This is an Accepted Manuscript of an article published by Taylor & Francis in Quantitative Finance on 4 Feb 2021, available online: https://www.tandfonline.com/doi/full/10.1080/14697688.2020.1865558.

DISCRETE TIME AFFINE TERM STRUCTURE MODELS WITH SQUARED GAUSSIAN SHOCKS (DTATSM-SGS)

Dr Marco Realdon

Economics and Finance Department, Brunel University London,

Kingstone Lane, Uxbridge, London UB8 3PH, UK,

e-mail: marco.realdon@gmail.com

First version 26/3/2018; this version 21/7/2020

Abstract

The tractability of discrete time affine term structure models (DTATSM) is fully preserved when adding squared Gaussian shocks (SGS) to factor processes. SGS guarantee non-negative factors under parameter restrictions that do not affect market prices of risk. Feller conditions are not needed. Changes of measure can alter the conditional covariance of factors and yields through the flexible second order Esscher transform. Non-negative factors can be conditionally correlated under the real measure even if they are not under the risk-neutral measure. The empirical evidence from US Treasury yields shows that SGS models tend to predict yields conditional volatility, yields unconditional moments and term premia better than corresponding autoregressive gamma (AG) models.

Key words: squared Gaussian shocks, discrete time affine term structure models, stochastic volatility, second order Esscher transform, affine autoregressive gamma models.

JEL classification: G12; G13.

1 Introduction and literature

In recent years discrete time affine term structure models based on autoregressive gamma processes (DTATSM-AG) have become popular (Gourieroux and Jasiak (2006), Le, Dai and Singleton (2010), Gourieroux, Monfort, Pegoraro and Renne (2014), Creal and Wu (2015), Monfort, Pegoraro, Renne and Roussellet (2017)). DTATSM-AG have provided the discrete time counterparts of the entire family of continuous time affine models $\mathbb{A}_M(n)$ of Dai and Singleton (2000, 2002), where n is the total number of factors and M (with $0 \leq M \leq n$) is the number of volatility factors. DTATSM-AG encompass $\mathbb{A}_M(n)$ as special cases, i.e. as their continuous time limits, admit more flexible specifications of market prices of risk and are just as tractable as $\mathbb{A}_M(n)$. This paper presents other new tractable DTATSM with similar merits and new desirable features. These new DTATSM add squared Gaussian shocks (SGS) to discrete time affine factor processes and therefore we refer to them as DTATSM-SGS.

DTATSM-SGS too encompass models $\mathbb{A}_M(n)$ as continuous time limits and are just as tractable. Mild parameter restrictions, which do not constrain the market price of risk, can guarantee that factors and bond yields be non-negative. Feller conditions are not needed. The factors zero lower bound is a reflecting barrier, unlike in Sun's (1992) discrete time approximation to the CIR model.

In continuous time affine models the market price of risk only alters the drift, not the diffusion, of factors. Instead in DTATSM-AG and in DTATSM-SGS the market price of risk can alter both the conditional mean and conditional covariance of factors. Moreover in DTATSM-SGS this can be achieved through the flexible second order Esscher transform of Monfort and Pegoraro (2012), which gives DTATSM-SGS new freedom to correlate non-negative factors. The shocks to non-negative factors can be correlated under the real measure even if they are not under the risk-neutral measure.

DTATSM-SGS can be estimated through quasi-maximum likelihood, since the conditional moments of factors and yields are known in closed form. The empirical evidence from US yields shows that DTATSM-SGS tend to perform better than corresponding DTATSM-AG when predicting the conditional volatility of yields, term premia, the unconditional mean and standard deviation of yields, and when fitting linear projections of yields. This is so even when the tested SGS and AG models tend to the same continuous time limit, and **cannot be entirely explained by the** risk premia of SGS models that alter the conditional volatility and correlation of yields. Single factor versions of SGS and AG models perform similarly. SGS and AG models perform similarly also when fitting the cross sections of yields without regard to the times series of yields.

In the immense literature on dynamic term structure models, three papers are closest to this work, namely Le, Dai and Singleton (2010) (hereafter LSD), Creal and Wu (2015) (hereafter CW) and Gourieroux and Monfort (2011). LSD and CW present DTATSM-AG with flexible specifications of the market price of risk. While the models in LSD and CW are based on affine autoregressive gamma processes, the DTATSM-SGS of this paper are based on affine autoregressive processes with squared Gaussian shocks (SGS). Both processes converge to the continuous time affine diffusions of Dai and Singleton (2000, 2002). DTATSM-SGS are also related to the discrete time "bilinear" term structure model of Gourieroux and Monfort (2011). As in DTATSM-SGS, also in the bilinear model the shocks to the factors driving the short rate can be Gaussian or can be the product of Gaussian shocks, but the bilinear model does not encompass DTATSM-SGS whose continuous time limits are $\mathbb{A}_M(n)$ models, with $0 < M \leq n$.

2 Discrete Time Affine Term Structure Model with Squared Gaussian Shocks (DTATSM-SGS)

The single factor version of DTATSM-SGS is

$$\begin{aligned} r_t &= z_t \\ z_{t+1} &= z_t + \left(\mu^{\mathbb{Q}} + q^{\mathbb{Q}} z_t\right)\Delta + \sigma\sqrt{k + h z_t}\,\xi^{\mathbb{Q}}_{t+1}\sqrt{\Delta} + \psi\sigma^2\left(\xi^{\mathbb{Q}}_{t+1}\right)^2\Delta \qquad (1) \\ \xi^{\mathbb{Q}}_{t+1} &\backsim N(0,1) \end{aligned}$$

where: r_t is the continuously compounded default-free short interest rate at time t for the period $[t, t+1]$; Δ is the length of the time period; all time periods are of length Δ; z_t is the time t value of the scalar stochastic factor z; $q^{\mathbb{Q}}, \mu^{\mathbb{Q}}, \sigma, k, h, \psi$ are parameters, with $\mu^{\mathbb{Q}}, \sigma, k, h, \psi \geq 0$; $q^{\mathbb{Q}}$ and $\mu^{\mathbb{Q}}$ denote parameters under the risk-neutral pricing measure $\mathbb{Q}$; $\xi^{\mathbb{Q}}_{t+1}$ is the time $t+1$ random shock under the risk-neutral pricing measure $\mathbb{Q}$; $N(0,1)$ denotes the standard normal density with mean 0 and variance 1. When $\psi = 0$, z_{t+1} may turn negative and the DTATSM-SGS becomes the model proposed by Sun (1992). When $k = 0$ and $h \neq 0$, the DTATSM-SGS extends the discrete time approximate CIR model proposed by Sun (1992). As explained below, the parameters σ, k, h cannot all be separately identified.

It is often fitting to impose that r_{t+1} and z_{t+1} be non-negative and have zero lower bound. Throughout the paper we assume $\left(1 + \Delta q^{\mathbb{Q}}\right) > 0$, since, even if $q^{\mathbb{Q}} < 0$, Δ is small in practical applications. For example $\Delta = \frac{1}{260}$ if there are 260 trading days in one year and if Δ coincides with a trading day. Since $\mu^{\mathbb{Q}} > 0$ and $\left(1 + \Delta q^{\mathbb{Q}}\right) > 0$, the lower bound of z_{t+1} is exactly 0 if and only if

$$\psi = \frac{h}{4\left(1 + q^{\mathbb{Q}}\Delta\right)}, \quad k = \frac{h\mu^{\mathbb{Q}}\Delta}{1 + q^{\mathbb{Q}}\Delta}. \tag{2}$$

When conditions 2 hold, h is not an identifiable parameter and can be normalised by setting $h = 1$, and

$$\begin{aligned} z_{t+1} &= z_t + \left(\mu^{\mathbb{Q}} + q^{\mathbb{Q}} z_t\right)\Delta + \sigma\sqrt{\frac{\mu^{\mathbb{Q}}\Delta}{1 + q^{\mathbb{Q}}\Delta} + z_t \xi^{\mathbb{Q}}_{t+1}}\sqrt{\Delta} + \frac{\sigma^2 \left(\xi^{\mathbb{Q}}_{t+1}\right)^2 \Delta}{4\left(1 + q^{\mathbb{Q}}\Delta\right)} \qquad (3) \\ &= \left(\sqrt{z_t + \left(\mu^{\mathbb{Q}} + q^{\mathbb{Q}} z_t\right)\Delta} + \frac{\sigma \xi^{\mathbb{Q}}_{t+1}\sqrt{\Delta}}{2\sqrt{1 + q^{\mathbb{Q}}\Delta}}\right)^2. \end{aligned}$$

Therefore, when conditions 2 hold, z_{t+1} is never negative. In DTATSM-SGS the lower bound of z_{t+1} could be set at any level, but in the rest of the paper it is set at zero. In DTATSM-SGS the lower bound is a "reflecting barrier", therefore the short rate cannot persist at the zero lower bound for extended periods, unlike in the DTATSM-AG of Monfort, Pegoraro, Renne and Roussellet (2017) as applied to the Japanese yield curve. Under $\mathbb{Q}$ the time t conditional expectation and conditional variance of z_{t+1} in 3 are respectively

$$E_t^{\mathbb{Q}}\left[z_{t+1}-z_t\right]=\left(\mu^{\mathbb{Q}}+q^{\mathbb{Q}}z_t\right)\Delta+\psi\sigma^2\Delta, \qquad Var_t^{\mathbb{Q}}\left[z_{t+1}-z_t\right]=\sigma^2\left(k+z_t\right)\Delta+2\psi^2\sigma^4\Delta^2$$

and taking the continuous time limits we obtain

$$\lim_{\Delta\to 0} E_t^{\mathbb{Q}}\left[z_{t+1}-z_t\right]\to\left(\mu^{\mathbb{Q}}+q^{\mathbb{Q}}z_t+\frac{\sigma^2}{4}\right)dt, \qquad \lim_{\Delta\to 0} Var_t^{\mathbb{Q}}\left[z_{t+1}-z_t\right]\to\sigma^2 z_t dt$$

since $\Delta^2=o(\Delta)$. dt denotes the infinitesimal time increment. Therefore the continuous time limit of 3 is the following Feller-type stochastic differential equation

$$dz_t=\left(\mu^{\mathbb{Q}}+q^{\mathbb{Q}}z_t+\frac{\sigma^2}{4}\right)dt+\sigma\sqrt{z_t}dw_t^{\mathbb{Q}}. \tag{4}$$

dz_t is the time t stochastic differential of z. $dw_t^{\mathbb{Q}}$ is the time t increment of a scalar Wiener process under $\mathbb{Q}$. In dz_t the squared Gaussian shock disappears, while the new term $\frac{\sigma^2}{4}dt$ appears in the drift. When $k=\psi=0$ and $h=1$, the Feller condition $\mu^{\mathbb{Q}}\geq\frac{1}{2}\sigma^2$ guarantees that the continuous time process in 4 does not reach 0. Instead for process 1 no Feller condition was needed, but rather conditions 2. Conditions 2 restrict the parameters ψ and k, but these two parameters are additional parameters that are absent in a Feller process.

2.1 Multi-factor DTATSM-SGS

We now extend the above one factor DTATSM-SGS to n factors driving the short interest rate. Then we define $\mathbf{z}_t = (z_{1,t}, .., z_{n,t})'$. $z_{i,t}$ for $i = 1, .., n$ is the time t value of the i-th scalar factor. For $\mathbf{z}_t, \mathbf{z}_{t+1}$ belonging to some domain $D \subset \mathbb{R}^N$, assume that under the risk-neutral pricing measure $\mathbb{Q}$

$$r_t = \rho_0 + \rho_1' \mathbf{z}_t$$

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \left(\mu^{\mathbb{Q}} + \mathbf{Q}^{\mathbb{Q}} \mathbf{z}_t\right) \Delta + \mathbf{\Sigma} \sqrt{diag\left(\mathbf{k} + \mathbf{H} \mathbf{z}_t\right)} \xi_{t+1}^{\mathbb{Q}} \sqrt{\Delta} + \Delta \textstyle\sum_{i=1}^{n} \mathbf{e}_i \xi_{t+1}^{\mathbb{Q}\prime} \mathbf{\Sigma}' \mathbf{\Psi}(i) \mathbf{\Sigma} \xi_{t+1}^{\mathbb{Q}} \tag{5}$$

$$\xi_{t+1}^{\mathbb{Q}} = \left(\xi_{1,t+1}^{\mathbb{Q}}, .., \xi_{n,t+1}^{\mathbb{Q}}\right)', \quad \xi_{t+1}^{\mathbb{Q}} \backsim N\left(\mathbf{0}_{n\times1}, \mathbf{I}_n\right)$$

where: ρ_0 is a scalar parameter; ρ_1 is an $n \times 1$ vector of parameters; $\mathbf{Q}^{\mathbb{Q}}$, $\mathbf{\Psi}(i)$ and $\mathbf{\Sigma}$ are $n \times n$ matrixes of parameters, $\mu^{\mathbb{Q}}$ is an $n \times 1$ vector of parameters; $\mathbf{Q}^{\mathbb{Q}}$ and $\mu^{\mathbb{Q}}$ denote parameters under the $\mathbb{Q}$ probability measure; $diag\left(\mathbf{k} + \mathbf{H}\mathbf{z}_t\right)$ is a diagonal matrix whose diagonal is the vector $\mathbf{k} + \mathbf{H}\mathbf{z}_t$; $\mathbf{k} = (k_1, .., k_n)'$ and $\mathbf{H} = [h_{i,j}]$ is an $n \times n$ matrix whose element in the i-th row and j-th column is $h_{i,j}$; $\xi_{t+1}^{\mathbb{Q}}$ is an $n \times 1$ Gaussian vector with mean $\mathbf{0}_{n\times1}$ and covariance $\mathbf{I}_n$; $\mathbf{0}_{n\times1}$ is an $n \times 1$ vector of zeros; $\mathbf{I}_n$ is the $n \times n$ identity matrix; $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_n$; $\xi_{i,t+1}^{\mathbb{Q}}$ for $i = 1, .., n$ are time $t+1$ values of scalar Gaussian shocks under the $\mathbb{Q}$ measure; $\mathbf{\Sigma} = [\Sigma_{i,j}]$ is a $n \times n$ matrix whose element in the i-th row and j-th column is the parameter $\Sigma_{i,j}$; $\mathbf{\Psi}(i)$ for $i = 1, .., n$ denote a set of $n \times n$ matrixes; without loss in generality we assume that $\mathbf{\Psi}(i)$ are symmetric for all i; ι_n is an $n \times 1$ vector whose elements are all equal to 1. $D = \left\{\mathbf{0}_{n\times1}, \mathbb{R}_+^n\right\}$ under parameter restrictions that make process $\mathbf{z}$ non-negative. The $\mathbf{z}$ process of equation 5 is the multi-factor discrete time "affine" autoregressive process with Squared Gaussian Shocks (SGS) that is the focus of this paper. $\mathbf{z}$ is "affine" since its conditional Fourier/pricing transform is exponential affine in $\mathbf{z}_t$. An Internet Appendix presents the said transform and derives conditional and unconditional moments of the $\mathbf{z}$ process.

