

https://dspace.org/

USVAffine Models Without Derivatives A Bayesian TimeSeries Approach

Malefane Molibeli; Gary Van Vuuren

https://hdl.handle.net/10539/47008
SCOPUS

Downloaded from a DSpace repository. https://dspace.org/

Journal of
Risk and Financial Management

MDPI

Article

USV-Affine Models Without Derivatives: A Bayesian Time-Series Approach

Malefane Molibeli [1,*] and Gary van Vuuren [2,3]

[1] School of Economics and Finance, University of the Witwatersrand, Johannesburg 2000, South Africa
[2] Centre for Business Mathematics and Informatics, North-West University, Potchefstroom Campus, Potchefstroom 2520, South Africa; 12001333@nwu.ac.za
[3] National Institute for Theoretical and Computational Sciences (NITheCS), Pretoria 0001, South Africa
* Correspondence: mhmolibeli@gmail.com

Abstract

We investigate the affine term structure models (ATSMs) with unspanned stochastic volatility (USV). Our aim is to test their ability to generate accurate cross-sectional behavior and time-series dynamics of bond yields. Comparing the restricted models and those with USV, we test whether they produce both reasonable estimates for the short rate variance and cross-sectional fit. Essentially, a joint approach from both time series and options data for estimating risk-neutral dynamics in ATSMs should be followed. Due to the scarcity of derivative data in emerging markets, we estimate the model using only time-series of bond yields. A Bayesian estimation approach combining Markov Chain Monte Carlo (MCMC) and the Kalman filter is employed to recover the model parameters and filter out latent state variables. We further incorporate macro-economic indicators and GARCH-based volatility as external validation of the filtered latent volatility process. The $A_1(4)$USV performs better both in and out of sample, even though the issue of a tension between time series and cross-section remains unresolved. Our findings suggest that even without derivative instruments, it is possible to identify and interpret risk-neutral dynamics and volatility risk using observable time-series data.

Keywords: parameter and model identifiability; stochastic volatility; MCMC; unspanned stochastic volatility

check for updates

Academic Editor: Thanasis Stengos

Received: 24 June 2025
Revised: 13 July 2025
Accepted: 14 July 2025
Published: 17 July 2025

Citation: Molibeli, M., & van Vuuren, G. (2025). USV-Affine Models Without Derivatives: A Bayesian Time-Series Approach. *Journal of Risk and Financial Management*, *18*(7), 395. https://doi.org/10.3390/jrfm18070395

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Interest rate volatility is increasingly recognized as a distinguished feature in ATSMs and an independent driver of time-varying risk premia in fixed-income markets (Duffee, 2002). Typically, volatility is unobservable and potentially unspanned by the prices of traded bonds, raising significant challenges for both modeling and empirical inference. In particular, the inability to observe volatility directly—especially in markets with sparse derivatives trades due to a lack of liquidity—necessitates models where volatility is treated as a latent factor. Such models must be capable of distinguishing between spanned and unspanned components of risk, a subtle but crucial distinction for pricing, hedging, and risk management.

Unspanned Stochastic Volatility (USV) refers to volatility risk that influences the dynamics of yields and risk premia but cannot be perfectly hedged or inferred from the cross-section of bond prices alone; see Duffee (2002). It results from a tension that exists between fitting a cross-section of yields and matching the time-series dynamics of bond

prices; see Collin-Dufresne and Goldstein (2002), Collin-Dufresne et al. (2008), Piazzesi (2010), and Riva (2024) among others. This introduces fundamental incompleteness into the fixed-income market, making it impossible to replicate all sources of risk through portfolios made up solely of bonds—suggesting the presence of portfolio volatility risk (Collin-Dufresne & Goldstein, 2002). The presence of USV can explain persistent empirical puzzles such as the failure of (ATSMs) to match observed option-implied volatilities or the dynamic behavior of the market price of risk (MPR) Duffee (2002), Bikbov and Chernov (2004), and Collin-Dufresne et al. (2009).

One of our central objectives is to evaluate the role of model restrictions in enabling or suppressing the presence of USV. In doing so, we aim to test affine models both with USV, leveraging off the maximal models from Dai and Singleton (2000) and Collin-Dufresne et al. (2008), which in turn, are characterized by a greater number of identifiable parameters. The representation that we follow leads to specifications that are more flexible than the canonical form identified by Dai and Singleton (2000). These are referred to as a maximal $A_M(N)$ form of ATSMs, which in some cases have more identifiable parameters. Our focus is on the impact of parameter restrictions on these maximal ATSMs, the three-factor $A_1(3)$ and four-factor $A_1(4)$ models (Collin-Dufresne et al., 2008). Subsequently, we compare and contrast both the $A_1(3)$USV and $A_1(4)$USV in terms of their time series and yield curve dynamics.

As in most emerging markets, we anticipate that market data constraints are significant due to non-liquid derivative trading. In such markets, data on derivatives may be sparse and irregular. Riva (2024) points out that this approach to test the presence of USV does not rely on derivatives market data to estimate the underlying risk factors under the physical measure. We consider the yield curve itself to reveal information about latent volatility. While the cross-section of yields provides insight about the first moment, higher-order dynamics—such as yield curvature, and term premium variation,—changes in the covariance structure of yields may be indirectly embedded within the latent volatility. Given our choice of models, we test whether the structures permit the identification of volatility as a latent but influential factor, even in the absence of explicit derivative prices.

There is a challenge of understanding the 'dual role' of matching the interest rate volatility within the ATSMs. Collin-Dufresne et al. (2008) reported significant discrepancies when comparing the stochastic volatility from non-Gaussian, square-root ATSMs to the observed state variable proxy of volatility. We explore the dual role of interest rate variance by considering the following aspects. First, we establish whether the three-factor $A_1(3)$ produces a time-varying volatility and whether the variance is also a factor for the yield curve. The $A_1(3)$ model was a preferred specification choice in Dai and Singleton (2000) for its ability to characterize the unconditional volatility and a flexible correlation structure. Second, there is a need to investigate the joint dynamics of level, slope, curvature and volatility.

The market price of risk specification should be chosen carefully to ensure that the tension in the simultaneous fitting of certain cross-sectional and time-series properties in the ATSMs is circumvented. The three-factor ATSMs with completely affine specification by Dai and Singleton (2000) were found to be poor in preserving the affine time-series and cross-sectional properties of bond prices. On the other hand, the essentially affine specification allows greater flexibility in fitting the time-varying price of interest rate risk over time, while preserving the affine time-series and cross-sectional properties of bond prices. This is despite the need for a trade-off between time-varying conditional variances and the flexibility in fitting time-varying market price of risk (Duffee, 2002).

Estimation methods will have a problem dealing with the explanation of bond yields under USV. Whereas USV can match some aspects of time series and cross-section of

yields, it is unclear whether USV restrictions will affect the model's ability to capture the cross-section and time series of yields (Collin-Dufresne et al., 2009). Econometrics methods, including generalized method of moments (GMM), simulated methods of estimation (SME), among others, are unsuitable for investigating models with USV (Collin-Dufresne et al., 2009) and references therein. Following their approach, we consider the Bayesian Markov Chain Monte Carlo (MCMC), with a Kalman filter integrated into it. The Kalman filter is preferred for its ability to address nonlinear processes. It is only used here as a computational devise to evaluate the likelihood given a path of the latent part V_t of the state variable.

Ultimately, we test the time series and term structure dynamics using multiple volatility representations: stochastic volatility (SV) from the models, GARCH-type dynamics, and macro variables such as oil shocks and FX volatility. This comparative approach helps us understand the robustness of results across different volatility specifications and provides evidence for or against the empirical importance of USV in fixed-income markets. To test the model performance and goodness-of-fit, we apply both the root-mean-square (RMSE) and biasness tests, including the pairwise comparison of Diebold and Mariano (2002) (DM) to the models.

The rest of this paper is structured as follows: In Section 2 we review the relevant literature. Section 3 covers the model establishment. Section 4 discusses a data collection method. In Sections 5–7 we discuss the model implementation and analysis of results. In Section 8 we conclude.

2. Literature Review

There are ongoing attempts aimed at successfully addressing the dynamic term structure models (DTSM) that exhibit USV. The presence of USV renders the fixed-income markets incomplete, therefore presenting some challenges to the pricing and hedging of bond prices and their derivatives. In recent studies, linear-rational square-root (LRSQ) models are being evaluated for the ability to capture cross-sectional and time-series dynamics simultaneously. Hansen (2025) specifies a five-factor model exhibiting USV, where two of the factors are not spanned by the yield curve. Their specification confirms the presence of USV in conditional yield variance and bond risk premia, which in turn are also linked to macro-economic uncertainty. Andreasen et al. (2025) tests the predictive power of the spread between short and long-term U.S government yields using both the quadratic term structure and the autoregressive gamma-zero models in comparison to LRSQ. Their results provide evidence that LRSQ allows USV, accurately fitting the cross-section of yields and capturing the time-series properties of the yield variance.

Specifically, the focus on ATSMs suggests other alternatives to the trade-off between fitting the cross-section of yields and capturing their time-series behavior. Riva (2024) considers the process where a quadratic variation of yields is tied to the cross-section of average yields. He implements arbitrary portfolios of bonds to test the presence of USV using jump-robust estimators of the diffusive variance of Nelson-Siegel factors. His approach follows Andersen and Benzoni (2010) on quadratic diffusive and affine jump-diffusive models, although they found them to be incapable of accommodating the observed yield volatility dynamics. Their model exhibits evidence for USV across all factors, which includes oil shocks and tax policy shocks, also driving part of the USV, although a greater part of the variation still requires some explanation.

Collin-Dufresne et al. (2009) study the "dual role" of the short rate variance as predicted by the stochastic volatility models. They provide evidence that the three-factor ATSMs explain the shape of the yield curve but fail to explain both the quadratic time-variation in spot rate as estimated from GARCH and the implied variance from options.

Evidence from the four-factor model with USV suggests that the model generates both realistic short rate volatility estimates and a good cross-sectional fit. The study suggests further that short rate volatility cannot be extracted from the cross-section of bond prices. They propose an alternative representation of ATSM by Dai and Singleton (2000) which is essentially characterized by—(1) physical interpretations of state variables such as level, slope and curvature, (2) affine dynamics and tractability are preserved, (3) models are econometrically identifiable, and (4) necessary and sufficient conditions lead to parameters restrictions under which the model exhibits USV (Collin-Dufresne et al., 2004). In their companion paper, Collin-Dufresne and Goldstein (2002), they derive these conditions under which restriction exhibits USV. The same paper provides evidence on additional explanatory power possessed by latent variables for explaining the time-series and cross-sectional behavior of bond prices. The presence of USV enforces the requirement that bond and derivative prices are required for the estimation of parameters. Finally, USV presents some challenges on how fixed-income derivatives should be hedged.

Bikbov and Chernov (2004) explores the differences among Gaussian, stochastic volatility, and USV models in the context of the conditional volatility, which is an inherent feature in ATSMs. They follow Collin-Dufresne and Goldstein (2002) approach and restrictions that ensure that models exhibit USV. They estimate the models using both Eurodollar Futures and options data. Gaussian and stochastic volatility models match the conditional mean and volatility of the term structure well. They require both the yield curve dynamics and options data for differences to be distinguished. USV models fail to resolve the tension between futures and options fits. Additional factors, such as jumps, should be considered for the model performance to be improved. Heidari and Wu (2002) maintains USV by allowing a clear separation between factors dedicated to term structure fitting and those for options fitting; see Bikbov and Chernov (2004) and references therein. Similarly, they also reported the tension in pricing and parallel tension in matching lower and higher-order moments, suggesting the direction of adding jump components. Ang et al. (2006) uses the term structure model with inflation and economic growth factors, together with latent variables, to investigate the effect of macro variables on bond prices and the yield curve dynamics. PCA and variance decomposition show that macro factors explain up to 85% of the variation in bond yields. They explain movements at the short end and middle of the yield curve, while latent factors still account for most of the movement at the long end of the yield curve. They conclude that incorporating macro factors in a term structure model improves forecasts.

Another contributing factor to the tension between the time-series dynamics and yield curve properties, hence the presence of USV, is the misspecification of the MPR. Duffee (2002) proposes the essentially affine model for MPR as they are found to allow greater flexibility in fitting variations in the price of interest rate risk over time, while retaining the affine time-series and cross-sectional properties of bond prices. Cheridito et al. (2007) studies a no-arbitrage specification for MPR, with their improved fit coming from the time-series rather than cross-sectional features of the yield curve. A separate set of parameters, with one being a risk-neutral measure and physical measure for the other set, adds flexibility to their specification. The specification relieves the tension between matching the time-series behavior of the interest rate process and matching the cross-sectional shape of the yield curve.

Collin-Dufresne et al. (2004) describes the concepts of identification, identifiability, and maximality. These are the centers for local and global identifiability of models and parameters, whose presence determines the existence of economic interpretation. Identification deals with how and whether the state vector and parameter vector can be inferred from a particular data set. Identifiability deals with the issue of whether the state vector

and parameter vector can be inferred from observing all conceivable financial data, as frequently as necessary. Maximality is about the most general model within $A_M(N)$ as discussed by Dai and Singleton (2000), which is identifiable given sufficiently informative data. Affine models are written in terms of latent variables with no clear economic meaning independently from the model and parameters, leading to representations that are locally but not globally identifiable and often meaningless. Collin-Dufresne et al. (2008) proposes a representation in which the state vector is written in terms of theoretically observable state variables that have unambiguous economic interpretations. To circumvent these challenges, a process is required whereby latent state variables are rotated to observable state variables so that they have economic meaning. Collin-Dufresne et al. (2008) proposes a new representation of affine models in which the state vector comprises infinitesimal maturity yields and their quadratic covariations. In contrast to the invariant transformation rotations by Duffie and Kan (1996) and Dai and Singleton (2000), their representation accommodates the USV (Singleton, 2006).

Finally, the issue of parameter uncertainty and model flexibility compels us to consider a probabilistic method of estimation, the MCMC. We are also concerned about the global identifiability of our models and parameters, and therefore economic meaning. López-Pérez et al. (2025) performs a comparative study between MCMC, Particle filter, and the Kalman filter. Key challenges are the discretization bias and the presence of latent factors. While all procedures are computationally demanding, they found the Kalman filter a computational efficient estimation method to use in goodness-of-fit testing procedures. Incorporating their process to stochastic volatility and application to observed data suggests that the volatility depends on an additional factor that varies independently of the short rate level.

Aït-Sahalia et al. (2024) proposes an efficient and flexible method to compute the maximum likelihood estimators of continuous-time models when part of the state vector is latent, considering stochastic volatility and term structure models. In contrast to MCMC, their approach relies on closed-form approximations to estimate parameters and simultaneously infer the distribution of filters—of the latent states conditioning on observations. The computational complexity of these methods may present some challenges to our intended purpose. MCMC enables us to estimate the parameters under assumptions of prior parameter distributions and conditional likelihood functions to determine posterior distributions of parameter samples and latent variables with reasonable precision.

