GMM Estimation of Affine Term Structure Models

Jaroslava Hlouskova
Leopold Sögner

IRIHS INSTITUTIONAL REPOSITORY AT IHS
Vienna

INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

GMM Estimation of Affine Term Structure Models[*]

Jaroslava Hlouskova   Leopold Sögner[†]

August 26, 2015

Abstract

This article investigates parameter estimation of affine term structure models by means of the generalized method of moments. Exact moments of the affine latent process as well as of the yields are obtained by using results derived for $p-$polynomial processes. Then the generalized method of moments, combined with Quasi-Bayesian methods, is used to get reliable parameter estimates and to perform inference. After a simulation study, the estimation procedure is applied to empirical interest rate data.

Keywords: Affine term-structure models, GMM.

JEL: C01, C11, G12.

1 Introduction

This article is concerned with parameter estimation and inference in affine term structure models. We use results of Cuchiero et al. (2012) on $p-$polynomial processes to obtain the exact conditional moments of a latent affine process driving the term structure. By assuming a stationary affine process, we obtain not only the exact moments of a vector of yields with various maturities but also the first-order auto-covariance matrices of the yields and the squared yields. Then we estimate the model parameters by means of the *Generalized Method of Moments* (GMM) introduced in Hansen (1982), where Quasi-Bayesian methods (see Chernozhukov and Hong, 2003) are used to minimize the GMM distance function. A further contribution of this paper is a rigorous study on testing market price of risk specifications discussed in quantitative finance literature. By considering the Wald test, we observe that test statistics obtained from output provided by Quasi-Bayesian methods strongly outperform test statistics which are obtained by standard procedures with respect to power and size.

Affine term structure models have their origin in the univariate models of Vasicek (1977) and Cox et al. (1985). The performance of these models and similar univariate setups were already investigated for example in Aït-Sahalia (1996a) and Aït-Sahalia (1996b). The articles show that these univariate parametric models inadequately describe the interest rate dynamics. Based on this finding Aït-Sahalia (1996a), Aït-Sahalia (1996b) as well as Stanton (1997) proposed non-parametric interest rates models. As an alternative, Dai and Singleton (2000) and Dai and Singleton (2003) favored multivariate settings to circumvent the shortcomings of univariate models. This alternative modeling approach has the advantage that a mathematical framework, where bonds and derivatives can be priced in a straightforward way, is available.

Let us briefly discuss some literature on the performance of different estimation approaches: Regarding parameter estimation, Zhou (2001) studied the efficient method of moments (EMM), the GMM, the quasi-maximum likelihood estimation ($QMLE$) and the maximum likelihood estimation (MLE) for the Cox et al. (1985) model. In his study the author assumes that the instantaneous interest rate, driven by a square root process, can be observed. The most efficient results are observed for the MLE, which is

followed by the $QMLE$ and the EMM.[1] Regarding the GMM, this method performs well if the sample size is sufficiently large. In addition, Zhou (2003) constructed a GMM estimator by deriving moments for univariate latent processes by applying Ito's formula (under the same assumption that the instantaneous interest rate can be observed). This estimator has been compared to the ML estimator. In contrast to Zhou (2001), in this setup the GMM estimator performs quite well in the finite sample compared to the maximum likelihood estimation.

More recent literature has proposed different frequentist and Bayesian approaches to estimate the parameters of multivariate affine term structure models. Bayesian methods have been applied almost recently in Chib and Ergashev (2009), an earlier application is e.g. Frühwirth-Schnatter and Geyer (1996). Regarding Bayesian estimation methods, Jones (2003) pointed out that strong priors are necessary to estimate the parameters in the case of a low degree of mean reversion (i.e., high persistence) of the stochastic process. MLE has been performed in a three factor Gaussian model (an $\mathbb{A}_0(3)$ model in the terminology of Dai and Singleton, 2000) by Hamilton and Wu (2012).

Additional articles on parameter estimation for affine models are e.g. Diebold et al. (2006), Duffee (2011), Aït-Sahalia and Kimmel (2010), Egorov et al. (2011) and Joslin et al. (2010). An overview is provided in Piazzesi (2010). A further approach is to approximate the transition density of the affine process via approximations of the Chapman/Kolmogorov forward equation. This approach has been explored in series of papers by Aït-Sahalia (see, e.g., Aït-Sahalia, 2002; Aït-Sahalia and Kimmel, 2010). Filipović et al. (2013) used the moments obtained in Cuchiero et al. (2012) to construct additional likelihood expansions.

In contrast to a lot of other approaches already used in the literature, we use the *exact* moments of the yields observed, arising from a multivariate affine term structure model. Neither an approximation of the moments (such as an approximation via the solution of the stochastic differential equation) nor an approximation of the likelihood is required. Since we have to minimize a GMM distance function in more than twenty parameters, GMM estimation is nontrivial. To account for this problem, we use Quasi-Bayesian methods developed in Chernozhukov and Hong (2003). As standard errors of parameter estimates are byproducts of this estimation routine, we apply them in parameter testing, where we observe

---

[1]For stochastic volatility models Andersen et al. (1999) have shown that the EMM estimator has almost the same efficiency as the maximum likelihood estimator.

rejection rates of the true null hypothesis to be close to the theoretical significance levels. By contrast, when using standard routines to estimate the asymptotic covariance matrix of the unknown parameter vector, the performance of the Wald test, measured in terms of power and size, is very poor.

This paper is organized as follows: Section 2 introduces affine term structure models. Section 3 applies results obtained in mathematical finance literature to calculate the moments of the latent process driving the yields and then derives the moments of the yields observed. Section 4 describes the small sample properties of the GMM estimator, while Section 5 applies the estimator to empirical data. Finally, Section 6 offers conclusions.

2 Affine Models

This section provides a brief description of affine models, which is mainly based on Filipović (2009). Consider the state space $\mathscr{S} = \mathbb{R}_+^m \times \mathbb{R}^n \subset \mathbb{R}^d$, where $m, n \geq 0$, $m + n = d$, and the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. With $\mathbf{X}(t) \in \mathbb{R}^d$, the stochastic process in continuous time $(\mathbf{X}(t))_{t \geq 0}$ is generated by the following affine stochastic differential equation

$$d\mathbf{X}(t) = \left(\mathbf{b}^P + \boldsymbol{\beta}^P \mathbf{X}(t) \right) dt + \boldsymbol{\rho}(\mathbf{X}(t)) d\mathbf{W}^P(t) , \tag{1}$$

where $\mathbf{b}^P$ is a d−dimensional vector and $\boldsymbol{\beta}^P$ and $\boldsymbol{\rho}(\mathbf{x})$ are $d \times d$ matrices. The $d \times d$ diffusion term $\mathbf{a}(\mathbf{x})$ is defined such that $\mathbf{a}(\mathbf{x}) = \boldsymbol{\rho}(\mathbf{x})\boldsymbol{\rho}(\mathbf{x})' = \mathbf{a} + \sum_{i=1}^d x_i \boldsymbol{\alpha}_i$, where $\mathbf{a}$ and $\boldsymbol{\alpha}_i$, $i = 1, \ldots, d$, are $d \times d$ matrices. $\mathbf{W}^P(t)$ is a d−dimensional standard Brownian motion. For more details the reader is referred to Appendix A. In an affine environment the *instantaneous interest rate (short rate, $r(t) \in \mathbb{R}$)* follows from

$$r(t) = \gamma_0 + \boldsymbol{\gamma}_x' \mathbf{X}(t) , \tag{2}$$

where γ_0 is a scalar and $\boldsymbol{\gamma}_x$ is a d−dimensional vector. We consider an arbitrage free market, where $\mathbb{P}$ is the empirical measure and $\mathbb{Q}$ is an equivalent martingale measure. We assume that the process $(\mathbf{X}(t))_{t \geq 0}$

is affine also in the measure $\mathbb{Q}$, such that

$$d\mathbf{X}(t) = \left(\mathbf{b}^Q + \boldsymbol{\beta}^Q \mathbf{X}(t)\right) dt + \boldsymbol{\rho}(\mathbf{X}(t))d\mathbf{W}^Q(t) , \qquad (3)$$

where $\mathbf{W}^Q(t)$ is a $d-$dimensional standard Brownian motion under $\mathbb{Q}$ measure.

By equations (1) and (3), the stochastic process $(\mathbf{X}(t))_{t \geq 0}$ is affine in both measures. While the diffusion parameters $(\mathbf{a}, \boldsymbol{\alpha}_i, i = 1, \ldots, d)$ remain the same under both measures, we have to consider parameters $\mathbf{b}^P$, $\boldsymbol{\beta}^P$, $\mathbf{b}^Q$ and $\boldsymbol{\beta}^Q$, in both measures $\mathbb{P}$ and $\mathbb{Q}$. This specification, namely equations (1) and (3), is called *the extended affine market price of risk specification*, and its mathematical foundation is provided in Cheridito et al. (2007). These authors also show by means of the Girsanov theorem that $\mathbf{W}^Q(t) = \mathbf{W}^P(t) + \int_0^t \boldsymbol{\phi}(\mathbf{X}(s))ds$. For the affine class

$$\boldsymbol{\phi}(\mathbf{X}(t)) = (\boldsymbol{\rho}(\mathbf{X}(t)))^{-1} \left(\mathbf{b}^P - \mathbf{b}^Q + \left(\boldsymbol{\beta}^P - \boldsymbol{\beta}^Q\right) \mathbf{X}(t)\right) , \qquad (4)$$

where $\boldsymbol{\phi}(\mathbf{X}(t)) \in \mathbb{R}^d$. The stochastic process $(\boldsymbol{\phi}(\mathbf{X}(t)))_{t \geq 0}$, is called *market price of risk process*.

Remark 1. To observe how the market price of risk process $(\boldsymbol{\phi}(\mathbf{X}(t)))_{t \geq 0}$ is connected to risk premia, Cochrane (2005)[p. 339] provides a formal relationship between the process $(\boldsymbol{\phi}(\mathbf{X}(t)))_{t \geq 0}$ and the (instantaneous) Sharpe ratio.

We also assume that the process $(\mathbf{X}(t))$ satisfies the admissibility conditions (under both measures), which ensure that the process $(\mathbf{X}(t))$ does not leave the state space $\mathscr{S}$ (see Filipović, 2009, Theorem 10.2 and Appendix E). Next, we define the index sets $I = \{1, \ldots, m\}$ and $J = \{m+1, \ldots, m+n\}$, where $m + n = d$. Let $\mathbf{b}_I = (b_1, \ldots, b_m)'$ and $\boldsymbol{\beta}_{II} = \boldsymbol{\beta}_{1:m,1:m}$.[2] This notation, the admissibility restrictions (see Appendix E), the short-rate model (2) and the condition $\mathbb{E}\left(\exp(-\int_0^{\bar{\tau}} r(z)dz)\right) < +\infty$, for some $\bar{\tau} \in \mathbb{R}_+$, imply that there exists a unique solution $(\Phi(t,\mathbf{u}), \boldsymbol{\Psi}(t,\mathbf{u})')' \in \mathbb{C} \times \mathbb{C}^d$ of the system of Riccati differential

---

[2]In this article we apply the following notation: For vectors and matrices we use boldface. If not otherwise stated, the vectors considered are column vectors. Given a $r_M \times c_M$ matrix $\mathbf{M}$, the term $\mathbf{M}_{r_a:r_b,c_a:c_b}$ stands for "from row r_a to row r_b and from column c_a to column c_b of matrix $\mathbf{M}$". The abbreviation $\mathbf{M}_{r_a:r_b,:}$ stands for "from row r_a to row r_b of matrix $\mathbf{M}$", while "$:$" stands for all columns, i.e. columns 1 to c_M. In addition, $\mathbf{M}_{r_a:r_b,c_a}$ extracts the elements r_a to r_b of the column c_a. In addition, β_{ij} stands for $[\boldsymbol{\beta}]_{ij}$; $\mathbf{0}_{a \times b}$ and $\mathbf{e}_{a \times b}$ stand for $a \times b$ matrices of zeros and ones; $\mathbf{0}_a$ and $\mathbf{e}_a$ is used to abbreviate $\mathbf{0}_{a \times 1}$ and $\mathbf{e}_{a \times 1}$; $\mathbf{I}_a$ is the $a \times a$ identity matrix, while $\mathbb{I}_{(.)}$ stands for an indicator function. Given a vector $\mathbf{x} \in \mathbb{R}^n$, $diag(\mathbf{x})$ transforms $\mathbf{x}$ into a $n \times n$ diagonal matrix. 2 E-3 stands for $2 \cdot 10^{-3} = 0.002$.

equations

$$\partial_t \Phi(t, \mathbf{u}) = \tfrac{1}{2} \left(\mathbf{\Psi}_J(t, \mathbf{u}) \right)' \mathbf{a}_{JJ} \mathbf{\Psi}_J(t, \mathbf{u}) + \left(\mathbf{b}^Q \right)' \mathbf{\Psi}(t, \mathbf{u}) - \gamma_0; \qquad \Phi(0, \mathbf{u}) = 0 \; ,$$

$$\partial_t \Psi_i(t, \mathbf{u}) = \tfrac{1}{2} \left(\mathbf{\Psi}(t, \mathbf{u}) \right)' \boldsymbol{\alpha}_i \mathbf{\Psi}(t, \mathbf{u}) + \left(\boldsymbol{\beta}_i^Q \right)' \mathbf{\Psi}(t, \mathbf{u}) - \gamma_{xi}; \qquad \text{for } i \in I \; , \qquad (5)$$

$$\partial_t \mathbf{\Psi}_J(t, \mathbf{u}) = \left(\boldsymbol{\beta}_{JJ}^Q \right)' \mathbf{\Psi}_J(t, \mathbf{u}) - \boldsymbol{\gamma}_{xJ}; \qquad \mathbf{\Psi}(0, \mathbf{u}) = \mathbf{u} \; ,$$

where $t \in [0, \bar{\tau}]$, $\mathbf{u} \in \imath \mathbb{R}^d$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$, with $\boldsymbol{\beta}_i$ being a $d-$dimensional vector, $i = 1, \ldots, d$ (see Filipović, 2009, Theorem 10.4).[3] This system of ordinary differential equations is used to calculate the time t *price of a zero coupon bond*, $\pi^0(t, \tau)$, with time to maturity τ. The arbitrage free zero coupon model prices $\pi^0(t, \tau)$ and the *model yields* $y^0(t, \tau)$ follow from Filipović (2009)[Corollary 10.2]. That is

$$\pi^0(t, \tau) = \exp \left(\Phi(\tau, \mathbf{0}) + \mathbf{\Psi}(\tau, \mathbf{0})' \mathbf{X}(t) \right) \text{ and}$$

$$y^0(t, \tau) = -\frac{1}{\tau} \log \left(\pi^0(t, \tau) \right) = -\frac{1}{\tau} \left(\Phi(\tau, \mathbf{0}) + \mathbf{\Psi}(\tau, \mathbf{0})' \mathbf{X}(t) \right). \qquad (6)$$

The time to maturity, τ, and $\mathbf{u} = \mathbf{0}$ are the arguments of the functions $\Phi(t, \mathbf{u})$ and $\mathbf{\Psi}(t, \mathbf{u})$ described in (5). The parameters under $\mathbb{Q}$ have to be used to derive $\Phi(\tau, \mathbf{0})$ and $\mathbf{\Psi}(\tau, \mathbf{0})$.

3 Moments and Polynomial Processes

Since the goal of this paper is to estimate the model parameters by means of the GMM, we have to obtain the moments of the yields. Section 3.1 uses a recent theory for polynomial processes to obtain a closed form expression for the moments of the latent process $(\mathbf{X}(t))_{t \geq 0}$. In Section 3.2 we derive the exact moments for the model yields of an affine term structure model with diagonal diffusion term. Finally, Section 3.3 deals with the case of empirical data, when the number of yields observed is larger than the dimension of $(\mathbf{X}(t))_{t \geq 0}$ and thus the yields observed cannot be matched exactly with the model yields derived in (6).

---

[3]Ordinary differential equations similar to (5) have already been investigated in Duffie and Kan (1996) and Duffie et al. (2000).

3.1 Polynomial Processes

Based on the results of Cuchiero et al. (2012) on $p-$polynomial Markov processes, this subsection derives the conditional moments of the latent process $(\mathbf{X}(t))_{t\geq 0}$. Let us consider a time homogeneous Markov processes $(\mathbf{X}(t))_{t\geq 0}$, started at $\mathbf{X}(0) = \mathbf{x} \in \mathscr{S}$, where the state space $\mathscr{S}$ is a closed subset of $\mathbb{R}^d$. The *semigroup* $(\mathscr{P}_t)_{t\geq 0}$ described by

$$\mathscr{P}_t f(\mathbf{x}) = \mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(0) = \mathbf{x}) = \int_{\mathscr{S}} f(\boldsymbol{\zeta})\nu_t(\mathbf{x}, d\boldsymbol{\zeta}) \tag{7}$$

is defined on all integrable functions $f\colon \mathscr{S} \to \mathbb{R}$ with respect to the Markov kernels $\nu_t(\mathbf{x}, \cdot)$. For an affine term structure model we need moments of $(\mathbf{X}(t))$ for a process "started" at $\mathbf{X}(s) = \mathbf{x}$; $t > s$. Given the filtration $(\mathcal{F}_t)_{t\geq 0}$ and the assumption that $(\mathbf{X}(t))$ is a homogeneous Markov process, the conditional expectation of $f(\mathbf{X}(t))$, when the process is started at $\mathbf{X}(s) = \mathbf{x}$, is given by $\mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(s) = \mathbf{x}) = \mathscr{P}_{t-s}f(\mathbf{x})$ (see, e.g., Klenke, 2008, Theorem 17.9).

Next, let $\mathcal{P}_{\leq p}(\mathscr{S})$ be the finite dimensional vector space of polynomials on $\mathscr{S}$ up to degree $p \geq 0$, i.e.

$$\mathcal{P}_{\leq p}(\mathscr{S}) = \left\{ \sum_{k=0}^{p} \boldsymbol{\kappa}_k' \mathbf{x}^k \, \middle| \, \mathbf{x} \in \mathscr{S}, \boldsymbol{\kappa}_k \in \mathbb{R}^{d_k} \right\}$$

$$\text{where } \mathbf{x}^k = \left(\Pi_{j=1}^d x_j^{l_{1j}^{(k)}}, \Pi_{j=1}^d x_j^{l_{2j}^{(k)}}, \ldots, \Pi_{j=1}^d x_j^{l_{d_k j}^{(k)}} \right)' \in \mathbb{R}^{d_k} \text{ and } d_k = \begin{pmatrix} k+d-1 \\ k \end{pmatrix}. \tag{8}$$

For $i = 1, \ldots, d_k$ and $j = 1, \ldots, d$ the exponents $l_{ij}^{(k)}$ in the expression for $\mathbf{x}^k$ satisfy $l_{ij}^{(k)} \in \mathbb{N}_0$ as well as $\sum_{j=1}^d l_{ij}^{(k)} = k$.[4] In affine term structure models the basis of $\mathcal{P}_{\leq p}(\mathscr{S})$ is given by $(1, \mathbf{x}', (\mathbf{x}^2)', \ldots, (\mathbf{x}^p)')'$ and thus its dimension is $N = \sum_{k=0}^p d_k$. In addition, the Markov process $(\mathbf{X}(t))_{t\geq s}$ with $\mathbf{X}(s) = \mathbf{x} \in \mathscr{S}$ is called *p-polynomial* if for all $f(\mathbf{x}) \in \mathcal{P}_{\leq p}(\mathscr{S})$ and $t \geq s$

$$\mathscr{P}_{t-s}f(\mathbf{x}) = \mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(s) = \mathbf{x}) \in \mathcal{P}_{\leq p}(\mathscr{S}). \tag{9}$$

---

[4]For example, for $d = 3$ and $k = 2$ we have the following: $\mathbf{x}^2 = \left(x_1^2, x_1 x_2, x_1 x_3, x_2^2, x_2 x_3, x_3^2 \right)'$, $d_2 = 6$ and thus (i) $l_{11}^{(2)} = 2, l_{12}^{(2)} = 0, l_{13}^{(2)} = 0$, (ii) $l_{21}^{(2)} = 1, l_{22}^{(2)} = 1, l_{23}^{(2)} = 0$, (iii) $l_{31}^{(2)} = 1, l_{32}^{(2)} = 0, l_{33}^{(2)} = 1$, (iv) $l_{41}^{(2)} = 0, l_{42}^{(2)} = 2, l_{43}^{(2)} = 0$, (v) $l_{51}^{(2)} = 0, l_{52}^{(2)} = 1, l_{53}^{(2)} = 1$, (vi) $l_{61}^{(2)} = 0, l_{62}^{(2)} = 0, l_{63}^{(2)} = 2$.

That is to say, if $f(\mathbf{x})$ is polynomial, then the $\mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(s) = \mathbf{x})$ is polynomial as well. Cuchiero et al. (2012)[Theorem 2.7] have shown that a time homogeneous Markov processes $(\mathbf{X}(t))$ is $p-$polynomial if and only if there exists a linear map $\mathbf{A}$ on $\mathcal{P}_{\leq p}(\mathscr{S})$ such that $\mathscr{P}_{t-s}$ restricted on $\mathcal{P}_{\leq p}$ can be written as $\mathscr{P}_{t-s}|_{\mathcal{P}_{\leq p}} = \exp((t-s)\mathbf{A})$.[5] Equipped with this mathematical tool and by means of (7), the conditional expectation $\mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(s) = \mathbf{x})$, for $t > s$ and $f(\mathbf{x}) \in \mathcal{P}_{\leq p}(\mathscr{S})$ can be derived by means of

$$\mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(s) = \mathbf{x}) = \exp((t-s)\mathbf{A})f(\mathbf{x}). \tag{10}$$

The conditional expectations of $f(\mathbf{X}(t))$ given $\mathbf{X}(s) = \mathbf{x}$, can be derived by obtaining the $N \times N$ matrix $\mathbf{A}$, where $N = \sum_{k=1}^{p} d_k$, from the generator (see Cuchiero et al., 2012, Theorem 2.9)

$$\begin{aligned}
\mathcal{G}f(\mathbf{x}) &= \sum_{i=1}^{d} \left(b_i^P + \left[\boldsymbol{\beta}^P\mathbf{x}\right]_i\right) \frac{\partial f(\mathbf{x})}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} [\mathbf{a}(\mathbf{x})]_{ij} \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \\
&= \sum_{i=1}^{d} \left(b_i^P + \boldsymbol{\beta}_{i,1:d}^P\mathbf{x}\right) \frac{\partial f(\mathbf{x})}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} [\mathbf{a}(\mathbf{x})]_{ij} \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}.
\end{aligned} \tag{11}$$

To obtain the moments of $(\mathbf{X}(t))$ we set $f(\mathbf{X}(t)) = \left[\mathbf{X}(t)^k\right]_i$ for $k = 1, \ldots, p$ and $i = 1, \ldots, d_k$. As already stated above, if the dimension of $\mathbf{X}(t)$ is larger than one, then $\mathbf{X}(t)^k = \left(\Pi_{j=1}^{d} X(t)_j^{l_{1j}^{(k)}}, \cdots, \Pi_{j=1}^{d} X_j^{l_{d_k j}^{(k)}}\right)'$, where $l_{ij}^{(k)} \in \mathbb{N}_0$, $\sum_{j=1}^{d} l_{ij}^{(k)} = k \leq p$, $i = 1, \ldots, d_k$ and $j = 1, \ldots, d$. In more detail, we consider the basis $(\mathfrak{e}_1, \ldots, \mathfrak{e}_N) = (1, \mathbf{x}', (\mathbf{x}^2)', \ldots, (\mathbf{x}^p)')$. By applying the extended generator $\mathcal{G}$ to the basis element $\mathfrak{e}_i$, we get the i-th row of the $N \times N$ matrix $\mathbf{A}$ by means of

$$\mathcal{G}\mathfrak{e}_i = \sum_{j=1}^{N} \mathbf{A}_{ij}\mathfrak{e}_j. \tag{12}$$

The left hand side has been calculated by applying (11) to the corresponding basis element. Then $\mathbf{A}_{ij}$

---

[5]Note that $\mathscr{P}_{t-s}|_{\mathcal{P}_{\leq p}} = \exp((t-s)\mathbf{A})$ also solves the Kolmogorov backward equation $\frac{\partial \mathbf{u}(t-s,\mathbf{x})}{\partial t} = \mathcal{G}\mathbf{u}(t-s,\mathbf{x})$, where $\mathcal{G}$ is an extended generator as described in Cuchiero et al. (2012)[Definition 2.3]. This follows from the proof of Cuchiero et al. (2012)[Theorem 2.7].

8

follows from (12) simply by comparing coefficients. This finally results in

$$\mathbb{E}(\mathbf{X}(t)^k \,|\, \mathbf{X}(s) = \mathbf{x}) \;\; = \;\; \left(\mathbf{0}_{d_k \times \sum_{j=0}^{k-1} d_j}, \mathbf{I}_{d_k}, \mathbf{0}_{d_k \times N - \sum_{j=0}^{k} d_j}\right) \exp((t-s)\mathbf{A}) \left(1, \mathbf{x}', (\mathbf{x}^2)', \ldots, (\mathbf{x}^p)'\right)' ,(13)$$

where $\mathbf{I}_{d_k}$ is the $d_k \times d_k$ identity matrix and $t > s$.

3.2 Dai and Singleton (2000)-Models and Moments of the Latent Process

To proceed with an identified model, we work with affine models where the diffusion term can be diago-nalized. For this sub-class, Dai and Singleton (2000) provided sufficient conditions for identification.[6] In this case, the affine process $(\mathbf{X}(t))_{t \geq 0}$ follows the stochastic differential equation

$$d\mathbf{X}(t) \;\; = \;\; (\mathbf{b}^Q + \boldsymbol{\beta}^Q \mathbf{X}(t))dt + \boldsymbol{\Sigma} \sqrt{\mathbf{S}(\mathbf{X}(t))} d\mathbf{W}^Q(t), \quad \text{where}$$

$$S_{ii}(\mathbf{X}(t)) = \mathcal{B}_i^0 + (\mathcal{B}_i^x)'\mathbf{X}(t), \;\; S_{ij}(\mathbf{X}(t)) = 0 \;, \;\; \text{for } i, j = 1, \ldots, d, \; i \neq j \;, \qquad (14)$$

$$\text{and } \boldsymbol{\Sigma} = \text{diag}\left(\Sigma_1, \ldots, \Sigma_d\right) \;\; \text{such that} \;\; \Sigma_i = [\boldsymbol{\Sigma}]_{ii} > 0.$$

Equation (14) is a special case of (3). The elements of the $d-$dimensional vector $\mathcal{B}^0$ are $\mathcal{B}_i^0$. $\mathcal{B}^x$ is a $d \times d$ matrix, where $d \times 1$ vector $\mathcal{B}_i^x$ is the i-th column of this matrix; i.e., $\mathcal{B}^x = (\mathcal{B}_1^x, \ldots, \mathcal{B}_d^x)$ with $\mathcal{B}_i^x = (\mathcal{B}_{1i}^x, \ldots, \mathcal{B}_{di}^x)'$, $i = 1, \ldots, d$. Since $\boldsymbol{\Sigma}$ and $\mathbf{S}(\mathbf{X}(t))$ are diagonal matrices we obtain $\mathbf{a}(\mathbf{X}(t)) = \boldsymbol{\Sigma}^2 \mathbf{S}(\mathbf{X}(t)) = \boldsymbol{\Sigma}^2 \text{diag}\left(\mathcal{B}^0 + (\mathcal{B}^x)'\mathbf{X}(t)\right)$. The diagonal elements of the $d \times d$ diagonal matrix $\mathbf{a}$ are given by $\mathbf{a}_{ii} = \Sigma_i^2 \mathcal{B}_i^0$, $i = 1, \ldots, d$ and the diagonal elements of the $d \times d$ diagonal matrices $\boldsymbol{\alpha}_i$, $i = 1, \ldots, d$, are $\Sigma_1^2 \mathcal{B}_{i1}^x, \Sigma_2^2 \mathcal{B}_{i2}^x, \ldots, \Sigma_d^2 \mathcal{B}_{id}^x$. For $\boldsymbol{\beta}^Q$ and $\mathcal{B}^x$ Dai and Singleton (2000) require

$$\boldsymbol{\beta}^Q = \begin{pmatrix} \boldsymbol{\beta}_{II}^Q & \mathbf{0}_{m \times n} \\ \boldsymbol{\beta}_{JI}^Q \geq 0 & \boldsymbol{\beta}_{JJ}^Q \end{pmatrix} \;\; \text{and} \;\; \mathcal{B}^x = \begin{pmatrix} \mathbf{I}_m & \mathcal{B}_{IJ}^x \geq 0 \\ \mathbf{0}_{n \times m} & \mathbf{0}_{n \times n} \end{pmatrix} \;, \qquad (15)$$

---

[6]For example, the $\mathbb{A}_1(3)$ model, which will be presented in equation (18), has 19 parameter under $\mathbb{Q}$. Dai and Singleton (2000) have shown that the same term structure can be obtained with different parameters. I.e. the model is not identified. Given the Dai and Singleton (2000) conditions for identification, only 14 parameters are allowed to be free parameters. Regarding the diagonal diffusion matrix, Cheridito et al. (2008)[Theorem 2.1] provide conditions where a transformation of a general affine model (1) to an affine model with diagonal $\mathbf{a}(\mathbf{x})$ exists. For $d \leq 3$ this is always the case.

where $m + n = d$. The matrix $\boldsymbol{\beta}_{II}^{Q}$ is of dimension $m \times m$, $\boldsymbol{\beta}_{JI}^{Q}$ is of dimension $n \times m$, $\boldsymbol{\beta}_{JJ}^{Q}$ is of dimension $n \times n$ and $\mathcal{B}_{IJ}^{x}$ is of dimension $m \times n$. As we use findings of Dai and Singleton (2000) we need to relate our notation to the notation of Dai and Singleton (2000), where the drift term of the process $(\mathbf{X}(t))_{t \geq 0}$ is considered in the form $-\boldsymbol{\beta}^{Q}(\boldsymbol{\theta}^{Q} - \mathbf{X}(t))dt$ and thus $\mathbf{b}^{Q} = -\boldsymbol{\beta}^{Q}\boldsymbol{\theta}^{Q}$. In the following, $\boldsymbol{\theta}^{Q} = -\left(\boldsymbol{\beta}^{Q}\right)^{-1}\mathbf{b}^{Q}$ is a vector of dimension d, partitioned into $\boldsymbol{\theta}_{I}^{Q}$ and $\boldsymbol{\theta}_{J}^{Q}$, where the first term is of dimension m while the second term is of dimension n; i.e., $\boldsymbol{\theta}_{I}^{Q} \in \mathbb{R}^{m}$, $\boldsymbol{\theta}_{J}^{Q} \in \mathbb{R}^{n}$, and thus $\boldsymbol{\theta}^{Q} = \left(\left(\boldsymbol{\theta}_{I}^{Q}\right)', \left(\boldsymbol{\theta}_{J}^{Q}\right)'\right)' \in \mathbb{R}^{d}$. The same partition is applied also to $\mathbf{X}(t)$. This yields to the following.