2.2 DTATSM-SGS and discrete time quadratic term structure models (DTQTSM)

DTATSM-SGS **and** DTQTSM are similar, but different, and this subsection compares DTATSM-SGS with DTQTSM. Under the risk-neutral pricing measure $\mathbb{Q}$, the one factor DTQTSM is

$$r_{t+1} = \overline{z}_{t+1}^2 = \left(\mu^{\mathbb{Q}}\Delta + \left(1 + \overline{q}^{\mathbb{Q}}\Delta\right)\overline{z}_t + \overline{\sigma} \cdot \xi_{t+1}^{\mathbb{Q}}\sqrt{\Delta}\right)^2 \tag{6}$$

$\overline{z}_t$ is the time t value of the scalar stochastic factor $\overline{z}$ and $\overline{q}^{\mathbb{Q}}, \overline{\sigma}$ are constants. Then under the risk-neutral pricing measure $\mathbb{Q}$, the one factor DTATSM-SGS can be written as

$$r_{t+1} = z_{t+1} = \left(\sqrt{\mu^{\mathbb{Q}}\Delta + \left(1 + q^{\mathbb{Q}}\Delta\right)z_t} + \overline{\sigma} \cdot \xi_{t+1}^{\mathbb{Q}}\sqrt{\Delta}\right)^2 \tag{7}$$

$$\overline{\sigma} = \frac{\sigma}{2\sqrt{1 + q^{\mathbb{Q}}\Delta}}.$$

If we set $z_{t+1} = \overline{z}_{t+1}^2$ and $\left(1 + \overline{q}^{\mathbb{Q}}\Delta\right) = \sqrt{(1 + q^{\mathbb{Q}}\Delta)}$, process 6 is the same as process 7 only when $\mu^{\mathbb{Q}} = 0$, whereas if $\mu^{\mathbb{Q}} \neq 0$ process 6 is not a special case of 7 and process 7 is not a special case of 6. When $r_t = \overline{z}_t^2$, bond prices **are** exponential functions of $\overline{z}_t$ and $\overline{z}_t^2$, and the conditional central moments of r_t are:

- $E_t^{\mathbb{Q}}\left[\overline{z}_{t+1}^2\right] = \left(\mu^{\mathbb{Q}}\Delta + \left(1+\overline{q}^{\mathbb{Q}}\Delta\right)\overline{z}_t\right)^2 + \overline{\sigma}^2\Delta$; therefore $E_t^{\mathbb{Q}}\left[\overline{z}_{t+1}^2\right]$ is driven by $\overline{z}_t$ and $\overline{z}_t^2$;

- $Var_t^{\mathbb{Q}}\left[\overline{z}_{t+1}^2\right] = 4\left(\mu^{\mathbb{Q}}\Delta + \left(1+\overline{q}^{\mathbb{Q}}\Delta\right)\overline{z}_t\right)^2\overline{\sigma}^2\Delta + 2\overline{\sigma}^4\Delta^2$; therefore $Var_t^{\mathbb{Q}}\left[\overline{z}_{t+1}^2\right]$ is driven by $\overline{z}_t$ and $\overline{z}_t^2$.

Instead when $r_t = z_t$, bond prices that are exponential **functions** of z_t, as shown below, and the conditional central moments of r_t are:

- $E_t^{\mathbb{Q}}\left[z_{t+1}\right] = \mu^{\mathbb{Q}}\Delta + \left(1+q^{\mathbb{Q}}\Delta\right)z_t + \overline{\sigma}^2\Delta$; therefore $E_t^{\mathbb{Q}}\left[z_{t+1}\right]$ is driven only by z_t;

- $Var_t^{\mathbb{Q}}\left[z_{t+1}\right] = 4\left(\mu^{\mathbb{Q}}\Delta + \left(1+q^{\mathbb{Q}}\Delta\right)z_t\right)\overline{\sigma}^2\Delta + 2\overline{\sigma}^4\Delta^2$; therefore $Var_t^{\mathbb{Q}}\left[z_{t+1}\right]$ is driven only by z_t.

Comparing two factor models further highlights the difference between DTATSM-SGS and DTQTSM. Under the risk-neutral pricing measure $\mathbb{Q}$, a typical two factor DTQTSM assumes

$$
\begin{aligned}
\overline{z}_{1.t+1} &= \overline{q}_{12}^{\mathbb{Q}}\overline{z}_{2,t}\Delta + \mu_1^{\mathbb{Q}}\Delta + \left(1+\overline{q}_1^{\mathbb{Q}}\Delta\right)\overline{z}_{1,t} + \overline{\sigma}_1\xi_{1,t+1}^{\mathbb{Q}}\sqrt{\Delta} \\
\overline{z}_{2,t+1} &= \mu_2^{\mathbb{Q}}\Delta + \left(1+\overline{q}_2^{\mathbb{Q}}\Delta\right)\overline{z}_{2,t} + \overline{\sigma}_2\xi_{2,t+1}^{\mathbb{Q}}\sqrt{\Delta} \\
r_{t+1} &= \overline{z}_{1,t+1}^2 = \left(\overline{q}_{12}^{\mathbb{Q}}\overline{z}_{2,t}\Delta + \mu_1^{\mathbb{Q}}\Delta + \left(1+\overline{q}_1^{\mathbb{Q}}\Delta\right)\overline{z}_{1,t} + \overline{\sigma}_1\xi_{1,t+1}^{\mathbb{Q}}\sqrt{\Delta}\right)^2
\end{aligned}
$$

and a corresponding two factor DTATSM-SGS with non-negative factors assumes

$$z_{1,t+1} = \left(\sqrt{q_{12}^{\mathbb{Q}} z_{2,t} \Delta + \mu_1^{\mathbb{Q}} \Delta + \left(1 + q_1^{\mathbb{Q}} \Delta\right) z_{1,t}} + \overline{\sigma}_1 \xi_{1,t+1}^{\mathbb{Q}} \sqrt{\Delta} \right)^2$$

$$z_{2,t+1} = \left(\sqrt{\mu_2^{\mathbb{Q}} \Delta + \left(1 + q_2^{\mathbb{Q}} \Delta\right) z_{2,t}} + \overline{\sigma}_2 \xi_{2,t+1}^{\mathbb{Q}} \sqrt{\Delta} \right)^2$$

$$r_{t+1} = z_{1,t+1}.$$

$q_{12}^{\mathbb{Q}}, \overline{q}_{12}^{\mathbb{Q}}, \mu_1^{\mathbb{Q}}, \mu_2^{\mathbb{Q}}, \overline{q}_1^{\mathbb{Q}}, \overline{q}_2^{\mathbb{Q}}, q_1^{\mathbb{Q}}, q_2^{\mathbb{Q}}, \overline{\sigma}_1, \overline{\sigma}_2$ are parameters and $\xi_{1,t+1}^{\mathbb{Q}}, \xi_{2,t+1}^{\mathbb{Q}} \backsim N(0,1)$ are independent standard Gaussian shocks under $\mathbb{Q}$. Note that $z_{1,t+1}$ still differs from $\overline{z}_{1,t+1}^2$, even if $\mu_1^{\mathbb{Q}} = 0$ and/or $\overline{q}_1^{\mathbb{Q}}, q_1^{\mathbb{Q}} = 0$. **Therefore also** multi-factor DTATSM-SGS with non-negative factors differ from multi-factor DTQTSM.

The difference in the empirical performance of DTATSM-SGS and DTQTSM can be expected to be similar to the difference in performance reported in the literature for continuous time affine and quadratic models. For example Ahn, Dittmar and Gallant (2002) compare continuous time affine models, which are the continuous time limits of DTATSM-SGS, and continuous time quadratic models, which are the continuous time limits of DTQTSM. They report the better performance of quadratic models, partly because factors can be freely correlated only in quadratic models. Leippold and Wu (2003) confirm that "factor interactions" in quadratic models are needed to match the dynamics of bond yields.

Quadratic models also have the merit of not needing non-negativity constraints on factors. Instead affine models that feature non-negative factors, including DTATSM-SGS, are hampered by the constraint

that factors be non-negative. In-sample this constraint typically consists in a penalty to the likelihood function that is triggered when non-negative factors do turn negative. Out-of-sample it is not even obvious how to impose such non-negativity constraint on the non-negative factors.

Despite these merits, quadratic models are less tractable than affine models: first, perfectly observed yields are difficult to "invert" to determine the latent factors; second, bond prices are computed less quickly, especially when the number of factors is four or more; third, computing and re-computing interest rate swap prices as factors change over time is much quicker with affine models than with quadratic models, if the factors are three or more.

3 Discount bond prices

Let $P_{t,m}$ denote the price of a discount bond at time t with maturity at $(t+m)$ and face value 1, so that $P_{t+m,0}=1$. Under the assumption of equation 5 this section shows that $P_{t,m}=\exp\left(A_m+\mathbf{B}_m'\mathbf{z}_t\right)$, where A_m is a scalar that only depends on m and $\mathbf{B}_m=(B_{1,m},..,B_{n,m})'$ is an $n\times 1$ vector whose elements are functions of m. To rule out arbitrage we impose

$$\exp\left(A_m+\mathbf{B}_m'\mathbf{z}_t\right)=E_t^{\mathbb{Q}}\left[\exp\left(-r_t\Delta+A_{m-1}+\mathbf{B}_{m-1}'\mathbf{z}_{t+1}\right)\right]$$

$$A_0=0,\quad \mathbf{B}_0=\mathbf{0}_{n\times 1}$$

where $E_t^{\mathbb{Q}}[..]$ denotes time t conditional expectation under the risk-neutral pricing measure $\mathbb{Q}$. $\mathbf{0}_{n\times 1}$ is an $n \times 1$ vector of zeros. To determine A_m and $\mathbf{B}_m$ we take logs of the last equation and obtain

$$A_m + \mathbf{B}_m'\mathbf{z}_t = -\left(\rho_0 + \rho_1'\mathbf{z}_t\right)\Delta + A_{m-1} + \mathbf{B}_{m-1}'\left(\mathbf{z}_t + \left(\mu^{\mathbb{Q}} + \mathbf{Q}^{\mathbb{Q}}\mathbf{z}_t\right)\Delta\right) +$$
$$+\ln E_t^{\mathbb{Q}}\left[\exp\left(\mathbf{B}_{m-1}'\left(\boldsymbol{\Sigma}\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\xi_{t+1}^{\mathbb{Q}}\sqrt{\Delta} + \Delta\textstyle\sum_{i=1}^{n}\mathbf{e}_i\xi_{t+1}^{\mathbb{Q}\prime}\boldsymbol{\Sigma}'\boldsymbol{\Psi}\left(i\right)\boldsymbol{\Sigma}\xi_{t+1}^{\mathbb{Q}}\right)\right)\right]$$

$$\ln\left(E_t^{\mathbb{Q}}\left[e^{\mathbf{F}'\xi_{t+1}^{\mathbb{Q}} + \xi_{t+1}^{\mathbb{Q}\prime}\boldsymbol{\Sigma}'\cdot\mathbf{C}_{m-1}\cdot\boldsymbol{\Sigma}\xi_{t+1}^{\mathbb{Q}}}\right]\right) = \ln\left(abs\left(|\gamma|\right)\right) + \frac{1}{2}\mathbf{F}'\gamma\gamma'\mathbf{F}$$
$$\mathbf{F}' = \mathbf{B}_{m-1}'\boldsymbol{\Sigma}\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\sqrt{\Delta}, \qquad \gamma = \left(\mathbf{I}_n - 2\cdot\boldsymbol{\Sigma}'\mathbf{C}_{m-1}\boldsymbol{\Sigma}\right)^{-1/2}, \qquad \mathbf{C}_{m-1} = \sum_{i=1}^{n} B_{i,m-1}\boldsymbol{\Psi}\left(i\right)\Delta$$
$$\xi_{\mathbf{t+1}}^{\mathbb{Q}\prime}\boldsymbol{\Sigma}'\mathbf{C}_{m-1}\boldsymbol{\Sigma}\xi_{\mathbf{t+1}}^{\mathbb{Q}} = \mathbf{B}_{m-1}'\cdot\Delta\textstyle\sum_{i=1}^{n}\mathbf{e}_i\xi_{t+1}^{\mathbb{Q}\prime}\boldsymbol{\Sigma}'\boldsymbol{\Psi}\left(i\right)\boldsymbol{\Sigma}\xi_{t+1}^{\mathbb{Q}}.$$

$\ln\left(abs\left(|\gamma|\right)\right)$ is the natural logarithm of the absolute value of the determinant of the matrix γ. It follows that

$$A_m + \mathbf{B}_m'\mathbf{z}_t = -\left(\rho_0 + \rho_1'\mathbf{z}_t\right)\Delta + A_{m-1} + \mathbf{B}_{m-1}'\left(\mathbf{z}_t + \left(\mu^{\mathbb{Q}} + \mathbf{Q}^{\mathbb{Q}}\mathbf{z}_t\right)\Delta\right) + \ln\left(abs\left(|\gamma|\right)\right) + \frac{1}{2}\mathbf{F}'\gamma\gamma'\mathbf{F} \tag{8}$$

$$\mathbf{F}'\gamma\gamma'\mathbf{F} = \mathbf{B}_{m-1}'\boldsymbol{\Sigma}\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\cdot\left(\mathbf{I}_n - 2\cdot\boldsymbol{\Sigma}'\mathbf{C}_{m-1}\boldsymbol{\Sigma}\right)^{-1}\cdot\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\boldsymbol{\Sigma}'\mathbf{B}_{m-1}\Delta. \tag{9}$$

As $\boldsymbol{\Psi}(i)$ is symmetric so is $\mathbf{C}_{m-1}$. We also assume that all the eigenvalues of $\boldsymbol{\Psi}(i)$ are non-negative. Then γ is normally real valued, since normally $B_{i,m-1} < 0$ for all i and m. However it is sometimes possible that under some parameter values $B_{i,m-1} > 0$, which may make γ complex valued if the time step Δ is large enough. Therefore in estimating the model we impose that parameters be such that γ is real valued. A related issue concerns the moment generating function. For example in the one factor model $E_t^{\mathbb{Q}}\left[\exp\left(z_{t+1}\right)\right]$ is well defined only if $1-2\sigma^2\psi\Delta>0$. This condition, as we impose $h=1$ and $\psi=\frac{1}{4(1+q^{\mathbb{Q}}\Delta)}$ as above, becomes $\frac{\Delta\sigma^2}{1+q^{\mathbb{Q}}\Delta}\frac{1}{2}\leq 1$ and is violated only when σ is extremely high. To ensure that this condition is not violated, we can either shrink the time step Δ or constrain the parameters of the $\mathbf{z}$ process under $\mathbb{Q}$. A similar issue affects also discrete time quadratic term structure models.

In what follows we concentrate on the case where $\rho_0=0$, $\rho_1=\iota_n$ and on conditions that guarantee $r_t=\iota_n'\mathbf{z}_t\geq 0$, where ι_n is an $n\times 1$ vector whose elements are all equal to 1. For all types of DTATSM-SGS $\mathbf{Q}^{\mathbb{Q}}$ must be of full rank. As discussed in the Internet Appendix, additional restrictions to identify the parameters of $\mathbf{Q}^{\mathbb{Q}}$ are not needed. We next consider further restrictions to identify and normalise the parameters of the $\mathbf{z}$ process under the $\mathbb{Q}$ measure for DTATSM-SGS families $\mathbb{A}_n^{sgs}(n)$, $\mathbb{A}_M^{sgs}(n)$, whose respective continuous time limits are the model families $\mathbb{A}_n(n)$, $\mathbb{A}_M(n)$ of Dai and Singleton (2000,2002).

3.1 $\mathbb{A}_n^{sgs}(n)$ models

Let $\mathbf{H}^{(i)}$ be the i-th row of $\mathbf{H}$. When $\sqrt{k_i+\mathbf{H}^{(i)}\mathbf{z}_t}\neq\sqrt{k_j+\mathbf{H}^{(j)}\mathbf{z}_t}$ for $i\neq j$, to retain exponential affine solutions for bond prices and transforms, γ needs to be diagonal, which entails that also $\mathbf{\Sigma}$ and $\mathbf{\Psi}(i)$ for all i be diagonal. In such case the continuous time limit of the resulting DTATSM-SGS belongs to the $\mathbb{A}_n(n)$ models of Dai and Singleton (2000, 2002). For $\mathbb{A}_n^{sgs}(n)$ models we set:

1) $\mathbf{\Sigma}=diag(\sigma)$ with $\sigma=(\sigma_1,..,\sigma_n)'$; therefore $\mathbf{\Sigma}$ is the diagonal matrix with i-th diagonal element σ_i;

2) $\mathbf{\Psi}(i) = \mathbf{e}_i \cdot \mathbf{e}_i' \cdot \psi_i$; ψ_i is a scalar parameter; $\mathbf{e}_i$ is the i-th column of $\mathbf{I}_n$; the matrix $\mathbf{e}_i \cdot \mathbf{e}_i' \cdot \psi_i$ is an $n \times n$ diagonal matrix with all elements equal to 0, except for the i-th diagonal element, which is ψ_i.

