GAUSSIAN MODELS FOR EURO HIGH GRADE GOVERNMENT YIELDS

Dr Marco Realdon
Newcastle Business School
Northumbria University
Newcastle Upon Tyne
City Campus East, NE1 8ST, UK
tel: +39/338/2875027
marco.realdon@gmail.com

27/11/2015 (this version 1/12/2015)

Abstract

This paper tests affine, quadratic and Black-type Gaussian models on Euro area triple A Government bond yields for maturities up to 30 years. Quadratic Gaussian models beat affine Gaussian models both in sample and out of sample. A Black-type model best fits the shortest maturities and the extremely low yields since 2013, but worst fits the longest maturities.

Even for quadratic models we can infer the latent factors from some yields observed without errors, which makes quasi maximum likelihood (QML) estimation feasible. New specifications of quadratic models fit the longest maturities better than does the "classic" specification of Ahn-Dittmar-Gallant (2002), but the opposite is true for the shortest maturities. These new specifications are more suitable to QML estimation.

Overall quadratic models seem preferable to affine Gaussian models, because of superior empirical performance, and to Black-type models, because of superior tractability.

This paper also proposes the vertical method of lines (MOL) to solve numerically partial differential equations (PDE's) for pricing bonds under multiple non-independent stochastic factors. "Splitting" the PDE drastically reduces computations. Vertical MOL can be considerably faster and more accurate than finite difference methods.

Key words: affine Gaussian models, quadratic Gaussian models, Black model, vertical method of lines, sequential splitting, quasi-maximum likelihood, Extended Kalman Filter.

JEL classification: G12; G13.

1

1 Introduction

This paper tests linear and non-linear Gaussian term structure models on Euro area triple A rated Government bond yields for long term maturities up to 30 years. By Gaussian term structure models we mean that the short interest rate is a function of latent factors whose stochastic process is Gaussian. The exclusive focus on Gaussian models is due to their tractability and relevance in practice. Moreover Gaussian models are not plagued by the "admissibility" issues explained in Duffie-Kan (1996) for affine models whose factors' volatilities are driven by "square roots" of the factors.

All the models we tests assume three Gaussian latent factors and are either affine **Gaussian** models, or quadratic models or an extension of Black's model (1995). **In this paper "affine Gaussian" model means that all model factors are homoscedastic and Gaussian.**

One question is whether affine Gaussian models, simple as they are, suffice or whether the complication of the non-linear Gaussian models is worthwhile. The main empirical finding is that indeed the complication of non-linear Gaussian models seems worthwhile, because of their superior fit to observed yields. Quadratic Gaussian models beat affine Gaussian models both in sample and out of sample, especially when yields are close to zero. The extended Black model beats even quadratic models out of sample and seems the best model to match the extremely low short term yields observed since 2013, but is the worst model in fitting the longest maturities.

Another question is whether new specifications of quadratic models can better fit yields than the "classic" specification of Ahn-Ditmar-gallant (2002). **Indeed some of these new specifications better fit the longest yield maturities than the "classic" specification**, but the "classic" specification better fits the shortest maturities.

Estimation through quasi maximum likelihood (QML), which assumes that some yields are observed without error, is feasible even for quadratic models, **especially under the new specifications,** and largely confirms the results of Extended Kalman Filter estimation. However QML is not reliable for the extended Black model.

Overall, to price Euro area default-free bonds, quadratic models seem preferable to affine **Gaussian** models, because of superior empirical performance, and to the Black-type models, for now, because of superior tractability.

This paper also proposes the vertical method of lines (MOL) to solve numerically partial differential equations (PDE's) for pricing bonds according to Black's model. In the presence of multiple non-independent stochastic factors, "splitting" the PDE drastically reduces computations and can be considerably faster and more accurate than finite difference methods.

