Preprints.org

South African Government Bond Yields and the Specifications of Affine Term Structure Models

Malefane Harry Molibeli [*] and Gary van Vuuren [*]

Article

South African Government Bond Yields and the Specifications of Affine Term Structure Models

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

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 and National Institute for Theoretical and Computational Sciences (NITheCS), Pretoria 0001, South Africa.
* Correspondence: mhmolibeli@gmail.com

Abstract: This study adopts the affine term structure three-factor models outlined by [1], aiming to analyse South African (SA) government bond yields across various maturities. The primary objective is to evaluate whether these models offer robust pricing capabilities—being both admissible and flexible—while capturing the conditional correlations and volatilities of yield factors specific to SA bond yields. For a model to be considered admissible, it must also demonstrate economic identification and maximal flexibility. We thus investigate the short-, medium-, and long-term dynamics of bond yields concurrently. Model estimation involves deriving joint conditional densities through the inversion of the Fourier transform applied to the characteristic function of the state variables. This enables the use of maximum likelihood estimation as an efficient method. We assume that the market prices of risk are proportional to the volatilities of the state variables. The analysis reveals negative correlations between factors. Among the models tested, the A1(3) model outperforms the A2(3) model in terms of fit, both in-sample and out-of-sample.

Keywords: affine term structure models; conditional correlation and volatility; three-factor models; market price of risk

1. Literature Review

This paper explores the behaviour of the SA interest rates in terms of historical time series and a cross-section of yields across a maturity spectrum. Inspired by the seminal work of [1], we proceed by implementing their model and its maximal counterpart. Among three options of models tested on US Treasury swaps, their $A_1(3)$ was found to perform better than other models, followed by their $A_2(3)$. However, high time-varying conditional volatility was exhibited at the expense of conditional correlations. Initially, they consider a comprehensive framework for the specification, analysis, and classification of Affine Term Structure Models (ATSMs). They provide a complete characterisation of admissible and identified ATSMs from which it is required that sufficient general conditions exist; see [2], who describe the regular affine process. They also characterise the sufficient general conditions that must be met for a process to be affine; see [3,4] among others.

ATSMs are among popular models in the vast literature on interest rates term structure and bond pricing. Few examples are the early generation consisting of a single-factor Gaussian of [5], and a square-root process by [6], extended by [7] into a multi-factor. The next generation are the correlated mixture affine models of [1,8], among others. The reason for their popularity is the ability to accommodate stochastic volatility, jumps and correlations among factors driving the asset returns; and lead to computationally tractable closed-form prices, and estimation through moment equations; see [4]. Among research problems addressed using ATSMs is the description and treatment of the co-movement of short and long-term bond yields. An affine process Y is defined as one in which a conditional mean μ and variance $\sigma\sigma'$ are affine functions of Y. The process is further defined and characterised by [2] as regular affine process, a class of time-homogeneous Markov processes. They

consider a state space $D = \mathbb{R}_+^M \times \mathbb{R}^N$, for integers $M \geq 0$ and $N \geq 0$, from which the logarithm of a characteristic function of a transitional probability $p(x_t)$ of such process is affine with respect to the initial state $x \in D$. [4] conveniently formalises it in terms of their exponential-affine Fourier (for continuous-time) and Laplace (for discrete-time) transforms. The affine relationship is defined by coefficients which are solved by a family of ordinary differential equations (ODEs). These ODEs are the essence of tractability of regular affine processes. [8] apply the ODEs as time-dependent drivers of the solution to a zero-coupon bond, provided the parameters are admissible. An inverted form of these zero-coupon bond gives rise to a yield as a state variable. They also exploit the idea of a yield-only analysis without including additional economic variables as latent factors.

[9,10] are among several authors who have approached the application of ATSMs in discrete-time although they are known to have less popularity compared to their continuous-time counterparts. Earlier models exhibited a tendency of perfectly correlated returns of bonds of all maturities, which is an unrealistic behaviour and unsuitable for hedging; see [11]. Several authors extended these one-factor Markov representation of a short-rate by introducing a range of multi-factor models with the long-run mean $\theta(t)$, and the stochastic volatility $v(t)$ of $r(t)$ that are affine functions of $(r(t), \theta(t), v(t))$ for which [1] explores several specifications. [12] endorse a parsimonious representation of the yield curve matching the time series and cross-sectional variation of bond yields through three-factor models. They develop a simple estimation approach by exploiting the exponential-affine structure of these models; see also [13] on the stochastic mean and stochastic volatility and three-factor model of the term structure of interest rates and its applications in derivatives pricing and risk management

A specification of an ATSM should be "admissible" and therefore lead to well-defined bond prices. The admissibility property is completely characterised by [2] in the "canonical" state space $D = \mathbb{R}_+^M \times \mathbb{R}^N$, with a non-negative diagonal matrix. However, this property has a problem of imposing parameter restrictions on the affine process to ensure that it is well defined. One typical scenario is the restriction of parameters to ensure that the conditional variance of a state variable remains non-negative. The requirements for admissibility become more complex as the number of state variables determining conditional variances increase; see [4]. The admissibility condition ensures that the process does not exit the domain $D \in \mathbb{R}^N$. A family of $A_M(N)$ models with a domain $D = \mathbb{R}_+^M \times \mathbb{R}^{N-M}$ are a common admissible family of models; where M factors evolve in a positive state space while $N - M$ evolve in an unrestricted space; see [14]. [1] verifies this easily through admissible $N-$factor ATSMs that are uniquely classified into $N + 1$ non-nested subfamilies.

Admissible models should also be canonical, meaning that they are economically identified, and maximally flexible; see [4]. As a result, the $A_M(N)$ benchmark ATSM models should have a canonical representation and also satisfy the non-negative and non-explosive solution of ([15]). Their drift should satisfy a Lipschitz condition, and the diffusion should satisfy the uniqueness condition of [16]; see [3]. These conditions have an effect of restricting the correlation structure of the affine diffusions. Exploiting the Gaussian and square-root form of diffusions, there still appear to be non-satisfaction of the regularity conditions of non-explosive growth and uniqueness, giving rise to need for a Feller condition;[1] see [3]. A multi-dimensional extension of a Feller condition was implemented by [8], which was found to handle the general correlated affine diffusions. The condition ensures that only positive factors enter the volatility $\sigma(y)$. This involves restrictions on the state variables that prevent the instantaneous conditional variances $S_{ii}(t)$ from becoming negative. This condition is sufficient for the existence of a unique solution to the affine SDE according to [2].

For each of the $N + 1$ subfamilies, there exists a maximal model that is econometrically plausible for all other models within this subfamily. They describe further the maximal models in relation to the $N + 1 = 4$ classification; and highlight an interaction within the family of ATSMs between

---

[1] A condition is met when $2\kappa\theta \geq \Sigma^2$; where κ represents the mean reversion speed, and θ the mean reversion rate. It ensures that the drift is sufficiently large to guarantee a positive variance.

the dependence of the conditional variance of each $Y_i(t)$ on $Y(t)$ and the admissible structure of the correlation matrix for Y. A key advantage of maximal models is that of overcoming the overidentifying restrictions that are imposed on yield curve dynamics; see [1]. The admissibility property is also confirmed by the no-arbitrage solution for a zero-coupon bond following [8].

[1] specification applied the continuous-time approach to the ATSMs which is popular to a majority of empirical literature. They explore the structural differences and relative goodness-of-fits of ATSMs. They refer to a trade-off between flexibility in modeling the conditional correlations and volatilities of the risk factors. They classify a family of $N-$factor affine into $N+1$ non-nested subfamilies of models. From their three-factor ATSMs, empirical analysis suggests that some subfamilies of ATSMs are better suited than others to explaining historical interest rate behaviour.

The focus of the research is to implement the specifications of [1] to test the pricing of zero-coupon bonds and forecasting the yield curve dynamics when using the SA bond yield. It also attempts to extract the latent factors from the yield itself, without any consideration for other economic factors; see [8]. ATSMs are proven to dominate both theoretical and empirical frameworks in term structure modelling; see [3]. A link between the cross-sectional and time series properties is made consistent by the ATSMs. Evolution of unobserved factors from the risk-neutral dynamics of the yield are proved to have both the drift and the diffusion coefficients as affine functions of such factors by the ATSMs; see [3]. Several methods of estimation are available and require mostly the knowledge of the joint conditional density of yields. In this study, we follow the estimation method of Fourier inversion for the characteristic function of a state variable, which is assumed to lead to a conditional density. This method leads to a closed-form solution where the maximum likelihood is an efficient estimator.

2. Model Establishment

We discuss the model in the context of admissibility of ATSMs. In the absence of arbitrage opportunities, a zero-coupon bond that matures at time T is priced as

$$P(t,\tau) = E_t^Q \left[e^{-\int_t^\tau r(s)ds} \right] \tag{1}$$

where:
$P(t,\tau)$ is the price of a bond at time t maturing at time τ.
t is the current or initial time at which the bond is evaluated.
τ is the maturity date, at which the bond pays its face value.
s is a continuous time variable at which the interest rate process $r(s)$ evolves.[2]
$E_t^Q[\cdot]$ denotes the conditional expectation under the risk-neutral measure Q given the information available at time t.

