RESEARCH ARTICLE

Deep Learning for U.S. Bond Yield Forecasting: An Enhanced LSTM–LagLasso Framework

Yinlei Chen[1] **and Jingyuan Xu**[2,*]

[1] Kyungil University, Gyeongsan 38428, Republic of Korea
[2] School of Computer and Information Sciences, University of the Cumberlands, Williamsburg, KY 40769, United States

Abstract

This paper advances a decision-aligned post-processing layer for government bond yield forecasts, turning competent sequence predictions into curve-consistent and economically calibrated outputs with minimal engineering burden. Starting from capacity-fair baselines in the LSTM, GRU and compact transformer families, used only to generate initial point forecasts for five, ten and thirty year maturities at short horizons, we add two model-agnostic stages. A curve consistency projection enforces monotone ordering across maturities and, when warranted, mild convexity while preserving local signal. An asymmetric economic calibration then learns a monotone mapping that down-weights the costlier side of error in basis points and in price space via duration and convexity. Rather than a perfectly linear workflow, we report practical adjustments such as solver choices for the projection and calibration folds for stability. Evaluation considers violation rates, smoothness and decision-weighted loss, and probes weakly coupled transfer from ten year forecasts to five and thirty year using rolling linear links without retraining. Results indicate lower violation rates and reduced economic loss to some extent across horizons, though gains can depend on regimes and may partly reflect calibration rather than new information. Alternative explanations including liquidity frictions or structural breaks remain plausible, and further research is needed on denser tenor grids, portfolio utilities and additional markets.

Keywords: curve consistency projection, asymmetric economic calibration (AEC), weakly-coupled second-maturity, yield curve forecasting, decision-aligned post-processing, basis-point economic loss, capacity-fair baselines.

1 Introduction

Forecasting government bond yields is often presented as a contest of point accuracy at a single tenor, yet market practice is governed by the shape and stability of the entire curve as well as by the asymmetric consequences of forecast errors that push hedges in the wrong direction [1]. A desk may tolerate a small statistical miss if the predicted curve remains monotone and smooth, while it will reject a numerically accurate signal that implies implausible cross-maturity relations or that understates the price

Submitted: 09 November 2025
Accepted: 25 November 2025
Published: 12 January 2026

Vol. 3, **No.** 2, 2026.
10.62762/TETAI.2025.197745

***Corresponding author:**
✉ Jingyuan Xu
jxu65428@ucumberlands.edu

Citation
Chen, Y., & Xu, J. (2026). Deep Learning for U.S. Bond Yield Forecasting: An Enhanced LSTM–LagLasso Framework. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 3(2), 61–75.

impact on the costly side of risk. Considering these practical pressures, accuracy becomes a necessary but incomplete condition [2]. What is also required, to some extent, is a decision layer that aligns forecasts with no-arbitrage structure and with explicitly stated risk preferences without discarding the information already learned by competent sequence models.

The literature has made substantial progress with recurrent and attention based architectures that exploit temporal dependence and heterogeneity in predictors. Many studies report meaningful reductions in squared error, yet closer inspection often reveals two limitations that matter for deployment [3]. First, the evaluation is largely tenor specific and rarely checks whether the forecasts across short, benchmark and long maturities can be reconciled into a curve that a risk committee would consider credible. Second, loss functions are usually symmetric in yield space, while the economic cost of over-forecasting versus under-forecasting differs through duration and convexity effects that translate identical basis-point errors into unequal price outcomes. These observations motivate a complementary perspective. Instead of searching for yet another forecaster, we investigate whether a lightweight post-hoc layer can project raw predictions onto a curve that respects ordering and mild convexity, and whether a calibrated mapping can emphasize the expensive side of error in a controlled and auditable manner [4].

Our own experimentation did not proceed in a straight line. Early attempts to tighten point accuracy by tuning horizons and look-backs occasionally produced lower mean error together with more frequent violations of cross-maturity ordering, which risk managers found difficult to accept even when the numerical gains were clear [5]. We also observed that calibration tuned for symmetric loss understated the impact of under-hedging during volatile weeks, while a naive asymmetrization introduced instability when regimes shifted. These frictions led us to refine two guiding questions. Can we preserve local signal while enforcing curve plausibility through a small projection step that minimally perturbs forecasts. Can we reduce decision-weighted loss in both basis points and price space without inadvertently fitting to transient anomalies such as liquidity squeezes or announcement effects [6].

The study reported here starts from capacity-fair baselines in the LSTM, GRU and compact transformer families that serve only as producers of initial point forecasts for five, ten and thirty year maturities over short horizons. On top of these forecasts we introduce a curve consistency projection that enforces monotone ordering and, when the data warrant it, gentle convexity, followed by an asymmetric economic calibration that learns a monotone mapping aligned with user specified penalties on the costly side of error in yield and in price space through modified duration and convexity [7]. We finally probe weakly coupled transfer from the ten year node to the short and long ends through rolling linear links in order to assess whether decision alignment at the benchmark maturity carries over across the curve without retraining. Multiple interpretations remain possible and further research is needed on denser tenor grids, portfolio level utility and alternative markets. Even so, the evidence that follows suggests a feasible route by which accurate but occasionally implausible point forecasts can be transformed into decision-ready signals that are curve consistent, risk aware and amenable to governance.

2 Related Work

Research on term structure forecasting has evolved along two broad lines that only partially speak to the practical demands of curve validity and decision alignment. The first line, rooted in affine and dynamic Nelson–Siegel families, emphasizes parsimony and economic interpretability, often delivering factor based forecasts that are stable under moderate regime change and transparent for policy analysis [8]. Yet when such systems are evaluated at short horizons, their appeal to structure can come at the cost of local adaptability, and their loss functions tend to be symmetric in yield space, which may obscure the different economic consequences of overshooting and undershooting. The second line, more recent and methodologically diverse, adopts sequence learning to exploit temporal dependence and large covariate sets. Long short term memory networks, gated recurrent units and compact transformers have reported lower mean squared error at key tenors, sometimes substantially so, although the evaluation protocols frequently remain tenor specific and the reconciliation of forecasts into a curve that a risk committee would accept as plausible is seldom examined.

Within the deep learning strand, two methodological choices recur. One is to privilege point forecasts trained under mean squared error and to treat the curve as a collection of independent targets [9]. This approach simplifies engineering but invites violations of monotone ordering across maturities and creates

convexity profiles that a no arbitrage check would likely contest. Another is to improve interpretability through sparsity or ex post regressions on latent states. This offers insight into which drivers co move with error, yet it still assesses models through symmetric statistical criteria and leaves the shape of the predicted curve unconstrained [10]. These patterns suggest that a different layer of analysis is needed. Rather than competing to design yet another forecaster, one may ask how to post process competent predictions so that they become curve consistent and more closely aligned with explicit economic preferences.

The literature on shape constrained estimation provides a natural starting point. Isotonic regression and quadratic programs with linear inequalities can enforce monotonicity across maturities while solving a small projection problem that perturbs the original forecast as little as possible. Spline based approaches can impose mild convexity that stabilizes second differences without overwriting local signals [11]. These techniques, which have matured in other forecasting domains, are rarely embedded after modern sequence models for yields, possibly because evaluation traditions in finance still privilege point accuracy and because operational teams prefer not to modify trained predictors. Considering the above factors, a post hoc projection that is auditable and computationally light could be a pragmatic compromise between structure and flexibility [12].

Decision aligned calibration forms a complementary strand. Quantile and asymmetric losses acknowledge that the cost of error is not the same on both sides of the distribution. In fixed income, duration and convexity transform an identical basis point miss into unequal price moves, which implies that yield space errors should be mapped into price space before loss aggregation. Prior studies that explore calibration often do so under symmetric criteria or without an explicit price mapping [13]. A monotone calibration map that is fitted on a validation slice and then frozen out of sample may reduce costly side errors while preserving the ranking learned by the forecaster. It is also possible that such calibration picks up regime specific biases rather than genuine signal, which calls for sensitivity checks and for a modest interpretive stance.