2 Literature

The literature on dynamic term structure models cannot be here summarised. Dai and Singleton (2003) give a good overview. This paper belongs to the branch of such literature that focuses on Gaussian term structure models and makes no reference to the macro-economy. To such branch belong also the pioneering study of Langetieg (1980), the empirical Kalman filter test of Babbs and Nowman (1999), the work of Dai and Singleton (2002) and many others. More recently Joslin et al. (2011) provide canonical identifiable affine Gaussian models whereby the factors are observable portfolios of yields. Duffee (2011) shows that the yields of affine Gaussian models cannot detect variation in US Government bond risk-premia.

This paper also belongs to the branch of literature on quadratic Gaussian models. After some studies on quadratic models in the nineties, e.g. Costantinides (1992), Ahn et al. (2002), Leippold and Wu (2002, 2003) and Chen et al. (2004) further develop quadratic models. Ahn et al. (2002) show the empirical advantage of quadratic models over affine models. Gourieroux and Sufana (2003) and Realdon (2006) present discrete time quadratic models. Li and Zhao (2006) use quadratic models to provide evidence of unspanned stochastic volatility in derivatives prices.

This paper is also directly related to Gaussian models of the type first proposed by Black (1995). These are non-linear Gaussian models in which the short rate cannot turn negative. Gorovoi and Linetski (2004) present a closed form solution to Black's model. Realdon (2009) tests a two factor version of Black's model. Kim and Singleton (2012) test various extensions of Black's model.

We can contrast the present paper with the existing literature as follows. Most of the term structure literature focuses on US Government yields, whereas this study considers the triple A Government yields in the Euro area. This is of interest, apart from the sheer size of the Euro area economy, because of protracted periods of extremely low and even negative Euro yields.

Most of the term structure literature focuses on yield maturities up to ten years, whereas this study considers maturities up to thirty years. This seems a much more severe test for term structure models.

Most of the term structure literature focuses on affine or quadratic models, whereas this study also focuses on a new extension of the Black (1995) model. This paper also tests new specifications of quadratic models.

Most of the term structure literature that tests quadratic Gaussian models employs Kalman Filter estimation, whereas this study also employs Quasi Maximum Likelihood (QML) estimation for quadratic models. QML assumes that some yields are observed without error, so that the latent factors can be inferred. It turns out this is feasible even for quadratic Gaussian models, not only for affine ones.

The literature that implements the Black model tends **to** use finite difference methods for pricing bonds. Finite difference methods have been used for decades, while in the past decade finite element methods have become more popular in finance, a recent example of which is Rambeerich and others (2013).

Finite difference methods are easier to apply, since they typically involve the intuitive discretisation of the partial derivatives of the pricing equations, but are less stable and less efficient than finite element methods. Much of the interest in finite elements methods is due to the fact that they do not discretise the pricing equations in the time dimension and that they use fast and accurate algorithms for solving the resulting matrix exponential. This paper proposes an algorithm **to solve** the bond pricing equations of the Black model **by discretising the** partial derivatives **in the space dimensions**, as in finite difference methods, **but not in the time dimension, which again** requires the solution of a matrix exponential as in finite element methods. As such the proposed **algorithm** retains some of the key benefits of both finite difference methods and finite element methods. Such **algorithm** is the vertical method of lines (MOL) and involves "sequential splitting" when the pricing equation involves more than one stochastic factor. Vertical MOL poses no stability problems **when** the stochastic factors are not instantaneously correlated. If the factors are instantaneously correlated, vertical MOL is not unconditionally stable.

3   The extended Black model

We consider an extension of the Black (1995) model in which the time t default-free instantaneous short interest rate r_t is a function of the time t value of three latent factors $x_{1,t}, x_{2,t}, x_{3,t}$ so that, in a filtered probability space with the usual properties,

$$r_t = \alpha + \max\left(x_{1,t}, 0\right)^{q_1} + \max\left(x_{2,t}, 0\right)^{q_2} + \max\left(x_{3,t}, 0\right)^{q_3}. \tag{1}$$

$$dx_{1,t} = \kappa_1 \cdot \left(\mu_1 + p \cdot x_{2,t} - x_{1,t}\right) \cdot dt + \sigma_1 \cdot dw_{1,t}^Q \tag{2}$$