Assumptions 1) and 2) imply that

$$\gamma = \left(\mathbf{I}_n - 2\Delta \cdot diag\left(\sigma_i^2 \psi_i B_{i,m-1}\right)\right)^{-1/2}$$

where $diag\left(\sigma_i^2 \psi_i B_{i,m-1}\right)$ is an $n \times n$ diagonal matrix with i-th diagonal element $\sigma_i^2 \psi_i B_{i,m-1}$. Under assumptions 1) and 2) models $\mathbb{A}_n^{sgs}(n)$ are such that

$$\mathbf{F}'\gamma\gamma'\mathbf{F} = \Delta \cdot \sum_{i=1}^{n} \left(B_{i,m-1} \cdot \sigma_i \cdot \gamma_{i,i}\right)^2 \cdot \left(k_i + \mathbf{H}^{(i)} \cdot \mathbf{z}_t\right)$$

$$A_m = -\rho_0\Delta + A_{m-1} + \mathbf{B}_{m-1}'\mu^{\mathbb{Q}}\Delta + \ln\left(abs\left(|\gamma|\right)\right) + \frac{1}{2}\Delta\sum_{i=1}^{n} \frac{B_{i,m-1}^2 \cdot \sigma_i^2 \cdot k_i}{1 - 2 \cdot \sigma_i^2 \cdot \Delta \cdot B_{i,m-1} \cdot \psi_i}$$

$$\mathbf{B}_m' = -\rho_1'\Delta + \mathbf{B}_{m-1}'\left(\mathbf{I}_n + \mathbf{Q}^{\mathbb{Q}}\Delta\right) + \frac{1}{2}\Delta\sum_{i=1}^{n} \frac{B_{i,m-1}^2 \cdot \sigma_i^2 \cdot \mathbf{H}^{(i)}}{1 - 2 \cdot \sigma_i^2 \cdot \Delta \cdot B_{i,m-1} \cdot \psi_i}.$$

$\gamma_{i,i}$ is the i-th diagonal element of γ. Moreover we impose that for $i = 1, .., n$

$$\mathbf{H}^{(i)} \cdot \mathbf{z}_t = z_{i,t} + \mathbf{Q}^{(i)\mathbb{Q}}\mathbf{z}_t\Delta, \qquad \psi_i = \frac{h_{i,i}}{4\left(1 + q_{i,i}^{\mathbb{Q}}\Delta\right)}, \quad k_i = \frac{\mu_i^{\mathbb{Q}}\Delta h_{i,i}}{1 + q_{i,i}^{\mathbb{Q}}\Delta} \tag{10}$$

$$\mu^{\mathbb{Q}} \geq \mathbf{0}_{n\times 1}, \qquad q_{i,j\neq i}^{\mathbb{Q}} \geq 0 \text{ for } i, j = 1, .., n.$$

$q_{i,i}^{\mathbb{Q}}$ is the i-th diagonal element of the matrix $\mathbf{Q}^{\mathbb{Q}}$. $\mathbf{Q}^{(i)\mathbb{Q}}$ is the i-th row of $\mathbf{Q}^{\mathbb{Q}}$. $h_{i,i}$ is the i-th diagonal element of $\mathbf{H}$. According to conditions 10, the off-diagonal elements of $\mathbf{Q}^{\mathbb{Q}}$, namely $q_{i,j\neq i}^{\mathbb{Q}}$, need to be non-negative. $q_{i,j}^{\mathbb{Q}}$ is the element in the i-th row and j-th column of the matrix $\mathbf{Q}^{\mathbb{Q}}$. Also the elements of $\mu^{\mathbb{Q}}$ need to be non-negative. Equations 10 encompass conditions 2. When equations 10 hold, $h_{i,i}$ is not identifiable and we normalise it by setting $h_{i,i} = 1$. Then, if conditions 10 hold, the lower bound of $\mathbf{z}$ is exactly $\mathbf{0}_{n\times 1}$ and for all i

$$z_{i,t+1} = \left(\sqrt{\Delta \mu_i^{\mathbb{Q}} + z_{i,t} + \mathbf{Q}^{(i)\mathbb{Q}} \mathbf{z}_t \Delta} + \frac{\sigma_i \cdot \xi_{i,t+1}^{\mathbb{Q}} \sqrt{\Delta}}{2\sqrt{1 + q_{i,i}^{\mathbb{Q}} \Delta}} \right)^2 . \tag{11}$$

The continuous time limit of this process is

$$d\mathbf{z}_t = \left(\mu^{\mathbb{Q}} + \mathbf{Q}^{\mathbb{Q}} \mathbf{z}_t\right) dt + diag\left(\sigma\right) \cdot \sqrt{diag\left(\mathbf{z}_t\right)} d\mathbf{w}_t^{\mathbb{Q}}$$

$$d\mathbf{z}_t = \left(dz_{1,t}, .., dz_{n,t}\right)', \qquad d\mathbf{w}_t^{\mathbb{Q}} = \left(dw_{1,t}^{\mathbb{Q}}, .., dw_{n,t}^{\mathbb{Q}}\right)'.$$

We can summarise the above parameter restrictions for $\mathbb{A}_n^{sgs}(n)$ models as follows: $\rho_0 = 0$ and $\rho_1 = \iota_n$; $\mathbf{\Sigma}$ is diagonal; $\mathbf{H} = \mathbf{I}_n$ and $\mathbf{\Psi}(i) = \mathbf{e}_i \cdot \mathbf{e}_i' \cdot \psi_i$ for $i = 1, .., n$; finally to ensure that the lower bound of $\mathbf{z}$ is $\mathbf{0}_{n\times 1}$, we impose conditions 10 for all i, that $\mu^{\mathbb{Q}} \geq \mathbf{0}_{n\times 1}$ and that the off-diagonal elements of $\mathbf{Q}^{\mathbb{Q}}$ be non-negative.

3.2 $\mathbb{A}_M^{sgs}(n)$ models

The continuous time limits of $\mathbb{A}_M^{sgs}(n)$ fall in the $\mathbb{A}_M(n)$ family of Dai-Singleton (2000, 2002). In $\mathbb{A}_M^{sgs}(n)$ models the first M factors are non-negative and drive yields volatility, while the last $n - M$ factors do not drive yields volatility. In $\mathbb{A}_M^{sgs}(n)$ models $r_t = \rho_0 + \rho_1' \mathbf{z}_t$ and

$$\mathbf{z}_{t+1} = \left(\mathbf{z}_t + \left(\mu^{\mathbb{Q}} + \mathbf{Q}^{\mathbb{Q}} \mathbf{z}_t\right)\Delta\right) + \mathbf{\Sigma}\sqrt{diag\left(\mathbf{k} + \mathbf{H}\mathbf{z}_t\right)}\xi_{t+1}^{\mathbb{Q}}\sqrt{\Delta} + \Delta \textstyle\sum_{i=1}^n \mathbf{e}_i \xi_{t+1}^{\mathbb{Q}\prime} \mathbf{\Sigma}' \mathbf{\Psi}(i) \mathbf{\Sigma} \xi_{t+1}^{\mathbb{Q}}$$

$$\mathbf{z}_t = \begin{pmatrix} \mathbf{v}_t \\ \overline{\mathbf{z}}_t \end{pmatrix}, \quad \mu^{\mathbb{Q}} = \begin{pmatrix} \mu_v^{\mathbb{Q}} \\ \overline{\mu}^{\mathbb{Q}} \end{pmatrix}, \quad \mathbf{Q}^{\mathbb{Q}} = \begin{pmatrix} \mathbf{Q}_v^{\mathbb{Q}} & \mathbf{0}_{M\times(n-M)} \\ \overline{\mathbf{Q}}_v^{\mathbb{Q}} & \overline{\mathbf{Q}}^{\mathbb{Q}} \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} diag\left(\sigma_M\right) & \mathbf{0}_{M\times(n-M)} \\ \mathbf{0}_{(n-M)\times M} & \overline{\mathbf{\Sigma}} \end{pmatrix},$$

$$\sigma_M = \left(\sigma_1, .., \sigma_M\right)'$$

where: $\rho_0 = 0$ and $\rho_1 = \iota_n$; $\mathbf{v}_t$ is the vector of the M factors that drive yields volatility; $\overline{\mathbf{z}}_t$ is the vector of the $n-M$ factors that do not drive yields volatility; $\mathbf{Q}^{\mathbb{Q}}$ is a block triangular matrix, whose block $\mathbf{0}_{M\times(n-M)}$ is necessary for $\mathbf{v}_{t+1} \geq \mathbf{0}_{M\times 1}$; $q^{\mathbb{Q}}_{v,i,j}$ denotes the element of the i-th row and j-th column of the $M \times M$ matrix $\mathbf{Q}^{\mathbb{Q}}_v$; **all the eigenvalues of the matrix $\left(\mathbf{I}_M + \mathbf{Q}^{\mathbb{Q}}_v \Delta\right)$ must be less than 1 in absolute value in order for $\mathbf{v}_t$ to be stationary**; to ensure that $\mathbf{v}_{t+1} \geq \mathbf{0}_{M\times 1}$ it is necessary that $1 + q^{\mathbb{Q}}_{v,i,i}\Delta \geq 0$ for $i = 1,..,M$, that the off-diagonal elements of $\mathbf{Q}^{\mathbb{Q}}_v$ be non-negative, i.e. $q^{\mathbb{Q}}_{i,j,v} \geq 0$ for $i \neq j$, and that $\mu^{\mathbb{Q}}_v \geq \mathbf{0}_{M\times 1}$, where $\mu^{\mathbb{Q}}_v$ is an $M \times 1$ column vector of parameters; $\overline{\mu}^{\mathbb{Q}}$ is an $(n-M) \times 1$ column vector of parameters; $\mathbf{\Sigma}$ is a block diagonal matrix; $diag\left(\sigma_M\right)$ is an $M \times M$ diagonal matrix with i-th diagonal element σ_i; $\overline{\mathbf{\Sigma}}$ is a $(n-M) \times (n-M)$ matrix that is lower triangular without loss in generality; $k_i = k, \mathbf{H}^{(i)} = \overline{\mathbf{h}}'$ for $i = M+1,..,n$; $\overline{\mathbf{h}} \geq \mathbf{0}_{n\times 1}$ is an $n \times 1$ vector of parameters whose last $n-M$ elements are all equal to 0; $k \geq 0$ is a non-negative scalar parameter; $\mathbf{\Psi}\left(i\right)$ for $i = 1,..,n$ is a set of n diagonal matrixes each of size $n \times n$; we assume $\mathbf{\Psi}\left(i\right) = \mathbf{e}_i \cdot \mathbf{e}_i' \cdot \psi_i$, so that $\sum_{i=1}^{n} B_{i,m-1}\mathbf{\Psi}\left(i\right) = diag\left(B_{i,m-1}\psi_i\right)$, where $diag\left(B_{i,m-1}\psi_i\right)$ is a diagonal matrix with i-th diagonal element $B_{i,m-1}\psi_i$. These assumptions about $\mathbf{\Psi}\left(i\right)$ cover most cases of practical interest for $\mathbb{A}^{sgs}_M\left(n\right)$ models. As the factors $\overline{\mathbf{z}}$ have no squared Gaussian shocks, $\psi_i = 0$ for $i = M+1,..,n$. Under these assumptions the Internet Appendix shows that for

$\mathbb{A}_M^{sgs}(n)$ models

$$\mathbf{F}'\gamma\gamma'\mathbf{F} = w_1 + \mathbf{w}_2' \cdot \mathbf{z}_t$$

$$w_1 = \Delta \sum_{i=1}^{M} \frac{\sigma_i^2 B_{v,i,m-1}^2 k_i}{1 - 2\Delta\psi_i B_{v,i,m-1}\sigma_i^2} + \Omega k \Delta$$

$$\mathbf{w}_2' = \Delta \sum_{i=1}^{M} \frac{\sigma_i^2 B_{v,i,m-1}^2 \mathbf{H}^{(i)}}{1 - 2\Delta\psi_i B_{v,i,m-1}\sigma_i^2} + \Omega \overline{\mathbf{h}}' \Delta$$

$$\Omega = \mathbf{B}_{\overline{z},m-1}' \overline{\mathbf{\Sigma}} \left(\mathbf{I}_{n-M} - 2\Delta \overline{\mathbf{\Sigma}}' diag\left(\psi_i B_{\overline{z},i,m-1}\right) \overline{\mathbf{\Sigma}} \right)^{-1} \overline{\mathbf{\Sigma}}' \mathbf{B}_{\overline{z},m-1}$$

$$\mathbf{B}_{m-1}' \mathbf{z}_t = \left(\mathbf{B}_{v,m-1}', \mathbf{B}_{\overline{z},m-1}' \right) \begin{pmatrix} \mathbf{v}_t \\ \overline{\mathbf{z}}_t \end{pmatrix}$$

$$\mathbf{B}_{v,m-1}' = \left(B_{v,1,m-1}, .., B_{v,M,m-1}\right), \quad \mathbf{B}_{\overline{z},m-1}' = \left(B_{\overline{z},M+1,m-1}, .., B_{\overline{z},n,m-1}\right)$$

where $B_{v,i,m-1}$ is the i-th element of $\mathbf{B}_{v,m-1}$, w_1 is a scalar, $\mathbf{w}_2$ is a $n \times 1$ row vector and $\mathbf{I}_{n-M}$ is the $(n-M) \times (n-M)$ identity matrix. It follows that

$$A_m = -\rho_0 \Delta + A_{m-1} + \mathbf{B}_{m-1}' \mu^{\mathbb{Q}} \Delta + \ln\left(abs\left(|\gamma|\right)\right) + \frac{1}{2} w_1$$

$$\mathbf{B}_m' = -\rho_1' \Delta + \mathbf{B}_{m-1}' \left(\mathbf{I} + \mathbf{Q}^{\mathbb{Q}} \Delta \right) + \frac{1}{2} \mathbf{w}_2'.$$

To ensure that the lower bound of $\mathbf{v}$ is $\mathbf{0}_{M\times 1}$, we further impose conditions 10 so that $\mathbf{H}^{(i)}\mathbf{z}_t = v_{i,t} + \mathbf{Q}_v^{(i)\mathbb{Q}} \mathbf{v}_t \Delta$, $\psi_i = \frac{h_{i,i}}{4\left(1+q_{v,i,i}^{\mathbb{Q}}\Delta\right)}$, $k_i = \frac{\mu_i^{\mathbb{Q}} \Delta h_{i,i}}{1+q_{v,i,i}^{\mathbb{Q}}\Delta}$ for $i \leq M$. Again $h_{i,i}$ is the i-th element of the row vector $\mathbf{H}^{(i)}$. When conditions 10 hold, each $h_{i,i}$ for $i = 1, .., M$ is not identifiable and can be normalised by setting $h_{i,i} = 1$. It follows that

$$\sum_{i=1}^{M} \mathbf{\Psi}(i) = \begin{pmatrix} diag\left(\frac{1}{4\left(1+q_{v,i,i}^{\mathbb{Q}}\Delta\right)} \right) & \mathbf{0}_{M\times(n-M)} \\ \mathbf{0}_{(n-M)\times M} & \mathbf{0}_{(n-M)\times(n-M)} \end{pmatrix}. \tag{12}$$

Then to identify the parameters of the $\mathbf{z}$ process under the $\mathbb{Q}$ measure in $\mathbb{A}_M^{sgs}(n)$ models with $M < n$ we further impose:

- that $\overline{\mathbf{Q}}^{\mathbb{Q}}$ be such that $\mathbf{Q}^{\mathbb{Q}}$ has full rank; except for this requirement the matrixes $\overline{\mathbf{Q}}_v^{\mathbb{Q}}$ and $\overline{\mathbf{Q}}^{\mathbb{Q}}$ inside the block matrix $\mathbf{Q}^{\mathbb{Q}}$ are unconstrained;

- only the first element of $\overline{\mu}^{\mathbb{Q}}$ differs from zero while all other elements are zero; as explained in the Internet Appendix, $\mu^{\mathbb{Q}}$ so normalised is identifiable.