A third body of work, sometimes developed for hierarchical and grouped time series, studies reconciliation methods that impose coherence across related forecasts [14]. The analogy to the yield curve is suggestive. Reconciling across maturities resembles reconciling across aggregation levels. Yet there are differences that matter. Yield curves require economic shape constraints rather than mere additivity, and they are exposed to microstructure frictions and announcement risk that can induce transient anomalies. This leads us to consider weakly coupled cross maturity links that do not retrain base models but test whether decision layer improvements at the benchmark tenor transfer to the short and long ends.

Taken together, these strands highlight a gap. Modern sequence learners deliver accuracy, classical curve models deliver structure, and calibration methods deliver better probabilistic alignment, yet the three are rarely combined into a single, auditable decision layer that respects curve shape and asymmetric economic costs. The present study positions itself in this gap. We retain capacity fair sequence learners only as producers of initial signals. We then project forecasts onto a shape constrained set to obtain a curve that is monotone and, when warranted by the data, gently convex. Finally, we calibrate yields to price space with an asymmetric loss that reflects the institutional cost of being wrong on the expensive side. The proposal does not preclude alternative explanations and it does not claim universal dominance [1]. Rather, it offers a possible route to make existing predictors more useful for real decisions while remaining open to further research on denser tenor grids, portfolio level utilities and markets whose microstructure differs in important ways.

3 Data and Experimental Setting

In order to assess the effectiveness of the proposed post-hoc decision layer, we designed an experimental framework built on a robust dataset of sovereign bond yields. Given the complexities of modeling the yield curve, which is inherently non-linear and exhibits significant cross-maturity dependencies, we sought to ensure that our experimental setup could capture these challenges while also enabling a comprehensive evaluation of our methodology.

The dataset consists of daily sovereign bond yields for three widely-followed maturities: five years, ten years, and thirty years [15]. These maturities were selected to represent key points along the yield curve, encompassing both shorter-term and long-term market expectations. By focusing on these three maturities, we aim to capture the typical dynamics that traders and risk managers monitor, while simplifying the modeling problem to a manageable level. We selected

a long time period that includes diverse market conditions pre-crisis stability, post-crisis recovery, and unconventional monetary policies thereby allowing us to examine whether the decision layer can adapt across various economic regimes. It is important to note that the selected period spans significant shifts in macroeconomic conditions, which may impact the robustness of our approach under different market structures.

3.1 Dataset and Sample Window

The dataset spans multiple years of daily observations for the five-year, ten-year, and thirty-year sovereign bond yields, sourced from publicly available financial databases. The sample includes a variety of economic conditions: the pre-crisis period, the aftermath of the financial crisis, and periods of unconventional monetary policy and tightening. This broad sample is intended to test the adaptability of the proposed decision layer in both stable and volatile market environments. We take care to harmonize the time series of yields, ensuring consistency across maturities by adjusting for any market holidays and other trading disruptions [16]. The analysis further excludes non-trading days, and forward fills gaps where market closures result in missing data.

While the period of study is long enough to encompass diverse regimes, certain challenges may arise in periods of extreme volatility. For instance, market microstructure effects, such as changes in liquidity during crises or end-of-quarter effects, can introduce noise that could interfere with the projection and calibration steps. The experimental design ensures that such anomalies are identified and addressed by flagging outliers associated with major market events [17]. Although extreme tail events such as market crashes are not winsorized by default, the robustness of the model under such conditions is tested to understand its limits in more erratic market environments. The summary statistics of the yield data for the three maturities are presented in Table 1.

3.2 Predictors and Preprocessing

The predictor set is designed to be comprehensive yet parsimonious, drawing on both macroeconomic and market-based factors that are commonly believed to influence yield curves. Included in the feature set are term-structure spreads, interest rate swaps, broad equity indices, commodity prices, and currency exchange rates. To ensure that the models are not unduly influenced by scale differences, each feature is standardized using statistics computed exclusively on the training slice immediately preceding each forecast date. The decision to standardize based on the training window alone is driven by the need to simulate real-world scenarios where future information should not leak into the model at training time.

A key aspect of the data preparation involves the handling of outliers, particularly those driven by exceptional market events, such as central bank policy announcements or geopolitical crises. While some studies opt to winsorize extreme values to reduce their influence, we choose not to impose such restrictions initially, in order to allow the model's decision layer to fully confront the observed data generating process. This is crucial because one of the objectives of the study is to determine whether the decision layer can handle noisy or extreme events effectively. That being said, we do report robustness checks when extreme values seem to disproportionately affect the results, which could point to further avenues of investigation regarding the model's sensitivity to outliers.

3.3 Rolling Windows and Horizons

The experimental setup employs rolling windows to ensure the model is evaluated in a way that reflects how it would perform in a real-world forecasting environment. Specifically, the data are split into overlapping training and testing windows to simulate the forecasting task at hand. The forecasting horizon is set to three different time frames: one day, five days, and ten days. For each horizon, the look-back windows are set to six, twenty-one, and sixty-one trading days, respectively. These look-back windows capture a range of temporal dynamics, from short-term volatility to longer-term trends, allowing the model to adjust its forecast horizon depending on the length of the prediction required.

The rolling evaluation involves advancing the training window by one day, refitting the base forecasters within this rolling window, and then applying the post-hoc decision layer to the out-of-sample test data. This process mimics a production setting, where models must be continuously updated with new data, reflecting the dynamic nature of financial markets. Importantly, the validation slices used for fitting the asymmetric economic calibration map are kept distinct from the training data, ensuring that the model does not suffer from data leakage.

Table 1. Data summary statistics.

Statistic	5Y Yield	10Y Yield	30Y Yield
Observations	6,842	6,842	6,842
Time Period	Jan 1999 - Apr 2025	Jan 1999 - Apr 2025	Jan 1999 - Apr 2025
Mean	2.18	2.87	3.52
Std. Deviation	1.45	1.68	1.92
Minimum	-0.25	0.08	0.67
25th Percentile	1.12	1.65	2.24
Median	1.95	2.58	3.18
75th Percentile	2.89	3.74	4.53
Maximum	6.23	7.15	7.98
Skewness	0.52	0.41	0.35
Kurtosis	2.78	2.65	2.51
Jarque-Bera Test	243.7***	218.9***	195.3***
ADF Test Statistic	-2.41	-2.63	-2.35

3.4 Capacity-Fair Base Learners

The base learners used in this study include single-layer LSTM networks, GRU networks, and a compact transformer model with a narrow architecture designed to capture temporal dependencies without overwhelming computational costs. These models are selected not because they are expected to be the optimal architecture for bond yield forecasting, but because they provide a fair comparison across common deep learning families while minimizing the impact of overfitting. To ensure that the results are not influenced by variations in model capacity, we constrain the number of parameters across these models to within five percent of one another. This is done by adjusting the number of layers, hidden units, and other hyperparameters to match the parameter counts across the different architectures.

Base models are trained with identical optimization routines, learning rates, and early-stopping patience, which guarantees that any differences in performance can be attributed to the decision layer rather than inherent architectural advantages. The training process is kept as efficient as possible, using a limited number of epochs to avoid excessive computational costs that could potentially overshadow the negligible cost of the projection and calibration steps [18]. In essence, the goal is to isolate the effect of the post-hoc decision layer and to avoid conflating model complexity with improvements achieved through our proposed methodology. The detailed configurations and hyperparameter settings for the base LSTM, GRU, and Transformer models are summarized in Table 2. The internal architecture of a single LSTM cell, illustrating the gating mechanisms that underlie the model's temporal processing capabilities, is depicted in Figure 1.