$$dx_{2,t} = \kappa_2 \cdot \left(\mu_2 - x_{2,t}\right) \cdot dt + \sigma_2 \cdot dw_{2,t}^Q \tag{3}$$

$$dx_{3,t} = \kappa_3 \cdot \left(\mu_3 - x_{3,t}\right) \cdot dt + \sigma_3 \cdot dw_{3,t}^Q \tag{4}$$

$$dw_{1,t}^Q \cdot dw_{2,t}^Q = \rho \cdot dt, \quad dw_{1,t}^Q \cdot dw_{3,t}^Q = dw_{2,t}^Q \cdot dw_{3,t}^Q = 0. \tag{5}$$

$q_1, q_2, q_3 \geq 1$ are constants. α is also a constant and can be negative so as to enable the model to match the slightly negative yields we observe in the sample. Equation 1 implies that $r_t \geq \alpha$ and yields for maturities longer than the instantaneous maturity are guaranteed to be higher than α, even when the factors have negative values. $dx_{1,t}$ is the stochastic differential of the factor x_1 and $dw_{1,t}^Q$ the stochastic differential of a Wiener process in the risk-neutral measure Q over the infinitesimal time interval $[t, t+dt]$. $dx_{2,t}$, $dw_{2,t}^Q$, $dx_{3,t}$, $dw_{3,t}^Q$ have similar meanings. The Wiener processes are not correlated except for $dw_{1,t}^Q \cdot dw_{2,t}^Q = \rho \cdot dt$. $\kappa_1, \sigma_1, \mu_1, \kappa_2, \sigma_2, \mu_2, \kappa_3, \sigma_3, \mu_3$ are all constant parameters. p is a constant and links the long term mean reversion level of x_1 to x_2. Therefore even when $\rho = 0$, x_1 and x_2 are not independent. All parameters are identifiable in estimation.

Let $V(\tau)$ with $\tau = T - t$ be the value at time t of a discount bond with maturity T and face value 1, so that $V(0) = 1$. Absent arbitrage, from the

above assumptions it follows that, dropping unnecessary time subscripts,

$$V(\tau) = \exp(-\alpha \cdot \tau) \cdot W \cdot U \tag{6}$$

$$\frac{\partial W}{\partial \tau} = \frac{\partial^2 W}{\partial x_1^2} \frac{1}{2} \sigma_1^2 + \frac{\partial W}{\partial x_1} \kappa_1 (\mu_1 - x_1) - W \cdot \max(x_1, 0)^{q_1} \tag{7}$$

$$+ \frac{\partial^2 W}{\partial x_1 x_2} \rho \sigma_1 \sigma_2 + \frac{\partial^2 W}{\partial x_2^2} \frac{1}{2} \sigma_2^2 + \frac{\partial W}{\partial x_2} \kappa_2 (\mu_2 - x_2) - W \cdot \max(x_2, 0)^{q_2}$$

$$\frac{\partial U}{\partial \tau} = \frac{\partial^2 U}{\partial x_3^2} \frac{1}{2} \sigma_3^2 + \frac{\partial U}{\partial x_3} \kappa_3 (\mu_3 - x_3) - U \cdot \max(x_3, 0)^{q_3} \tag{8}$$

$$\lim_{x_1 \to -\infty} \frac{\partial W}{\partial x_1^2} \to 0, \quad \lim_{x_1 \to \infty} \frac{\partial W}{\partial x_1^2} \to 0, \lim_{x_2 \to -\infty} \frac{\partial W}{\partial x_2^2} \to 0, \quad \lim_{x_2 \to \infty} \frac{\partial W}{\partial x_2^2} \to 0, \ W(0) = 1$$

$$\lim_{x_3 \to -\infty} \frac{\partial U}{\partial x_3^2} \to 0, \quad \lim_{x_3 \to \infty} \frac{\partial U}{\partial x_3^2} \to 0, \ U(0) = 1.$$