In state–space models, the use of MCMC methods introduces sampling noise due to the discreteness of the generated samples and the inherent gaps between each pair of measurements. Elerian et al. (2001) proposes the estimation framework, which relies on the introduction of latent auxiliary data to complete the missing diffusion between each pair of measurements. This data augmentation is synonymously referred to as smoothing. Stroud et al. (2003) proposes a simulation-based smoothing technique and the auxiliary mixture model. The auxiliary mixture model consists of state-dependent weights and efficient block sampling algorithms to jointly update all unobserved states given latent mixture indicators (Collin-Dufresne et al., 2009) and references therein. The same blocking approach is followed by Collin-Dufresne et al. (2009) with the parameters vector broken into three blocks— those affecting the dynamics under the risk-neutral measure, risk-premium parameters, and the measurement error standard deviations.

In estimating the USV models, bond prices alone may not be sufficient to identify all the parameters (Bikbov & Chernov, 2004; Collin-Dufresne & Goldstein, 2002). In contrast, Collin-Dufresne et al. (2009) adopts the "time-series only" approach and augments the observed yields with PCA factors. The estimated parameters are fitted to the restricted models that produce USV. Regression analysis is conducted between the short rate volatility produced by these models and the market at-the-money implied volatility to detect the

sensitivity of USV to options implied volatility. Our study follows their approach by estimating such parameters from "time series-only", augmenting with the PCA factors. Instead of regressing the short rate volatility against the options implied volatility, we consider the macro variables as a proxy for the implied volatility and test their sensitivity to the USV. This approach may also be considered to be an alternative practice, particularly in the emerging markets where market options data may be limited due to illiquidity and sparsity of trades.

3. Model Establishment

ATSMs are described in terms of an N-dimensional Markov process of a state variable X and its dynamics under the risk-neutral measure $\mathbb{Q}$ are written as:

$$dX_t = \mathcal{K}^{\mathbb{Q}}\left(\theta^{\mathbb{Q}} - X_t\right)dt + \Sigma\sqrt{S_t}dW_t^{\mathbb{Q}} \tag{1}$$

where $\theta^{\mathbb{Q}}$ is the long-term mean of the process X_t under the risk-neutral measure, W_t is an independent Brownian motion, κ and Σ are $N \times N$ matrices, and S_t is a diagonal matrix with the ith element given by

$$S_{ii,t} = \alpha_i + \beta_i^T \cdot X_t \tag{2}$$

spot rate is an affine function of X_t and is written as

$$r_t = \delta_0 + \delta_1 \dot{X}_t \tag{3}$$

where X_t is a Markov N-dimensional state vector at time t, δ_0 is a scalar and δ_i is an N-dimensional vector for $i = 1, ..., N$.

Provided that parameters are admissible, it is known from Duffie and Kan (1996) that a zero-coupon bond has a solution

$$P_t(\tau) = e^{A(\tau) - B(\tau)^T \cdot X_t} \tag{4}$$

where $\tau \equiv (T - t)$ and $A(\tau)$ and $B(\tau)$ satisfy the following ordinary differential equations (ODEs), also known as Ricatti equations

$$\frac{dA(\tau)}{d\tau} = -\theta^{\mathbb{Q}^T}\mathcal{K}^{\mathbb{Q}^T}B(\tau) + \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma^T B(\tau)\right]_i^2 \alpha_i - \delta_0 \tag{5}$$

$$\frac{dA(\tau)}{d\tau} = -\theta^T\mathcal{K}^{\mathbb{Q}^T}B(\tau) - \frac{1}{2}\sum_{i=1}^{N}\left[\Sigma^T B(\tau)\right]_i^2 \beta_i + \delta_y \tag{6}$$

A solution to these ODEs is found through numerical integration, starting from the initial conditions $A(0)$=$B(0)_{N\times 1}$.

By inverting (4) a related yield y_t is computed as

$$y_t(\tau) = -\frac{logP_t(\tau)}{\tau} = \frac{A(\tau)}{\tau} + \frac{B(\tau)'X_t}{\tau} \tag{7}$$

We adopt a framework of Collin-Dufresne et al. (2009) in which they use a classification scheme of Dai and Singleton (2000) for $A_1(3)$ and $A_1(4)$ models, each with $M = 1$ conditional volatility vector. They choose a representation where, for a state vector X, each state variable has a clear economic interpretation. They derive the state variables below using the Taylor series expansions with respect to maturity τ

$$Y_t(\tau) = Y_t^0(\tau) + \tau Y_t^1(\tau) + \frac{1}{2}\tau^2 Y_t^2 + \cdots \tag{8}$$

They assign the state vector $X = [r, \mu^Q, V]$ for the three-factor model. An additional state variable θ^Q, being three times the curvature, is introduced in the four-factor model so that the state vector becomes $X = [r, \mu^Q, \theta^Q, V]$. These state variables are defined as follows:

$$r_t = Y_t(0) \tag{9}$$

$$\mu_t^Q = 2\frac{\partial Y_t(\tau)}{\partial \tau}\bigg|_{\tau=0} \tag{10}$$

$$\theta_t^Q = 3\frac{\partial^2 Y_t(\tau)}{\partial \tau^2}\bigg|_{\tau=0} \tag{11}$$

$$V_t = \frac{1}{dt} dr_t^2 \tag{12}$$

Following Collin-Dufresne et al. (2008), alternative ways of representing (1) were evaluated to identify the most general and identifiable form than the $\Sigma\sqrt{S_t}$ for the diffusion matrix. Here, the stochastic components of the model are expressed in terms of a covariance matrix rather than as Iˆo diffusions and are found to be more suitable for introducing the parameters that have clear economic interpretations.

3.1. The $A_1(3)$ Model

Model $A_1(3)$ is a three-factor ATSM that allows the presence of stochastic volatility of interest rates. Its canonical form is expressed by a state vector $S = [x, r, \mu_1]$ can be rotated into three state variables short rate, risk-neutral drift of the short rate, and the variance of the short rate, resulting into an observable state vector $X = [r, \mu^Q, V]$. This is regardless of whether there is or is not USV; see Singleton (2006). The risk-neutral dynamics for the state vector X can be written as
Instantaneous mean:

$$\frac{1}{dt}\mathbb{E}^{\mathbb{Q}}[dX_t] = \begin{bmatrix} m_0 + m_r r_t + m_\mu \mu_t^{\mathbb{Q}} + m_V V_t \\ \mu_t^{\mathbb{Q}} \\ \gamma_V - \mathcal{K}_V V_t \end{bmatrix} \tag{13}$$

where:

- m_0 = is a constant term
- $r_t = \delta_0 + \delta_1^\top X_t$ = feedback from the short rate
- $\mu_t^{\mathbb{Q}}$ = slope of the yield curve
- m_V = volatility feedback
- γ_V = the long-term mean or level of the volatility process.
- $\mathcal{K}_V$ = the mean-reversion speed of the volatility process.

and Covariance matrix:

$$\frac{1}{dt} cov\left(dX_t, dX_t^T\right) \equiv \Omega_t = \Omega_0 + \Omega_V(V_t - \underline{V}) \tag{14}$$

where $\underline{V}$ is set to be a lower bound for V_t. Admissibility requirements are also met provided that Ω_0 and Ω_V are positive semidefinite and positive definite, respectively.

$$\Omega_0 = \begin{bmatrix} \underline{V} & c_{r\mu} & 0 \\ c_{r\mu} & \sigma_\mu & 0 \\ 0 & 0 & 0 \end{bmatrix} and \quad \Omega_V = \begin{bmatrix} 1 & c_{r\mu} & c_{rV} \\ c_{r\mu} & \sigma_\mu & c_{\mu V} \\ c_{rV} & c_{\mu V} & \sigma_V \end{bmatrix} \tag{15}$$

There are six parameters in the drift and eight in the covariance matrix, resulting in 14 risk-neutral parameters in total. This implies that the model is a Q-maximal model

because of the greater number of parameters, according to Dai and Singleton (2000) and Collin-Dufresne et al. (2008).

3.2. Market Price of Risk

The pricing kernel, which is also known as the stochastic discount factor (SDF), is defined as

$$\frac{d\Lambda_t}{\Lambda_t} = -r_t dt - \lambda_t^\top dW_t \tag{16}$$

Theessentially affine form of Duffee (2002) specifies the market price of risk (MPR) Λ_t as a nonlinear function of volatility

$$\Lambda_t = -\sqrt{S_t}\lambda_1 + \sqrt{S_t^{-1}}\lambda_2 \tag{17}$$

where S_t is a non-negative function of the volatility state V_t. This structure introduces both direct and inverse sensitivity to volatility in the risk premia. It captures effects such as increased compensation for high volatility environments, and asymmetric or leverage effects through $\sqrt{S_t^{-1}}$.

3.3. Model $A_1(3)$ Under Physical Measure

No-arbitrage conditions require that both the risk-neutral measure $\mathbb{Q}$ and the physical or historical measure $\mathbb{P}$ be equivalent. It is implied that the volatility or diffusion part of the SDE does not change in either measure and that only the drift will be influenced by the risk premia. The drift for $A_1(3)$ state vector under physical measure $\mathbb{P}$ is specified as

$$\frac{1}{dt}E^P[dX_t] = \begin{bmatrix} \lambda_{r0} + \lambda_{rr} r_t + (1 + \lambda_{r\mu}\mu_t^Q + \lambda_{rV} V_t) \\ (m_0 + \lambda_{\mu 0}) + (m_r + \lambda_{\mu r}) r_t + (m_\mu + \lambda_{\mu\mu})\mu_t^Q + (m_V + \lambda_{\mu V}) V_t \\ (\gamma_V + \lambda_{V0}) - (\mathcal{K}_V - \lambda_{VV}) V_t \end{bmatrix} \tag{18}$$

See Appendix A for a detailed derivation:

3.4. The $A_1(4)$ Model

In (11) we defined the state variable θ_t^Q, which represents the curvature of the yield at short maturities. Adding this to the state vector results in $X = [r, \mu_t^Q, \theta_t^Q, V_t]$. The result is a four-factor model with risk-neutral dynamics represented by the instantaneous mean and covariance matrix as

$$\frac{1}{dt}E^Q[dX_t] = \begin{bmatrix} \mu_t^Q \\ \theta_t^Q + V_t \\ a_0 + a_r r_t + a_\mu \mu_t^Q + a_V V_t \\ \gamma_V - \mathcal{K}_V V_t \end{bmatrix} \tag{19}$$

and

$$\frac{1}{dt} cov(dX_t, dX_t^T) \equiv \Omega_t = \Omega_0 + \Omega_V (V_t - \underline{V}) \tag{20}$$

where

$$\Omega_0 = \begin{bmatrix} \underline{V} & c_{r\mu} & c_{r\theta} & 0 \\ c_{r\mu} & \sigma_\mu & c_{\mu\theta} & 0 \\ c_{r\mu} & c_{\mu\theta} & \sigma_\theta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ and } \Omega_V = \begin{bmatrix} 1 & c_{r\mu} & c_{r\theta} & c_{rV} \\ c_{r\mu} & \sigma_\mu & c_{\mu\theta} & c_{uV} \\ c_{r\theta} & c_{\mu\theta} & \sigma_\theta & c_{\theta V} \\ c_{rV} & c_{\mu V} & c_{\theta V} & \sigma_V \end{bmatrix} \tag{21}$$

Similar to $A_1(3)$ model, admissibility requirements are met provided that Ω_0 and Ω_V are positive semidefinite and positive definite, respectively. We note that the model has a total of 22 free risk-neutral parameters, qualifying the model to be a maximal $A_1(4)$.

3.5. Parameter Restrictions

Collin-Dufresne et al. (2004) describethe necessary and sufficient conditions for parameter restrictions under which an $A_1(3)$ model can display USV. A model exhibits USV if the state variables driving volatility risk cannot be hedged away by trading the bond prices alone. These conditions must also lead to the admissibility of the model (Collin-Dufresne et al., 2004).

3.5.1. $A_1(3)$

Implementing the following restricted parameters characterize the $A_1(3)$ as a model with USV (Collin-Dufresne et al., 2004)

$m_r = -2c_V^2$

$m_\mu = 3c_V$

$m_V = 1$

$\sigma_V^2 = c_V^2$

3.5.2. $A_1(4)$

Collin-Dufresne et al. (2009) proposes an extension to the $A_1(3)$ restrictions and derive the following for the $A_1(4)$ model to display USV $a_r = -2c_{r\mu}^2(3c_{r\mu} - a_\theta)$

$a_\mu = 7c_{r\mu}^2 - 3c_{r\mu}a_\theta$

$a_V = 3c_{r\mu}$

$\sigma_\mu = c_{r\mu}^2$

$\sigma_\theta = c_{r\mu}^4$

$c_{r\theta} = c_{r\mu}^2$

$c_{\mu\theta} = c_{r\mu}^3$

3.6. Estimation Strategy

Bayesian estimation procedures for continuous-time finance are well documented in many sources, including Stroud et al. (2003), Collin-Dufresne et al. (2009) and references therein, and Johannes and Polson (2010). We aim to infer the posterior distribution of unknown parameters given observed data. This process can be computationally challenging, especially when the posterior has a complex or high-dimensional structure. Two powerful tools that can simplify these challenges are data augmentation and a Gibbs-like posterior sampler. Data augmentation is a technique by which a latent auxiliary variable is introduced to fill the missing diffusion between each observed pair of measurements; see Elerian et al. (2001). The Gibbs sampler is the simplest algorithm that is possible to use directly to sample iteratively from all complete conditionals (Θ, X), where Θ is a parameter vector and X a latent variable (Johannes & Polson, 2010).

Collin-Dufresne et al. (2009) considers two data augmentation approaches. First, augmenting with unobservable high-frequency data, which enables the use of the Euler approximation, and provides a Gaussian density that is easy to work with. This introduces discretization bias, especially when time steps h are not sufficiently small, and the true continuous-time model deviates from the linear or Gaussian assumptions underlying Euler methods. Second, augmenting the observed yield data with the theoretically observable term structure factors—state variables X. Its advantage over the former is that it treats the main sources of variation in the yield curve as observable proxies for the latent state process. This approach assumes that the latent factors driving the term structure are extracted directly from cross-sectional yield data. It results in a much simpler posterior structure and avoids discretization altogether, enabling more stable and efficient inference.

We let $\mathcal{P} = \mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_T$ represent a time series of principal components of the yields Y_t. The posterior $p(\Theta|\mathcal{P})$ is approximated using PCA-augmented data as

$$p(\Theta|\mathcal{P}) \propto p(\mathcal{P}|\Theta)p(\Theta)$$

where Θ represents a vector of parameters.

The first term on the right-hand side is the likelihood function, and the second is the prior distribution of parameters. The likelihood function is intractable, making it difficult to evaluate the posterior. This would require the use of several techniques from the MCMC. The observable data $\mathcal{P}$ are augmented with the term structure factor data $X = \{X_1, X_2, ..., X_T\}$. These factors are theoretically observable as they can be interpreted independently from the model being considered. In practice, they may not be directly observed as there are only a finite maturity yields available. Uncertainty in these state variables is therefore integrated out using a Gibb-like posterior that alternates between drawing $p(\Theta|\mathcal{P}, X)$ and $p(X|\mathcal{P}, \Theta)$ (Collin-Dufresne et al., 2009).

True dynamics from SDE (1) are then approximated as[1]

$$(X_{t+h} \sim X_t) \sim \mathcal{N}(h(a + bX_t), h\Omega_t) \tag{22}$$

where:

- $h = dt$ is the discretization step (e.g., $\frac{1}{52}$ for weekly data),
- $a + bX_t$ is the drift under $\mathbb{P}$,
- Ω_t is the state-dependent instantaneous covariance matrix which depends on V_t.

