

Article

Not peer-reviewed version

Affine Term Structure Models with Unspanned Stochastic Volatility in Illiquid Derivative Markets

[Malefane Molibeli](#) * and [Gary van Vuuren](#) *

Posted Date: 30 June 2025

doi: 10.20944/preprints202506.2324.v1

Keywords: parameter and model identifiability; stochastic volatility; MCMC; unspanned 16 stochastic volatility



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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.

Article

Affine Term Structure Models with Unspanned Stochastic Volatility in Illiquid Derivative Markets

Malefane Molibeli ^{1,*}  and Gary van Vuuren ^{2,3} 

¹ School of Economics and Finance, University of the Witwatersrand, Johannesburg, South Africa

² Centre for Business Mathematics and Informatics, North-West University, Potchefstroom Campus, South Africa

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 between the restricted models versus 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 are employed to recover the model parameters and filter out latent state variables. We further incorporate macroeconomic 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-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

1. Introduction

Interest rate volatility is increasingly recognised as a distinguished feature in ATSMs and also an independent driver of time-varying risk premia in fixed income markets [1]. Typically, volatility is unobservable and potentially unspanned by the prices of traded bonds, raising significant challenges for both modelling and empirical inference. In particular, the inability to observe volatility directly—especially in markets with sparse derivatives trades due to 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 [1]. It results from a tension that exists between fitting a cross-section of yields and matching the time series dynamics of bond prices; see [2], [3], [4] and [5] 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 on bonds—suggesting the presence of portfolio volatility risk [2]. The presence of USV can explain persistent empirical puzzles such as the failure of (ATSMs) to match observed option-implied volatilities or the dynamic behaviour of the market price of risk (MPR) [1], [6] and [7].

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 [8] and [3] which in turn, are characterised 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 [8]. 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 [3]. 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. [5] points out that his 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. [3] 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 [8] for its ability to characterise 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 simultaneous fitting of certain cross-sectional and time-series properties in the ATSMs is circumvented. The three-factor ATSMs with completely affine specification by [8] was 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 [1].

Estimation methods will have a problem dealing with 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 [7]. Econometrics methods, including generalised method of moments (GMM), simulated methods of estimation (SME), among others, are unsuitable for investigating models with USV [7] 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 the non-linear processes. It is only used here as a computational device 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 [9] (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 render the fixed income markets to be incomplete, therefore, present some challenges to 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 a cross-sectional and time series dynamics simultaneously. [10] 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. [11] test 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 suggest other alternatives to the trade-off between fitting the cross-section of yields and capturing their time-series behaviour. [5] 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 [12] on quadratic diffusive and affine jump-diffusive models, although they found them to be incapable of accommodating the observed yield volatility dynamics. Their model exhibit 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 variation still requires some explanation.

[7] 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 fails 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 suggest further that short rate volatility cannot be extracted from the cross-section of bond prices. They propose an alternative representation of ATSM by [8] which is essentially characterised 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 [13]. In their companion paper, [2] 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 behaviour of bond prices. The presence of USV enforce the requirement that bond and derivative prices are required for estimation of parameters. Finally, USV presents some challenges on how fixed income derivatives should be hedged.

[6] explore 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 [2] 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 among others, jumps should be considered for the model performance to be improved. [14] maintains USV by allowing a clear separation between factors dedicated to term structure fitting and those for options fitting; see [6] and references therein. Similarly, they also reported the tension in pricing and parallel tension in matching lower and higher order moments, also suggesting the direction of adding jump components. [15] use 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. [1] 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. [16] 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 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 behaviour of the interest rate process and matching the cross-sectional shape of the yield curve.

[13] describes the concepts of identification, identifiability, and maximality. These are the centre for local and global identifiability of models and parameters, whose presence determine 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 [8], that 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. [3] 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 variable so that they have economic meaning. [3] 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 [17] and [8], their representation accommodates the USV [18].

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. [19] performs a comparative study between MCMC, Particle Filter and the Kalman Filter. Key challenges are the discretisation bias and the presence of a 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 suggest that the volatility depends on an additional factor that varies independently of the short rate level.

[20] 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. Computational complexity of these methods may present some challenges to our intended purpose. MCMC enable 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. [21] 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. [22] propose 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 [7] and

references therein. The same blocking approach is followed by [7] with the parameters vector broken into three blocks— those affecting the dynamics under risk-neutral measure, risk-premium parameters and the measurement error standard deviations.

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:

$$dX_t = \kappa^{\mathbb{Q}} (\theta^{\mathbb{Q}} - X_t) dt + \Sigma \sqrt{S_t} dW_t^{\mathbb{Q}} \quad (1)$$

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 i th element given by

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

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

$$r_t = \delta_0 + \delta_1 X_t \quad (3)$$

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 $i = 1, \dots, N$.

Provided that parameters are admissible, it is known from [17] that a zero-coupon bond has a solution

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

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

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

$$\frac{dB(\tau)}{d\tau} = -\theta' \mathcal{K}^{\mathbb{Q}} B(\tau) - \frac{1}{2} \sum_{i=1}^N [\Sigma' B(\tau)]_i^2 \beta_i + \delta_y \quad (6)$$

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

$$y_t(\tau) = -\frac{\log P_t(\tau)}{\tau} = \frac{A(\tau)}{\tau} + \frac{B(\tau)' X_t}{\tau} \quad (7)$$

We adopt a framework of [7] in which they use a classification scheme of [8] 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

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

They assign the state vectors $X = [r, \mu^Q, V]$ and $X = [r, \mu^Q, \theta^Q, V]$ for three-factor and four-factor models, respectively, where state variables are defined as follows:

$$r_t = Y_t(0) \quad (9)$$

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

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

$$V_t = \frac{1}{dt} dr_t^2 \quad (12)$$

Following [3], 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 $I \hat{v}$ diffusions and is 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 which 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 no USV; see [18]. The risk-neutral dynamics for the state vector X can be written as

Instantaneous mean:

$$\frac{1}{dt} \mathbb{E}^Q[dX_t] = \begin{bmatrix} m_0 + m_r r_t + m_\mu \mu_t^Q + m_V V_t \\ \mu_t^Q \\ g_V - k_V V_t \end{bmatrix} \quad (13)$$

where:

- m_0 = is a constant term
- $r_t = \delta_0 + \delta_1^\top X_t$ = feedback from the short rate
- μ_t^Q = slope of the yield curve
- m_V = volatility feedback

and Covariance matrix:

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

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.