W is a function of x_1, x_2 and τ. U is a function of x_3 and τ. $W(0) = U(0) = 1$ are the values of W and U when $\tau = 0$. V tends to be linear in the factors x_1, x_2, x_3 as the factors tend to plus or minus infinity. Discount bond yields are computed as $\frac{-\ln V(\tau)}{\tau}$. The fact that x_3 is independent of the other factors simplifies the numerical solution to the bond pricing equation. Instead of solving for V on a grid with three "space" dimensions, we solve for W and U on separate grids. The grid for W has two "space" dimensions and the grid for U one "space" dimension. In the empirical part the partial differential equations for W and U are each solved through the vertical method of lines (MOL). The next section explains **the details of** vertical MOL. We also assume that in the physical probability measure

$$dx_{1,t} = \kappa_1^* \cdot (\mu_1^* + p^* \cdot x_{2,t} - x_{1,t}) \cdot dt + \sigma_1 \cdot dw_{1,t}^*$$
$$dx_{2,t} = \kappa_2^* \cdot (\mu_2^* - x_{2,t}) \cdot dt + \sigma_2 \cdot dw_{2,t}^*$$
$$dx_{3,t} = \kappa_3^* \cdot (\mu_3^* - x_{3,t}) \cdot dt + \sigma_3 \cdot dw_{3,t}^*$$
$$dw_{1,t}^* \cdot dw_{2,t}^* = \rho \cdot dt, \quad dw_{1,t}^* \cdot dw_{3,t}^* = dw_{2,t}^* \cdot dw_{3,t}^* = 0.$$

$dw_{1,t}^*, dw_{2,t}^*, dw_{3,t}^*$ are differentials of Wiener processes in the physical probability measure. The parameters with the superscript "$*$" signify that such parameters refer to the real probability measure.

Let $t = 0, 1, 2, 3, .., M$ denote the set of M dates on which we observe the yield curve. Δ is the time between consecutive observations and is approximately equal to the inverse of number trading days in one year. Therefore $\Delta = \frac{1}{261}$ since we observe about 261 daily prices per year in the data. Then in the empirical tests we approximate, using the Euler discretisation, the above stochastic differential equations. As a result of the said discretisation, the approximate physical conditional transition density of $\mathbf{x}_{t \cdot \Delta} = (x_{1,t \cdot \Delta}, x_{2,t \cdot \Delta}, x_{3,t \cdot \Delta})'$ given $\mathbf{x}_{(t-1) \cdot \Delta} = (x_{1,(t-1) \cdot \Delta}, x_{2,(t-1) \cdot \Delta}, x_{3,(t-1) \cdot \Delta})'$ is

$$N\left(\boldsymbol{\eta} + (\mathbf{I}_3 - \boldsymbol{\phi})\,\mathbf{x}_{t-1}, \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right) \tag{9}$$

$$\boldsymbol{\phi} = \begin{pmatrix} -\kappa_1^* & \kappa_1^* \cdot p^* & 0 \\ 0 & -\kappa_2^* & 0 \\ 0 & 0 & -\kappa_3^* \end{pmatrix} \cdot \Delta, \; \boldsymbol{\eta} = \begin{pmatrix} \mu_1^* \cdot \kappa_1^* \\ \mu_2^* \cdot \kappa_2^* \\ \mu_3^* \cdot \kappa_3^* \end{pmatrix} \cdot \Delta,$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ \rho_{12} \cdot \sigma_2 & \sqrt{1 - \rho_{12}^2} \cdot \sigma_2 & 0 \\ \rho_{13} \cdot \sigma_3 & \frac{\rho_{32} - \rho_{12} \cdot \rho_{13}}{\sqrt{1 - \rho_{12}^2}} \cdot \sigma_3 & \sqrt{1 - \rho_{13}^2 - \frac{(\rho_{32} - \rho_{12} \cdot \rho_{13})^2}{1 - \rho_{12}^2}} \cdot \sigma_3 \end{pmatrix} \sqrt{\Delta}$$

with $\rho = \rho_{12}, \rho_{13} = \rho_{32} = 0$. $N\left(\boldsymbol{\eta} + (\mathbf{I}_3 - \boldsymbol{\phi})\,\mathbf{x}_{t-1}, \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right)$ is the multivariate normal density with mean $\boldsymbol{\eta} + (\mathbf{I}_3 - \boldsymbol{\phi})\,\mathbf{x}_{t-1}$ and covariance $\boldsymbol{\Sigma}\boldsymbol{\Sigma}'$. $\mathbf{I}_3$ is the 3×3 identity matrix.