The empirical tests below consider models $\mathbb{A}_M^{sgs+}(n)$, which are the same as $\mathbb{A}_M^{sgs}(n)$ except that $k_i = k, \mathbf{H}^{(i)} = \overline{\mathbf{h}}'$ for $i = 1, .., n$ not just for $i = M+1, .., n$, and except that $\mathbf{\Sigma}$ is lower triangular not block triangular. Then for $\mathbb{A}_M^{sgs+}(n)$ models

$$\mathbf{F}'\gamma\gamma'\mathbf{F} = \mathbf{B}'_{m-1}\mathbf{\Sigma}\left(\mathbf{I}_n - 2\cdot\mathbf{\Sigma}'\mathbf{C}_{m-1}\mathbf{\Sigma}\right)^{-1}\mathbf{\Sigma}'\mathbf{B}_{m-1}\left(k + \overline{\mathbf{h}}'\mathbf{z}_t\right)\Delta = w_1^+ + \mathbf{w}_2^{+\prime}\mathbf{z}_t$$

$$w_1^+ = \mathbf{B}'_{m-1}\mathbf{\Sigma}\left(\mathbf{I}_n - 2\cdot\mathbf{\Sigma}'\mathbf{C}_{m-1}\mathbf{\Sigma}\right)^{-1}\mathbf{\Sigma}'\mathbf{B}_{m-1}k\Delta$$

$$\mathbf{w}_2^{+\prime} = \mathbf{B}'_{m-1}\mathbf{\Sigma}\left(\mathbf{I}_n - 2\cdot\mathbf{\Sigma}'\mathbf{C}_{m-1}\mathbf{\Sigma}\right)^{-1}\mathbf{\Sigma}'\mathbf{B}_{m-1}\overline{\mathbf{h}}'\Delta$$

$$A_m = -\rho_0\Delta + A_{m-1} + \mathbf{B}'_{m-1}\mu^{\mathbb{Q}}\Delta + \ln\left(abs\left(|\gamma|\right)\right) + \frac{1}{2}w_1^+$$

$$\mathbf{B}'_m = -\rho'_1\Delta + \mathbf{B}'_{m-1}\left(\mathbf{I} + \mathbf{Q}^{\mathbb{Q}}\Delta\right) + \frac{1}{2}\mathbf{w}_2^{+\prime}.$$

4 DTATSM-SGS under the real measure $\mathbb{P}$

As with DTATSM-AG, also with DTATSM-SGS much freedom is available in specifying the market price of risk to switch from the $\mathbb{Q}$ measure to the $\mathbb{P}$ measure, without losing closed form solutions for the conditional moments of factors and yields. We now specify the market price of risk to derive the $\mathbf{z}$ process under $\mathbb{P}$, and still assume the $\mathbf{z}$ process under $\mathbb{Q}$ of equation 5. The time t stochastic discount factor $\mathbb{M}_t$ is such that

$$\mathbb{M}_{t+1} = \mathbb{M}_t \cdot e^{-r_t \Delta} \cdot \left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)_{t+1}, \quad \left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)_{t+1} = e^{-\left(\ln(abs(|\gamma_\Phi|)) + \frac{1}{2}\mathbf{\Lambda}_t' \gamma_\Phi' \gamma_\Phi \mathbf{\Lambda}_t \Delta\right) - \mathbf{\Lambda}_t' \xi_{t+1}^{\mathbb{P}} \sqrt{\Delta} - \xi_{t+1}^{\mathbb{P}\prime} \mathbf{\Phi}_t \Delta \xi_{t+1}^{\mathbb{P}}}$$

$$\mathbf{\Lambda}_t = (\mathbf{I}_n + 2\mathbf{\Phi}_t \Delta)^{\frac{1}{2}} \cdot \mathbf{\Lambda}_t^*, \quad \gamma_\Phi = (\mathbf{I}_n + 2\mathbf{\Phi}_t \Delta)^{-\frac{1}{2}}.$$

$\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)_{t+1}$ denotes the Radon-Nykodim derivative between the $\mathbb{Q}$ and the $\mathbb{P}$ measures at time $t+1$. The market price of risk is determined by $\mathbf{\Lambda}_t$ and $\mathbf{\Phi}_t$. $\mathbf{\Phi}_t$ is an $n \times n$ matrix at time t of continuous functions of $\mathbf{z}_t$. $\mathbf{\Lambda}_t^*$ is a vector of size $n \times 1$ at time t of continuous functions of $\mathbf{z}_t$. $\mathbf{\Phi}_t$ is symmetric without loss in generality. An attractive property of process 5 is that, if the lower bound of the affine SGS process $\mathbf{z}_t$ under $\mathbb{Q}$ is $\mathbf{0}_{n\times 1}$, so is also the lower bound of $\mathbf{z}_{t+1}$ under $\mathbb{P}$, irrespective of the choice of $\mathbf{\Lambda}_t$ and $\mathbf{\Phi}_t$. A peculiarity of $\mathbb{M}_{t+1}$ is the term $\xi_{t+1}^{\mathbb{P}\prime} \mathbf{\Phi}_t \Delta \xi_{t+1}^{\mathbb{P}}$, which adds new flexibility in specifying the market price of risk through the matrix $\mathbf{\Phi}_t$. This term was first introduced in the discrete time second-order Esscher transform of the option pricing model of Monfort and Pegoraro (2012). Thus DTATSM-SGS can exploit the flexibility of the second-order Esscher transform in switching from $\mathbb{Q}$ to $\mathbb{P}$. $\mathbf{\Lambda}_t$ and $\mathbf{\Phi}_t$ can be chosen with much freedom as long as they are compatible with an equivalent martingale measure to rule out arbitrage. $\mathbf{\Lambda}_t$ and $\mathbf{\Phi}_t$ should also preserve the ergodicity of $\mathbf{z}$ under the $\mathbb{P}$ measure, which in turn generally preserves the large sample properties of the estimators of the $\mathbf{z}$ process parameters under $\mathbb{P}$. LSD (2010) detail conditions for ergodicity. For model estimation we need the time t conditional expectation and covariance of $\mathbf{z}_{t+1}$ under $\mathbb{P}$, which we respectively denote $E_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$ and $Cov_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$. While $\mathbf{z}$ follows an affine SGS process under

$\mathbb{Q}$, $\mathbf{z}$ may not follow an affine SGS process under $\mathbb{P}$, but, even so, $E_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$ and $Cov_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$ are usually known in closed form. Then it can be shown that $E_t^{\mathbb{P}}\left[\frac{\mathbb{M}_{t+1}}{\mathbb{M}_t}\right] = e^{-r_t\Delta}$ and

$$\ln E_t^{\mathbb{P}}\left[e^{-\mathbf{\Lambda}_t'\xi_{t+1}^{\mathbb{P}}\sqrt{\Delta}-\xi_{t+1}^{\mathbb{P}\prime}\mathbf{\Phi}_t\Delta\xi_{t+1}^{\mathbb{P}}}\right] = \ln\left(abs\left(|\gamma_\Phi|\right)\right)+\frac{1}{2}\mathbf{\Lambda}_t'\gamma_\Phi'\gamma_\Phi\mathbf{\Lambda}_t\Delta = \ln\left(abs\left(|\gamma_\Phi|\right)\right)+\frac{1}{2}\mathbf{\Lambda}_t^{*\prime}\mathbf{\Lambda}_t^*\Delta$$

where $E_t^{\mathbb{P}}[..]$ is the time t conditional expectation under the $\mathbb{P}$ measure. The Internet Appendix shows that

$$\xi_{t+1}^{\mathbb{Q}} = \left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}}\left(\xi_{t+1}^{\mathbb{P}} + \left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{-1}\mathbf{\Lambda}_t\sqrt{\Delta}\right) = \left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}}\xi_{t+1}^{\mathbb{P}}+\mathbf{\Lambda}_t^*\sqrt{\Delta} \tag{13}$$

where $\xi_{t+1}^{\mathbb{Q}} \backsim N\left(\mathbf{0}_{n\times 1}, \mathbf{I}_n\right)$ are Gaussian shocks under the $\mathbb{Q}$ measure and $\xi_{t+1}^{\mathbb{P}} \backsim N\left(\mathbf{0}_{n\times 1}, \mathbf{I}_n\right)$ are Gaussian shocks under the $\mathbb{P}$ measure. The change of measure from $\mathbb{Q}$ to $\mathbb{P}$ and vice versa changes not only the conditional mean, but also the conditional covariance of the shocks. The measure change turns process 5 under $\mathbb{Q}$ into the following process under $\mathbb{P}$

$$\begin{aligned}\mathbf{z}_{t+1} = {} & \mathbf{z}_t + \left(\mu^{\mathbb{Q}} + \mathbf{Q}^{\mathbb{Q}}\mathbf{z}_t\right)\Delta + \mathbf{\Sigma}\cdot\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\mathbf{\Lambda}_t^*\Delta \\ & + \Delta^2\cdot\textstyle\sum_{i=1}^n \mathbf{e}_i\cdot\mathbf{\Lambda}_t^{*\prime}\mathbf{\Sigma}'\mathbf{\Psi}(i)\,\mathbf{\Sigma}\mathbf{\Lambda}_t^* \\ & + 2\Delta^{\frac{3}{2}}\cdot\textstyle\sum_{i=1}^n \mathbf{e}_i\cdot\mathbf{\Lambda}_t^{*\prime}\mathbf{\Sigma}'\mathbf{\Psi}(i)\,\mathbf{\Sigma}\left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}}\xi_{t+1}^{\mathbb{P}} \\ & + \mathbf{\Sigma}\cdot\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}}\xi_{t+1}^{\mathbb{P}}\sqrt{\Delta} \\ & + \Delta\cdot\textstyle\sum_{i=1}^n \mathbf{e}_i\cdot\xi_{t+1}^{\mathbb{P}\prime}\left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}\prime}\mathbf{\Sigma}'\mathbf{\Psi}(i)\,\mathbf{\Sigma}\left(\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}}\xi_{t+1}^{\mathbb{P}}.\end{aligned} \tag{14}$$

In equation 14 $\mathbf{\Lambda}_t^*$ drives the drift of $\mathbf{z}$ and $\mathbf{\Phi}_t$ drives its conditional covariance mainly through the term in the fourth line. However, even when $\mathbf{\Phi}_t = \mathbf{0}_{n\times n}$ the conditional covariance of $\mathbf{z}_{t+1}$ still changes with the measure change, because of the term in the third line of equation 14. In practice the terms of order $O\left(\Delta^{\frac{3}{2}}\right)$ and $O\left(\Delta^2\right)$ in equation 14 are often negligible in estimation and would disappear in continuous time as $\Delta \to 0$. In particular with monthly time steps $\Delta = \frac{1}{12}$ and terms of order $\Delta^{\frac{3}{2}}$ and Δ^2 may be omitted with little loss in estimation accuracy. Then, if we neglect terms of order $o(\Delta)$

$$
\begin{aligned}
&E_t^{\mathbb{P}}[\mathbf{z}_{t+1}] \backsimeq \mu^{\mathbb{Q}}\Delta + \left(\mathbf{I}_n + \mathbf{Q}^{\mathbb{Q}}\Delta\right)\mathbf{z}_t + \mathbf{\Sigma}\cdot\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\mathbf{\Lambda}_t^*\Delta + \\
&+\Delta\textstyle\sum_{i=1}^{n}\mathbf{e}_i\cdot\mathrm{tr}\left(\left(\mathbf{I}_n+2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}\prime}\mathbf{\Sigma}'\mathbf{\Psi}(i)\,\mathbf{\Sigma}\left(\mathbf{I}_n+2\mathbf{\Phi}_t\Delta\right)^{\frac{1}{2}}\right) \\
&Cov_t^{\mathbb{P}}[\mathbf{z}_{t+1}] \backsimeq \mathbf{\Sigma}\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\cdot\left(\mathbf{I}_n+2\mathbf{\Phi}_t\Delta\right)\cdot\sqrt{diag\left(\mathbf{k}+\mathbf{H}\mathbf{z}_t\right)}\mathbf{\Sigma}'\cdot\Delta.
\end{aligned}
$$

$\mathbf{\Lambda}_t^*$ affects $E_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$ but not so much $Cov_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$. $\mathbf{\Phi}_t$ affects both $E_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$ and $Cov_t^{\mathbb{P}}[\mathbf{z}_{t+1}]$ even when $\mathbf{\Lambda}_t^* = \mathbf{0}_{n\times 1}$.

$\mathbf{z}$ is stationary under $\mathbb{Q}$ as long as all the eigenvalues of the feedback matrix $\left(\mathbf{I}_n + \mathbf{Q}^{\mathbb{Q}}\Delta\right)$ are all less than 1 in absolute value. In this paper we focus on specifications whereby the $\mathbf{z}$ process is affine not only under $\mathbb{Q}$ but also under $\mathbb{P}$, so that similar restrictions on the eigenvalues of the feedback matrix of the $\mathbf{z}$ process under $\mathbb{P}$ guarantee process stationarity also under $\mathbb{P}$. More generally, when the $\mathbf{z}$ process is not affine under $\mathbb{P}$, $\mathbf{\Lambda}_t^*$ and $\mathbf{\Phi}_t$ should be chosen so that $\mathbf{z}$ under $\mathbb{P}$ satisfies the sufficient conditions of Mokkadem's Lemma for geometric ergodicity. These conditions are discussed by Le, Dai and Singleton (2010, page 2217) and can be satisfied by many non-affine processes that exhibit SGS.