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

There are six parameters in the drift and eight in the covariance matrix, resulting into 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 [8] and [3].

3.2. Market Price of Risk

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

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

The essentially affine form of [1] specifies the market price of risk (MPR) Λ_t as a nonlinear function of volatility

$$\Lambda_t = -\sqrt{S_t}\lambda_1 + \sqrt{S_t^{-1}}\lambda_2 \quad (17)$$

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 risk-neutral measure \mathbb{Q} and physical or historical measure \mathbb{P} be equivalent. It is implied that volatility or diffusion part of the SDE does not change in either measure and that only drift will be influenced by the risk premia. The drift for $A_1(3)$ state vector under physical measure \mathbb{P} is specified as

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

See appendix A for a detailed derivation:

3.4. The $A_1(4)$ Model

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

$$\frac{1}{dt}E^{\mathbb{Q}}[dX_t] = \begin{bmatrix} \mu_t^{\mathbb{Q}} \\ \theta_t^{\mathbb{Q}} + V_t \\ a_0 + a_r r_t + a_\mu \mu_t^{\mathbb{Q}} + a_V V_t \\ \gamma_V - \kappa_V V_t \end{bmatrix} \quad (19)$$

and

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

where

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

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

[2] describe the necessary and sufficient conditions for parameter restrictions under which an $A_1(3)$ model can display USV.

3.5.1. $A_1(3)$

Implementing the following restricted parameters characterise the $A_1(3)$ as a model with USV

$$m_r = -2c_V^2$$

$$m_\mu = 3c_V$$

$$m_V = 1$$

$$\sigma_V^2 = c_V^2$$

3.5.2. $A_1(4)$

[7] proposes an extension to the $A_1(3)$ restrictions and derive the following for the $A_1(4)$ model to display USV [7].

$$a_r = -2c_{r\mu}^2(3c_{r\mu} - a_\theta)$$

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

$$a_V = 3c_{r\mu}$$

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

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

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

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

3.6. Estimation STRATEGY

Bayesian estimation procedures for continuous-time finance are well documented in many sources including [22], [7] and references therein, and [23]. 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 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 [21]. The Gibbs sampler is the most 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 [23].

[7] considers two data augmentation approaches. First, augmenting with unobservable high frequency data, which enables the use of the Euler approximation, and provide a Gaussian density that is easy to work with. This introduces discretisation 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 with 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 discretisation altogether, enabling more stable and efficient inference.

We let $\mathcal{P} = \mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_T$ 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 $X = \{X_1, X_2, \dots, X_T\}$. 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 $p(X|\mathcal{P}, \Theta)$ [7].

True dynamics from SDE (1) are then approximated as¹

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

where Ω_t represents the instantaneous covariance matrix, and h is a discretisation time step, for which a choice of a smaller number is suitable for approximating to a normal distribution, in avoidance of using a non-central chi-square distribution of the square root factors.

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

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

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

$$\mathcal{P}_t = K + LX_t + \epsilon_t \quad (23)$$

(23) is further defined as a "measurement equation" in a state-space model, which we shall use in Kalman filtering to estimating 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:

$$\alpha = \min\left(1, \frac{\pi(X^*)p(X^{(t)}|X^*)}{\pi(X^{(t)})p(X^*|X^{(t)})}\right) \quad (24)$$

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

¹ [7] decomposes the state variable X into $X_t = \begin{bmatrix} X_t^0 \\ V_t \end{bmatrix}$; 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_t^0 and \mathcal{P} in linear Gaussian state space form. The draws involving V can be done using relatively inefficient MH.

between the parameters and latent states resulting in slow convergence. [23] recommends the use of artificial data to check the efficiency and convergence of the algorithm. [22] 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 C discusses briefly the estimation methodology and algorithm as adapted from [7] approach; where parameters are broken into three blocks, 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 ATSMs where zero-coupon bond prices satisfy

$$P(t, T) = E_t \left[\exp\left(-\int_t^T r_s ds\right) \right] \quad (25)$$

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 Riccati 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 paths, 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 yields data. It is noted that the first three components have a cumulative variance of 99.97% which agrees with the empirical evidence of [24]; 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 others.

	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

Table 1. Cont.

	Principal components						
	1	2	3	4	5	6	7
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.

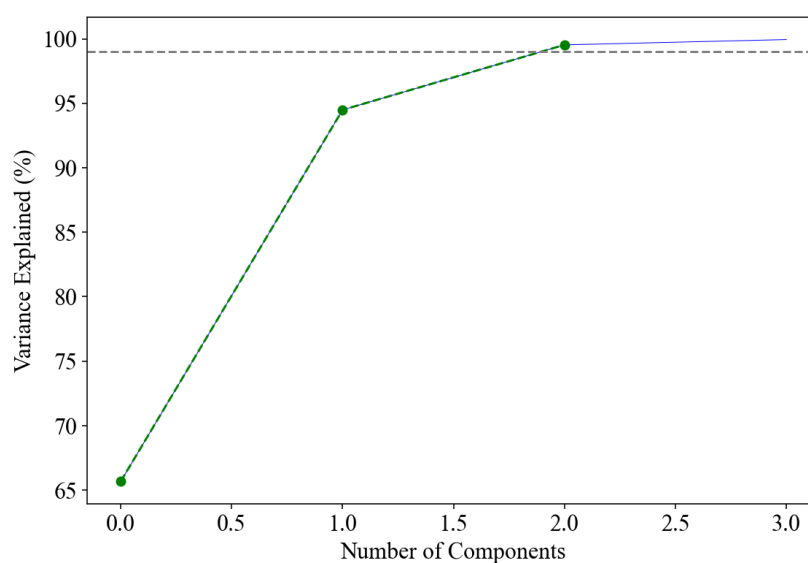


Figure 1. Cumulative explained variance in percentages, is plotted against number of principal components. The green dotted line with dots and blue line represent the cumulative variance for models $A_1(3)$ and $A_1(4)$ fitted yields, respectively. A grey dotted line is the maximum level of 99.97% beyond 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 represent 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 present several challenges including the volatility and trading risk, risk premia and market completeness. At the heart of term structure modelling, is the need to study both the term structure and time series simultaneously. ATSMs have been successfully used in this area as they are characterised by tractability, closed-form solutions, efficient approximation and closed-form moment conditions for empirical analysis [13] 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 parameters 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 $A_1(3)$ model is unable to play the dual role [13]. We compare the $A_1(4)$ USV with $A_1(3)$ USV.