The likelihood function is then specified in terms of the relationship between data and the state vector as

$$\mathcal{P}_t = PC\ loadings \times Y_t \tag{23}$$

It is clear from (7) that there is a linear relation between the principal components and state variables. Adding a Gaussian error vector $\epsilon_t \sim N(0, \Omega_T)$, we obtain

$$\mathcal{P}_t = K + LX_t + \epsilon_t \tag{24}$$

(24) is further defined as a "measurement equation" in a state–space model, which we shall use in Kalman filtering to estimate the latent factor V_t.

When full conditionals (Θ, X) are not tractable for a Gibbs-sample, Metropolis–Hastings (MH) steps within the Gibbs sampler become an alternative technique. The MH algorithm generates a Markov chain with a target posterior density $\pi(X)$. At each iteration t:

- It proposes $X^* \sim p(X^*|X^(t))$
- Accept X^* with the following probability:

$$\alpha = min\left(1, \frac{\pi(X^*)p(X^{(t)})|X^*)}{\pi(X^{(t)})p(X^*)p(X^*|X^{(t)})}\right) \tag{25}$$

where:
X^* represents a proposed state or candidate in the MH algorithm
$p(X^*|X^{(t)})$ is the proposal distribution that defines how the candidate state X^* is chosen given the current-state $X^{(t)}$.
α is the acceptance probability in the MH algorithm. It determines whether the proposed state X^* is accepted as the new state $X^{(t)}$. If X^* is accepted, the chain moves to X^*; otherwise, it stays at $X^{(t)}$.

There are several challenges and limitations for the MCMC, such as computational burden, weak parameter identification, and poor mixing. The latter is caused by the

presence of high correlation between the parameters and latent states, resulting in slow convergence. Johannes and Polson (2010) recommends the use of artificial data to check the efficiency and convergence of the algorithm. Stroud et al. (2003) conducted a comparative analysis on three alternative MCMC algorithms in terms of their autocorrelation functions. The option with fast decay in autocorrelation was found to have a faster convergence to the other alternatives.

Appendix B discusses briefly the Bayesian algorithm and workflow, as adapted from the Collin-Dufresne et al. (2009) approach. The algorithm, posterior structure, and convergence diagnostics are fully documented in Appendix B. They are augmented by a workflow diagram and pseudo-code in Appendices C and D. A three-block approach is followed, with parameters broken into ϕ^Q, ϕ^Λ, and ϕ^λ for risk-neutral measure, measurement error standard deviations, and risk-premium parameters, respectively.

4. Data Collection

We use a sample of weekly SA government treasury bond prices spanning the period October 2013 to September 2024 with maturities of 3 months, 5, 10, 12, 20, 25, and 30 years. For out-sample analysis, we use data for the periods October 2023–September 2024 with the same maturities. The data were retrieved from the Thomson Reuters database. We extract the approximated zero-coupon yields by inverting the bond pricing equation. Specifically, for a bond with maturity τ and price P, we compute the implied yield using the simplified inversion Formula (7).

Traditionally, zero-coupon yields are typically bootstrapped from swap rates, LIBOR, or any of its replacements. Swap-based bootstrapping depends on credit assumptions, interpolation, and curve construction techniques that may not reflect the true market dynamics. These instruments are derived, not directly traded, and do not represent actual cash instruments. In contrast, coupon bond prices that are observed in the market contain rich maturity and liquidity information, and are directly priced by supply–demand dynamics. We therefore use coupon bond data as the basis for our term structure modeling.

To ensure internal consistency and eliminate arbitrage opportunities, we employ a no-arbitrage ATSM where zero-coupon bond prices satisfy

$$P(t,T) = E_t\left[exp(-\int_t^T r_s ds)\right] \tag{26}$$

Assuming an affine structure for the short rate and state variables X_t, the price takes the form of (4). Here, X_t is a vector of latent state variables, and $A(\tau)$ and $B(\tau)$ are functions satisfying Ricatti differential equations, which ensure closed-form and arbitrage-free bond pricing.

We simulate the evolution of interest rates using the SDE (1). At time t, given each simulation path, bond prices are computed, followed by zero yields. These zero yields produce a time series of term structures $y(t,\tau)$, each curve reflecting the simulated latent state at time t for a maturity τ.

To extract the structure from the evolving yield curves, we apply the PCA. The zero-yield data are then decomposed into orthogonal components to capture the majority of variance with a few statistical factors: level, slope, and curvature. Table 1 presents the PCA for our zero-yield data. It is noted that the first three components have a cumulative variance of 99.97%, which agrees with the empirical evidence of Litterman et al. (1991); see Figure 1 below.

Table 1. Principal components of the zero-coupon yields for the SA government treasury bond over maturities 3 months, 5, 10, 12, 20, 25, and 30 years. The second row from the bottom provides details about explained variances for individual components, which, by observation, are declining from 66.06% for the first component, followed by 30.14% for the second one, thereafter dropping to almost zero. The last row represents the cumulative percentages, confirming that the first three components contribute more to variation than the remaining ones.

	Principal Components						
	1	**2**	**3**	**4**	**5**	**6**	**7**
3-month	−0.10	0.86	−0.33	0.13	−0.13	−0.33	−0.06
5-year	0.02	0.36	−0.08	−0.50	0.44	0.58	0.28
10-year	0.17	−0.03	0.04	−0.37	−0.51	−0.25	0.71
12-year	0.30	−0.11	−0.28	0.52	0.52	−0.18	0.51
20-year	0.60	0.00	−0.37	0.21	−0.43	0.50	−0.15
25-year	0.45	0.34	0.80	0.20	0.05	0.02	0.01
30-year	0.56	−0.08	−0.16	−0.50	0.27	−0.46	−0.36
Explained Variance (%)	66.06	30.14	3.77	0.03	0	0	0
Cumulative Variance (%)	66.06	96.2	99.97	100	100	100	100

We also compute the FX volatility and oil shock using Brent crude and USDZAR prices, respectively, for the period October 2013 to September 2024.

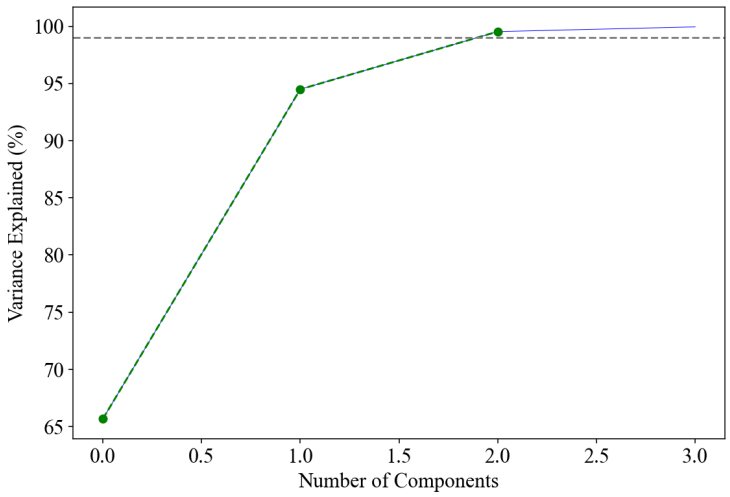


Figure 1. Cumulative explained variance in percentages is plotted against the number of principal components. The green dotted line with dots and blue line represent the cumulative variance for models $A_1(3)$ and $A_1(4)$ fitted yields, respectively. A gray dotted line is the maximum level of 99.97% beyond which the remaining 0.03% represented by a blue line is negligible.

5. Scenario Determination

We apply the ATSM to extract the interest rate volatility from a cross-section of bond prices. ATSM represents both the yield curve term structure and the time series of bond prices over several trading periods. This multi-period and cross-sectional nature of the interest rates presents several challenges, including the volatility and trading risk, risk premia, and market completeness. At the heart of term structure modeling is the need to study both the term structure and time series simultaneously. ATSMs have been successfully used in this area as they are characterized by tractability, closed-form solutions, efficient approximation, and closed-form moment conditions for empirical analysis (Collin-Dufresne et al., 2004) and references therein. Our study is based on the following choice of scenarios, some of which are already supported by the empirical evidence:

- The test whether state vectors and parameter vectors have a unique economic interpretation. This requires a suitable representation among ATSMs where latent state vectors are translated into observable factors. Both the state vector and model parameter vector should be globally identifiable so that their values can be compared directly across different countries, periods, and even models.
- Among the three and four-factor stochastic volatility models, evaluate their capability to break the dual role of predicting the variance of the short rate and simultaneously a linear combination of yields and the quadratic variation of the spot rate. There is empirical evidence that the $A_1(3)$ model is unable to play the dual role (Collin-Dufresne et al., 2004). We compare the $A_1(4)$ USV with $A_1(3)$USV.
- To determine whether estimation can be based on time series only. In the absence of option price data, we test the macro variables as sources of variation and a substitute for option data when estimating USV models.

Among several estimation approaches, we proceed with the Bayesian approach, which combines the MCMC and Kalman filter. This probabilistic approach is more suitable where there are parameter uncertainty and model flexibility issues.

6. Model Implementation

We use (4) to invert zero yields from the SA government treasury bond. The zero yields are further transformed into a nonlinear state–space form. The purpose is to implement the three-factor and four-factor affine models to extract the short rate variance. The final outcome is to determine whether the affine models can explain simultaneously the time series and cross-sectional properties of bond prices. It should be noted that the ATSMs that we selected are based on the maximal forms of both $A_(3)$ and $A_1(4)$.

Collin-Dufresne et al. (2008) propose an alternative approach to those of Duffie and Kan (1996) (DK) and Dai and Singleton (2000) to rotate from latent to observable state vectors. Whereas DK's technique involves inverting the term structure with respect to latent factors and cannot be implemented, Collin-Dufresne et al. (2008) writes these term structure dynamics in a nonlinear state–space. The model parameters are then estimated by Bayesian and MCMC methods.

It is noted that the volatility state variable V_t does not enter the bond pricing equation for those models exhibiting USV. A four-factor model with a state vector $X = [r, \mu_t, \theta_t, V_t]$ effectively becomes a three-factor model by excluding the state variable V_t. V_t is, therefore, an additional volatility factor that is free to explain the time-series patterns (Collin-Dufresne et al., 2004). USV can also be regarded as a latent component V_t that influences the risk premia and yield variance but is not directly priced by the cross-section of bond yields.

Despite the exclusion of V_t from the bond price equation, it continues to influence the following, thus fulfilling the conditions for USV:

- The conditional variance of state transitions $\Sigma(X_{t+1}|X_t)$
- The market price of risk via $\Lambda_t = -\sqrt{S_t}\lambda_1 + \sqrt{S_t^{-1}}\lambda_2$

Since V_t is not spanned by the cross-section of bond yields, it must be identified through time-series variation and external instruments.

In our empirical implementation, we filter V_t using the observed term structure, which, by way of a robustness check, we compare against the variation of macro signals such as oil shocks, and GARCH-based FX volatility. This, we believe, also assists in ascertaining the identifiability and economic relevance of the USV factor in markets with limited market options data due to sparse trades.

We consider a state–space model:

$$y_t = K + LX_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon) \tag{27}$$

$$X_{t+1} = \mu + \Phi X_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q(V_t)) \tag{28}$$

X_t is an unknown state vector, and y_t is the observed data vector. Both equations represent the Kalman filter. We apply the Kalman filter within the MCMC procedure to extract the posterior distribution of V_t.

7. Analysis of Results

This section presents the results from the two competing models $A_1(3)$ and $A_1(4)$, both with USV, focusing on both in-sample and out-of-sample performance. We begin with the analysis of the posterior distribution for parameters to assess the point estimates and associated intervals for estimation consistency. This is followed by an assessment of the models' ability to fit the yield curve in-sample, and thereafter, an evaluation of their forecasting performance out-of-sample. Subsequently, we provide a closer examination of the time-series behavior of the key latent factor, volatility, term structure implications, various sensitivity and robustness diagnostics.

7.1. Posterior Distributions of Key Parameters

Tables 2 and 3 report the parameter point estimates for risk-neutral and risk premia, respectively. We applied the discretization value $h = \frac{1}{52}$, which produced reasonable estimates. For each point estimate, based on mean values, there are corresponding credible interval bounds, which appear consistent across the two models. We observe that $A_1(4)$USV exhibits substantially narrower posterior distributions across most parameters compared to $A_1(3)$USV. This suggests that $A_1(4)$USV constrains the parameter estimates better, potentially due to improved model specification. In contrast, the broader intervals in $A_1(3)$USV may indicate underfitting and that its structure might not capture the complexity in the data sufficiently. The tighter credible intervals in $A_1(4)$USV should be interpreted as increased certainty.

There may be a concern over a possible underfitting observed in $A_1(4)$USV, suggesting that a higher-dimensional ATSM, such as a five-factor model, might improve the fit. It should be noted, as stated by (Collin-Dufresne et al., 2004), that additional factors do not necessarily translate into materially different yield curve dynamics. The four-factor USV model effectively behaves like a three-factor model in terms of the yield curve, since the USV factor does not impact bond prices directly. Therefore, increasing the number of factors may not always lead to meaningful gains in model performance, though this remains an avenue for future research.

In addition, for a sample of risk-neutral drift parameters, which includes the physical γ_V^P and $\mathcal{K}_V^P$, we analyzed the posterior plots for consistency.[2] Figures 2 and 3 plot the histograms and kernel densities for drift parameters for models $A_1(3)$ USV and $A_1(4)$ USV, respectively. As expected, the shapes of the histograms do not exhibit a multi-model but instead, are all unimodal. This is a sign of convergence and good mixing of the MCMC algorithm.