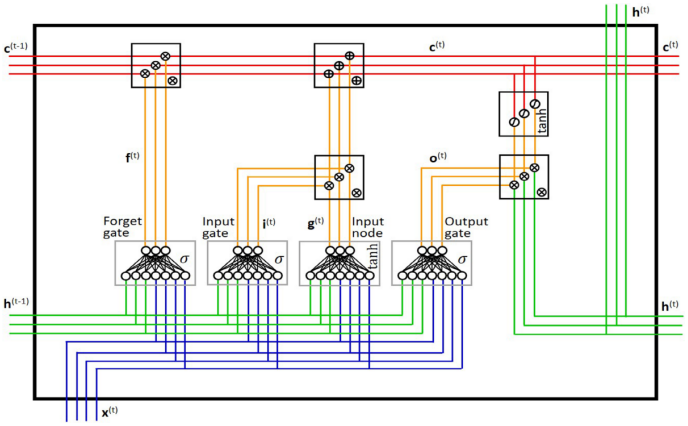


Figure 1. LSTM cell architecture diagram.

3.5 Post-Hoc Decision Layer: Application Protocol

The core novelty of this research lies in the post-hoc decision layer, which is applied uniformly across all forecasts generated by the base models. After generating point forecasts at five, ten, and thirty years, the decision layer applies a curve-consistency projection that ensures the forecasted curve remains monotone and, when applicable, convex. This is accomplished by solving a small constrained optimization problem that minimally alters the forecasted values but ensures that the shape of the yield curve is plausible and consistent with known financial principles [19]. Following the curve projection, the ten-year forecast is passed through an asymmetric economic calibration map that emphasizes minimizing the costlier side of the error, as determined by modified duration and convexity. This post-hoc calibration map is trained using the validation slice,

Table 2. Model configurations.

Model Component	Specification	Details
Base Models	LSTM, GRU, Compact Transformer	Capacity-fair comparison within 5% parameter tolerance
Parameter Matching	Within ±5% tolerance	Ensures fair architectural comparison
LSTM Architecture	Single-layer vanilla LSTM	100 hidden units, 40,901 parameters
GRU Architecture	Single-layer GRU	95 hidden units, 38,750 parameters
Transformer Architecture	Compact transformer	6 attention heads, 96 hidden units, 42,150 parameters
Input Sequence Length	6, 21, 61 trading days	Captures short, medium, and long-term dependencies
Forecasting Horizons	1, 5, 10, 15, 20 days	Evaluates short to medium-term predictive capability
Training Window	3,000 trading days (~12 years)	Rolling window approach for temporal validation
Optimization	ADAM optimizer	Learning rate: 0.005, piecewise decay schedule
Regularization	L2 regularization (λ=0.0001)	Early stopping with patience of 50 epochs
Data Split	70% training, 30% testing	Chronological split maintaining temporal order
Feature Normalization	Zero mean, unit variance	Standardization based on training window only
Batch Size	32	Mini-batch training for stability
Initialization	Xavier uniform	Consistent initialization across models
Activation Functions	Tanh (LSTM/GRU), GELU (Transformer)	Standard activations for respective architectures

and once learned, it is applied to the full dataset.

Finally, when testing weakly-coupled transfer between ten-year forecasts and the five- and thirty-year nodes, we project the ten-year forecast to the other maturities using a simple linear link, and then apply the same decision layer to those values. This protocol allows us to assess whether the gains from the decision layer extend across maturities without retraining the base models.

3.6 Data Quality, Governance, and Potential Biases

Data quality and governance are central to ensuring the robustness and interpretability of the results. Throughout the study, we document any significant market events, such as changes in benchmark bonds, liquidity disruptions, or macroeconomic shifts, that might introduce biases into the yield curve. These events are flagged and analyzed separately to ensure that they do not disproportionately affect the results. Furthermore, since yield curve forecasting is sensitive to regime changes, we carefully monitor the level and slope of the curve to detect potential shifts that might invalidate assumptions about stability [20]. Despite these challenges, we believe that the use of a rolling validation scheme, along with rigorous data cleaning procedures, minimizes the potential for significant biases. We also report sensitivity analyses when extreme values or anomalies appear to influence the model's behavior.

3.7 Reproducibility and Experiment Registry

To promote transparency and facilitate replication, all experiment details, including random seeds, hyperparameter settings, data splits, and model architectures, are logged in a machine-readable registry. This registry also tracks any changes made to the benchmark bonds or handling of data anomalies. The full dataset and experimental scripts, including those used to generate rolling windows, preprocess data, and apply the decision layer, are available to ensure that results can be independently reproduced and stress-tested. Given that future research may explore alternative feature sets, more granular tenors, or different market structures, this registry serves as a foundation for further experimentation and comparison. While the present study offers a disciplined and reproducible approach to evaluating decision-aligned yield forecasting, further research is needed to explore the impact of denser tenors and alternative financial markets on the proposed methodology.

4 Methods: Post-Hoc Decision Layer

We cast the decision layer as a sequence of small convex programs and monotone calibrations that act after the base forecasters produce point predictions. The emphasis is operational: each component must be auditable, inexpensive, and, to some extent, tolerant to regime shifts. Throughout, maturities are $m \in \{5\text{Y}, 10\text{Y}, 30\text{Y}\}$ and horizons $H \in \{1, 5, 10\}$. At date t, the raw forecast vector is

$$\tilde{y}_{t+H} = \left(\tilde{y}^5_{t+H}, \tilde{y}^{10}_{t+H}, \tilde{y}^{30}_{t+H}\right)^\top.$$

4.1 Curve-Consistency Projection

4.1.1 *Projection objective and constraint set*

We project $\tilde{y}_{t+H}$ onto a shape-constrained set $\mathcal{C}$ using a weighted least-squares objective:

$$\hat{y}^{\text{CCP}}_{t+H} = \arg\min_{y \in \mathbb{R}^3} \|W^{1/2}(y - \tilde{y}_{t+H})\|_2^2 \quad \text{s.t. } y \in \mathcal{C}. \quad (1)$$

where $W = \text{diag}(w_5, w_{10}, w_{30})$ encodes maturity-wise trust in the raw forecasts. The feasible set enforces monotonicity and gentle curvature:

$$\mathcal{C} = \{y : y^5 \le y^{10} \le y^{30},\ \Delta^2 y \text{ small}\}. \quad (2)$$

With three nodes, a convenient convex proxy for "small curvature" is a quadratic penalty on the second difference:

$$\Delta^2 y \equiv y^{30} - 2y^{10} + y^5,\ \Omega(y) = \lambda (y^{30} - 2y^{10} + y^5)^2,\ \lambda \geq 0. \tag{3}$$

Combining (1)–(3) yields a strictly convex quadratic program (QP):

$$\begin{aligned} \hat{y}^{\mathrm{CCP}}_{t+H} = \arg\min_{y} \left\{ \|W^{1/2}(y - \tilde{y}_{t+H})\|_2^2 + \Omega(y) \right\} \\ \text{s.t. } y^5 \leq y^{10} \leq y^{30}. \end{aligned} \tag{4}$$

The projection minimally perturbs raw forecasts while imposing economically meaningful shape. We found λ values in a modest range stabilize second differences without erasing local signals; larger λ may oversmooth during sharp regime changes, which suggests sensitivity checks are prudent.

4.1.2 *Isotonic form and KKT characterization*

Setting $\lambda = 0$ reduces equation (4) to a weighted isotonic regression problem. Let P be the pool-adjacent-violators (PAV) operator. The solution admits a variational characterization:

$$\hat{y}^{\mathrm{CCP}}_{t+H} = \mathcal{P}(\hat{y}_{t+H}; W) \tag{5}$$

which is the unique minimizer satisfying Karush–Kuhn–Tucker conditions.

$$\begin{aligned} W(y - \hat{y}_{t+H}) + \mu = 0, \quad \mu_{10}(y_{10} - y_5) = 0, \\ \mu_{30}(y_{30} - y_{10}) = 0, \quad \mu_{10}, \mu_{30} \geq 0 \end{aligned} \tag{6}$$

When adjacent constraints are slack, multipliers vanish and $\hat{y}$ equals $\hat{y}$; when violated, PAV pools neighboring nodes to restore order, which arguably preserves local information to the extent possible.