- To determine whether estimation based on bond price only or a combination of both bond price and options data produce best results. In the absence of option price data, we test as an alternative the simulated at-the-money bond futures implied volatility, and the macro variables as sources of variation and the substitute for options 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 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 $A_1(3)$ and $A_1(4)$.

[3] propose an alternative approach to those of [17](DK) and [8] 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, [3] write these term structure dynamics in a nonlinear state space. The model parameters are then estimated by Bayesian and MCMC.

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 [13]. 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 robustness check we compare against the variation of macro signals such as oil shocks, and GARCH-based FX volatility. This we believe, also assist to ascertain 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:

$$y_t = K + LX_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \Sigma_\epsilon) \quad (26)$$

$$X_{t+1} = \mu + \Phi X_t + \eta_t, \quad \eta_t \sim \mathcal{N}(0, Q(V_t)) \quad (27)$$

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 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, thereafter, an evaluation of their forecasting performance out-of-sample. Subsequently, we provide a closer examination of the time series behavior of key the 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. For each point estimate, based on mean values there are a corresponding credible interval bounds, which appear consistent across the two models. Point estimates lie well within the interval bounds. The variation between models is expected due to differences in model structure, but the overall consistency suggests stability in estimation. We applied the discretisation value $h = \frac{1}{52}$, which produced reasonable estimates.

In addition, for a sample of risk-neutral drift parameters, which includes the physical γ_V^P and K_V^P we analysed the posterior plots for consistency.² Figure 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.

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 undelined 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	-0.3588 [-0.4171, -0.3005]	-0.0780 [-0.1147, -0.0414]
m_μ	-0.8077 [-1.1711, -0.4443]	-0.0576 [-0.0676, -0.0476]
m_θ		-0.1113 [-0.1242, -0.0984]
m_V	-0.3152 [-1.0949, 0.4645]	-0.0196 [-0.0203, -0.0189]
γ_V^P	0.0009 [[0.0009, 0.0009]	0.0020 [0.0019, 0.0021]
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]	0.0190 [0.0016, 0.0365]
σ_θ		0.0101 [0.0098, 0.0104]

² Risk-neutral parameters K_V and γ_V are not identifiable under USV, hence they are replaced by $\gamma_V^P = \gamma_V + \lambda_{V0}$, and $K_V^P = K_V - \lambda_{VV}$ [7].

Table 2. Cont.

Parameter	$A_1(3)$ USV	$\$A_1(4)$ USV
σ_V	0.0003 [-0.0007, 0.0013]	0.0001 [-0.0000, 0.0002]
\underline{V}	1.0e-06 [1.0e-06, 1.0e-06]	0.0001 [0.0001, 0.0001]
$c_{r\mu}$	-0.4622 [-0.5098, -0.4146]	-0.3086 [-0.3226, -0.2945]
$c_{r\theta}$		-0.0503 [-0.0524, -0.0482]
$c_{\mu\theta}$		-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]	-0.0951 [-0.0989, -0.0914]
$c_{\theta V}$		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]

Table 3. Cont.

Parameter	$A_1(3)$ USV	$A_1(4)$ USV
$\lambda_{\theta\mu}$		0.0211 [0.0200, 0.0221]
$\lambda_{\theta V}$		0.0193 [0.0181, 0.0204]

To gain more insight on the effectiveness of the Bayesian inference obtained through MCMC sampling, we used the python software function `arviz.summary()` of [25] 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