Definition 1 (Dai and Singleton (2000)-*canonical representation* of an $\mathbb{A}_{m}(d)$ model). *Consider (14) with diagonal diffusion matrix and the short-rate model (2). Admissibility and identification require the following:*

(i)-(a) *For $m > 0$ is $\boldsymbol{\beta}^{Q}$ of structure given by (15), where in addition $\beta_{ij}^{Q} \geq 0$ for $1 \leq j \leq m$ and $i \neq j$. Furthermore, $\boldsymbol{\theta}_{I}^{Q} \geq 0$, $\boldsymbol{\theta}_{J}^{Q} = \mathbf{0}$ and $\boldsymbol{\beta}_{II}^{Q}\boldsymbol{\theta}_{I}^{Q} < \mathbf{0}$.*

(i)-(b) *For $m = 0$ is $\boldsymbol{\beta}^{Q}$ a lower (or upper) triangular matrix (Dai and Singleton, 2000, p. 1948).*

(ii) *$\boldsymbol{\Sigma} = \mathbf{I}_{d}$.*

(iii) *γ_{0} and γ_{xi} are unrestricted for $i \in I$, while $\gamma_{xj} \geq 0$ for $j \in J$.*

(iv) *$\mathcal{B}^{0} = (\mathbf{0}_{1 \times m}, \mathbf{e}_{1 \times n})'$ and $\mathcal{B}^{x}$ is of structure provided by (15).*

If the admissibility conditions (i)-(iv) *for the affine process $(\mathbf{X}(t))_{t \geq 0}$ are satisfied, then model (14) with diagonal diffusion term will be called $\mathbb{A}_{m}(d)$ model.*

Definition 1(i)-(a) implies that $b_{i}^{Q} = -\sum_{j=1}^{m}\beta_{ij}^{Q}\theta_{j}^{Q} > 0$, for $i = 1, \ldots, m$, and thus the first m elements of $\mathbf{b}^{Q}$ are strictly positive and the last n elements of $\mathbf{b}^{Q}$ are negative. Namely

$$\mathbf{b}^{Q} = \begin{pmatrix} \mathbf{b}_{I}^{Q} \\ \mathbf{b}_{J}^{Q} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\beta}_{II}^{Q}\boldsymbol{\theta}_{I}^{Q} & > \mathbf{0} \\ -\boldsymbol{\beta}_{JI}^{Q}\boldsymbol{\theta}_{I}^{Q} & \leq \mathbf{0} \end{pmatrix}. \tag{16}$$

This implies that the diagonal elements of $\boldsymbol{\beta}_{II}$ are negative. We slightly deviate from the canonical representation in Definition 1 by assuming $\boldsymbol{\Sigma}$ to be a diagonal matrix with entries $\Sigma_{i} > 0$ and $\boldsymbol{\gamma}_{x} = \mathbf{e}_{d}$.[7]

---

[7]Note that the canonical representation of Dai and Singleton (2000) is one of many representations where the admissibility and identification conditions are met. The Appendix of Dai and Singleton (2000) presents affine linear transformations $\Lambda_{A}\mathbf{X}(t) = \mathbf{L}_{A}\mathbf{X}(t) + \mathbf{l}_{A}$ where the model is still admissible and identified.

Since $\boldsymbol{\theta}_J^Q$ is restricted to zero, not all elements of $\boldsymbol{\beta}^Q$ and $\mathbf{b}^Q$ can be unrestricted. In the estimation procedure we account for this fact by using $\boldsymbol{\theta}^Q$ as a parameter. Then $\mathbf{b}^Q = -\boldsymbol{\beta}^Q \boldsymbol{\theta}^Q$.

Now we apply the tools developed in Section 3.1 to $\mathbb{A}_m(d)$ models. To observe how this works we first derive matrix $\mathbf{A}$ for the Vasicek (1977) and the Cox et al. (1985) model. Then we calculate $\mathbf{A}$ for an $\mathbb{A}_m(d)$ model for arbitrary $0 \le m \le d$ and $d \le 3$. Matrix $\mathbf{A}$, for $d = 3$, is presented in Appendix B, as for $p = 4$ moments its dimension becomes large (35×35).

Let us start with the Vasicek (1977) model, where $d = 1$ and $m = 0$ such that $(X(t))$ follows an Ornstein-Uhlbeck process $dX(t) = (b^P + \beta^P X(t))dt + \Sigma\, dW^P(t)$. For this model the generator of Markov-transition probabilities $\mathcal{G}$ is given by

$$\mathcal{G}f(x) = \left(b^P + \beta^P x\right)\frac{df(x)}{dx} + \frac{1}{2}\Sigma^2\frac{d^2 f(x)}{dx^2}. \tag{17}$$

Consider the basis $1, x, x^2, \ldots, x^p$. The linear map $\mathbf{A}$ used to derive the moments $\le p$ (under $\mathbb{P}$) is given by the $(p+1) \times (p+1)$ matrix

$$\mathbf{A} = \begin{pmatrix} 0 & \ldots & & & & & & \\ b^P & \beta^P & 0\ldots & & & & & \\ \Sigma^2 & 2b^P & 2\beta^P & 0 & \ldots & & & \\ 0 & 3\Sigma^2 & 3b^P & 3\beta^P & 0 & \ldots & & \\ & & & & \ddots & & & \\ 0 & \ldots & 0 & \frac{k(k-1)}{2}\Sigma^2 & kb^P & k\beta^P & & \\ & & & & \ddots & \ddots & \ddots & \\ 0 & \ldots & \ldots & \ldots & 0 & \frac{p(p-1)}{2}\Sigma^2 & pb^P & p\beta^P \end{pmatrix}.$$

For the Cox et al. (1985) model, where $d = 1$ and $m = 1$, $(X(t))$ follows a square-root process $dX(t) = (b^P + \beta^P X(t))dt + \Sigma\sqrt{X(t)}dW^P(t)$. The generator of Markov-transition probabilities $\mathcal{G}$ is given by

$$\mathcal{G}f(x) = (b^P + \beta^P x)\frac{df(x)}{dx} + \frac{1}{2}\Sigma^2 x\frac{d^2 f(x)}{dx^2} \,,$$

11

such that the linear map $\mathbf{A}$ is given by the $(p+1) \times (p+1)$ matrix

$$
\mathbf{A} = \begin{pmatrix}
0 & \cdots & & & & & \\
b^P & \beta^P & 0 \cdots & & & & \\
0 & 2b^P + \Sigma^2 & 2\beta^P & 0 & & \cdots & \\
0 & 0 & 3b^P + 3\Sigma^2 & 3\beta^P & 0 & & \cdots \\
& & & & \ddots & & \\
0 & \cdots & & \cdots & 0 & kb^P + \frac{k(k-1)}{2}\Sigma^2 & k\beta^P \\
& & & & & \ddots & & \ddots \\
0 & \cdots & & \cdots & \cdots & \cdots & 0 & pb^P + \frac{p(p-1)}{2}\Sigma^2 & p\beta^P
\end{pmatrix},
$$

where $1 \le k \le p$. For an $\mathbb{A}_1(3)$ model, where $d = 3$ and $m = 1$, $(\mathbf{X}(t))$ follows a stochastic process containing one square root component. Let us start with the model under $\mathbb{Q}$

$$
\begin{aligned}
d\mathbf{X}(t) = &\left(\begin{pmatrix} b_1^Q = -\beta_{11}^Q \theta_1^Q > 0 \\ b_2^Q = -\beta_{21}^Q \theta_1^Q \le 0 \\ b_3^Q = -\beta_{31}^Q \theta_1^Q \le 0 \end{pmatrix} + \begin{pmatrix} \beta_{11}^Q < 0 & 0 & 0 \\ \beta_{21}^Q \ge 0 & \beta_{22}^Q & \beta_{23}^Q \\ \beta_{31}^Q \ge 0 & \beta_{32}^Q & \beta_{33}^Q \end{pmatrix} \mathbf{X}(t) \right) dt \\
&+ \begin{pmatrix} \Sigma_1 \sqrt{X_1(t)} & & \\ & \Sigma_2 \sqrt{1 + \mathcal{B}_{12}^x X_1(t)} & \\ & & \Sigma_3 \sqrt{1 + \mathcal{B}_{13}^x X_1(t)} \end{pmatrix} d\mathbf{W}^Q(t). \quad (18)
\end{aligned}
$$

The Dai and Singleton (2000) restrictions discussed above yield: $\theta_1^Q > 0$ and $\beta_{11}^Q < 0$, $\mathcal{B}_{12}^x, \mathcal{B}_{13}^x \ge 0$, and $\Sigma_1, \Sigma_2, \Sigma_3 > 0$. Note that (18) has 13 parameters while under $\mathbb{Q}$ we can identify 14 parameters. These parameters are the thirteen parameters in (18) and γ_0 arising in (2).[8] The same structure is assumed under $\mathbb{P}$. Based on Cheridito et al. (2007) this *extended affine market price of risk* specification is mathematically well defined given that $b_I^P = b_1^P \ge 0$, $b_J^P = (b_2^P, b_3^P)' \le 0$, and eight additional parameters $\beta_{11}^P \le 0$, $\beta_{21}^P \ge 0$, $\beta_{31}^P \ge 0$, $\beta_{22}^P, \beta_{32}^P, \beta_{23}^P, \beta_{33}^P$, contained in $\boldsymbol{\beta}^P$, and $\theta_1^P \ge 0$ contained in $\boldsymbol{\theta}^P$, where $\theta_2^P = \theta_3^P = 0$. Then

---

[8]In more detail: $\boldsymbol{\beta}^Q$ (7 parameters), θ_1^Q (1 parameter; which is $\theta_1^Q \ge 0$ while $\theta_2^Q = \theta_3^Q = 0$, and thus $\mathbf{b}^Q = -\boldsymbol{\beta}^Q \boldsymbol{\theta}^Q = -[\beta_{11}^Q, \beta_{21}^Q, \beta_{31}^Q]' \theta_1^Q)$, $\boldsymbol{\Sigma}$ (3 parameters, only the elements in the main diagonal are positive, the other parameters are zero), $\mathcal{B}_{12}^x \ge 0$ and $\mathcal{B}_{13}^x \ge 0$.

$\mathbf{b}^P = -\boldsymbol{\beta}^P\boldsymbol{\theta}^P$. Since $\boldsymbol{\theta}_{2:3}^P = \boldsymbol{\theta}_{2:3}^Q = \mathbf{0}_2$ for the $\mathbb{A}_1(3)$ model considered, we write θ^Q and θ^P instead for θ_1^Q and θ_1^P in the following. By collecting these parameters (not subject to an equality restriction), we obtain the vector of model parameters $\boldsymbol{\vartheta}_{\mathbb{A}_{1(3)}} \in \mathbb{R}^{22}$.

By means of (18) and the extended affine market price of risk assumption the generator becomes

$$\mathcal{G}f(\mathbf{x}) = \sum_{i=1}^{3} \left(b_i^P + \boldsymbol{\beta}_i^P\mathbf{x}\right)\frac{\partial f(\mathbf{x})}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{3}\Sigma_i^2\left(\mathcal{B}_i^0 + \mathcal{B}_{1i}^x x_1\right)\frac{\partial^2 f(x)}{\partial x_i^2}. \tag{19}$$

The conditional expectation $\mathbb{E}(f(\mathbf{X}(t))|\mathbf{X}(s) = \mathbf{x})$ for $f(\mathbf{x}) \in \mathcal{P}_{\leq p}(\mathscr{S})$ follows from Section 3.1. In particular, the conditional moments $\mathbb{E}(\mathbf{X}(t)^k|\mathbf{X}(s) = \mathbf{x})$, $t > s$, can be derived by means of (13), where $\mathbf{A}$ is a matrix of dimension $N \times N$. We shall consider the first four moments, i.e., $p = 4$. The number of moments, N, follows from the multinomial coefficients. Regarding the basis elements $\mathfrak{e}_j$, $j = 1, \ldots, N$, of our polynomial, we choose the basis $\left(1 \,|\, x_1, x_2, x_3 \,|\, x_1^2, \ldots, x_3^2 \,|\, x_1^3, \ldots, x_3^3 \,|\, x_1^4, \ldots, x_3^4\right)$. In this expression we have separated the terms of different power by means of $|$. Matrix $\mathbf{A}$ is derived by comparing coefficients, such that $\mathcal{G}\mathfrak{e}_j = \sum_{l=1}^{N} A_{jl}\mathfrak{e}_l$, for $j = 1, \ldots, N$, where $A_{jl} = [\mathbf{A}]_{jl}$. With $(\mathbf{X}(t))$ of dimension 3, we get one term for $k = 0$, three terms for $k = 1$, six terms for $k = 2$, ten terms for $k = 3$ and fifteen terms for $k = 4$. Therefore $N = 35$. Restricting the corresponding model parameters provides us with the matrix $\mathbf{A}$ for an $\mathbb{A}_1(3)$ model.

In the remaining part of this article we stick to following assumption.

Assumption 1. *The background driving process $(\mathbf{X}(t))$ is stationary.*

Sufficient conditions for a stationary process $(\mathbf{X}(t))$ are provided in Glasserman and Kim (2010). For $\mathbb{A}_m(d)$ models, when $d \leq 3$, sufficient conditions for a stationary process are also reported in Aït-Sahalia and Kimmel (2010) and in Appendix E.

For a stationary $(\mathbf{X}(t))$, we get $\mathbb{E}(\mathbf{X}(t)) = \boldsymbol{\theta}^P$. In addition, to obtain higher order moments, we use the following abbreviations: $\tilde{\mathbf{x}} = (1, (\mathbf{x}^1)', (\mathbf{x}^2)', \ldots, (\mathbf{x}^p)')'$, which is of dimension N, while $\tilde{\mathbf{x}}_{2:N} = ((\mathbf{x}^1)', (\mathbf{x}^2)', \ldots, (\mathbf{x}^p)')'$ is a $(N-1)$−dimensional vector. $\tilde{\mathbf{X}}(t)$ and $\tilde{\mathbf{X}}(t)_{2:N}$ are defined in the same way.

Since $\mathbb{E}\left(\tilde{\mathbf{X}}(t)\right) = \mathbb{E}\left(\mathbb{E}(\tilde{\mathbf{X}}(t)|\mathbf{X}(s))\right)$, for $0 \leq s < t$, by the tower rule we obtain

$$
\begin{aligned}
\mathbb{E}\left(\tilde{\mathbf{X}}(t)\right) &= \begin{pmatrix} 1 \\ \mathbb{E}\left(\tilde{\mathbf{X}}(t)_{2:N}\right) \end{pmatrix} = \mathbb{E}\left([\exp((t-s)\mathbf{A})]\,\tilde{\mathbf{X}}(t)\right) = [\exp((t-s)\mathbf{A})]\mathbb{E}\left(\tilde{\mathbf{X}}(t)\right) \\
&= \begin{pmatrix} 1 & \mathbf{0}_{1\times N-1} \\ [\exp((t-s)\mathbf{A})]_{2:N,1} & [\exp((t-s)\mathbf{A})]_{2:N,2:N} \end{pmatrix} \begin{pmatrix} 1 \\ \mathbb{E}\left(\tilde{\mathbf{X}}(t)_{2:N}\right) \end{pmatrix},
\end{aligned}
\tag{20}
$$

where the $N \times N$ matrix $\exp((t-s)\mathbf{A})$ can be partitioned into four blocks: (i) north-western $[\exp((t-s)\mathbf{A})]_{11} = 1$, (ii) north-eastern $[\exp((t-s)\mathbf{A})]_{1,2:N} = \mathbf{0}_{1\times N-1}$, (iii) south-western $[\exp((t-s)\mathbf{A})]_{2:N,1}$, and (iv) south-eastern $[\exp((t-s)\mathbf{A})]_{2:N,2:N}$.[9] Hence, the (unconditional) moments of order 1 to p follow from

$$
\mathbb{E}\left(\tilde{\mathbf{X}}(t)_{2:N}\right) = \left(\mathbf{I}_{N-1} - [\exp((t-s)\mathbf{A})]_{2:N,2:N}\right)^{-1} [\exp((t-s)\mathbf{A})]_{2:N,1}.
\tag{21}
$$

3.3 Moments of the Observed Yields

The previous Section 3.2 provided us with the moments of the latent process $(\mathbf{X}(t))$. By means of (6) the *model yields* are

$$
y^0(t,\tau) = -\frac{1}{\tau}\left(\Phi(\tau,\mathbf{0}) + \mathbf{\Psi}(\tau,\mathbf{0})'\mathbf{X}(t)\right).
$$

Now we have to account for the fact that real world data cannot be observed on a continuous time scale, but only on a discrete grid $\Delta, 2\Delta, \ldots, \mathtt{t}\Delta, \ldots, T\Delta$, where T is the time series dimension and Δ is the step-width. We set $\Delta = 1$ and assume that $\mathbf{X}_{\mathtt{t}}$ stands for $\mathbf{X}(\mathtt{t}\Delta)$. Additionally, the maturities τ available are given by $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_M)'$, where M is the number of maturities observed. For model yields with a maturity $\tau_i \in \{\tau_1, \ldots, \tau_M\}$ observed at $t = \mathtt{t}\Delta$ we use the notation $y^0_{\mathtt{t}i}$, $i = 1, \ldots, M$. Since M yields

---

[9]Note that $\exp((t-s)\mathbf{A})$ and $\mathbf{A}$ are of the same structure. This follows from the power series representation of the matrix exponential $\exp((t-s)\mathbf{A}) = \sum_{v=0}^{\infty} \frac{1}{v}((t-s)\mathbf{A})^v$. In addition, the existence of $\left(\mathbf{I}_{N-1} - [\exp((t-s)\mathbf{A})]_{2:N,2:N}\right)^{-1}$ follows from the properties of the matrix exponential.

cannot be matched exactly by $d < M$ factors, we add the noise term $\varepsilon_{\mathsf{t}i}$ and arrive at the *yields observed*

$$y_{\mathsf{t}i} \;=\; y_{\mathsf{t}i}^0 + \varepsilon_{\mathsf{t}i} = -\frac{1}{\tau_i}\left(\Phi(\tau_i, \mathbf{0}) + \boldsymbol{\Psi}(\tau_i, \mathbf{0})'\mathbf{X}_{\mathsf{t}}\right) + \varepsilon_{\mathsf{t}i}, \quad i = 1, \ldots, M, \quad \mathsf{t} = 1, \ldots, T.$$

With M maturities $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_M)$ we define

$$\tilde{\boldsymbol{\Phi}} \;=\; \begin{pmatrix} -\Phi(\tau_1, \mathbf{0})/\tau_1 \\ \vdots \\ -\Phi(\tau_M, \mathbf{0})/\tau_M \end{pmatrix} \in \mathbb{R}^M \;, \quad \tilde{\boldsymbol{\Psi}} = \begin{pmatrix} -\boldsymbol{\Psi}(\tau_1, \mathbf{0})'/\tau_1 \\ \cdots \\ -\boldsymbol{\Psi}(\tau_M, \mathbf{0})'/\tau_M \end{pmatrix} \in \mathbb{R}^{M \times d} \;\; \text{and} \;\; \boldsymbol{\varepsilon}_{\mathsf{t}} = \begin{pmatrix} \varepsilon_{\mathsf{t}1} \\ \vdots \\ \varepsilon_{\mathsf{t}M} \end{pmatrix} \in \mathbb{R}^M \;,$$

such that the $M-$dimensional vector of yields, $\mathbf{y}_{\mathsf{t}} = (y_{\mathsf{t}1}, \ldots, y_{\mathsf{t}M})'$, is given by

$$\mathbf{y}_{\mathsf{t}} \;=\; \tilde{\boldsymbol{\Phi}} + \tilde{\boldsymbol{\Psi}}\mathbf{X}_{\mathsf{t}} + \boldsymbol{\varepsilon}_{\mathsf{t}} \in \mathbb{R}^M. \tag{22}$$

Based on (22) we observe that the moments of $y_{\mathsf{t}i}$ have to follow from the moments of $\mathbf{X}_{\mathsf{t}}$. For the noise term $\varepsilon_{\mathsf{t}i}$ we apply the following assumption.

Assumption 2. *Let $\varepsilon_{\mathsf{t}i}$, $\mathsf{t} = 1, \ldots, T$, $i = 1, \ldots, M$, be independent with zero mean, variance $0 < \sigma_i^2 < +\infty$ and $\mathbb{E}(\varepsilon_{\mathsf{t}i}^4) < +\infty$. In addition $|\mathbb{E}(\varepsilon_{\mathsf{t}i}^p)| < +\infty$ for $i = 1, \ldots, M$ and $\mathbb{E}(\varepsilon_{\mathsf{t}i}^{2\iota-1}) = 0$ for $\iota = 1, \ldots, \lfloor p/2 \rfloor$, where $\lfloor p/2 \rfloor$ is the largest integer smaller or equal to $p/2$.*

Note that by Assumption 2 all maturities are assumed to be observed with noise. In addition, $\mathbb{E}(\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j}) = 0$ for $i \neq j$, $i, j = 1, \ldots, M$ and $\mathbb{E}(\varepsilon_{\mathsf{t}i}^4) < +\infty$. By means of equation (22) and Assumption 2 we derive the moments of the empirical yields $\mathbb{E}(y_{\mathsf{t}i}^k y_{\mathsf{t}j}^l) = \mathbb{E}(([\tilde{\boldsymbol{\Phi}} + \tilde{\boldsymbol{\Psi}}\mathbf{X}_{\mathsf{t}} + \boldsymbol{\varepsilon}_{\mathsf{t}}]_i)^k ([\tilde{\boldsymbol{\Phi}} + \tilde{\boldsymbol{\Psi}}\mathbf{X}_{\mathsf{t}} + \boldsymbol{\varepsilon}_{\mathsf{t}}]_j)^l)$, where $0 \leq k + l \leq p$ and $[\cdot]_i$ extracts the i-th element of a vector. Hence, we derive the first four moments of the yields observed, i.e. $\mathbb{E}(y_{\mathsf{t}i}^k)$, $k = 1, \ldots, 4$. In addition, applications in finance often take the auto-covariance of the yields, $\mathbb{E}(y_{\mathsf{t}i}y_{\mathsf{t}-1i})$, and the auto-covariance of the squared yields, $\mathbb{E}(y_{\mathsf{t}i}^2 y_{\mathsf{t}-1i}^2)$, into consideration ("indicator for volatility clustering" - see, e.g., the discussion in Piazzesi (2010)[p. 649]). Therefore also the terms $\mathbb{E}(y_{\mathsf{t}i}y_{\mathsf{t}-1i})$ and $\mathbb{E}\left(y_{\mathsf{t}i}^2 y_{\mathsf{t}-1i}^2\right)$ are calculated. Since this part is straightforward, but tedious algebraic manipulations were necessary to obtain all these moments, we present the results in Appendix C. We put the noise parameters necessary to obtain the moments of the observed yields into the parameter vector $\boldsymbol{\vartheta}_\sigma$.

The dimension of $\boldsymbol{\vartheta}_\sigma$ depends on how σ_i^2 is specified and on the moments used in the estimation. If σ_i^2 is different for each maturity, we have M parameters for the second moments of the noise. If, in addition, the fourth moments of the yields are calculated, the fourth moments of the noise enter into the calculations as well, i.e. we get another M parameters for the moments of the noise. In this case the dimension of $\boldsymbol{\vartheta}_\sigma$ is $2M$. Since the dimension of the model parameter $\boldsymbol{\vartheta}_{\mathbb{A}_{1(3)}}$ is already over twenty, we continue with a more parsimonious specification of the noise, where $\sigma_i^2 = \sigma^2$ and $\mathbb{E}(\varepsilon_{\mathsf{t}i}^4) = \tilde{\sigma}^4$ for all $i = 1, \ldots, M$. Hence, the dimension of $\boldsymbol{\vartheta}_\sigma$ is two if fourth moments are required in the calculation of the yields observed, otherwise it is one. This results in the model parameter vector $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_{\mathbb{A}_{1(3)}}, \boldsymbol{\vartheta}'_\sigma)'$ of dimension $\mathfrak{p}$, which is contained in the parameter space $\Theta \in \mathbb{R}^{\mathfrak{p}}$, where due to the Dai and Singleton (2000) and stationarity restrictions, Θ is proper subset of $\mathbb{R}^{\mathfrak{p}}$. The components of $\boldsymbol{\vartheta}$ are introduced by the first column of Table 1.

The calculation of the moments also requires to solve the Riccati equations (29). For the Vasicek and the Cox-Ingersol-Ross model closed form solutions are available, as e.g. presented in Filipović (2009)[Chapter 10.3.2]. For $\mathbb{A}_m(d)$ models, however, Φ and $\boldsymbol{\Psi}$ have to be derived by means of numerical tools in general.[10] In this paper we follow a computationally efficient way proposed by Grasselli and Tebaldi (2008), to obtain an (almost) closed form solution for $\Phi(t, \mathbf{u})$ and $\boldsymbol{\Psi}(t, \mathbf{u})$. This methodology requires the matrix $\boldsymbol{\beta}_{II}^Q$ to be diagonal. Given Dai and Singleton (2000) setup, this implies no further restrictions for $m \leq 1$, while for $m \geq 2$ the off-diagonal parts of $\boldsymbol{\beta}_{II}^Q$ have to be set to zero. Appendix D shows how Φ and $\boldsymbol{\Psi}$ could be derived for an $\mathbb{A}_m(d)$ model with diagonal $\boldsymbol{\beta}_{II}$ in a numerically parsimonious way.

4 Parameter Estimation and Finite Sample Properties

4.1 Parameter Estimation

By observing yields for maturities τ_i, $i = 1, \ldots, M$, in periods $\mathsf{t} = 1, \ldots, T$, we obtain M−variate vectors $\mathbf{y}_{\mathsf{t}} = (y_{\mathsf{t}1}, \ldots, y_{\mathsf{t}M})'$, $\mathsf{t} = 1, \ldots, T$, the observations of M−variate time series $\mathbf{y}_{1:T} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$, as well as $\tilde{\mathfrak{q}}$−dimensional vectors $\tilde{\mathbf{m}}_{(\mathsf{t})}(\mathbf{y}_{1:T}) = \left(y_{\mathsf{t}1}, \ldots, y_{\mathsf{t}M}^p, y_{\mathsf{t}1}y_{\mathsf{t}-1,1}, \ldots, y_{\mathsf{t}M}^2 y_{\mathsf{t}-1,M}^2\right)'$ and $\tilde{\mathbf{m}}_T(\mathbf{y}_{1:T}) = \left(\frac{1}{T}\sum_{\mathsf{t}=1}^T y_{\mathsf{t}1}, \ldots, \frac{1}{T}\sum_{\mathsf{t}=1}^T y_{\mathsf{t}M}^p, \frac{1}{T-1}\sum_{\mathsf{t}=2}^T y_{\mathsf{t}1}y_{\mathsf{t}-1,1}, \ldots, \frac{1}{T-1}\sum_{\mathsf{t}=2}^T y_{\mathsf{t}M}^2 y_{\mathsf{t}-1,M}^2\right)'$.

Let $\tilde{\boldsymbol{\mu}}(\boldsymbol{\vartheta}) = \left(\mathbb{E}(y_{\mathsf{t}1}), \ldots, \mathbb{E}(y_{\mathsf{t}M}^p), \mathbb{E}(y_{\mathsf{t}1}y_{\mathsf{t}-1,1}), \ldots, \mathbb{E}(y_{\mathsf{t}M}^2 y_{\mathsf{t}-1,M}^2)\right)'$ stand for the corresponding vector

---

[10]See also Duffie and Kan (1996); Dai and Singleton (2000); Chen and Joslin (2012).

of moments as a function of the unknown parameter vector $\boldsymbol{\vartheta} \in \Theta \subset \mathbb{R}^{\mathfrak{p}}$. The components of the vector $\tilde{\boldsymbol{\mu}}(\boldsymbol{\vartheta})$ are provided in Appendix C (see equations (34), (38), (39), (40), (41), (44) and (45)).