To obtain an $N-$factor ATSMs it is assumed that an instantaneous short-rate $r(t)$ is an affine function of a vector of N unobservable state variables $Y(t) = Y_1(t), Y_2(t), ..., Y_N(t)$, written as

$$\begin{aligned} r(t) &= \delta_0 + \sum_{i=1}^N \delta_i Y_i(t) \\ &= \delta_0 + \delta_y' Y(t) \end{aligned} \tag{2}$$

where $\delta_0 \in \mathbb{R}$ and $\delta_y \in \mathbb{R}^N$.

---

[2]   The state variable Y is a Markov process, therefore the future state y_s at time s depends only on the current state at time t and not on the history before time t. Y satisfies the Markov property; see definition 5.1 in [4].

Another assumption is that $Y(t)$ follows an affine diffusion

$$dY(t) = \mathcal{K}^Q(\theta^Q - Y(t))dt + \Sigma\sqrt{S(t)}dW^Q(t) \tag{3}$$

$\mathcal{K}^Q$ and θ^Q represent the reversion rate and central tendency (long-term mean) parameters under risk-neutral measure, respectively. $W^Q(t)$ is an $N-$ dimensional independent Brownian motion under the risk-neutral measure Q; $\mathcal{K}$, and Σ are $N \times N$ matrices, which may be asymmetric or non-diagonal. $S(t)$ is a diagonal matrix with the ith diagonal elements written as

$$[S(t)]_{ii} = \alpha_i + \beta_i'Y(t) \tag{4}$$

where $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}^N$.
The drifts in (3) and conditional variances in (4) are both affine in $Y(t)$.

[8] has the following time dependent solution to the price of a zero-coupon bond, provided that parameters are admissible.

$$P(t,\tau) = e^{A(\tau)-B(\tau)'Y(t)} \tag{5}$$

and the related yield is computed as

$$y(t,\tau) = -\frac{logP(t,\tau)}{\tau} = \frac{A(\tau)}{\tau} + \frac{B(\tau)'Y(t)}{\tau} \tag{6}$$

where $A(\tau)$ and $B(\tau)$ are coefficients whose solution satisfies the following ODEs (Ricatti equations)

$$\frac{dA(\tau)}{d\tau} = -\theta'\mathcal{K}^{Q'}B(\tau) + \frac{1}{2}\sum_{i=1}^{N}[\Sigma'B(\tau)]_i^2\alpha_i - \delta_0 \tag{7}$$

$$\frac{dA(\tau)}{d\tau} = -\theta'\mathcal{K}^{Q'}B(\tau) - \frac{1}{2}\sum_{i=1}^{N}[\Sigma'B(\tau)]_i^2\beta_i + \delta_y \tag{8}$$

A solution to these ODEs is found through numerical integration, starting from the initial conditions $A(0)=B(0)_{N\times1}$. Risk-neutral dynamics of the short rate $r(t)$ in (2) through to (4) determine this specification of the ODEs.

To use the closed-form representation of (1) in the empirical study of ATSMs, it is required that the distributions of $P(t,\tau)$ and $Y(t)$ under actual physical measure P be known. To this end, a market price of risk $\Lambda(t)$ is introduced as

$$\Lambda(t) = \sqrt{S(t)}\lambda \tag{9}$$

where λ is an $N \times 1$ vector of constants. The process $Y(t)$ under physical measure P, therefore also has an affine form [3]

$$dY(t) = \mathcal{K}(\Theta - Y(t))dt + \Sigma\sqrt{S(t)}dW(t) \tag{10}$$

Note that a superscript Q has been removed. $W(t)$ is an $N-$dimensional vector of independent Brownian motion under P, $\mathcal{K} = \mathcal{K}^Q - \Sigma\Phi$, $\Theta = \mathcal{K}^{-1}(\mathcal{K}^Q\theta^Q + \Sigma\psi)$. Φ comprises of $\lambda_i\beta_i'$ in its ith row, and ψ is an $N-$vector with $\lambda_i\alpha_i$ as its ith element.

---

[3] Continuous-time SDE are better treated in discrete form using methods such as the Euler approach. A discretised version also ensures a positive truncation for the variance; We approximate $Y(t)$ as $Y_{n+1} = Y_n + \kappa(\theta - Y_n)\Delta t + \Sigma\sqrt{S(t)}\Delta t$; where Y_n is the approximation of $Y(t)$ at time $t_n = n\Delta t$; $\Delta t \sim N(0,1)$ is a standard normal variable. Several authors discuss these discretisation schemes; see [17] : page 229 and [18] : page 177.

3. A Canonical Representation of ATSMs

According to [1], a general specification for (10) may not always lead to a positive conditional variances over a range of Y, given an arbitrary set of parameters $\psi = (\mathcal{K}, \Theta, \Sigma, \mathcal{B}, \alpha)$. However, admissibility requires that parameters restrict $S_{ii}(t)$ in (10) to be strictly positive for all i; where $\mathcal{B}$ denotes the matrix of coefficients on Y in $S_{ii}(t)$.

From (4), there is a special case where there is no admissibility problem when $\beta_i = 0$, for all i, since the instantaneous conditional variances are all constant. Outside the special case, it is necessary to impose constraints on the drift parameters $\mathcal{K}$ and Θ, and diffusion coefficients Σ and $\mathcal{B} = (\beta_1, \beta_2, \cdots, \beta_N)$. The requirements for admissibility become more restrictive as a number of state variables determining $S_{ii}(t)$ increases.

They consider a case where there are M state variables driving the instantaneous conditional variance of the $N-$vector Y, such that $M = rank(\mathcal{B})$. They further propose a set of $N+1$ benchmark models $A_M(N)$ as the most flexible econometrically identified affine DTSM on the state space $R_+^M \times R^{N-M}$; see also([2]). It is only when the admissibility conditions are met that a canonical representation may be defined.

Definition 3.1: For each M, Y_t is partitioned as $Y' = (Y'^V, Y'^D)$ where Y^V is $M \times 1$, and Y^D is $(N-M) \times 1$; where V and D represent the volatility sources and the dependent factor, respectively. The canonical representation of the benchmark model $A_M(N)$ is defined as a special case of (3) with

$$\mathcal{K} = \begin{pmatrix} K_{M \times M}^{VV} & 0_{M \times (N-M)} \\ K_{(N-M) \times m}^{DV} & K_{(N-M) \times (N-M)}^{DD} \end{pmatrix} \tag{11}$$

for $M > 0$, and $\mathcal{K}$ is either lower or upper triangular for $M = 0$.

The canonical representation of $\mathcal{K}$ is the mean-reversion matrix, with diagonal terms expected to pull the mean level to non-negativity, thus influencing positive variances. Its off-diagonal terms on the other hand reflect how different state variables influence each other, indicating potential dependencies or interactions that could affect the overall system behaviour. The matrix $\mathcal{K}$ therefore captures both the stabilising effects of the mean-reversion rates and the dynamic interplay between different state variables. In the three-factor analysis, this trade-off between non-negative variance and correlations requires a special attention. It also has an impact on the choice of M, number of state variables entering volatility and the interactions among $N = 3$ factors

$$\Theta = \begin{pmatrix} \Theta_{M \times 1}^V \\ 0_{(N-M) \times 1} \end{pmatrix} \tag{12}$$

$$\Sigma = I \tag{13}$$

$$\oslash = \begin{pmatrix} 0_{M \times 1} \\ 1_{(N-M) \times 1} \end{pmatrix} \tag{14}$$

$$\mathcal{B} = \begin{pmatrix} I_{M \times M} & V_{M \times (N-M)}^{VD} \\ 0_{(N-M) \times M} & 0_{(N-M) \times (N-M)} \end{pmatrix} \tag{15}$$

The following parameter restrictions are imposed:

$$\delta_i \geq 0, \quad M+1 \leq i \leq N \tag{16}$$

$$\mathcal{K}_i \Theta \equiv \sum_{j=1}^m \mathcal{K}_{ij} \Theta_j > 0, \quad 1 \leq i \leq M \tag{17}$$

$$\mathcal{K}_{ij} \leq 0, \quad 1 \leq j \leq M, \ j \neq i \tag{18}$$

$$\Theta_i \geq 0, \quad 1 \leq i \leq M \tag{19}$$

$$\mathcal{B}_{ij} \geq 0, \quad 1 \leq i \leq M \;\; M+1 \leq j \leq N \tag{20}$$

[1] define a subfamily $A_M(N)$; of affine DTSM as nested special cases of the *Mth* canonical model or its invariant transformation; where $M = (0, \cdots, N)$. Equivalent affine models are obtained under invariant transformations that preserve admissibility and identification and leave the observable quantities like short rate unchanged. Details on invariant transformation are discussed in Appendix A of [1].