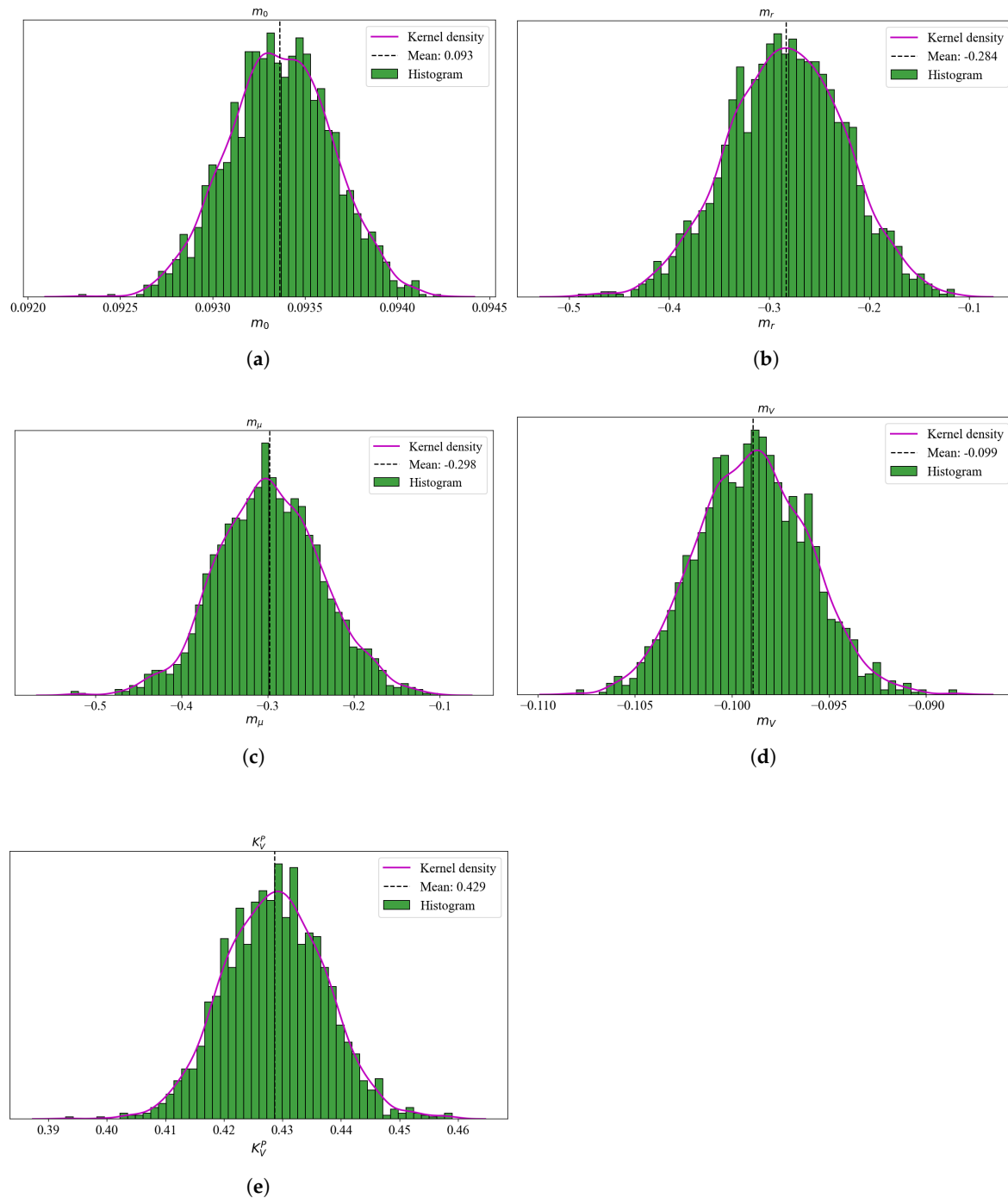


Figure 2. A sample of posterior parameter histograms for the $A_1(3)$ USV model. These are the risk-neutral drift parameters except for the γ_V^p and K_V^p . The magenta line represents a theoretical Gaussian density plot. The histograms exhibit a unimodal shape in most cases, which confirm a successful convergence and mixing.

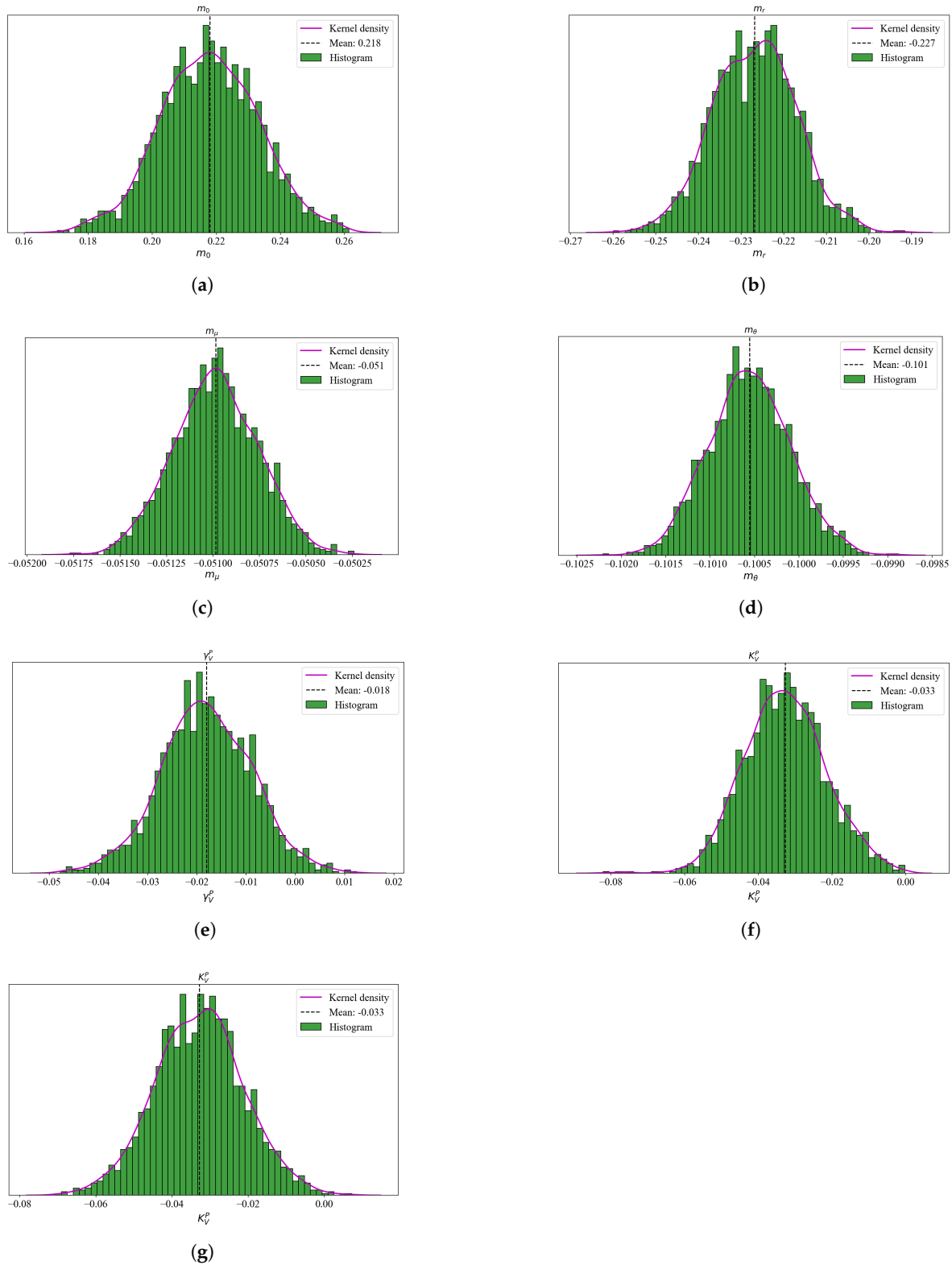


Figure 3. A sample of posterior parameter histograms for the $A_1(4)$ USV model are plotted. Specifically, these are the risk-neutral drift parameters except for the γ_V^P and K_V^P . For each histogram, a plot in magenta, represents a theoretical Gaussian density plot. The histograms exhibit a unimodal shape in most cases, which confirm a successful convergence and mixing.

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 [7] 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 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 as zero yields.

Table 6 presents the yield curve fit analysis results in terms of RMSE and bias, both in and out-sample³. The yield curve fitting performance of two models is assessed using both RMSE and forecast bias, evaluated in-sample and out-of-sample across maturities 0.25, 5, 10, 12, 20, 25, 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 suggest that $A_1(4)$ USV captures the cross-sectional variation in yields more accurately. Model $A_1(3)$ USV exhibits significantly a 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 exhibit 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 biaseness whose decision depend 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 [7], 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. 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	

³ We use the [9] 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 standardised per-point loss differentials d_i to highlight localized forecast performance differences. Each per-point z-score is defined as:

$$z_i = \frac{d_i - \bar{d}}{s_d}, \text{ where } d_i \text{ is the pointwise difference in forecast losses, } \bar{d} \text{ is the mean loss difference, and } s_d \text{ is the sample standard deviation. Significance levels per point are marked by: } ***|z| > 2.756, **|z| > 1.960, *|z| > 1.645$$

Table 6. Cont.