The generalized method of moments demands for $\mathfrak{q} \geq \mathfrak{p}$ moments to be selected. By means of a $\mathfrak{q} \times \tilde{\mathfrak{q}}$ selector matrix $\mathcal{M}$, where $[\mathcal{M}]_{ij} = 1$ if the corresponding moment is used and zero otherwise, we obtain $\boldsymbol{\mu}(\boldsymbol{\vartheta}) = \mathcal{M}\, \tilde{\boldsymbol{\mu}}(\boldsymbol{\vartheta}) \in \mathbb{R}^{\mathfrak{q}}$, $\mathbf{m}_{(\mathtt{t})}(\mathbf{y}_{1:T}) = \mathcal{M}\, \tilde{\mathbf{m}}_{(\mathtt{t})}(\mathbf{y}_{1:T}) \in \mathbb{R}^{\mathfrak{q}}$ and $\mathbf{m}_T(\mathbf{y}_{1:T}) = \mathcal{M}\, \tilde{\mathbf{m}}_T(\mathbf{y}_{1:T}) \in \mathbb{R}^{\mathfrak{q}}$. Next we define $\mathbf{h}_{(\mathtt{t})}(\boldsymbol{\vartheta}; \mathbf{y}_{1:T}) = \mathbf{m}_{(\mathtt{t})}(\mathbf{y}_{1:T}) - \boldsymbol{\mu}(\boldsymbol{\vartheta})$ and $\mathbf{h}_T(\boldsymbol{\vartheta}; \mathbf{y}_{1:T}) = \mathbf{m}_T(\mathbf{y}_{1:T}) - \boldsymbol{\mu}(\boldsymbol{\vartheta})$ as well as the GMM distance function

$$Q_T(\boldsymbol{\vartheta}; \mathbf{y}_{1:T}) = \mathbf{h}_T(\boldsymbol{\vartheta}; \mathbf{y}_{1:T})'\, \mathbf{C}_T\, \mathbf{h}_T(\boldsymbol{\vartheta}; \mathbf{y}_{1:T}). \tag{23}$$

The GMM estimate $\widehat{\boldsymbol{\vartheta}}$ (of $\boldsymbol{\vartheta}$) minimizes $Q_T(\cdot)$ in (23), where $\mathbf{C}_T$ is a $\mathfrak{q} \times \mathfrak{q}$ symmetric positive semi-definite weighting matrix (see, e.g., Ruud, 2000, Chapters 21-22). In particular, the *continuous updating estimator* (CUE) is used to obtain an efficient GMM estimate. That is, we run an iterative procedure with iteration steps $\mathtt{m} = 1, \ldots, \mathtt{M}$, where we commute between (i) augmenting the parameter-estimate to $\boldsymbol{\vartheta}^{(\mathtt{m})}$ based on $Q_T(\cdot)$ given C_T and (ii) updating C_T given $\boldsymbol{\vartheta}^{(\mathtt{m}-1)}$ from the previous iteration step $\mathtt{m}-1$. The weighting matrix applied is $\mathbf{C}_T = \left(\hat{\boldsymbol{\Lambda}}_T(\boldsymbol{\vartheta}^{(\mathtt{m}-1)})\right)^{-1}$, with $\hat{\boldsymbol{\Lambda}}_T\left(\boldsymbol{\vartheta}^{(\mathtt{m}-1)}\right) = \frac{1}{T-1} \sum_{\mathtt{t}=2}^{T} \mathbf{h}_{(\mathtt{t})}(\boldsymbol{\vartheta}^{(\mathtt{m}-1)}; \mathbf{y}_{1:T})\, \mathbf{h}_{(\mathtt{t})}(\boldsymbol{\vartheta}^{(\mathtt{m}-1)}; \mathbf{y}_{1:T})'$. For regularity conditions and further issues on GMM estimation see, e.g., Hansen (1982); Altonji and Segal (1996); Pötscher and Prucha (1997); Windmeijer (2005); Guggenberger and Smith (2005); Newey and Windmeijer (2009).

To satisfy the order condition, the inequality "$\mathfrak{q} \geq \mathfrak{p}$" has to be fulfilled. For the $\mathbb{A}_1(3)$ model considered in Section 3, the dimension of the parameter vector $\boldsymbol{\vartheta}$ is 23 ($\mathfrak{p} = 23$), if moments of order smaller than four are used. Including fourth order moments of the yields results in $\mathfrak{p} = 24$. The number of maturities M available is around ten. Therefore, by using the moments $\mathbb{E}(y_{\mathtt{t}i})$, $\mathbb{E}(y_{\mathtt{t}i}^2)$ and $\mathbb{E}(y_{\mathtt{t}i}y_{\mathtt{t}-1,i})$ for $i = 1, \ldots, M$, we are already equipped with $3M$ moment conditions. Hence, for $M \geq 8$ the order condition $\mathfrak{q} \geq \mathfrak{p}$ is already met. By using the first four moments ($p = 4$) and the auto-covariances (for $M = 10$), the number of moments is much larger than the number of parameters.

To obtain parameter estimates, a high-dimensional nonlinear minimization problem has to be solved and $\mathfrak{q}$ moment conditions have to be selected from the set of moments available. Regarding the latter issue, it turned out that the instability of the parameter estimates is amplified if higher order moments

are added. Due to this instability, using the Wald and the distance difference tests to test for redundant moment conditions (testing for over-identifying restrictions; see, e.g., Ruud, 2000, Chapter 22.2) provide us with very ambiguous results. Hence, the selection of these moments was performed by means of simulation experiments. Based on the simulation results, we work with $\mathtt{q} = 27$ moment conditions, namely, $\mathbb{E}(y_{\mathtt{t}i})$, $\mathbb{E}(y_{\mathtt{t}i}y_{\mathtt{t}-1,i})$, $i = 1, \ldots, M = 10$, and $[\mathbb{E}(\mathbf{y_t}\mathbf{y'_t})]_{ij}$, for $(i, j) = (1, 1), (2, 2), (3, 2), (5, 5), (7, 7), (9, 10)$ and $(10, 10)$.

Regarding the minimization of the GMM distance function, we observe that standard minimization procedures designed to find local minima do not result in reliable parameter estimates. In more detail, to investigate the properties of our estimation routine we performed Monte Carlo experiments with simulated yields where $M = 10$, $T = 500$ and the number of simulation runs is $1,000$. The parameter vector $\boldsymbol{\vartheta}$ used to generate the yields is presented in the second column of Table 1. The initial values for the GMM estimation, $\boldsymbol{\vartheta}^{(\mathtt{m_0})}$, are generated as follows: $[\boldsymbol{\vartheta}^{(\mathtt{m_0})}]_j = [\boldsymbol{\vartheta}]_j + c_{\vartheta}[|\boldsymbol{\vartheta}|]_j\zeta_j$ for coordinate j, when the support is the real axis, while $[\boldsymbol{\vartheta}^{(\mathtt{m_0})}]_j = \exp\left(\log[|\boldsymbol{\vartheta}|]_j + c_{\vartheta}\zeta_j\right)\mathrm{sgn}\left([\boldsymbol{\vartheta}]_j\right)$ is used for the elements j living on the non-positive or non-negative part of the real axis. ζ_j is iid standard normal and distortion parameter c_{ϑ} is set to 0, 0.1, 0.25, 0.5 and 1. Then, parameter estimates are obtained by means of the MATLAB minimization routine `fminsearch` based on the Nelder-Mead algorithm.[11] With this algorithm an estimate $\widehat{\boldsymbol{\vartheta}}$ is provided by $\boldsymbol{\vartheta}^{(\mathtt{M})}$, where – in this case — $\mathtt{M}$ is the last iteration step. We observe that the parameters can be estimated easily by means of this standard minimization tool when $c_{\vartheta} \leq 0.25$; i.e., when the optimization is started sufficiently close to the true parameter $\boldsymbol{\vartheta}$. However, the parameter estimation with $c_{\vartheta} = 0.5$ or $c_{\vartheta} = 1$ becomes a difficult problem.[12]

To cope with this problem, we combine multistart random search methods with Quasi-Bayesian methods (see, e.g., Törn and Zilinskas, 1989; Chernozhukov and Hong, 2003). For each Monte Carlo run ℓ, where $\ell = 1, \ldots, \mathtt{L} = 200$, we proceed as follows: First, parameter estimation is started with the random draws $\boldsymbol{\vartheta}^{(\mathtt{n})}$, where $\mathtt{n} = 1, \ldots, \mathtt{N} = 2,000$. The samples $\boldsymbol{\vartheta}^{(\mathtt{n})}$ are generated in the same way as $\boldsymbol{\vartheta}^{(\mathtt{m_0})}$ in the above paragraph with distortion parameter $c_{\vartheta} = 1$. Then $\boldsymbol{\vartheta}^{(\mathtt{n})}$ with the smallest GMM distance function

---

[11]See `http : //www.mathworks.de/de/help/matlab/ref/fminsearch.html`

[12]By combining multistart random search methods (see, e.g., Törn and Zilinskas, 1989) with the Nelder-Mead algorithm, we observe that the parameter estimates improve. However, performing inference still remains a difficult problem. For more details see Appendix F.

is used as the starting value of the Quasi-Bayesian sampler. Appendix F describes how the draws, $\boldsymbol{\vartheta}^{(\mathtt{m})}$, from an ergodic Markov Chain are obtained.[13] Finally, parameter estimates $\widehat{\boldsymbol{\vartheta}}_\ell$ as well as the estimates of the variance $\hat{\mathbb{V}}_{BM}\left(\left[\widehat{\boldsymbol{\vartheta}}_\ell\right]_\iota\right)$, $\iota = 1, \ldots, \mathfrak{p}$, are derived from these draws, where the latter are obtained by applying a batch mean estimator (see Flegal and Jones, 2010, in particular, Equation (6)).

Tables 1 and 2 present results from our Monte Carlo experiments. In both tables the true parameter vector $\boldsymbol{\vartheta}$ is provided in the second column. In Table 1 the data are generated such that $\theta^P = 1.5 \neq 10 = \theta^Q$, while $\theta^P = \theta^Q = 1.5$ in Table 2. In all Monte Carlo experiments an unrestricted model is estimated. That is, we obtain separate estimates for θ^P and θ^Q, respectively. We force our multistart random search routine to generate samples such that $\left(\theta^P\right)^{(\mathtt{n})} = \left(\theta^Q\right)^{(\mathtt{n})}$ as well as $\left(\theta^P\right)^{(\mathtt{n})} \neq \left(\theta^Q\right)^{(\mathtt{n})}$ (for both experiments presented in Tables 1 and 2, respectively). In addition, a reversible jump move, based on Green (1995) and Richardson and Green (1997) is included in the Bayesian sampler. The reversible jump move turned out to be useful in the case when $\theta^P = \theta^Q$ (see Appendix F for more details).

From estimates $\widehat{\boldsymbol{\vartheta}}_\ell$, $\ell = 1, \ldots, \mathtt{L} = 200$, we obtain the sample *mean*, *median*, minimum (*min*), maximum (*max*), standard deviation (*std*), skewness (*skew*) and kurtosis (*kurt*). These descriptive statistics are reported in columns three to nine of Tables 1 and 2. The last column presents the absolute difference between the sample mean of the estimates and the true parameter value.

Comparing results based on Quasi-Bayesian methods (see Table 1) for the case when $\theta^P \neq \theta^Q$ to results based on a standard minimization procedure (see Table 5 in Supplementary Material F), we see that the Quasi-Bayesian approach reduces the standard deviations of the point estimates for most parameters. For example, the standard deviation of the point estimate of θ^Q is reduced from 6.05 (see Table 5) to approximately 3.05 (see Table 1). Similar effects are observed for the estimates of the terms driving volatility, i.e., Σ_1, Σ_2, Σ_3 and σ_ε^2, which are difficult to estimate. By considering the smallest and the largest point estimates (*min* and *max* in the corresponding tables), we observe a substantially smaller dispersion in the point estimates of $\boldsymbol{\vartheta}$ for the Quasi-Bayesian approach. Note that an estimate of θ^P is an estimate of the expected value of the first component of the process $(\mathbf{X}(t))_{t \geq 0}$. Since the serial correlation of $(\mathbf{X}_\mathtt{t})_{\mathtt{t} \in \mathbb{N}_0}$ is quite high, we know from estimating means of an autoregressive process, that the standard error of the estimator of the mean becomes large (e.g., when the Fisher-information matrix of an $AR(1)$

---

[13]In our analysis $\mathtt{m} = 1, 2, \ldots, \mathtt{M} = 20,000$.

process is calculated). Similar results are presented in Table 2 for the $\theta^P = \theta^Q$ case.

$\boldsymbol{\vartheta}$		$mean$ $\widehat{\boldsymbol{\vartheta}}$	$median$	min	max	std	$skew$	$kurt$	$\|\boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}\|$
θ^Q	10	8.8660	8.5486	0.1534	19.2247	3.5008	0.6071	4.2157	1.1340
θ^P	1.5	1.6610	1.4883	0.0042	2.5920	1.0643	0.4911	-0.3916	0.1610
β_{11}^Q	-1	-1.6418	-1.2797	-9.2212	-0.4173	1.5798	-3.2923	11.8217	0.6418
β_{21}^Q	0.2	0.1817	0.1524	0.0025	0.3591	0.1299	1.8759	6.5029	0.0183
β_{31}^Q	0.02	0.0350	0.0214	1.86E-5	0.3473	0.0457	3.1329	14.2229	0.0150
β_{22}^Q	-1	-1.4731	-1.0671	-8.1154	-0.4823	1.1478	-2.7690	10.1519	0.4731
β_{32}^Q	0.04	0.0373	0.0219	-0.0662	0.2711	0.0606	2.3781	10.2813	0.0027
β_{23}^Q	0	0.0006	-0.0003	-0.0840	0.0266	0.0176	1.5436	16.5725	0.0006
β_{33}^Q	-0.8	-1.5327	-1.2070	-7.8466	-0.6308	1.2389	-2.5704	8.1375	0.7327
β_{11}^P	-1	-1.5069	-0.9650	-7.0168	-0.1670	1.4929	-1.5812	1.8702	0.5069
β_{21}^P	0.02	0.0288	0.0037	3.67E-6	0.0170	0.0778	5.4759	35.4115	0.0088
β_{31}^P	0.01	0.0099	0.0032	4.44E-7	0.0006	0.0206	5.0431	32.4652	0.0001
β_{22}^P	-0.7	-1.1194	-0.6085	-7.5792	-0.1400	1.2938	-2.2389	6.0933	0.4194
β_{32}^P	0.01	-1.1194	-0.6085	-7.5792	-0.1400	1.2938	-2.2389	6.0933	1.1294
β_{23}^P	0	-0.0015	0.0000	-0.0551	0.0017	0.0104	-1.1382	8.4369	0.0015
β_{33}^P	-0.7	-0.9059	-0.4692	-6.5051	-0.1844	1.1881	-2.7918	7.8669	0.2059
$\mathcal{B}_{12}^x$	0.1	0.0623	0.0123	2.43E-6	0.0493	0.1652	5.6178	35.3695	0.0377
$\mathcal{B}_{13}^x$	0.01	0.1045	0.0352	8.08E-7	0.8676	0.1856	3.3315	13.2906	0.0945
γ_0	2	1.7855	1.8939	-0.0070	3.2115	0.8411	-0.0384	-0.0520	0.2145
Σ_1	0.7	0.5921	0.5238	0.2002	1.3639	0.3121	0.9334	0.4182	0.1079
Σ_2	1	0.4704	0.3714	0.1060	0.9983	0.3336	1.3636	1.4000	0.5296
Σ_3	0.8	0.4563	0.3447	0.1071	1.0514	0.3451	1.4573	1.8141	0.3437
σ_ε^2	0.0067	0.0113	0.0096	0.0053	0.0176	0.0047	0.7672	-0.5302	0.0046

Table 1: Parameter estimates for the $\mathbb{A}_1(3)$ based on Quasi-Bayesian methods. Data simulated with $M = 10$, $T = 500$ and $\theta^Q \neq \theta^P$. $c_\vartheta = 1$ is controlling for the noise in the generation of the starting value of the optimization routine. Statistics are obtained from $\mathtt{L} = 200$ simulation runs. $mean$, $median$, min, max, std, $skew$ and $kurt$ stand for the sample mean, median, minimum, maximum, standard deviation, skewness and kurtosis of the point estimates $\widehat{\boldsymbol{\vartheta}}_\ell$, $\ell = 1, \ldots, \mathtt{L}$. $\|\boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}}\|$ stands for absolute value of the mean deviation from the true parameter. The true parameter values $\boldsymbol{\vartheta}$ are reported in the second column.

4.2 Inference

The asymptotic distribution of $\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right)$ is a normal distribution with mean vector $\mathbf{0}_{\mathfrak{p}}$ and the asymptotic covariance matrix $\mathbf{V}$ (for more details and regularity conditions see, e.g., Hansen, 1982; Pötscher and Prucha, 1997; Newey and McFadden, 1994; Ruud, 2000). As our test statistics rely on asymptotic

results, we have to investigate the finite sample properties of our tests. Since a lot of parameters are considered and various restrictions can be constructed, we focus now on the restriction $\theta^P = \theta^Q$, which is often discussed in finance literature.

To test for parameter restrictions, we assume that the null hypothesis consists of $\mathfrak{r}_{\mathfrak{p}}$ restrictions. Suppose that these restrictions are described by a twice continuously differential function $r(\boldsymbol{\vartheta}) : \mathbb{R}^{\mathfrak{p}} \to \mathbb{R}^{\mathfrak{r}_{\mathfrak{p}}}$ and the $\mathfrak{r}_{\mathfrak{p}} \times \mathfrak{p}$ matrix of partial derivatives

$$\mathbf{R} = \mathbf{D}_{\vartheta} \mathbf{r}(\widehat{\boldsymbol{\vartheta}}) = \begin{pmatrix} \frac{\partial r_1(\widehat{\boldsymbol{\vartheta}})}{\partial \vartheta_1} & \cdots & \frac{\partial r_1(\widehat{\boldsymbol{\vartheta}})}{\partial \vartheta_{\mathfrak{p}}} \\ \cdots & \cdots & \cdots \\ \frac{\partial r_{\mathfrak{r}_{\mathfrak{p}}}(\widehat{\boldsymbol{\vartheta}})}{\partial \vartheta_1} & \cdots & \frac{\partial r_{\mathfrak{r}_{\mathfrak{p}}}(\widehat{\boldsymbol{\vartheta}})}{\partial \vartheta_{\mathfrak{p}}} \end{pmatrix}, \tag{24}$$

which has rank $\mathfrak{r}_{\mathfrak{p}}$. Under the null hypothesis we have $\mathbf{r}(\boldsymbol{\vartheta}) = \mathbf{0}_{\mathfrak{r}_{\mathfrak{p}}}$ and thus the Wald-statistic becomes

$$\mathscr{W} \;\; = \;\; T \mathbf{r}(\widehat{\boldsymbol{\vartheta}})' \left(\mathbf{R} \hat{\mathbf{V}}_T \mathbf{R}' \right)^{-1} \mathbf{r}(\widehat{\boldsymbol{\vartheta}}) , \tag{25}$$

where $\hat{\mathbf{V}}_T$ is an estimate of the asymptotic covariance matrix of $\sqrt{T}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta})$. Under the null hypothesis the Wald-statistic $\mathscr{W}$ follows a χ^2-distribution with $\mathfrak{r}_{\mathfrak{p}}$ degrees of freedom. The null hypothesis is rejected if $\mathscr{W} > \chi^2_{\mathfrak{r}_{\mathfrak{p}}, 1-\alpha_S}$, where α_S is the significance level and $\chi^2_{\mathfrak{r}_{\mathfrak{p}}, 1-\alpha_S}$ is the $1-\alpha_S$ percentile of a χ^2-distribution with $\mathfrak{r}_{\mathfrak{p}}$ degrees of freedom. In particular, if the goal is to test the null hypothesis $\theta^P = \theta^Q$ against the alternative $\theta^P \neq \theta^Q$, then $\mathfrak{r}_{\mathfrak{p}} = 1$, $r(\boldsymbol{\vartheta}) = (1, -1, 0, \ldots, 0) \boldsymbol{\vartheta} = \theta^Q - \theta^P$ and $\mathbf{R} = (1, -1, 0, \ldots, 0).$[14]

Appendix F demonstrates that the performance of the Wald test implemented in a standard way (as well as the distance difference test) is poor. In particular with $c_{\vartheta} = 1$, substantial undersizing is observed for the Wald test while the power is very low. With the distance difference test we observe only minor oversizing, and even if it's power is already better than the power of the Wald test, is is still low (approximately 55% rejection rate on a 5% significance level).[15] To implement a "standard" Wald or distance difference test, the $\mathfrak{p} \times \mathfrak{p}$ covariance matrix $\mathbf{V}$ is estimated by means of the "standard GMM covariance matrix estimate" (see, e.g., Ruud, 2000, Chapters 21 and 22, for a "standard" implementation

---

[14]The components of the parameter vector $\boldsymbol{\vartheta}$ are presented in the first column of Table 1.

[15]We used here the same simulation designs as in Tables 1 and 2.

of the Wald and the distance difference test). That is, when the following estimate is applied

$$
\hat{\mathbf{V}}_T = \left(\hat{\mathbf{H}}_T' \hat{\mathbf{\Lambda}}_T^{-1} \hat{\mathbf{H}}_T \right)^{-1}, \text{ where}
$$

$$
\hat{\mathbf{H}}_T = \frac{1}{T-1} \sum_{\mathtt{t}=2}^{T} \mathbf{D}_{\boldsymbol{\vartheta}} \mathbf{h}_{(\mathtt{t})} \left(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}_{1:T} \right) \in \mathbb{R}^{\mathtt{q} \times \mathtt{p}} \text{ and}
$$

$$
\hat{\mathbf{\Lambda}}_T = \frac{1}{T-1} \sum_{\mathtt{t}=2}^{T} \mathbf{h}_{(\mathtt{t})} \left(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}_{1:T} \right) \mathbf{h}_{(\mathtt{t})} \left(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}_{1:T} \right)' \in \mathbb{R}^{\mathtt{q} \times \mathtt{q}}. \tag{26}
$$

Note that in (26) matrices of dimension $\mathtt{p} \times \mathtt{p}$ (with $\mathtt{p} \geq 23$) have to be inverted and partial derivatives in matrix $\mathbf{D}_{\boldsymbol{\vartheta}} \mathbf{h}_{(\mathtt{t})} \left(\widehat{\boldsymbol{\vartheta}}; \mathbf{y}_{1:T} \right)$ have to be derived numerically. Hence, estimating the covariance matrix $\mathbf{V}$ by means of (26) is numerically demanding. Additionally, $\hat{\mathbf{H}}_T$ as well as $\hat{\mathbf{\Lambda}}_T$ also depend on $\mathbf{y}_{1:T}$, and therefore are subject to the variation of the finite samples.

To cope with this problem, we use the output of the Bayesian sampler to perform inference. Based on Chernozhukov and Hong (2003), asymptotic normality still holds and the draws from an ergodic Markov Chain, $\boldsymbol{\vartheta}^{(\mathtt{m})}$, can be used to estimate the covariance matrix $\mathbf{V}$. In particular, to estimate the asymptotic variance of $\hat{\theta}^P - \hat{\theta}^Q = (1, -1, 0, \ldots, 0) \, \widehat{\boldsymbol{\vartheta}}$, we use Markov-Chain Monte Carlo output and the batch mean estimator (see Flegal and Jones, 2010, Equation (6)). For the Wald test, rejection rates of the true and the false null-hypothesis are provided in Table 3. We observe that the rejection rates of the true null-hypothesis $\theta^Q = \theta^P$ are quite close to their theoretical values α_S.

5 Parameter Estimation in Empirical Data

This section applies the estimator developed in the previous sections to empirical data. We downloaded H-15 interest rate data from the Federal Reserve.[16] In particular, we used weekly data (measured every Friday) of "Treasury constant maturity" yields. The time period considered is August 3, 2001 to August 30, 2013. An almost full panel of maturities from one month to thirty years is available for these periods. Since the thirty year maturity time series exhibits a lot of missing values this maturity has been excluded. Thus, we have $M = 10$ maturities such that $\boldsymbol{\tau} = \{1/12, 1/4, 1/2, 1, 2, 3, 5, 7, 10, 20\}$ and $T = 631$ observations

---

[16] http://federalreserve.gov/releases/h15/data.htm

per yield. Although the H-15 data set can only be seen as a proxy for the risk-free term structure, we follow the related literature (see, e.g., Chib and Ergashev, 2009) and work with this dataset.

In contrast to the analysis in Section 4, where L draws from the data generating process were considered, this section investigates one panel of interest rate data. The purpose of running the GMM estimation procedure $\mathtt{L}$−times with the same data, is to check for the stability of our estimation routine in the empirical data.[17] By doing this, we observe that in all simulation runs, $\ell = 1, \ldots, \mathtt{L} = 5$, the intervals $\left[\widehat{\boldsymbol{\vartheta}}_\ell\right]_\iota \pm \hat{\mathbb{V}}_{BM}\left(\left[\widehat{\boldsymbol{\vartheta}}_\ell\right]_{\iota\iota}\right)^{0.5}$, $\iota = 1, \ldots, \mathfrak{p}$, overlap. Without the Quasi-Bayesian algorithm, this stability result would not have been attained. In addition, in all simulation runs the p-values for the test $\theta^Q = \theta^P$ against the two-sided alternative $\theta^Q \neq \theta^P$ are smaller than 0.05. Hence, we reject the null hypothesis $\theta^Q = \theta^P$ at the significance level $\alpha_S = 0.05$.

To obtain parameter estimates, the draws of the Bayesian sampler $\boldsymbol{\vartheta}^{(\mathtt{m})}$, $\mathtt{m} = 5,001, \ldots, 20,000$, are used from which we obtain the sample mean $\widehat{\boldsymbol{\vartheta}}$ and the vector of sample standard deviations $\left(\left[\hat{\mathbb{V}}_{BM}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]_{11}^{0.5}, \ldots, \left[\hat{\mathbb{V}}_{BM}\left(\widehat{\boldsymbol{\vartheta}}\right)\right]_{\mathfrak{p}\mathfrak{p}}^{0.5}\right)'$, where again the batch mean estimator Flegal and Jones (2010)[Equation 6] is applied. In contrast to Tables 1 and 2, where the descriptive statistics based on the various point estimates $\widehat{\boldsymbol{\vartheta}}_\ell$ are presented, we now obtain the $\widehat{median}_{\boldsymbol{\vartheta}}$, sample minimum, $\widehat{min}_{\boldsymbol{\vartheta}}$, sample maximum, $\widehat{max}_{\boldsymbol{\vartheta}}$, sample standard deviation, $\widehat{std}_{\boldsymbol{\vartheta}}$, sample skewness, $\widehat{skew}_{\boldsymbol{\vartheta}}$ and sample kurtosis, $\widehat{kurt}_{\boldsymbol{\vartheta}}$ from the draws of one particular chain $\boldsymbol{\vartheta}^{(\mathtt{m})}$, $\mathtt{m} = 5,001, \ldots, 20,000$. These descriptive statistics are presented in Table 4.

Following mathematical finance literature (see, e.g., Cheridito et al., 2007; Cochrane, 2005), a usual way to investigate how the market demands for a compensation (*risk premium*) for the risk generated by $\mathbf{W}^P(t)$, is to consider the market price of risk process $(\boldsymbol{\phi}(\mathbf{X}(t)))_{t \geq 0}$ described in (4). This process depends on the model parameters $\boldsymbol{\vartheta}$. If $\mathbf{b}^P = \mathbf{b}^Q$ and $\boldsymbol{\beta}^P = \boldsymbol{\beta}^Q$, then $\boldsymbol{\phi}(\mathbf{X}(t)) = \mathbf{0}_d$. In terms of the parametrization used in this article, $\boldsymbol{\phi}(\mathbf{X}(t)) = \mathbf{0}_d$ if $\theta^P = \theta^Q$ and $\boldsymbol{\beta}^P = \boldsymbol{\beta}^Q$, while if $\theta^P \neq \theta^Q$ or $\boldsymbol{\beta}^P \neq \boldsymbol{\beta}^Q$, then $\boldsymbol{\phi}(\mathbf{X}(t)) \neq \mathbf{0}_d$ (almost surely). In the following we test whether this is the case.

By considering the estimates $\hat{\theta}^Q = 12.0667$ and $\hat{\theta}^P = 0.0682$ and their estimated standard deviations $\hat{\mathbb{V}}_{BM}\left(\hat{\theta}^Q\right)^{0.5} = 2.0573$ and $\hat{\mathbb{V}}_{BM}\left(\hat{\theta}^P\right)^{0.5} = 0.0728$, respectively, we observe that the difference in the parameter estimates is relatively large, compared to their estimated standard deviations. We obtained

---

[17]For the mulitstart random search, the vector of parameters presented in the second row of Table 1 is used.

the Wald statistic $\mathscr{W} = 4.32546$ with p-value being 0.03056. Based on this, the null hypothesis $\theta^Q = \theta^P$ is rejected at the significance level $\alpha_S = 0.05$ for this empirical data set.