4 Vertical MOL

The above extended Black model has no closed form solution. When a pricing **partial differential equation (PDE)** cannot be solved in closed form, it is solved numerically usually with some finite difference method**, whereas** this paper **proposes** the vertical method of lines (MOL). Vertical MOL discretises the PDE **in** the "space" dimensions, but not **in** the "time" dimension, whereas horizontal MOL discretises the PDE **in** the "time" dimension, but not **in** the "space" dimensions. **The "space" discretisation of vertical MOL** reduces the pricing PDE to a system of ordinary differential equations (ODE's). When the pricing PDE is linear, as is typically the case in finance, the resulting ODE system can be solved by computing a matrix exponential, which can be done through well known and quick algorithms.

As a preliminary step, **we consider** how vertical MOL **is** an alternative to finite difference methods for solving pricing PDE's. As it involves no discretisation in the "time" dimension, vertical MOL can be quicker and more accurate than finite difference methods. For pricing contracts whose value is driven by one stochastic factor, only a single matrix exponential needs solving, posing no stability issues. **For contracts driven by two stochastic factors**, vertical MOL involves computing the exponential of very large matrixes, which can often be prohibitive. For example, assuming 100 steps in both "space" dimensions, the matrix exponential that needs solving is of size 10000×10000. With a relatively fast computer, Matlab can compute a 1000×1000 matrix exponential in about 3 seconds, but a 10000×10000 matrix exponential takes about 340 seconds, which is prohibitively slow. To avoid this computation burden, this paper uses "sequential equation splitting", which provides quick and unconditionally stable solutions, if the Wiener processes driving the two factors are not instantaneously correlated. When the two Wiener processes are correlated, "splitting" may sometimes provide unstable solutions and this seems the main unsolved problem of vertical MOL. Vertical MOL with sequential splitting is

very convenient to price contracts when the **factors** may "jump" and when **parameters** may "switch regime" over time.

4.1 Vertical method of lines (MOL) for PDE's with one space dimension

This section explains vertical MOL can be used to price $U(x_3, \tau)$, or more simply U, whose value is driven by just one stochastic factor x_3 and time to maturity $\tau = T - t$. To simplify notation, in this section, and only in this section, we write x to mean x_3. Dropping subscripts, in the **risk-neutral** measure x satisfies the stochastic differential equation

$$dx = \kappa_3 (\mu_3 - x) \, dt + \sigma_3 \cdot dw_3^Q \tag{10}$$

and we set $c(x) = \max(x, 0)^{q_3}$. $x \in (-\infty, \infty)$ but PDE 8 is solved in the finite region $x_1 \le x \le x_I$. We define

$$x_i = i \cdot \delta x + x_0, \quad \delta x = \frac{x_I - x_0}{I}$$

$$a_i = \kappa_3 (\mu_3 - x_i), \quad b_i = \sigma_3, \quad c_i = c(x_i)$$

$$\text{for} \quad i = 1, 2, .., I.$$

x_0 is the lower bound of the solution region. We approximate the exact solution U to PDE 8 with the MOL solution $u_i(\tau)$ or more simply u_i. $u_i(0)$ denotes u_i when $\tau = 0$, i.e. at maturity. Then PDE 8 and its conditions are approximated as

$$\frac{\partial u_i}{\partial \tau} = \frac{u_{i+1} - 2u_i + u_{i-1}}{(\delta x)^2} \frac{(a_i)^2}{2} + \frac{u_{i+1} - u_{i-1}}{2 \cdot \delta x} b_i + c_i u_i$$