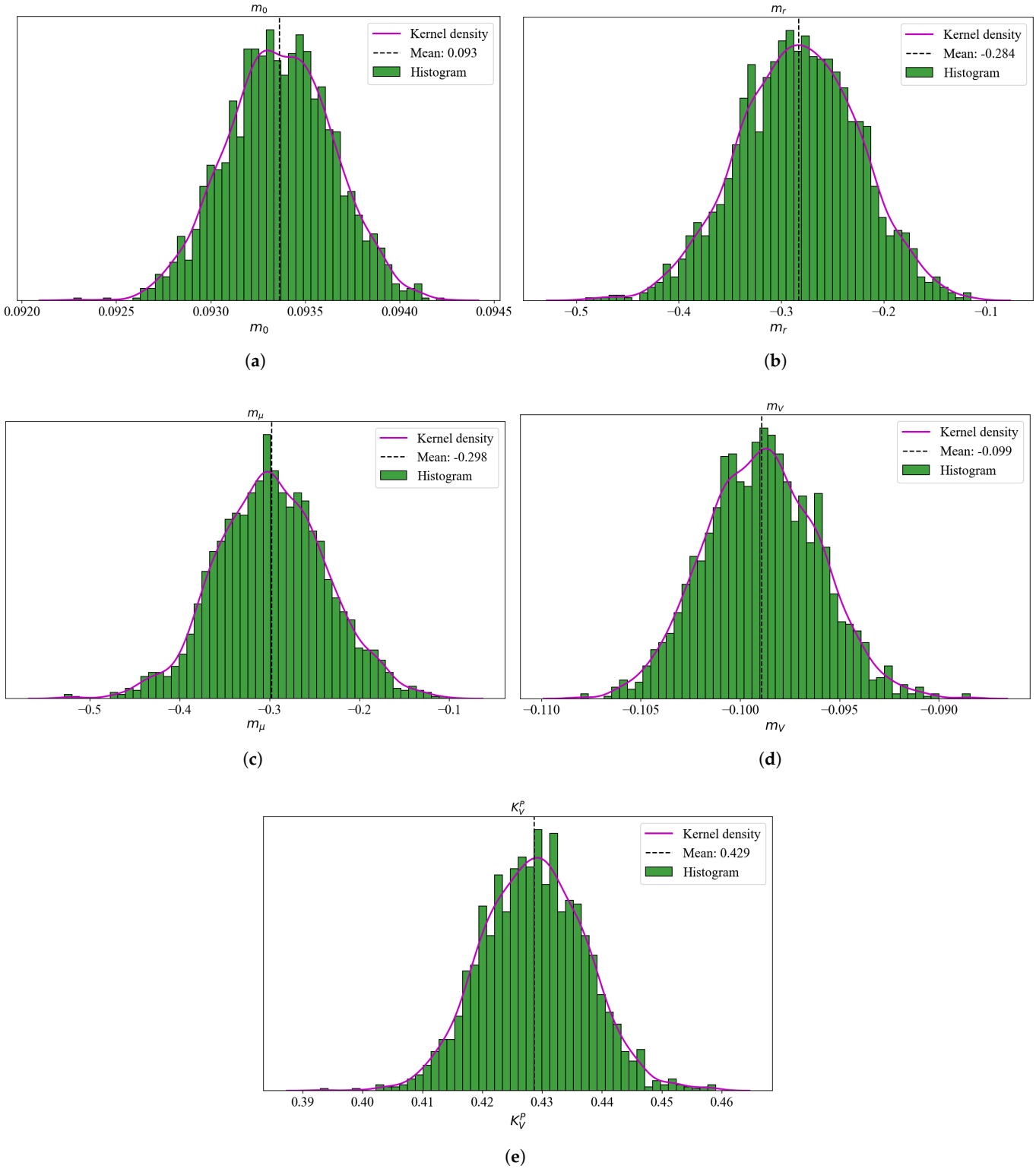


Figure 2. A sample of posterior parameter histograms for the $A_1(3)$ USV model. Each subplot depicts a distribution for each parameter in this order **(a)** m_0 **(b)** m_r **(c)** m_μ **(d)** m_V **(e)** $\mathcal{K}_V^P$. These are the risk-neutral drift parameters except for the γ_V^P and $\mathcal{K}_V^P$. The magenta line represents a theoretical Gaussian density plot. The histograms exhibit a unimodal shape in most cases, which confirms a successful convergence and mixing.

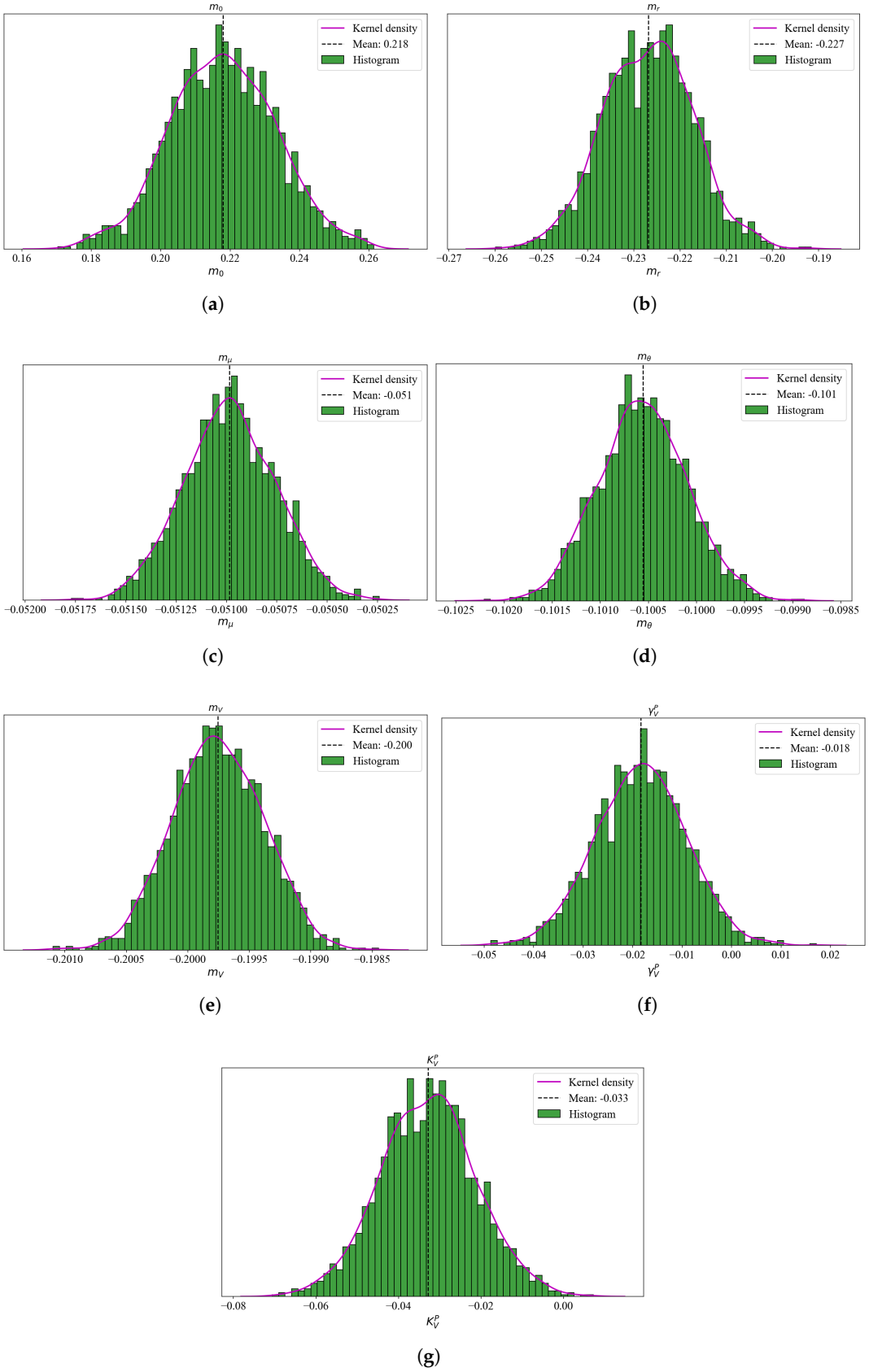


Figure 3. A sample of posterior parameter histograms for the $A_1(4)$ USV model are plotted. Each subplot depicts a distribution for each parameter in this order (**a**) m_0 (**b**) m_r (**c**) m_μ (**d**) m_θ (**e**) m_V (**f**) γ_V^P (**g**) $\mathcal{K}_V^P$. Specifically, these are the risk-neutral drift parameters except for the γ_V^P and $\mathcal{K}_V^P$. For each histogram, a plot in magenta represents a theoretical Gaussian density plot. The histograms exhibit a unimodal shape in most cases, which confirms a successful convergence and mixing.

Table 2. Posterior distribution for key risk-neutral drift and covariance parameters for model $A_1(3)$ compared to $A_1(4)$. The parameters are fitted to the weekly approximated zero yields for the period October 2013 to September 2024. Both models are distinguished from the unrestricted maximal model by restricted parameters. These restricted parameters are shown as underlined and are a function of normal parameters. The table consists of point estimates, based on mean values of the posterior parameter distribution, and confidence interval bounds (in square brackets). Confidence interval bounds for the point estimates are computed as 2.5 and 97.5 percentiles of the posterior distribution.

Parameter	$A_1(3)$ USV	$A_1(4)$ USV
m_0	0.0012 [0.0009, 0.0015]	0.0010 [0.0010, 0.0010]
m_r	$\underline{-0.3588}$ [−0.4171, −0.3005]	$\underline{-0.0780}$ [−0.1147, −0.0414]
m_μ	$\underline{-0.8077}$ [−1.1711, −0.4443]	$\underline{-0.0576}$ [−0.0676, −0.0476]
m_θ		−0.1113 [−0.1242, −0.0984]
m_V	$\underline{-0.3152}$ [−1.0949, 0.4645]	$\underline{-0.0196}$ [−0.0203, −0.0189]
γ_V^P	0.0009 [[0.0009, 0.0009]	0.0020 [0.0019, 0.0021]
$\mathcal{K}_V^P$	0.4457 [0.4414, 0.4500]	1.0450 [1.0129, 1.0770]
σ_r	0.0071 [−0.0048, 0.0189]	0.0258 [−0.0230, 0.0746]
σ_μ	0.0816 [−0.1823, 0.3454]	$\underline{0.0190}$ [0.0016, 0.0365]
σ_θ		$\underline{0.0101}$ [0.0098, 0.0104]
σ_V	0.0003 [−0.0007, 0.0013]	0.0001 [−0.0000, 0.0002]
$\underline{V}$	1.0×10^{-6} $[1.0 \times 10^{-6}, 1.0 \times 10^{-6}]$	0.0001 [0.0001, 0.0001]
$c_{r\mu}$	−0.4622 [−0.5098, −0.4146]	$\underline{-0.3086}$ [−0.3226, −0.2945]
$c_{r\theta}$		$\underline{-0.0503}$ [−0.0524, −0.0482]
$c_{\mu\theta}$		$\underline{-0.0507}$ [−0.0523, −0.0491]
c_{rV}	−0.0507 [−0.0533, −0.0481]	−0.0973 [−0.1017, −0.0929]
$c_{\mu V}$	−0.2237 [−0.2456, −0.2018]	$\underline{-0.0951}$ [−0.0989, −0.0914]
$c_{\theta V}$		$\underline{0.1043}$ [0.0992, 0.1094]

Table 3. Posterior distribution for key risk-premium parameters for model $A_1(3)$ compared to $A_1(4)$. The parameters are fitted to the weekly approximated zero yields for the period October 2013 to September 2024. The table consists of point estimates, based on mean values of the posterior parameter distribution, and confidence interval bounds (in square brackets). Confidence interval bounds for the point estimates are computed as 2.5 and 97.5 percentiles of the posterior distribution.

Parameter	$A_1(3)$ USV	$A_1(4)$ USV
λ_{r0}	−0.0122 [−0.0144, −0.0100]	−0.0102 [−0.0104, −0.0099]
λ_{rr}	−0.0506 [−0.0543, −0.0469]	−0.0478 [−0.0490, −0.0465]
$\lambda_{r\mu}$	−0.0475 [−0.0524, −0.0427]	−0.0495 [−0.0510, −0.0480]
$\lambda_{\mu 0}$	−0.0098 [−0.0107, −0.0089]	−0.0102 [−0.0105, −0.0099]
$\lambda_{\mu r}$	0.0947 [0.0872, 0.1021]	0.0495 [0.0478, 0.0512]
$\lambda_{\mu\mu}$	0.2783 [0.2518, 0.3047]	0.1025 [0.0961, 0.1089]
$\lambda_{\mu V}$	47.4552 [42.8097, 52.1008]	10.2594 [9.7983, 10.7205]
λ_{V0}	−0.0000 [−0.0000, −0.0000]	−0.0001 [−0.0001, −0.0001]
λ_{VV}	0.0579 [0.0499, 0.0659]	0.0941 [0.0890, 0.0992]
λ_{rV}	−10.5910 [−11.5637, −9.6183]	−1.9671 [−2.0473, −1.8868]
$\lambda_{r\theta}$	0.0516 [0.0485, 0.0548]	0.0201 [0.0196, 0.0206]
$\lambda_{\mu\theta}$		0.0201 [0.0197, 0.0205]
$\lambda_{\theta\theta}$		0.0193 [0.0183, 0.0203]
$\lambda_{\theta r}$		0.0208 [0.0201, 0.0215]
$\lambda_{\theta\mu}$		0.0211 [0.0200, 0.0221]
$\lambda_{\theta V}$		0.0193 [0.0181, 0.0204]

To gain more insight into the effectiveness of the Bayesian inference obtained through MCMC sampling, we used the ARVIZ version 0.22 python software function arviz.summary() of (Kumar et al., 2019) to generate the Tables 4 and 5. These tables present posterior summaries for key model parameters obtained via MCMC sampling for models $A_1(3)$USV and $A_1(4)$USV, respectively. For each parameter, we report the posterior mean, standard deviation, and 95% Highest Density Interval (HDI), along with diagnostic statistics including Monte Carlo standard errors (MCSE), effective sample sizes (ESS), and the Gelman–Rubin convergence statistic ($\hat{R}$). All $\hat{R}$ values are equal to 1, indicating satisfactory convergence. Effective sample sizes (ESS) were used to assess sampling efficiency and convergence. In Table 4, bulk ESS values ranged from 975 to 2550 across parameters, with all but one parameter exceeding 2000. The lowest bulk ESS (975) is still well above common diagnostic thresholds ≥ 400, which is indicative of adequate sampling.

Tail ESS values ranged from 674 to 1500, demonstrating reliable estimation of posterior distribution tails and credible intervals. Similar trend for bulk ESS and tail ESS is also observed in Table 5. These values suggest that the Markov chains mixed well and that the

posterior estimates are statistically stable. The effective sample sizes are sufficiently large, suggesting reliable estimation. Notably, all point estimates lie well within their respective HDI bounds, consistent with stable posterior inference.

Table 4. The results of estimation for a sample of drift parameters for the model $A_1(3)$ are presented below. The purpose of the report is to determine the reliability of the posterior parameters.

	Mean	Std	HDI (2.5%)	HDI (97.5%)	MCSE (Mean)	MCSE (Std)	ESS (Bulk)	ESS (Tail)	$\hat{R}$
m_0	0.001	0	0.001	0.001	0	0	2326	1346	1
m_r	−0.284	0.059	−0.398	−0.171	0.001	0.001	2160	1106	1
m_μ	−0.299	0.063	−0.421	−0.177	0.001	0.001	2555	1429	1
m_V	−0.099	0.003	−0.105	−0.093	0	0	2312	1495	1
K_V^P	0.428	0.008	0.411	0.445	0	0	2113	1149	1
γ_V^P	0.007	0.005	0	0.016	0	0	975	674	1

Table 5. The results of estimation for a sample of drift parameters for the model $A_1(4)$ are presented below. The purpose of the report is to determine the reliability of the posterior parameters.

	Mean	Std	HDI (2.5%)	HDI (97.5%)	MCSE (Mean)	MCSE (Std)	ESS (Bulk)	ESS (Tail)	$\hat{R}$
m_0	0.093	0	0.093	0.094	0	0	3775	1808	1
m_r	−0.227	0.01	−0.247	−0.209	0	0	2626	1503	1
m_μ	−0.051	0	−0.051	−0.051	0	0	2549	1383	1
m_θ	−0.101	0	−0.101	−0.100	0	0	2241	1407	1
m_V	−0.200	0	−0.201	−0.199	0	0	3182	1548	1
γ_V^P	−0.018	0.009	−0.036	−0.001	0	0	2589	1617	1
K_V^P	−0.032	0.011	−0.052	−0.010	0	0	2162	1650	1

7.2. Yield Curve Fit

In this section, we evaluate the in-sample and out-sample model performance by way of the root-mean-square error (RMSE) and bias for each maturity. Model-fitted yields are derived from the point estimates in Tables 2 and 3. Similar to Collin-Dufresne et al. (2009), we avoid the process of integrating over the posterior distribution for computational burden. Posterior samples are therefore rerun only for sampled state variables and parameter values held fixed.