The following issues are further noted from [1]:

The assumed structure of $\mathcal{B}$ ensures that $rank(\mathcal{B}) = M$ for the *Mth* canonical representation. To verify that M resides in $A_M(N)$, instantaneous conditional correlations among $Y^V(t)$ are zero, whereas the instantaneous correlations among $Y^D(t)$ are determined by parameters $\mathcal{B}_{ij}$ because $\Sigma = I$. Admissibility is established provided (20) holds, and that the conditional covariance matrix of Y depends only on Y^V. Zero restrictions in the upper right $M \times (N - M)$ block of $\mathcal{K}$ and the constraints in (18) and (19) ensure that Y^V is positive. Stationarity is also assured by ensuring that all the eigenvalues of $\mathcal{K}$ are strictly positive; see also Appendix B in [1].

In addition to an admissible canonical representation, in which the minimal known sufficient condition for admissibility were imposed, minimal normalisations for econometric identification are imposed to derive a "maximal" model in $A_M(N)$. A more unique class of maximal $A_M(N)$, referred to as the equivalence class of $AM_M(N)$ model is obtained by invariant transformation of the canonical representation; see Appendix A in [1].

[1] further points that the canonical representation of $A_M(N)$ models may not always be a practical way for analysing state variables in ATSMs. Often, existing literature opted for parameterising ATSMs with the riskless rate r as a state variable, resulting in "affine in r" (Ar) representation. This can be rewritten as an "affine in Y" (AY), where $r(t)$ can be expressed as an unobserved state vector $Y(t)$. As a result, a thorough specification analysis for $N-$factor ATSMs necessitates evaluating $N + 1$ non-nested, maximal models, and ensuring that a thorough understanding of the model's structure and implications is obtained.

4. The Three-Factor Affine Term Structure Models

Three-factor models are used to describe the historical behaviour of the term structure of interest rates. Traditionally, these factors are unobserved (latent) and can only be defined statistically using techniques such as principal component analysis to convey economic meaning. Popular yield-curve fitting approaches such as the dynamic Nelson-Siegel model apply PCA loadings to fit a yield curve; see [19]. These approaches appear to fit and forecast well but lack the theoretical rigour to enforce some no-arbitrage restrictions. Contrary to the yield-curve fitting approaches, the empirical approaches to the factor models such as the ATSMs are worth pursuing. They consider the maximal parameterisation through which in general the economic identification of factors can be revealed. The paper by [12] is among early works that are based on enforcing the no-arbitrage restrictions by implementing the three-factors models. They constructed a simple affine model with short term interest rate, mean rate and volatility as three factors, which are easy to estimate. They further conclude that the short rate plays and important role in yield curve modeling, following their observation that it could not be dominated by any other factor across all maturities.

[1] explores various forms of the canonical ATSMs and their maximal counterparts, as influenced by the number conditional volatility and correlation of factors. Fixing these factors into $N = 3$ gives rise to their three-factor models which posit mainly the representation of the short-rate itself, its mean rate and volatility as the three-factors. Analysis and comparisons are made of the Gaussian versus the square root diffusion forms of the models, even though the latter appear to be preferred as it imposes the non-negative variance restrictions.

Three-factor models were derived from the notation $A_M(N)$ where M is the number of state variables that enter volatility $S(t)$ according to [1]. Emphasis has been put on the trade-off between conditional volatility and correlation as a focus for the analysis of the term structure of interest rates. As previously discussed, [8] introduced a multi-dimensional Feller condition. It ensures that negative state factors do not enter the volatility $S_{ii}(t)$ by restricting correlations; see also [3]. We have previously also discussed in a similar context, the role of the mean-reversion rate matrix $\mathcal{K}$, its non-negative diagonal terms restrictions and interactions among state variables through its off-diagonal terms.

A number M of the factors that drive the process which enters a volatility $S(t)$ become the main argument on the choice of an $A_M(N)$ model, depending on the purpose of the study. [20] point out that more volatility factors result in less flexibility in allowing risk premium and correlation structure. As a result, they are in favour of a one conditional volatility factor models $A_1(N)$ such as the $A_1(3)$ by [1]. [21] also favour the $A_1(N)$ with $N = 3$ and $N = 4$ for the same purpose of allowing flexibility for risk premium and correlation. Their focus is to impose restrictions on the parameters of $A_1(3)$ such that the volatility factor $v(t)$ disappears from the bond pricing equation. In our approach, we analyse the admissibility of parameters and cross-equation restrictions that result from interactions among the factors $Y(t) = (v(t), \theta(t), r(t))$. As previously discussed, the mean-reversion rate matrix $\mathcal{K}$ has elements with either negative of positive magnitudes playing a role of ensuring that factors are pulled from entering the variance only for non-negative values, otherwise non-negative correlations are the result. This is also applicable in the case of a three-factor model.

In this study we focus on the $A_1(3)$, $A_2(3)$ models and their maximal counterparts $AM_1(3)$ and $AM_2(3)$ to determine both the fit and estimation when applied to the SA bond yield curve.

4.1. $A_1(3)$

These models are characterised by one factor Y as a source of conditional volatility. As a result, $M = 1$ gives rise to the model form of $A_1(3)$. From the original [12] BDFS model, the $A_1(3)$ according to the notation of [1] is specified as

$$
\begin{aligned}
dv(t) &= \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dB_v(t) \\
d\theta(t) &= v(\bar{\theta} - \theta(t))dt + \zeta dB_\theta(t) \\
dr(t) &= \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}dB_r(t) + \sigma_{rv}\eta\sqrt{v(t)}dB_v(t)
\end{aligned}
\tag{21}
$$

where:
$\sigma_{rv} = \frac{\rho_{rv}}{\eta\sqrt{1-\rho_{rv}^2}}$

The state variables $v(t)$, $\theta(t)$ and $r(t)$ are the stochastic volatility for $r(t)$, central tendency or long-run mean of $r(t)$ and short rate processes, respectively. The volatility affects the short rate through its volatility factor η. The coefficient κ represents the rate at which the short rate reverts to the central tendency. The stochastic volatility $v(t)$ also enters $r(t)$ and it is also instantaneously correlated with $r(t)$ as noted in the last term $\sigma_{rv}\eta\sqrt{v(t)}dB_v(t)$.

The maximal model is best suited for interpreting the parameter restrictions. As a result, [1] prefer the following model in (23) as a maximal $AM_1(3)$ which is affine in r. They determine their $A_1(3)$ by relaxing the parameters $\sigma_{\theta r}$ and $\sigma_{r\theta}$ in order to accommodate a non-zero correlation between the short rate and central tendency. All the other parameters inside the square boxes are set to zero to impose significant restrictions on the dynamics of interest rates and their volatility.

$$
r(t) = \delta_0 + \boxed{\delta_1} Y_1(t) + Y_2(t) + Y_3(t)
\tag{22}
$$

$$dv(t) = \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dB_v(t)$$

$$d\theta(t) = v(\bar{\theta} - \theta(t))dt + \sqrt{\zeta^2 + \boxed{\beta_\theta}v(t)}dB_\theta(t)$$

$$+ \boxed{\sigma_{\theta v}}\eta\sqrt{v(t)}dB_v(t) + \boxed{\sigma_{\theta r}}\sqrt{\alpha_r + v(t)}dB_r(t)$$

$$dr(t) = \boxed{\kappa_{rv}}(\bar{v} - v(t))dt + \kappa(\theta(t) - r(t))dt$$

$$+ \sqrt{\boxed{\alpha_r} + v(t)}dB_r(t) + \sigma_{rv}\eta\sqrt{v(t)}dB_v(t)$$

$$+ \boxed{\sigma_{r\theta}}\sqrt{\zeta^2 + \beta_\theta v(t)}dB_\theta(t)$$

$$(23)$$

where:

$v(t)$ serves as stochastic volatility for $r(t)$, but also enters the drift of $r(t)$, and correlated to $r(t)$ as noted in the term σ_{rv}; $\theta(t)$ is the central tendency of r and κ is the rate at which the short rate reverts to its central tendency. Appendix E in [1] describes the transformation framework from which a test for admissibility and canonical representations in $AM_1(3)$ can be achieved.