$$u_i(0) = 1, \ u_{I+1} = 2u_I - u_{I-1}, \ u_0 = 2u_1 - u_2$$

for $i = 1, .., I$ giving the ordinary differential equations (ODE) system

$$\frac{\partial u_i}{\partial \tau} = u_{i-1} \cdot \mathcal{A}_i + u_i \cdot \mathcal{B}_i + u_{i+1} \cdot \mathcal{C}_i \quad \text{for } i = 2, .., I - 1 \tag{11}$$

$$\frac{\partial u_I}{\partial \tau} = u_{I-1} (\mathcal{A}_I - \mathcal{C}_I) + u_I (\mathcal{B}_I + 2\mathcal{C}_I)$$

$$\frac{\partial u_1}{\partial \tau} = u_1 (2\mathcal{A}_1 + \mathcal{B}_1) + u_2 (\mathcal{C}_1 - \mathcal{A}_1)$$

$$\mathcal{A}_i = \frac{a_i}{2(\delta x)^2} - \frac{b_i}{2 \cdot \delta x}, \ \mathcal{B}_i = c_i - \frac{a_i}{(\delta x)^2}, \ \mathcal{C}_i = \frac{a_i}{2(\delta x)^2} + \frac{b_i}{2 \cdot \delta x} \text{ for } i = 1, .., I.$$

Then we can rewrite ODE system 11 as

$$\frac{\partial \mathbf{u}}{\partial \tau} = \mathbf{M} \cdot \mathbf{u} \tag{12}$$

$$\mathbf{u}(0) = \left(u_1(0), u_2(0), .., u_I(0)\right)' = \mathbf{1}_I$$

$$\frac{\partial \mathbf{u}}{\partial \tau} = \left(\frac{\partial u_1}{\partial \tau}, \frac{\partial u_2}{\partial \tau}, .., \frac{\partial u_I}{\partial \tau}\right)'$$

$$\mathbf{u} = (u_1, u_2, .., u_I)'.$$

$\mathbf{1}_I$ is a column vector with I elements, each equal to 1. Also $\frac{\partial \mathbf{u}}{\partial \tau}$ and $\mathbf{u}$ are column vectors with I elements. M is an $I \times I$ matrix, such that

$$\mathbf{M} = \begin{bmatrix} (2\mathcal{A}_1 + \mathcal{B}_1) & (\mathcal{C}_1 - \mathcal{A}_1) & 0 & 0 & 0 & .. & 0 \\ \mathcal{A}_2 & \mathcal{B}_2 & \mathcal{C}_2 & .. & .. & .. & 0 \\ 0 & .. & .. & .. & .. & .. & .. \\ .. & .. & .. & .. & \mathcal{A}_{I-1} & \mathcal{B}_{I-1} & \mathcal{C}_{I-1} \\ 0 & .. & .. & .. & .. & (\mathcal{A}_I - \mathcal{C}_I) & (\mathcal{B}_I + 2\mathcal{C}_I) \end{bmatrix}.$$

The solution to system 11 is

$$\mathbf{u}(\tau) = \exp(\tau \cdot \mathbf{M}) \cdot \mathbf{u}(0). \tag{13}$$

The matrix exponential $\exp(\tau \cdot \mathbf{M})$ can easily be computed. For example Matlab computes the matrix exponential with the Padé approximation with scaling and squaring of Higham (2005), which proves quick and reliable.

The solution in 13 is more accurate and faster than finite difference methods that use the same grid points in the space "dimension", because finite difference methods discretise also in the "time" dimension and their computations grow with the number of time steps. Moreover discretisation in the "time" dimension may pose stability problems for some finite difference methods, but stability is not an issue for the solution in 13.