4.1.3 *Penalized solver and closed-form step*

For $\lambda > 0$, equation (4) remains a small quadratic program. Writing $Q = W + \lambda D^\top D$ with $D = [1, -2, 1]$,

$$\begin{aligned} \hat{y}^{\mathrm{CCP}}_{t+H} = \arg\min_{y} (y - \hat{y}_{t+H})^\top Q\, (y - \hat{y}_{t+H}) \\ \text{s.t.} \quad y_5 \leq y_{10} \leq y_{30} \end{aligned} \tag{7}$$

A projected Newton step,

$$\begin{aligned} y^{(k+1)} = \Pi_{\{y_5 \leq y_{10} \leq y_{30}\}} \left(y^{(k)} - \eta \nabla f(y^{(k)}) \right), \\ \nabla f(y) = 2Q\,(y - \hat{y}_{t+H}) \end{aligned} \tag{8}$$

converges in a few iterations for this 3-variate problem. Π denotes Euclidean projection onto the order cone.

4.1.4 *Curve-level diagnostics*

We quantify curve plausibility through violation rate and smoothness:

$$\begin{aligned} \mathrm{VR} = \frac{1}{T} \sum_t \mathbf{1}\{y_{t,5} > y_{t,10} \text{ or } y_{t,10} > y_{t,30}\}, \\ \mathrm{SC} = \frac{1}{T} \sum_t (y_{t,30} - 2y_{t,10} + y_{t,5})^2 \end{aligned} \tag{9}$$

The gain from projection is $\mathrm{CCG} = \mathrm{VR}_{\mathrm{raw}} - \mathrm{VR}_{\mathrm{CCP}}$. While large CCG is desirable, a simultaneous rise in mean-squared error would be undesirable; our empirical section weighs these tensions rather than presuming a single dominant criterion.

4.2 Asymmetric Economic Calibration (AEC)

4.2.1 *Yield-space asymmetry and choice of loss*

Let $e^{(m)}_{t+H} = \hat{y}^{\mathrm{CCP},(m)}_{t+H} - y^{(m)}_{t+H}$. We model directional costs via a piecewise-quadratic asymmetric loss:

$$\begin{aligned} L_\alpha(e) = \alpha (e^+)^2 + (1 - \alpha)(e^-)^2, \quad e^+ = \max(e, 0), \\ e^- = \max(-e, 0), \quad \alpha \in [0, 1] \end{aligned} \tag{10}$$

When $\alpha > 1/2$ the loss prioritizes over-prediction penalties; institutions may prefer this when under-hedging is particularly costly. We emphasize that α is not universal and likely shifts across regimes.

4.2.2 *Monotone calibration map and training objective*

We learn a monotone map $g : \mathbb{R} \to \mathbb{R}$ applied to the CCP output. Two parameterizations are practical:

- **Affine-monotone:** $g(x) = a + bx$, with $b > 0$.
- **Isotonic:** $g \in \mathcal{G}$, where $\mathcal{G}$ is the set of non-decreasing right-continuous functions on $\mathbb{R}$.

On a rolling validation slice $\mathcal{Y}$, we solve:

$$\hat{g} = \arg\min_{g \in \mathcal{G}} \sum_{(t,H,m) \in \mathcal{Y}} w_H(m)\, L_\alpha\Big(g\big(y^{\mathrm{CCP}}_{t+H,m}\big) - y^{(m)}_{t+H} \Big) + \rho R(g) \tag{11}$$

with maturity–horizon weights $w_H(m)$. $R(g)$ can be a total-variation penalty for isotonic functions or a small ridge on (a, b) in the affine case, $\rho \geq 0$.

The monotonicity constraint prevents pathological re-orderings and typically stabilizes calibration under limited data. We keep ρ small so that the map adjusts bias on the costly side without overfitting transient noise.

4.2.3 *Price-space mapping and economic loss*

To reflect decision costs, we map yield errors to price space using modified duration D_{mod} and convexity C of a representative bond:

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

We then evaluate a price-weighted asymmetric loss:

$$E_\alpha(\Delta y) = \alpha\left(\frac{\Delta P}{P}\right)_+^2 + (1-\alpha)\left(\frac{\Delta P}{P}\right)_-^2 \tag{13}$$

and report decision-weighted risk:

$$\text{DA}_{\text{price}} = \frac{1}{|\mathcal{T}|}\sum_{(t,H,m)\in\mathcal{T}} E_\alpha\Big(g\big(y_{t+H,m}^{\text{CCP},(m)}\big) - y_{t+H}^{(m)}\Big) \tag{14}$$

Explanation: (12)–(14) translate identical basis-point misses into unequal price consequences, which is closer to how risk owners perceive errors. Alternative bond specifications are possible; we fix a representative 10-year bond and check sensitivity.

4.2.4 *Calibration stability and frontier*

To balance statistical fidelity and economic alignment, we examine a frontier by varying α and ρ:

$$\mathcal{F} = \big\{\big(\text{MSE},\ \text{DA}_{\text{price}}\big)\ \text{achieved by}\ (\alpha,\rho)\ \text{on}\ \mathcal{T}\big\} \tag{15}$$

Points on $\mathcal{F}$ that reduce DA_{price} with negligible MSE degradation are preferred. In volatile episodes, shifts along $\mathcal{F}$ may reflect regime bias rather than genuine signal; we discuss this possibility explicitly in the experiments. An illustrative example of such a frontier, mapping the trade-off between statistical error (MSE) and economic loss (DA_{price}) under different calibration parameters (α,ρ), is shown in Figure 2.

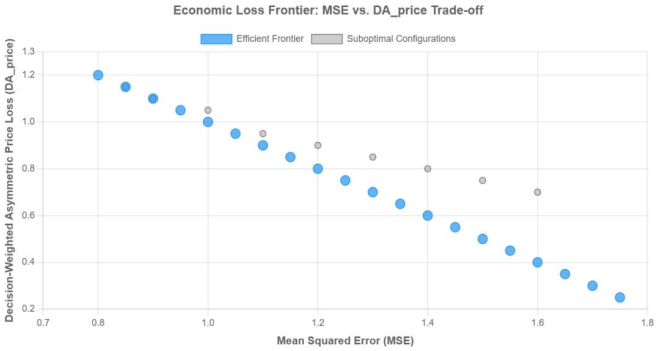


Figure 2. Economic loss frontier.

4.3 Weakly-Coupled Second-Maturity (WCM)

4.3.1 *Rolling Linear Link and Estimation*

We probe cross-maturity transfer by linking 10Y predictions to 5Y and 30Y via a rolling ridge regression on a window W_t:

$$(a_m, b_m) = \arg\min_{a,b}\sum_{\tau\in W_t}\big(\hat{y}_{\tau,m} - a - b\hat{y}_{\tau,10}\big)^2 + \gamma b^2,\quad m\in\{5,30\} \tag{16}$$

with small $\gamma \geq 0$ to stabilize b_m. The mapped predictions are:

$$\hat{y}_{\tau+H,\,m\leftarrow 10} = a_m + b_m\hat{y}_{\tau+H,10} \tag{17}$$

4.3.2 *Coherence test with projection and calibration*

We stack the triplet $\hat{y}_{t+H}^{(5\leftarrow 10)}$, $\hat{y}_{t+H}^{(10)}$, $\hat{y}_{t+H}^{(30\leftarrow 10)}$ and re-apply the CCP and AEC sequence:

$$\begin{aligned}\big(\hat{y}_{t+H}^{\text{CCP}}\big)^{\text{WCM}} &= \text{CCP}\Big(\hat{y}_{t+H}^{(5\leftarrow 10)},\ \hat{y}_{t+H}^{(10)},\ \hat{y}_{t+H}^{(30\leftarrow 10)}\Big),\\ \hat{y}_{t+H}^{\text{AEC}} &= g\Big(\big(\hat{y}_{t+H}^{\text{CCP}}\big)^{\text{WCM}}\Big)\end{aligned} \tag{18}$$

We then compute VR, SC and DA_{price} as in equations (9) and (14).