4.2. $A_2(3)$

These models are characterised by two factors of Y as a source of conditional volatility. As a result, $M = 2$ gives rise to the model form of $A_2(3)$. The [13] model is the member of this sub-class of models, and it is represented as

$$dv(t) = \mu(\bar{v} - v(t))dt + \eta\sqrt{v(t)}dW_1(t)$$

$$d\theta(t) = v(\bar{\theta} = \theta(t))dt + \zeta\sqrt{\theta(t)}dW_2(t)$$

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sqrt{v(t)}dW_3(t)$$

$$(24)$$

W_1, W_2 and W_3 are independent Brownian motions. The θ follows a square-root diffusion unlike in the case of the BDFS model. Other parameters v, η and κ are the same as in the above models. These leads us to the convenient maximal model for $A_2(3)$ which is represented as

$$r(t) = \boxed{\delta_0} + \boxed{\delta_1}Y_1(t) + Y_2(t) + Y_3(t)$$

$$(25)$$

$$dv(t) = \mu(\bar{v} - v(t))dt + \kappa_{v\theta}(\bar{\theta} - \theta(t))dt + \eta\sqrt{v(t)}dW_1(t),$$

$$d\theta(t) = v(\theta - \theta(t))dt + \kappa_{\theta v}(\bar{v} - v(t))dt + \zeta\sqrt{\theta(t)}dW_2(t),$$

$$dr(t) = \boxed{\kappa_{rv}}(\bar{v} - v(t))dt + \boxed{\kappa_{r\theta}}(\bar{\theta} - \theta(t))dt + \kappa(\boxed{\bar{r}} - r(t))dt$$

$$+ \boxed{\sigma_{rv}}\eta\sqrt{v(t)}dW_1(t) + \boxed{\sigma_{r\theta}}\zeta\sqrt{\theta(t)}dW_2(t)$$

$$+ \sqrt{\boxed{\alpha_r} + \boxed{\beta_\theta}\theta(t) + v(t)}dW_3(t).$$

$$(26)$$

[1] relaxes the restrictions on $\kappa_{\theta v}$, κ_{rv} and σ_{rv}, while other parameters within a square box are restricted to zero.

5. Estimation for Affine Models

The conditional likelihood function is not always known for affine diffusion models. However, a closed-form solution to the ATSMs provides formidable grounds to base the estimation on the conditional characteristic function of a state variable even when the functional form of the density is

not known. [22] applied the Fourier inversion to estimate the conditional-likelihood function for a state variable, which leads to an asymptotically efficient estimator of parameters. He also shows that a method-of-moments estimator works well to approximate the efficiency of maximum likelihood of ATSMs from the conditional characteristic function; see also [23].

The conditional density function f of Y_{t+1} is known up to an inverse Fourier transform of $\psi_t(\phi; \gamma)$

$$f(Y_{t+1}|Y_t) = \frac{1}{(2\pi)^N} \int_{R^N} e^{-i\phi' Y_{t+1}} \psi_t(\phi; \gamma) d\phi \tag{27}$$

The charaterisic function $\psi_t(\phi; \gamma)$ for Y_{t+1} given Y_t

$$\psi(Y_{t+1}) = E[e^{i\phi' Y_{t+1}}|Y_t] \tag{28}$$

From the Proposition 1 of [24], it can be shown that under suitable regulations (5) is the conditional characteristic function of Y_{t+1}, with $A(\tau)$ and $B(\tau)$ derived from the solution of (7) and (8) for $\tau = 1$. Therefore, the conditional characteristic function becomes

$$\psi(\tau) = e^{A(1) + B(1)' Y_{t+1}} \tag{29}$$

The log-likelihood form for (27) becomes

$$l_T(\gamma) = \frac{1}{T} \sum_{t=1}^{T} log \left\{ \frac{1}{(2\pi)^N} \int_{R^N} e^{-i\phi' Y_{t+1}} \psi_t(\phi; \gamma) d\phi \right\} \tag{30}$$

By conjecturing the parameters γ and computing the Fourier inversion, maximum likelihood can be obtained by maximising (30), to obtain a maximum likelihood estimator by characteristic function (ML-CCF); see [4].

[4] considers densities of individual columns Y_i, for $i = 1, ..., N$. A selector vector l_j is assigned an entry 1 and zero elsewhere. The density f of $y_{j,t+1} = l_j \cdot y_{t+1}$ given the entire y_t is the inverse Fourier transform of $\psi_t(\phi l_j; \gamma)$

$$f(Y_{t+1}|Y_t) = \frac{1}{(2\pi)^N} \int_{R^N} e^{-i\phi l_j' Y_{t+1}} \psi_t(\phi l_j; \gamma) d\phi \tag{31}$$

Estimation of (31) is based on one-dimensional N integrations instead of $N-$dimensional integrations.

The general method of moments (GMM) using a characteristic function is achieved by the residual

$$\epsilon(t+1, \phi; \gamma) = e^{-i\phi' Y(t+1)} - \psi(t, \phi; \gamma) \tag{32}$$

For an arbitary instrument $z(t, \phi)$; the estimator becomes

$$l_T = \frac{1}{T} \sum_t \int_{R^N} z(t, \phi) e^{-i\phi' Y_{t+1}} - \psi(t, \phi; \gamma) \tag{33}$$

The GMM approach is a beter alternative to a multi-dimensional Fourier inversion. However, as a grid of ϕ becomes finer, correlations among moments become increasingly large, leading to a singular distance matrix; see [4].

6. Data Collection

Weekly SA government yields from Thomson Reuters for all the active treasury bonds over the periods October 2013 to September 2024 with maturities 3 months, 5, 10, 12, 20, 25, and 30 years were used. Our in-sample and out-sample data were based on the periods October 2013 - Sep 2023 and October 2023 - September 2023, respectively. The out-sample will be best-suited for forecasting and validation. A summary of descriptive statistics for yields across maturities is presented in Table 1.

Mean values range between 6.6% to 10.3% exhibiting an upward slope which is also convex in shape. Recent study by [25] reported a similar behaviour for the average yields. The highest weekly standard deviation of 1.3% is observed for the 3-months maturity which is typical in the short end of the yield curve, suggesting that yields may react quickly to changes in monetary policy or market sentiment. It is followed by a drop to 0.7% and 0.6% for the 5-year and 10-year maturities, respectively. The 20-year maturity exhibits a rise in weekly standard deviation to 1.1%, which may be due to some unexpected economic event. This is also confirmed by a very high kurtosis and negative skewness of 30.69 and -3.23, respectively. After the 20-year maturity, a drop in the weekly standard deviations of 0.9% is observed. This suggests lower volatility and relative stability for the long end of the yield curve.

Table 1. Statistical summary of in-sample yields for the SA Treasury bond by maturity caption. Data were retrieved from Thomson Reuters.

	3 mths	5 yrs	10 yrs	12 yrs	20 yrs	25 yrs	30 yrs
count	418	418	418	418	418	418	418
mean	0.066	0.084	0.094	0.098	0.103	0.104	0.104
std	0.013	0.007	0.006	0.007	0.011	0.009	0.009
min	0.035	0.066	0.084	0.085	0.000	0.088	0.088
25%	0.058	0.080	0.090	0.093	0.097	0.097	0.097
50%	0.069	0.085	0.092	0.096	0.101	0.101	0.101
75%	0.074	0.089	0.096	0.101	0.110	0.111	0.110
max	0.094	0.105	0.117	0.122	0.127	0.127	0.127
skew	-0.406	-0.361	1.162	0.939	-3.232	0.561	0.583
kurtosis	-0.278	0.116	1.349	0.633	30.693	-0.536	-0.457

Table 2 presents the correlations across maturities of yields. The short end of the yield curve is characterised by weak correlations. The 3-month and 5-year terms exhibit negative correlations with their long end counterparts. This may suggest a difference and diversity in dynamics between the short and long end of the yield curve. From the 10-year maturity, higher correlations ranging from 0.608 to 0.999 are observed, suggesting stability as the yield curve approaches its long end.

Table 2. Correlation matrix of in-sample yields across maturities. Data were retrieved from Thomson Reuters.

	3 mths	5 yrs	10 yrs	12 yrs	20 yrs	25 yrs	30 yrs
3 mths	1.000	0.713	0.302	0.046	-0.111	-0.152	-0.152
5 yrs	0.713	1.000	0.613	0.269	-0.035	-0.063	-0.051
10 yrs	0.302	0.613	1.000	0.918	0.608	0.718	0.724
12 yrs	0.046	0.269	0.918	1.000	0.781	0.931	0.934
20 yrs	-0.111	-0.035	0.608	0.781	1.000	0.838	0.836
25 yrs	-0.152	-0.063	0.718	0.931	0.838	1.000	0.999
30 yrs	-0.152	-0.051	0.724	0.934	0.836	0.999	1.000

Figure 1 presents the first three principal components (PC) of the yields over maturities. They are explained by the variance of about 96.32%, which is close to the 98% empirical finding according to [26]; see also [3,19]. The first PC represents a key rate shift or level change in rates. It is the result of volatility causing rates of all maturities to fluctuate by almost the same amount. The observation is that short end is associated with high volatility and increasing rates. At mid-term around the 10-year maturity they reach a peak, followed by stability as they approach the long end of the yield curve. The second PC represents a slope which exhibits a hump in the short end associated with rate increase, followed by a drop in rates in mid-term region and stable but falling rates towards the long end. Volatility forces a fall in rates in the short end followed by a slight increase up to mid-term, 10-year maturity, thereafter, increases until it falls and stabilises towards the long end.