Using the short rate observed during the short term while assuming the constant yield to maturity, we compute the discount rate for each maturity, which we then apply to convert the observed yields into par bonds. From the par bonds, we compute the zero yields using (7). This exercise is necessary so that we may compare both the model-fitted yields and the observed ones with zero yields.

Table 6 presents the yield curve fit analysis results in terms of RMSE and bias, both in and out-sample[3]. The yield curve fitting performance of the two models is assessed using both RMSE and forecast bias, evaluated in-sample and out-of-sample across maturities of 0.25, 5, 10, 12, 20, 25, and 30 years. We applied the DM test to compare forecast errors $\epsilon_t = Y_t - \hat{Y}_t$ and detect statistically significant differences, where Y_t and $\hat{Y}_t$ represent the observed and forecast yields, respectively.

In-sample, model $A_1(4)$ USV provides a significantly better fit to the yield curve in terms of RMSE, with a DM test statistic of 6.1233 and p-value of 0.0009. This suggests that $A_1(4)$ USV captures the cross-sectional variation in yields more accurately. Model $A_1(3)$ USV exhibits a significantly lower forecast bias, particularly at longer maturities, with a DM statistic of −8.00 and p-value of 0.0002, suggesting that the model aligns more closely with observed yields on average, despite being less precise.

Out-sample, the results are more nuanced. Model $A_1(4)$ USV exhibits lower RMSE at shorter maturities, with a statistically significant difference observed at the first maturity. Across the full maturity spectrum, the difference in RMSE between the models is not statistically significant with a DM statistic of 0.9999 and p-value of 0.3559. A similar pattern is seen in forecast bias with model $A_1(3)$ USV consistently showing lower bias, but the overall difference is again not significant with a DM statistic of 0.9571 and p-value of 0.3755.

Finally, the results suggest that $A_1(4)$ USV fits the in-sample yield curve more accurately, whereas $A_1(3)$ USV exhibits less bias and potentially more stable performance out-of-sample. There is a trade-off between precision and bias, whose decision depends on the application, as we shall also discuss under the volatility regression below. Although previous studies have associated model misspecification with high autocorrelation in residuals (Collin-Dufresne et al., 2009), we did not perform an autocorrelation analysis in this study. Future research could incorporate this test to further assess model adequacy.

Table 6. In-sample and out-sample zero yields Y_t are compared to model-fitted estimates $\hat{Y}_t$ for the 0.25, 5,10, 12, 20, 25, and 30-year maturities. The fifth column reflects the statistical significance at 5% and 1%, labelled with ** and *, respectively. Bias is determined in terms of the null hypothesis that bias is zero, while RMSE is based on a pairwise analysis of the models. The inequality sign provides the direction of a loss.

Maturity	$A_1(3)$ USV	Loss Direction	$A_1(4)$ USV	Significance
	In-sample RMSE [1]			
0.25	0.0703	<	0.0683	*
5	0.0868	<	0.0856	
10	0.0958	<	0.0946	
12	0.0994	<	0.0984	
20	0.1045	<	0.1040	
25	0.1053	<	0.1043	
30	0.1046	<	0.1040	
	In-sample bias [2]			
0.25	0.0660		0.0661	
5	0.0841		0.0842	
10	0.0938		0.0939	
12	0.0978		0.0979	
20	0.1035		0.1036	
25	0.1039		0.1041	**
30	0.1035		0.1036	
	Out-sample RMSE [3]			
0.25	15.3984	<	0.086407	**
5	0.2219	<	0.0896	
10	0.0780	>	0.1012	
12	0.0612	>	0.1095	
20	0.0389	>	0.1233	
25	0.0505	>	0.1241	
30	0.0256	>	0.1232	
	Out-sample bias [4]			
0.25	−12.0832		0.0848	**
5	0.0855		0.0879	
10	0.0078		0.0996	
12	−0.0549		0.1079	
20	−0.0369		0.1218	
25	−0.0495		0.1227	
30	−0.0201		0.1218	

[1] DM t-statistic = 6.1233, p-value = 0.0009 → Reject H_0. Better model: $A_1(4)$ USV (based on lower RMSE). [2] DM t-statistic = −8.0000, p-value = 0.0002 → Reject H_0. Better model: $A_1(3)$ USV is significantly less biased. [3] DM t-statistic = 0.9999, p-value = 0.3559 → Fail to reject H_0. Better model: $A_1(4)$ USV (based on lower RMSE). [4] DM t-statistic = 0.9571, p-value = 0.3755 → Fail to reject H_0. Better model: $A_1(4)$ USV (based on lower bias).

7.3. Time-Series Dynamics

In this section, we evaluate the properties of model-implied[4] state variables. To estimate them, we run a posterior sampler by holding the parameters in Tables 2 and 3 fixed. Smoothed estimates are obtained by averaging the results of the draws of state variables (Collin-Dufresne et al., 2009). Both models produce a time series of state variables X_t, including a filtered volatility V_t. From these model-implied state variables, we compute various correlations against the actual yield time-series variables. A 26-week rolling standard deviation of the log returns for the 3-month actual yields is fitted. As a comparative means, we also repeated the computation of the 26-week rolling standard deviation for the 12-year and 30-year maturities. Yield curve variables, slope was computed as a function of the 30-year and 3-month yield as $\left(Y_{30y} - Y_{0.25y}\right)$, and the curvature as $\left(Y_{30y} - 2Y_{12y} + Y_{0.25y}\right)$. Other variables include the GARCH(1,1) volatility of the same 3-month log returns data used for computing the 26-week rolling standard deviation, FX volatility for the SA Rand Dollar currency pair, and Brent crude volatility.

Table 7 compares models $A_1(3)$USV and $A_1(4)$USV in terms of the various correlations between model-implied variables and the actual time-series variables for 3-month, 12-year, and 30-year maturities. The $A_1(3)$ model shows strong alignment with yield curve decomposition literature of Litterman et al. (1991), by capturing over 99% of the variance in the yield data. All three components, level, slope, and curvature, show very high correlations with their empirical counterparts, with over 96%, especially curvature with 98.9%, consistent with theoretical expectations. Figure 1 illustrates how the first three components contribute 99.97% to total variation. As a result, 0.03% variance explained is well below any reasonable threshold and negligible.

Table 7. Model-implied volatility proxy for the weekly yield returns computed as a 26-week rolling standard deviation. Initially, the rolling statistic is computed for the 3-month (short-term) yields, then compared to the 12-year (mid-term) and 30-year (long-term) yields. The model-implied volatility is derived as filtered V_t from each model. Other time-series variables for the same maturities are the GARCH, Brent crude, USDZAR, and the three PCA factors: level, slope, and curvature. The Spearman correlation is calculated between the model-implied volatility and these variables.

	$A_1(3)$			$A_1(4)$		
Maturities	**0.25**	**12**	**30**	**0.25**	**12**	**30**
Actual vs. Model Average Yield	0.964	0.964	0.964	1.000	1.000	1.000
Actual vs. Model Sope	0.974	0.974	0.974	0.919	0.933	0.933
Actual vs. Model Curvature	0.989	0.989	0.989	0.733	0.749	0.749
Rolling vs. Model Volatility	0.054	0.002	−0.037	0.550	0.157	−0.049
SA Rand Dollar vs. Model Volatility	0.007	−0.024	0.058	0.408	0.122	0.613
Brent Crude vs. Model Volatility	0.050	0.064	0.000	0.216	0.055	0.478
GARCH vs. Model Volatility	0.026	−0.040	−0.055	0.128	0.145	−0.019
Curvature vs. Model Volatility	−0.119	0.034	−0.034	−0.421	−0.049	−0.297
Curvature vs. Model Variance	−0.119	0.034	−0.034	−0.421	−0.049	−0.297

The $A_1(4)$USV model introduces a marginal component explaining only 0.03% of variance. Including this factor raises the average yield fit to $r_t = 1.0$ due to overfitting, but significantly worsens the curvature correlation—down to 0.733, likely due to numerical instability introduced by a non-meaningful fourth factor. Although the $A_1(4)$USV model slightly improves average yield, it does so at the cost of degrading the structural interpretation of key term structure components.

The $A_1(4)$USV model exhibits a larger correlation of 0.550 between the model-implied and 26-week rolling for the short term, even though it decreases for both mid and long term. The $A_1(3)$USV model, on the other hand, exhibits very low or even negative correlations of

0.054, 0.002, and −0.037 for the short, mid, and long term, respectively. The GARCH(1,1) displays the results that are almost similar, except that surprisingly some slight increase in correlations is observed with the $A_1(4)$USV model.

The left panel of Figure 4 plots the 26-rolling volatility and the model-implied volatility V_t against the period 2016–2024. Subplot (a) displays the inability of the $A_1(3)$USV model volatility to track the variation of short rate volatility with a 26-week rolling; instead, the model exhibits a flat pattern. The $A_1(4)$USV volatility, on the other hand, tracks the variation of the short rate volatility closely with the 26-week rolling volatility.

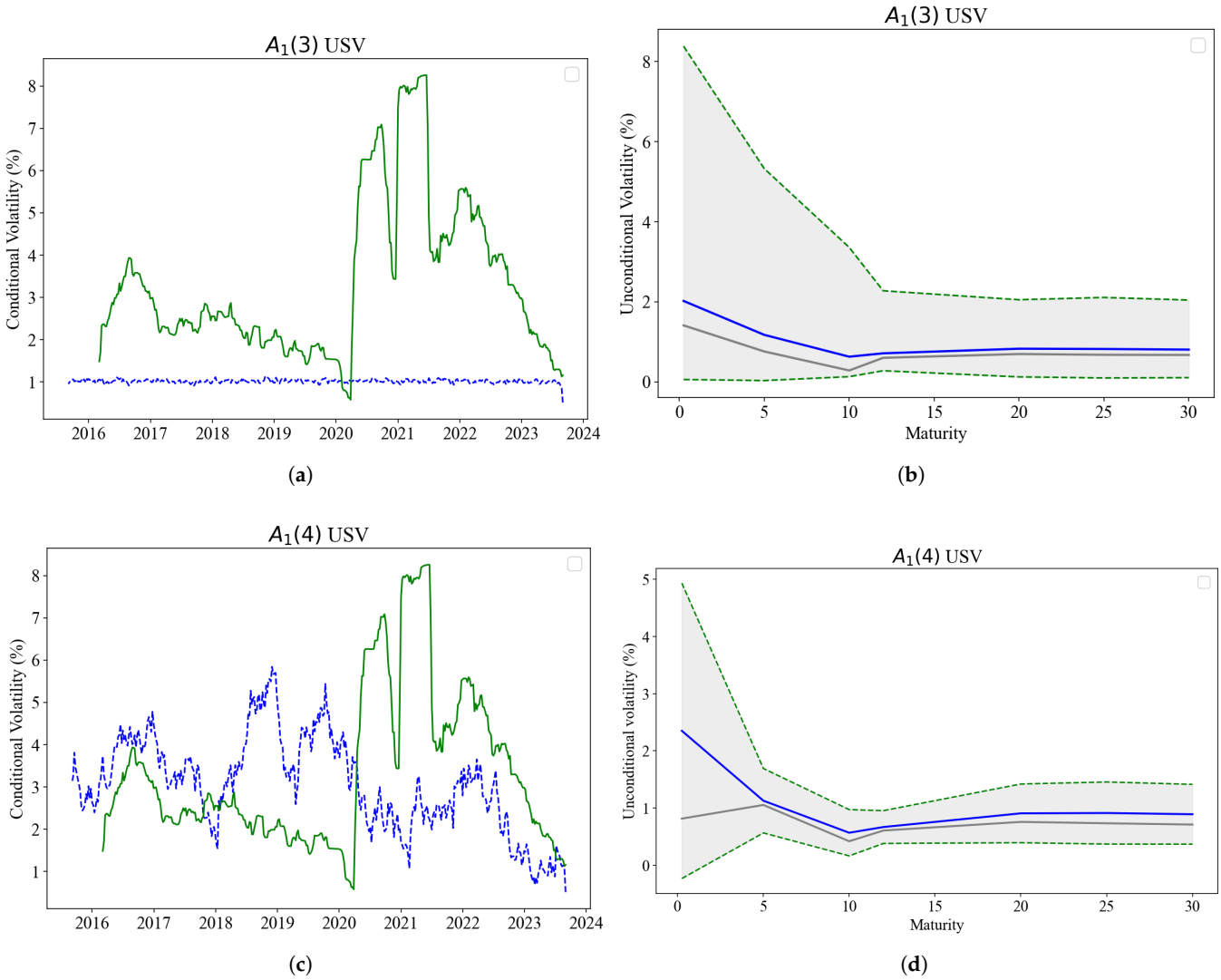


Figure 4. Conditional and unconditional volatility are plotted with the probability distributions of parameters. (**a**) $A_1(3)$USV model-implied conditional volatility V_t in blue, plotted with a 26-week rolling volatility in green. (**b**) $A_1(3)$USV model-implied unconditional volatility in blue and shown within the 95% confidence bounds, is plotted with a 26-week rolling volatility in green. (**c**) $A_1(4)$USV model-implied conditional volatility V_t in blue, plotted with a 26-week rolling volatility in green. (**d**) $A_1(4)$USV model-implied unconditional volatility in blue and shown within the 95% confidence bounds, is plotted with a 26-week rolling volatility in green. Conditional volatilities are plotted against a time period 2016–2024, while the unconditional volatilities are plotted against the maturities.

As a robustness check, we compare the correlations between the model-implied volatility V_t and FX volatility, and Brent crude. Correlations for the 3-month display a positive relationship with the model-implied volatility V_t of 0.408 for the SA Rand Dollar, and 0.216 for the Brent crude volatilities. A poor picture is exhibited for the mid and

short end of the yield curve regarding the correlations. For the yield curve factors level, slope, and curvature, there is generally a positive relationship with the model-implied volatility V_t.

The curvature factor exhibits negative correlation with model-implied volatility for both models, with $A_1(3)$USV showing slightly higher correlation of -0.119 when compared to more negative -0.421 for the $A_1(4)$USV models. The negative correlations between the model-implied volatility and the curvature factors—a yield curve component—confirm the existence of the tension between time-series dynamics and yield curve dynamics.

$A_1(4)$USV fits the yield curve more accurately, as demonstrated by stronger sensitivities to level, slope, and curvature factors and a slightly better R^2 at short maturities. This suggests it effectively captures the dominant yield dynamics at the front end, where traditional factors drive movements. USV restrictions on $A_1(3)$ appear to reduce performance slightly at the short end—likely due to overfitting or noise from volatility interactions, they provide more stable and potentially more robust behavior over the long term. $A_1(3)$ USV capture the persistent stochastic volatility effects and improve out-of-sample stability, despite a lower in-sample R^2. $A_1(4)$USV exhibits stronger level, slope, and curvature loadings at the short end, while $A_1(3)$USV shows nuanced volatility-related dynamics impacting longer maturities. This highlights that USV dynamics interact differently with yield curve factors across maturities. Ultimately, $A_1(4)$USV excels at capturing short-term yield behavior, whereas $A_1(3)$USV's richer structure better reflects long-term volatility features, suggesting a trade-off between in-sample fit and out-of-sample stability. These findings underscore the importance of carefully modeling factor-volatility interactions in yield curve models, especially when testing for USV.

