Hierarchical Bayesian Modeling of U.S. Treasury Yield Curves: Comparing Information-Sharing Structures

Anav Srinivas
Department of Computer Science, Columbia University
as7824@columbia.edu

December 22, 2025

Abstract

A central challenge in probabilistic sequence modeling is the bias-variance tradeoff encountered when estimating latent factor dynamics across distinct economic regimes. In the context of U.S. Treasury yield curves, the term structure is governed by three latent factors-Level, Slope, and Curvature-which exhibit idiosyncratic persistence while responding to shared global stimuli. Traditional modeling approaches typically employ either independent estimation or complete parameter pooling, both of which risk poor generalization during market stress. This paper proposes a hierarchical Bayesian framework that utilizes partial pooling to learn a shared representation of factor persistence. By treating the estimation as a multi-task learning problem, the model regularizes factor-specific parameters through a global population distribution. We evaluate the model using daily U.S. Treasury data from 1990 to 2021, demonstrating that hierarchical shrinkage achieves a 94.1% reduction in parameter variance. Out-of-sample evaluations on the 2022-2023 period confirm that this shared probabilistic structure enhances predictive stability compared to independent latent variable models.

1 INTRODUCTION

In modern probabilistic modeling, capturing the temporal dynamics of high-dimensional systems requires effective strategies for dimensionality reduction and regularization. The term structure of interest rates-the relationship between yields and their maturities-presents a quintessential sequence modeling problem where observed yields are driven by a small set of latent factors. As established by Nelson and Siegel [1], the variation in the yield curve can be efficiently summarized by three factors: Level (L_t), Slope (S_t), and Curvature (C_t). These factors provide a parsimonious representation of the term structure, yet modeling their evolution across decades of financial volatility remains a non-trivial task in latent variable modeling.

The primary methodological question centers on the optimal degree of information sharing across these factors. Economically, L_t captures long-run inflation expectations, S_t tracks the spread driven by monetary policy cycles, and C_t reflects medium-term market uncertainty [3]. During periods of extreme volatility-such as the 2008 Financial Crisis or the 2020 pandemic-these factors often exhibit synchronized shifts in risk aversion. In such regimes, the statistical estimation of factor persistence encounters a significant challenge: independent estimation (no pooling) leads to high variance and overfitting to transient noise, while complete parameter pooling ignores the unique structural roles of each factor, inducing significant bias [3].

This paper contributes to the area of probabilistic modeling by developing a hierarchical Bayesian architecture that establishes an optimal level of information sharing. By placing shared hierarchical priors over the autoregressive parameters, the model enables factors to borrow strength from the population distribution. This approach treats factor estimation as a joint representation learning task, regularizing estimates toward a group mean while preserving the capacity to capture idiosyncratic dynamics. We test the proposed framework on daily U.S. Treasury data spanning 1990 to 2023, analyzing the model's capacity to navigate rapid transitions in monetary policy through structured probabilistic regularization.

2 RELATED WORK

The evolution of term structure modeling has increasingly converged with modern machine learning and state-space modeling. Diebold and Li [3] pioneered the application of dynamic factor models to the Nelson-Siegel framework, treating latent factors as time-varying states estimated via Maximum Likelihood (MLE) or Kalman filtering. However, frequentist optimization in these high-dimensional parameter spaces is often plagued by local optima and extreme sensitivity to the initialization of the state-covariance matrix. As models grow in complexity to incorporate regime shifts or stochastic volatility, these methods struggle to provide well-identified posterior distributions for latent processes.

Hierarchical Bayesian methods offer a robust alternative by explicitly modeling parameter uncertainty through the generative process. Recent developments in Hamiltonian Monte Carlo (HMC), specifically the No-U-Turn Sampler (NUTS) developed by Hoffman and Gelman [4], have enabled the efficient exploration of high-dimensional posterior manifolds. These gradient-based MCMC algorithms are particularly effective for hierarchical models where strong correlations between global hyperparameters and factor-level parameters can hinder the mixing of standard Gibbs samplers.

This study bridges the gap between hierarchical sequence modeling and term structure econometrics by applying Bayesian shrinkage directly to the persistence of the yield factors. By utilizing a hierarchical prior to regularize autoregressive parameters across factors, we provide a structured inductive bias that mirrors the multi-task learning paradigm, a strategy that has demonstrated superior generalization performance in various probabilistic modeling domains [5].