The behaviour of PCs is also associated with the correlations as discussed earlier, where short end is associated with weak correlations whereas in the long end of the yield curve, strong correlations were observed. These patterns are suitable for trading in swaps and correlation-based hedging strategies. Our focus being the ATSMs, we believe that these PCs are somehow closely related to the latent factors derived by the solution of coefficients $A(\tau)$ and $B(\tau)$ in (6). The three-factor models with three labels short-rate, volatility and central tendency should exhibit nearly a similar pattern to PCs; see [3].
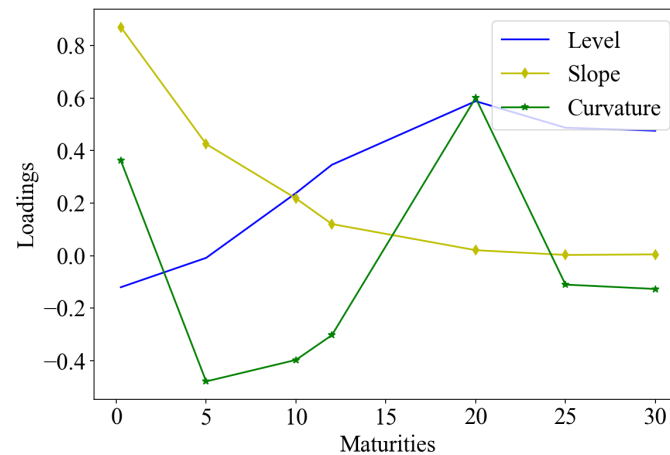
Figure 1. Loadings of the first three principal components of the yields over maturities. Data were retrieved from Thomson Reuters.

Figure 2 presents a selection of observed yields from the SA Treasury bonds plotted against maturities of up to 30 years. A spread between 5 and 20 years is also plotted and expected to represent a slope. Crossovers are observed among individual plots from time to time, indicative of either positive or negative (inverted) yield curves. Initial unobserved inputs to the three-factor simulation of $v(t), \theta(t), r(t)$ are selected from a 3-month yield, a spread between 5-year and 20-year yields, and a 30-year yield. We apply these rates together with the initially guessed parameters to simulate the state variables from (10). These are further used as inputs to the solution of ODE (7) and (8) from which coefficients $A(\tau)$ and $B(\tau)$ are obtained. Thereafter, a model-based set of zero-coupon bonds and zero-yields are obtained from (6). The selection is also guided by the observations from the PCA, suggesting that our selection is a proxy for level, slope and curvature, taking into consideration the PCA features for short end, mid-term and long end.
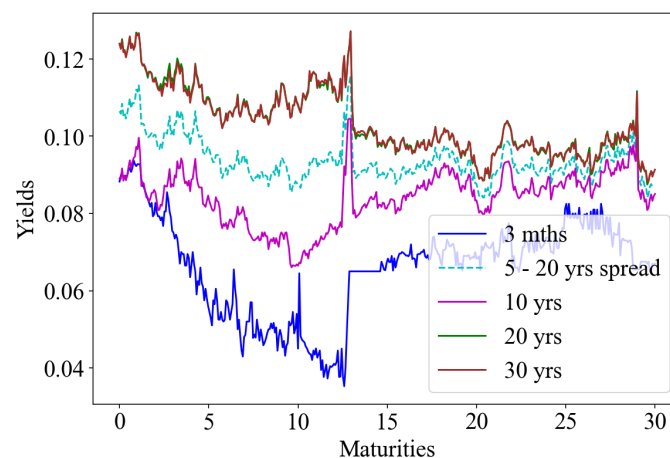

Figure 2. 3 months, 5, 10, 20, and 30 years observed yields on are plotted together with the spread between 5 and 20 years, against maturities. The 5 - 20 year spread is indicative of a slope while there are overlaps from one line to others suggesting either positive of inverted yields. Data were retrieved from Thomson Reuters.

7. Scenario Determination

Our objective is to fit the ATSMs using the SA bond yield curve to extract the latent factors by way of zero-yields. From these zero-yields, we pursue a three-factor approach to model specifications in the following scenarios.

- Evaluate the performance of model $A_1(3)$ and its maximal counterpart from in-sample. Ensure that parameters are admissible and that models meet the canonical form in-sample
- Evaluate the performance of model $A_2(3)$ and its maximal counterpart from in-sample. Ensure that parameters are admissible and that models meet the canonical form in-sample
- Estimate both models and their counterparts out-sample and evaluate their performance.

Further, we draw more insight from optimisation results and statistical analysis. In all the scenarios, we assume the market price of risk is completely affine and of the form $\Lambda(t) = \sqrt{S(t)}\lambda(t)$.

8. Model Implementation

8.1. Three-Factor Models

The first three principal components are selected as proxies to the required three factors. Their initial values are inputs to both the $A_{(}3)$ and $A_2(3)$ and their maximal counterparts. The observed market yields are therefore selected from 3-month, spread between 5-year and 20-year, and 30-year yields. The observed yields are used to calibrate parameters, which are then applied to estimate the models. Solution to SDE is initialised by the initial state vector and parameters to obtain $v(t)$, θ and $r(t)$. The market price of risk associated with each of these factors are also computed by derived equations, following a transformation process; see Appendix E in [1].

The three-factors models are further calibrated using the ML-CCF which is an efficient estimator for the log-likelihood function. The calibration process identifies optimal parameters and the maximum of the log-likelihood function. In addition, the scores of likelihood functions are used to identify the market price of risk parameters; see [22,27]. The same assumptions of completely affine price of risk are made regarding market price of risk. Individual parameters λ_v, λ_θ and λ_r are assumed to be easy to identify in the case where time-varying volatility exists. The scores of likelihood function are also applicable to identify the market price of risk as the parameters λ_v, λ_θ and λ_r are expected to be non-constant nor collinear. From the invariant transformations, [1] derives the equations to compute these parameters in Appendix E.

9. Analysis of Results

9.1. In-Sample Analysis

9.1.1. $A_1(3)$ and $AM_1(3)$ Models

In-sample data were applied to implement both models through SDE's, from which variables $v(t)$, $\theta(t)$ and $r(t)$ for stochastic volatility, central tendency and short rate, respectively were produced. Market price of risk parameters λ_v, λ_θ and λ_r, are also obtained from the results of these SDE according to the formulation from Appendix E in [1]. A linear combination of these variables result in a short rate that is fitted according to the affine function (22). The process is followed by implementation of optimisation techniques from which an efficient maximum likelihood and parameters estimates are obtained.

Table 3 presents the estimation results for both $A_1(3)$ model and its maximal counterpart $AM_1(3)$. Estimation results are based on the Fourier inversion of the characteristic function ML-CCF of the factor $Y(t)$ from (30). Variables $v(t)$, $\theta(t)$ and $r(t)$ were simulated from (23) where initial guess parameters, and the initial vector of factors $Y(i,t)$ for $i = 1,3$, at time t, were inputs. According to [1] the BDFS model is a special case of (23) with the parameters in square brackets restricted to zero. However, in their specification, parameters $\sigma_{\theta r}$ and σ_{rv} are relaxed and can be assumed take any other non-zero value.

The Nelder-Mead algorithm was used as it does not require a computation of gradients or second derivatives. This is despite its significant limitation that it cannot directly produce standard errors for the parameter estimates due to its inability to compute the Hessian matrix. To evaluate model goodness-of-fit and performance, we relied on the Chi-square (χ^2) statistics. Both AIC and BIC indicate how model complexity was penalised, taking into account both fit and the number of parameters.

Table 3. The estimators reported here are based on the log-likelihood computed from the Fourier inversion of the characteristic function of yields. Computation is based on variables $v(t)$, $\theta(t)$ and $r(t)$ of both the model-based and observed yields. Parameters in the first column are the same as those used in (23), second column are initial guesses while the 3rd and 4th columns are calibrated from the models $A_1(3)$ and $AM_1(3)$, respectively. Figures in square brackets refer to initial parameter values which are restricted to zero in terms of the model assumptions.

Parameter	Initial	Estimates	
		$A_1(3)$	$AM_1(3)$
μ	0.365	0.365	0.366
$\bar{v}$	0.015	0.015	0.008
η	0.000	0.000	0.000
$\bar{\theta}$	0.083	0.083	0.083
ζ^2	0.000	0.000	0.000
β_θ	[0.000]	0.000	0.000
σ_{rv}	4.270	4.275	4.212
$\sigma_{\theta v}$	[0.000]	0.000	0.0213
$\sigma_{\theta r}$	-0.094	-0.094	-0.089
$\sigma_{r\theta}$	-3.420	-3.509	-3.786
α_r	[0.000]	0.000	0.000
κ_{rv}	[0.000]	0.000	0.035
κ	17.400	17.421	18.006
$\bar{r}$	0.050	0.050	0.047
δ_0	0.050	0.050	0.050
δ_1	0.481	0.483	0.484
δ_2	0.666	0.666	0.669
δ_3	0.884	0.883	0.887
λ_v		0	0.000
λ_θ		0.025	0.000
λ_r		0.233	0.0442
AIC		-866.21	-866.33
BIC		-769.36	-769.48
Degree of freedom		389	393
Critical value at 95% confidence level		435.99	440.22
χ^2		57.16	87.57
P-value		1	1
Log-ikelihood function		-456.96	-457.02
Iterations		1498	1376

AIC and BIC are both lower for $AM_1(3)$ compared to $A_1(3)$ indicating $AM_1(3)$ to be slightly better at balancing model complexity while it also has slightly more parameters. However, differences between AIC and BIC values, such as -866.21 vs.-866.33 for AIC, and -769.36 vs. -769.48 for BIC are quite small, suggesting that the models might be very similar in performance. Nonetheless, $AM_1(3)$ seems to be marginally better. The model with a lower χ^2 of 57.16, which is the $A_1(3)$ appears to have a better fit than the $AM_1(3)$ with a χ^2 of 87.57.