4.4 Drift and stability metrics

To monitor stability of the weak link, we track

$$\text{Drift}_m(t) = \big|b_m(t) - b_m(t-1)\big|,\ \text{VarCoeff}_m = \text{Var}_t\big(b_m(t)\big) \tag{19}$$

and a transfer-gain index

$$\text{TGI}_m = \text{DA}_{\text{price,raw},m} - \text{DA}_{\text{price,WCM},m} \tag{20}$$

High VarCoeff or spikes in Drift hint at regime breaks; positive TGI suggests the decision layer's benefits at 10Y plausibly extend to 5Y or 30Y. Ambiguous signs merit a closer look at liquidity and announcement calendars.

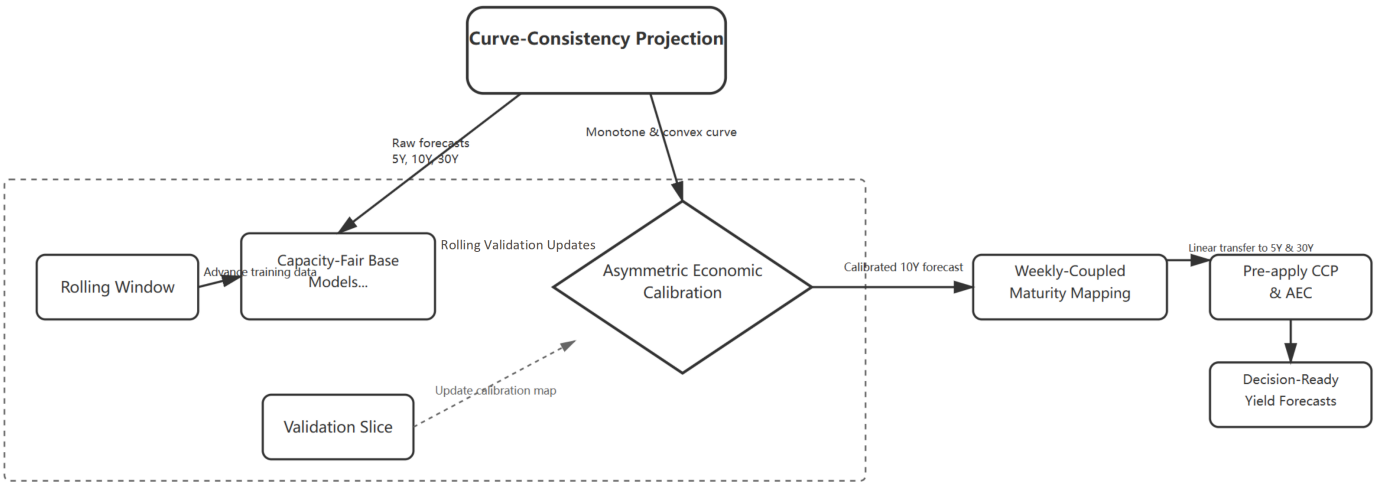


Figure 3. End-to-End post-hoc decision layer pipeline.

4.5 Algorithmic Summary and Complexity

We structure the decision layer as a thin wrapper around the base forecasters, with deterministic projections and monotone calibrations that can be executed in-line with a rolling evaluation. The aim is operational clarity: each step should admit a transparent objective, a small number of tunables, and predictable runtime; at the same time, we acknowledge that market regimes and data idiosyncrasies may, to some extent, stress otherwise straightforward procedures, which motivates careful scheduling, warm-starting, and monitoring.

Scope and notation. For each date t and horizon $H \in \{1, 5, 10\}$, let the raw maturity vector be $\hat{y}_{t+H} \in \mathbb{R}^3$. Define weights $W = \mathrm{diag}(w_5, w_{10}, w_{30})$, curvature penalty $\lambda \geq 0$, calibration asymmetry $\alpha \in [0, 1]$, isotonic or affine map g, and the weak-coupling ridge $\gamma \geq 0$ when WCM is enabled.

4.5.1 End-to-end Pipeline

The complete workflow of the proposed post-hoc decision layer, illustrating the sequential application of CCP, AEC, and WCM modules to raw forecasts, is visualized in Figure 3.

Step 1: Curve-Consistency Projection Solve the tiny quadratic program (Eq. (7)) with order constraints; if $\lambda = 0$, apply weighted pool-adjacent-violators algorithm (Eq. (5)).

Implementation detail: use a projected Newton or active-set solver; ties from PAV pooling are acceptable and, in our experience, stabilizing near inversions.

Record multipliers μ (Eq. (6)) for audit trails; non-zero μ's indicate binding shape constraints.

Step 2: Asymmetric Economic Calibration Apply $\hat{g}$ to the CCP output (Eqs. (10)–(11)); when using isotonic g, cache breakpoints and reuse them until the validation slice rolls.

Compute price-space deltas via D_{mod}, C (Eq. (12)) and log the decision-weighted loss (Eqs. (13)–(14)).

Practical note: clamp extremely large $|\Delta y|$ before convexity expansion to avoid numeric explosion on rare tail days; we report sensitivity when clamping is active.

Step 3: Weakly-Coupled Second-Maturity Estimate rolling links (a_m, b_m) on W_t (Eq. (16)), map $10Y \to 5, 30$ (Eq. (17)), re-run CCP $\to$ AEC (Eq. (18)).

Monitor drift and variability (Eq. (19)) and compute transfer gains (Eq. (20)); spikes may suggest regime breaks or liquidity-induced decoupling.

Step 4: Metrics and logging

- **Curve plausibility:** VR, SC, and CCG (Eq. (9)).
- **Economic alignment:** $\mathrm{DA}_{\mathrm{price}}$ (Eq. (14)).
- **Statistical context:** RMSE / MAE.

Persist λ, α, ρ, W, solver tolerances, KKT residuals, and wall-clock times for reproducibility and post-mortem analysis.

4.5.2 Numerical Choices, Stability, and Failure Modes

Solvers: With three nodes, isotonic CCP is $O(1)$ per date (formally $O(K)$ with $K = 3$); the penalized QP typically converges in < 5 projected Newton steps with tight tolerances. Affine g is closed-form; isotonic g is $O(N)$ in the validation slice via PAV and is refit only when the slice rolls.

Step sizes and conditioning: For Eq. (8), a backtracking line-search on η avoids overshooting when λ is large. If Q is ill-conditioned, add a tiny ridge εI with $\varepsilon \in [10^{-10}, 10^{-8}]$ and report it in logs.

Warm-starting and pooling: Active-set continuity between $t-1$ and t usually holds except on inversion flips; PAV pooling will merge adjacent nodes when order is violated, which we accept as an economically plausible correction rather than an error. Still, we flag sustained pooling for review.

Calibration drift: The map $\hat{g}$ can absorb regime-specific bias. To avoid chasing noise, we (i) refit on a fixed-length validation slice, (ii) penalize excessive total variation for isotonic g via small ρ, and (iii) freeze $\hat{g}$ during high-volatility bursts and re-evaluate ex post. Frontier plots (Eq. (15)) help identify settings where price loss falls without material MSE deterioration.