7.4. Volatility Forecasting and Regression

In this section, we analyze the in-sample and out-sample performance using two volatility proxies, the absolute one-week of changes in yield returns $E[\|\Delta Y\|]$ and realized volatility $\hat{\sigma}$. We use the same data for parameter point estimates from Tables 2 and 3 to extract our weekly forecasts.

Following the methodology adopted in Collin-Dufresne et al. (2009), we implement a rolling-window approach using a two-year or approximately 104-week estimation window to construct one-week-ahead forecasts of the state variable. At each step, model parameters obtained from a posterior sampler are held fixed. The rolling process is used to re-estimate the latent state variables based on incoming data. This recursive structure allows for dynamic updating of state variables while maintaining stable parameter estimates. Forecasts are generated for both a 104-week in-sample period and a 45-week out-of-sample period. This design enables robust comparison of forecast accuracy across different subsamples and simulates a realistic forecasting scenario.

To generate each forecast, we first estimate the current values of the latent state variables using only information available up to time t. This is conducted in a manner consistent with the previous section, except that in the forecasting context, no data beyond time t is used in the state filtering step. Importantly, the parameter estimates used in this process are fixed at their posterior modes or means, obtained from the full sample estimation. Given these current-state estimates, we simulate 10,000 paths of the model forward one week to generate the predictive distribution of the state variables at time $t+1$. From this simulated distribution, we then construct a forecast distribution for each yield and volatility proxy of interest per maturity.

The second proxy, realized volatility is given as

$$\hat{\sigma_{t,\tau}} = \sqrt{\sum_{i=1}^{N} \Delta Y_{t,i,\tau}^2} \tag{29}$$

where $Y_{t,i,\tau}$ is the τ-maturity yield on the ith week following the observation at t.

Traditionally, realized volatility is better estimated using high-frequency return data, such as daily, to increase the certainty over the reliability and accuracy of the forecast. In this study, we use weekly yield data to construct both proxies for volatility. Our choice is motivated both by data availability and by the structure of the domestic bond market, where SA government treasury bonds are traded during weekly auctions. It is our view that these auctions anchor the pricing of yields, meaning that weekly data capture economically meaningful shifts in investor expectations, interest rate risk, and macro-economic signals. Intermediate level frequencies, such as weekly or monthly, are better than both high-frequency and low-frequency levels in producing sign dependence in volatility, which is the source of forecastability in asset returns (Christoffersen & Diebold, 2002).

7.4.1. Forecasting and Model Performance

We report both in-sample and out-of-sample RMSEs in Table 8. In each case, the DM test is employed to assess the statistical significance of differences in forecast accuracy between models. The in-sample RMSE of $\|\Delta Y\|$ exhibits a mixed picture. For the first three maturities, model $A_1(4)$USV shows a greater reduction in RMSEs than $A_1(3)$USV, indicating an improved fit. For the remaining four maturities, $A_1(3)$USV performs better. Despite this maturity-dependent performance, the overall DM test yields a test statistic of 0.053 with a p-value of 0.959, indicating that the observed differences are not statistically significant at conventional levels.

Turning to the in-sample RMSE of the estimated volatility series $\hat{\sigma}$, $A_1(4)$USV consistently outperforms $A_1(3)$USV across all maturities. The DM statistic of 3.976 and a p-value of 0.007 support this result with significance at the 1% level. This suggests that $A_1(4)$USV more accurately captures the in-sample volatility dynamics of yields.

Out-sample, $A_1(4)$USV again shows improved performance in forecasting $\|\Delta Y\|$, with uniformly lower RMSEs. The DM statistic of 2.404 and a p-value of 0.053 suggest a statistical significance at the 5% level. This result supports $A_1(4)$USV's superior ability to forecast short-term volatility in yield changes. Out-sample performance of the volatility forecasts is even more decisive. $A_1(4)$USV demonstrates markedly lower RMSEs in forecasting $\hat{\sigma}$ with a DM statistic of 6.791 and a p-value of 0.001. This result is highly significant and highlights the robustness of $A_1(4)$ USV in capturing the underlying volatility dynamics beyond the estimation window.

It is worth noting, a tension between time-series volatility performance and cross-sectional yield curve dynamics. While $A_1(4)$USV consistently outperforms $A_1(3)$ USV in forecasting volatility, both in-sample and out-sample, the initial inconsistency across maturities in the RMSE of $\|\Delta Y\|$ reflects a possible trade-off. Specifically, models that are optimized for volatility dynamics may not always align perfectly with those tailored for fitting the cross-section of the yield curve.

This tension is not surprising. Yield curve models often prioritize cross-sectional fit at a point in time, while volatility models emphasize dynamic consistency and predictive power over time. The results here suggest that $A_1(4)$USV, potentially incorporating richer dynamics or additional latent volatility structure, sacrifices some cross-sectional fit in the short term, as seen in maturities of 12, 20, 25, and 30 years, but gains substantially in time series forecasting accuracy.

Ultimately, the decision between models depends on the application, such as pricing and hedging, for which cross-sectional accuracy may dominate. For risk management and forecasting, superior volatility dynamics as achieved by $A_1(4)$USV are likely to be more valuable.

Table 8. In-Sample and Out-Sample one-week forecast of two different volatility proxies $E[\|\Delta Y\|]$ and $\hat{\sigma}$ for the 0.25, 5, 10, 12, 20, 25, and 30-year maturities. The fifth column reflects the statistical significance at 5% labelled with **. respectively. Bias is determined in terms of the null hypothesis that bias is zero, while RMSE is based on a pairwise analysis of the models. The inequality sign provides the direction of a loss.

Maturity	$A_1(3)$ USV	Loss Direction	$A_1(4)$ USV	Significance
In-Sample RMSE of weekly $\|\Delta Y\|$ (bps) [5]				
0.25	131.83	<	37.53	**
5	50.38	<	25.22	
10	39.66	<	20.54	
12	31.97	>	45.51	
20	27.15	>	79.48	
25	27.50	>	81.28	
30	26.69	>	78.63	
In-Sample RMSE of $\hat{\sigma}$ (bps) [6]				
0.25	44.68	<	39.64	**
5	28.42	<	25.92	
10	26.91	<	25.22	
12	26.92	<	24.97	
20	27.75	<	24.72	
25	26.85	<	23.90	
30	27.10	<	24.26	
Out-Sample RMSE of weekly $\|\Delta Y\|$ (bps) [7]				
0.25	18.35	<	14.87	
5	33.54	<	33.22	
10	41.45	<	41.39	
12	45.38	<	44.91	
20	56.44	<	49.60	
25	57.51	<	50.00	
30	58.16	<	50.18	
Out-Sample RMSE of $\hat{\sigma}$ (bps) [8]				
0.25	15.60	<	14.41	
5	29.58	<	28.56	
10	38.52	<	37.26	
12	41.57	<	40.24	
20	39.71	<	38.18	
25	40.57	<	38.88	
30	40.56	<	38.82	

[5] DM t-statistic = 0.0527, p-value = 0.9597 $\rightarrow$ Fail to reject H_0. Better model: $A_1(4)$ USV (based on lower RMSE). [6] DM t-statistic = 2.4035, p-value = 0.0530 * $\rightarrow$ Fail to reject H_0. Better model: $A_1(4)$ USV (based on lower RMSE). [7] DM t-statistic = 6.7912, p-value = 0.0005 $\rightarrow$ Reject H_0. Better model: $A_1(4)$ USV (based on lower RMSE). [8] DM t-statistic = 6.7912, p-value = 0.0005 $\rightarrow$ Reject H_0. Better model: $A_1(4)$ USV (based on lower RMSE).

We also notice that both models match the unconditional volatility of yield changes, with $A_2(4)$USV exhibiting a more refined pattern. The right panel of Figure 4 displays the unconditional volatility plotted against maturities, standard deviation of model-fitted yield changes ΔY_t in blue and lying within the light-gray bound distribution of model-implied yield volatilities as determined by the point estimates. The model-fitted unconditional volatility clearly fits along the mean and median of the distribution. In the bottom plot (d), we note that $A_1(4)$ USV appears to be more of an improvement from $A_1(3)$ USV in the top

plot (b). In $A_1(4)$ USV, the model-fitted unconditional volatility (blue line) matches the pattern of the distribution average (gray line) better when compared to $A_1(3)$ USV. Both plots exhibit a snake-shaped curve as discussed by Piazzesi (2010) for US Treasury yields and swaps. From an improved plot (d), $A_1(4)$ USV, the pattern commencing with the back of the snake, followed by a hump towards 5-year maturities, a hump and drops towards the 10-year maturity, thereafter decreasing in a stable manner towards longer maturities. High volatility over the period between 3-month to 5-year maturities, also termed by Piazzesi (2010) the back-of-the-snake, is due to reactions from monetary policy events, changes in liquidity premia, and macro-economic dynamics, among others. It is also a key to the factor correlations.

7.4.2. Regression

To evaluate the extent to which model-derived volatility V_t can be explained by both term structure dynamics and exogenous volatility sources, we estimate the following linear regression model:

$$V_t = \alpha + \beta^T X_t + \gamma_i \Sigma_{m_{i,t}} + \epsilon_t \tag{30}$$

where X_t denotes the yield curve factors derived from either the three-factor model with level, slope, and curvature or a four-factor model including latent volatility V_t in the term structure model. $\Sigma_{m_{i,t}}$ represent the M observable macro variable proxies at time t for $i = 1, ..., M$, and γ_i is the coefficient for the i-th macro variable.

Table 9 presents the results of volatility regressions for the 3-month and 30-year volatilities. The GARCH model exhibits moderate explanatory power, with R^2 values between 0.221 and 0.477, suggesting economically meaningful relationships in both short- and long-term volatility regressions, regardless of the inclusion of PCA factors. Applying PCA on both models $A_1(3)$USV and $A_1(4)$USV significantly improves the model performance, increasing R^2 values from approximately 0.022 to 0.428 for both models, with minimal differences observed between them. Since both models achieve similar explanatory power after PCA with $R^2$0.43, R^2 alone cannot be used to select the better model.

Stochastic volatility models, $A_1(3)$USV and $A_1(4)$USV exhibit weak explanatory power in their original specification with R^2 values of 0.022 and 0.027, but performance improves markedly after applying PCA to R^2 values of 0.428 and 0.427, indicating that latent yield curve factors significantly enhance the models' ability to capture volatility dynamics. Despite similar fit post-PCA, $A_1(3)$USV demonstrates a much stronger negative relationship between volatility and the yield curve factors with β coefficient of -6.876, while $A_1(4)$USV exhibits a modest positive relationship with β of 0.810, suggesting fundamentally different volatility structures. The estimated loadings on the level, slope, and curvature components further reveal that $A_1(4)$USV is more responsive to changes in yield curve shape, especially in terms of slope –0.173 and curvature -0.364. These findings imply that while both models satisfy the USV condition by showing significant latent factor influence, $A_1(4)$USV may better align with USV theory by isolating volatility innovations not captured by the term structure, though $A_1(3)$USV captures a stronger overall volatility response. In the second regression, $A_1(3)$USV and $A_1(4)$USV exhibit a beta of -2.144 and 0.302, respectively, suggesting that $A_1(4)$USV is still positively related to realized volatility and aligns better. A drop in β coefficient from 0.810 to 0.302 only suggests that the model $A_1(4)$USV performs better in the short term than the longer term.

Table 9. Regressions for weekly realized volatilities $\hat{\sigma}$ 3-month and 30-year maturities. These regressions are performed on different variables. They are reflected as systematic risks according to β coefficients for each variable and standard errors in square brackets. For each model, there are two rows: the first without PCA factors and the second with PCA factors, level, slope, and curvature. These regressions were performed on the time-series variables for the period 2013 to 2024. The top panel is the regression results for the 3-month weekly realized volatility, and the bottom panel is for the 30-year weekly realized volatility.

Variable	Intercept(α)	Volatility(β)	R^2	Level	Slope	Curvature
			3-Month Yield Volatilities			
GARCH(1,1)	0.202 [0.015]	0.395 [0.036]	0.240			
	0.210 [0.013]	0.228 [0.036]	0.477	−0.025 [0.005]	−0.112 [0.009]	−0.279 [0.025]
$A_1(3)$ USV	1.196 [0.2619]	−8.473 [2.576]	0.027			
	0.977 [0.203]	−6.876 [1.999]	0.428	−0.027 [0.005]	−0.132 [0.009]	−0.334 [0.024]
$A_1(4)$ USV	0.069 [0.090]	0.810 [0.274]	0.022			
	0.696 [0.126]	−1.321 [0.3956]	0.427	−0.043 [0.007]	−0.173 [0.016]	−0.364 [0.025]
			30-Year Yield Volatilities			
GARCH(1,1)	0.068 [0.007]	0.472 [0.045]	0.221			
	0.067 [0.007]	0.447 [0.045]	0.241	−0.005 [0.002]	−0.005 [0.004]	−0.023 [0.009]
$A_1(3)$ USV	0.350 [0.085]	−2.144 [0.835]	0.017			
	0.317 [0.084]	−1.862 [0.8262]	0.062	−0.007 [0.002]	−0.006 [0.004]	−0.032 [0.010]
$A_1(4)$ USV	0.038 [0.025]	0.302 [0.080]	0.035			
	−0.293 [0.046]	1.393 [0.150]	0.221	0.010 [0.003]	0.037 [0.006]	−0.032 [0.009]

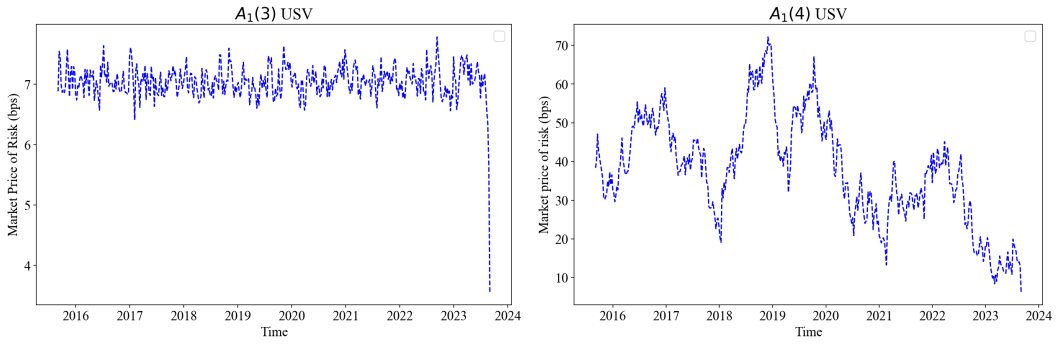
The purpose of this regression is to evaluate the extent to which the stochastic volatility models capture information embedded in the yield curve, and whether they span the term structure or USV properties. Two key insights emerge from both the 3-month and 30-year maturity regressions. First, time-series information, extracted through PCA, proves substantially more informative than raw cross-sectional data. In their unaugmented forms, both models exhibit weak explanatory power with R^2 approximately 0.022–0.027, indicating that volatility is poorly explained by contemporaneous yield levels alone. However, once the latent factors of level, slope, and curvature are introduced, explanatory power improves substantially, with R^2 increasing to 0.428 for the 3-month volatility forecast and up to 0.221 for the 30-year forecasts. This highlights the importance of dynamic yield curve behavior in driving interest rate volatility.

Second, although both models are influenced by latent yield curve factors, they diverge in structure and responsiveness. $A_1(3)$ USV displays a strong contemporaneous volatility response, with large negative coefficients on latent factors. However, it shows limited sensitivity to yield curve shape in the long-term regression. This implies that $A_1(3)$ USV primarily reacts to broad level shifts, lacking a nuanced view of the term structure, particularly over longer horizons.