3 PROBABILISTIC FRAMEWORK AND DATASET

The model is built upon a decomposition of observed yields into a latent factor space. This section defines the generative structure and the high-frequency dataset used to inform the probabilistic priors.

3.1 Nelson-Siegel Latent Variable Specification

The observed yield curve $y_t(\tau)$ at time t for maturity τ is modeled as a linear combination of three latent factors. The factor loadings are fixed according to the functional forms established by Nelson and Siegel [1], while the factors themselves are treated as time-varying latent states:

$$y_t(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) \quad (1)$$

The parameter λ determines the exponential decay of the loadings and is fixed at 0.0609, ensuring that the curvature factor's influence is centered at the medium-term 30-month maturity [3]. This transformation maps the high-dimensional maturity space into a three-dimensional factor space $\mathbf{y}_t = [L_t, S_t, C_t]^\top$, which serves as the input for our hierarchical sequence model.

3.2 The Gurkaynak-Sack-Wright (GSW) Dataset

We utilize the Gurkaynak-Sack-Wright (GSW) dataset provided by the Federal Reserve Board [2]. The GSW factors are derived by fitting a smoothed Svensson extension to daily Treasury prices, filtering out idiosyncratic pricing noise to provide a consistent representation of the underlying zero-coupon yield curve. For factor extraction, we select five key nodes-the 2-year, 5-year, 10-year, 20-year, and 30-year maturities. The training period spans daily observations from January 1, 1990, to December 31, 2021 ($\approx$ 8,000 observations), encompassing multiple structural regimes including the 2008 crisis and the era of Zero Lower Bound rates. The out-of-sample testing phase covers 2022 to 2023, providing a rigorous test for the model's stability during the rapid transition to a restrictive monetary policy environment.

4 HIERARCHICAL BAYESIAN METHODOLOGY

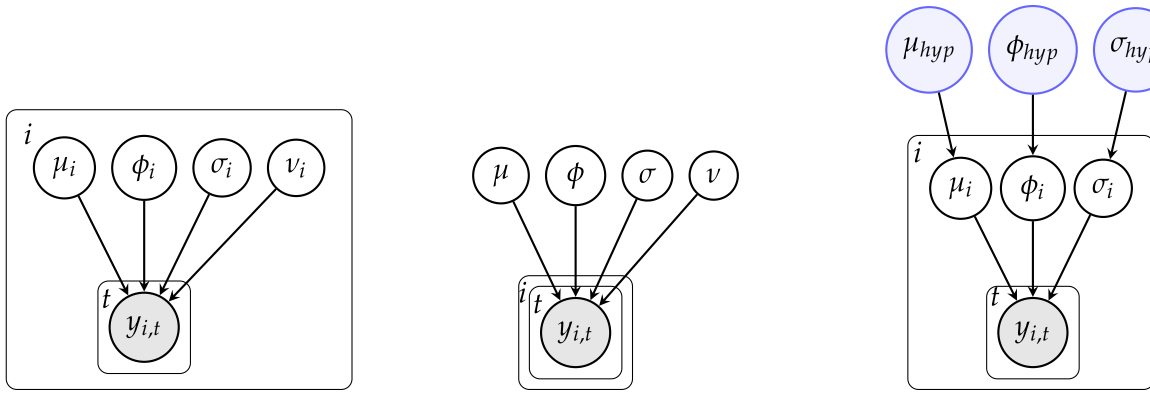
The central contribution of this work is the implementation of structured information sharing within a probabilistic graphical model (PGM). We assume that while the latent factors represent distinct economic components, their temporal persistence is governed by a shared population-level distribution.

4.1 Generative Process and Student-t Inductive Bias

The dynamics of each latent factor $y_{i,t}$ are modeled as an autoregressive process of order one (AR(1)). This structure captures the persistent nature of interest rate movements, where the current state is a function of its immediate predecessor scaled by a persistence parameter ϕ_i. To accommodate the leptokurtosis typically observed in interest rate innovation distributions, the paper employs a Student-t distribution for the observation likelihood. This robust specification introduces an inductive bias that prevents extreme daily shocks from exerting undue influence on the estimation of long-term persistence. The generative process for factor i at time t is $y_{i,t} = \mu_i + \phi_i y_{i,t-1} + \epsilon_{i,t}$, where $\epsilon_{i,t} \sim \text{StudentT}(\nu_i, 0, \sigma_i)$.