Parameters κ_{rv}, $\sigma_{\theta v}$, α_r and β_θ are restricted to zero for the $A_1(3)$ model, whereas they are relaxed for the maximal model $AM_1(3)$. In addition, some negative correlations were introduced for $\sigma_{r\theta}$ and $\sigma_{\theta r}$. A small movement in κ_{rv} after calibration from zero to 0.035 for $AM_1(3)$ is exhibited. This suggest that the model has incorporated a weak positive relationship between short rate and volatility, which could reflect realistic market dynamics, though the effect is not large enough to dominate the model's outcomes.

For model $A_1(3)$ there is no change in $\sigma_{\theta r}$ from -0.094 to -0.094 after calibration which is interpreted as an almost unchanged relationship between the volatility of the central tendency θ and short rate r. For $AM_1(3)$ a change to 0.089 reflects a slightly weak inverse relationship between the volatility of θ and r. $AM_1(3)$ is therefore slightly less sensitive to the impact of central tendency volatility on short rate volatility.

In the case of $\sigma_{r\theta}$, $A_1(3)$ has calibrated from -3.420 to -3.509 suggesting a modest increase in the negative relationship between the volatility of the short rate r and and the central tendency θ. This implies a slightly more negative correlation between the central tendency volatility and the volatility of the short rate. For $AM_1(3)$ a change from -3.420 to -3.786 indicates a significant increase in the negative correlation between the volatility of r and θ. In summary, $A_1(3)$ exhibits a weaker relationship with volatilities while in the case of $AM_1(3)$ a slightly stronger negative correlation between short rate and central tendency is observed.

Further analysis was performed to test the residuals resulting from model-implied factors from both $A_1(3)$ and $AM_1(3)$ versus market observed yields. It is expected that the residuals are characterised by a normal distribution. Figure 3 presents qq-plots in respect of residuals for the $A_1(3)$ three factors. Despite some skewness and variations as exhibited by a red line, there seem to be less significant deviation from symmetry as can be seen from slight diversions of the red from the blue line. Therefore, all the subplots suggest a normal distribution for residuals. Based on the χ^2 test, a p-value of 1 for both models also confirm that the residuals perfectly match the expected distribution. However, there are some concerns that a p-value of 1 is unusual and unlikely and could suggest overfitting.
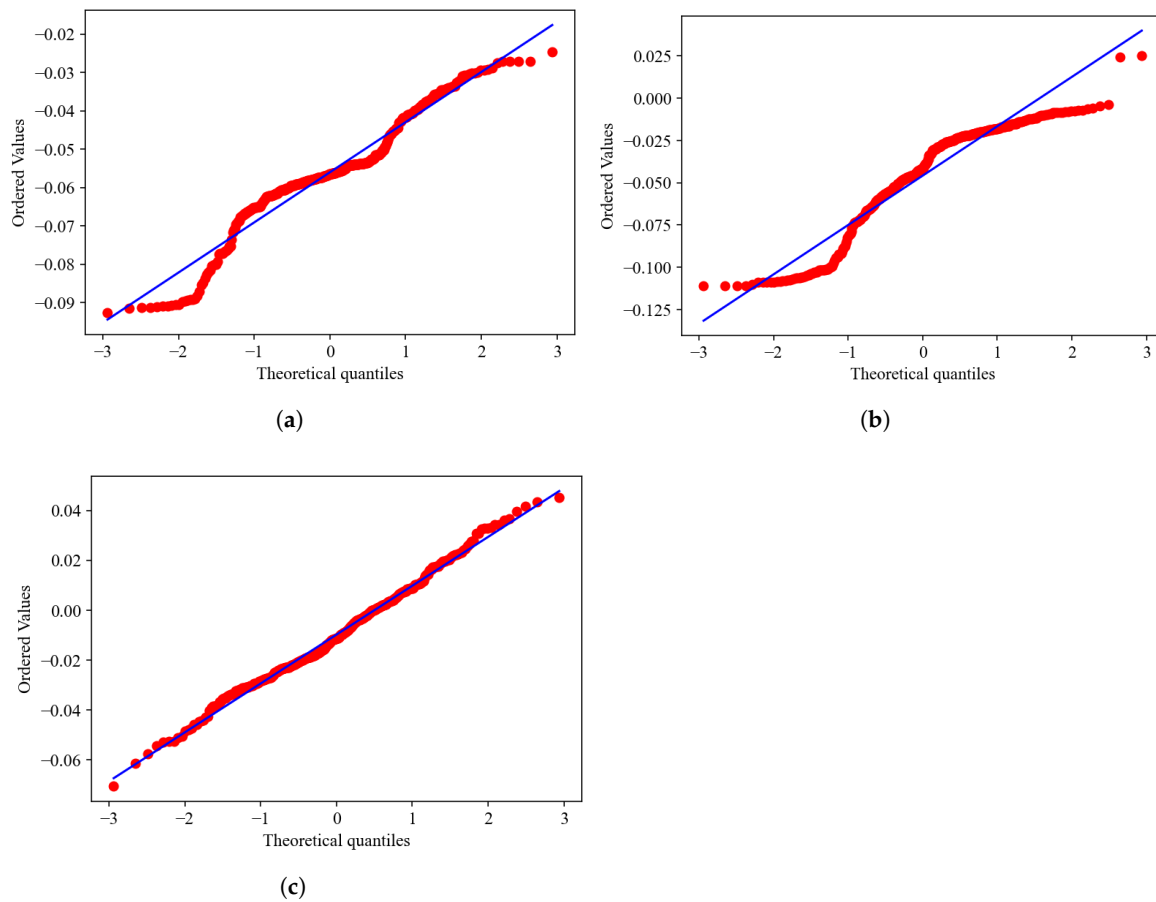
Figure 3. (**a**) Factor 1 residuals. (**b**) Factor 2 residuals. (**c**) Factor 3 residuals. QQ-plots for the residuals derived from model $A_1(3)$ computed factors and observed yields. For each factor the deviation of the red line from the blue exhibit asymmetry. The slight skewness in the residuals may not be alarming but does indicate that the models are not perfectly capturing the underlying distribution as confirmed generally by the stylised facts for asset returns. Data were retrieved from Thomson Reuters.

9.1.2. $A_2(3)$ and $AM_2(3)$ Models

Table 4 presents the results of parameter estimation for both models $A_2(3)$ and $AM_2(3)$. The second column list the initial guesses of parameters[4]. The third and fourth columns report the estimates of parameters that result from the optimisation.

Nelder-mead algorithm converged after 1574 and 1471 for models $A_2(3)$ and $AM_2(3)$, respectively. Both AIC and BIC are almost the same, suggesting that either one of them is worth a selection from a model complexity perspective. However, individual differences of -3455.43 vs -4483.99 for AIC, and -3350.51 vs -4379.07 for BIC, model $AM_2(3)$ to be slightly better at balancing model complexity while it also has slightly lower parameters. χ^2 statistics of 77.26 and 96.01 for models $A_2(3)$ and $AM_2(3)$, confirm $A_2(3)$ to be a good fit. The same applies to the p-values of 1 for each model, which may suggest a good fit, except for the issue of possible overfitting.

Parameter $\kappa_{\theta v}$ under model $A_2(3)$ moves from -33.900 to -33.989 suggests a stronger mean reversion which may lead to lower stochastic volatility and a more stable behaviour of the short rate. For model $AM_2(3)$ it moves to -12.427, indicative of weaker mean reversion, higher stochastic volatility and more variability in the short rate. For $A_2(3)$, κ_{rv} exhibit a small movement from -35.300 to -35.393

---

4   The parameters we used as initial guesses are based on the estimates according to [1]. We find these parameters to be the best starting point as they have empirically been tested to result in convergence. The same approach was also applied in the case of our $A_1(3)$ and $AM_1(3)$ models

(a difference of 0.0931). This suggest a relatively stable relationship between short rate and volatility, and a slight increase in volatility. A large negative value in κ_{rv} for $AM_2(3)$ is observed, meaning that even a small change in the short rate would have a large effect on volatility.

The covariance parameters σ_{rv} in $A_2(3)$ exhibits a move from -182.000 to -182.420 which is a modest increase in the negative relationship between the short rate r and volatility v. This implies a slight increase in negative correlation between the r and v. For $AM_2(3)$ a change from -182.000 to -133.403 indicates a significant increase in the negative correlation between the r and v. It is apparent that $A_2(3)$ exhibits a slight increase in relationship with volatilities while in the case of $AM_2(3)$ a slight increase in negative correlation between r and v is observed.