Next, we perform the test $\boldsymbol{\beta}^P = \boldsymbol{\beta}^Q$ against the alternative $\boldsymbol{\beta}^P \neq \boldsymbol{\beta}^Q$, where $\boldsymbol{\beta}^{\cdot}$ contains seven parameters. In more details, we test the null hypothesis $\left(\beta_{11}^Q, \beta_{21}^Q, \beta_{31}^Q, \beta_{22}^Q, \beta_{32}^Q, \beta_{23}^Q, \beta_{33}^Q\right)$ $= \left(\beta_{11}^P, \beta_{21}^P, \beta_{31}^P, \beta_{22}^P, \beta_{32}^P, \beta_{23}^P, \beta_{33}^P\right)$ against the two sided alternative $\left(\beta_{11}^Q, \beta_{21}^Q, \beta_{31}^Q, \beta_{22}^Q, \beta_{32}^Q, \beta_{23}^Q, \beta_{33}^Q\right)$ $\neq \left(\beta_{11}^P, \beta_{21}^P, \beta_{31}^P, \beta_{22}^P, \beta_{32}^P, \beta_{23}^P, \beta_{33}^P\right)$. By estimating $\left(\beta_{11}^Q, \beta_{21}^Q, \beta_{31}^Q, \beta_{22}^Q, \beta_{32}^Q, \beta_{23}^Q, \beta_{33}^Q\right)'$ $-$ $\left(\beta_{11}^P, \beta_{21}^P, \beta_{31}^P, \beta_{22}^P, \beta_{32}^P, \beta_{23}^P, \beta_{33}^P\right)'$ and its covariance matrix from Monte Carlo output, we obtain the Wald statistic $\mathscr{W} = 38.7047$ with a corresponding p-value of 2.223 E-6. That is, also the null hypothesis $\left(\beta_{11}^Q, \beta_{21}^Q, \beta_{31}^Q, \beta_{22}^Q, \beta_{32}^Q, \beta_{23}^Q, \beta_{33}^Q\right) = \left(\beta_{11}^P, \beta_{21}^P, \beta_{31}^P, \beta_{22}^P, \beta_{32}^P, \beta_{23}^P, \beta_{33}^P\right)$ is rejected on significance levels $\alpha_S \geq 0.01$. Summing up, since the null hypothesis $\theta^P = \theta^Q$ and $\boldsymbol{\beta}^P = \boldsymbol{\beta}^Q$ are rejected, the market price of risk process is significantly different from zero.

6 Conclusions

In this article we developed a new method allowing for parameter estimation based on the exact moments of the yields for affine term structure models. By applying the results of Cuchiero et al. (2012) on $p-$polynomial processes the conditional moments are derived. By assuming a stationary process, we obtain the exact moments of the yields as well as the first order auto-covariance of the yields and the squared yields. By means of these moments, the model parameters can be estimated by the generalized method of moments.

Since the number of parameters is relatively large and the moments are non-linear in the model parameters, the implementation of the generalized method of moments becomes a non-trivial problem. We observe that standard minimization routines perform poorly. To cope with this problem, we use random search methods combined with Quasi-Bayesian methods to minimize the GMM distance function as proposed in Chernozhukov and Hong (2003). By these techniques parameter estimation becomes more stable. The standard deviations as well as the dispersions of the point estimates decrease for most parameters, compared to parameter estimation based on a standard minimization of the GMM distance function. For some parameters this decline is substantial.

Another main contribution of this article is a rigorous investigation of the testing problem, whether parameters controlling for the mean of the latent affine process in the empirical and in the equivalent martingale measure are different. We observe substantial undersizing, when implementing a Wald test based on standard estimates of the covariance matrix of the unknown parameter. By applying methods developed by Chernozhukov and Hong (2003), the standard errors of the corresponding components of the parameter vector can be obtained from the draws provided by a Bayesian sampler. We observe that in this case the rejection rates of the true null hypothesis are close to theoretically correct levels.

In a final step, our estimation methodology is applied to empirical term structure data. By applying the testing procedure developed in this article, the null hypothesis of equal parameters controlling for the mean of the latent affine process, in the empirical as well as in the equivalent martingale measure, is rejected. Our estimates support a significant market price of risk.

ϑ		$mean$ $\widehat{\vartheta}$	$median$	min	max	std	$skew$	$kurt$	$\lvert\vartheta - \widehat{\vartheta}\rvert$
θ^Q	1.5	1.7127	1.2500	0.0148	5.4034	1.5225	2.4322	6.6231	0.2127
θ^P	1.5	1.4298	1.4745	0.0218	2.1810	0.5370	-0.2753	0.6087	0.0702
β_{11}^Q	-1	-0.9482	-0.7216	-9.3936	-0.2657	1.1017	-5.9892	42.4434	0.0518
β_{21}^Q	0.2	0.2760	0.1745	0.0082	0.5801	0.3184	2.7465	8.6887	0.0760
β_{31}^Q	0.02	0.0365	0.0188	0.0001	0.0271	0.0501	3.5544	16.0667	0.0165
β_{22}^Q	-1	-1.4434	-1.1180	-8.5167	-0.6810	1.1585	-2.6154	10.2604	0.4434
β_{32}^Q	0.04	0.0391	0.0280	-0.0514	0.0828	0.0483	1.8007	6.8699	0.0009
β_{23}^Q	0	-0.0013	-0.0001	-0.0562	0.0295	0.0108	-0.9656	8.6647	0.0013
β_{33}^Q	-0.8	-1.3134	-1.0069	-6.6218	-0.5230	1.0095	-2.1866	6.7837	0.5134
β_{11}^P	-1	-1.8616	-1.4688	-6.8225	-0.7239	1.5857	-0.8374	0.4486	0.8616
β_{21}^P	0.02	0.2610	0.1233	0.0017	0.8445	0.4000	4.3370	26.3265	0.2410
β_{31}^P	0.01	0.0314	0.0127	0.0001	0.0602	0.0489	3.4009	15.1149	0.0214
β_{22}^P	-0.7	-1.1592	-0.8226	-6.6295	-0.1769	1.0613	-1.5312	4.0173	0.4592
β_{32}^P	0.01	0.0383	0.0207	-0.1791	0.0872	0.0606	1.8303	6.8339	0.0283
β_{23}^P	0	-0.0010	0.0002	-0.2496	0.0425	0.0231	-4.7594	70.9923	0.0010
β_{33}^P	-0.7	-1.3493	-1.0871	-6.3858	-0.1818	1.2570	-1.5028	3.0873	0.6493
$\mathcal{B}_{12}^x$	0.1	0.0769	0.0326	0.0007	0.0456	0.1328	3.8884	18.7903	0.0231
$\mathcal{B}_{13}^x$	0.01	0.1262	0.0677	0.0019	0.2241	0.1816	4.0900	22.9321	0.1162
γ_0	2	1.9495	1.9564	0.0111	2.1332	0.5018	-2.0313	7.2117	0.0505
Σ_1	0.7	0.8427	0.9478	0.0186	1.1315	0.3418	-0.7662	-0.4319	0.1427
Σ_2	1	0.6263	0.5797	0.0225	1.0547	0.3573	0.4775	-0.4588	0.3737
Σ_3	0.8	0.5591	0.4891	0.0182	1.2413	0.3631	0.6926	-0.4573	0.2409
σ_ε^2	0.0067	0.0106	0.0093	0.0009	0.0215	0.0049	0.6131	-0.2135	0.0039

Table 2: Parameter estimates for the $\mathbb{A}_1(3)$ based on Quasi-Bayesian methods. Data simulated with $M = 10$, $T = 500$ and $\theta^Q = \theta^P$. $c_\vartheta = 1$ is controlling for the noise in the generation of the starting value of the optimization routine. Statistics are obtained from $\mathtt{L} = 200$ simulation runs. *mean*, *median*, *min*, *max*, *std*, *skew* and *kurt* stand for the sample mean, median, minimum, maximum, standard deviation, skewness and kurtosis of the point estimates $\widehat{\vartheta}_\ell$, $\ell = 1, \ldots, \mathtt{L}$. $\lvert\vartheta - \widehat{\vartheta}\rvert$ stands for absolute value of the mean deviation from the true parameter. The true parameter values ϑ are reported in the second column.

α_S	$\theta^Q = 10 \neq 1.5 = \theta^P$	$\theta^Q = \theta^P = 1.5$
0.01	1.0000	0.0286
0.05	1.0000	0.0476
0.10	1.0000	0.0857

Table 3: Parameter tests based on the Wald test (25): Data simulated with $M = 10$ and $T = 500$; α_S stands for the significance level; $c_\theta = 1$ controls for the noise in the generation of the starting value of the optimization routine. The null hypothesis is $\theta^Q = \theta^P$, which is tested against the two sided alternative $\theta^Q \neq \theta^P$. The draws of the Quasi-Bayesian sampler are used to estimate θ^Q, θ^P as well as the asymptotic variance of $\hat{\theta}^Q - \hat{\theta}^P$. The quantities presented are rejection rates of the null hypothesis given the significance level α_S. Statistics are obtained from $\mathtt{L} = 200$ simulation runs.

$\boldsymbol{\vartheta}$	$\widehat{\boldsymbol{\vartheta}}$	$\widehat{median}_{\boldsymbol{\vartheta}}$	$\widehat{min}_{\boldsymbol{\vartheta}}$	$\widehat{max}_{\boldsymbol{\vartheta}}$	$\widehat{std}_{\boldsymbol{\vartheta}}$	$\widehat{skew}_{\boldsymbol{\vartheta}}$	$\widehat{kurt}_{\boldsymbol{\vartheta}}$
θ^Q	12.0667	11.6886	6.1634	15.2185	2.0573	-0.5395	3.0115
θ^P	0.0682	0.0460	0.0174	0.3855	0.0728	2.7690	10.3944
β_{11}^Q	-0.1036	-0.1005	-0.1754	-0.1000	0.0108	-3.8346	17.1485
β_{21}^Q	0.0331	0.0245	0.0128	0.1169	0.0245	1.8071	5.5307
β_{31}^Q	0.0108	0.0088	0.0060	0.0283	0.0052	1.6536	4.6572
β_{22}^Q	-1.4100	-1.3316	-2.3193	-0.8051	0.5064	-0.3689	1.6129
β_{23}^Q	0.0925	0.0915	0.0782	0.1203	0.0073	0.5622	3.3120
β_{32}^Q	-0.0096	-0.0092	-0.0143	-0.0060	0.0020	-0.4681	2.4442
β_{33}^Q	-0.8124	-0.7957	-1.1401	-0.7108	0.0720	-2.4422	9.4258
β_{11}^P	-0.7390	-0.5119	-2.1933	-0.1430	0.5474	-0.8072	2.3157
β_{21}^P	0.0542	0.0475	0.0191	0.1164	0.0231	0.6122	2.4151
β_{31}^P	0.0196	0.0196	0.0083	0.0379	0.0053	0.2360	2.9440
β_{22}^P	-2.9191	-2.9900	-5.5775	-1.1761	1.0701	-0.2561	2.1254
β_{23}^P	0.0047	0.0049	0.0017	0.0088	0.0019	0.0118	1.4428
β_{32}^P	-0.0019	-0.0020	-0.0030	-0.0010	0.0005	0.2345	2.0680
β_{33}^P	-0.4352	-0.4247	-0.8304	-0.3137	0.0704	-2.0444	9.6661
$\mathcal{B}_{12}^x$	0.0324	0.0295	0.0155	0.0570	0.0101	0.4526	2.0998
$\mathcal{B}_{13}^x$	0.0871	0.0842	0.0508	0.1412	0.0215	0.4104	2.1585
γ_0	1.8155	1.8188	1.5172	1.9797	0.0662	-0.7029	4.5135
Σ_1	0.2004	0.2002	0.2000	0.2024	0.0004	2.2614	8.7845
Σ_2	1.1715	1.1476	0.7876	1.4995	0.2554	0.0554	1.4457
Σ_3	1.4936	1.4957	1.4258	1.4999	0.0074	-3.3418	21.3889
σ_ε^2	0.0119	0.0119	0.0106	0.0137	0.0006	-0.0545	3.4937

Table 4: Parameter estimates for empirical H-15 interest rate data for the $\mathbb{A}_1(3)$ model. Statistics are obtained from $\texttt{M} = 20,000$ draws with $\texttt{M}_b = 5,000$ burn-in steps. $\widehat{\boldsymbol{\vartheta}}$ stands for sample mean, $\widehat{median}_{\boldsymbol{\vartheta}}$ for sample median, $\widehat{min}_{\boldsymbol{\vartheta}}$ for sample minimum, $\widehat{max}_{\boldsymbol{\vartheta}}$ for sample maximum, $\widehat{std}_{\boldsymbol{\vartheta}}$ for sample standard deviation, $\widehat{skew}_{\boldsymbol{\vartheta}}$ for sample skewness and $\widehat{kurt}_{\boldsymbol{\vartheta}}$ for sample kurtosis obtained from the draws of the chain $\left(\boldsymbol{\vartheta}^{(\texttt{m})} : \texttt{m} = \texttt{M}_b + 1, \ldots, \texttt{M} \right)$.

A Affine Models

The following paragraphs - based on Filipović (2009) - describe *affine processes*. Let us assume the following: The state space is given by $\mathscr{S} \subset \mathbb{R}^d$, $\mathbf{W}(t)$ stands for $d-$dimensional standard Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ and for any initial value $\mathbf{X}(0) = \mathbf{x}$, $\mathbf{x} \in \mathscr{S}$, there exists a unique solution $(\mathbf{X}(t))$ for the stochastic differential equation

$$d\mathbf{X}(t) = \tilde{\boldsymbol{\beta}}^Q(\mathbf{X}(t))dt + \boldsymbol{\rho}(\mathbf{X}(t))d\mathbf{W}(t), \quad \text{where } \tilde{\boldsymbol{\beta}}^Q(\mathbf{x}) \in \mathbb{R}^d \text{ and } \boldsymbol{\rho}(\mathbf{x}) \in \mathbb{R}^{d\times d}. \tag{27}$$

An affine stochastic process is defined as follows:

Definition 2 (Affine Process). *Consider $\mathbf{X}(t) \in \mathbb{R}^d$. $(\mathbf{X}(t))_{t\geq 0}$ described by the stochastic differential equation (27) is called affine stochastic process if the $\mathcal{F}_s$ conditional characteristic function of $\mathbf{X}(t)$ is exponentially affine in $\mathbf{X}(s)$, $0 \leq s \leq t$. Thus, there exist functions $\Phi(t, \mathbf{u}) \in \mathbb{C}$ and $\boldsymbol{\Psi}(t, \mathbf{u}) \in \mathbb{C}^d$, with jointly continuous t-derivatives, such that*

$$\mathbb{E}\left(\exp(\mathbf{u}'\mathbf{X}(t))|\mathcal{F}_s\right) = \exp\left(\Phi(t-s, \mathbf{u}) + \boldsymbol{\Psi}(t-s, \mathbf{u})'\mathbf{X}(s)\right) \tag{28}$$

for all $\mathbf{u} \in \imath\mathbb{R}^d$ and $s \leq t$.

As the conditional characteristic function is bounded by one, the real part of the exponent $\Phi(t-s, \mathbf{u}) + \boldsymbol{\Psi}(t-s, \mathbf{u})'\mathbf{X}(s)$ is negative. The functions $\Phi(t, \mathbf{u})$ and $\boldsymbol{\Psi}(t, \mathbf{u})$ are uniquely determined by (28) for $t \geq 0$ and $\mathbf{u} \in \imath\mathbb{R}^d$ and satisfy the initial conditions $\Phi(0, \mathbf{u}) = 0$ and $\boldsymbol{\Psi}(0, \mathbf{u}) = \mathbf{u}$.

If $(\mathbf{X}(t))_{t\in\mathbb{R}_+}$ is affine, then the *drift term* $\tilde{\boldsymbol{\beta}}^Q(\mathbf{X}(t))$ and the (positive definite) *diffusion matrix* $\mathbf{a}(\mathbf{X}(t)) = \boldsymbol{\rho}(\mathbf{X}(t))\boldsymbol{\rho}(\mathbf{X}(t))'$ are affine functions in $\mathbf{X}(t)$ (see Filipović (2009)[Definition 10.1 and Theorem 10.1]); i.e., $\tilde{\boldsymbol{\beta}}^Q(\mathbf{x}) = \mathbf{b}^Q + \sum_{i=1}^d x_i \boldsymbol{\beta}_i^Q$ and $\mathbf{a}(\mathbf{x}) = \mathbf{a} + \sum_{i=1}^d x_i \boldsymbol{\alpha}_i$ where $\mathbf{b}^Q$, $\boldsymbol{\beta}_i^Q$ and $\mathbf{x}$ are vectors of dimension d and $\mathbf{a}(\mathbf{x})$, $\mathbf{a}$ and $\boldsymbol{\alpha}_i$ are $d \times d$ matrices. $\boldsymbol{\beta}^Q = (\boldsymbol{\beta}_1^Q, \ldots, \boldsymbol{\beta}_d^Q)$ is a $d \times d$ matrix. In addition,

$\Phi(t, \mathbf{u})$ and $\boldsymbol{\Psi}(t, \mathbf{u})$ solve the following system of Riccati equations; see Filipović (2009)[Eq. 10.4][18]

$$\begin{aligned}
\partial_t \Phi(t, \mathbf{u}) &= \frac{1}{2} \boldsymbol{\Psi}(t, \mathbf{u})' \mathbf{a} \boldsymbol{\Psi}(t, \mathbf{u}) + (\mathbf{b}^Q)' \boldsymbol{\Psi}(t, \mathbf{u}), & \Phi(0, \mathbf{u}) &= 0, \\
\partial_t \boldsymbol{\Psi}_i(t, \mathbf{u}) &= \frac{1}{2} \boldsymbol{\Psi}(t, \mathbf{u})' \boldsymbol{\alpha}_i \boldsymbol{\Psi}(t, \mathbf{u}) + (\boldsymbol{\beta}_i^Q)' \boldsymbol{\Psi}(t, \mathbf{u}), & \boldsymbol{\Psi}(0, \mathbf{u}) &= \mathbf{u},
\end{aligned} \tag{29}$$

$i = 1, \ldots, d$ and $\mathbf{u} \in \imath \mathbb{R}^d$.

B Matrix A for $\mathbb{A}_m(3)$ Models

This section derives the matrix $\mathbf{A}$ for an arbitrary $\mathbb{A}_m(3)$ setting; where $0 \leq m \leq 3$. In the first step we ignore all the restrictions arising from admissibility, the boundary conditions, stationarity and identification, and calculate $\mathbf{A}$ for a model with diagonal diffusion, where all elements in $\mathbf{b}^P$, $\boldsymbol{\beta}^P$, $\boldsymbol{\Sigma}$, $\mathcal{B}^0$ and $\mathcal{B}^x$ are free parameters. To obtain $\mathbf{A}$ for a particular $\mathbb{A}_m(3)$ model, the corresponding parameter restrictions have to be taken into consideration. Moreover, restrictions like $\beta_{ij}^Q = \beta_{ij}^P$ for some ij can also be included. This allows a joint treatment of all models.

For the first four moments $\mathbf{x}^k$, $k = 1, \ldots, p = 4$, we choose the basis $\left(1 \mid x_1, x_2, x_3 \mid x_1^2, \ldots, x_3^2 \mid x_1^3, \ldots, x_3^3 \mid x_1^4, \ldots, x_3^4\right)$. In this expression we have separated the terms of different power by means of $\mid$. I.e., with $d = 3$, we get one term for $k = 0$, three for $k = 1$, six for $k = 2$, ten for $k = 3$ and fifteen for $k = 4$. Therefore $N = 35$. The elements of matrix $\mathbf{A}$ not presented are zero by the model assumptions.

In the following we use (19) and start with $k = 0$: Here we immediately observe that the first row of $\mathbf{A}$ is $\mathbf{A}_{1,:} = \mathbf{0}_{1 \times N}$. With $k = 1$ we obtain the rows 2 to $d + 1$ of the matrix $\mathbf{A}$ as follows: With $f(\mathbf{x}) = x_i$ we get $\frac{\partial x_i}{\partial x_i} = 1$, $\frac{\partial x_j}{\partial x_i} = 0$ and $\frac{\partial^2 x_i}{\partial x_i^2} = 0$. Hence, $\mathcal{G}(x_i) = b_i^P + \boldsymbol{\beta}_i^P \mathbf{x}$, $i = 1, \ldots, d$. This yields

$$\mathbf{A}_{2:4,:} = \left(\begin{array}{c|ccc|c} b_1^P & \beta_{11}^P & \beta_{12}^P & \beta_{13}^P & 0 \ldots \\ b_2^P & \beta_{21}^P & \beta_{22}^P & \beta_{23}^P & 0 \ldots \\ b_3^P & \beta_{31}^P & \beta_{32}^P & \beta_{33}^P & 0 \ldots \end{array} \right).$$

---

[18]Extensions with jumps are possible - for some theory see Keller-Ressel and Mayerhofer (2012), Mayerhofer et al. (2010), Duffie et al. (2000), Duffie et al. (2003).

Next, for $k = 2$ we have to consider $d(d+1)/2 = d_2$ basis elements, corresponding to rows $d+2$ to $d+1+\frac{d(d+1)}{2}$ of $\mathbf{A}$. We arrange the basis elements as follows $\mathbf{x}^2 = \left(x_1^2,\ x_1 x_2,\ x_1 x_3,\ x_2^2,\ x_2 x_3,\ x_3^2\right)$. Since the diffusion matrix is diagonal we have only non-zero elements in the generator for $i = j$.

The first partial derivatives with respect to x_1 are $2x_1$, x_2, x_3, 0, 0, 0 for these basis elements. The second partial derivatives with respect to x_1 are 2, 0 ,0, 0, 0, 0, etc. For x_2 and x_3 we proceed in the same way.

For example, consider $f(\mathbf{x}) = x_1^2$ and thus equation (19) yields $\mathcal{G}(x_1^2) = \left(b_i^P + \boldsymbol{\beta}_i^P \mathbf{x}\right) 2x_1 + \frac{1}{2}\sum_{j=1}^d \left(\Sigma_1^2 \left(\mathcal{B}_1^0 + \mathcal{B}_{j1}^x x_j\right)\right) 2$. For $f(\mathbf{x}) = x_1 x_2$, where $\frac{\partial (x_1 x_2)}{\partial x_1} = x_2$, $\frac{\partial (x_1 x_2)}{\partial x_2} = x_1$ and $\frac{\partial^2 (x_1 x_2)}{\partial x_1 x_2} = 1$, (19) and the fact that $\mathbf{S}(\mathbf{X}(t))$ and $\boldsymbol{\Sigma}$ are diagonal matrices,[19] result in $\mathcal{G}(x_1 x_2) = \left(b_1^P + \boldsymbol{\beta}_1^P \mathbf{x}\right) x_2 + \left(b_2^P + \boldsymbol{\beta}_2^P \mathbf{x}\right) x_1 + \frac{1}{2}\left[\boldsymbol{\Sigma}\mathbf{S}\right]_{12} \cdot 1 + \frac{1}{2}\left[\boldsymbol{\Sigma}\mathbf{S}\right]_{21} \cdot 1$. With $x_1 x_3, \ldots, x_3^2$ we proceed in the same way. This results in

$\mathbf{A}_{5:10,1:10}$

$$
= \left(
\begin{array}{c|ccc|ccccc|cc}
\Sigma_1^2 \mathcal{B}_1^0 & 2b_1^P + \Sigma_1^2 \mathcal{B}_{11}^x & \Sigma_1^2 \mathcal{B}_{21}^x & \Sigma_1^2 \mathcal{B}_{31}^x & 2\beta_{11}^P & 2\beta_{12}^P & 2\beta_{13}^P & 0 & 0 & 0 & 0 & \ldots \\
0 & b_2^P & b_1^P & 0 & \beta_{21}^P & \beta_{11}^P + \beta_{22}^P & \beta_{23}^P & \beta_{12}^P & \beta_{13}^P & 0 & 0 & \ldots \\
0 & b_3^P & 0 & b_1^P & \beta_{31}^P & \beta_{32}^P & \beta_{11}^P + \beta_{33}^P & 0 & \beta_{12}^P & \beta_{13}^P & 0 & \ldots \\
\Sigma_2^2 \mathcal{B}_2^0 & \Sigma_2^2 \mathcal{B}_{12}^x & 2b_2^P + \Sigma_2^2 \mathcal{B}_{22}^x & \Sigma_2^2 \mathcal{B}_{32}^x & 0 & 2\beta_{21}^P & 0 & 2\beta_{22}^P & 2\beta_{23}^P & 0 & 0 & \ldots \\
0 & 0 & b_3^P & b_2^P & 0 & \beta_{31}^P & \beta_{21}^P & \beta_{32}^P & \beta_{22}^P + \beta_{33}^P & \beta_{23}^P & 0 & \ldots \\
\Sigma_3^2 \mathcal{B}_3^0 & \Sigma_3^2 \mathcal{B}_{13}^x & \Sigma_3^2 \mathcal{B}_{23}^x & 2b_3^P + \Sigma_3^2 \mathcal{B}_{33}^x & 0 & 0 & 2\beta_{31}^P & 0 & 2\beta_{32}^P & 2\beta_{33}^P & 0 & \ldots
\end{array}
\right).
$$

For $k = 3$ we obtain $\binom{5}{3} = 10 = d_3$ elements. Therefore we consider the rows 11 to 20. The basis elements are $\mathbf{x}^3 = \left(x_1^3,\ x_1^2 x_2,\ x_1^2 x_3,\ x_1 x_2^2,\ x_1 x_2 x_3,\ x_1 x_3^2,\ x_2^3,\ x_2^2 x_3,\ x_2 x_3^2,\ x_3^3\right)$. Then,

$\mathbf{A}_{11:20,1:10} =$

$$
\left(
\begin{array}{c|ccc|ccc|ccc}
0 & 3\Sigma_1^2 \mathcal{B}_1^0 & 0 & 0 & 3b_1^P + 3\Sigma_1^2 \mathcal{B}_{11}^x & 3\Sigma_1^2 \mathcal{B}_{21}^x & 3\Sigma_1^2 \mathcal{B}_{31}^x & 0 & 0 & 0 \\
0 & 0 & \Sigma_1^2 \mathcal{B}_1^0 & 0 & b_2^P & 2b_1^P + \Sigma_1^2 \mathcal{B}_{11}^x & 0 & \Sigma_1^2 \mathcal{B}_{21}^x & \Sigma_1^2 \mathcal{B}_{31}^x & 0 \\
0 & 0 & 0 & \Sigma_1^2 \mathcal{B}_1^0 & b_3^P & 0 & 2b_1^P + \Sigma_1^2 \mathcal{B}_{11}^x & 0 & \Sigma_1^2 \mathcal{B}_{21}^x & \Sigma_1^2 \mathcal{B}_{31}^x \\
0 & \Sigma_2^2 \mathcal{B}_2^0 & 0 & 0 & \Sigma_2^2 \mathcal{B}_{12}^x & 2b_2^P + \Sigma_2^2 \mathcal{B}_{22}^x & \Sigma_2^2 \mathcal{B}_{32}^x & b_1^P & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & b_3^P & b_2^P & 0 & b_1^P & 0 \\
0 & \Sigma_3^2 \mathcal{B}_3^0 & 0 & 0 & \Sigma_3^2 \mathcal{B}_{13}^x & \Sigma_3^2 \mathcal{B}_{23}^x & 2b_3^P + \Sigma_3^2 \mathcal{B}_{33}^x & 0 & 0 & b_1^P \\
0 & 0 & 3\Sigma_2^2 \mathcal{B}_2^0 & 0 & 0 & 3\Sigma_2^2 \mathcal{B}_{12}^x & 0 & 3b_2^P + 3\Sigma_2^2 \mathcal{B}_{22}^x & 3\Sigma_2^2 \mathcal{B}_{32}^x & 0 \\
0 & 0 & 0 & \Sigma_2^2 \mathcal{B}_2^0 & 0 & 0 & \Sigma_2^2 \mathcal{B}_{12}^x & b_3^P & 2b_2^P + \Sigma_2^2 \mathcal{B}_{22}^x & \Sigma_2^2 \mathcal{B}_{32}^x \\
0 & 0 & \Sigma_3^2 \mathcal{B}_3^0 & 0 & 0 & \Sigma_3^2 \mathcal{B}_{13}^x & 0 & \Sigma_3^2 \mathcal{B}_{23}^x & 2b_3^P + \Sigma_3^2 \mathcal{B}_{33}^x & b_2^P \\
0 & 0 & 0 & 3\Sigma_3^2 \mathcal{B}_3^0 & 0 & 0 & 3\Sigma_3^2 \mathcal{B}_{13}^x & 0 & 3\Sigma_3^2 \mathcal{B}_{23}^x & 3b_3^P + 3\Sigma_3^2 \mathcal{B}_{33}^x
\end{array}
\right).
$$

---

[19] Where $S_{ii}(\mathbf{X}(t)) = \mathcal{B}_i^0 + (\mathcal{B}_i^x)' \mathbf{X}(t)$ and $S_{ij}(\mathbf{X}(t)) = 0$, $i, j = 1, \ldots, d$.

and

$$\mathbf{A}_{11:20,11:20} =$$

$$\left(\begin{array}{cccccccccc|cc}
3\beta_{11}^P & 3\beta_{12}^P & 3\beta_{13}^P & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
\beta_{21}^P & 2\beta_{11}^P+\beta_{22}^P & \beta_{23}^P & 2\beta_{12}^P & 2\beta_{13}^P & 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\
\beta_{31}^P & \beta_{32}^P & 2\beta_{11}^P+\beta_{33}^P & 0 & 2\beta_{12} & 2\beta_{13} & 0 & 0 & 0 & 0 & 0 & \ldots \\
0 & 2\beta_{21}^P & 0 & \beta_{11}^P+2\beta_{22}^P & 2\beta_{23}^P & 0 & \beta_{12}^P & \beta_{13}^P & 0 & 0 & 0 & \ldots \\
0 & \beta_{31}^P & \beta_{21}^P & \beta_{32}^P & \beta_{11}^P+\beta_{22}^P+\beta_{33}^P & \beta_{23}^P & 0 & \beta_{12}^P & \beta_{13}^P & 0 & 0 & \ldots \\
0 & 0 & 2\beta_{31}^P & 0 & 2\beta_{32}^P & \beta_{11}^P+2\beta_{33}^P & 0 & 0 & \beta_{12}^P & \beta_{13}^P & 0 & \ldots \\
0 & 0 & 0 & 3\beta_{21}^P & 0 & 0 & 3\beta_{22}^P & 3\beta_{23}^P & 0 & 0 & 0 & \ldots \\
0 & 0 & 0 & \beta_{31}^P & 2\beta_{21}^P & 0 & \beta_{32}^P & 2\beta_{22}^P+\beta_{33}^P & 2\beta_{23}^P & 0 & 0 & \ldots \\
0 & 0 & 0 & 0 & 2\beta_{31}^P & \beta_{21}^P & 0 & 2\beta_{32}^P & 2\beta_{33}^P+\beta_{22}^P & \beta_{23}^P & 0 & \ldots \\
0 & 0 & 0 & 0 & 0 & 3\beta_{31}^P & 0 & 0 & 3\beta_{32}^P & 3\beta_{33}^P & 0 & \ldots
\end{array}\right).$$