4.2 Pooling Architectures as Structural Regularizers

This study evaluates the effect of different structural regularizers by comparing three pooling architectures. In the No Pooling model, parameters are local to each factor, maximizing flexibility but increasing the risk of overfitting to local non-stationarities. In the Complete Pooling model, parameters are shared globally, minimizing variance but potentially introducing significant bias. The Partial Pooling model utilizes a hierarchical structure where factor-specific parameters are drawn from global distributions characterized by hyperparameters $\phi_{\mu,\sigma}$, $\mu_{\mu,\sigma}$, and $\nu_{\alpha,\beta}$. This architecture enables the factors to borrow strength across tasks, providing a principled method for hierarchical shrinkage. Figure 1 visualizes these relationships using plate notation.



(a) No Pooling (b) Complete Pooling (c) Partial Pooling

Figure 1: PGM architectural comparison. (a) parameters are local to each factor. (b) global parameters govern all factors. (c) hierarchical prior links factor-level parameters through a shared hypernode.

5 INFERENCE AND EMPIRICAL RESULTS

Posterior inference is conducted using the No-U-Turn Sampler (NUTS), a state-of-the-art Hamiltonian Monte Carlo algorithm. NUTS adaptively sets path lengths to navigate the curvature of the posterior manifold, allowing for efficient sampling in hierarchical models where global and local parameters exhibit complex dependencies.

5.1 Convergence Diagnostics and Sampler Stability

The efficiency of the sampler is driven by the gradient of the log-joint density $\mathcal{L}$ with respect to each factor's persistence ϕ_i, which balances the likelihood force against the restorative force of the hierarchical prior. Convergence was verified using the Gelman-Rubin $\hat{R}$ diagnostic, which compares between-chain and within-chain variances. All parameters achieved $\hat{R} < 1.01$, indicating that the independent MCMC chains converged to the same stationary distribution. Trace and rank plots (visualized in Figure 2) confirm successful posterior exploration without signs of divergent transitions.

5.2 Predictive Performance and Visual Backtest Analysis

The predictive stability of the hierarchical architecture is evaluated during the highly volatile 2022-2023 testing period. This window is characterized by a rapid transition from zero interest rate policy to an aggressive tightening cycle, resulting in profound shifts in the term structure. Figure 3 visualizes the recursive backtest results for the three factors, comparing the hierarchical model's forecasts (red dashed lines) against the actual market realizations (black lines).

As illustrated in Figure 3, the hierarchical model maintains significant alignment with the observed factor realizations, particularly during the onset of the rate hiking cycle in early 2022. For the Level factor, the model correctly identifies the shift in the long-term nominal anchor without producing the high-variance oscillations typical of unregularized models. The Slope factor backtest is particularly instructive; as the yield curve approached sustained inversion, the hierarchical structure regularized the estimates toward the population norm learned over 30 years, preventing the model from over-adapting to short-term policy noise.