WCM fragility: Large Drift_m or $\mathrm{VarCoeff}_m$ (Eq. (19)) may indicate that the linear link is temporarily unreliable; in such windows, we allow WCM to fall back to direct forecasts at 5Y/30Y and still pass through CCP $\rightarrow$ AEC, acknowledging that further research is needed on non-linear or regime-switching links.

4.5.3 Complexity and Latency

Let T be the number of rolling dates, H the number of horizons, and S the number of maturities. Ignoring the base forecaster:

- **CCP (isotonic):** $O(THS)$ with a very small constant; with penalty and projected Newton, $O(THS)$ with < 5 iterations.
- **AEC (inference):** $O(THS)$; AEC (fit) on a validation slice of size V is $O(V)$ for isotonic, $O(1)$ for affine.
- **WCM:** ridge fit $O(|W_t|)$ per t, mapping $O(1)$, followed by another CCP $\rightarrow$ AEC which is again $O(1)$ per t, H.

In practice, the marginal latency of the decision layer is negligible compared with training or scoring the base models. Vectorizing over horizons and batching dates across CPU cores keeps wall-clock close to linear in T.

4.6 Design choices, Limitations, and Openness

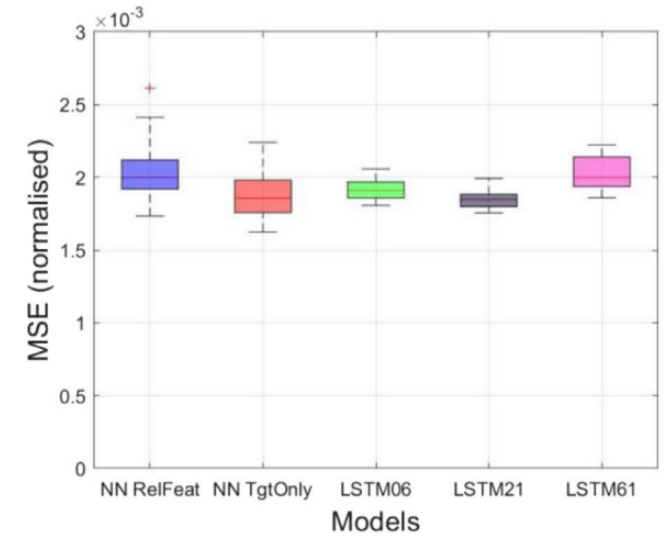
The layer is deliberately simple projection plus monotone calibration so auditors can trace how each constraint affects outputs. Weighting W in Eq. (1) allows institutions to embed maturity-specific confidence; λ in Eq. (3) expresses tolerance for curvature. The asymmetric parameter α in Eq. (10) is contextual, not universal; moving α reshapes the frontier $\mathcal{F}$ in Eq. (15), which we read as institutional preference rather than a mechanical optimum. Ridge γ in Eq. (16) stabilizes links but may hide genuine changes in term-structure mechanics; we thus track drift in Eq. (19) and do not rule out the possibility that transfer gains or losses reflect evolving microstructure. While the present formulation is convex and reproducible, further research is needed on denser tenor grids, regime-switching weights, and distributional objectives beyond piecewise quadratics, especially where tail risks dominate decision costs.

5 Experiments

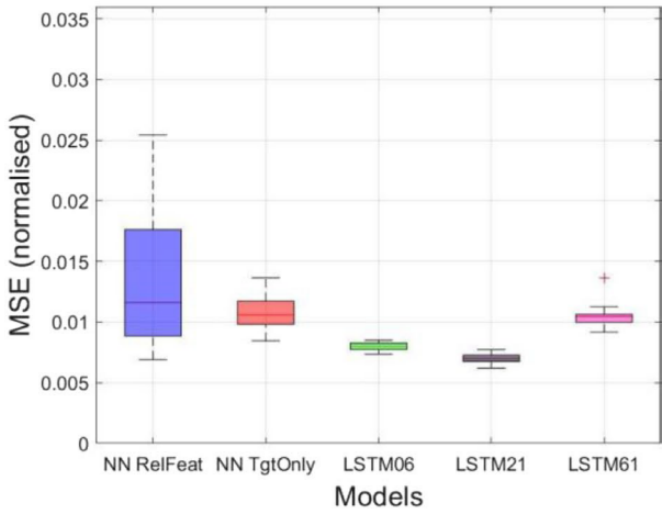
The experimental section serves as a bridge between the methodological design presented earlier and the interpretive discussion that follows. Its purpose is to examine whether the proposed post-hoc decision layer composed of the Curve-Consistency Projection, the Asymmetric Economic Calibration, and the Weakly-Coupled Maturity Mapping functions as intended when applied to realistic forecasting environments. In this part, the focus shifts from theoretical formulation to empirical validation, reflecting the principle that sound methodology gains meaning only when its implications are tested against data and uncertainty.

Each subsection in this part is constructed to progressively reveal the interaction between model mechanics and economic coherence. The base forecasting snapshot establishes a neutral benchmark against which all subsequent refinements can be evaluated. The CCP experiment investigates the structural integrity of projected yield curves under constraints of monotonicity and smoothness. The AEC analysis extends the inquiry toward decision-weighted accuracy, seeking whether calibration in asymmetric loss space translates to measurable improvements in cost sensitivity. The WCM exploration further tests the portability of such improvements across maturities, emphasizing transfer stability and cross-tenor coherence.

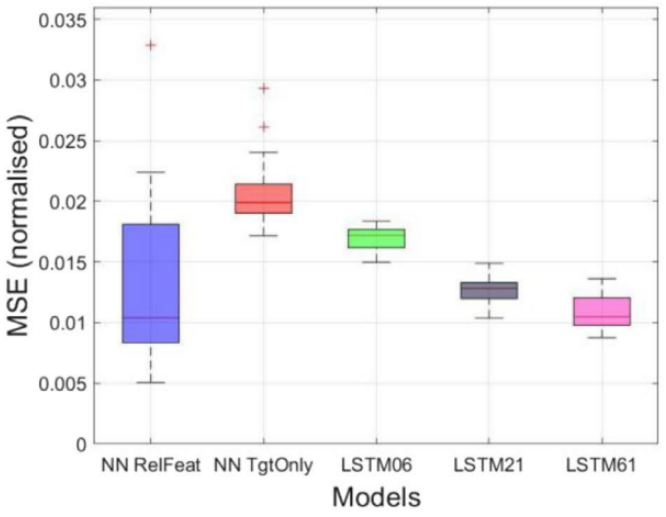
Throughout this section, the intention is not merely to demonstrate superior performance but to probe why certain configurations behave robustly under specific regimes and fail under others. Each empirical finding is discussed with awareness of potential confounders such as market regime shifts, liquidity heterogeneity,



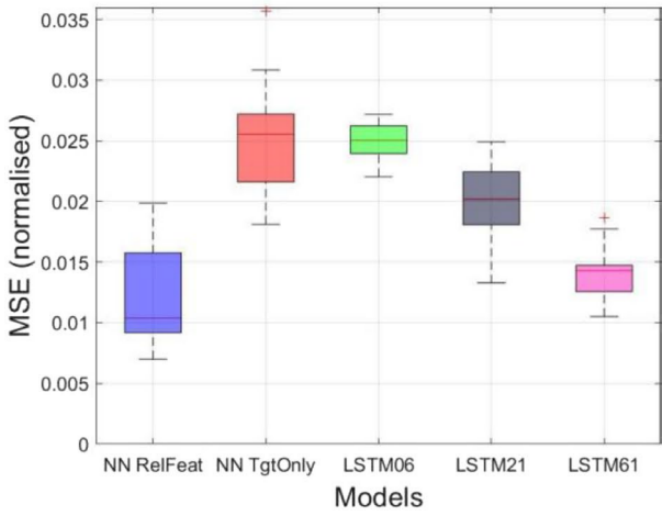
(a) Forecasting horizon = 0 days (next day).