Last but not least, with $k = 4$ we have $d_4 = 15$ basis elements

$\mathbf{x}^4 = \left(x_1^4,\ x_1^3 x_2,\ x_1^3 x_3,\ x_1^2 x_2^2,\ x_1^2 x_2 x_3,\ x_1^2 x_3^2,\ x_1 x_2^3,\ x_1 x_2^2 x_3,\ x_1 x_2 x_3^2,\ x_1 x_3^3,\ x_2^4,\ x_2^3 x_3,\ x_2^2 x_3^2,\ x_2 x_3^3,\ x_3^4\right)$. Then we obtain:

31

$$\mathbf{A}_{21:35,1:10} =$$

$$
\left(
\begin{array}{cccc|cccccc}
0 & 0 & 0 & 0 & 6\Sigma_1^2\mathcal{B}_1^0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 3\Sigma_1^2\mathcal{B}_1^0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3\Sigma_1^2\mathcal{B}_1^0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \Sigma_2^2\mathcal{B}_2^0 & 0 & 0 & \Sigma_1^2\mathcal{B}_1^0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Sigma_1^2\mathcal{B}_1^0 & 0 \\
0 & 0 & 0 & 0 & \Sigma_3^2\mathcal{B}_3^0 & 0 & 0 & 0 & 0 & \Sigma_1^2\mathcal{B}_1^0 \\
0 & 0 & 0 & 0 & 0 & 3\Sigma_2^2\mathcal{B}_2^0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \Sigma_2^2\mathcal{B}_2^0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \Sigma_3^2\mathcal{B}_3^0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 3\Sigma_3^2\mathcal{B}_3^0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 6\Sigma_2^2\mathcal{B}_2^0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3\Sigma_2^2\mathcal{B}_2^0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \Sigma_3^2\mathcal{B}_3^0 & 0 & \Sigma_2^2\mathcal{B}_2^0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3\Sigma_3^2\mathcal{B}_3^0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6\Sigma_3^2\mathcal{B}_3^0
\end{array}
\right)
\tag{30}
$$

$$\mathbf{A}_{21:35,\,11:20} =$$

$$
\begin{pmatrix}
4b_1^P+6\Sigma_1^2\mathcal{B}_{11}^x & 6\Sigma_1^2\mathcal{B}_{21}^x & 6\Sigma_1^2\mathcal{B}_{31}^x & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
b_2^P & 3b_1^P+3\Sigma_1^2\mathcal{B}_{11}^x & 0 & 3\Sigma_1^2\mathcal{B}_{21}^x & 0 & 0 & 0 & 0 & 0 & 0 \\
b_3^P & 0 & 3b_1^P+3\Sigma_1^2\mathcal{B}_{11}^x & 3\Sigma_1^2\mathcal{B}_{21}^x & 3\Sigma_1^2\mathcal{B}_{31}^x & 0 & 0 & 0 & 0 & 0 \\
\Sigma_2^2\mathcal{B}_{12}^x & 2b_2^P+\Sigma_2^2\mathcal{B}_{22}^x & \Sigma_2^2\mathcal{B}_{32}^x & 0 & 0 & 0 & \Sigma_1^2\mathcal{B}_{31}^x & \Sigma_1^2\mathcal{B}_{31}^x & \Sigma_1^2\mathcal{B}_{31}^x & \Sigma_1^2\mathcal{B}_{31}^x \\
0 & b_3^P & b_2^P & 2b_1^P+\Sigma_1^2\mathcal{B}_{11}^x & 0 & 0 & \Sigma_1^2\mathcal{B}_{21}^x & \Sigma_1^2\mathcal{B}_{21}^x & \Sigma_1^2\mathcal{B}_{21}^x & 0 \\
\Sigma_3^2\mathcal{B}_{13}^x & \Sigma_3^2\mathcal{B}_{23}^x & 2b_3^P+\Sigma_3^2\mathcal{B}_{33}^x & 0 & 2b_1^P+\Sigma_1^2\mathcal{B}_{11}^x & 0 & 0 & b_1^P & b_1^P & b_1^P \\
0 & 3\Sigma_2^2\mathcal{B}_{12}^x & 0 & 3\Sigma_2^2\mathcal{B}_{32}^x & 3\Sigma_2^2\mathcal{B}_{32}^x & 4b_2^P+6\Sigma_2^2\mathcal{B}_{22}^x & b_1^P & 0 & 0 & 0 \\
0 & 0 & \Sigma_2^2\mathcal{B}_{12}^x & 2b_2^P+\Sigma_2^2\mathcal{B}_{22}^x & b_2^P & b_3^P & 0 & 0 & 0 & 0 \\
0 & \Sigma_3^2\mathcal{B}_{13}^x & 0 & 2b_3^P+\Sigma_3^2\mathcal{B}_{33}^x & 3b_3^P+3\Sigma_3^2\mathcal{B}_{33}^x & \Sigma_3^2\mathcal{B}_{23}^x & 3\Sigma_2^2\mathcal{B}_{32}^x & 6\Sigma_2^2\mathcal{B}_{32}^x & 3\Sigma_2^2\mathcal{B}_{32}^x & 0 \\
0 & 0 & 3\Sigma_3^2\mathcal{B}_{13}^x & 3\Sigma_3^2\mathcal{B}_{23}^x & 0 & 0 & 2b_2^P+\Sigma_2^2\mathcal{B}_{22}^x & 3b_2^P+3\Sigma_2^2\mathcal{B}_{22}^x & 2b_2^P+\Sigma_2^2\mathcal{B}_{22}^x & \Sigma_2^2\mathcal{B}_{32}^x \\
0 & 0 & 0 & 3\Sigma_2^2\mathcal{B}_{12}^x & 0 & \Sigma_2^2\mathcal{B}_{12}^x & 3b_3^P+3\Sigma_3^2\mathcal{B}_{33}^x & 2b_3^P+\Sigma_3^2\mathcal{B}_{33}^x & 3b_3^P+3\Sigma_3^2\mathcal{B}_{33}^x & b_2^P \\
0 & 0 & 0 & 0 & \Sigma_2^2\mathcal{B}_{12}^x & 0 & 6\Sigma_3^2\mathcal{B}_{23}^x & 3\Sigma_3^2\mathcal{B}_{23}^x & 6\Sigma_3^2\mathcal{B}_{23}^x & 4b_3^P+6\Sigma_3^2\mathcal{B}_{33}^x \\
0 & 0 & 0 & 3\Sigma_3^2\mathcal{B}_{13}^x & 0 & 6\Sigma_3^2\mathcal{B}_{13}^x & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \Sigma_3^2\mathcal{B}_{13}^x & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.
$$

$$
A_{21:35,\,21:35} =
\begin{pmatrix}
4\beta_{11} & \beta_{12} & \beta_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\[2pt]
4\beta_{21} & 3\beta_{11}+\beta_{22} & \beta_{23} & 2\beta_{12} & \beta_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\[2pt]
4\beta_{31} & \beta_{32} & 3\beta_{11}+\beta_{33} & 0 & \beta_{12} & 2\beta_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\[2pt]
0 & 3\beta_{21} & 0 & 2\beta_{11}+2\beta_{22} & \beta_{23} & 0 & 3\beta_{12} & \beta_{13} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\[2pt]
0 & 3\beta_{31} & 0 & 2\beta_{32} & 2\beta_{11}+\beta_{22}+\beta_{33} & 2\beta_{23} & 0 & 2\beta_{12} & 2\beta_{13} & 0 & 0 & 0 & 0 & 0 & 0 \\[2pt]
0 & 0 & 3\beta_{31} & 0 & \beta_{32} & 2\beta_{11}+2\beta_{33} & 0 & 0 & \beta_{12} & 3\beta_{13} & 0 & 0 & 0 & 0 & 0 \\[2pt]
0 & 0 & 0 & 2\beta_{21} & 0 & 0 & \beta_{11}+3\beta_{22} & \beta_{23} & 0 & 0 & 4\beta_{12} & \beta_{13} & 0 & 0 & 0 \\[2pt]
0 & 0 & 0 & 2\beta_{31} & 2\beta_{21} & 0 & 3\beta_{32} & \beta_{11}+2\beta_{22}+\beta_{33} & 2\beta_{23} & 0 & 0 & 3\beta_{12} & 2\beta_{13} & 0 & 0 \\[2pt]
0 & 0 & 0 & 0 & 2\beta_{31} & 2\beta_{21} & 0 & 2\beta_{32} & \beta_{11}+\beta_{22}+2\beta_{33} & 3\beta_{23} & 0 & 0 & 2\beta_{12} & 3\beta_{13} & 0 \\[2pt]
0 & 0 & 0 & 0 & 0 & 2\beta_{31} & 0 & 0 & \beta_{32} & \beta_{11}+3\beta_{33} & 0 & 0 & 0 & \beta_{12} & 4\beta_{13} \\[2pt]
0 & 0 & 0 & 0 & 0 & 0 & \beta_{21} & 0 & 0 & 0 & 4\beta_{22} & \beta_{23} & 0 & 0 & 0 \\[2pt]
0 & 0 & 0 & 0 & 0 & 0 & \beta_{31} & \beta_{21} & 0 & 0 & 4\beta_{32} & 3\beta_{22}+\beta_{33} & 2\beta_{23} & 0 & 0 \\[2pt]
0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{31} & \beta_{21} & 0 & 0 & 3\beta_{32} & 2\beta_{22}+2\beta_{33} & 3\beta_{23} & 0 \\[2pt]
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{31} & \beta_{21} & 0 & 0 & 2\beta_{32} & \beta_{22}+3\beta_{33} & 4\beta_{23} \\[2pt]
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_{31} & 0 & 0 & 0 & \beta_{32} & 4\beta_{33}
\end{pmatrix}
$$

C Moments of the Observed Yields

The following paragraphs obtain the first four moments of the yields observed, i.e. $\mathbb{E}\left(y_{\mathsf{t}i}^k\right)$, $k = 1, \ldots, 4$, the auto-covariance of the yields, $\mathbb{E}\left(y_{\mathsf{t}i}y_{\mathsf{t}-1,i}\right)$, and the auto-covariance of the squared yields, $\mathbb{E}\left(y_{\mathsf{t}i}^2 y_{\mathsf{t}-1,i}^2\right)$. Assumption 2 specifies the moments $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}^k \varepsilon_{\mathsf{t}i}^l\right)$.[20] If p moments of $\mathbf{y_t}$ should be considered, we get the number of moments by summing over the multinomial coefficients, i.e.

$$N_y = \sum_{j=1}^{p} \binom{j + M - 1}{j}. \tag{31}$$

Powers of sums can be obtained by means of the *multinomial formula*. With $k = \sum_{i=1}^{d} l_i$, $l_i \geq 0$, we get

$$(x_1 + x_2 + \cdots + x_d)^k = \sum_{l_1 + l_2 + \cdots + l_d = k} \binom{k}{l_1, l_2, \ldots, l_d} \prod_{1 \leq i \leq d} x_i^{l_i}, \tag{32}$$

where $\binom{k}{l_1, l_2, \ldots, l_d} = \frac{k!}{l_1! l_2! \cdots l_d!}$. Let

$$d_{(i,K)} = \binom{i + K - 1}{i} \tag{33}$$

for $K \in \mathbb{N}$ and $i \leq p$. In accordance with equation (8), we write $d_i \equiv d_{(i,d)}$; when $K = d$. That is, the notation is simplified when $K = d$. Note that d_i calculates the dimension of conditional moments $\mathbb{E}(\mathbf{X}(t)^i|\mathbf{X}(s) = \mathbf{x})$. In addition, $N_i = \sum_{j=0}^{i} d_j$ corresponds to the sum of the conditional moments smaller or equal to i.

We shall derive the first four moments, which implies that $p = 4$ in the following. From (22) we get

---

[20]Here we derive the 1st moments with $k = 1$, $l = 0$, where $\mathbb{E}(\varepsilon_{\mathsf{t}i}) = 0$ for $i = 1, \ldots, M$. For the 2nd moments: $k = 2$, $l = 0$, such that $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}^2\right) = \sigma_i^2$ for all i; with $k = l = 1$, $i \neq j$ we get $\mathbb{E}(\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j}) = 0$, $i \neq j$. For the 3rd moments: $k = 3$, $l = 0$, all i, $k = 2$, $l = 1$, $i \neq j$, and $k = 1$, $l = 2$, $i \neq j$. All these terms are zero by assumption, i.e. $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}j}\right) = 0$, and $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j}^2\right) = 0$, $i \neq j$. For the 4th moments: $k = 4$, $l = 0$, all i, $k = 3$, $l = 1$, $i \neq j$, $k = l = 2$, $i \neq j$, $k = 1$, $l = 3$, $i \neq j$, $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}^3\varepsilon_{\mathsf{t}j}\right) = 0$, $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}j}^2\right) = 0$, and $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j}^3\right) = 0$, $i \neq j$, $\mathbb{E}\left(\varepsilon_{\mathsf{t}i}^4\right) = \sigma_i^4$. Note that σ_i^4 stands for the fourth moment of $\varepsilon_{\mathsf{t}i}$, where in general $(\sigma_i^2)^2 \neq \sigma_i^4$.

the first moments by means of

$$\mathbb{E}(\mathbf{y_t}) = \tilde{\boldsymbol{\Phi}} + \tilde{\boldsymbol{\Psi}}\mathbb{E}(\mathbf{X_t}) \, ,$$

$$\mathbb{E}(\mathbf{y_{ti}}) = \Phi_i + \boldsymbol{\Psi}_i' \mathbb{E}(\mathbf{X_t}) = \Phi_i + \boldsymbol{\Psi}_i' \mathbb{E}(\tilde{\mathbf{X}}_{\mathbf{t},1:3}) \, , \tag{34}$$

where $\Phi_i = -\frac{1}{\tau_i}\Phi(\tau_i,\mathbf{0}) \in \mathbb{R}$, $\boldsymbol{\Psi}_i = -\frac{1}{\tau_i}\boldsymbol{\Psi}(\tau_i,\mathbf{0}) \in \mathbb{R}^d$, $\tilde{\boldsymbol{\Phi}} \in \mathbb{R}^M$, $\mathbf{X_t} \in \mathbb{R}^d$, $\mathbf{y_t} \in \mathbb{R}^M$ and $\tilde{\boldsymbol{\Psi}} \in \mathbb{R}^{M \times d}$.

The second moments of the yields are given by:

$$\mathbb{E}(y_{\mathbf{t}i}y_{\mathbf{t}j}) = \Phi_i\Phi_j + \left(\Phi_i\boldsymbol{\Psi}_j' + \Phi_j\boldsymbol{\Psi}_i'\right)\mathbb{E}(\mathbf{X_t}) + \boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X_t}\mathbf{X_t'})\boldsymbol{\Psi}_j + \mathbb{E}(\varepsilon_{\mathbf{t}i}\varepsilon_{\mathbf{t}j})$$

$$= \Phi_i\Phi_j + \left(\Phi_i\boldsymbol{\Psi}_j' + \Phi_j\boldsymbol{\Psi}_i'\right)\mathbb{E}(\tilde{\mathbf{X}}_{\mathbf{t},1:d}) + \boldsymbol{\Psi}_i'\mathbb{E}\left(vech^{-1}(\tilde{\mathbf{X}}_{\mathbf{t},d+1:d+d_2})\right)\boldsymbol{\Psi}_j + \mathbb{E}(\varepsilon_{\mathbf{t}i}\varepsilon_{\mathbf{t}j}) \, , \tag{35}$$

for $i,j = 1,\ldots,M$. In (35) we need the function $vech^{-1}$. The purpose of this function is to transform the $d(d+1)/2 \times 1$ vector $\tilde{\mathbf{X}}_{\mathbf{t},d+1:d+d_2}$ into the symmetric $d \times d$ matrix $\mathbf{X_t}\mathbf{X_t'}$. In more details, $\tilde{\mathbf{X}}_{\mathbf{t},d+1:d+d_2} = vech\left(\mathbf{X_t}\mathbf{X_t'}\right)$, where $vec\left(\mathbf{X_t}\mathbf{X_t'}\right)$ vectorizes the $d \times d$ matrix $\mathbf{X_t}\mathbf{X_t'}$ and $vech\left(\mathbf{X_t}\mathbf{X_t'}\right)$ eliminates the supra-diagonal elements from the $d^2 \times 1$ vector $vec\left(\mathbf{X_t}\mathbf{X_t'}\right)$ (see, e.g., Poirier, 1995, page 646). Hence $vech\left(\mathbf{X_t}\mathbf{X_t'}\right)$ is a $d(d+1)/2 \times 1$ vector. The function $vech^{-1}$ takes us back to $\mathbf{X_t}\mathbf{X_t'}$, i.e. $vech^{-1}$ maps the $d(d+1)/2 \times 1$ vector $\mathbb{E}(\tilde{\mathbf{X}}_{\mathbf{t},d+1:d+d_2})$ to the symmetric $d \times d$ matrix $\mathbb{E}(\mathbf{X_t}\mathbf{X_t'})$. For $d = 3$ this works as follows:

$$vech^{-1}\begin{pmatrix} a_1 \\ \ldots \\ a_6 \end{pmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_4 & a_5 \\ a_3 & a_5 & a_6 \end{bmatrix} \tag{36}$$

and thus $vech^{-1}(\tilde{\mathbf{X}}_{\mathbf{t},d+1:d+d_2}) = \mathbf{X_t}\mathbf{X_t'}$. By Assumption 2 we obtain $\mathbb{E}(X_{\mathbf{t}l}\varepsilon_{\mathbf{t}i}) = 0$ for $l = 1,\ldots,d$, $i = 1,\ldots,M$ and

$$\mathbb{E}(\varepsilon_{\mathbf{t}i}\varepsilon_{\mathbf{t}j}) = \begin{cases} \sigma_i^2, & \text{for } i = j \, , \\ 0, & \text{for } i \neq j \, , \end{cases} \tag{37}$$

for $i, j = 1, \ldots, M$. Based on this, (35) can be written as

$$\mathbb{E}(y_{\mathsf{t}i}y_{\mathsf{t}j}) = \Phi_i\Phi_j + \left(\Phi_i\boldsymbol{\Psi}_j' + \Phi_j\boldsymbol{\Psi}_i'\right)\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},1:d}) + (\mathbf{m}_2^{ij})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2}) + \mathbb{E}(\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j}), \tag{38}$$

where for $d = 3$ we define $\mathbf{m}_2^{ij} = (\Psi_{i1}\Psi_{j1}, \Psi_{i1}\Psi_{j2} + \Psi_{i2}\Psi_{j1}, \Psi_{i1}\Psi_{j3} + \Psi_{i3}\Psi_{j1}, \Psi_{i2}\Psi_{j2}, \Psi_{i2}\Psi_{j3} + \Psi_{i3}\Psi_{j2}, \Psi_{i3}\Psi_{j3})' = \mathbf{m}_2^{ji} \in \mathbb{R}^{d_2}$ and thus $(\mathbf{m}_2^{ij})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},4:9}) = \boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}\mathbf{X}_{\mathsf{t}}')\boldsymbol{\Psi}_j$. Regarding the third moments we observe:

$$
\begin{aligned}
\mathbb{E}(y_{\mathsf{t}i}^2 y_{\mathsf{t}j}) &= \mathbb{E}\left(\left(\Phi_i + \boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}} + \varepsilon_{\mathsf{t}i}\right)^2\left(\Phi_j + \boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}} + \varepsilon_{\mathsf{t}j}\right)\right) \\
&= \mathbb{E}\left((\Phi_i + \boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})^2(\Phi_j + \boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}}) + (\Phi_i + \boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})^2\varepsilon_{\mathsf{t}j} + 2(\Phi_i + \boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})(\Phi_j + \boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}})\varepsilon_{\mathsf{t}i}\right) \\
&\quad + \mathbb{E}\left(2(\Phi_i + \boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j} + (\Phi_j + \boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}})\varepsilon_{\mathsf{t}i}^2 + \varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}j}\right) \\
&= \mathbb{E}\left(\Phi_i^2\Phi_j + \Phi_i^2(\boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}}) + 2\Phi_i\Phi_j(\boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}}) + 2\Phi_i(\boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})(\boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}}) + \Phi_j(\boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})^2 + (\boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})^2(\boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}})\right) \\
&\quad + 2(\Phi_i + \boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}))\sigma_i^2\mathbb{I}_{(i=j)} + (\Phi_j + \boldsymbol{\Psi}_j'\mathbb{E}(\mathbf{X}_{\mathsf{t}}))\sigma_i^2 \\
&= \Phi_i^2\Phi_j + (\Phi_i^2\boldsymbol{\Psi}_j' + 2\Phi_i\Phi_j\boldsymbol{\Psi}_i')\mathbb{E}(\mathbf{X}_{\mathsf{t}}) + 2\Phi_i\boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}\mathbf{X}_{\mathsf{t}}')\boldsymbol{\Psi}_j + \Phi_j\boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}\mathbf{X}_{\mathsf{t}}')\boldsymbol{\Psi}_i + \mathbb{E}((\boldsymbol{\Psi}_i'\mathbf{X}_{\mathsf{t}})^2\boldsymbol{\Psi}_j'\mathbf{X}_{\mathsf{t}}) \\
&\quad + 2(\Phi_i + \boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}))\sigma_i^2\mathbb{I}_{(i=j)} + (\Phi_j + \boldsymbol{\Psi}_j'\mathbb{E}(\mathbf{X}_{\mathsf{t}}))\sigma_i^2 \\
&= \left(\Phi_i^2 + \sigma_i^2\right)\Phi_j + \left(\Phi_i^2\boldsymbol{\Psi}_j' + 2\Phi_i\Phi_j\boldsymbol{\Psi}_i' + \sigma_i^2\boldsymbol{\Psi}_j'\right)\mathbb{E}(\mathbf{X}_{\mathsf{t}}) + \left(2\Phi_i(\mathbf{m}_2^{ij})' + \Phi_j(\mathbf{m}_2^{ii})'\right)\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2}) \\
&\quad + (\mathbf{m}_3^{i^2j})'\mathbb{E}\left(\tilde{\mathbf{X}}_{\mathsf{t},d+d_2+1:d+d_2+d_3}\right) + 2(\Phi_i + \boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}))\sigma_i^2\mathbb{I}_{(i=j)} , \tag{39}
\end{aligned}
$$

where $\mathbf{m}_3^{i^2j} = \left(\Psi_{i1}^2\Psi_{j1}, \; \Psi_{i1}^2\Psi_{j2} + 2\Psi_{i1}\Psi_{i2}\Psi_{j1}, \; \Psi_{i1}^2\Psi_{j3} + 2\Psi_{i1}\Psi_{i3}\Psi_{j1}, \; \Psi_{i2}^2\Psi_{j1} + 2\Psi_{i1}\Psi_{i2}\Psi_{j2}, \; \Psi_{i3}^2\Psi_{j1} + 2\Psi_{i1}\Psi_{i3}\Psi_{j3}, 2(\Psi_{i1}\Psi_{i2}\Psi_{j3} + \Psi_{i1}\Psi_{i3}\Psi_{j2} + \Psi_{i2}\Psi_{i3}\Psi_{j1}), \; \Psi_{i2}^2\Psi_{j2}, \; \Psi_{i2}^2\Psi_{j3} + 2\Psi_{i2}\Psi_{i3}\Psi_{j2}, \; \Psi_{i3}^2\Psi_{j2} + 2\Psi_{i2}\Psi_{i3}\Psi_{j3}, \; \Psi_{i3}^2\Psi_{j3}\right)' \in \mathbb{R}^{d_3}$ and $\mathbb{I}_{(.)}$ stands for an indicator function. By Assumption 2 we get $\mathbb{E}(\mathbf{X}_{\mathsf{t}l}\varepsilon_{\mathsf{t}i}) = 0$, $\mathbb{E}(X_{\mathsf{t}l}^2\varepsilon_{\mathsf{t}i}) = 0$, $\mathbb{E}(X_{\mathsf{t}l}\varepsilon_{\mathsf{t}i}\varepsilon_{\mathsf{t}j}) = 0$, for $i \neq j$, and $\mathbb{E}(\varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}j}) = 0$ for $l = 1, \ldots, d$ and $i, j = 1, \ldots, M$. In a similar way and under the same assumptions, it can be shown that

$$
\begin{aligned}
\mathbb{E}\left(y_{\mathsf{t}i}y_{\mathsf{t}j}^2\right) &= \left(\Phi_j^2 + \sigma_j^2\right)\Phi_i + \left(\Phi_j^2\boldsymbol{\Psi}_i' + 2\Phi_i\Phi_j\boldsymbol{\Psi}_j' + \sigma_j^2\boldsymbol{\Psi}_i'\right)\mathbb{E}(\mathbf{X}_{\mathsf{t}}) + \left(2\Phi_j(m_2^{ij})' + \Phi_i(\mathbf{m}_2^{jj})'\right)\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2}) \\
&\quad + (\mathbf{m}_3^{ij^2})'\mathbb{E}\left(\tilde{\mathbf{X}}_{\mathsf{t},d+d_2+1:d+d_2+d_3}\right) + 2(\Phi_i + \boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathsf{t}}))\sigma_i^2\mathbb{I}_{(i=j)} , \tag{40}
\end{aligned}
$$

where $\mathbf{m}_3^{ij^2} = \left(\Psi_{i1}\Psi_{j1}^2, \; \Psi_{i2}\Psi_{j1}^2 + 2\Psi_{i1}\Psi_{j1}\Psi_{j2}, \; \Psi_{i3}\Psi_{j1}^2 + 2\Psi_{i1}\Psi_{j1}\Psi_{j3}, \; \Psi_{i1}\Psi_{j2}^2 + 2\Psi_{i2}\Psi_{j1}\Psi_{j2}, \; \Psi_{i1}\Psi_{j3}^2 + 2\Psi_{i3}\Psi_{j1}\Psi_{j3}, 2(\Psi_{i1}\Psi_{j2}\Psi_{j3} + \Psi_{i2}\Psi_{j1}\Psi_{j3} + \Psi_{i3}\Psi_{j1}\Psi_{j2}), \; \Psi_{i2}\Psi_{j2}^2, \; \Psi_{i3}\Psi_{j2}^2 + 2\Psi_{i2}\Psi_{j2}\Psi_{j3}, \; \Psi_{i2}\Psi_{j3}^2 + 2\Psi_{i3}\Psi_{j2}\Psi_{j3},$

$\left. \Psi_{i3}\Psi_{j3}^2 \right)' \in \mathbb{R}^{d_3}$ for $d = 3$. For the fourth moment we obtain

$$
\begin{aligned}
\mathbb{E}\left(y_{\mathsf{t}i}^2 y_{\mathsf{t}j}^2\right) &= \mathbb{E}\left(\left(\Phi_i + \boldsymbol{\Psi}_i'\mathbf{X_t} + \varepsilon_{\mathsf{t}i}\right)^2 \left(\Phi_j + \boldsymbol{\Psi}_j'\mathbf{X_t} + \varepsilon_{\mathsf{t}j}\right)^2\right) \\
&= \mathbb{E}\left(\left(\Phi_i^2 + (\boldsymbol{\Psi}_i'\mathbf{X_t})^2 + \varepsilon_{\mathsf{t}i}^2 + 2\Phi_i\boldsymbol{\Psi}_i'\mathbf{X_t} + 2\Phi_i\varepsilon_{\mathsf{t}i} + 2\boldsymbol{\Psi}_i'\mathbf{X_t}\varepsilon_{\mathsf{t}i}\right)\right. \\
&\quad \times \left. \left(\Phi_j^2 + (\boldsymbol{\Psi}_j'\mathbf{X_t})^2 + \varepsilon_{\mathsf{t}j}^2 + 2\Phi_j\boldsymbol{\Psi}_j'\mathbf{X_t} + 2\Phi_j\varepsilon_{\mathsf{t}j} + 2\boldsymbol{\Psi}_j'\mathbf{X_t}\varepsilon_{\mathsf{t}j}\right)\right) \\
&= \Phi_i^2\Phi_j^2 + \sigma_i^2\Phi_j^2 + \sigma_j^2\Phi_i^2 + \mathbb{E}(\varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}j}^2) \\
&\quad + 2\Phi_i\boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X_t})(\Phi_j^2 + \sigma_j^2) + 2\Phi_j\boldsymbol{\Psi}_j'\mathbb{E}(\mathbf{X_t})(\Phi_i^2 + \sigma_i^2) \\
&\quad + (\mathbf{m}_2^{ii})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2})(\Phi_j^2 + \sigma_j^2) \\
&\quad + (\mathbf{m}_2^{jj})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2})(\Phi_i^2 + \sigma_i^2) + 4\Phi_i\Phi_j(\mathbf{m}_2^{ij})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2}) \\
&\quad + 2\Phi_i(\mathbf{m}_3^{ij^2})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+d_2+1:d+d_2+d_3}) + 2\Phi_j(\mathbf{m}_3^{i^2j})'\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+d_2+1:d+d_2+d_3}) \\
&\quad + \mathbf{m}_4^{i^2j^2}\mathbb{E}(\tilde{\mathbf{X}}_{\mathsf{t},d+d_2+d_3+1:d+d_2+d_3+d_4}) \\
&\quad + 4\sigma_i^2\left[\Phi_i^2 + 2\Phi_i\boldsymbol{\Psi}_i'\mathbb{E}(\mathbf{X_t}) + (\mathbf{m}_2^{ii})'\tilde{\mathbf{X}}_{\mathsf{t},d+1:d+d_2}\right]\mathbb{I}_{(i=j)}\,,
\end{aligned}
\tag{41}
$$

where $\mathbf{m}_4^{i^2j^2}\mathbb{E}\left(\tilde{\mathbf{X}}_{\mathsf{t},d+d_2+d_3+1:d+d_2+d_3+d_4}\right) = \mathbb{E}\left((\boldsymbol{\Psi}_i'\mathbf{X_t})^2(\boldsymbol{\Psi}_j'\mathbf{X_t})^2\right)$. By sticking to Assumption 2 the expectation $\mathbb{E}(\varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}i}^2) = \mathbb{E}(\varepsilon_{\mathsf{t}i}^4)$ and $\mathbb{E}(\varepsilon_{\mathsf{t}i}^2\varepsilon_{\mathsf{t}j}^2) = \sigma_i^2\sigma_j^2$ for $j \neq i$. Moreover, $\mathbf{m}_4^{i^2j^2} = \left(\Psi_{i1}^2\Psi_{j1}^2,\right.$ $2\Psi_{i1}\Psi_{j1}(\Psi_{i2}\Psi_{j1} + \Psi_{i1}\Psi_{j2}), 2\Psi_{i1}\Psi_{j1}(\Psi_{i1}\Psi_{j3} + \Psi_{i3}\Psi_{j1}), \Psi_{i1}^2\Psi_{j2}^2 + \Psi_{i2}^2\Psi_{j1}^2 + 4\Psi_{i1}\Psi_{i2}\Psi_{j1}\Psi_{j2}, 4\Psi_{i1}\Psi_{j1}(\Psi_{i2}\Psi_{j3} + \Psi_{i3}\Psi_{j2}) + 2(\Psi_{i1}^2\Psi_{j2}\Psi_{j3} + \Psi_{i2}\Psi_{i3}\Psi_{j1}^2), \Psi_{i1}^2\Psi_{j3}^2 + \Psi_{i3}^2\Psi_{j1}^2 + 4\Psi_{i1}\Psi_{i3}\Psi_{j1}\Psi_{j3}, 2\Psi_{i2}\Psi_{j2}(\Psi_{i2}\Psi_{j1} + \Psi_{i1}\Psi_{j2}), 4\Psi_{i2}\Psi_{j2}(\Psi_{i3}\Psi_{j1} + \Psi_{i1}\Psi_{j3}) + 2(\Psi_{i1}\Psi_{i3}\Psi_{j2}^2 + \Psi_{i2}^2\Psi_{j1}\Psi_{j3}), 4\Psi_{i3}\Psi_{j3}(\Psi_{i1}\Psi_{j2} + \Psi_{i2}\Psi_{j1}) + 2(\Psi_{i1}\Psi_{i2}\Psi_{j3}^2 + \Psi_{i3}^2\Psi_{j1}\Psi_{j2}), 2\Psi_{i3}\Psi_{j3}(\Psi_{i3}\Psi_{j1} + \Psi_{i1}\Psi_{j3}), \Psi_{i2}^2\Psi_{j2}^2, 2\Psi_{i2}\Psi_{j2}(\Psi_{i3}\Psi_{j2} + \Psi_{i2}\Psi_{j3}), \Psi_{i2}^2\Psi_{j3}^2 + \Psi_{i3}^2\Psi_{j2}^2 + 4\Psi_{i2}\Psi_{i3}\Psi_{j2}\Psi_{j3}, 2\Psi_{i3}\Psi_{j3}(\Psi_{i2}\Psi_{j3} + \Psi_{i3}\Psi_{j2}), \left.\Psi_{i3}^2\Psi_{j3}^2\right)' \in \mathbb{R}^{d_4}$.