7.4.3. Market Price of Risk

Figure 5 plots a comparison of MPR components from the $A_1(3)$ and $A_1(4)$ USV models. The $A_1(3)$ model yields relatively stable and low-magnitude MPRs, while the $A_1(4)$ USV model exhibits consistently higher values and greater variability, suggesting that the additional factor introduces a more pronounced risk compensation mechanism. This difference, largely a result of model specification and scale, highlights how additional factors can amplify perceived risk premia. Given that both sets of MPRs are derived from Bayesian posterior point estimates under the risk-neutral measure, the results reflect not

just data fit but also the influence of prior beliefs. Overall, the 4-factor model implies a richer, possibly more realistic, structure for pricing risk.



(**a**) (**b**)

Figure 5. Market price of risk in basis points is plotted against the time series for the period 2013–2024. The first sub-figure (**a**) plots the model $A_1(3)$ USV and second (**b**) plots $A_1(4)$ USV both against time. They are based on the proportional relationship between $\sqrt{S_t}$ and the essentially affine $\Lambda = \lambda_1 + \lambda_2$.

8. Conclusions

We assume that restricted ATSMs with USV can be effectively estimated through Bayesian methods in emerging markets using only time-series data. As a result, we do not follow a traditional joint approach of estimation using both bond price and options data. We evaluate the sensitivity of oil shock and exchange rate volatility to the interest rate volatility. We test how the restricted models $A_1(3)$USV and $A_1(4)$USV respond to the tension between cross-sectional and time-series dynamics of bond prices.

The tension between time series and yield curve dynamics does not go away completely. Evidence of this is highlighted by a negative correlation between the curvature versus both volatility and variance. Model $A_1(4)$USV performs better than $A_1(3)$USV in capturing the time-series dynamics. There is a necessary trade-off between the time series or yield curve fitting, which depends on the application—option pricing, portfolio management, risk management and hedging, and policy formulation.

Model flexibility and parameter uncertainty issues are reduced by the Bayesian MCMC estimation strategy. The result is that parameters become locally and globally identifiable and economically meaningful, such that they are easily compared to parameters used in data from other countries.

Our study did not assess the autocorrelation of residuals in bias and RMSE due to computational burden. We did not factor the macro variables into a joint modeling but only used them as a robustness test mechanism. Future research should consider a joint modeling of macro variables as a means to compensate for the role of options data in a joint estimation and evaluate their impact on parameter estimation in the context of Bayesian modeling.

Author Contributions: Conceptualization, M.M. and G.v.V.; methodology, G.v.V.; software, M.M.; validation, M.M. and G.v.V.; formal analysis, M.M. and G.v.V.; investigation, M.M. and G.v.V.; resources, M.M. and G.v.V.; data curation, M.M. and G.v.V.; writing—original draft preparation, M.M.; writing—review and editing, G.v.V.; visualization, M.M.; supervision, G.v.V.; project administration, G.v.V.; funding acquisition, None. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the authors upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ATSM	Affine Term Structure Models
BIC	Bayesian Information Criterion
BDFS	Balduzzi P, Das SR, Foresi S
DM	Diebold–Mariano
DK	Duffie and Kahn
DTSM	Dynamic Term Structure Models
ESS	Effective Sample Size
FX	Foreign Exchange
GMM	Generalized Method of Moments
HDI	High-Density Interval
MCSE	Monte Carlo Standard Error
MCMC	Markov Chain Monte Carlo
LRSQ	Linear-Rational Square-Root
MH	Metropolis–Hastings
ODE	Ordinary Differential Equation
PCA	Principal Component Analysis
RMSE	Root-Mean-Square Error
SA	South African
SME	Simulated Method of Estimation
SDE	Stochastic Differential Equation
SV	Stochastic Volatility
USV	Unspanned Stochastic Volatility
USDZAR	SA Rand Dollar

Appendix A. Derivation of the Physical Measure Drift

It is assumed that the market price of risk Λ_t takes an "essentially affine form" of Duffee (2002) in the state variables and is written as

$$\Lambda_t = \sqrt{S_t}\lambda_1 + \sqrt{S_t^{-1}}\lambda_2 X_t$$

where $\lambda_1 \in R^N$ and $\lambda_2 \in R^{N\times N}$ such that

$$\lambda_1 = \begin{bmatrix} \lambda_{r0} \\ \lambda_{\mu 0} \\ \lambda_{V0} \end{bmatrix}, \quad \lambda_2 = \begin{bmatrix} \lambda_{rr} & \lambda_{r\mu} & \lambda_{rV} \\ \lambda_{\mu r} & \lambda_{\mu\mu} & \lambda_{\mu V} \\ \lambda_{Vr} & \lambda_{V\mu} & \lambda_{VV} \end{bmatrix} \tag{A1}$$

To derive drift under a physical measure $\mathbb{P}$, we invoke the Girsanov's theorem: Firstly, the link between Brownian motions under both measures Q and P is established by

$$W_t^{\mathbb{Q}} = W_t^{\mathbb{P}} + \int_{0^t} \Lambda_t dt \tag{A2}$$

By expectation and dividing by dt on both sides we find

$$\frac{1}{dt}\mathbb{E}^{\mathbb{Q}}[dX_t] = \frac{1}{dt}\mathbb{E}^{\mathbb{P}}[dX_t] + \Lambda_t \tag{A3}$$

Substituting the first term on the right-hand side of (A3) with (13) we obtain the following

$$\frac{1}{dt}\mathbb{E}^{\mathbb{P}}[dX_t] = \begin{bmatrix} m_0 + m_r r_t + m_\mu \mu_t^{\mathbb{Q}} + m_V V_t \\ \mu_t^{\mathbb{Q}} \\ g_V - k_V V_t \end{bmatrix} + \left(\sqrt{S_t}\lambda_1 + \sqrt{S_t^{-1}}\lambda_2 \begin{bmatrix} r_t \\ \mu_t^{\mathbb{Q}} \\ V_t \end{bmatrix} \right) \tag{A4}$$

We now compute each part of the sum explicitly:
Step 1: Risk-neutral drift

$$\mu^{\mathbb{Q}}(X_t) = \begin{bmatrix} m_0 + m_r r_t + m_\mu \mu_t^{\mathbb{Q}} + m_V V_t \\ \mu_t^{\mathbb{Q}} \\ g_V - k_V V_t \end{bmatrix}$$

Step 2: Market price of risk term
Let us denote:

$$\Lambda_t = \Lambda_1 + \Lambda_2 \tag{A5}$$

where

$$\Lambda_1 = \sqrt{S_t}\lambda_1 = \begin{bmatrix} \sqrt{S_t}\lambda_{r0} \\ \sqrt{S_t}\lambda_{\mu 0} \\ \sqrt{S_t}\lambda_{V0} \end{bmatrix}, \quad \Lambda_2 = \sqrt{S_t^{-1}}\lambda_2 X_t = \sqrt{S_t^{-1}} \cdot \begin{bmatrix} \lambda_{rr} r_t + \lambda_{r\mu}\mu_t^{\mathbb{Q}} + \lambda_{rV} V_t \\ \lambda_{\mu r} r_t + \lambda_{\mu\mu}\mu_t^{\mathbb{Q}} + \lambda_{\mu V} V_t \\ \lambda_{Vr} r_t + \lambda_{V\mu}\mu_t^{\mathbb{Q}} + \lambda_{VV} V_t \end{bmatrix}$$

Assuming consistent scaling (or setting $\sqrt{S_t} = 1$ and $\sqrt{S_t^{-1}} = 1$ without loss of generality, as is common in affine models), we obtain:

$$\Lambda_t = \begin{bmatrix} \lambda_{m0} + \lambda_{rr} r_t + \lambda_{r\mu}\mu_t^{\mathbb{Q}} + \lambda_{rV} V_t \\ \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu\mu}\mu_t^{\mathbb{Q}} + \lambda_{\mu V} V_t \\ \lambda_{V0} + \lambda_{Vr} r_t + \lambda_{V\mu}\mu_t^{\mathbb{Q}} + \lambda_{VV} V_t \end{bmatrix} \tag{A6}$$

Step 3: Adding the terms to obtain a physical drift
Now, summing $\mu^{\mathbb{Q}}(X_t)$ and Λ_t:

$$\frac{1}{dt}\mathbb{E}^{\mathbb{P}}[dX_t] = \begin{bmatrix} m_0 + \lambda_{r0} + (m_r + \lambda_{rr}) r_t + (m_\mu + \lambda_{r\mu})\mu_t^{\mathbb{Q}} + (m_V + \lambda_{rV}) V_t \\ \mu_t^{\mathbb{Q}} + \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu\mu}\mu_t^{\mathbb{Q}} + \lambda_{\mu V} V_t \\ g_V + \lambda_{V0} - (k_V - \lambda_{VV}) V_t + \lambda_{Vr} r_t + \lambda_{V\mu}\mu_t^{\mathbb{Q}} \end{bmatrix} \tag{A7}$$

To simplify notation and group terms, define:

$$\tilde{m}_0 = m_0 + \lambda_{r0}, \quad \tilde{m}_r = m_r + \lambda_{rr}, \quad \tilde{m}_\mu = m_\mu + \lambda_{r\mu}, \quad \tilde{m}_V = m_V + \lambda_{rV}$$
$$\tilde{g}_V = g_V + \lambda_{V0}, \quad \tilde{k}_V = k_V - \lambda_{VV}$$

Finally, drift under a physical measure $\mathbb{P}$ becomes

$$\frac{1}{dt}\mathbb{E}^{\mathbb{P}}[dX_t] = \begin{bmatrix} \tilde{m}_0 + \tilde{m}_r r_t + \tilde{m}_\mu \mu_t^{\mathbb{Q}} + \tilde{m}_V V_t \\ (1 + \lambda_{\mu\mu})\mu_t^{\mathbb{Q}} + \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu V} V_t \\ \tilde{g}_V - \tilde{k}_V V_t + \lambda_{Vr} r_t + \lambda_{V\mu}\mu_t^{\mathbb{Q}} \end{bmatrix} \tag{A8}$$

Appendix B. Bayesian Estimation Algorithm

This appendix outlines the Bayesian estimation procedure for the ATSM, implemented via a Metropolis–Hastings (MH) within a Gibbs sampler. The algorithm follows the block sampling structure proposed in Collin-Dufresne et al. (2009); see also Singleton (2006) and Johannes and Polson (2010) for a detailed discussion on MCMC.

Appendix B.1. Latent State Structure and Volatility

The latent state vector X_t is partitioned as:

$$X_t = \begin{bmatrix} X_t^0 \\ V_t \end{bmatrix}, \quad \text{where } X_t^0 = \begin{bmatrix} r_t \\ \mu_t \end{bmatrix} \tag{A9}$$

Here, V_t is the latent stochastic volatility component, constrained below by $\underline{V}$. We define the deviation variable:

$$U_t = \max(V_t - \underline{V}, 0) \tag{A10}$$

Sampling is conducted over U_t, from which $V_t = \underline{V} + U_t$ is recovered.

Appendix B.2. Model Discretization

The continuous-time dynamics of the state vector under the physical measure $\mathbb{P}$ are approximated using the Euler–Maruyama scheme in (22):

Appendix B.3. State–Space Representation

The model is cast in the following discrete-time state–space form:

$$\mathcal{P}_t = K + LX_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_\varepsilon) \tag{A11}$$

$$X_{t+1} = m + FX_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q(V_t)) \tag{A12}$$

where:

- $\mathcal{P}_t \in \mathbb{R}^N$: observed yield vector by PC loading, see (23),
- $X_t \in \mathbb{R}^d$: latent state,
- $K \in \mathbb{R}^N, L \in \mathbb{R}^{N\times d}$: observation equation parameters,
- $m \in \mathbb{R}^d, F \in \mathbb{R}^{d\times d}$: state equation parameters,
- $Q(V_t)$: state-dependent transition covariance matrix,
- Σ_ε: observation noise covariance.

Appendix B.4. Prior Distributions

The parameter vector is partitioned into three blocks:

- $\phi^{\mathbb{Q}}$: risk-neutral dynamics parameters,
- ϕ^{λ}: risk premia parameters,
- ϕ^{Λ}: measurement error variances.

Priors are assigned as:

$$\begin{aligned} \phi^{\mathbb{Q}} &\sim \mathcal{N}(\mu_Q, \Sigma_Q) \\ \phi^{\lambda} &\sim \mathcal{N}(\mu_\lambda, \Sigma_\lambda) \\ \sigma^2_{\varepsilon,i} &\sim \text{Inverse-Gamma}(a_\varepsilon, b_\varepsilon), \quad i = 1, \ldots, N \\ X_0 &\sim \mathcal{N}(\mu_0, \Sigma_0) \end{aligned}$$

Appendix B.5. Posterior Structure

Let $\mathcal{P}_{1:T}$ denote the observed data expressed in terms of the linear relation between the yields and principal components and augmented with the term structure factor data $X_{1:T}$ (Collin-Dufresne et al., 2009). The posterior distribution is:

$$\begin{aligned} p(U_{1:T}, X^0_{1:T}, \phi^\Lambda, \phi^\lambda, \phi^Q \mid \mathcal{P}_{1:T}) \propto \\ p(\mathcal{P}_{1:T} \mid X^0_{1:T}, U_{1:T}, \phi^\Lambda) \cdot p(X^0_{1:T} \mid U_{1:T}, \phi^Q) \cdot p(U_{1:T} \mid \phi^Q) \\ \cdot p(\phi^\Lambda) \cdot p(\phi^\lambda) \cdot p(\phi^Q) \end{aligned}$$

Appendix B.6. Sampling Algorithm

We implement a Gibbs sampler with Metropolis–Hastings steps embedded for non-conjugate blocks. The full parameter space is divided into:

- $U_{1:T}$ (sampled in blocks),
- ϕ^Λ (measurement error),
- ϕ^λ (risk premia),
- ϕ^Q and X_0 (risk-neutral dynamics and initial condition),
- $X^0_{1:T}$ (Gaussian latent states).