4.1 Market prices of risk in DTQTSM and in DTATSM-SGS

Quadratic models, be they in continuous time or in discrete time, need no Feller conditions and no admissibility conditions. Also DTATSM-SGS need no Feller conditions that restrict the market price of risk, but DTATSM-SGS need non-negativity conditions that are similar to admissibility conditions for the non-negative factors that appear under a "square root" and that drive yields volatilities. These non-negativity conditions only need imposing under one measure, either $\mathbb{P}$ or $\mathbb{Q}$, and they will automatically ensure non-negativity also under the other measure, thanks to the SGS, and this is an advantage of DTATSM-SGS over continuous time affine models. For example, we saw above that the one factor DTATSM-SGS and the one factor DTQTSM respectively assume

$$r_{t+1} = z_{t+1} = \left(\sqrt{\mu^{\mathbb{Q}}\Delta + (1+q^{\mathbb{Q}}\Delta)\, z_t} + \overline{\sigma}\xi^{\mathbb{Q}}_{t+1}\sqrt{\Delta}\right)^2$$

$$r_{t+1} = \overline{z}^2_{t+1} = \left(\mu^{\mathbb{Q}}\Delta + \left(1+\overline{q}^{\mathbb{Q}}\Delta\right)\overline{z}_t + \overline{\sigma}\xi^{\mathbb{Q}}_{t+1}\sqrt{\Delta}\right)^2.$$

Then, since $\xi^{\mathbb{Q}}_{t+1} = \sqrt{1+2\Phi_t\Delta}\xi^{\mathbb{P}}_{t+1} + \Lambda^*_t\sqrt{\Delta}$, it follows that $\overline{z}^2_{t+1}$ and z_{t+1} are non-negative, if real valued, irrespective of the market prices of risk Φ_t, Λ^*_t, which are two scalars now. Therefore both DTQTSM and DTATSM-SGS have the merit that market prices of risk cannot turn the short rate process r negative. However the DTATSM-SGS needs the conditions $1+2\Phi_t\Delta \geq 0$, $\mu^{\mathbb{Q}}\Delta + \left(1+q^{\mathbb{Q}}\Delta\right) z_t \geq 0$, in order for z_{t+1} to be real valued. Then the market price of risk Λ^*_t, which also needs to be real valued, can be both positive and negative both in DTATSM-SGS and in DTQTSM. For example, if

$$\Lambda_t^* = \frac{1}{\overline{\sigma}\Delta}\left(\sqrt{\mu^{\mathbb{P}}\Delta + (1+q^{\mathbb{P}}\Delta)\, z_t} - \sqrt{\mu^{\mathbb{Q}}\Delta + (1+q^{\mathbb{Q}}\Delta)\, z_t}\right)$$

$$\mu^{\mathbb{P}}\Delta + \left(1+q^{\mathbb{P}}\Delta\right) z_t \geq 0$$

where $\mu^{\mathbb{P}}, q^{\mathbb{P}}$ are parameters under $\mathbb{P}$, it follows that

$$\begin{aligned} z_{t+1} &= \left(\sqrt{\mu^{\mathbb{Q}}\Delta + (1+q^{\mathbb{Q}}\Delta)\, z_t} + \overline{\sigma}\sqrt{1+2\Phi_t\Delta}\xi_{t+1}^{\mathbb{P}}\sqrt{\Delta} + \overline{\sigma}\Lambda_t^*\Delta\right)^2 \\ &= \left(\sqrt{\mu^{\mathbb{P}}\Delta + (1+q^{\mathbb{P}}\Delta)\, z_t} + \overline{\sigma}\sqrt{1+2\Phi_t\Delta}\xi_{t+1}^{\mathbb{P}}\sqrt{\Delta}\right)^2 . \end{aligned}$$

Therefore Λ_t^* can be both positive and negative, so that expected excess bond returns can be both positive and negative, just as in the continuous time essentially affine models of Duffee (2002). Also DTQTSM can have both positive and negative market prices of risk, as is well known. For example, if $\Lambda_t^* = \frac{1}{\overline{\sigma}}\left(\mu^{\mathbb{P}} + \overline{q}^{\mathbb{P}}\overline{z}_t - \left(\mu^{\mathbb{Q}} + \overline{q}^{\mathbb{Q}}\overline{z}_t\right)\right)$, it follows that

$$\begin{aligned} r_{t+1} = \overline{z}_{t+1}^2 &= \left(\mu^{\mathbb{Q}}\Delta + \left(1+\overline{q}^{\mathbb{Q}}\Delta\right)\overline{z}_t + \overline{\sigma}\sqrt{1+2\Phi_t\Delta}\xi_{t+1}^{\mathbb{P}}\sqrt{\Delta} + \overline{\sigma}\Lambda_t^*\Delta\right)^2 \\ &= \left(\mu^{\mathbb{P}}\Delta + \left(1+\overline{q}^{\mathbb{P}}\Delta\right)\overline{z}_t + \overline{\sigma}\sqrt{1+2\Phi_t\Delta}\xi_{t+1}^{\mathbb{P}}\sqrt{\Delta}\right)^2 . \end{aligned}$$

Again Λ_t^* can be both positive and negative. Then also DTATSM-AG can have market prices of risk that can be both positive and negative. Finally the second order Esscher transform implies that the market price of risk Φ_t can cause factors conditional covariance to differ under $\mathbb{P}$ and $\mathbb{Q}$. Unlike DTATSM-AG, DTATSM-SGS seem ideal for the second order Esscher transform, because of their Gaussian shocks. However, the second order Esscher transform can be applied to DTQTSM just as well as to DTATSM-SGS. We next **consider** the $\mathbf{z}$ process **under** $\mathbb{P}$ for $\mathbb{A}_n^{sgs}(n)$ models.

4.2 $\mathbb{A}_n^{sgs}(n)$ models under the real measure

Models $\mathbb{A}_n^{sgs}(n)$ assume process 11 under $\mathbb{Q}$. Then, if $\mathbf{\Phi}_t$ is diagonal and $\Phi_{i,i}$ is its i-th diagonal element and is a parameter, and if

$$\mathbf{\Lambda}_t^* = \left(\Lambda_{1,t}^*, .., \Lambda_{n,t}^*\right)'$$

$$\Lambda_{i,t}^* = \left(\sqrt{\Delta\mu_i^{\mathbb{P}} + z_{i,t} + \mathbf{Q}^{(i)\mathbb{P}}\mathbf{z}_t\Delta} - \sqrt{\Delta\mu_i^{\mathbb{Q}} + z_{i,t} + \mathbf{Q}^{(i)\mathbb{Q}}\mathbf{z}_t\Delta}\right)\frac{2\sqrt{1+q_{i,i}^{\mathbb{Q}}\Delta}}{\Delta\sigma_i}, \quad i = 1, .., n$$

it follows that $\xi_{i,t+1}^{\mathbb{Q}} = \sqrt{1+2\Phi_{i,i}\Delta}\xi_{i,t+1}^{\mathbb{P}} + \Lambda_{i,t}^*\sqrt{\Delta}$ and that the $\mathbf{z}$ process under $\mathbb{P}$ can be written as $diag\left(\mathbf{z}_{t+1}\right)\cdot\iota_n$, where

$$diag\left(\mathbf{z}_{t+1}\right) = \left(\sqrt{diag\left(\mathbf{z}_t + \left(\mu^{\mathbb{P}} + \mathbf{Q}^{\mathbb{P}}\mathbf{z}_t\right)\Delta\right)} + \frac{1}{2}diag\left(\sigma\right)\sqrt{diag\left(\frac{1+2\Phi_{i,i}\Delta}{1+q_{i,i}^{\mathbb{Q}}\Delta}\right)}diag\left(\xi_{t+1}^{\mathbb{P}}\right)\sqrt{\Delta}\right)^2. \tag{15}$$

Recall that $diag\left(x_i\right)$ denotes the diagonal matrix whose i-th diagonal element is x_i. A model similar to this is $\mathbb{A}_3^{sgs}(3)$**,** which is tested empirically below. The continuous time limit of 15 (in vector format) is

$$d\mathbf{z}_t = \left(\mu^{\mathbb{P}} + \mathbf{Q}^{\mathbb{P}}\mathbf{z}_t\right)dt + diag\left(\sigma\right)\cdot\sqrt{diag\left(\mathbf{z}_t\right)}d\mathbf{w}_t^{\mathbb{P}}$$

$$d\mathbf{w}_t^{\mathbb{P}} = \left(dw_{1,t}^{\mathbb{P}}, .., dw_{n,t}^{\mathbb{P}}\right)'.$$

Other things equal, if $\mathbf{\Phi}_t$ is unconstrained, then

$$diag\left(\mathbf{z}_{t+1}\right) = \left(\sqrt{diag\left(\mathbf{z}_t + \left(\mu^{\mathbb{P}} + \mathbf{Q}^{\mathbb{P}}\mathbf{z}_t\right)\Delta\right)} + \frac{1}{2}\left(diag\left(1+q_{i,i}^{\mathbb{Q}}\Delta\right)\right)^{-1/2}\cdot diag\left(\mathbf{S}\xi_{t+1}^{\mathbb{P}}\right)\sqrt{\Delta}\right)^2 \tag{16}$$

with $\mathbf{S} = diag\left(\sigma\right)\sqrt{\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta}$. Then we can set

$$\sqrt{\mathbf{I}_n + 2\mathbf{\Phi}_t\Delta} = diag\left(\mathbf{p}\right)\cdot\mathbf{\Xi}, \qquad \mathbf{p} = \left(p_1, .., p_n\right)'. \tag{17}$$

$\mathbf{p}$ is a vector of n risk-premium parameters. Ξ is an $n \times n$ lower triangular matrix and $\Xi\Xi' = [\rho_{i,j}]$. $[\rho_{i,j}]$ is an $n \times n$ matrix whose element in the i-th row and j-th column is $\rho_{i,j}$, which is the correlation coefficient between $\Xi^{(i)}\xi_{t+1}^{\mathbb{P}}$ and $\Xi^{(j)}\xi_{t+1}^{\mathbb{P}}$, where $\Xi^{(i)}$ and $\Xi^{(j)}$ denote the i-th row and j-th row of Ξ. Ξ is computed through a Choleski decomposition of matrix $[\rho_{i,j}]$. Note that 17 implies that factors shocks are correlated under $\mathbb{P}$ even as they are uncorrelated under $\mathbb{Q}$. If we neglect terms of order $o(\Delta)$, **17** implies that

$$Cov_t^{\mathbb{P}}[\mathbf{z}_{t+1}] \backsimeq diag(\mathbf{z}_t) \cdot diag(\sigma) \cdot diag(\mathbf{p}) \cdot \Xi\Xi' \cdot diag(\mathbf{z}_t) \cdot diag(\sigma) \cdot diag(\mathbf{p}) \cdot \Delta$$

and the continuous time limit of the $\mathbf{z}$ process under $\mathbb{P}$ (in vector format) becomes $d\mathbf{z}_t = \left(\mu^{\mathbb{P}} + \mathbf{Q}^{\mathbb{P}}\mathbf{z}_t\right)dt + \sqrt{diag(\mathbf{z}_t)} \cdot diag(\sigma) \cdot diag(\mathbf{p}) \cdot \Xi \cdot d\mathbf{w}_t^{\mathbb{P}}$. Processes 11 and 16 imply that $\mathbf{z}$ is non-negative both under $\mathbb{P}$ and under $\mathbb{Q}$ irrespective of the market price of risk. Feller conditions effectively restrict market prices of risk, while the non-negativity conditions of DTATSM-SGS do not. A model similar to 16 and 17 is model $\mathbb{A}_3^{sgs}(3)\,v$ tested below.

5 The tested models

This section details the DTATSM-SGS and DTATSM-AG that are later tested. Comparisons of DTATSM-SGS with the many continuous time term structure models in the literature are beyond the scope of this paper. This is not just due to space limits, since LSD (2010) and others have already explained the benefits of discrete time models and these benefits have already made DTATSM-AG quite popular. Then the comparison between DTATSM-SGS and DTATSM-

AG seems the most pressing one, because these models seem the only discrete time affine counterparts, if we exclude Sun (1992), to the popular continuous time models $\mathbb{A}_n(n), \mathbb{A}_M(n)$. For all models we use monthly observations so that $\Delta = 1/12$ since time is measured in years. All tested models have three factors. For all tested DTATSM-SGS we assume $\mathbf{\Psi}(i) = \mathbf{e}_i \cdot \mathbf{e}_i' \cdot \psi_i$ for $i = 1, 2, 3$.

5.1 Models $\mathbb{A}_1^{sgs+}(3)$, $\mathbb{A}_1^{sgs+}(3)\,v$

Model $\mathbb{A}_1^{sgs+}(3)$ assumes

$$r_t = \overline{z}_{3,t}$$

$$v_{1,t+1} = v_{1,t} + \left(\mu_1^{\mathbb{Q}} - q_1^{\mathbb{Q}} v_{1,t}\right)\Delta + \sigma_1 \sqrt{v_{1,t}}\xi_{1,t+1}^{\mathbb{Q}}\sqrt{\Delta} + \psi_1^{\mathbb{Q}}\sigma_1^2\left(\xi_{1,t+1}^{\mathbb{Q}}\right)^2\Delta$$

$$\overline{z}_{2,t+1} = \overline{z}_{2,t} + \left(v_{1,t} - q_2^{\mathbb{Q}}\overline{z}_{2,t}\right)\Delta + \sigma_2\sqrt{v_{1,t}}\xi_{2,t+1}^{\mathbb{Q}}\sqrt{\Delta}$$

$$\overline{z}_{3,t+1} = \overline{z}_{3,t} + \left(\overline{z}_{2,t} - q_3^{\mathbb{Q}}\overline{z}_{3,t}\right)\Delta + \sigma_3\sqrt{v_{1,t}}\xi_{3,t+1}^{\mathbb{Q}}\sqrt{\Delta}$$

$$Cov_t\left[\xi_{t+1}^{\mathbb{Q}}\right] = Cov_t\left[\xi_{t+1}^{\mathbb{P}}\right] = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}, \ \xi_{t+1} = \left(\xi_{1,t+1}, \xi_{2,t+1}, \xi_{3,t+1}\right)'$$

$$\mu_1^{\mathbb{Q}} \geq 0, \ \psi_1^{\mathbb{Q}} = \frac{1}{4\left(1 - q_1^{\mathbb{Q}}\Delta\right)}, \ 1 - q_1^{\mathbb{Q}}\Delta \geq 0, \ q_1^{\mathbb{Q}} \geq 0.$$

$\rho_{13}, \rho_{23}, \rho_{12}$ are correlation coefficients. Then under $\mathbb{P}$ all is the same except that $\overline{z}_{3,t+1} = \overline{z}_{3,t} + \left(\mu_3^{\mathbb{P}} + \overline{z}_{2,t} - q_3^{\mathbb{P}}\overline{z}_{3,t}\right)\Delta + \sigma_3\sqrt{v_{1,t}}\xi_{3,t+1}^{\mathbb{Q}}\sqrt{\Delta}$. Note that v_1 is mean reverting both under $\mathbb{P}$ and $\mathbb{Q}$. Model $\mathbb{A}_1^{sgs+}(3)\,v$ is the same as $\mathbb{A}_1^{sgs+}(3)$ except that $\overline{z}_{3,t+1} = \overline{z}_{3,t} + \left(\mu_3^{\mathbb{P}} + \overline{z}_{2,t} - q_3^{\mathbb{P}}\overline{z}_{3,t}\right)\Delta + p_3\sigma_3\sqrt{v_{1,t}}\xi_{3,t+1}^{\mathbb{P}}\sqrt{\Delta}$ under $\mathbb{P}$, where p_3 is a risk premium parameter. Therefore in $\mathbb{A}_1^{sgs+}(3)\,v$ the conditional variance of r_t differs under $\mathbb{P}$ and $\mathbb{Q}$.

5.2 Models $\mathbb{A}_3^{sgs}(3)$ and $\mathbb{A}_3^{sgs}(3)\,v$

In model $\mathbb{A}_3^{sgs}(3)$

$$r_t = v_{3,t}$$

$$v_{1,t+1} = v_{1,t} + \left(\mu_1^{\mathbb{Q}} - q_1^{\mathbb{Q}} v_{1,t}\right)\Delta + \sigma_1\sqrt{v_{1,t}}\xi_{1,t+1}^{\mathbb{Q}}\sqrt{\Delta} + \psi_1^{\mathbb{Q}}\sigma_1^2\left(\xi_{1,t+1}^{\mathbb{Q}}\right)^2\Delta$$

$$v_{2,t+1} = v_{2,t} + q_2^{\mathbb{Q}}\left(v_{1,t} - v_{2,t}\right)\Delta + \sigma_2\sqrt{v_{2,t}}\xi_{2,t+1}^{\mathbb{Q}}\sqrt{\Delta} + \psi_2^{\mathbb{Q}}\sigma_2^2\left(\xi_{2,t+1}^{\mathbb{Q}}\right)^2\Delta$$

$$v_{3,t+1} = v_{3,t} + \left(q_3^{\mathbb{Q}}\left(v_{2,t} - v_{3,t}\right)\right)\Delta + \sigma_3\sqrt{v_{3,t}}\xi_{3,t+1}^{\mathbb{Q}}\sqrt{\Delta} + \psi_3^{\mathbb{Q}}\sigma_3^2\left(\xi_{3,t+1}^{\mathbb{Q}}\right)^2\Delta$$

$$\mu_3^{\mathbb{Q}} = 0,\ \mu_1^{\mathbb{Q}} \geq 0,\ Cov_t\left[\xi_{t+1}^{\mathbb{Q}}\right] = Cov_t\left[\xi_{t+1}^{\mathbb{P}}\right] = \mathbf{I}_3,\ \xi_{t+1} = \left(\xi_{1,t+1}, \xi_{2,t+1}, \xi_{3,t+1}\right)'$$

$$\psi_i^{\mathbb{Q}} = \frac{1}{4\left(1 - q_i^{\mathbb{Q}}\Delta\right)},\ 1 - q_i^{\mathbb{Q}}\Delta \geq 0,\ q_i^{\mathbb{Q}} \geq 0 \text{ for } i = 1, 2, 3.$$

Under $\mathbb{P}$ all is the same except that $v_{3,t+1} = v_{3,t} + \left(\mu_3^{\mathbb{P}} + q_3^{\mathbb{P}}\left(v_{2,t} - v_{3,t}\right)\right)\Delta + \sigma_3\sqrt{v_{3,t}}\xi_{3,t+1}^{\mathbb{P}}\sqrt{\Delta} + \frac{\sigma_3^2\left(\xi_{3,t+1}^{\mathbb{P}}\right)^2\Delta}{4\left(1-q_3^{\mathbb{P}}\Delta\right)}$, with $1 - q_3^{\mathbb{P}}\Delta \geq 0$ and $q_3^{\mathbb{P}} \geq 0$. v_1, v_2, v_3 are non-negative and mean reverting both under $\mathbb{Q}$ and $\mathbb{P}$. Model $\mathbb{A}_3^{sgs}(3)\,v$ is the same as $\mathbb{A}_3^{sgs}(3)$ except that $Cov_t\left[\xi_{t+1}^{\mathbb{P}}\right] = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$. In $\mathbb{A}_3^{sgs}(3)\,v$ factors shocks are correlated under $\mathbb{P}$ but not under $\mathbb{Q}$.