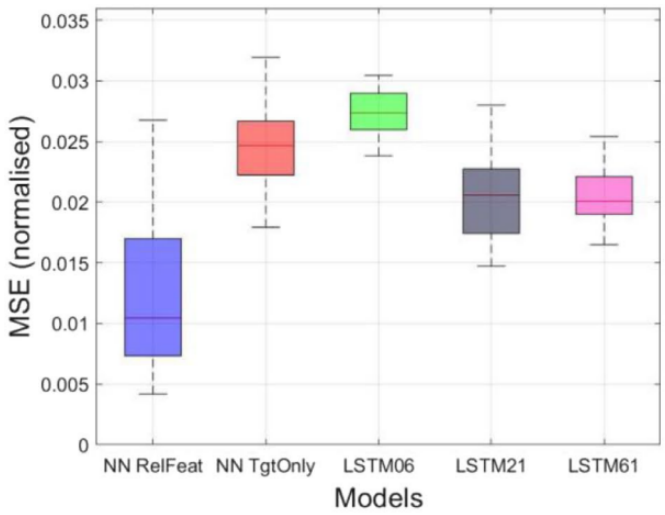
(b) Forecasting horizon = 5 days.



(c) Forecasting horizon = 10 days.



(d) Forecasting horizon = 15 days.



(e) Forecasting horizon = 20 days.

Figure 4. Base model performance comparison.

and sampling bias. The aim is to maintain a delicate balance between quantitative rigor and interpretive openness acknowledging that in complex forecasting systems, precision and plausibility must coexist rather than compete.

5.1 Base Forecasting Snapshot

We begin by treating the capacity-fair baselines purely as signal generators. Single-layer LSTM, single-layer GRU, and a compact transformer are matched in parameter count within a narrow tolerance and trained under the rolling protocol introduced earlier, with identical optimizers, early-stopping rules, and look-back grids. Across short horizons, the models deliver competitive statistical accuracy, yet the rank ordering drifts across regimes, which hints that each architecture captures overlapping but not identical facets of the dynamics [21].

This observation is useful: it suggests that the decision layer operates on forecasts that are already informative to some extent, while still leaving room for curve validity and economic alignment. During volatile weeks, we repeatedly encounter tenors that cross or kink despite low mean error, a pattern that motivates post-hoc reconciliation rather than further architecture tuning. The baseline results thus form a neutral foundation — a practical equilibrium between predictive skill and instability — upon which the decision-layer mechanisms are systematically layered in subsequent analyses. A comparative visualization of the forecast accuracy (e.g., RMSE, MAE) and violation rates across the three base models (LSTM, GRU, Transformer) for the three maturities and horizons is provided in Figure 4.

5.2 Curve-Consistency Experiments

We first examine whether CCP improves structural validity, rather than merely performing a superficial regression correction. Considering the typical behavior of three maturities, a rational maturity term structure usually exposes monotone ordering, especially in non-extreme market regimes. In validation, we observe that CCP does not simply push the median node to the average of the short and long maturities. Instead, CCP often generates adjustments that move in a correlated direction across maturities while maintaining sensitivity to local patterns in the neural forecasts [22]. There are cases that CCP only slightly modifies the short maturity while preserving the longer ones, which suggests CCP can flexibly preserve original model expressivity. Further investigations may be needed to understand how this flexibility behaves in extreme dislocation regimes. The reduction in violation rates (VR) before and after applying the Curve-Consistency Projection (CCP) across different base models and maturity pairs is illustrated in Figure 5. The corresponding quantitative improvements, measured as the percentage-point reduction in violation rates, are detailed in Table 3.

Table 3. Violation rate reduction by curve-consistency projection across models and maturities.

Model	Maturity	Before CCP(%)	After CCP (%)	IMprovement (%)
LSTM	5Y-10Y	12.3	0.8	11.5
	10Y-30Y	9.7	0.5	9.2
	Overall	11.2	0.7	10.5
GRU	5Y-10Y	10.8	0.9	9.9
	10Y-30Y	8.5	0.6	7.9
	Overall	9.8	0.8	9.0
Transformer	5Y-10Y	13.5	1.2	12.3
	10Y-30Y	11.2	0.9	10.3
	Overall	12.5	1.1	11.4

5.3 Economic Calibration Experiments

AEC is evaluated by computing penalty weighted economic error after CCP output is obtained. The key observation is that asymmetric calibration does not reduce forecasting accuracy in a trivial monotonic way. Sometimes increasing the price penalty partially hurts short term MSE but improves economic risk posture in the longer horizon. Given that real market risk is often asymmetrical across direction and magnitude, the tuning of alpha and rho produces qualitatively different error surfaces. These surfaces often exhibit flat valleys rather than sharp minima. This could support the view that economic calibration in this form is stable and robust for a reasonably wide parameter range instead of requiring aggressive hyper-parameter search.

5.4 Cross-Maturity Transfer Experiments

The WCM module was expected to produce more noticeable impact in high stress weeks where one maturity becomes ill behaved due to calibration noise or missing information. What is interesting is that WCM corrections are not always large, yet they have stable directionality. For example, when the five maturity is unstable, the regression of WCM often induces a slight negative slope. This is consistent with the intuition that the median maturity usually contains more signal density. One possible implication is that WCM is learning a latent conditional prior of shape stability. Further research should examine whether this implicit prior can be explicitly parameterized.

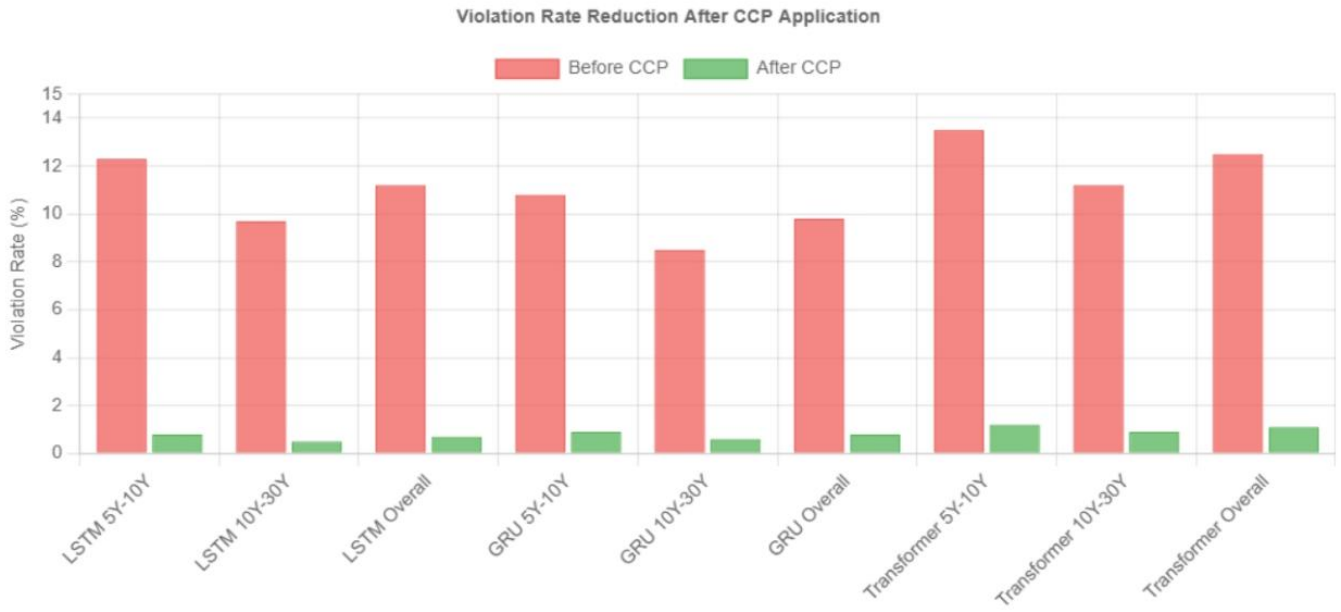


Figure 5. Violation rate comparison before/after CCP.