For the auto-covariance of the yields and the auto-covariance of the squared yields we have to calculate $\mathbb{E}(\mathbf{X_t^v}(\mathbf{X_s^w})')$ (which will become clear later). Before we proceed with these moments we obtain the result that only $\mathbf{x_s^\iota}$ with exponents $\iota \leq v$ enter into the calculation of the conditional moment $\mathbb{E}(\mathbf{X_t^v}|\mathbf{X_s} = \mathbf{x_s})$. In addition we derive a result on the structure of $\exp((t-s)\mathbf{A})$, which is presented in the following lemma:

Lemma 1. *Let $\mathbf{D}$ and $\mathbf{B}$ be $n \times n$ lower-block triangular matrices such that:* $\mathbf{D}_{m_i:n_i,n_i+1:n} = \mathbf{B}_{m_i:n_i,n_i+1:n} = 0$ *where $m_i \leq n_i$ for $i = 1, \ldots, k$, $k \leq n$, $n_k = n$ and $n_i < n_{i+1}$ for $i = 1, \ldots, k-1$. Then the*

matrix $\mathbf{C} = \mathbf{DB}$ *is of the same structure, namely* $\mathbf{C}_{m_i:n_i,n_i+1:n} = 0$ *where* $m_i \leq n_i$ *and* $n_i < n_{i+1}$ *for* $i = 1, \ldots, k-1$.

Proof: Let j and l be such that there exists $i \in \{1, \ldots, k\}$ such that $m_i \leq j \leq n_i$, $n_i + 1 \leq l \leq n$. Then $C_{jl} = (D_{j1}, \ldots, D_{jn_i}, 0, \ldots, 0)(0, \ldots, 0, B_{n_i+1,l}, \ldots, B_{nl})' = 0$. $\qquad\square$

Note that for a square matrix $\mathbf{B}$, $\exp(\mathbf{B}) = \sum_{i=0}^{+\infty} \frac{\mathbf{B}^i}{i!}$. Thus, if $\mathbf{B}$ is a matrix of the structure described in the lemma then $\exp(\mathbf{B})$ has the same structure as well. As the matrix $(t-s)\mathbf{A}$ is of the structure described in Lemma 1, this and the definition of $\exp((t-s)\mathbf{A})$ imply that also the matrix $\exp((t-s)\mathbf{A})$ is of that same structure. Thus,

$$\exp((t-s)\mathbf{A})_{N_v-1:N_v,N_v+1:N} = \mathbf{0},$$

which gives

$$\exp((t-s)\mathbf{A})_{N_{v-1}+1:N_v,:}\left[1, (\mathbf{x}^1)', (\mathbf{x}^2)', \ldots, (\mathbf{x}^p)'\right]'$$
$$= \left[\exp((t-s)\mathbf{A})_{N_{v-1}+1:N_v,1:N_v}, \exp((t-s)\mathbf{A})_{N_{v-1}+1:N_v,N_v+1:N}\right]\left[1, (\mathbf{x}^1)', \ldots, (\mathbf{x}^v)', (\mathbf{x}^{v+1})', \ldots, (\mathbf{x}^p)'\right]'$$
$$= \left[\exp((t-s)\mathbf{A})_{N_{v-1}+1:N_v,1:N_v}, \mathbf{0}_{d_v \times N-N_v}\right]\left[1, (\mathbf{x}^1)', \ldots, (\mathbf{x}^v)', (\mathbf{x}^{v+1})', \ldots, (\mathbf{x}^p)'\right]'$$
$$= \exp((t-s)\mathbf{A})_{N_{v-1}+1:N_v,1:N_v}\left[1, (\mathbf{x}^1)', \ldots, (\mathbf{x}^v)'\right]' + \mathbf{0}_{d_v \times N-N_v} \times \left[(\mathbf{x}^{v+1})', \ldots, (\mathbf{x}^p)'\right]'$$
$$= \exp((t-s)\mathbf{A})_{N_{v-1}+1:N_v,1:N_v}\left[1, (\mathbf{x}^1)', \ldots, (\mathbf{x}^v)'\right]'.$$

(13) and the above calculations show that: Only $\mathbf{x}^\iota$ with $\iota \leq v$ enter into the calculation of the conditional moment $\mathbb{E}(\mathbf{X}_t^v|\mathbf{X}_s = \mathbf{x})$. The conditional expectation of the v-th moment of $\mathbf{X}_t$ with respect to $\mathbf{X}_s = \mathbf{x}$ is

$$\mathbb{E}(\mathbf{X}_t^v|\mathbf{X}_s = \mathbf{x}) = \mathbb{E}\left(\tilde{\mathbf{X}}_{t,N_{v-1}+1:N_v}|\mathbf{X}_s = \mathbf{x}\right) = \exp((t-s)\Delta\mathbf{A})_{N_{v-1}+1:N_v,1:N_v}\left[1, (\mathbf{x}^1)', \ldots, (\mathbf{x}^v)'\right]'$$
$$= \exp((t-s)\Delta\mathbf{A})_{N_{v-1}+1:N_v,1} + \exp((t-s)\Delta\mathbf{A})_{N_{v-1}+1:N_v,2:1+d_1}\mathbf{x}^1 + \cdots$$
$$\cdots + \exp((t-s)\Delta\mathbf{A})_{N_{v-1}+1:N_v,N_{v-1}+1:N_v}\mathbf{x}^v, \tag{42}$$

which is of dimension $d_v \times 1$; $\Delta \in \mathbb{R}_{++}$ is the step-width already defined in Section 3.3. This implies:[21]

$$
\begin{aligned}
\mathbb{E}(\mathbf{X}_{\mathtt{t}}^v (\mathbf{X}_{\mathtt{s}}^w)') &= \mathbb{E}\left(\mathbb{E}(\mathbf{X}_{\mathtt{t}}^v | \mathbf{X}_{\mathtt{s}})(\mathbf{X}_{\mathtt{s}}^w)'\right) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{N_{v-1}+1:N_v,1}\mathbb{E}((\mathbf{X}_{\mathtt{s}}^w)') + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{N_{v-1}+1:N_v,2:1+d_1}\mathbb{E}(\mathbf{X}_{\mathtt{s}}(\mathbf{X}_{\mathtt{s}}^w)') \\
&\quad + \cdots + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{N_{v-1}+1:N_v,N_{v-1}+1:N_v}\mathbb{E}(\mathbf{X}_{\mathtt{s}}^v(\mathbf{X}_{\mathtt{s}}^w)').
\end{aligned}
\tag{43}
$$

Then for $\mathtt{t} > \mathtt{s}$ we obtain

$$
\begin{aligned}
\mathbb{E}(y_{\mathtt{t}i} y_{\mathtt{s}i}) &= \mathbb{E}\left((\Phi_i + \mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}} + \varepsilon_{\mathtt{t}i})(\Phi_i + \mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}} + \varepsilon_{\mathtt{s}i})\right) \\
&= \Phi_i^2 + 2\Phi_i\mathbf{\Psi}_i'\mathbb{E}(X_{\mathtt{t}}) + \mathbf{\Psi}_i'\mathbb{E}(\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')\mathbf{\Psi}_i \\
&= \Phi_i^2 + 2\Phi_i\mathbf{\Psi}_i'\mathbb{E}(X_{\mathtt{t}}) + \mathbf{\Psi}_i'\exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{2:1+d,1}\mathbb{E}(\mathbf{X}_{\mathtt{t}}')\mathbf{\Psi}_i + \mathbf{\Psi}_i'\exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{2:1+d,2:1+d}\mathbb{E}(\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{t}}')\mathbf{\Psi}_i \\
&= \Phi_i^2 + 2\Phi_i\mathbf{\Psi}_i'\mathbb{E}(\tilde{\mathbf{X}}_{\mathtt{t},1:d}) + \mathbf{\Psi}_i'\exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{2:1+d,1}\mathbb{E}(\tilde{X}_{\mathtt{t},1:d}')\mathbf{\Psi}_i \\
&\quad + \mathbf{\Psi}_i'\exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{2:1+d,2:1+d}\mathbb{E}(vech^{-1}(\tilde{\mathbf{X}}_{\mathtt{t},1+d:d+d_2}))\mathbf{\Psi}_i, \quad \text{and} \\
\mathbb{E}(y_{\mathtt{t}i}^2 y_{\mathtt{s}i}^2) &= \mathbb{E}\left((\Phi_i + \mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}} + \varepsilon_{\mathtt{t}i})^2(\Phi_i + \mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}} + \varepsilon_{\mathtt{s}i})^2\right) \\
&= \mathbb{E}\left((\Phi_i^2 + (\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2 + \varepsilon_{ti}^2 + 2\Phi_i\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}} + 2\Phi_i\varepsilon_{\mathtt{t}i} + 2\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}}\varepsilon_{\mathtt{t}i}\right) \\
&\quad \left(\Phi_i^2 + (\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})^2 + \varepsilon_{si}^2 + 2\Phi_i\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}} + 2\Phi_i\varepsilon_{\mathtt{s}i} + 2\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}}\varepsilon_{\mathtt{s}i}\right)\right) \\
&= \Phi_i^4 + 2\Phi_i^2\mathbb{E}((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})^2) + 2\Phi_i^2\sigma_i^2 + 4\Phi_i^3\mathbf{\Psi}_i'\mathbb{E}(\mathbf{X}_t) + \mathbb{E}((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2(\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})^2) \\
&\quad + 2\sigma_i^2\mathbb{E}((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2) + 2\Phi_i\mathbb{E}\left((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}} + \mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}}(\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})^2\right) + 4\Phi_i\sigma_i^2\mathbf{\Psi}_i'\mathbb{E}(\mathbf{X}_t) + 4\Phi_i^2\mathbb{E}\left((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})(\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})\right) \\
&= \Phi_i^4 + 2(\Phi_i^2 + \sigma_i^2)\mathbb{E}\left((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2\right) + 2\Phi_i^2\sigma_i^2 + \mathbb{E}\left((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2(\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})^2\right) \\
&\quad + 2\Phi_i\mathbb{E}\left((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})^2\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}} + \mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}}(\mathbf{\Psi}_i'X_{\mathtt{s}})^2\right) + 4(\Phi_i^2 + \sigma_i^2)\Phi_i\mathbf{\Psi}_i'\mathbb{E}(X_{\mathtt{t}}) + 4\Phi_i^2\mathbb{E}\left((\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{t}})(\mathbf{\Psi}_i'\mathbf{X}_{\mathtt{s}})\right).
\end{aligned}
\tag{44, 45}
$$

To complete the calculation of these moments, the quantities $\mathbb{E}\left((\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})^2\right)$, $\mathbb{E}\left((\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})\right)$, $\mathbb{E}\left((\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})^2\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}}\right)$, $\mathbb{E}\left(\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}}(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})^2\right)$ and $\mathbb{E}\left((\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})^2(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})^2\right)$ have to be derived. To simplify the notation, we omit the index i in $\mathbf{\Psi}_i$ in the following expressions; $\Psi_l \in \mathbb{R}$ is the element l of the $\mathbf{\Psi} \in \mathbb{R}^d$ (when

---

[21]A step width $\Delta = 1$ was already assumed in the main text. To derive the following moments with a different step-width if necessary, Δ will be included in the following expressions.

the index i is still included this would be Ψ_{il}). Note that

$$
\begin{aligned}
\mathbb{E}\left((\boldsymbol{\Psi}'\mathbf{X_t})^2\right) &= \boldsymbol{\Psi}'\mathbb{E}(\mathbf{X_t}\mathbf{X_t}')\boldsymbol{\Psi} = \boldsymbol{\Psi}'\mathbb{E}(vech^{-1}(\mathbf{X_t^2}))\boldsymbol{\Psi} \text{ and} \\
\mathbb{E}\left((\boldsymbol{\Psi}'\mathbf{X_t})(\boldsymbol{\Psi}'\mathbf{X_s})\right) &= \boldsymbol{\Psi}'\mathbb{E}(\mathbf{X_t}\mathbf{X_s}')\boldsymbol{\Psi} = \boldsymbol{\Psi}'\left(\mathbb{E}\left(\mathbb{E}(\mathbf{X_t}|\mathbf{X_s})\mathbf{X_s}'\right)\right)\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}'\left[\exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{2:1+d,1}\mathbb{E}(\mathbf{X_t}') + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{2:1+d,2:1+d}\mathbb{E}(\mathbf{X_t}\mathbf{X_t}')\right]\boldsymbol{\Psi} \\
&= \boldsymbol{\Psi}'\left[\exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{2:1+d,1}\mathbb{E}(\mathbf{X_t}') + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{2:1+d,2:1+d}\mathbb{E}(vech^{-1}(\mathbf{X_t^2}))\right]\boldsymbol{\Psi}.
\end{aligned}
\tag{46}
$$

In addition, we obtain

$$
\mathbb{E}\left((\boldsymbol{\Psi}'\mathbf{X_t})^2\boldsymbol{\Psi}'\mathbf{X_s}\right) = \boldsymbol{\Psi}'\mathbb{E}(\mathbf{X_t}\boldsymbol{\Psi}'\mathbf{X_t}\mathbf{X_s}')\boldsymbol{\Psi} = \boldsymbol{\Psi}'\mathbb{E}\left((\mathbf{X_t}\mathbf{X_s}')(\boldsymbol{\Psi}'\mathbf{X_t})\right)\boldsymbol{\Psi},
\tag{47}
$$

for $\mathtt{t} > \mathtt{s}$. Here we observe the following equality:

$$
(\mathbf{X_t}\mathbf{X_s}')(\boldsymbol{\Psi}'\mathbf{X_t}) = \begin{bmatrix} X_{\mathtt{t}1}X_{\mathtt{s}1} & X_{\mathtt{t}1}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}1}X_{\mathtt{s}d} \\ X_{\mathtt{t}2}X_{\mathtt{s}1} & X_{\mathtt{t}2}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}2}X_{\mathtt{s}d} \\ \cdots & \cdots & \cdots & \cdots \\ X_{\mathtt{t}d}X_{\mathtt{s}1} & X_{\mathtt{t}d}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}d}X_{\mathtt{s}d} \end{bmatrix} \sum_{l=1}^{d} \Psi_l X_{\mathtt{t}l}.
\tag{48}
$$

Equation (47) requires the derivation of $\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{t}l})$ where $i,j,l \in \{1,\ldots,d\}$. To simplify the notation, the following functions $g_i(\cdot)$, $i = 2,3,4$, are introduced to facilitate tracking specific elements of the moments vectors. We obtain

$$
g_2(i,j) = (i-1)\left(d - \frac{i}{2}\right) + j,
\tag{49}
$$

for $i,j \in \mathbb{N}$ and $i \leq j \leq d$. Moreover we derive

$$
g_3(i,j,m) = \sum_{k=1}^{i-1} d_{d-k} + \frac{j-i}{2}(2d-i-j+3) + m - j + 1
\tag{50}
$$

$$
g_4(i,j,m,n) = \sum_{k=1}^{i-1} d_{d-k} + \sum_{k=1}^{j-i} d_{(d-i+1-k,d-1)} + \left(d+1 - \frac{m+j-1}{2}\right)(m-j) + n - m + 1,
\tag{51}
$$

for $i,j,m \in \mathbb{N}$ and $i \leq j \leq m \leq d$ and for $i,j,m \in \mathbb{N}$ and $i \leq j \leq m \leq n \leq d$, respectively. While d was

the dimension of the process $(\mathbf{X}(t))$, $d_{(\cdot,\cdot)}$ is the function already defined in (33). For $d = 3$ this yields

$$
g_2(i,j) = \begin{cases}
1, & \text{if } i = 1, \ j = 1 \\
2, & \text{if } i = 1, \ j = 2 \\
3, & \text{if } i = 1, \ j = 3 \\
4, & \text{if } i = 2, \ j = 2 \\
5, & \text{if } i = 2, \ j = 3 \\
6, & \text{if } i = 3, \ j = 3 \ ,
\end{cases}
\tag{52}
$$

$$
g_3(i,j,m) = \begin{cases}
1, & \text{if } i = 1, \ j = 1, \ m = 1 \\
2, & \text{if } i = 1, \ j = 1, \ m = 2 \\
3, & \text{if } i = 1, \ j = 1, \ m = 3 \\
4, & \text{if } i = 1, \ j = 2, \ m = 2 \\
5, & \text{if } i = 1, \ j = 2, \ m = 3 \\
6, & \text{if } i = 1, \ j = 3, \ m = 3 \\
7, & \text{if } i = 2, \ j = 2, \ m = 2 \\
8, & \text{if } i = 2, \ j = 2, \ m = 3 \\
9, & \text{if } i = 2, \ j = 3, \ m = 3 \\
10, & \text{if } i = 3, \ j = 3, \ m = 3 \ ,
\end{cases}
\tag{53}
$$

and

$$
g_4(i,j,m,n) = \begin{cases}
1, & \text{if } i=1,\ j=1,\ m=1,\ n=1 \\
2, & \text{if } i=1,\ j=1,\ m=1,\ n=2 \\
3, & \text{if } i=1,\ j=1,\ m=1,\ n=3 \\
4, & \text{if } i=1,\ j=1,\ m=2,\ n=2 \\
5, & \text{if } i=1,\ j=1,\ m=2,\ n=3 \\
6, & \text{if } i=1,\ j=1,\ m=3,\ n=3 \\
7, & \text{if } i=1,\ j=2,\ m=2,\ n=2 \\
8, & \text{if } i=1,\ j=2,\ m=2,\ n=3 \\
9, & \text{if } i=1,\ j=2,\ m=3,\ n=3 \\
10, & \text{if } i=1,\ j=3,\ m=3,\ n=3 \\
11, & \text{if } i=2,\ j=2,\ m=2,\ n=2 \\
12, & \text{if } i=2,\ j=2,\ m=2,\ n=3 \\
13, & \text{if } i=2,\ j=2,\ m=3,\ n=3 \\
14, & \text{if } i=2,\ j=3,\ m=3,\ n=3 \\
15, & \text{if } i=3,\ j=3,\ m=3,\ n=3.
\end{cases}
\tag{54}
$$

Let $\mathbf{e}$ be a vector of ones, $\mathbf{e} = (1, \ldots, 1)'$,[22] and $\tilde{\mathbf{e}} = (1, 2, \ldots, d)'$. Then $\mathbf{M}$, $\mathbf{M}^j$ and $\mathbf{M}^{j,l}$ are the following $d_2 \times 2$, $d_2 \times 3$ and $d_2 \times 4$ matrices:

$$
\mathbf{M} = \begin{pmatrix}
\mathbf{e} & \tilde{\mathbf{e}} \\
2\mathbf{e} & \tilde{\mathbf{e}}_{2:d} \\
\cdots & \cdots \\
i\mathbf{e} & \tilde{\mathbf{e}}_{i:d} \\
\cdots & \cdots \\
d & d
\end{pmatrix},
\quad \text{which for } d = 3 \text{ is} \quad
\mathbf{M} = \begin{pmatrix}
1 & 1 \\
1 & 2 \\
1 & 3 \\
2 & 2 \\
2 & 3 \\
3 & 3
\end{pmatrix},
\tag{55}
$$

---

[22]It's dimension is not specified on purpose as it will vary and will be clear from the context.

$$\mathbf{M}^j = (\mathbf{M}, j\mathbf{e}) \quad \text{and} \quad \mathbf{M}^{j,l} = (\mathbf{M}^j, l\mathbf{e}) = (\mathbf{M}, j\mathbf{e}, l\mathbf{e}) \;, \tag{56}$$

where $\mathbf{e}$ is here a vector of ones of the dimension $d_2 \times 1$. Thus, for $\mathtt{t} > \mathtt{s}$

$$
\begin{aligned}
\mathbb{E}\left(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{t}l}\right) &= \mathbb{E}\left(\mathbb{E}\left(X_{\mathtt{t}i}X_{\mathtt{t}l}|\mathbf{X}_{\mathtt{s}}\right)X_{\mathtt{s}j}\right) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{k,1}\mathbb{E}(X_{\mathtt{s}j}) + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{k,2:1+d}\mathbb{E}(\mathbf{X}_{\mathtt{s}}X_{\mathtt{s}j}) \\
&\quad + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{k,2+d:2+d+d_2}\mathbb{E}(\mathbf{X}_{\mathtt{s}}^2 X_{\mathtt{s}j}) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{k,1}\mathbb{E}(X_{\mathtt{t}j}) + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{k,2:1+d}\mathbb{E}\left(\mathbf{X}_{g_2([\tilde{\mathbf{e}},j\mathbf{e}]),\mathtt{t}}^2\right) \\
&\quad + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{k,2+d:2+d+d_2}\mathbb{E}\left(\mathbf{X}_{g_3(\mathbf{M}^j),\mathtt{t}}^3\right) \;, \tag{57}
\end{aligned}
$$

where $k = 1 + d + g_2(i,l)$. Thus, the (i,j) element, $i,j = 1,\dots,d$, of matrix in (47) is

$$\left[\mathbb{E}\left((\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}}\right)\right]_{ij} = \sum_{l=1}^d \Psi_l\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{t}l}) \;,$$

where $\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{t}l})$ is given by (57). Let $\mathtt{t} > \mathtt{s}$, then

$$\mathbb{E}\left(\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}}(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})^2\right) = \mathbf{\Psi}'\mathbb{E}(\mathbf{X}_t\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}}\mathbf{X}_{\mathtt{s}}')\mathbf{\Psi} = \mathbf{\Psi}'\mathbb{E}\left((\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})\right)\mathbf{\Psi} \;, \tag{58}$$

where

$$
\begin{aligned}
(\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}}) &= \begin{bmatrix} X_{\mathtt{t}1}X_{\mathtt{s}1} & X_{\mathtt{t}1}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}1}X_{\mathtt{s}d} \\ X_{\mathtt{t}2}X_{\mathtt{s}1} & X_{\mathtt{t}2}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}2}X_{\mathtt{s}d} \\ \cdots & \cdots & \cdots & \cdots \\ X_{\mathtt{t}d}X_{\mathtt{s}1} & X_{\mathtt{t}d}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}d}X_{\mathtt{s}d} \end{bmatrix} \sum_{i=l}^d \Psi_l X_{\mathtt{s}l} \\
&= \sum_{i=l}^d \Psi_l \begin{bmatrix} X_{\mathtt{t}1}X_{\mathtt{s}1}X_{\mathtt{s}l} & X_{\mathtt{t}1}X_{\mathtt{s}2}X_{\mathtt{s}l} & \cdots & X_{\mathtt{t}1}X_{\mathtt{s}d}X_{\mathtt{s}l} \\ X_{\mathtt{t}2}X_{\mathtt{s}1}X_{\mathtt{s}l} & X_{\mathtt{t}2}X_{\mathtt{s}2}X_{\mathtt{s}l} & \cdots & X_{\mathtt{t}2}X_{\mathtt{s}d}X_{\mathtt{s}l} \\ \cdots & \cdots & \cdots & \cdots \\ X_{\mathtt{t}d}X_{\mathtt{s}1}X_{\mathtt{s}l} & X_{\mathtt{t}d}X_{\mathtt{s}2}X_{\mathtt{s}l} & \cdots & X_{\mathtt{t}d}X_{\mathtt{s}d}X_{\mathtt{s}l} \end{bmatrix} \;. \tag{59}
\end{aligned}
$$

44

For expression (59) one needs to know $\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{s}l})$ where $i, j, l \in \{1, \ldots, d\}$. Thus, for $\mathtt{t} > \mathtt{s}$

$$
\begin{aligned}
\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{s}l}) &= \mathbb{E}\left(\mathbb{E}(X_{\mathtt{t}i}|X_{\mathtt{s}})X_{\mathtt{s}j}X_{\mathtt{s}l}\right) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{i+1,1}\mathbb{E}(X_{\mathtt{s}j}X_{\mathtt{s}l}) + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{i+1,2:1+d}\mathbb{E}(X_{\mathtt{s}}X_{\mathtt{s}j}X_{\mathtt{s}l}) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{i+1,1}\mathbb{E}\left(\mathbf{X}^2_{g_2(j,l),\mathtt{t}}\right) + \exp((\mathtt{t}-\mathtt{s})\Delta\mathbf{A})_{i+1,2:1+d}\mathbb{E}\left(\mathbf{X}^3_{g_3([\bar{e},je,le]),\mathtt{t}}\right).
\end{aligned}
\tag{60}
$$

The (i,j) element, $i,j = 1, \ldots, d$, of matrix in (59) is

$$
\left[\mathbb{E}\left((\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}}\right)\right]_{ij} = \sum_{l=1}^{d}\Psi_l\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{s}l}),
$$

where $\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{s}l})$ is given by (61). Finally

$$
\mathbb{E}\left((\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})^2(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})^2\right) = \mathbf{\Psi}'\mathbb{E}\left(\mathbf{X}_{\mathtt{t}}(\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})\mathbf{X}_{\mathtt{s}}'\right)\mathbf{\Psi} = \mathbf{\Psi}'\mathbb{E}\left((\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')(\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}})\right)\mathbf{\Psi}
\tag{61}
$$

and

$$
\begin{aligned}
(\mathbf{X}_{\mathtt{t}}\mathbf{X}_{\mathtt{s}}')(\mathbf{\Psi}'\mathbf{X}_{\mathtt{t}})(\mathbf{\Psi}'\mathbf{X}_{\mathtt{s}}) &= \begin{bmatrix} X_{\mathtt{t}1}X_{\mathtt{s}1} & X_{\mathtt{t}1}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}1}X_{\mathtt{s}d} \\ X_{\mathtt{t}2}X_{\mathtt{s}1} & X_{\mathtt{t}2}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}2}X_{\mathtt{s}d} \\ \cdots & \cdots & \cdots & \cdots \\ X_{\mathtt{t}d}X_{\mathtt{s}1} & X_{\mathtt{t}d}X_{\mathtt{s}2} & \cdots & X_{\mathtt{t}d}X_{\mathtt{s}d} \end{bmatrix}\sum_{i=1}^{d}\sum_{j=1}^{d}\Psi_i\Psi_j X_{\mathtt{t}i}X_{\mathtt{s}j} \\
&= \sum_{i=1}^{d}\sum_{j=1}^{d}\Psi_i\Psi_j \begin{bmatrix} X_{\mathtt{t}1}X_{\mathtt{s}1}X_{\mathtt{t}i}X_{\mathtt{s}j} & X_{\mathtt{t}1}X_{\mathtt{s}2}X_{\mathtt{t}i}X_{\mathtt{s}j} & \cdots & X_{\mathtt{t}1}X_{\mathtt{s}d}X_{\mathtt{t}i}X_{\mathtt{s}j} \\ X_{\mathtt{t}2}X_{\mathtt{s}1}X_{\mathtt{t}i}X_{\mathtt{s}j} & X_{\mathtt{t}2}X_{\mathtt{s}2}X_{\mathtt{t}i}X_{\mathtt{s}j} & \cdots & X_{\mathtt{t}2}X_{\mathtt{s}d}X_{\mathtt{t}i}X_{\mathtt{s}j} \\ \cdots & \cdots & \cdots & \cdots \\ X_{\mathtt{t}d}X_{\mathtt{s}1}X_{\mathtt{t}i}X_{\mathtt{s}j} & X_{\mathtt{t}d}X_{\mathtt{s}2}X_{\mathtt{t}i}X_{\mathtt{s}j} & \cdots & X_{\mathtt{t}d}X_{\mathtt{s}d}X_{\mathtt{t}i}X_{\mathtt{s}j} \end{bmatrix}.
\end{aligned}
\tag{62}
$$

Then for $\mathtt{t} > \mathtt{s}$ we have

$$
\begin{aligned}
\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{t}m}X_{\mathtt{s}n}) &= \mathbb{E}\left(\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{t}m}|\mathbf{X}_\mathtt{s})X_{\mathtt{s}j}X_{\mathtt{s}n}\right) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{k,1}\mathbb{E}\left(X_{\mathtt{s}j}X_{\mathtt{s}n}\right) + \exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{k,2:1+d}\mathbb{E}\left(\mathbf{X}_\mathtt{s}X_{\mathtt{s}j}X_{\mathtt{s}n}\right) \\
&\quad + \exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{k,2+d:1+d+d_2}\mathbb{E}\left(X_\mathtt{s}^2 X_{\mathtt{s}j}X_{\mathtt{s}n}\right) \\
&= \exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{k,1}\mathbb{E}\left(\mathbf{X}_{g_2(j,n),\mathtt{t}}^2\right) + \exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{k,2:1+d}\mathbb{E}\left(\mathbf{X}_{g_3(\tilde{e},je,ne),\mathtt{t}}^3\right) \\
&\quad + \exp((\mathtt{t}-\mathtt{s})\Delta\cdot\mathbf{A})_{k,2+d:1+d+d_2}\mathbb{E}\left(\mathbf{X}_{g_4(\mathbf{M}^{j,n}),\mathtt{t}}^4\right) , \quad (63)
\end{aligned}
$$

where $k = 1 + d + g_2(i,m)$. The (i,j) element, $i,j = 1,\ldots,d$, of the matrix in (61) is

$$
\left[\mathbb{E}\left((\mathbf{X}_\mathtt{t}\mathbf{X}_\mathtt{s}')(\mathbf{\Psi}'\mathbf{X}_\mathtt{t})(\mathbf{\Psi}'\mathbf{X}_\mathtt{s})\right)\right]_{ij} = \sum_{k=1}^{d}\sum_{l=1}^{d}\Psi_k\Psi_l\mathbb{E}(X_{\mathtt{t}i}X_{\mathtt{s}j}X_{\mathtt{t}k}X_{\mathtt{s}l}) ,
$$

with expectations being given by (63).