A plot for residuals suggests a not so poor a model fit despite a slight deviation from normality due to a slight skewness and asymmetry. Figure 4 presents a qq-plots for residuals based on the observed yields and the first three factors. Variation and skewness suggest a slight deviation from normality even though acceptable in terms of the stylised facts for assets and returns.
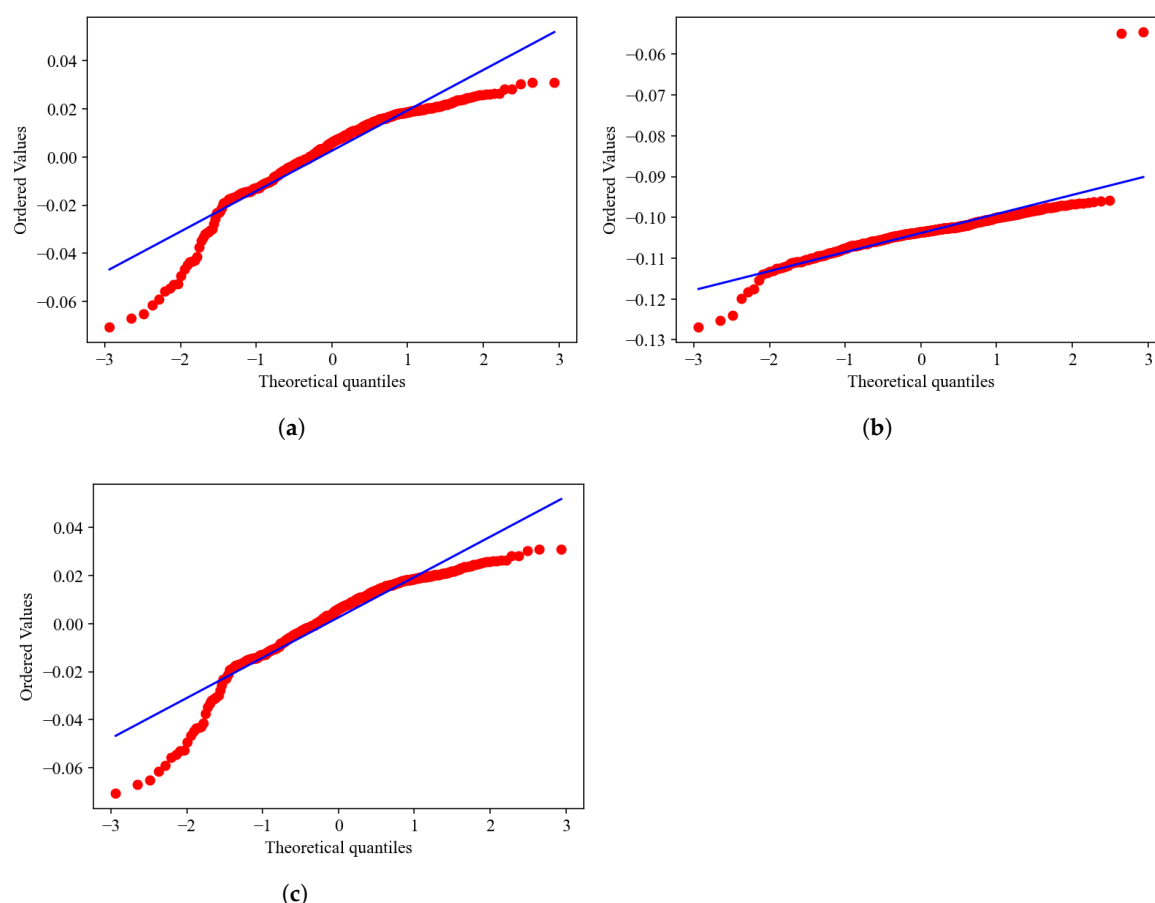

Figure 4. (**a**) Factor 1 residuals. (**b**) Factor 2 residuals. (**c**) Factor 3 residuals. QQ-plots for the residuals derived from models $A_2(3)$ computed factors and observed yields. For each factor the deviation of the red line from the blue exhibit asymmetry. The slight skewness in the residuals may not be alarming but does indicate that the models are not perfectly capturing the underlying distribution as confirmed generally by the stylised facts for asset returns. Data were retrieved from Thomson Reuters.

Table 4. The estimators reported here are based on the log-likelihood computed from the Fourier inversion of the characteristic function of yields. Computaion is based on moment labels $v(t)$, $\theta(t)$ and $r(t)$ of both the model-based and observed yields. Parameters in the first column are the same as those used in (23), second column are initial guesses while the 3rd and 4th columns are calibrated from the models $A_2(3)$ and $AM_2(3)$, respectively. Figures in square brackets refer to initial parameter values which are restricted to zero in terms of the model assumptions.

Parameter	Initial	Estimates	
		$A_2(3)$	$AM_2(3)$
μ	0.636	0.634	0.292
$\kappa_{\theta v}$	-33.900	-33.989	-12.427
κ_{rv}	-35.300	-35.393	-274.799
$\kappa_{v\theta}$	0.000	0.000	-0.002
$\kappa_{r\theta}$	[0.000]	0.000	3.561
κ	2.700	2.708	3.539
$\bar{v}$	0.000	0.000	0.000
$\bar{\theta}$	0.026	0.027	0.014
$\bar{r}$	0.026	0.026	0.053
σ_{rv}	-182.000	-182.420	-133.403
$\sigma_{r\theta}$	[0.000]	0.000	1.001
$\sigma_{\theta v}$	0.000	0.000	0.095
η	0.000	0.000	0.000
ζ^2	0.003	0.003	0.002
β_θ	[0.000]	0.000	0.000
α_r	[0.000]	0.000	0.000
δ_0	0.050	0.050	0.050
δ_1	0.496	0.495	0.496
δ_2	0.314	0.315	0.315
δ_3	0.626	0.628	0.628
λ_v		0.000	0.000
λ_θ		-2.603	0.000
λ_r		0.235	0.796
AIC:		-3455.43	-4483.99
BIC:		-3350.51	-4379.07
Degree of freedom		387	391
Critical value at 95% confidence		433.87	438.11
χ^2 statistic:		77.26	96.01
P-value:		1	1
Degrees of Freedom:		391	391
Log-likelihood function		-1753.71	-2267.99
Iterations		1574	1471

9.2. Instanteneous Short Rate

    Table 5 presents the average weekly short rate, standard deviation and annualised volatility for model-implied instantaneous rates that were computed from (22) and (25). The rates were computed for subfamilies $A_1(3)$, $A_2(3)$ and their"maximal counterparts. They are compared to the related statistics for a three-month yields (a proxy for short rate), as reported in Table 1. Model $A_2(3)$ exhibits an average short rate value of 5.1% which is closer to the weekly mean value of 6.6% for the observed three-month yields. The $A_1(3)$ model exhibits highest weekly standard deviation and annualised volatility of 2.8% and 20.3%, respectively. This is followed by 1.9% and 13.9% weekly standard

deviation and annualised volatility, respectively, exhibited by the maximal model $AM_1(3)$. The $A_1(3)$ and $AM_1(3)$ suggest higher volatility than the $A_2(3)$ and $AM_2(3)$ subfamilies.

Table 5. Average short rates, standard deviations and weekly volatilities for the observed three-month maturity are compared to the model-implied short rate averages, standard deviation and weekly volatilities for $A_1(3)$, $A_2(3)$ and their maximal counterparts. Model-implied instantaneous rates were computed from (22) and (25).

	Observed	$A_1(3)$	$AM_1(3)$	$A_2(3)$	$AM_2(3)$
mean	0.066	0.119	0.081	0.051	0.089
std deviation	0.013	0.028	0.019	0.010	0.015
weekly volatility	0.094	0.203	0.139	0.073	0.111

Figure 5 plots the model-implied instantaneous rate against maturities. Both $A_1(3)$ and $AM_1(3)$ in blue and magenta, respectively. The plots are based on estimated parameters fitted to the models. $A_2(3)$ and its maximal counterpart are plotted against maturities in Figure 6.