Maturity	$A_1(3)$ USV	Loss direction	$A_1(4)$ USV	Significance
	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 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 [7]. 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 $(Y_{30y} - Y_{0.25y})$, and the curvature as $(Y_{30y} - 2Y_{12y} + Y_{0.25y})$. 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 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 [24], 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 result, 0.03% variance explained is well below any reasonable threshold and negligible.

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. Whereas, 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 end, even though it decreases for both mid and long end. 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 end, respectively. The GARCH(1,1) displays the results that are almost similar except that suprisingly 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. Figure (a) displays the inability of $A_1(3)$ USV model volatility to track the variation of short rate volatility with 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.

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 displays a positive relationship with the model-implied volatility V_t of 0.408 for 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 existense of the tension between time series dynamics and yield curve dynamics.

Table 7. Model-implied volatility proxy for the weekly yield returns computed as a 26-weeks rolling standard deviation. Initially, the rolling statistic is computed for the 3-month (short-end) yields, then compared to the 12-year (mid-term) and 30-year (long-end). 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 speaman correlation is calculated between the model-implied volatility and these variables.

Maturities	$A_1(3)$			$A_1(4)$		
	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

7.4. Volatility Forecasting and Regression

In this section we analyse the in-sample and out-sample performance using two volatility proxies, the absolute one-week of changes in yield returns $E[|\Delta Y|]$ and realised 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 [7], 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 done in a manner consistent with the previous section, except that in the forecasting context, no data beyond time t are 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 10000 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, realised volatility is given as

$$\hat{\sigma}_{t,\tau} = \sqrt{\sum_{i=1}^N \Delta Y_{t,i,\tau}^2} \quad (28)$$

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

Traditionally, realised volatility is better estimated using high-frequency return data such as daily to increase the certainty over 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 macroeconomic 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 [26].

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 show a more 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 near-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 optimised 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 prioritise cross-sectional fit at a point in time, while volatility models emphasise 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 near term, as seen in maturities 12, 20, 25 and 30-year 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.

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-grey bound distribution of model-implied yield volatilities as determined by the point estimates. The model-fitted unconditional volatility is clearly fitting 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 (grey line) better when compared to $A_1(3)$ USV. Both plots exhibit a snake-shaped curve as discussed by [4] 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 [4] 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.

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% 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. 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	

Table 8. Cont.

Maturity	$A_1(3)$ USV	Loss Direction	$A_1(4)$ USV	Significance
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 → Fail to reject H_0 . Better model: $A_1(4)$ USV (based on lower RMSE).

⁶DM t-statistic = 2.4035, p-value = 0.0530* → Fail to reject H_0 . Better model: $A_1(4)$ USV (based on lower RMSE).

⁷DM t-statistic = 6.7912, p-value = 0.0005*** → Reject H_0 . Better model: $A_1(4)$ USV (based on lower RMSE).

⁸ DM t-statistic = 6.7912, p-value = 0.0005*** → Reject H_0 . Better model: $A_1(4)$ USV (based on lower RMSE).

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:

$$V_t = \alpha + \beta^T X_t + \gamma_i \Sigma_{m_{i,t}} + \epsilon_t \quad (29)$$

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

Table 9 presents the results of volatility regressions for the 3-months 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-end 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 exhibits a beta of -2.144 and 0.302 , respectively, suggesting that $A_1(4)$ USV is still positively

related to realised 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- than longer end.

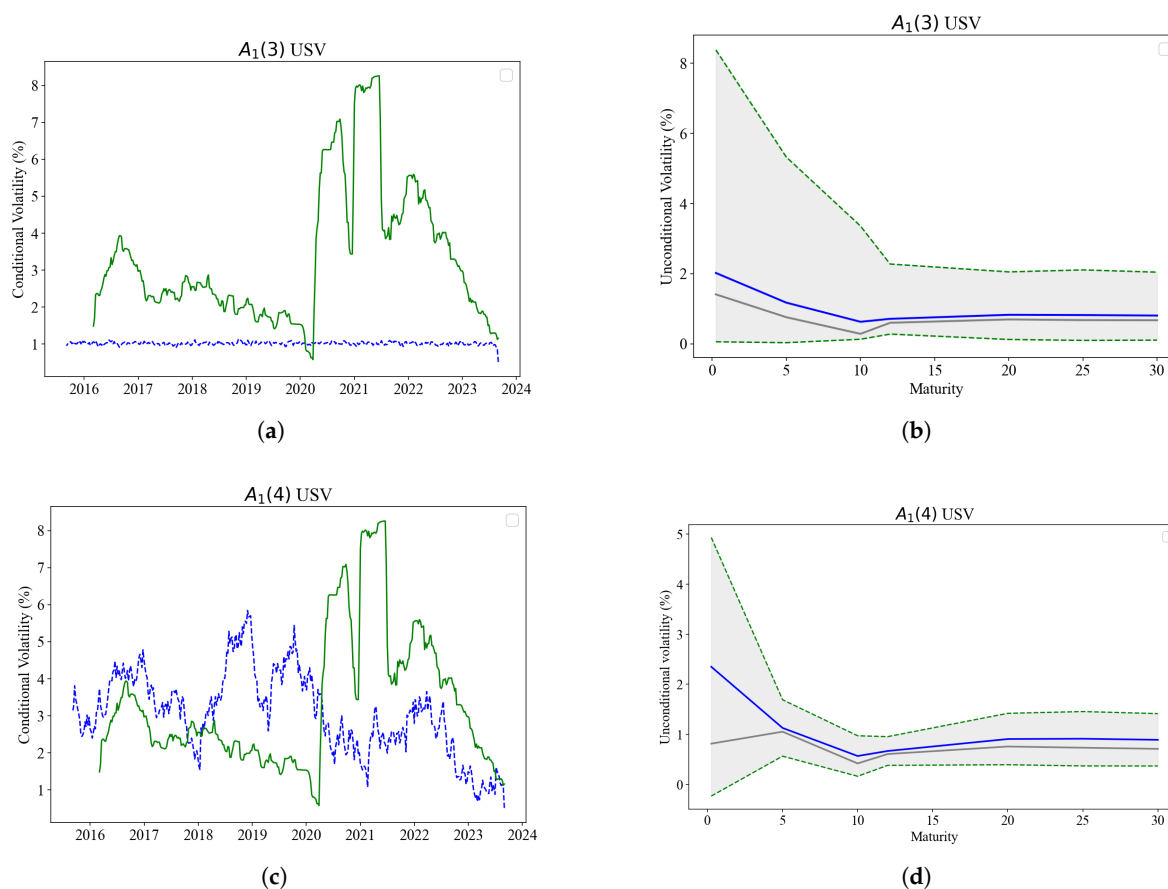


Figure 4. Conditional and unconditional volatility are plotted with the probability distributions of parameters. (a) $A_1(3)$ USV model-implied conditional volatility V_t in blue, plotted with a 26-week rolling volatility in green. (b) $A_1(3)$ USV model-implied unconditional volatility in blue and shown 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.