D   Solving for $\Phi(t,\mathbf{u})$ and $\mathbf{\Psi}(t,\mathbf{u})$

This section derives the functions $\Phi(t,\mathbf{u})$ and $\mathbf{\Psi}(t,\mathbf{u})$ of the Riccati differential equations described by (29) for an $\mathbb{A}_m(d)$ model with diagonal $\boldsymbol{\beta}_{II}$. By equation (6), which is based on Filipović (2009)[Theorem 10.4 and Corollary 10.2], $\mathbf{\Psi}(t,\mathbf{u})$ and $\Phi(t,\mathbf{u})$ evaluated at $t = \tau_i$, $i = 1,\ldots,M$ and $\mathbf{u} = \mathbf{0}_d$ are necessary to compute the zero coupon prices $\pi^0(t,\tau_i)$ and corresponding model yields. For the Vasicek (1977) and the Cox et al. (1985) model the solutions are presented e.g. in Filipović (2009)[p. 162-163].

Now we apply the results obtained in Grasselli and Tebaldi (2008)[Section 3.4.1] for $\mathbb{A}_m(d)$ models with diagonal $m \times m$ matrix $\boldsymbol{\beta}_{II}$. In the first step we have to solve the linear ODE for the J components. I.e. we consider[23]

$$
\begin{aligned}
\partial_t\mathbf{\Psi}_J(t,\mathbf{u}) &= \left(\beta_{JJ}^Q\right)'\mathbf{\Psi}_J(t,\mathbf{u}) - \boldsymbol{\gamma}_{xJ}; \\
\mathbf{\Psi}_J(0,\mathbf{u}) &= \mathbf{u}_J , \quad \boldsymbol{\gamma}_{xJ} = \mathbf{e}_{n\times 1}. \quad (64)
\end{aligned}
$$

---

[23]Note that the dimension of Ψ_J is n.

A particular solution of (64) is of the structure

$$\mathbf{\Psi}_J(t, \mathbf{u}) = \exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)\mathbf{c}_1 + \mathbf{c}_2 , \tag{65}$$

with

$$\mathbf{\Psi}_J(0, \mathbf{u}) = \mathbf{c}_1 + \mathbf{c}_2 = \mathbf{u}_J. \tag{66}$$

Then (65) implies

$$\partial_t \mathbf{\Psi}_J(t, \mathbf{u}) = \left(\boldsymbol{\beta}_{JJ}^Q\right)' \exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)\mathbf{c}_1. \tag{67}$$

Plugging (65) and (67) into (64) yields

$$\left(\boldsymbol{\beta}_{JJ}^Q\right)' \exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)\mathbf{c}_1 = \left(\boldsymbol{\beta}_{JJ}^Q\right)' \left(\exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)\mathbf{c}_1 + \mathbf{c}_2\right) - \boldsymbol{\gamma}_{xJ} ,$$

which gives $\boldsymbol{\gamma}_{xJ} = \left(\boldsymbol{\beta}_{JJ}^Q\right)' \mathbf{c}_2$ and thus $\mathbf{c}_2 = \left(\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)^{-1} \boldsymbol{\gamma}_{xJ}$. This and (66) imply that $\mathbf{c}_1 = \mathbf{u}_J - \left(\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)^{-1} \boldsymbol{\gamma}_{xJ}$. Plugging the last expression and $\mathbf{c}_2$ into (65) gives[24]

$$\begin{aligned}\mathbf{\Psi}_J(t, \mathbf{u}) &= \exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)\mathbf{u}_J - \left(\exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right) - \mathbf{I}_n\right)\left(\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)^{-1}\boldsymbol{\gamma}_{xJ} \\ &= \exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)\mathbf{u}_J - \left(\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right)^{-1}\left(\exp\left(t\left(\boldsymbol{\beta}_{JJ}^Q\right)'\right) - \mathbf{I}_n\right)\boldsymbol{\gamma}_{xJ}. \end{aligned} \tag{68}$$

In a second step the solution of the subsystem $\mathbf{\Psi}_J(t, \mathbf{u})$ is plugged in into the ODEs for the square root

---

[24]Equation (68) also follows from Perko (1991)[Theorem 1, p. 60]. The matrix product in the last expression of (68) can be exchanged by the properties of the matrix exponential. I.e. $(\boldsymbol{\beta}_{JJ}^{Q'})^{-1}\exp(t\boldsymbol{\beta}_{JJ}^{Q'}) = \exp(t\boldsymbol{\beta}_{JJ}^{Q'})(\boldsymbol{\beta}_{JJ}^{Q'})^{-1}$ follows from $\exp(\mathbf{Y}\mathbf{X}\mathbf{Y}^{-1}) = \mathbf{Y}\exp(\mathbf{X})\mathbf{Y}^{-1}$.

terms $\boldsymbol{\Psi}_I$. Thus, the Riccati equations for the I components are

$$
\begin{aligned}
\partial_t \Psi_i(t, \mathbf{u}) &= \frac{1}{2} \Sigma_i^2 \Psi_i^2(t, \mathbf{u}) + \beta_{ii}^Q \Psi_i(t, \mathbf{u}) - \tilde{\gamma}_{xi} \\
&\tilde{\gamma}_{xi}(t, \mathbf{u}) = \gamma_{xi} - \sum_{j=1}^n \beta_{m+j,i}^Q (\boldsymbol{\Psi}_J(t, \mathbf{u}))_j - \frac{1}{2} \sum_{j=1}^n \Sigma_{m+j}^2 \mathcal{B}_{i,m+j}^x \left[\boldsymbol{\Psi}_J(t, \mathbf{u})\right]_j^2 , \\
\Psi_i(0, \mathbf{u}) &= u_i , \quad \gamma_{xi} = 1 , i = 1, \ldots, m.
\end{aligned}
\tag{69}
$$

As (69) is a time inhomogeneous Riccati equation, it can be solved in the following way: The ODE of interest is $\partial_t \Psi_i = \frac{1}{2} \Sigma_i^2 \Psi_i^2 + \beta_{ii}^Q \Psi_i - \tilde{\gamma}_{xi}$ for $i = 1, \ldots, m$. After the substitution $\nu_i = \Sigma_i^2 \Psi_i$, $i = 1, \ldots, m$, we get $\partial_t \nu_i = \frac{1}{2} \nu_i^2 + \beta_{ii}^Q \nu_i - \Sigma_i^2 \tilde{\gamma}_{xi}$. A solution for an inhomogenous Riccati ODE of this structure is provided in Grasselli and Tebaldi (2008)[Section 3.4.1]. The solution for ν_i is

$$
\begin{aligned}
\nu_i(t, \mathbf{u}) &= \frac{M_1^{(i)}(t, \mathbf{u}) \, u_i + M_2^{(i)}(t, \mathbf{u})}{M_3^{(i)}(t, \mathbf{u}) \, u_i + M_4^{(i)}(t, \mathbf{u})} , \quad \text{where} \\
\mathbf{M}^{(i)}(t, \mathbf{u}) &= \begin{pmatrix} M_1^{(i)}(t, \mathbf{u}) & M_2^{(i)}(t, \mathbf{u}) \\ M_3^{(i)}(t, \mathbf{u}) & M_4^{(i)}(t, \mathbf{u}) \end{pmatrix} = \exp \begin{pmatrix} t\beta_{ii}^Q & -\Sigma_i^2 \int_0^t \tilde{\gamma}_{xi}(s, \mathbf{u}) ds \\ -t/2 & 0 \end{pmatrix} .
\end{aligned}
\tag{70}
$$

At the end $\Psi_i = \frac{\nu_i}{\Sigma_i^2}$ for $i = 1, \ldots, m$. With $\mathbf{u} = \mathbf{0}_{d \times 1}$ we get

$$
\Psi_i(t, \mathbf{0}) = \frac{1}{\Sigma_i^2} \frac{M_2^{(i)}(t, \mathbf{0})}{M_4^{(i)}(t, \mathbf{0})} , \quad \text{for } i = 1, \ldots, m.
\tag{71}
$$

To derive $\mathbf{M}^{(i)}(t, \mathbf{u})$ the integral $\int_0^t \tilde{\gamma}_{xi}(s, \mathbf{u}) ds$ has to be solved, where

$$
\int_0^t \tilde{\gamma}_{xi}(s, \mathbf{u}) ds = \gamma_{xi} t - \sum_{j=1}^n \beta_{m+j,i}^Q \int_0^t \left[\boldsymbol{\Psi}_J(s, u)\right]_j ds - \frac{1}{2} \sum_{j=1}^n \Sigma_{m+j}^2 \mathcal{B}_{i,m+j}^x \int_0^t \left[\boldsymbol{\Psi}_J^2(t, \mathbf{u})\right]_j ds.
\tag{72}
$$

The second term in (72) can be derived by means of

$$
\int_0^t \boldsymbol{\Psi}_J(s,0)ds = -\int_0^t \left[\left(\left(\boldsymbol{\beta}_{JJ}^Q \right)' \right)^{-1} \left(\exp\left(s \left(\boldsymbol{\beta}_{JJ}^Q \right)' \right) - \mathbf{I}_n \right) \boldsymbol{\gamma}_{xJ} \right] ds
$$

$$
= -\left(\left(\boldsymbol{\beta}_{JJ}^Q \right)' \right)^{-1} \left[\left(\left(\boldsymbol{\beta}_{JJ}^Q \right)' \right)^{-1} \left(\exp\left(t \left(\boldsymbol{\beta}_{JJ}^Q \right)' \right) - \mathbf{I}_n \right) - t\mathbf{I}_n \right] \boldsymbol{\gamma}_{xJ} \qquad (73)
$$

using (68). The third term in (72) can be derived numerically as well as the whole expression (72). It remains to calculate $\Phi(t,\mathbf{0})$, where by (29)

$$
\partial_t \Phi(t,\mathbf{u}) = \frac{1}{2} \boldsymbol{\Psi}(t,\mathbf{u})' \mathbf{a} \boldsymbol{\Psi}(t,\mathbf{u}) + \left(\mathbf{b}^Q \right)' \boldsymbol{\Psi}(t,\mathbf{u}) , \qquad \Phi(0,\mathbf{u}) = 0 ,
$$

$$
= \frac{1}{2} diag(\boldsymbol{\Sigma}_J^2) \boldsymbol{\Psi}_J(t,\mathbf{u})^2 + \left(\mathbf{b}^Q \right)' \boldsymbol{\Psi}(t,\mathbf{u}).
$$

We can express $\Phi(t,\mathbf{0})$ by means of $\Phi(t,\mathbf{0}) = \Phi_I(t,\mathbf{0}) + \Phi_J(t,\mathbf{0})$. The J components of the $d \times d$ matrix $\mathbf{a}$ are equal to a $n \times n$ diagonal matrix having $\Sigma_{m+1}^2, \ldots, \Sigma_d^2$ along the main diagonal such that

$$
\Phi_J(t,\mathbf{0}) = \frac{1}{2} \int_0^t \boldsymbol{\Psi}_J(s,\mathbf{0})' \begin{pmatrix} \Sigma_{m+1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_d^2 \end{pmatrix} \boldsymbol{\Psi}_J(s,\mathbf{0}) \, ds + \int_0^t (b_{m+1}^Q, \ldots, b_d^Q) \boldsymbol{\Psi}_J(s,\mathbf{0}) \, ds.
$$

For $\Phi_I(t,\mathbf{0})$ we obtain

$$
\Phi_I(t,\mathbf{0}) = \left(\mathbf{b}_I^Q \right)' \int_0^t \boldsymbol{\Psi}_I(s,\mathbf{0}) \, ds. \qquad (74)
$$

$\Phi_I(t,\mathbf{0})$, $\Phi_J(t,\mathbf{0})$ and $\boldsymbol{\Psi}(t,\mathbf{0})$ can be easily obtained by means of numerical integration. To do this we generate a grid $\Gamma = \{t_0, t_1, \ldots, t_G\}$ with $G+1$ grid points. We set $t_0 = 0$ and $t_G = \max(\tau_l) = \tau_M$. By including the maturities τ_l, $l = 1, \ldots, M$, in Γ we know that for each maturity we have $t_{k_l} = \tau_l$ for some $k_l \in \{1, \ldots, G+1\}$.[25] The step-widths are given by $\Delta_k = t_k - t_{k-1}$, $k = 1, \ldots, G+1$ (if some elements of Γ coincide with τ_l this does not cause any problems since $\Delta_k = 0$ for

---

[25]This has been implemented as follows: (i) generate an equally spaced grid, (ii) include the M maturities, (iii) sort all these points in ascending order.

such grid-points). Then we evaluate $\Phi_J(t,\mathbf{0})$ at each $t = t_k$, $k = 1,\ldots,G+1$. By calculating

the sums $\frac{1}{2}\sum_{k=1}^{k_l-1}\boldsymbol{\Psi}_J(t_k,\mathbf{0})'\begin{pmatrix} \Sigma_{m+1}^2 & 0 \\ 0 & \Sigma_d^2 \end{pmatrix}\boldsymbol{\Psi}_J(t_k,\mathbf{0})\Delta_k + \sum_{k=1}^{k_l-1}(b_{m+1}^Q,\ldots,b_d^Q)\boldsymbol{\Psi}_J(t_k,\mathbf{0})\Delta_k$ (left

Riemann sums), $\frac{1}{2}\sum_{k=2}^{k_l}\boldsymbol{\Psi}_J(t_k,\mathbf{0})'\begin{pmatrix} \Sigma_{m+1}^2 & 0 \\ 0 & \Sigma_d^2 \end{pmatrix}\boldsymbol{\Psi}_J(t_k,0)\Delta_k + \sum_{k=2}^{k_l}(b_{m+1}^Q,\ldots,b_d^Q)\boldsymbol{\Psi}_J(t_k,\mathbf{0})\Delta_k$

(right Riemann sums), or $\frac{1}{2}\sum_{k=2}^{k_l}\frac{\boldsymbol{\Psi}_J(t_{k-1},\mathbf{0})'+\boldsymbol{\Psi}_J(t_k,\mathbf{0})'}{2}\begin{pmatrix} \Sigma_{m+1}^2 & 0 \\ 0 & \Sigma_d^2 \end{pmatrix}\frac{\boldsymbol{\Psi}_J(t_{k-1},\mathbf{0})'+\boldsymbol{\Psi}_J(t_k,\mathbf{0})'}{2}\Delta_k +$

$\sum_{k=2}^{k_l}(b_{m+1}^Q,\ldots,b_d^Q)\frac{\boldsymbol{\Psi}_J(t_{k-1},\mathbf{0})'+\boldsymbol{\Psi}_J(t_k,\mathbf{0})'}{2}\Delta_k$ (trapeze-rule) we get a numerical approximation of $\Phi_J(\tau_l,\mathbf{0})$,

$k_l = G(+1)$ for $\tau_l = \tau_M$. In our code right sums were implemented. Since integrals of $\boldsymbol{\Psi}_J$ and $\boldsymbol{\Psi}_J^2$

are necessary to obtain $\int_0^{\tau_l}\tilde{\gamma}_{xi}(t,\mathbf{0})dt$, we use numerical integration also to obtain $\int_0^{\tau_l}\tilde{\gamma}_{xi}(t,\mathbf{0})dt$. These

proxies are then used in (71) to calculate $\boldsymbol{\Psi}_i(\tau_l,\mathbf{0})$, $i = 1,\ldots,m$. Equipped with $\boldsymbol{\Psi}_I(t_k,\mathbf{0})$, $k = 1,\ldots,G+1$

we are also able to obtain a numerical approximation of $\Phi_I(\tau_l,\mathbf{0})$.

E Restrictions on the Parameters

First we present the conditions for admissibility which guarantee that $(\mathbf{X}(t))$ remains with in the state

space $\mathscr{S}$. All these restriction are applied in both measures, $\mathbb{P}$ and $\mathbb{Q}$, respectively.

Admissibility conditions (see Filipović, 2009, Theorem 10.2): $\mathbf{a}$, $\boldsymbol{\alpha}_i$ are symmetric and positive semidef-

inite. $\mathbf{a}_{II} = \mathbf{0}_{m\times m}$, $\mathbf{a}_{IJ} = \mathbf{a}'_{JI} = \mathbf{0}_{m\times n}$, $\boldsymbol{\alpha}_j = \mathbf{0}_{n\times n}$ for all $j = m+1,\ldots,m+n$. $\alpha_{i,kl} = \alpha_{i,lk} = 0$ for

$k \in I \setminus \{i\}$ for all $1 \le i,l \le d$, $\mathbf{b}^{\cdot} \in \mathscr{S}$, $\boldsymbol{\beta}_{IJ} = \mathbf{0}_{m\times n}$ and $\boldsymbol{\beta}_{II}$ has non-negative off-diagonal elements.

In a model with diagonal diffusion matrix the admissibility restrictions are met if the Dai and Singleton

(2000) conditions presented in Definition 1 are met. To keep the process $(\mathbf{X}(t))$ off the boundaries of the

state space $\mathscr{S}$ we can impose the *Boundary conditions/Feller conditions* (see Aït-Sahalia and Kimmel,

2010, Eq. 15-17): $b_i^{\cdot} \ge \frac{1}{2}\Sigma_i^2$ for $i = 1,\ldots,m$. (The conditions $\boldsymbol{\beta}_{IJ}^{\cdot} = \mathbf{0}_{m\times n}$ and $\boldsymbol{\beta}_{II}^{\cdot}$ having non-negative

off-diagonal elements are already included in the admissibility conditions.)

Last but not least, we have some further restrictions for stationarity:

Stationarity conditions (see Aït-Sahalia and Kimmel, 2010, Table 1): The real part of the eigenvalues of

$\boldsymbol{\beta}$ is smaller than zero. A more general treatment regarding stationarity is provided in Glasserman and

Kim (2010).

F *GMM*-Estimation

For our model it turned out that minimizing the *GMM* distance function (23) is non-trivial. By using a standard minimization routine, as the MATLAB minimization routine `fminsearch` based on the Nelder-Mead algorithm,[26] we observed that the estimation procedure preforms poorly.[27] Therefore, as being described in `Step 1` below, we include multistart random search methods in our minimization procedure (see, e.g., Törn and Zilinskas, 1989). Compared to working with the above minimization routine only, this procedure improves parameter estimation, especially when looking at the means and the absolute deviation from the mean in percentage terms. Some results are presented in Table 5.

In addition, we apply classical tests such as the *Wald* and the *distance difference* test (see, e.g., Ruud, 2000; Newey and McFadden, 1994). We observe that these tests do not perform well. Some results for tests of the null hypothesis $\theta^P = \theta^Q$ against the alternative $\theta^P \neq \theta^Q$ are presented in Table 6, where it can be seen that power and size of these tests do not fulfill "the usual quality standards". We explain this behavior by the problem of estimating a relatively large (23×23) covariance matrix and a matrix of gradients with the Wald test (see also equation (26)). Regarding the distance difference test, we observe that the $(\mathfrak{q} \times \mathfrak{q})$ weighting matrix $\mathbf{C}_T = \hat{\mathbf{\Lambda}}^{-1}$ has a strong impact on the results of the tests, which in turn introduces potential inaccuracies in case $\hat{\mathbf{\Lambda}}$ was not estimated accurately enough.

To further improve the properties of the estimation routine, we combine multistart random search methods with Quasi-Bayesian methods (see Chernozhukov and Hong, 2003). To apply Bayesian tools a prior $\tilde{\pi}(\boldsymbol{\vartheta})$ has to be specified. The parameter space Θ is a subset of $\mathbb{R}^{\mathfrak{p}}$. It is a proper subset, since some parameters are strictly positive, nonnegative, etc. by the model assumptions. In addition, admissibility and stationarity further restrict the parameter space. Hence, the prior $\tilde{\pi}(\boldsymbol{\vartheta}) = 0$ for all $\boldsymbol{\vartheta} \notin \Theta$. In addition, to implement a random search method on a computer and to add "prior information" we restrict Θ to $\Theta_0 \subset \Theta$, where $\tilde{\pi}(\boldsymbol{\vartheta}) = 0$ for all $\boldsymbol{\vartheta}$ not contained in Θ_0.

---

[26]See $\mathtt{http : //www.mathworks.de/de/help/matlab/ref/fminsearch.html}$
[27]Detailed results of these simulation experiments can be obtained from the authors on request.

The subset Θ_0 is constructed as follows: For Σ_i the lower bound is set to 0.1, while the upper bound is set to 2. The upper bound follows from variances of the yields observed, the lower bound from the assumption that the variance of each component is not too small. For the unrestricted $\mathcal{B}_{ij}^x$ we assume that $\mathcal{B}_{ij}^x \in [0, 2]$, where $\mathcal{B}_{ij}^x \geq 0$ follows from the models assumptions, while $\mathcal{B}_{ij}^x \leq 2$ is used to keep the impact of the square root term on the other volatilities bounded. In addition, $\sigma_\varepsilon^2 \in [0.005, 0.025]$. This is motivated by the argument that the observation error is small compared to the variance of the yields. The observation error can be due to market-microstructure noise (see, e.g., Campbell et al., 1997; Chen et al., 2007). The lower bound is based on the assumption that at least 10 basis points can be attributed to the noise. To ensure that the matrices $\boldsymbol{\beta}^Q$ and $\boldsymbol{\beta}^P$ are sufficiently far away from a singular matrix, we assume $\beta_{ii} \leq -0.1$. To cope with the high degree of serial correlation of the yields, we demand for $\beta_{ii} \geq -50$. For β_{ij}, $i \neq j$ we apply a lower bound of -10 and an upper bound of 10. The differences in the matrix exponential of $\boldsymbol{\beta}$ become small, when values outside these intervals are used.

Since θ^P and γ_0 determine the mean of the instantaneous spot rate $\mathbb{E}(r_{\mathbf{t}}) = \gamma_0 + \theta^P$ defined by a stationary $(\mathbf{X}(t))$ (see equation (2)), we assume that $\frac{1}{c}[\tilde{\mathbf{m}}_T(\mathbf{y}_{1:T})]_1 \leq \gamma_0 + \theta^P \leq c[\tilde{\mathbf{m}}_T(\mathbf{y}_{1:T})]_1$, where $c = 1.45$ is applied in the Bayesian sampler. Since the sample mean of the instantaneous short rate cannot be observed, we use the sample mean of the shortest maturity, which in terms of our notation is $[\tilde{\mathbf{m}}_T(\mathbf{y}_{1:T})]_1$.

In addition, the conditions on stationarity, identification and admissibility have to be met. Given these restrictions and the uniform prior on the components of Θ_0, the prior $\tilde{\pi}(\boldsymbol{\vartheta})$ is proportional to $\mathbb{I}_{(\text{Stationarity,Identification,Admissibility})}\mathbb{I}_{\left(\frac{1}{c}[\tilde{\mathbf{m}}_T(\mathbf{y}_{1:T})]_1 \leq \gamma_0 + \theta^P \leq c[\tilde{\mathbf{m}}_T(\mathbf{y}_{1:T})]_1\right)}$, where the term $\mathbb{I}_{(\cdot)}$ stands for an indicator function. Summing up, all the above restrictions result in the set Θ_0. For all elements $\boldsymbol{\vartheta}$ contained in Θ_0 we use a uniform prior and for all $\boldsymbol{\vartheta} \notin \Theta_0$ we set $\tilde{\pi}(\boldsymbol{\vartheta}) = 0$.

After the prior has been specified, parameter estimates are obtained as follows.

Step 1:	Run multistart random search methods, generate $\boldsymbol{\vartheta}^{(\mathtt{n})}$, where $\mathtt{n} = 1,\ldots,\mathtt{N} = 2,000$.
Step 2:	Run $MCMC$:

For each $MCMC$-step $\mathtt{m}$, where $\mathtt{m} = 1,\ldots,\mathtt{M} = 20,000$,

update $\boldsymbol{\vartheta}^{(\mathtt{m})}$ block-wise by means of the Metropolis-Hastings algorithm:

$MCMC$ Sub-Step 1: update block $\mathbb{J}_1$

$\vdots$

$MCMC$ Sub-Step K: update block $\mathbb{J}_{\mathtt{K}}$

$MCMC$ Sub-Step: reversible jump step (with a probability of 90%).

Obtain an estimate $\widehat{\boldsymbol{\vartheta}}$ from the draws $\boldsymbol{\vartheta}^{(\mathtt{m})}$, where $\mathtt{m} = \mathtt{M}_b + 1,\ldots,\mathtt{M} = 20,000$.