5.3 Models $\mathbb{AG}_3(3)$, $\mathbb{AG}_1(3)$

Model $\mathbb{AG}_3(3)$ follows Le, Dai and Singleton (2010) and is the AG counterpart to $\mathbb{A}_3^{sgs+}(3)$. The continuous time limit of $\mathbb{AG}_3(3)$ is

$$
\begin{aligned}
r_t &= v_{3,t} \\
dv_{1,t} &= \left(\mu_1^{\mathbb{Q}} - q_1^{\mathbb{Q}} v_{1,t}\right) dt + \sigma_1 \sqrt{v_{1,t}} dw_{1,t}^{\mathbb{Q}} \\
dv_{2,t} &= \left(q_2^{\mathbb{Q}} \left(v_{1,t} - v_{2,t}\right)\right) dt + \sigma_2 \sqrt{v_{2,t}} dw_{2,t}^{\mathbb{Q}} \\
dv_{3,t} &= \left(\mu_3^{\mathbb{Q}} + q_3^{\mathbb{Q}} \left(v_{2,t} - v_{3,t}\right)\right) dt + \sigma_3 \sqrt{v_{3,t}} dw_{3,t}^{\mathbb{Q}}
\end{aligned}
$$

with $\mu_3^{\mathbb{Q}} = 0, q_1^{\mathbb{Q}}, q_2^{\mathbb{Q}}, q_3^{\mathbb{Q}}, \mu_1^{\mathbb{Q}} \geq 0$. $dv_{i,t}$ are stochastic differentials and $dw_{i,t}^{\mathbb{Q}}$ are differentials of Wiener processes for $i = 1,2,3$ under $\mathbb{Q}$. dt is the infinitesimal time increment. In $\mathbb{AG}_3(3)$ the variable $\frac{v_{i,t+1}}{c_i}$ for $i = 1,2,3$ is distributed according to a Poisson mixture of standard gamma distributions, i.e. $\frac{v_{i,t+1}}{c_i} \sim gamma\left(\mathfrak{w}_i^{\mathbb{Q}} + \mathfrak{l}\right)$ and $\mathfrak{l}$ given $\mathbf{v}_t$ is Poisson distributed, i.e. $\mathfrak{l} \sim poisson\left(\frac{\rho^{(i)}\mathbf{v}_t}{c_i}\right) = \frac{\left(\frac{\rho^{(i)}\mathbf{v}_t}{c_i}\right)^{\mathfrak{l}} e^{-\left(\frac{\rho^{(i)}\mathbf{v}_t}{c_i}\right)}}{\mathfrak{l}!}$ where $\mathbf{v}_t = (v_{1,t}, v_{2,t}, v_{3,t})$, $\rho^{(i)}$ is the i-th row of ρ, $\rho = \mathbf{I}_3 + \mathbf{Q}^{\mathbb{Q}}\Delta$, $\mathbf{Q}^{\mathbb{Q}} = \begin{pmatrix} -q_1^{\mathbb{Q}} & 0 & 0 \\ q_2^{\mathbb{Q}} & -q_2^{\mathbb{Q}} & 0 \\ 0 & q_3^{\mathbb{Q}} & -q_3^{\mathbb{Q}} \end{pmatrix}$ and $\frac{1}{\Delta} > q_i^{\mathbb{Q}} > 0, c_i = \frac{\sigma_i^2}{2}\Delta, \mathfrak{w}_i^{\mathbb{Q}} = \frac{2\mu_i^{\mathbb{Q}}}{\sigma_i^2}$ for $i = 1,2,3$. Under the $\mathbb{P}$ measure all is the same except that $\mathbf{Q}^{\mathbb{P}} = \begin{pmatrix} -q_1^{\mathbb{Q}} & 0 & 0 \\ q_2^{\mathbb{Q}} & -q_2^{\mathbb{Q}} & 0 \\ 0 & q_3^{\mathbb{P}} & -q_3^{\mathbb{P}} \end{pmatrix}$ replaces $\mathbf{Q}^{\mathbb{Q}}$ and $\mu_1^{\mathbb{P}} = \mu_1^{\mathbb{Q}}, \mu_3^{\mathbb{P}} \neq \mu_3^{\mathbb{Q}}, \mathfrak{w}_i^{\mathbb{P}} = \frac{2\mu_i^{\mathbb{P}}}{\sigma_i^2}$ for $i = 1,2,3$. To improve model performance we do not impose the Feller conditions $\mathfrak{w}_i^{\mathbf{Q}}, \mathfrak{w}_i^{\mathbf{P}} \geq 1$ for $i = 1,2,3$. Then according to $\mathbb{AG}_3(3)$ the price of a discount bond is $P_{t,m} = \exp\left(A^{ag}_m + \mathbf{B}^{ag\prime}_m \mathbf{z}_t\right)$ with

$$A^{ag}_m = A^{ag}_{m-1} + \sum_{i=1}^{3} \ln\left(1 - c_i \cdot B^{ag}_{i,m-1}\right)^{-\mathfrak{w}^{\mathbb{Q}}_i}$$

$$\mathbf{B}^{ag\prime}_m = -\left(0,0,1\right)\Delta + \sum_{i=1}^{3} \frac{B^{ag}_{i,m-1} \cdot \rho^{(i)}}{1 - c_i \cdot B^{ag}_{i,m-1}}$$

$$A^{ag}_0 = 0, \mathbf{B}^{ag\prime}_0 = \left(0,0,0\right).$$

Model $\mathbb{AG}_1(3)$ is the counterpart of $\mathbb{A}^{sgs+}_1(3)$. The continuous time limit of $\mathbb{AG}_1(3)$ is

$$r_t = \overline{z}_{3,t}$$

$$dv_{1,t} = \left(\mu^{\mathbb{Q}}_1 - q^{\mathbb{Q}}_1 v_{1,t}\right)dt + \sigma_1\sqrt{v_{1,t}}dw^{\mathbb{Q}}_{1,t}$$

$$d\overline{z}_{2,t} = \left(v_{1,t} - q^{\mathbb{Q}}_2 \overline{z}_{2,t}\right)dt + \sigma_2\sqrt{v_{1,t}}dw^{\mathbb{Q}}_{2,t} + \theta_1\left(dv_{1,t} - E^{\mathbb{Q}}_t\left[dv_{1,t}\right]\right)$$

$$d\overline{z}_{3,t} = \left(\mu^{\mathbb{Q}}_3 + \overline{z}_{2,t} - q^{\mathbb{Q}}_3 \overline{z}_{3,t}\right)dt + \sigma_3\sqrt{v_{1,t}}dw^{\mathbb{Q}}_{3,t} + \theta_2\left(dv_{1,t} - E^{\mathbb{Q}}_t\left[dv_{1,t}\right]\right)$$

where $\mu^{\mathbb{Q}}_3 = 0$, θ_1, θ_2 are two constants and $dw^{\mathbb{Q}}_{2,t} dw^{\mathbb{Q}}_{3,t} = \rho_{23}dt$. Now $\frac{v_{1,t+1}}{c_1} \sim gamma\left(\mathfrak{w}^{\mathbb{Q}}_1 + \mathfrak{l}\right)$ and $\mathfrak{l} \sim poisson\left(\frac{\rho^{(1)}\mathbf{z}_t}{c_1}\right) = \frac{\left(\frac{\rho^{(1)}\mathbf{z}_t}{c_1}\right)^{\mathfrak{l}} e^{-\left(\frac{\rho^{(1)}\mathbf{z}_t}{c_1}\right)}}{\mathfrak{l}!}$; $z_t = (v_{1,t}, \overline{z}_{2,t}, \overline{z}_{3,t})$, $\rho^{(1)}$ is the first row of ρ, $\rho = \mathbf{I}_3 + \mathbf{Q}^{\mathbb{Q}}\Delta$, $\frac{1}{\Delta} > q^{\mathbb{Q}}_1 > 0$ and $c_1 = \frac{\sigma^2_1}{2}\Delta, \mathfrak{w}^{\mathbb{Q}}_1 = \frac{2\mu^{\mathbb{Q}}_1}{\sigma^2_1}$. Under the $\mathbb{P}$ measure all is the same except that $\mathbf{Q}^{\mathbb{P}}$ replaces $\mathbf{Q}^{\mathbb{Q}}$, $\mathbf{Q}^{\mathbb{P}} = \begin{pmatrix} -q^{\mathbb{Q}}_1 & 0 & 0 \\ 1 & -q^{\mathbb{Q}}_2 & 0 \\ 0 & 1 & -q^{\mathbb{P}}_3 \end{pmatrix}$, $\mathfrak{w}^{\mathbb{P}}_1 = \frac{2\mu^{\mathbb{P}}_1}{\sigma^2_1}, \mu^{\mathbb{P}}_1$ respectively replace and are the same as $\mathfrak{w}^{\mathbb{Q}}_1, \mu^{\mathbb{Q}}_1$, while $\mu^{\mathbb{P}}_3 \neq \mu^{\mathbb{Q}}_3$. We do not impose the Feller conditions $\mathfrak{w}^{\mathbf{Q}}_1, \mathfrak{w}^{\mathbf{P}}_1 \geq 1$. The formulae for bond prices and factors conditional moments according to $\mathbb{AG}_1(3)$ are reported in Creal and Wu (2015). In $\mathbb{AG}_1(3)$ only one of the three factors drives factors stochastic volatility, while in $\mathbb{AG}_3(3)$ all three factors drive stochastic volatility. The symbols $v_{1,t}, v_{2,t}, v_{3,t}, \overline{z}_{2,t}, \overline{z}_{3,t}$ that denote the factors in the AG models highlight the similarity between AG models and SGS models, but such factors follow processes that differ from the SGS processes.

5.4 The Gaussian models

Gaussian models $\mathbb{A}_0(3)$ assumes that under $\mathbb{Q}$

$$\begin{aligned}
r_t &= \overline{z}_{3,t} \\
\overline{z}_{1,t+1} &= \overline{z}_{1,t} + \left(\mu_1^{\mathbb{Q}} - q_1^{\mathbb{Q}}\overline{z}_{1,t}\right)\Delta + \sigma_1\xi_{1,t+1}^{\mathbb{Q}}\sqrt{\Delta} \\
\overline{z}_{2,t+1} &= \overline{z}_{2,t} + q_2^{\mathbb{Q}}\left(\overline{z}_{1,t} - \overline{z}_{2,t}\right)\Delta + \sigma_2\xi_{2,t+1}^{\mathbb{Q}}\sqrt{\Delta} \\
\overline{z}_{3,t+1} &= \overline{z}_{3,t} + q_3^{\mathbb{Q}}\left(\overline{z}_{2,t} - \overline{z}_{3,t}\right)\Delta + \sigma_3\xi_{3,t+1}^{\mathbb{Q}}\sqrt{\Delta} \\
Cov_t\left[\xi_{t+1}^{\mathbb{Q}}\right] &= Cov_t\left[\xi_{t+1}^{\mathbb{P}}\right] = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}.
\end{aligned}$$

In $\mathbb{A}_0(3)$ the $\mathbf{z}$ process under $\mathbb{P}$ is the same as under $\mathbb{Q}$, except that $\overline{z}_{3,t+1} = \overline{z}_{3,t} + \left(\mu_3^{\mathbb{P}} + q_3^{\mathbb{P}}\left(\overline{z}_{2,t} - \overline{z}_{3,t}\right)\right)\Delta + \sigma_3\xi_{3,t+1}^{\mathbb{P}}\sqrt{\Delta}$. Finally, as term of comparison, we estimate a discrete time quadratic Gaussian model, denoted as $\boldsymbol{Q}(3)$, which assumes $r_t = \overline{z}_{3,t}^2$ while all else is as in $\mathbb{A}_0(3)$. This DTQTSM is a special case of the one in Realdon (2006).

5.5 The market price of risk

The market price of risk of all the above models is similar and affects the drift of factors z_3 or $\overline{z}_3$ that drive the short rate r. Thus only yields "level risk" is priced, a parsimonious specification inspired by past literature, e.g. **Cochrane and Piazzesi (2008) and** Bauer (2018). $\mathbb{A}_1^{sgs}(3)\,v$ additionally assumes that the conditional variance of z_3 differs under $\mathbb{P}$ and $\mathbb{Q}$. Instead $\mathbb{A}_3^{sgs}(3)\,v$ additionally assumes that factors correlation differs under $\mathbb{P}$ and $\mathbb{Q}$, while factors conditional variance is the same under $\mathbb{P}$ and $\mathbb{Q}$.

6 First empirical tests

This section presents tests of SGS and AG models using US Treasury bond yields. Yields are continuously compounded and observed monthly for all yearly maturities from one year to twenty years. Yields are computed from discount functions provided by Thompson-Reuters Eikon and the discount functions are implied from Treasury bond prices. Table 1 reports yields descriptive statistics for the sample period, i.e. from January 1995 to November 2017. All models are estimated through (extended) Kalman Filters. All estimated models predict ergodic yields as we constrain all model factors to be mean-reverting rather than mean-averting. The prior distribution of the latent factors at the start of the sample is Gaussian with mean and covariance equal to the unconditional mean and covariance of the factors.

[Table 1]

In the Tables "A1(3)v" denotes $\mathbb{A}_1^{sgs+}(3)\,v$, "A3(3)v" denotes $\mathbb{A}_3^{sgs}(3)\,v$ and the names of the other SGS models are shortened similarly. Table 2 summarises forecasting performance by reporting root mean squared errors (RMSE's) for each estimated model and yield maturity. For example RMSE for the five year maturity is computed as $\sqrt{\frac{\sum_{t=2}^{276}\left(y_{5\cdot 12,t}-E_{t-1}^{\mathbb{P}}(y_{5\cdot 12,t})\right)^2}{275}}$, where $y_{t,m} = -\frac{\ln(P_{t,m})}{\Delta m}$. Recall that $P_{t,m}$ is the price of a discount bond at the end of month t with maturity at the end of month $t+m$ and that $\Delta = 1/12$ is the monthly time step. Therefore $y_{5\cdot 12,t}$ is the five year (i.e. 60 month) discount bond yield at the end of month t and $E_{t-1}^{\mathbb{P}}\left(y_{5\cdot 12,t}\right)$ is the Kalman Filter prediction of $y_{5\cdot 12,t}$ made at the end of month $t-1$. There are 276 months in the sample. The row "Average" of Table 2 computes average RMSE across the twenty maturities. The average RMSE of $\mathbb{A}_3^{sgs}(3)\,v$ is 28 basis points and the lowest, while those of $\mathbb{AG}_1(3)\,c$ and $\mathbb{AG}_3(3)$ are 59 and 77 basis points respectively and are the largest. This highlights the relative weakness of the AG models. For the other models Average RMSE's range between 30 and 41 basis points.