The model-implied instantaneous short rate functions from (22) and (25) are made up of a linear combination of factors $(Y_1(t), Y_2(t), Y_3(t))$ mapped to $v(t)$, $\theta(t)$ and $r(t)$ and representing volatility, central tendency and short rate, respectively. They exhibit an average 11.9% and 8.1% short rates for $A_1(3)$ and $AM_1(3)$, respectively. The $A_1(3)$ and $AM_1(3)$ models exhibits higher volatility than the $A_2(3)$ and $AM_2(3)$ models. As the instantaneous rate evolves, mean-reversion persists throughout the process. Both subfamilies of models suggest that a positive but stable economic growth is forecasted, exept for a higher volatility as reflected in $A_1(3)$ and $AM_1(3)$ plot.
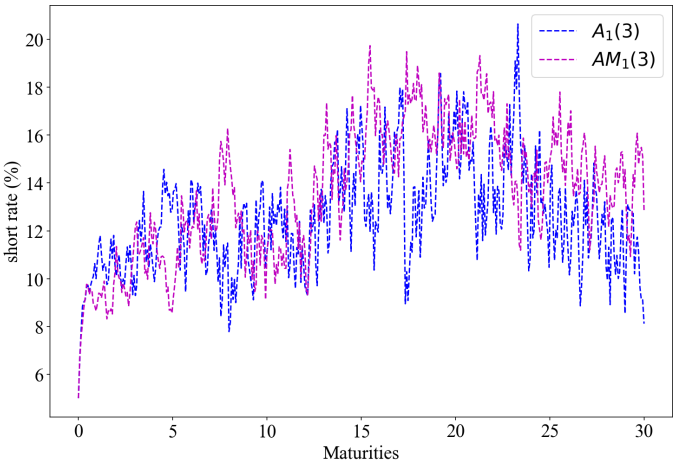

Figure 5. Model-implied instantaneous rate in percentages is plotted against maturities. $A_1(3)$ and $AM_1(3)$ appear in blue and magenta colours, respectively. Data were retrieved from Thomson Reuters.
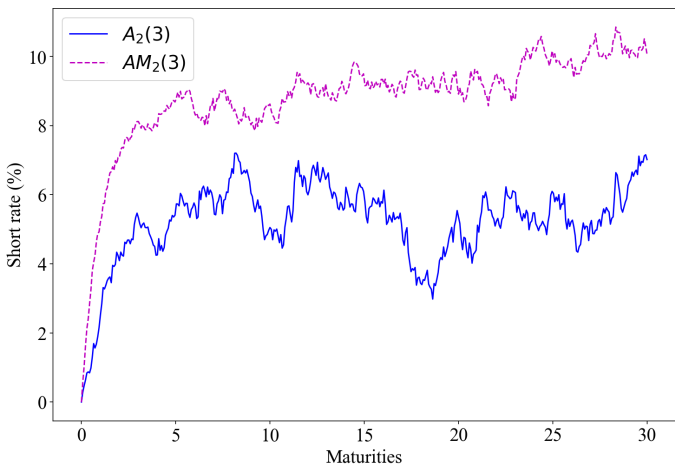
Figure 6. Model-implied instantaneous rate in percentages is plotted against maturities. $A_2(3)$ and $AM_2(3)$ appear in blue and magenta colours, respectively. Data were retrieved from Thomson Reuters.

9.3. Out-Sample Analysis

Table 6 presents a comparison of the out-sample performance for both models $A_1(3)$ and $A_2(3)$ and their maximal counterparts. Model $A_1(3)$ appears to be have a good fit according to the χ^2 statistic of 8.26 versus 9.20 of the $AM_1(3)$.These are also considered relative to their respective degrees of freedom. Vast differences between the out-sample figures when compared to in-sample might be due to the size of data series. However, the out-sample perfomance of the models fit well. Morover, the p-values exhibit a promising improvement which suggest a better fit for the models.

Table 6. A comparison is made of the $A_1(3)$ and $A_2(3)$ models together with their maximal counterparts. Performance analysis results are presented in terms of statistical results of the model estimation based on the optimised parameters obtained from in-sample analysis.

	$A_1(3)$	$AM_1(3)$	$A_2(3)$	$AM_2(3)$
AIC:	77.37	77.45	-52.71	-58.28
BIC:	120.73	120.81	-5.73	-11.31
df	16	20	14	18
χ^2	8.26	9.20	10.46	10.11
p-value	1	1	0.73	0.93
Log-likelihood	14.69	14.73	-52.35	-55.14
Iterations	1447	1444	1532	1505

10. Conclusion

The historical behaviour of the term structure of interest rates for the SA treasury bond was analysed. The purpose was to establish whether the ATSMs specifications of the three-factor models were suitable for the data. The study also considered the conditional volatilities and correlations of the state variables as they are essential components. It has therefore been essential to thoroughly examine the interactions among the state variables—stochastic volatility, central tendency, and the short rate—within the framework of a three-factor approach to the ATSM.

Model $A_1(3)$ is a suitable specification for the data, even though it suggests a weaker relationship between the short rate and its stochastic volatility. Based on the Chi-square test, the model out-performs its maximal counterpart $AM_1(3)$. $AM_1(3)$ exhibits a stronger negative correlation between short rate volatility and central tendency volatility. Similar circumstances prevail for the $A_2(3)$ model and its counterpart, although negative correlation tends to strengthen more from modest level with $A_2(3)$ to even stronger with $AM_2(3)$

As an initial guess to the parameters, we adapted the empirical estimates in [1]. This includes both their assumptions in regard to both parameters that were fixed to (zero) and those allowed to be unrestricted. The reason behind unrestricted parameters is to enable flexibility of interactions between conditional volatility and correlations. Resulting post-calibration estimates suggest a good fit for our data.

There were some limitations to the approach, which should be considered for future research. First, the inability to perform a detailed analysis on the market price of risk impact to the specifications. Second, the models were based only on the diffusion process and do not explicitly incorporate the impact of jumps.

Abbreviations

The following abbreviations are used in this manuscript:

ATSM	Affine term structure models
AIC	Akaike information criterion
BIC	Bayesian information criterion
BDFS	Balduzzi P, Das SR, Foresi S
DTSM	Dynamic term structure models
GMM	Generalised method of moments
ML-CCF	Maximum likelihood estimator by conditional characteristic function
ODE	Ordinary differential equation
PCA	Principal component analysis
SDE	Stochastic differential equation

References

1. Dai, Q.; Singleton, K.J. Specification analysis of affine term structure models. *The journal of finance* **2000**, *55*, 1943–1978.
2. Duffie, D.; Filipović, D.; Schachermayer, W. Affine processes and applications in finance. *The Annals of Applied Probability* **2003**, *13*, 984–1053.
3. Piazzesi, M. Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*; Elsevier, 2010; pp. 691–766.
4. Singleton, K.J. *Empirical dynamic asset pricing: model specification and econometric assessment*; Princeton University Press, 2006.
5. Vasicek, O. An equilibrium characterization of the term structure. *Journal of financial economics* **1977**, *5*, 177–188.

6. Cox, J.C.; Ingersoll, J.E.; Ross, S.A. An analysis of variable rate loan contracts. *The Journal of Finance* **1980**, *35*, 389–403.

7. Langetieg, T.C. A multivariate model of the term structure. *The Journal of Finance* **1980**, *35*, 71–97.

8. Duffie, D.; Kan, R. A yield-factor model of interest rates. *Mathematical finance* **1996**, *6*, 379–406.

9. Dai, Q.; Le, A.; Singleton, K.J. Discrete-time dynamic term structure models with generalized market prices of risk **2006**.

10. Darolles, S.; Gourieroux, C.; Jasiak, J. Compound autoregressive processes. *Unpublished working paper. CREST* **2001**.

11. Aït-Sahalia, Y.; Kimmel, R.L. The Econometrics of Fixed-Income Markets. *Handbook of Fixed-Income Securities* **2016**, pp. 265–281.

12. Balduzzi, P.; Das, S.R.; Foresi, S. A simple approach three-factor affine term structure models **1996**.

13. Chen, L. *Stochastic mean and stochastic volatility: a three-factor model of the term structure of interest rates and its applications in derivatives pricing and risk management*; Blackwell publishers, 1996.

14. Tebaldi, C.; Veronesi, P. Risk-Neutral Pricing: Monte Carlo Simulations. *Handbook of Fixed-Income Securities* **2016**, pp. 435–468.

15. Ikeda, N.; Watanabe, S. *Stochastic differential equations and diffusion processes*; Elsevier, 2014.

16. Yamada, T.; Watanabe, S. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University* **1971**, *11*, 155–167.

17. Hirsa, A. *Computational methods in finance*; CRC Press Boca Raton, FL, 2013; pp. 11–22.

18. Rouah, F.D. *The Heston model and its extensions in Matlab and C*; John Wiley & Sons, 2013.

19. Diebold, F.X.; Piazzesi, M.; Rudebusch, G.D. Modeling bond yields in finance and macroeconomics. *American Economic Review* **2005**, *95*, 415–420.

20. Andersen, T.G.; Benzoni, L. Can bonds hedge volatility risk in the US treasury market? A specification test for affine term structure models. Technical report, Working paper, Kellogg School of Management, Northwestern University, 2005.

21. Bikbov, R.; Chernov, M. Term structure and volatility: Lessons from the Eurodollar markets. *Available at SSRN 562454* **2004**.

22. Singleton, K.J. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* **2001**, *102*, 111–141.

23. Carrasco, M.; Chernov, M.; Ghysels, E.; Florens, J.P. Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. *Available at SSRN 338961* **2002**.

24. Duffie, D.; Pan, J.; Singleton, K. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* **2000**, *68*, 1343–1376.

25. Shu, H.C.; Chang, J.H.; Lo, T.Y. Forecasting the term structure of South African government bond yields. *Emerging Markets Finance and Trade* **2018**, *54*, 41–53.

26. Litterman, R.B.; Scheinkman, J.; Weiss, L. Volatility and the yield curve. *The Journal of Fixed Income* **1991**, *1*, 49–53.

27. Carrasco, M.; Chernov, M.; Florens, J.P.; Ghysels, E. Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions **2000**.