Ad Step 1: Given the set Θ_0, we randomly generate initial points $\boldsymbol{\vartheta}^{(\mathtt{n})}$, $\mathtt{n} = 1,\ldots,\mathtt{N}= 2,000$, which are independently drawn by means of $[\boldsymbol{\vartheta}^{(\mathtt{n})}]_j = [\boldsymbol{\vartheta}]_j + c_\vartheta[|\boldsymbol{\vartheta}|]_j \varepsilon_j$ for elements j, $j \in \{1,\ldots,\mathfrak{p}\}$, when the support is the real axis and $\log[|\boldsymbol{\vartheta}^{(\mathtt{n})}|]_j = \log[|\boldsymbol{\vartheta}|]_j + c_\vartheta \varepsilon_j$ such that $[\boldsymbol{\vartheta}^{(\mathtt{n})}]_j = \exp\left(\log\left([|\boldsymbol{\vartheta}|]_j\right) + c_\vartheta \varepsilon_j\right)\mathrm{sgn}\left([\boldsymbol{\vartheta}]_j\right)$ for elements j from non-positive or non-negative part of the real axis. The random variables ε_j are *iid* standard normal and only $\boldsymbol{\vartheta}^{(\mathtt{n})}$ with $\tilde{\pi}\left(\boldsymbol{\vartheta}^{(\mathtt{n})}\right) > 0$ are used. In addition, as already stated in Section 4, our random search routine also generates samples, where $\left(\theta^P\right)^{(\mathtt{n})} = \left(\theta^Q\right)^{(\mathtt{n})}$. This is done by setting $\left(\theta^P\right)^{(\mathtt{n})}$ equal to the sampled $\left(\theta^Q\right)^{(\mathtt{n})}$ with a probability of 80%. By sorting $\boldsymbol{\vartheta}^{(\mathtt{n})}$ according to $Q_T(\boldsymbol{\vartheta}^{(\mathtt{n})}; \mathbf{y}_{1:T})$ in ascending order, we are equipped with the sorted draws $\boldsymbol{\vartheta}_{[j]}$ and distances $Q_T\left(\boldsymbol{\vartheta}_{[j]}; \mathbf{y}_{1:T}\right)$, where $Q_T\left(\boldsymbol{\vartheta}_{[1]}; \mathbf{y}_{1:T}\right) \leq Q_T\left(\boldsymbol{\vartheta}_{[2]}; \mathbf{y}_{1:T}\right) \leq \cdots \leq Q_T\left(\boldsymbol{\vartheta}_{[\mathtt{N}]}; \mathbf{y}_{1:T}\right)$. The GMM distance function $Q_T\left(\boldsymbol{\vartheta}; \mathbf{y}_{1:T}\right)$ is defined in (23), where $\mathbf{C}_T = \mathbf{I}_{\mathfrak{q}}$ for all $\mathtt{n} = 1,\ldots,\mathtt{N}$.

Ad Step 2: Based on the results in Chernozhukov and Hong (2003), the Metropolis-Hastings algorithm (see, e.g., Robert and Casella, 2004) can be used to minimize the $CUE - GMM$ criterion function $Q_T(\cdot)$. To do this we proceed as follows: Suppose that $\boldsymbol{\vartheta}^{(\mathtt{m}-1)}$ is available, where, just now, $\mathtt{m}$ stands for the index of the $MCMC$ step. For $\mathtt{m} = 1$ we start the Bayesian sampler at $\boldsymbol{\vartheta}_{[1]}$, that is $\boldsymbol{\vartheta}^{(0)} = \boldsymbol{\vartheta}_{[1]}$.

The parameter vector to be updated, $\boldsymbol{\vartheta}^{(\mathtt{m}-1)}$, is of dimension $\mathfrak{p}$, where the index set $\{1, \ldots, \mathfrak{p}\}$ is covered by the blocks $\mathbb{J}_{\mathtt{k}} \subset \{1, \ldots, \mathfrak{p}\}$, $\mathtt{k} = 1, \ldots, \mathtt{K} = 5$. The first block $\mathbb{J}_1$ consists of the first two parameters and the 19th parameter, which is γ_0, $\mathbb{J}_2 = \{3, \ldots, 9\}$, the third block $\mathbb{J}_3 = \{10, \ldots, 15\}$, while $\mathbb{J}_4 = \{16, 17, 18\}$. Finally, the fifth block $\mathbb{J}_5$ contains the volatility parameters. For the parameter odering see first column of Table 1.

Within updating step $\mathtt{m}$, we consider the sub-steps $\mathtt{k} = 1, \ldots, \mathtt{K}$, where $\boldsymbol{\vartheta}^{(\mathtt{m},\mathtt{k})}$ stands for the parameter vector in $MCMC$-step $\mathtt{m}$ at sub-step $\mathtt{k}$. Let $\boldsymbol{\vartheta}^{old} = \boldsymbol{\vartheta}^{(\mathtt{m}-1)} = \boldsymbol{\vartheta}^{(\mathtt{m}-1,\mathtt{K})}$ for $\mathtt{k} = 1$ and $\boldsymbol{\vartheta}^{old} = \boldsymbol{\vartheta}^{(\mathtt{m},\mathtt{k}-1)}$ for $\mathtt{k} = 2, \ldots, \mathtt{K}$.[28] When the block $\mathbb{J}_{\mathtt{k}}$ is considered, $[\boldsymbol{\vartheta}^{old}]_i$, $i \in \mathbb{J}_{\mathtt{k}}$, is updated. To update $[\boldsymbol{\vartheta}^{(\mathtt{m},\mathtt{k}-1)}]_i$, $i \in \mathbb{J}_{\mathtt{k}}$, a random walk proposal, with *proposal density* $q\left([\boldsymbol{\vartheta}^{new}]_i \mid [\boldsymbol{\vartheta}^{old}]_i\right) = f_{\mathcal{N}\left([\boldsymbol{\vartheta}^{old}]_i, \sigma_{RWi}^2\right)}([\boldsymbol{\vartheta}^{new}]_i)$, is used, where $f_{\mathcal{N}(\cdot)}(\cdot)$ stands for a normal density. In the random walk proposals, we use small standard deviations of the noise in relative terms. In particular, $\sigma_{RWi} = 0.01[|\boldsymbol{\vartheta}^{old}|]_i$, with a probability of 90%, for remaining 10% we set the standard deviation of this noise term equal to $\sigma_{RWi} = 0.005[|\boldsymbol{\vartheta}^{old}|]_i$. By applying these proposals to all elements $i \in \mathbb{J}_{\mathtt{k}}$, we get the parameter vector $[\boldsymbol{\vartheta}^{new}]_i$ and the proposal density $q\left(\boldsymbol{\vartheta}^{new}|\boldsymbol{\vartheta}^{old}\right) = \prod_{i \in \mathbb{J}_k} q\left([\boldsymbol{\vartheta}^{new}]_i \mid [\boldsymbol{\vartheta}^{old}]_i\right)$. For the remaining components $[\boldsymbol{\vartheta}^{new}]_\ell = [\boldsymbol{\vartheta}^{old}]_\ell$, where ℓ is not contained in the block $\mathbb{J}_k$. Equipped with $Q_T(\boldsymbol{\vartheta}^{new}; \mathbf{y}_{1:T})$ and $Q_T(\boldsymbol{\vartheta}^{old}; \mathbf{y}_{1:T})$, the prior $\tilde{\pi}(\cdot)$ and the proposal densities $q(\cdot)$, the Metropolis-Hastings algorithm can be used. Let $\mathscr{L}(\boldsymbol{\vartheta}) = \exp\left[-\frac{1}{2}TQ_T(\boldsymbol{\vartheta}; \mathbf{y}_{1:T})\right]$. The GMM distance function $Q_T(\boldsymbol{\vartheta}; \mathbf{y}_{1:T})$ is defined in (23), where $\mathbf{C}_T = \left(\hat{\mathbf{\Lambda}}_T(\boldsymbol{\vartheta}^{(\mathtt{m}-1)})\right)^{-1}$ with $\hat{\mathbf{\Lambda}}_T\left(\boldsymbol{\vartheta}^{(\mathtt{m}-1)}\right) = \frac{1}{T-1}\sum_{\mathtt{t}=2}^T \mathbf{h}_{(\mathtt{t})}(\boldsymbol{\vartheta}^{(\mathtt{m}-1)}; \mathbf{y}_{1:T})\mathbf{h}_{(\mathtt{t})}(\boldsymbol{\vartheta}^{(\mathtt{m}-1)}; \mathbf{y}_{1:T})'$. Then, a transition from $\boldsymbol{\vartheta}^{old}$ to $\boldsymbol{\vartheta}^{new}$ is accepted with probability

$$\varrho\left(\boldsymbol{\vartheta}^{old}, \boldsymbol{\vartheta}^{new}\right) = \min\left\{1, \frac{\mathscr{L}(\boldsymbol{\vartheta}^{new})}{\mathscr{L}(\boldsymbol{\vartheta}^{old})}\frac{\tilde{\pi}(\boldsymbol{\vartheta}^{new})}{\tilde{\pi}(\boldsymbol{\vartheta}^{old})}\frac{q\left(\boldsymbol{\vartheta}^{old}|\boldsymbol{\vartheta}^{new}\right)}{q\left(\boldsymbol{\vartheta}^{new}|\boldsymbol{\vartheta}^{old}\right)}\right\}. \tag{75}$$

To implement this Metropolis-Hastings step, we draw a $[0, 1]$ uniform random variable and accept $\boldsymbol{\vartheta}^{new}$, i.e. $\boldsymbol{\vartheta}^{(\mathtt{m},\mathtt{k})} = \boldsymbol{\vartheta}^{new}$, if this uniform random variable is smaller or equal to $\varrho\left(\boldsymbol{\vartheta}^{old}, \boldsymbol{\vartheta}^{new}\right)$, otherwise $\boldsymbol{\vartheta}^{(\mathtt{m},\mathtt{k})} = \boldsymbol{\vartheta}^{old}$. By our assumptions on the prior, it follows that $\tilde{\pi}(\boldsymbol{\vartheta}^{new}) = \tilde{\pi}(\boldsymbol{\vartheta}^{old})$ as long as $\boldsymbol{\vartheta}^{new} \in \Theta_0$. Whenever $\boldsymbol{\vartheta}^{new} \notin \Theta_0$, then the probability ϱ equals to zero. Due the random walk proposal described above, we observe that $q\left(\boldsymbol{\vartheta}^{old}|\boldsymbol{\vartheta}^{new}\right) = q\left(\boldsymbol{\vartheta}^{new}|\boldsymbol{\vartheta}^{old}\right)$. The next block is then updated such that $\boldsymbol{\vartheta}^{old}$ becomes equal

---

[28]The index of the sub-step $\mathtt{k}$ is not applied, when it is not essential.

to the current $\boldsymbol{\vartheta}^{(\mathrm{m},\mathrm{k})}$. After having performed these updating steps for all blocks, $\mathrm{k} = 1, \ldots, \mathrm{K}$, we obtain $\boldsymbol{\vartheta}^{(\mathrm{m})} = \boldsymbol{\vartheta}^{(\mathrm{m},\mathrm{K})}$.

To improve the properties of the Bayesian sampler in the case when $\theta^Q = \theta^P$ or when $\theta^Q \neq \theta^P$, a reversible jump move based on Green (1995) and Richardson and Green (1997) has been implemented. Suppose that $\boldsymbol{\vartheta}^{(\mathrm{m})}$ has been obtained by the above steps. Let $\boldsymbol{\vartheta}^{old} = \boldsymbol{\vartheta}^{(\mathrm{m})}$. With a probability of 90% we add the following step to sampling step m: Consider the state s_1, where $\theta^Q = \theta^P$ and state s_2, where $\theta^Q \neq \theta^P$. The state S is Bernoulli distributed random variable with prior probability $\mathbb{P}(S = s_1) = p_{s_1} = 0.90$. By applying Green (1995), transitions from $\{S = s_1\}$ to $\{S = s_2\}$ and vice versa can be performed by means of the Metropolis Hastings algorithm. In particular, consider the uniformly distributed random variable η, as well as the normal *iid* random variables u and u_γ. The proposal densities are $f_{\mathcal{N}(0,\sigma_u^2)}(u)$ and $f_{\mathcal{N}(0,\sigma_{u_\gamma}^2)}(u_\gamma)$. Let $\{S = s_1\}$, where $\theta^{old} = \theta^Q = \theta^P$. A possible split transition from $\{S = s_1\}$ to $\{S = s_2\}$ works as follows

$$
\begin{aligned}
\theta^{P,new} &= \theta^{old} - 2\eta u \ , \\
\theta^{Q,new} &= \theta^{old} + 2(1-\eta)u \ , \\
\gamma_0^{new} &= \gamma_0^{old} - 2\eta u + u_\gamma.
\end{aligned}
\tag{76}
$$

By replacing the corresponding elements in $\boldsymbol{\vartheta}^{old}$ by $\theta^{P,new}$, $\theta^{Q,new}$ and γ_0^{new}, we get the new parameter vector $\boldsymbol{\vartheta}^{new}$.[29] Let $\boldsymbol{\chi}^{old} = (s_1; \theta^{old}, \gamma_0^{old})$ and $\boldsymbol{\chi}^{new} = (s_2; \theta^{P,new}, \theta^{Q,new}\gamma_0^{new})$. By taking partial derivatives of the terms in (76), we obtain the Jacobian matrix

---

[29]Note that $\boldsymbol{\vartheta}^{old}$ contains the old parameters, where $\theta^P = \theta^Q$. In the notation of Green (1995), the dimension of the parameter of interest with state s_1 is $n_1 = 2$, (consisting of θ^{old} and γ_0^{old}), the dimension of the noise component is $m_1 = 3$ (due to η, u and u_γ). With s_2 we get $n_2 = 3$ (consisting of $\theta^{P,new}$, $\theta^{Q,new}$ and γ_0^{new}) and $m_2 = 2$ (due to η and u_γ). This yields, $n_1 + m_1 = n_2 + m_2$.

$$\mathbf{J} = \frac{\partial(\theta^P, \theta^Q, \eta, \gamma_0, u_\gamma)'}{\partial(\theta, \eta, \gamma_0, u, u_\gamma)'} = \begin{pmatrix} 1 & -2u & 0 & -2\eta & 0 \\ 1 & -2u & 0 & 2(1-\eta) & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -2u & 1 & -2\eta & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{77}$$

The determinant of the matrix $\mathbf{J}$ is equal to 2. Given the proposal densities $q(u) = f_{\mathcal{N}(0,\sigma_u^2)}(u)$ and $q(u_\gamma) = f_{\mathcal{N}(0,\sigma_{u_\gamma}^2)}(u_\gamma)$ for u and u_γ, a transition from $\boldsymbol{\chi}^{old}$ to $\boldsymbol{\chi}^{new}$ is accepted with probability (see Green, 1995, equation (7))

$$\begin{aligned} &\varrho\left(\boldsymbol{\chi}^{old}, \boldsymbol{\chi}^{new}\right) \\ &= \min\left(1, \frac{\mathcal{L}(\boldsymbol{\vartheta}^{new})}{\mathcal{L}(\boldsymbol{\vartheta}^{old})} \frac{\tilde{\pi}(\boldsymbol{\vartheta}^{new})}{\tilde{\pi}(\boldsymbol{\vartheta}^{old})} \frac{1 - p_{s_1}}{p_{s_1}} \frac{f_{\mathcal{N}(0,\sigma_{u_\gamma}^2)}\left(u_\gamma^{old}\right)}{f_{\mathcal{N}(0,\sigma_{u_\gamma}^2)}\left(u_\gamma^{new}\right) f_{\mathcal{N}(0,\sigma_u^2)}(u)} |\mathbf{J}|\right) \\ &= \min\left(1, \frac{\mathcal{L}(\boldsymbol{\vartheta}^{new})}{\mathcal{L}(\boldsymbol{\vartheta}^{old})} \frac{\tilde{\pi}(\boldsymbol{\vartheta}^{new})}{\tilde{\pi}(\boldsymbol{\vartheta}^{old})} \frac{1 - p_{s_1}}{p_{s_1}} \frac{2}{f_{\mathcal{N}(0,\sigma_u^2)}(u)}\right). \end{aligned} \tag{78}$$

Since $u_\gamma^{old} = \gamma_0^{new} - \gamma_0^{old} + 2\eta u$, the densities $f_{\mathcal{N}(0,\sigma_\gamma^2)}$ cancel out in (78). An equivalent Metropolis-Hastings move can be performed without an update of γ_0. A possible merge transition from $\{S = s_2\}$ to $\{S = s_1\}$ works as follows

$$\begin{aligned} \theta^{new} &= \theta^{Q,new} = \theta^{P,new} = (1-\eta)\theta^{Q,new} + \eta\theta^{Q,old} \\ \gamma_0^{new} &= \gamma_0^{old} - 2\eta u + u_\gamma \;, \text{ such that} \\ u &= \frac{\theta^P - \theta^{new}}{-2\eta} = \frac{\theta^Q + \theta^{new}}{2(1-\eta)}. \end{aligned} \tag{79}$$

By means of (79) we get θ^{new} and γ_0^{new}. Then, a transition from $\boldsymbol{\chi}^{old} = \left(s_2; \theta^{P,old}, \theta^{Q,old}, \gamma_0^{old}\right)$ to $\boldsymbol{\chi}^{new} = \left(s_1; \theta^{P,new} = \theta^{Q,new} = \theta^{new}, \gamma_0^{new}\right)$ is accepted with probability (Green, 1995, equation (7)):

$$\varrho\left(\boldsymbol{\chi}^{old}, \boldsymbol{\chi}^{new}\right) = \min\left\{1, \frac{\mathcal{L}(\boldsymbol{\vartheta}^{new})}{\mathcal{L}(\boldsymbol{\vartheta}^{old})} \frac{p_{s_1}}{1 - p_{s_1}} f_{\mathcal{N}(0,\sigma_u^2)}(u) \frac{1}{2}\right\}. \tag{80}$$

If either a split or a merge transition is accepted we set $\boldsymbol{\vartheta}^{(m)} = \boldsymbol{\vartheta}^{new}$. After a merge move $\theta^P = \theta^Q$ in updating sub-step $\mathtt{k} = 1$, until a split move takes place.

Parameter Estimation: To obtain the parameter estimates $\widehat{\boldsymbol{\vartheta}}$, we consider the draws $\boldsymbol{\vartheta}^{(\mathtt{m})}$, where $\mathtt{m} = \mathtt{M}_b + 1, \ldots, \mathtt{M}$ of the convergent part of the Markov chain. We work with $\mathtt{M}_b = 5,000$ and $\mathtt{M} = 20,000$. Then $\widehat{\boldsymbol{\vartheta}}$ is provided by the sample mean. Tables 1 and 2 show parameter estimates obtained by using the Bayesian algorithm described above.

In addition, as shown by Chernozhukov and Hong (2003), the draws after burn-in phase can also be used to estimate the asymptotic variance of the parameters. To do this, we can simply calculate the sample variance of $\widehat{\boldsymbol{\vartheta}}^{(m)}$, where $\mathtt{m} = \mathtt{M}_b + 1, \ldots, \mathtt{M}$. To account for the serial correlation observed with the Markov chain, we follow Bayesian literature to estimate the variance of the components of $\widehat{\boldsymbol{\vartheta}}$ by means of the batch-means approach described in Flegal and Jones (2010)[in particular, Equation (6) is used].

Monte Carlo Study: In the simulation studies described in Section 4, $\mathtt{Steps\ 1}$ and $\mathtt{2}$ are performed for each Monte-Carlo replication ($\mathtt{l} = 1, \ldots, \mathtt{L} = 200$).

Remark 2. The implementation of the Quasi-Bayesian sampler based Chernozhukov and Hong (2003) is not "free of cost". Running multistart random search methods and a standard minimization procedure and then performing the Wald test based on (26) takes approximately 20 minutes, while one full estimation step based on running a random search and then obtaining 20,000 draws from a Markov Chain lasts for approximately 24 hours on the same standard PC.

$\boldsymbol{\vartheta}$		mean $\widehat{\boldsymbol{\vartheta}}$	median	min	max	std	skew	kurt	$\lvert \boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}} \rvert$
θ^Q	10	10.3593	8.9022	1.0527	69.9629	6.0544	3.4352	23.6299	0.3593
θ^P	1.5	1.5046	1.2986	0.0676	6.4437	1.0296	1.3471	5.1909	0.0046
β_Q^{11}	-1	-1.2823	-1.0328	-7.6430	-0.1108	0.9617	-2.0173	9.4853	0.2823
β_Q^{21}	0.2	0.2523	0.1729	0.0099	2.8282	0.2549	3.2707	22.3280	0.0523
β_Q^{31}	0.02	0.0326	0.0204	0.0009	0.5962	0.0416	5.2132	49.7281	0.0126
β_Q^{22}	-1	-1.5493	-1.4686	-4.2679	-0.1046	0.7283	-0.5842	3.2132	0.5493
β_Q^{32}	0.04	0.0375	0.0354	-0.0734	0.1586	0.0404	0.1497	2.8928	0.0025
β_Q^{23}	0	-0.0005	-0.0002	-0.0343	0.0303	0.0097	0.0013	3.1280	0.0005
β_Q^{33}	-1	-1.5042	-1.4266	-4.7165	-0.0664	0.7906	-0.5289	3.0549	0.5042
β_P^{11}	-0.8	-1.6204	-0.8868	-43.8618	-0.0503	2.9373	-8.3139	99.3434	0.8204
β_P^{21}	0.02	0.0330	0.0210	0.0013	0.3927	0.0378	3.0352	17.4456	0.0130
β_P^{31}	0.01	0.0168	0.0102	0.0004	0.2022	0.0212	4.1593	27.5666	0.0068
β_P^{22}	-0.7	-0.9193	-0.8646	-3.1251	0.2598	0.5646	-0.5446	2.8395	0.2193
β_P^{32}	0.01	0.0094	0.0094	-0.0182	0.0433	0.0099	-0.5446	2.8395	0.0006
β_P^{23}	0	0.0000	-0.0003	-0.0316	0.0305	0.0099	0.0824	2.8274	0.0000
β_P^{33}	-0.7	-0.9199	-0.8383	-3.0770	0.2518	0.5418	-0.5914	3.0650	0.2199
$\mathcal{B}_{12}^x$	0.05	0.0791	0.0496	0.0029	1.2802	0.0964	4.3127	35.8452	0.0291
$\mathcal{B}_{13}^x$	0.1	0.1590	0.0978	0.0025	2.1414	0.1969	4.3398	32.1066	0.0590
γ_0	2	2.1224	2.1483	-4.1375	6.2455	1.6726	-0.2254	2.9938	0.1224
Σ_1	0.7	0.5450	0.4636	0.0176	2.7764	0.3771	1.6212	7.1421	0.1550
Σ_2	1	1.0037	0.7162	0.0267	5.8461	0.8681	1.9397	7.7289	0.0037
Σ_3	0.8	0.8538	0.6104	0.0253	8.2162	0.8849	3.6164	22.8961	0.0538
σ_ε^2	0.0067	0.0119	0.0068	0.0003	0.2908	0.0180	7.8621	102.5576	0.0051

Table 5: Parameter estimates for the $\mathbb{A}_1(3)$. Data simulated with $M = 10$ and $T = 500$. Estimation based on using `fminsearch`. $c_\vartheta = 1$ is controlling for the noise in the generation of the starting value of the optimization routine. Statistics are obtained from $1,000$ simulation runs. *mean*, *median*, *min*, *max*, *std*, *skew* and *kurt* stand for the sample mean, median, minimum, maximum, standard deviation, skewness and kurtosis of the point estimates $\widehat{\boldsymbol{\vartheta}}_\ell$, $\ell = 1, \ldots, 1,000$. $\lvert \boldsymbol{\vartheta} - \widehat{\boldsymbol{\vartheta}} \rvert$ stands for absolute value of the mean deviation from the true parameter. The true parameter values $\boldsymbol{\vartheta}$ are reported in the second column.

	$\theta^Q = 10 \neq 1.5 = \theta^P$		$\theta^Q = \theta^P = 1.5$	
α_S	Wald	DD	Wald	DD
0.01	0.018	0.545	0.015	0.057
0.05	0.028	0.583	0.021	0.062
0.10	0.043	0.623	0.025	0.065

Table 6: Parameter tests: Data are simulated with $M = 10$, $T = 500$ and $c_\vartheta = 1$. $[\vartheta]_1 = \theta^Q$ and $[\vartheta]_2 = \theta^P$ and the remaining elements of $\boldsymbol{\vartheta}$ are equal to those of the second column in Table 5. α_S stands for the significance level. c_ϑ controls for the noise in the generation of the starting value of the optimization routine. The null hypothesis is $\theta^Q = \theta^P$ against the two sided alternative $\theta^Q \neq \theta^P$. The parameters ϑ estimated by combining multistart random search methods and a standard minimization procedure. The Wald test as well as the distance difference test (DD) are implemented as described in Chapter 22 in Ruud (2000). Equation (26) is used to estimate the asymptotic variance of $\sqrt{T}\left(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\right)$ with the Wald test, while $\hat{\boldsymbol{\Lambda}}_T$, as presented in (26), is used with the distance difference test. The numbers in the table are rejection rates of the null hypothesis given the significance level α_S, when using a Wald test and a distance difference test. Statistics are obtained from $1,000$ simulation runs.

References

Aït-Sahalia, Y. (1996a). Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64:527–560.

Aït-Sahalia, Y. (1996b). Testing continuous-time models of the spot interest rate. *The Review of Financial Studies*, 9(2):385–426.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica*, 70:223–262.

Aït-Sahalia, Y. and Kimmel, R. L. (2010). Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions. *Journal of Financial Economics*, 98:113–144.

Altonji, J. G. and Segal, L. M. (1996). Small-Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business & Economic Statistics*, 14(3):353–66.

Andersen, T. G., Chung, H.-J., and Sorensen, B. E. (1999). Efficient method of moments estimation of a stochastic volatility model: A monte carlo study. *Journal of Econometrics*, 91(1):61–87.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

Chen, H. and Joslin, S. (2012). Generalized transform analysis of affine processes and applications in finance. *The Review of Financial Studies*, 25(7):2225–2256.

Chen, L., Lesmond, D. A., and Wei, J. (2007). Corporate yield spreads and bond liquidity. *The Journal of Finance*, 62(1):119–149.

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

Cheridito, P., Filipovic, D., and Kimmel, R. L. (2008). A Note on the Dai-Singleton Canonical Representation of Affine Term Structure Models. *SSRN eLibrary*.

Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115:293–346.

Chib, S. and Ergashev, B. (2009). Analysis of multifactor affine yield curve models. *Journal of the American Statistical Association*, 104(488):1324–1337.

Cochrane, J. (2005). *Asset Pricing*. Princeton University Press, revised edition.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Cuchiero, C., Teichmann, J., and Keller-Ressel, M. (2012). Polynomial processes and their application to mathematical finance. *Finance & Stochastics*, 16(4):711–740.

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models. *Journal of Finance*, 55(5):1943–1978.

Dai, Q. and Singleton, K. J. (2003). Term structure dynamics in theory and reality. *The Review of Financial Studies*, 16(3):631–678.

Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics*, 131:309–338.

Duffee, G. R. (2011). Information in (and not in) the term structure. *The Review of Financial Studies*, 24(9):2895–2934.

Duffie, D., Filipović, D., and Schachermayer, W. (2003). Affine processes and applications in finance. *Annals of Applied Probability*, 13:984–1053.

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

Duffie, D., Pan, J., and Singleton, K. J. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Egorov, A. V., Li, H., and Ng, D. (2011). A tale of two yield curves: Modeling the joint term structure of dollar and euro interest rates. *Journal of Econometrics*, 162(1):55–70.

Filipović, D. (2009). *Term-Structure Models: A Graduate Course.* Springer, Berlin.

Filipović, D., Mayerhofer, E., and Schneider, P. (2013). Transition density approximations for multivariate affine jump diffusion processes. *Journal of Econometrics*, 176:93–111.

Flegal, J. M. and Jones, G. L. (2010). Batch Means and Spectral Variance Estimators in Markov Chain Monte Carlo. *The Annals of Statistics*, 38(2):1034–1070.

Frühwirth-Schnatter, S. and Geyer, A. (1996). Bayesian estimation of economemtric multi-factor cox-ingersoll-ross-models of the term structure of interest rates via MCMC methods. Working paper, Vienna University of Economics and Business.

Glasserman, P. and Kim, K.-K. (2010). Moment Explosions and Stationary Distributions in Affine Diffusion Models. *Mathematical Finance*, 20(1):1–33.

Grasselli, M. and Tebaldi, C. (2008). Solvable affine term structure models. *Mathematical Finance*, 18(1):135–153.

Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732.

Guggenberger, P. and Smith, R. J. (2005). Generalized empirical likelihood estimators and tests under partial, weak, and strong identification. *Econometric Theory*, null:667–709.

Hamilton, J. D. and Wu, J. C. (2012). Identification and estimation of Gaussian affine term structure models. *Journal of Econometrics*, 168(2):315 – 331.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Jones, C. S. (2003). Nonlinear mean reversion in the short-term interest rate. *The Review of Financial Studies*, 16(3):793–843.

Joslin, S., Singleton, K. J., and Zhu, H. (2010). A new perspective on gaussian dynamic term structure models. *The Review of Financial Studies*, 24:926–970.

Keller-Ressel, M. and Mayerhofer, E. (2012). Exponential moments of affine processes. Technical report, Deutsche Bundesbank, Frankfurt.

Klenke, A. (2008). *Probability Theory - A Comprehensive Course.* Springer.

Mayerhofer, E., Pfaffel, O., and Stelzer, R. (2010). On strong solutions of matrix valued jump-diffusions. *preprint*.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam.

Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719.

Perko, L. (1991). *Differential Equations and Dynamical Systems.* Texts in Applied Mathematics, No. 7. Springer.

Piazzesi, M. (2010). *Affine Term Structure Models.* In Y. Aït-Sahalia and L. Hansen (Eds.), Handbook of Financial Econometrics, North-Holland, Amsterdam.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach.* MIT Press, Cambridge, Massachusetts.

Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic Nonlinear Econometric Models, Asymptotic Theory.* Springer, New York.

Richardson, S. and Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):731–792.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition.

Ruud, P. A. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press, New York.

Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance*, 52(5):1973–2002.

Törn, A. and Zilinskas, A. (1989). *Global Optimization*. Lecture Notes in Computer Science 350. Springer.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1):25–51.

Zhou, H. (2001). Finite sampler properties of EMM, GMM, QMLE, and MLE for a square-root interest rate diffusion model. *Journal of Computational Finance*, 5:89–122.

Zhou, H. (2003). Itô conditional moment generator and the estimation of short-rate process. *Journal of Financial Econometrics*, 1(2):250–271.