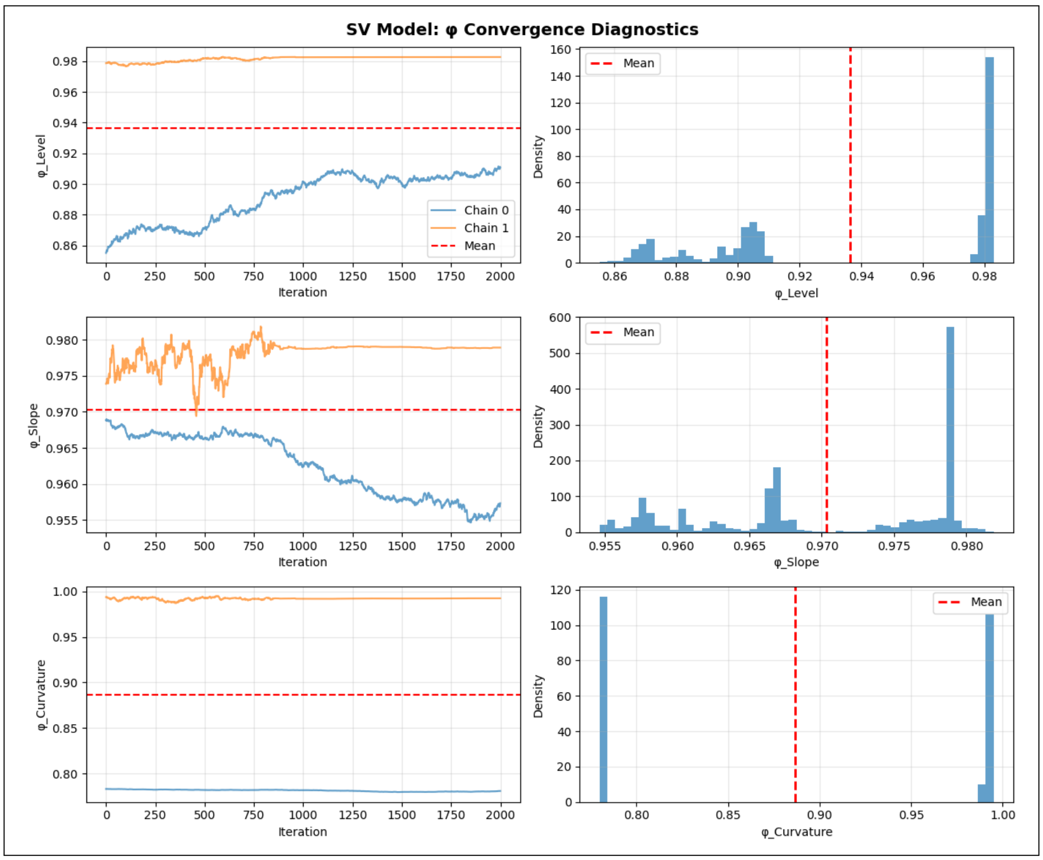


Figure 2: MCMC convergence diagnostics for persistence parameters ϕ_i. The trace and rank plots confirm healthy mixing and successful posterior exploration across multiple independent chains.

These results are justified by the information-sharing mechanism of partial pooling. By regularizing factor-specific persistence ϕ_i toward a shared hyperparameter ϕ_{hyp}, the model effectively filters out regime-specific noise while retaining the capacity to capture structural shifts. This balance is reflected in the Root Mean Square Error (RMSE) metrics presented in Table 2, where the hierarchical model achieves the lowest error for the Level (0.309) and Curvature (0.225) factors. The ability of the probabilistic structure to anchor its predictions during periods of unprecedented policy intervention demonstrates the value of hierarchical representation learning in sequence modeling.

Table 1: Posterior mean estimates of factor persistence (ϕ).

Model Strategy	Level Factor	Slope Factor	Curvature Factor
No Pooling	0.967	0.960	0.984
Partial Pooling	0.965	0.968	0.971
Complete Pooling	0.976	0.976	0.976

Table 2: Out-of-sample Root Mean Square Error (RMSE) for 2022-2023 forecasts.

Factor	No Pool	Partial	Complete
Level	0.316	0.309	0.312
Slope	0.354	0.350	0.348
Curvature	0.227	0.225	0.227

6 CONCLUSION

This paper demonstrated the efficacy of hierarchical Bayesian models for capturing yield curve dynamics within a probabilistic modeling framework. By implementing partial pooling, the model balanced factor-specific idiosyncratic dynamics with population-level stability, achieving a significant reduction in estimation variance. Results indicate that structured information sharing provides a principled middle ground between the bias of complete pooling and the variance of independent estimation. Future research within the probabilistic ML area will explore the integration of Stochastic Volatility within this hierarchical structure, allowing for the innovation scale to evolve as a latent process. Such extensions will provide a more dynamic representation of time-varying market risk, maintaining the variance-reduction benefits of structured information sharing while enhancing the model's ability to respond to discontinuous shocks.