The purpose of this regressions 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 latent factors level, slope, and curvature are introduced, explanatory power improves substantially with R^2 increasing to 0.428 for 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. Yet, it shows limited sensitivity to yield curve shape in the long end 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.

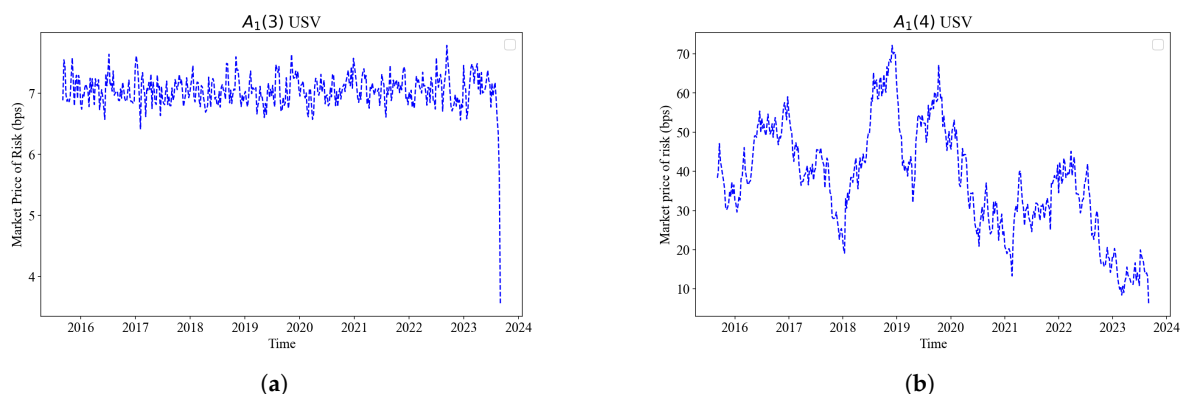


Figure 5. Market price of risk in basis points is plotted against the time series for the period 2013 - 2024. The first sub-figure (a) plots the model $A_1(3)$ USV and second (b) plots $A_1(4)$ USV both against time. They are based on the proportional relationship between $\sqrt{S_t}$ and the essentially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

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.

Table 9. Regressions for weekly realised 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 models there are two rows; 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. Top panel are the regression results for the 3-month weekly realised volatility and the bottom panel is for the 30-year weekly realised volatility.

Variable	Intercept(α)	Volatility(β)	R^2	Level	Slope	Curvature
3-Month Yield Volatilities						
GARCH(1,1)	0.202	0.395	0.240			
	[0.015]	[0.036]				
$A_1(3)$ USV	0.210	0.228	0.477	-0.025	-0.112	-0.279***
	[0.013]	[0.036]		[0.005]	[0.009]	[0.025]
$A_1(4)$ USV	1.196	-8.473	0.027			
	[0.2619]	[2.576]				
$A_1(4)$ USV	0.977	-6.876	0.428	-0.027	-0.132	-0.334
	[0.203]	[1.999]		[0.005]	[0.009]	[0.024]
$A_1(4)$ USV	0.069	0.810	0.022			
	[0.090]	[0.274]				
$A_1(4)$ USV	0.696	-1.321	0.427	-0.043	-0.173	-0.364
	[0.126]	[0.3956]		[0.007]	[0.016]	[0.025]

Table 9. Cont.

Variable	Intercept(α)	Volatility(β)	R^2	Level	Slope	Curvature
30-Year Yield Volatilities						
GARCH(1,1)	0.068	0.472	0.221			
	[0.007]	[0.045]				
$A_1(3)$ USV	0.067	0.447	0.241	-0.005	-0.005	-0.023
	[0.007]	[0.045]		[0.002]	[0.004]	[0.009]
$A_1(3)$ USV	0.350	-2.144	0.017			
	[0.085]	[0.835]				
$A_1(4)$ USV	0.317	-1.862	0.062	-0.007	-0.006	-0.032
	[0.084]	[0.8262]		[0.002]	[0.004]	[0.010]
$A_1(4)$ USV	0.038	0.302	0.035			
	[0.025]	[0.080]				
$A_1(4)$ USV	-0.293	1.393	0.221	0.010	0.037	-0.032
	[0.046]	[0.150]		[0.003]	[0.006]	[0.009]

8. Conclusion

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

Author Contributions: Conceptualization, M.M. and G.V.; methodology, G.V.; software, M.M.; validation, M.M. and G.V.; formal analysis, M.M., and G.V.; investigation, M.M and G.V.; resources, M.M and G.V.; data curation, M.M and G.V.; writing—original draft preparation, M.M; writing—review and editing, G.V.; visualization, M.M.; supervision, G.V.; project administration, G.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.

Acknowledgments: None.

Conflicts of Interest: None.