Step 1: Sampling U_t for $t \in \{1, 1+h, 1+2h, \ldots, T\}$

For each t, the target density follows from the model's Markov structure:

$$p(U_t \mid \cdot) \propto p(U_{t+h}, X^0_{t+h} \mid U_t, X^0_t, \phi^Q) \cdot p(U_t \mid U_{t-h}, X^0_{t-h}, X^0_t, \phi^Q)$$

The second term is Gaussian and serves as the **proposal** in a random-walk MH algorithm. At iteration s:

1. Propose $\tilde{U}_t \sim \mathcal{N}(U_t^{(s-1)}, \sigma^2_{\text{prop}})$, enforce $\tilde{U}_t \geq 0$
2. Evaluate acceptance ratio:

$$\alpha = \min\left(1, \frac{p(U_{t+h}, X^0_{t+h} \mid \tilde{U}_t, X^0_t) \cdot p(\tilde{U}_t \mid \cdot)}{p(U_{t+h}, X^0_{t+h} \mid U_t^{(s-1)}, X^0_t) \cdot p(U_t^{(s-1)} \mid \cdot)}\right)$$

3. Accept $\tilde{U}_t$ with probability α, otherwise retain $U_t^{(s)} = U_t^{(s-1)}$

Step 2: Sample ϕ^Λ (Measurement Error)

- Conditional on $X^0_{1:T}, U_{1:T}, \phi^Q$
- If priors are inverse-gamma, a conjugate update is available

Step 3: Sample ϕ^λ (Risk Premia)

- Conditional on $X^0_{1:T}, U_{1:T}, \phi^\Lambda, \phi^Q$
- Use MH step if non-conjugate
- Enforce admissibility and stationarity

Step 4: Sample ϕ^Q and X_0

- Conditional on $X^0_{1:T}, U_{1:T}, \phi^\Lambda, \phi^\lambda$
- Sample ϕ^Q via MH or conjugate draw, enforce stationarity
- Sample $X_0 \sim \mathcal{N}(\mu_0^*, \Sigma_0^*)$ via conditional normal update

Step 5: Sample $X^0_{1:T}$ (Gaussian Latent States)

- Conditional on all parameters and $U_{1:T}$
- Use Forward-Filtering Backward-Sampling (FFBS) or Kalman smoother

Step 6: Iterate

- Repeat Steps 1–5 for $s = 1, \ldots, S$
- Store posterior draws; discard burn-in period

Appendix B.7. Convergence Diagnostics and Tuning

- Acceptance Rate: A reasonable rate is 20–40% in Metropolis–Hastings blocks. Proposal variances should be adjusted accordingly.
- Trace Plots and Autocorrelation: They are used to monitor the mixing behavior of key parameters and latent states.
- Gelman–Rubin Statistic $\hat{R}$: For multiple chains, verify $\hat{R} < 1.1$ for all monitored parameters to ensure convergence.
- Effective Sample Size (ESS): It is computed for each parameter to assess how many independent draws are available:

$$ESS = \frac{S}{1 + 2\sum_{k=1}^{\infty} \rho_k} \tag{A13}$$

  where ρ_k is the autocorrelation at lag k, and S is the total number of post-burn-in samples.
- Highest Density Intervals (HDIs): Use posterior quantiles or kernel density estimates to compute 90% or 95% HDIs:

$$\text{HDI}_\alpha = \{\theta : p(\theta \mid \text{data}) \geq c\} \tag{A14}$$

  where c is the density threshold that ensures coverage probability α.
- Burn-in and Thinning: Discard the first B iterations (e.g., $B = 2000$), where B represents the number of Burn-in iterations. Thinning is optional but may reduce autocorrelation in stored draws. For a detailed discussion on convergence diagnostics, see Roy (2020) and (Kumar et al., 2019) for software applications.

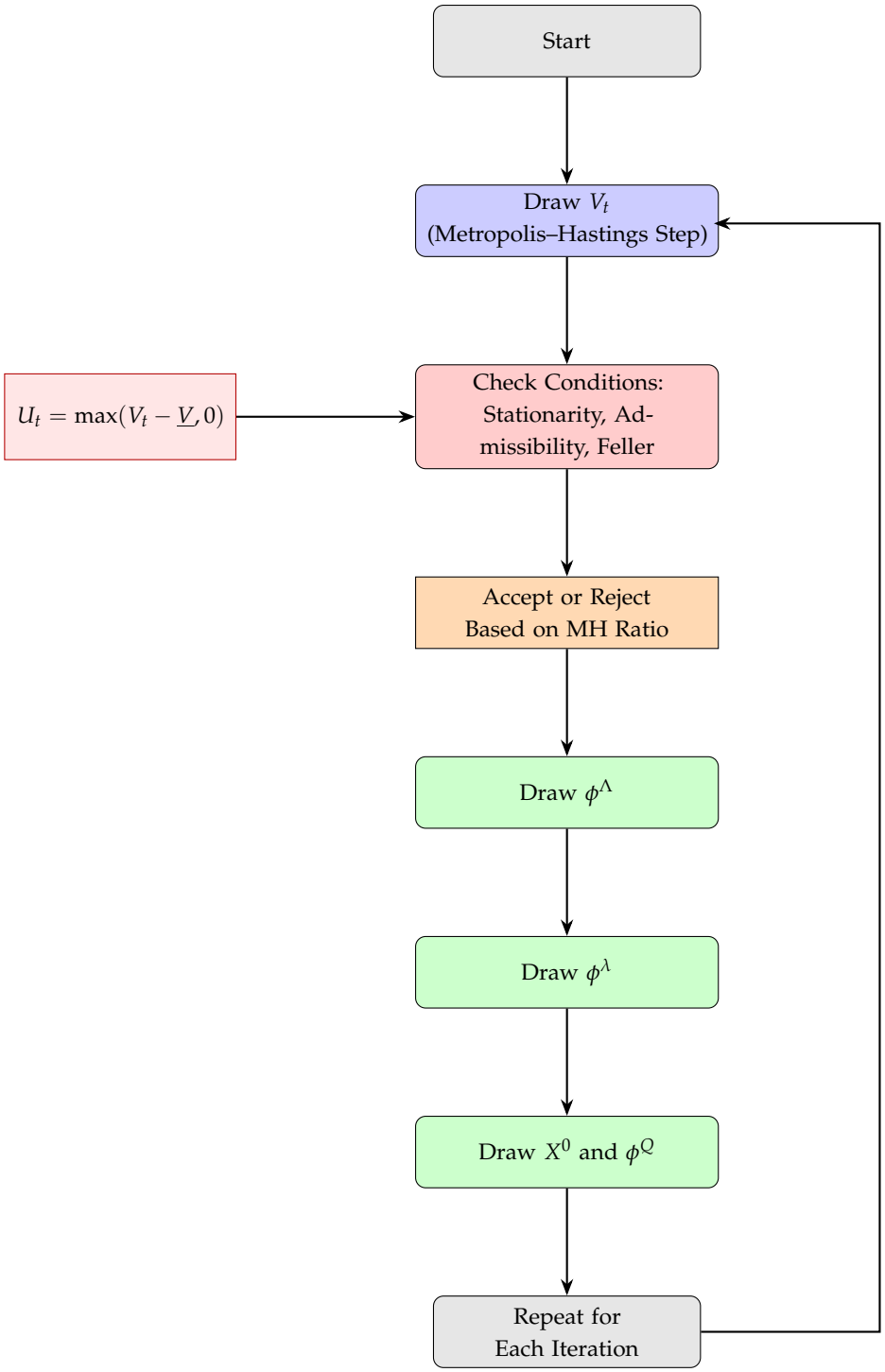
Appendix B.8. Alternative Samplers

More efficient sampling algorithms exist. For instance:

- Hamiltonian Monte Carlo (HMC) or No-U-Turn Sampler (NUTS):
  - They use gradient information to improve sampling efficiency
  - They are particularly useful for high-dimensional latent blocks
- Particle MCMC or Particle Gibbs:
  - It is well-suited to nonlinear and non-Gaussian state–space models

We opt for the MH-within-Gibbs algorithm for transparency, ease of constraint enforcement, and alignment with prior literature.

Appendix C. Workflow Diagram: Yield Curve Modeling and Inference



Appendix D. Algorithm: Yield Curve Inference via PCA, Kalman and MCMC

Algorithm A1 MCMC Algorithm with Block Sampling of U_t and Kalman Filtering

Require: Observed data Y, initial values $\Phi = \{\phi^{\mathbb{Q}}, \phi^{\Lambda}, \phi^{\lambda}\}$, initial state path $X^0_{1:T}$, and deviation path $U_{1:T}$
Ensure: Posterior samples of Φ, $X^0_{1:T}$, and $U_{1:T}$
for iteration = 1 to N **do**
  for each $t \in \{1, 1+h, 1+2h, \ldots, T\}$ **do**
    Propose $\tilde{U}_t \sim \mathcal{N}(U_t^{(s-1)}, \sigma^2_{\text{prop}})$, truncated below at 0
    Compute MH acceptance probability:

$$\alpha = \min\left(1, \frac{p(U_{t+h}, X^0_{t+h} \mid \tilde{U}_t, X^0_t) \cdot p(\tilde{U}_t \mid \cdot)}{p(U_{t+h}, X^0_{t+h} \mid U_t^{(s-1)}, X^0_t) \cdot p(U_t^{(s-1)} \mid \cdot)}\right)$$

    Accept or reject $\tilde{U}_t$ with probability α
    if constraint violations (e.g., Feller condition, stationarity) **then**
      Reject draw; set $U_t^{(s)} = U_t^{(s-1)}$
    else
      Accept draw; set $U_t^{(s)} = \tilde{U}_t$
    end if
  end for
  Draw $\phi^{\Lambda} \sim p(\phi^{\Lambda} \mid Y, X^0, U_{1:T})$
  Draw $\phi^{\lambda} \sim p(\phi^{\lambda} \mid Y, X^0, U_{1:T})$
  Sample latent states $X^0_{1:T}$:
  - Conditional on $U_{1:T}$, the system is linear-Gaussian
  - Apply Kalman filter and smoother to draw $X^0_{1:T}$ from its conditional posterior
  Draw $\phi^{\mathbb{Q}} \sim p(\phi^{\mathbb{Q}} \mid Y, X^0_{1:T}, U_{1:T})$
  Store current samples: Φ, $X^0_{1:T}$, $U_{1:T}$
end for
return Posterior draws of $\{\phi^{\mathbb{Q}}, \phi^{\Lambda}, \phi^{\lambda}\}$, $X^0_{1:T}$, $U_{1:T}$

Notes

[1] Collin-Dufresne et al. (2009) decomposes the state variable X into $X_t = \begin{bmatrix} X^0_t \\ V_t \end{bmatrix}$; where X^0 includes all the state variables r_t,μ_t, and θ_t, but exclude V_t. The reason is that V_t only affects the factor covariance matrix. They condition on the entire path of, write X^0_t and $\mathcal{P}$ in linear-Gaussian state–space form. The draws involving V can be done using relatively inefficient MH.

[2] Risk-neutral parameters $\mathcal{K}_V$ and γ_V are not identifiable under USV, hence they are replaced by $\gamma^P_V = \gamma_V + \lambda_{V0}$, and $\mathcal{K}^P_V = K_V - \lambda_{VV}$ (Collin-Dufresne et al., 2009).

[3] We use the Diebold and Mariano (2002) test to assess whether forecast $A_1(3)$ USV significantly outperforms $A_1(4)$ USV in terms of bias and RMSE. The global DM test statistic evaluates the null hypothesis of equal predictive accuracy across the full forecast horizon. Significance is indicated as follows: ** p-value < 0.01, * p-value < 0.05, p-value < 0.1.
In addition to the global DM statistic, we compute standardized per-point loss differentials d_i to highlight localized forecast performance differences. Each per-point z-score is defined as:
$z_i = \frac{d_i - \bar{d}}{s_d}$, where d_i is the pointwise difference in forecast losses, $\bar{d}$ is the mean loss difference, and s_d is the sample standard deviation. Significance levels per point are marked by: *** $|z| > 2.756$, ** $|z| > 1.960$, * $|z| > 1.645$

[4] By model-implied volatility we refer to the output from both $A_1(3)$ AND $A_1(4)$ USV models. Readers should not confuse this with the implied volatility surface (option-implied volatility).

References

Aït-Sahalia, Y., Li, C., & Li, C. X. (2024). Maximum likelihood estimation of latent Markov models using closed-form approximations. *Journal of Econometrics*, *240*(2), 105008. [CrossRef]

Andersen, T. G., & Benzoni, L. (2010). Do bonds span volatility risk in the US Treasury market? A specification test for affine term structure models. *The Journal of Finance*, *65*(2), 603–653. [CrossRef]

Andreasen, M. M., Jørgensen, K., & Meldrum, A. (2025). Bond risk premiums at the zero lower bound. *Journal of Econometrics*, *247*, 105939. [CrossRef]

Ang, A., Piazzesi, M., & Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, *131*(1–2), 359–403. [CrossRef]

Bikbov, R., & Chernov, M. (2004). Term structure and volatility: Lessons from the Eurodollar markets. *SSRN Electronic Journal* . [CrossRef]

Cheridito, P., Filipović, D., & Kimmel, R. L. (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics*, *83*(1), 123–170. [CrossRef]

Christoffersen, P., & Diebold, F. X. (2002, January 2). *Financial asset returns, market timing, and volatility dynamics.* Market Timing, and Volatility Dynamics. Available online: https://users.nber.org/~confer/2003/si2003/papers/efww/diebold.pdf (accessed on 25 June 2025).

Collin-Dufresne, P., & Goldstein, R. S. (2002). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *The Journal of Finance*, *57*(4), 1685–1730. [CrossRef]

Collin-Dufresne, P., Goldstein, R. S., & Jones, C. S. (2008). Identification of maximal affine term structure models. *The Journal of Finance*, *63*(2), 743–795. [CrossRef]

Collin-Dufresne, P., Goldstein, R. S., & Jones, C. S. (2009). Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics*, *94*(1), 47–66. [CrossRef]

Collin-Dufresne, P., Jones, C., & Goldstein, R. (2004). *Can interest rate volatility be extracted from the cross section of bond yields? An investigation of unspanned stochastic volatility.* Available online: https://www.epfl.ch/labs/sfi-pcd/wp-content/uploads/2021/07/Can-Interest-Rate-Volatility-be-Extracted-from-the-Cross-Section-of-Bond-Yields.pdf (accessed on 25 June 2025).

Dai, Q., & Singleton, K. J. (2000). Specification analysis of affine term structure models. *The Journal of Finance*, *55*(5), 1943–1978. [CrossRef]

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, *20*(1), 134–144.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, *57*(1), 405–443. [CrossRef]

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, *6*(4), 379–406. [CrossRef]

Elerian, O., Chib, S., & Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, *69*(4), 959–993. [CrossRef]

Hansen, J. W. (2025). Unspanned stochastic volatility in the linear-rational square-root model: Evidence from the Treasury market. *Journal of Banking & Finance*, *171*, 107354.

Heidari, M., & Wu, L. (2002, September 10). Term structure of interest rates, yield curve residuals, and the consistent pricing of interest rate derivatives. *Yield Curve Residuals, and the Consistent Pricing of Interest Rate Derivatives.* Available online: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c708679c36cd86175f128f6fdd5f33ffdc61e959 (accessed on 25 June 2025).

Johannes, M., & Polson, N. (2010). MCMC methods for continuous-time financial econometrics. In *Handbook of financial econometrics: Applications* (pp. 1–72). Elsevier.

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. (2019). ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software*, *4*(33), 1143. [CrossRef]

Litterman, R. B., Scheinkman, J., & Weiss, L. (1991). Volatility and the yield curve. *The Journal of Fixed Income*, *1*(1), 49–53. [CrossRef]

López-Pérez, A., Febrero-Bande, M., & González-Manteiga, W. (2025). Estimation and specification test for diffusion models with stochastic volatility. *Statistical Papers*, *66*(2), 40. [CrossRef]

Piazzesi, M. (2010). Affine term structure models. In *Handbook of financial econometrics: Tools and techniques* (pp. 691–766). Elsevier.

Riva, R. (2024). *How much unspanned volatility can different shocks explain? Available at SSRN 4878175.* Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4878175 (accessed on 25 June 2025).

Roy, V. (2020). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, *7*(1), 387–412. [CrossRef]

Singleton, K. J. (2006). *Empirical dynamic asset pricing: Model specification and econometric assessment*. Princeton University Press.

Stroud, J. R., Müller, P., & Polson, N. G. (2003). Nonlinear state-space models with state-dependent variances. *Journal of the American Statistical Association*, *98*(462), 377–386. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.