[Table 2]

[Table 3]

Table 3 presents the estimated parameters for all models in the columns "param" and the corresponding asymptotic standard errors in the columns "stdev". Standard errors are computed with the "Sandwich" estimator. In Tables 3 and 4 the parameters $\mu_{..}^{\mathbb{Q}}$, $q_{..}^{\mathbb{Q}}$, $\mu_{..}^{\mathbb{P}}$, $q_{..}^{\mathbb{P}}$ are denoted respectively as $\mu_{..}^{*}$, $q_{..}^{*}$, $\mu_{..}$, $q_{..}$. s is the estimated standard deviation of yields observation errors, which is assumed to be the same for all yield maturities to reduce parameter over-fitting and to highlight differences in model fit to observed yields. s is about 8 to 9 basis points for most models, but is significantly higher for the affine models that rule out negative factors, i.e. $\mathbb{A}_3^{sgs}(3)\,, \mathbb{A}_3^{sgs}(3)\,v, \mathbb{AG}_3(3)$. Unreported results show that, when the models were estimated under the assumption that s varies across maturities, for each model the average of the estimates of s across all maturities was very close to the estimate of s reported in Table 3.

For each model the row lk reports the value of the maximised Kalman quasi-likelihood and the row AIC reports the Akaike Information Criterion (AIC). AIC is best, i.e. lowest, for the quadratic model, which is followed by $\mathbb{A}_0(3)$ and then by $\mathbb{A}_1^{sgs}(3)\,v$ and $\mathbb{A}_1^{sgs}(3)$. Models whose name ends with "v" feature risk premia parameters related to the second order Esscher transform, namely p_3 for $\mathbb{A}_1^{sgs}(3)\,v$ and $\rho_{12}, \rho_{13}, \rho_{23}$ for $\mathbb{A}_3^{sgs}(3)\,v$. All these four parameters are significant at conventional levels in Tables 3. This evidence is in favour of the transform. Almost all estimates of $\rho_{12}, \rho_{13}, \rho_{23}$ are significant across models, thus supporting correlated factors. The Vuong likelihood ratio tests in Table 4 show a number of other results.

First, $\mathbb{A}_3^{sgs}(3)\,v$, whose factors are conditionally correlated under the real measure but not under the risk-neutral measure, fits yields significantly better better than $\mathbb{A}_3^{sgs}(3)$, whose factors are uncorrelated under both measures. This shows how the second order Esscher transform, which implies factors correlation under the real measure, "helps" the SGS model with non-negative factors. Then the Vuong tests show that $\mathbb{A}_3^{sgs}(3)\,v$ outperforms $\mathbb{AG}_3(3)$ with a p-value of 4%. However all the affine models that rule out negative factors, namely $\mathbb{A}_3^{sgs}(3), \mathbb{A}_3^{sgs}(3)\,v, \mathbb{AG}_3(3)$, tend to under-perform the other models.

Second, the Vuong tests show that SGS models $\mathbb{A}_1^{sgs}(3)\,v$ and $\mathbb{A}_1^{sgs}(3)$ outperform their counterpart AG model $\mathbb{AG}_1(3)$. Third, $\mathbb{A}_1^{sgs}(3)\,v$ outperforms

$\mathbb{A}_1^{sgs}(3)$ with a p-value of 3%, thanks to the Esscher transform and the risk premium parameters p_3, which causes the conditional volatility of the short rate r_t to differ under the real and risk-neutral measures. Finally the Vuong tests show that the quadratic model neither outperforms not is outperformed by any other model, while affine Gaussian model $\mathbb{A}_0(3)$ outperforms most other models.

Overall the Vuong tests show that SGS models can fit in-sample yields better than corresponding AG models. This is the case even though the tested SGS and AG models converge to the same continuous time affine models. The second order Esscher transform further enhances SGS models through risk premia that alter factors volatility and correlation[1].

[Table 4]

[1] The tests in Table 4 and the estimates in Table 3 use the Gaussian "quasi-likelihood" function of EKF. Note that the likelihood function of bond yields is not Gaussian according to affine SGS and AG models and according to quadratic models. Then the conditional variance of yields in affine SGS and AG models is computed by using the mean of latent factors as estimated by EKF. This too is an approximation since the exact value of latent factors is not known. As a result, maximising the EKF "quasi-likelihood" gives parameter estimates that generally lack asymptotic consistency to the true parameter values in the population, even for a well-specified model.

However the Vuong tests in Table 4 that compare the different models seem still relevant because all models are all estimated in the same way through EKF. Moreover Vuong tests can even compare mis-specified likelihood functions.

7 Further tests

This section presents further tests of the models reported in Table 3. These tests compare the one factor versions of the models, the performance of the models under the $\mathbb{Q}$ measure, the conditional volatility of yields and the term premia implied by the models, linear projection of yields based on the models and the unconditional mean and standard deviation of yields implied by the models.

7.1 Comparison of one factor models

To gain further insight, univariate models $\mathbb{A}_1^{sgs}(1)$, $\mathbb{AG}_1(1)$, $\mathbb{A}_0(1)$, $\boldsymbol{Q}(1)$ have been estimated through Kalman Filters using only the 1 year yield and the results are reported in Figures 1 and 2 and in Table 5. Figure 1 shows Monte Carlo simulated paths of the short rate r_t according to models $\mathbb{A}_1^{sgs}(1)$, $\mathbb{A}_0(1)$, $\boldsymbol{Q}(1)$. These paths are computed using the same random Gaussian shocks and the model parameters estimated through the Kalman Filter. The starting value of the latent factor is found by "inverting" the 1 year yield observed in the first month of the sample. The starting value of the latent factor of the quadratic models is a parameter estimated by the Kalman Filter. Figure 1 shows that models $\mathbb{A}_1^{sgs}(1)$ and $\boldsymbol{Q}(1)$ can predict persistently low and non-negative paths of the short rate r_t.

Figure 2 plots a histogram of the 1 year yield observed monthly and also plots one mixture of conditional densities for each of the models $\mathbb{A}_1^{sgs}(1)$, $\mathbb{AG}_1(1)$, $\mathbb{A}_0(1)$, $\boldsymbol{Q}(1)$. Each mixture is computed as $\sum_{t=1}^{276} \digamma\left(y_{1\cdot 12,t}; E_{t-1}^{\mathbb{P}}(z_{t-1}), \Theta\right)$, where Θ are the parameters estimated by the Kalman Filter and $\digamma\left(y_{1\cdot 12,t}, E_{t-1}^{\mathbb{P}}(z_{t-1}), \Theta\right)$

is the density of $y_{1\cdot 12,t}$, which is the one year (i.e. 12 month) yield at the end of month t, conditional on Θ and on the expected value $E^{\mathbb{P}}_{t-1}(z_{t-1})$ of the latent factor z estimated by the Kalman Filter after observing the 1 year yield at the end of month $t-1$. There are 276 monthly observations of yields in the sample. Each month $\digamma\left(y_{1\cdot 12,t}, E^{\mathbb{P}}_{t-1}(z_{t-1}), \Theta\right)$ is computed through a Monte Carlo simulation with 1000 iterations. Figure 2 shows the mixture of conditional densities also for $\mathbb{A}_0(1)$ and for $\boldsymbol{Q}(1)$, but for these models the latent factor is $\overline{z}$ instead of z. Figure 2 shows that the mixture of conditional densities for model $\mathbb{A}^{sgs}_1(1)$ and that for $\mathbb{AG}_1(1)$ match the histogram of the observed 1 year yield similarly well, in particular when the 1 year yield is close to zero. Instead the mixture of densities for $\boldsymbol{Q}(1)$ displays a hump between 6% and 9% that does not match the histogram of the observed 1 year yield.

Table 5 reports Vuong likelihood ratio tests of the in sample fit of $\mathbb{A}^{sgs}_1(1)$, $\mathbb{AG}_1(1)$, $\mathbb{A}_0(1)$, $\boldsymbol{Q}(1)$ to the observed 1 year yield. These Vuong tests show that $\mathbb{A}_0(1)$, which is the only one of these four models that allows the short rate to turn negative, is outperformed by the other three models. Among the other three models none significantly outperforms or under-performs. These results are confirmed by the AIC in Table 5, although quadratic model $\boldsymbol{Q}(1)$ fares slight better than the other models according to this metric. Panel C of Table 5 also reports another set of root mean squared errors (RMSE's), where the errors are the differences between the one 1 yield forecasted by a model (i.e. $\mathbb{A}^{sgs}_1(1)$, $\mathbb{AG}_1(1)$, $\mathbb{A}_0(1)$ or $\boldsymbol{Q}(1)$) and its observed value. The said forecast is produced by the Kalman Filter after observing the 1 year yield of the previous month. The RMSE's of Panel C of Table 5 are "out of sample" in that they are computed over the second half of the sample, i.e. the last 138 months, after the parameters of each model have been estimated in the first half of the sample, i.e. in the first 138 months. These "out of sample" RMSE's are lowest for $\mathbb{A}_1^{sgs}(1)$ and $\mathbb{AG}_1(1)$ at 18.1 basis points. Overall the one factor SGS and AG models seem to match the one year yield similarly well.

[Table 5]

[Figure 1]

[Figure 2]

7.2 Performance under the $\mathbb{Q}$ measure of calibrated three factor affine models

Panel D of Table 5 shows the RMSE's of three factor affine models $\mathbb{AG}_3(3)$, $\mathbb{AG}_1(3)$, $\mathbb{A}_3^{sgs}(3)$, $\mathbb{A}_1^{sgs}(3)$, $\mathbb{A}_0(3)$ when calibrated to the whole sample of observed yields. Calibration of a model disregards information in the time series of yields, i.e. the dynamics of yields under the $\mathbb{P}$ measure, and determines the model parameters under the $\mathbb{Q}$ measure that minimise the sum of squared pricing errors across all maturities and months in the sample, while assuming that the 1 year, 10 year and 20 year yields are observed perfectly, so as to infer the values of the latent factors for every month. Every month each model perfectly matches the 1 year, 10 year and 20 year yield, and predicts yields of all other maturities for that same month, not for future months. Therefore the RMSE's in Panel D refer only to the cross sectional yield predictions of the models. The errors are differences between observed yields and model predicted yields for all maturities. Panel D shows that AG models $\mathbb{AG}_3(3)$ and $\mathbb{AG}_1(3)$ are slightly more accurate than corresponding SGS models $\mathbb{A}_3^{sgs}(3)$ and $\mathbb{A}_1^{sgs}(3)$, but the differences between the RMSE's of these models are small, i.e. less than half a basis point. The two models that rule out negative yields, namely $\mathbb{A}_3^{sgs}(3)$ and $\mathbb{AG}_3(3)$, have lower RMSE's than the others models, which do not rule out negative yields.

Figure 3 plots the time series of another set of RMSE's for the same calibrated models $\mathbb{AG}_3(3)$, $\mathbb{AG}_1(3)$, $\mathbb{A}_3^{sgs}(3)$, $\mathbb{A}_1^{sgs}(3)$, $\mathbb{A}_0(3)$. These RMSE's are computed for each model across the 20 yield maturities each month. $\mathbb{AG}_3(3)$ displays the highest RMSE during part of the low yields period after the great recession of 2008-2009, but later also displays the lowest RMSE while yields are still close to zero. Note that $\mathbb{A}_3^{sgs}(3)\,v$, $\mathbb{A}_1^{sgs}(3)\,v$ are the same as $\mathbb{A}_3^{sgs}(3)$, $\mathbb{A}_1^{sgs}(3)$ for the purposes of the said calibration. **The** yields predicted by $\boldsymbol{Q}(3)$**, which are not monotonic with respect to the latent factors, are cumbersome to "invert" to determine the said latent factors** as required by **the** calibration. Overall three factor SGS and AG models seem to predict the cross section of yields similarly well.

[Figure 3]

7.3 Conditional volatility of yields

Figure 4 plots the monthly time series of the conditional volatility of the 1 year yield according to the data and the models. Figure 5 does the same for the conditional volatility of the 10 year yield. The (annualised) conditional volatility of the 1 year yield at the end of month t according to the data is computed as $\sigma^{data}_{1\cdot 12,t}\sqrt{252}$, where $\sigma^{data}_{1\cdot 12,t}$ is the standard deviation of daily changes in the 1 year (i.e. 12 month) yield computed over the last 21 trading days. Eikon provides daily observations of yields. On average there are about 252 trading days in one year in our sample. Effectively $\sigma^{data}_{1\cdot 12,t}$ is the daily volatility of the 1 year yield realised over the rolling window of the last 21 trading days at the end of month t. The conditional volatility of the 10 year yield according to the data is computed similarly.

The (annualised) conditional volatility of the 1 year yield implied by model $\mathbb{A}_3^{sgs}(3)$ at the end of month t is approximately computed as $\frac{-\mathbf{B}'_m \cdot \sqrt{diag(\mathbf{z}_t)} \cdot (\sigma_1, \sigma_2, \sigma_3)'}{m\Delta}$ with $m = 12$ and $\Delta = 1/12$. This approximation omits terms of order $O(\Delta^2)$ with little loss in accuracy. For the other models and for the 10 year maturity the model-implied conditional volatility of yields is computed in a similar way, mutatis mutandis.

Figures 4 and 5 show that the time series of conditional volatilities implied by $\mathbb{A}_1^{sgs}(3)\,v$ and $\mathbb{A}_1^{sgs}(3)$ are closer to realised volatilities than the conditional volatilities implied by $\mathbb{AG}_1(3)$. The latter are just too high. Then Figure 4 shows that the time series of conditional volatilities for the 1 year yield implied by $\mathbb{A}_3^{sgs}(3)$ are closer to realised volatilities than the conditional volatilities implied by $\mathbb{AG}_3(3)$ and by $\mathbb{A}_3^{sgs}(3)\,v$, which are just too low. Figure 5 shows that $\mathbb{A}_3^{sgs}(3)\,v$ predicts too low volatility also for the 10 year yield, while $\mathbb{A}_3^{sgs}(3)$ matches the realised volatility of the 10 year yield **slightly worse** than

$\mathbb{AG}_3(3)$. If we define the 10 year yield "volatility error" as the difference between $\sigma^{data}_{10\cdot12,t}\sqrt{252}$ and the model implied volatility of the 10-year yield at the end month t, the corresponding "volatility RMSE" for the 10 year maturity across the 276 months in the sample is **46.5** basis points for $\mathbb{A}^{sgs}_3(3)$ and **41.4** basis points for $\mathbb{AG}_3(3)$. **The** SGS models seem to match the rolling windows of realised volatilities **better** than their corresponding AG models**, as** confirmed by Panel **F** of Table 5, **but** $\mathbb{AG}_3(3)$ **tends to best match the volatility of long term yields, such as the 10 year yield**.

Panel E of Table 5 summarises Figures 4 and 5 through correlations between realised yield volatilities and model implied yield volatilities, and through regressions of the former volatilities on the latter volatilities for one year and ten year yields. For the one year yield the said correlations between market and model volatilities are in the range of 0.35 to 0.44 and are quite similar for SGS and corresponding AG models. Instead for the ten year yield $\mathbb{AG}_3(3)$ outperforms all other models with a correlation of 0.43, while the correlations of SGS models under-perform those of AG models. Panel E also reports the R^2 and point estimates of the intercept and slope coefficients from regressing the realised volatility of the one (ten) year yield on model implied one (ten) year yield volatilities. For the one year yield volatility the slope estimates are respectively closer to 1 for SGS models $\mathbb{A}^{sgs}_3(3), \mathbb{A}^{sgs}_1(3)$ than for the corresponding AG models $\mathbb{AG}_3(3), \mathbb{AG}_1(3)$, and the intercept estimate for $\mathbb{AG}_1(3)$ is more negative than for $\mathbb{A}^{sgs}_1(3)$. This

confirms that SGS models beat AG models in matching the one year yield volatility. Instead for the ten year yield volatility the slope estimate is much closer 1 for $\mathbb{AG}_3(3)$ than for $\mathbb{A}_3^{sgs}(3)$, thus confirming that the former better matches the dynamics of the ten year yield volatility.