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	Generalised 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 [1] in the state variables and 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

$$\lambda_1 = \begin{bmatrix} \lambda_{r0} \\ \lambda_{\mu 0} \\ \lambda_{V0} \end{bmatrix}, \quad \lambda_2 = \begin{bmatrix} \lambda_{rr} & \lambda_{r\mu} & \lambda_{rV} \\ \lambda_{\mu r} & \lambda_{\mu\mu} & \lambda_{\mu V} \\ \lambda_{Vr} & \lambda_{V\mu} & \lambda_{VV} \end{bmatrix} \quad (\text{A.1})$$

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

$$W_t^Q = W_t^P + \int_0^t \Lambda_t dt \quad (\text{A.2})$$

By expectation and dividing by dt on both sides we find

$$\frac{1}{dt} \mathbb{E}^Q[dX_t] = \frac{1}{dt} \mathbb{E}^P[dX_t] + \Lambda_t \quad (\text{A.3})$$

Substituting the first term on the right hand side of (A.3) with (13) we obtain the following

$$\frac{1}{dt} \mathbb{E}^{\mathbb{P}}[dX_t] = \begin{bmatrix} m_0 + m_r r_t + m_\mu \mu_t^{\mathbb{Q}} + m_V V_t \\ \mu_t^{\mathbb{Q}} \\ g_V - k_V V_t \end{bmatrix} + \left(\sqrt{S_t} \lambda_1 + \sqrt{S_t^{-1}} \lambda_2 \begin{bmatrix} r_t \\ \mu_t^{\mathbb{Q}} \\ V_t \end{bmatrix} \right) \quad (\text{A.4})$$

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:

$$\Lambda_t = \Lambda_1 + \Lambda_2 \quad (\text{A.5})$$

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 get:

$$\Lambda_t = \begin{bmatrix} \lambda_{m0} + \lambda_{rr} r_t + \lambda_{r\mu} \mu_t^{\mathbb{Q}} + \lambda_{rV} V_t \\ \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu\mu} \mu_t^{\mathbb{Q}} + \lambda_{\mu V} V_t \\ \lambda_{V0} + \lambda_{Vr} r_t + \lambda_{V\mu} \mu_t^{\mathbb{Q}} + \lambda_{VV} V_t \end{bmatrix} \quad (\text{A.6})$$

Step 3: Adding the terms to obtain a physical drift

Now, summing $\mu^{\mathbb{Q}}(X_t)$ and Λ_t :

$$\frac{1}{dt} \mathbb{E}^{\mathbb{P}}[dX_t] = \begin{bmatrix} m_0 + \lambda_{r0} + (m_r + \lambda_{rr}) r_t + (m_\mu + \lambda_{r\mu}) \mu_t^{\mathbb{Q}} + (m_V + \lambda_{rV}) V_t \\ \mu_t^{\mathbb{Q}} + \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu\mu} \mu_t^{\mathbb{Q}} + \lambda_{\mu V} V_t \\ g_V + \lambda_{V0} - (k_V - \lambda_{VV}) V_t + \lambda_{Vr} r_t + \lambda_{V\mu} \mu_t^{\mathbb{Q}} \end{bmatrix} \quad (\text{A.7})$$

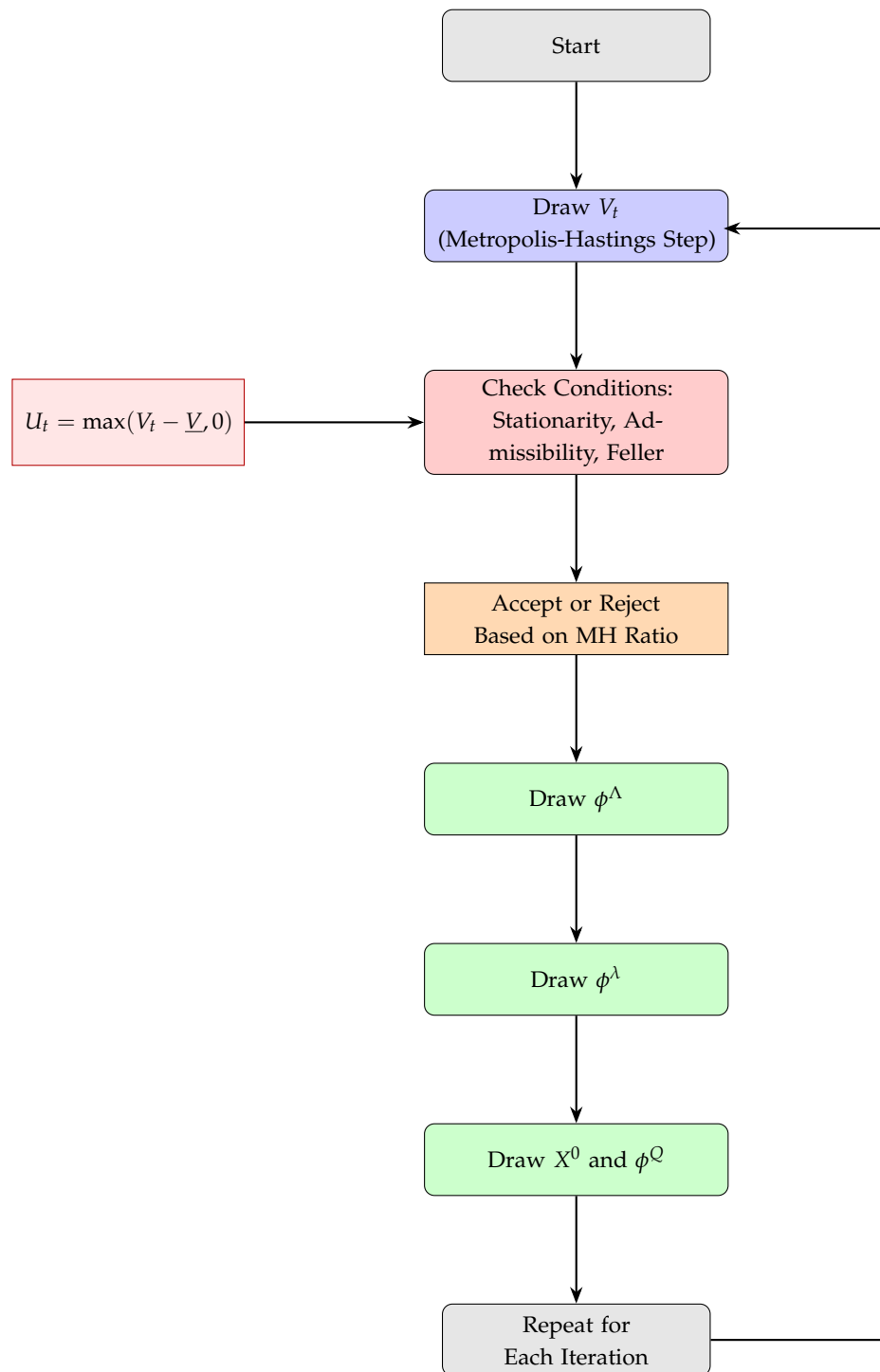
To simplify notation and group terms, define:

