RL environment: Our environment lacks many real-world complexities: liquidity constraints, transaction price impact, short-selling restrictions, and portfolio size limits. The deterministic variant with perfect foresight of defaults is particularly unrealistic.

Scope Limitations

This study focuses on prediction and portfolio allocation. We do not address causal questions about what drives sovereign defaults or evaluate policy interventions. Our models identify correlations between macroeconomic conditions and defaults but cannot distinguish causation from confounding.

We also do not explore certain advanced techniques that might improve deep learning performance: transfer learning from corporate defaults, pre-training on auxiliary tasks, or few-shot learning methods. These directions merit investigation but were beyond our scope.

Conclusion

This study provides a systematic comparison of traditional machine learning against deep learning for sovereign default prediction, with novel reinforcement learning for portfolio allocation. Our findings demonstrate that with limited training data, simpler methods substantially outperform complex architectures.

Random forest achieves AUC 0.828, significantly exceeding the two-tower neural network's AUC 0.675. Embedding analysis reveals catastrophic collapse to one dimension, indicating the network failed to learn meaningful representations from approximately 3,000 training observations and 78 positive examples. This provides direct evidence that deep learning sample size requirements exceed what historical sovereign default data can provide.

The reinforcement learning agent achieves 12.6% improvement over equal-weight baseline, but sensitivity analysis reveals pattern matching rather than dynamic economic reasoning. The agent learned historical risk patterns without responding to marginal changes in macroeconomic fundamentals.

These results have important practical implications. For sovereign default prediction, traditional machine learning methods remain more reliable than deep learning until substantially larger datasets become available. Practitioners should prioritize interpretable models, conduct comprehensive sensitivity analysis, and employ appropriate evaluation methods for rare-event prediction.

More broadly, our study contributes to understanding when deep learning offers advantages over traditional approaches. Architectural sophistication does not substitute for adequate training data. In macroeconomic contexts with limited historical observations and severe class imbalance, simpler methods that incorporate domain knowledge through careful feature engineering and regularization outperform flexible but data-hungry deep learning approaches.

Future work should explore transfer learning from related domains (corporate defaults, credit spreads), incorporation of text data from IMF reports and news articles, and causal methods for understanding default drivers rather than pure prediction. The integration of economic theory with machine learning through structural models or physics-informed neural networks may provide paths toward more data-efficient learning in macroeconomic applications.

Future Research Directions

Several promising avenues warrant investigation. First, transfer learning approaches that pre-train on corporate default data before fine-tuning on sovereign defaults may help overcome sample size limitations. Second, incorporating unstructured data from IMF Article IV reports, credit rating agency statements, and financial news through natural language processing could capture political and institutional factors missed by macroeconomic fundamentals alone. Third, causal inference methods such as synthetic control or instrumental variables could move beyond prediction to understand the causal mechanisms driving defaults. Finally, hybrid approaches that combine economic theory with machine learning through structural models or physics-informed neural networks may provide more interpretable and data-efficient learning. These directions could help bridge the gap between flexible machine learning methods and the fundamental data scarcity inherent to sovereign default prediction.

Reproducibility Statement

All code and data for this study are available in a public repository at https://github.com/thylinao1/Sovereign-Risk-and-Protfolio-Allocation.

The repository includes:

- Data collection scripts for World Bank and FRED APIs
- Sources for sovereign default databases
- Jupyter notebooks for all experiments
- Trained model weights for all architectures
- Scripts to reproduce all tables and figures
- Requirements file specifying exact package versions

For any questions regarding reproducibility, please contact the author at maksim.silchenko@bayes.city.ac.uk.

Acknowledgments

This research was conducted independently as part of the author's academic portfolio development. The views expressed are those of the author and do not represent the official position of Bayes Business School or City, University of London.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

The author declares no conflicts of interest.

Ethical Approval

This research uses publicly available aggregate macroeconomic data and does not involve human subjects. No ethical approval was required.

References

Aguiar, M., and Gopinath, G. (2006). Defaultable debt, interest rates and the current account. *Journal of International Economics*, 69(1):64–83.

Afonso, A., Gomes, P., and Rother, P. (2011). Short and long-run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, 16(1):1–15.