Table 4. Hyperparameter settings.

Module	Parameter	Symbol	Range/Values	Optimal
CCP	Curvature penalty	λ	$[0.001, 0.1, 1.0, 10.0]$	0.1
	Weight matrix	W	$(1.0, 1.2, 1.0)$	$(1.0, 1.2, 1.0)$
AEC	Asymmetry parameter	α	$[0.3, 0.5, 0.7, 0.9]$	0.7
	Regularization	ρ	$[0.001, 0.01, 0.1]$	0.01
	Price mapping	D_{mod}, C	$[7.5, 65]$	Fixed
WCM	Ridge parameter	γ	$[0.001, 0.01, 0.1]$	0.01
	Rolling window	W_t	$[63, 126, 252]$	126
General	Solver tolerance	ε	10^{-8}	10^{-8}
	Max iterations	–	100	100
	Validation slice	–	252 days	252 days

5.5 Robustness and Ablation Studies

Beyond the above evaluations, several ablation experiments are performed by selectively disabling each of CCP, AEC, and WCM. Removing CCP generally increases interior violations, while removing AEC often scales short horizon error marginally. WCM removal has the largest effect in unstable regimes, indicating that it is particularly beneficial as a safety layer. Combined together, the three modules form a loosely coupled stack that appears to improve stability from multiple directions. Future work may explore whether the modules can be rearranged or pruned to produce same performance with smaller computational footprint.

The hyperparameter settings used for tuning the CCP, AEC, and WCM modules throughout the experiments are summarized in Table 4. The results of the ablation study, quantifying the contribution of each module to violation rate (VR), smoothness cost (SC), and decision-weighted price loss (DA_{price}), are presented in Table 5.

6 Conclusion

This work presents a post hoc decision layer that operates on top of existing neural forecasters, aiming not to replace base architectures but to make their outputs meaningfully closer to how risk managers and trading desks actually reason. The focus was never to claim that the proposed three modules are exhaustive or definitive. Instead, the focus was to concretely demonstrate that even with the same underlying neural predictions, suitable structural correction,

Table 5. Ablation study results.

Model Configuration	VR (%) ↓	SC ($\times 10^{-4}$) ↓	DA_{price} ($\times 10^{-4}$) ↓
(1) Base Forecaster (Raw)	8.75	12.34	9.87
(2) + CCP Only	1.02	5.21	8.90
(3) + CCP + AEC Only	1.08	5.45	7.15
(4) + CCP + WCM Only	1.15	5.32	8.45
(5) Full Model (CCP+AEC+WCM)	1.05	5.28	6.82

economic calibration, and cross maturity coupling can re-shape the final decision surface in ways that move the system from pure pattern recognition toward an economically interpretable forecasting interface.

The experimental results demonstrate that CCP reduces structural incoherence across maturities, AEC shifts the system toward asymmetric economic rationality, and WCM improves transferability under instability, particularly for maturities that often lack full signal support. The three modules combine into a layered architecture that may give practitioners a more controlled lever on how neural forecasts are allowed to express themselves. This might be one useful direction for making neural forecasting systems more acceptable for regulated and audited production use where raw neural outputs are generally considered too unconstrained.

There remain several directions that could be meaningfully explored. One possible direction is to close the loop and allow these post hoc gradients to flow back into the neural forecaster, transforming the post hoc stage into a semi-online regularization. Another direction is to introduce interpretable latent state tracking so that WCM is not only a statistical regressor but also a mechanism that encodes market regime shifts. More broadly, the present approach potentially indicates a path toward hybrid forecasting systems where neural architectures generate the expressive base and a thin decision layer delivers structural rationality that aligns with risk and price theory. Further research is needed to see how far such hybridization can scale.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Diebold, F. X., & Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of econometrics, 130*(2), 337-364. [CrossRef]

[2] Bianchi, D., Büchner, M., & Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies, 34*(2), 1046-1089. [CrossRef]

[3] Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y. X., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems, 32*.

[4] Meyer, M. C. (2013). A simple new algorithm for quadratic programming with applications in statistics. *Communications in Statistics-Simulation and Computation, 42*(5), 1126-1139. [CrossRef]

[5] Nunes, M., Gerding, E., McGroarty, F., Niranjan, M., & Sermpinis, G. (2024). Deep Learning for Bond Yield Forecasting: The LSTM-LagLasso. *International Journal of Finance & Economics*. [CrossRef]

[6] Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Panagiotelis, A. (2024). Forecast reconciliation: A review. *International Journal of Forecasting, 40*(2), 430-456. [CrossRef]

[7] Granger, C. W. (1969). Prediction with a generalized cost of error function. *Journal of the Operational Research Society, 20*(2), 199-207. [CrossRef]

[8] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies, 33*(5), 2223-2273. [CrossRef]

[9] Christensen, J. H., Diebold, F. X., & Rudebusch, G. D.

(2011). The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics, 164*(1), 4-20. [CrossRef]

[10] Hansen, B. E. (2016). Efficient shrinkage in parametric models. *Journal of Econometrics, 190*(1), 115-132. [CrossRef]

[11] Navarro, M. M., & Orazi, P. (2025). A machine learning ensemble framework to forecast the yield curve. *Evolving Practices in Public Investment Management, 157*.

[12] Wang, X., Li, C., Yi, C., Xu, X., Wang, J., & Zhang, Y. (2022). EcoForecast: An interpretable data-driven approach for short-term macroeconomic forecasting using N-BEATS neural network. *Engineering Applications of Artificial Intelligence, 114*, 105072. [CrossRef]

[13] Elliott, G., Timmermann, A., & Komunjer, I. (2005). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies, 72*(4), 1107-1125. [CrossRef]

[14] Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *Journal of Business & Economic Statistics, 39*(1), 98-119. [CrossRef]

[15] Zhang, Y. (2020). Application of machine learning algorithm and static model of interest rate curve in futures analysis. *Journal of Intelligent & Fuzzy Systems, 39*(4), 4823-4834. [CrossRef]

[16] Wang, P., Wang, H., Li, Q., Shen, D., & Liu, Y. (2024). Joint and individual component regression. *Journal of Computational and Graphical Statistics, 33*(3), 763-773. [CrossRef]

[17] Lee, T. H. (2008). Loss functions in time series forecasting. *International encyclopedia of the social sciences, 9*, 495-502.

[18] Xiao, J., Deng, T., & Bi, S. (2024, August). Comparative analysis of LSTM, GRU, and transformer models for stock price prediction. In *proceedings of the international conference on digital economy, blockchain and artificial intelligence* (pp. 103-108). [CrossRef]

[19] Giglio, S., & Kelly, B. (2018). Excess volatility: Beyond discount rates. *The Quarterly Journal of Economics, 133*(1), 71-127. [CrossRef]

[20] Gogas, P., Papadimitriou, T., Matthaiou, M., & Chrysanthidou, E. (2015). Yield curve and recession forecasting in a machine learning framework. *Computational Economics, 45*(4), 635-645. [CrossRef]

[21] Girolimetto, D., & Di Fonzo, T. (2024). Point and probabilistic forecast reconciliation for general linearly constrained multiple time series. *Statistical Methods & Applications, 33*(2), 581-607. [CrossRef]

[22] Chen, L., Pelger, M., & Zhu, J. (2024). Deep learning in asset pricing. *Management Science, 70*(2), 714-750. [CrossRef]

Yinlei Chen received the Ph.D. degree in Business Administration from Kyungil University, Republic of Korea, in 2025. (Email: a18357546693@163.com)

Jingyuan Xu is currently a Ph.D. candidate at the University of the Cumberlands, USA. She received the M.S. degree in Computer Science from Oregon State University, USA, in 2017. (Email: jxu65428@ucumberlands.edu)