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

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

$$\frac{1}{dt} \mathbb{E}^{\mathbb{P}}[dX_t] = \begin{bmatrix} \tilde{m}_0 + \tilde{m}_r r_t + \tilde{m}_\mu \mu_t^{\mathbb{Q}} + \tilde{m}_V V_t \\ (1 + \lambda_{\mu\mu}) \mu_t^{\mathbb{Q}} + \lambda_{\mu 0} + \lambda_{\mu r} r_t + \lambda_{\mu V} V_t \\ \tilde{g}_V - \tilde{k}_V V_t + \lambda_{Vr} r_t + \lambda_{V\mu} \mu_t^{\mathbb{Q}} \end{bmatrix} \quad (\text{A.8})$$

Appendix B. Workflow Diagram: Yield Curve Modeling and Inference



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

Algorithm A1 MCMC Algorithm with Kalman Filtering for State-Space Models

Require: Observed data Y , initial values $\Phi = \{\phi^Q, \phi^\Lambda, \phi^\lambda\}$, X^0 , and V

Ensure: Posterior samples of Φ , X^0 , and V

- 1: **for** iteration = 1 to N **do**
- 2: **Draw** V_t^* from proposal distribution $q(V_t^* | V_t)$
- 3: **Compute MH acceptance probability:**

$$\alpha = \min\left(1, \frac{p(Y | X^0, V_t^*, \Phi)}{p(Y | X^0, V_t, \Phi)}\right)$$

- 4: **Accept or reject** V_t^* with probability α
 - 5: **if** any draw violates constraints **then**
 - 6: Reject draw and continue to next iteration
 - 7: ▷ Constraints: Stationarity, Admissibility, Feller, Non-degenerate covariance
 - 8: **end if**
 - 9: **Draw** $\phi^\Lambda \sim p(\phi^\Lambda | Y, X^0, V_t)$
 - 10: **Draw** $\phi^\lambda \sim p(\phi^\lambda | Y, X^0, V_t)$
 - 11: **Draw** X^0 and ϕ^Q :
 - 12: Conditional on V_t , dynamics of X^0 are linear-Gaussian
 - 13: Apply Kalman Filter and Smoother to sample X^0
 - 14: Sample $\phi^Q \sim p(\phi^Q | Y, X^0, V_t)$
 - 15: **Store** current samples: Φ , X^0 , V_t
 - 16: **end for**
 - 17: **return** Posterior samples of $\{\phi^Q, \phi^\Lambda, \phi^\lambda\}$, X^0 , V_t
-

References

1. Duffee, G.R. Term premia and interest rate forecasts in affine models. *The Journal of Finance* **2002**, *57*, 405–443.
2. Collin-Dufresne, P.; Goldstein, R.S. Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *The Journal of Finance* **2002**, *57*, 1685–1730.
3. Collin-Dufresne, P.; Goldstein, R.S.; Jones, C.S. Identification of maximal affine term structure models. *The Journal of Finance* **2008**, *63*, 743–795.
4. Piazzesi, M. Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*; Elsevier, 2010; pp. 691–766.
5. Riva, R. How much unspanned volatility can different shocks explain? Available at SSRN 4878175 **2024**.
6. Bikbov, R.; Chernov, M. Term structure and volatility: Lessons from the Eurodollar markets. Available at SSRN 562454 **2004**.
7. Collin-Dufresne, P.; Goldstein, R.S.; Jones, C.S. Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics* **2009**, *94*, 47–66.
8. Dai, Q.; Singleton, K.J. Specification analysis of affine term structure models. *The Journal of Finance* **2000**, *55*, 1943–1978.
9. Diebold, F.X.; Mariano, R.S. Comparing predictive accuracy. *Journal of Business & economic statistics* **2002**, *20*, 134–144.
10. Hansen, J.W. Unspanned stochastic volatility in the linear-rational square-root model: Evidence from the Treasury market. *Journal of Banking & Finance* **2025**, *171*, 107354.
11. Andreasen, M.M.; Jørgensen, K.; Meldrum, A. Bond risk premiums at the zero lower bound. *Journal of Econometrics* **2025**, *247*, 105939.
12. Andersen, T.G.; Benzoni, L. Do bonds span volatility risk in the US Treasury market? A specification test for affine term structure models. *The Journal of Finance* **2010**, *65*, 603–653.
13. Collin-Dufresne, P.; Jones, C.; Goldstein, R. Can interest rate volatility be extracted from the cross section of bond yields? An investigation of unspanned stochastic volatility **2004**.
14. Heidari, M.; Wu, L. 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 (September 10, 2002)* **2002**.

15. Ang, A.; Piazzesi, M.; Wei, M. What does the yield curve tell us about GDP growth? *Journal of Econometrics* **2006**, *131*, 359–403.
16. Cheridito, P.; Filipović, D.; Kimmel, R.L. Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics* **2007**, *83*, 123–170.
17. Duffie, D.; Kan, R. A yield-factor model of interest rates. *Mathematical finance* **1996**, *6*, 379–406.
18. Singleton, K.J. *Empirical dynamic asset pricing: model specification and econometric assessment*; Princeton University Press, 2006.
19. López-Pérez, A.; Febrero-Bande, M.; González-Manteiga, W. Estimation and specification test for diffusion models with stochastic volatility. *Statistical Papers* **2025**, *66*, 40.
20. Ait-Sahalia, Y.; Li, C.; Li, C.X. Maximum likelihood estimation of latent Markov models using closed-form approximations. *Journal of Econometrics* **2024**, *240*, 105008.
21. Elerian, O.; Chib, S.; Shephard, N. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica* **2001**, *69*, 959–993.
22. Stroud, J.R.; Müller, P.; Polson, N.G. Nonlinear state-space models with state-dependent variances. *Journal of the American Statistical Association* **2003**, *98*, 377–386.
23. Johannes, M.; Polson, N. MCMC methods for continuous-time financial econometrics. In *Handbook of financial econometrics: Applications*; Elsevier, 2010; pp. 1–72.
24. Litterman, R.B.; Scheinkman, J.; Weiss, L. Volatility and the yield curve. *The Journal of Fixed Income* **1991**, *1*, 49–53.
25. Kumar, R.; Carroll, C.; Hartikainen, A.; Martin, O. ArviZ a unified library for exploratory analysis of Bayesian models in Python. *Journal of Open Source Software* **2019**, *4*, 1143. <https://doi.org/10.21105/joss.01143>.
26. Christoffersen, P.; Diebold, F.X. Financial asset returns, market timing, and volatility dynamics. *Market Timing, and Volatility Dynamics (January 2, 2002)* **2002**.

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.