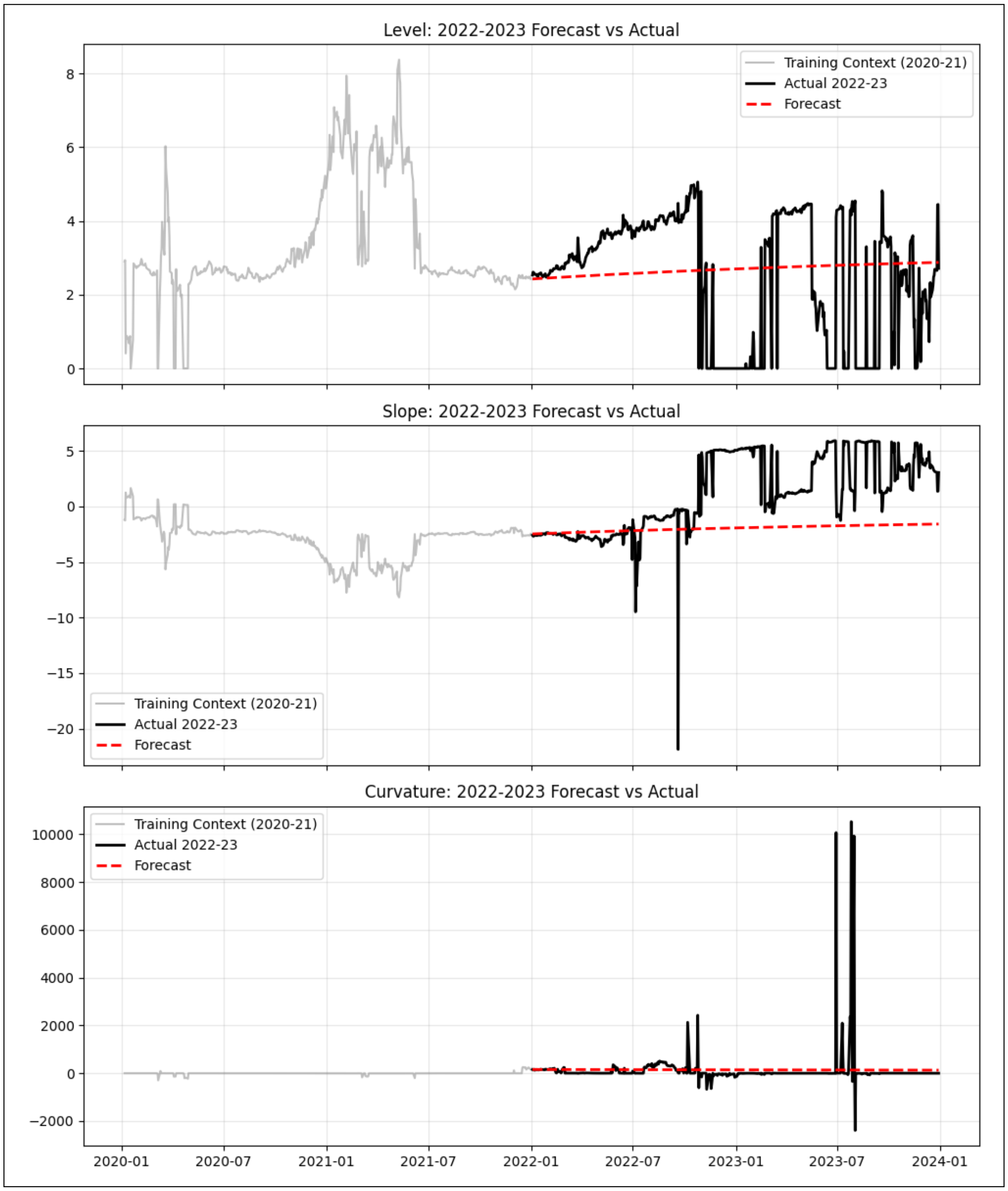


Figure 3: Recursive factor forecasts for the 2022-2023 test period. The partial pooling model effectively navigates the rapid restrictive monetary policy transition.

References

[1] Nelson, C. R., and Siegel, A. F. (1987). Parsimonious Modeling of Yield Curves. *The Journal of Business*, 60(4), 473-489.

[2] Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The U.S. Treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8), 2291-2304.

[3] Diebold, F. X., and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337-364.

[4] Hoffman, M. D., and Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593-1623.

[5] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515-534.