Arellano, C. (2008). Default risk and income fluctuations in emerging economies. American Economic Review, 98(3):690–712.

Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.

Arnott, R., Harvey, C. R., and Markowitz, H. (2019). A backtesting protocol in the era of machine learning. *The Journal of Financial Data Science*, 1(1):64–74.

Barboza, F., Kimura, H., and Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83:405–417.

Beers, D. T., and Nadeau, J.-S. (2017). Database of sovereign defaults, 2017. Bank of Canada Technical Report 101.

Belkin, M., Hsu, D., Ma, S., and Mandal, S. (2019). Reconciling modern machine learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854.

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120:70–83.

Bianchi, D., and Büchner, M. (2020). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Broner, F., Lorenzoni, G., and Schmukler, S. L. (2013). Why do emerging economies borrow short term? *Journal of the European Economic Association*, 11(s1):67–100.

Cantor, R., and Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Economic Policy Review*, 2(2):37–53.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794.

Chen, Z., Li, K., Li, W., and Shi, X. (2019). Deep learning for credit card fraud detection: A survey. *IEEE Access*, 7:109386–109398.

Ciarlone, A., and Trebeschi, G. (2007). Assessing sovereign credit risk in emerging markets. Bank of Italy Working Paper 624.

Cruces, J. J., and Trebesch, C. (2013). Sovereign defaults: The price of haircuts. *American Economic Journal: Macroeconomics*, 5(3):85–117.

Davis, J., and Goadrich, M. (2006). The relationship between precision-recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 233–240.

Edwards, S. (1986). The pricing of bonds and bank loans in international markets: An empirical analysis of developing countries' foreign borrowing. *European Economic Review*, 30(3):565–589.

Efron, B., and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Feder, G., and Just, R. E. (1985). A study of debt servicing capacity applying logit analysis. *Journal of Development Economics*, 14(1-2):25–38.

Fioramanti, M. (2008). Predicting sovereign debt crises using artificial neural networks: A comparative approach. *Journal of Financial Stability*, 4(2):149–164.

Forbes, K. J. (2002). Are trade linkages important determinants of country vulnerability to crises? In *Preventing Currency Crises in Emerging Markets*, pages 77–124. University of Chicago Press.

Frank, C. R., and Cline, W. R. (1971). Measurement of debt servicing capacity: An application of discriminant analysis. *Journal of International Economics*, 1(3):327–344.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232.

Gaillard, N. (2012). *A Century of Sovereign Ratings*. Springer.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2017). Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12.

Jiang, Z., Xu, D., and Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. arXiv preprint arXiv:1706.10059.

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3):e0194889.

Manasse, P., and Roubini, N. (2009). Rules of thumb for sovereign debt crises. *Journal of International Economics*, 78(2):192–205.

Moody, J., and Saffell, M. (1998). Reinforcement learning for trading systems and portfolios. In *Proceedings of the IEEE/IAFE Conference on Computational Intelligence for Financial Engineering*, pages 279–287.

Reinhart, C. M., and Rogoff, K. S. (2009). *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3):e0118432.

Savona, R., and Vezzoli, M. (2015). Fitting and forecasting sovereign defaults using multiple risk signals. *Oxford Bulletin of Economics and Statistics*, 77(1):66–92.

Schulman, J., Moritz, P., Levine, S., Jordan, M., and Abbeel, P. (2015). High-dimensional continuous control using generalized advantage estimation. arXiv preprint arXiv:1506.02438.

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

Sirignano, J., and Cont, R. (2016). Universal features of price formation in financial markets: Perspectives from deep learning. arXiv preprint arXiv:1803.06917.

Tanaka, K., Kinkyo, T., and Hamori, S. (2018). Random forests-based early warning system for bank failures. *Economics Letters*, 148:118–121.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*, 16(4):437–450.

Yi, X., Yang, J., Hong, L., Cheng, D. Z., Heldt, L., Kumthekar, A., Zhao, Z., Wei, L., and Chi, E. (2019). Sampling-bias-corrected neural modeling for large corpus item recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 269–277.

Zhang, Z., Zohren, S., and Roberts, S. (2020). Deep reinforcement learning for trading. *The Journal of Financial Data Science*, 2(2):25–40.