Finally Panel F of Table 5 reports clearly lower, i.e. better, "volatility" RMSE for SGS models than for AG models, as yields of all maturities are considered. When computing "volatility RMSE" **the** "volatility error" **for** the x-year maturity is again the difference between $\sigma_{x\cdot 12,t}^{data}\sqrt{252}$ and the model implied volatility of the x-year yield at the end of every month t. "**Volatility**" RMSE in Panel F of Table 5 is computed across all yield maturities and months for each model. In Panel F models $\mathbb{A}_1^{sgs}(3)$ and $\mathbb{A}_1^{sgs}(3)\,v$ display the lowest "volatility" RMSE, 41 basis points, followed by $\mathbb{A}_3^{sgs}(3)$ with 46 basis points and then by $\mathbb{AG}_3(3)$ with 60 basis points, while $\mathbb{AG}_1(3)$ seems very disappointing with 240 basis points. Relatively disappointing is also $\mathbb{A}_3^{sgs}(3)\,v$ with "volatility RMSE" of 79 basis points. The **Esscher** transform makes the factors of $\mathbb{A}_3^{sgs}(3)\,v$ correlated under $\mathbb{P}$ even if not under $\mathbb{Q}$, but this does not help $\mathbb{A}_3^{sgs}(3)\,v$ fit yields conditional volatility better then $\mathbb{A}_3^{sgs}(3)$, quite the opposite. The **Esscher** transform does not help $\mathbb{A}_1^{sgs}(3)\,v$ fit the conditional volatility of yields better than $\mathbb{A}_1^{sgs}(3)$ either, despite being model flexible. Panel F of Table 5 confirms that SGS models match the dynamics of yields volatility better than corresponding AG models, but this outperformance is not due to the **Esscher** transform.

Overall SGS models seem to match the conditional volatility of **all** yields more closely than AG models, even though no model closely matches the dynamics of yields volatility **in** the data.

[Figure 4]

[Figure 5]

7.4 The time series of term premia

Figures 6 displays the time series of the term premium for the 1 year maturity predicted by models $\mathbb{AG}_3(3)$, $\mathbb{AG}_1(3)$, $\mathbb{A}_3^{sgs}(3)$, $\mathbb{A}_1^{sgs}(3)$, $\mathbb{A}_3^{sgs}(3)\,v$, $\mathbb{A}_1^{sgs}(3)\,v$. Figure 7 does the same for the 10 year maturity term premium. The term premium at the end of month t for the discount bond whose maturity is at the end of month $t+m$ is

$$c_{t,m}^{*} = y_{t,m} - \frac{\Delta \cdot E_t^{\mathbb{P}}\left[\sum_{i=0}^{m-1} r_{t+i}\right]}{\Delta \cdot m}.$$

Therefore the term premium at the end of month t for the 10 year maturity is $y_{10\cdot 12,t} - \frac{\Delta \cdot E_t^{\mathbb{P}}\left[\sum_{i=0}^{10\cdot 12-1} r_{t+i}\right]}{10}$, where for the models in this paper $E_t^{\mathbb{P}}\left[\sum_{i=0}^{10\cdot 12-1} r_{t+i}\right]$ can be computed by solving a system of difference equations. The term premium for all other maturities is computed similarly, mutatis mutandis. The term premium just defined is not "pure", because it reflects also the effect of bond price convexity with respect to the latent factors, but is adopted in this paper because it is the definition of term premium typically used in tests using linear projection of yields, which are reported below.

The term premia predicted by $\mathbb{A}_0(3)$ and $\boldsymbol{Q}(3)$ are not reported because they were so disappointing as to fall out of the scale of Figures 6 and 7. Figures 6 and 7 show that the term premium predicted by $\mathbb{AG}_1(3)$ is often strongly negative and less plausible than that of the other models. Figure 6 also shows that the 1 year term premium predicted by $\mathbb{AG}_3(3)$ is persistently negative at around -20 basis points after the 2008-2009 great recession. Instead the 1 year term premium predicted by the SGS models drops close to zero, but usually not below zero, after the great recession. This evidence confirms that the 1 year maturity is often viewed as a safe heaven by investors "flying to quality and liquidity" and by investors who "park money without chasing yield". Then Figure 7 shows that $\mathbb{AG}_3(3)$ and, to a lesser extent, also $\mathbb{A}_1^{sgs}(3)\,v$ predict negative 10 year term premia after the great recession, which appear less plausible than the negative 1 year term premium, since 10 year bonds are not typically viewed as safe heavens. Apart from $\mathbb{A}_1^{sgs}(3)\,v$, the other SGS models predict non-negative 10 year term premia after the 2008-2009 great recession, although such premia are lower than before the great recession. Overall the time series of term premia for the 1 year and the 10 year maturities lend more support to SGS models than to corresponding AG models.

[Figure 6]

[Figure 7]

7.5 Linear projections of yields and unconditional moments of yields

For each of the affine models in Table 3, Table 6 presents the slope coefficient estimates from the linear projections of yields LPYi and LPYii of Dai and Singleton (2002). LPYi regression equations are

$$y_{t+1,m-1} - y_{t,m} = \delta_m + \phi_m \frac{y_{t,m} - y_{t,1}}{m-1} + \varepsilon_{t,m}$$

where δ_m and δ_m are two constants, and $\varepsilon_{t,m}$ is the regression error term with zero mean. Again m is the yield maturity measured in months. This regression predicts the monthly change in a discount bond yield using the slope of the yield curve $\frac{y_{t,m}-y_{t,1}}{m-1}$ as predictor. If the expected excess bond return is constant over time, i.e. if δ_m is constant over time in the population, then $\phi_m = 1$ in the population, but it it notorious that LPYi regression estimates of ϕ_m that use observed yields significantly differ from 1. LPYii regression equations are

$$y_{t+1,m-1} - y_{t,m} - \left(c^*_{t,m-1} - c^*_{t+1,m-1}\right) + \frac{p^*_{t,m-1}}{m-1} = \phi_m \frac{y_{t,m} - y_{t,1}}{m-1} + \varepsilon_{t,m}$$

$$p^*_{t,m} = f_{t,m} - E^{\mathbb{P}}_t \left[r_{t+m}\right], \quad f_{t,m} = -\frac{1}{\Delta} \ln \frac{P_{t,m+1}}{P_{t,m}}.$$

For each of the affine models in Table 3, the term premia $c^*_{t,m-1}, c^*_{t+1,m-1}$ and the forward premium $p^*_{t,m-1}$ are computed each month using the parameters estimated through Kalman Filter. LPYi regressions measure how well a model matches the dynamics of yields under the real measure $\mathbb{P}$, while LPYii regressions measure how well a model matches the dynamics of yields under the risk-neutral measure $\mathbb{Q}$. Following Dai and Singleton (2002), the population slope coefficients of the LPYi regressions were estimated for each affine model in Table 3 while assuming that the parameters estimated through Kalman Filter coincide with the population parameters. These estimates are reported for each model in Table 6 in the rows "LPYi population". Under the same assumption of estimated parameters coinciding with population parameters, 1000 time series of 276 monthly term structures of yields were simulated for each of the affine models in Table 3. 276 is the number of months in our sample. Then for each of these 1000 simulated time series an LPYi regression was run. Table 6 reports the mean, the 5th percentile and the 95th percentile of these 1000 estimates for each slope coefficient ϕ_m of LPYi regressions for each model and each yearly maturity. These are respectively reported for each model and each maturity in the rows "LPYi Mean, LPYi 5th percentile, LPYi 95th percentile".

For all models Table 6 reports large differences between population slope coefficient estimates and the corresponding mean estimates computed on simulated yields. These differences are due to small sample bias in the estimates of the LPYi regression slope coefficient. The said bias affects all models, but is smaller for SGS models than for AG models. Then the estimated population slope coefficient of the LPYi regression is generally outside the range between the 5th and the 95th percentile for the AG models, but inside the said range for the SGS models. This is the case for almost all maturities. For these reasons LPYi regressions seem to provide more support to SGS models than to AG models.

For each model we can also compute $dslope = \frac{1}{19} \cdot \sum_{i=2}^{20} abs\left(\phi_{12 \cdot i}^{LPYi-data} - \phi_{12 \cdot i}^{LPYi\ Mean-model}\right)$, where $abs\,(..)$ is the absolute value operator. *dslope* is average across 19 maturities of the absolute value of the difference between the LPYi slope estimated on the data, which is $\phi_{12\cdot i}^{LPYi-data}$ for the i-year maturity, and the mean of the LPYi slopes estimated on model simulated yields, which is $\phi_{12\cdot i}^{LPYi\ Mean-model}$ for the i-year maturity. *dslope* is 43.2 for $\mathbb{AG}_3\,(3)$, 3.7 for $\mathbb{AG}_1\,(3)$, 1.87 for $\mathbb{A}_1^{sgs}\,(3)$ and $\mathbb{A}_3^{sgs}\,(3)$, 1.61 for $\mathbb{A}_1^{sgs}\,(3)\,v$, 1.4 for $\mathbb{A}_3^{sgs}\,(3)\,v$ and 1.36 for $\mathbb{A}_0\,(3)$. Therefore *dslope* shows that SGS models match the slopes of LPYi regressions run on the data more closely than AG models. Again LPYi regressions seem to provide more support to SGS models than to AG models.

In Table 6 the rows "LPYii" report the estimates of ϕ_m from LPYii regressions for each maturity and each model. For a model that properly describes the dynamics of yields under $\mathbb{Q}$, the estimated term premia $c^*_{t,m-1}, c^*_{t+1,m-1}$ and forward premium $p^*_{t,m-1}$ should be such that the LPYii regressions give estimates of the slope coefficient ϕ_m close to 1 for all maturities. This is the case for none of the estimated models and is not surprising. Andreasen and Meldrum (2019) show that the protracted stay of US yields close to the zero lower bound after the great recession worsens the estimates of the slope coefficient ϕ_m of LPYi and LPYii regressions. However the slope coefficient estimates of LPYii regressions tends are closer to 1 for the SGS models than for the AG models for all maturities other than the one year maturity. In this sense LPYii regressions provide more support to SGS models than to AG models.

For each model in Table 3 and for each maturity, the rows "unconditional mean" and "unconditional stdev" in Table 8 report the unconditional mean and standard deviation of yields, which are computed using the parameters estimated through the Kalman Filter. For all models the unconditional mean and standard deviation of yields appear realistic, except for $\mathbb{AG}_3\,(3)$, whose unconditional standard deviation of yields is far too high and whose unconditional mean of yields also seems too high, i.e. between 16% and 24%. $\mathbb{A}_1^{sgs}\,(3)\,v$ exhibits the lowest unconditional standard deviation of yields for all maturities, around 2%. The said standard deviation is relatively high for $\mathbb{A}_0\,(3)$, around 10%. On the other hand $\mathbb{AG}_1\,(3)$ exhibits the lowest unconditional mean of yields, hovering around 0% for all maturities, which seems too low. We conclude that the unconditional mean and unconditional standard deviation of yields of the four SGS models seem more plausible than those of the AG models.

Overall the linear projections of yields and the unconditional moments of yields lend more support to SGS models than to AG models.

[Table 6]

7.6 Conclusions from the tests

The empirical results of this section highlight that the main advantage of SGS models over AG models is the better fit to the dynamics of yields under the real measure $\mathbb{P}$. AG models can fit the cross section of yields slightly better than SGS models, as shown in Panel D of Table 5. Therefore AG models can fit the dynamics of yields under the risk measure $\mathbb{Q}$ slightly better than SGS models, but this modest advantage seems to be outweighed by the better fit of SGS models

to the dynamics of yields under the real measure $\mathbb{P}$, when models are estimated through Kalman Filter. This conclusion emerges especially from the information criterion (AIC) in Table 3, from the Vuong tests in Table 4, from the fit to yield volatilities in Panel F of Table 5, and from the volatility plot in Figure 4. This same conclusion is also supported by the term premia in Figures 6 and 7, by the linear projections of yields and by the model implied unconditional mean and standard deviation of yields in Table 6.

Finally the SGS and AG models whose factors are all non-negative, namely $\mathbb{A}_3\,(3)$ and $\mathbb{AG}_3\,(3)$, **report markedly lower RMSE in the calibration of Panel D of Table 5. The reason is that these models, which rule out negative yields, can fit yields during the zero lower bound period better than the other affine models, which do not rule out negative yields.**

8 Conclusion

This paper has presented discrete time affine term structure models (DTATSM) based on affine factor processes with squared Gaussian shocks (SGS). The continuous time limit of SGS models is the same as that of popular autoregressive gamma (AG) models. In SGS models flexible risk premia can alter even the conditional correlation of factors and yields. The empirical evidence from US yields shows that SGS models tend to perform better than corresponding AG models when predicting the conditional volatility of yields, term premia and the unconditional mean and standard deviation of yields.

References

[1] **Ahn D., Dittmar R. and Gallant R., 2002, "Quadratic term structure models: theory and evidence", The Review of Financial Studies 15, n.1, 243-288**.

[2] Andreasen M.M. and Meldrum A., 2019, "A Shadow Rate or a Quadratic Policy Rule? The Best Way to Enforce the Zero Lower Bound in the United States", Journal of Financial and Quantitative Analysis 54, n. 5, 2261–2292.

[3] Bauer M.D., 2018, "Restrictions on Risk Prices in Dynamic Term Structure Models", Journal of Business & Economic Statistics 36, n.2, 196-211.

[4] **Cochrane J.H. and Piazzesi M., 2008, "Decomposing the Yield Curve", unpublished manuscript, Chicago Booth School of Business.**

[5] Creal D.D. and Wu J.C., 2015, "Estimation of affine term structure models with spanned or unspanned stochastic volatility", Journal of Econometrics 185, 60–81.

[6] Dai Q. and Singleton K., 2000, "Specification Analysis of Affine Term Structure Models", The Journal of Finance 55, n.5, 1943-1978.

[7] Dai Q. and Singleton K., 2002, "Expectation puzzles, time-varying risk premia, and affine models of the term structure", Journal of Financial Economics 63, 415–441.

[8] Gourieroux, C. and Jasiak J., 2006, "Autoregressive Gamma Processes", Journal of Forecasting 25, 129–52.

[9] Gourieroux C. and Monfort A., 2011, "Bilinear term structure model", Mathematical Finance 21, n. 1, 1–19.

[10] Gourieroux C., Monfort A., Pegoraro F. and Renne JP., 2014, "Regime Switching and Bond Pricing", Journal of Financial Econometrics 12, n. 2, 237–277.

[11] Kim D.H. and Singleton K.J., 2012, "Term structure models and the zero lower bound: an empirical investigation of Japanese yields", Journal of Econometrics 170, 32-49.

[12] Le A., Singleton K. and Dai Q., 2010, "Discrete-time Affine^Term Structure Models with Generalized Market Prices of Risk", Review of Financial Studies 23, 2184-227.

[13] **Leippold M. and Wu L., 2003, "Design and estimation of quadratic term structure models", European Finance Review 7, n.1, 47-73**.

[14] Monfort A. and Pegoraro F., 2012, "Asset pricing with Second-Order Esscher Transforms", Journal of Banking & Finance 36, 1678–1687.

[15] Monfort A., Pegoraro F., Renne J. and Roussellet G., 2017, "Staying at zero with affine processes: An application to term structure modelling", Journal of Econometrics 201, 348-366.

[16] Realdon M., 2006, "Quadratic term structure models in discrete time", Finance Research Letters 3, issue 4, 277-289.

[17] Sun T., 1992, "Real and nominal interest rates: a discrete time model and its continuous time limit", Review of Financial Studies 5, n.4, 581-611.