As a preparation for what follows, we also consider the case whereby pricing PDE 8 is non-homogeneous, so that

$$\frac{\partial U}{\partial \tau} = \frac{\partial^2 U}{\partial x^2} \frac{a(x)^2}{2} + \frac{\partial U}{\partial x} b(x) + c(x) U + q(x) \tag{14}$$

$$U(x,0) = 1, \quad \lim_{x \to -\infty} \frac{\partial^2 U}{\partial x^2} \to 0, \quad \lim_{x \to \infty} \frac{\partial^2 U}{\partial x^2} \to 0.$$

where $q(x)$ is a bounded function of x. Then the system of ODE's that results form applying vertical MOL becomes.

$$\frac{\partial \mathbf{u}}{\partial \tau} = \mathbf{M} \cdot \mathbf{u} + \mathbf{q} \tag{15}$$

$$\mathbf{q} = \left(q(x_1), q(x_2), .., q(x_{I-1}), q(x_I)\right)'$$

with solution

$$\mathbf{u}(\tau) = -\mathbf{M}^{-1} \cdot (\mathbf{I}_I - \exp(\tau \cdot \mathbf{M})) \cdot \mathbf{q} + \exp(\tau \cdot \mathbf{M}) \cdot \mathbf{u}(0). \tag{16}$$

This result is used below. $\mathbf{I}_I$ is the identity matrix of size $I \times I$.

4.2 PDE's with two space dimensions: vertical MOL with sequential splitting

To simplify notation, in this section, and only in this section, we write x to mean x_1 and y to mean x_2. Thus $W(x, y, \tau)$ or more simply W is a function of x, y **and** τ. Then, time dropping subscripts, in the risk-neutral measure x and y are such that

$$dx = b_x \cdot dt + a_x \cdot dw_{x,t}^Q \tag{17}$$

$$dy = b_y \cdot dt + a_y \cdot dw_{y,t}^Q \tag{18}$$

$$b_x = \kappa_1 \cdot (\mu_1 + p \cdot y - x) \tag{19}$$

$$b_y = \kappa_2 \cdot (\mu_2 - y) \tag{20}$$

$$a_x = \sigma_1, \ a_y = \sigma_2 \tag{21}$$

$$dw_{x,t}^Q \cdot dw_{y,t}^Q = \rho \cdot dt$$

$$c(x) + c(y) = \max(x, 0)^{q_1} + \max(y, 0)^{q_2}. \tag{22}$$

Then PDE 7 and its conditions can be re-written as

$$\frac{\partial W}{\partial \tau} = \frac{\partial^2 W}{\partial x \partial y} \rho a_x a_y + \frac{\partial^2 W}{\partial x^2} \frac{a_x^2}{2} + \frac{\partial W}{\partial x} b_x + \frac{\partial^2 W}{\partial y^2} \frac{a_y^2}{2} + \frac{\partial W}{\partial y} b_y + (c(x) + c(y)) W \tag{23}$$

$$W(x, y, 0) = 1, \ \lim_{x \to -\infty} \frac{\partial^2 W}{\partial x^2} \to 0, \ \lim_{x \to \infty} \frac{\partial^2 W}{\partial x^2} \to 0, \ \lim_{y \to -\infty} \frac{\partial^2 W}{\partial y^2} \to 0, \ \lim_{y \to \infty} \frac{\partial^2 W}{\partial y^2} \to 0$$

for $x, y \in (-\infty, \infty)$. However PDE 23 is solved over the finite region $[x_1, x_I] \times [y_1, y_I]$. We define

$$y_j = j \cdot \delta y + y_0, \ \ \delta y = \frac{y_I - y_0}{I}$$

$$a_{x,i,j} = \sigma_1, \ \ b_{x,i,j} = \kappa_1 \cdot (\mu_1 + p \cdot y_j - x_i), \ \ a_{y,i,j} = \sigma_2, \ \ b_{y,i,j} = \kappa_2 \cdot (\mu_2 - y_j),$$

$$c(x_i) = c_{x,i}, \ \ c(y_j) = c_{y,j}$$

for $i = 1, 2, .., I$ and $j = 1, 2, .., I$.

We now use vertical MOL to discretise the PDE in both "space" dimensions, but not in the time dimension, and then we "sequentially split" the discretised PDE. Define $\tau_k = k \cdot \delta \tau$ for $k = 0, 1, 2, .., K$. $\delta \tau$ is the size of a time step. $[0, K \cdot \delta \tau]$ is the time interval over which PDE 23 is solved. Using vertical MOL with sequential splitting, we approximate PDE 23 during the interval $[\tau_k, \tau_{k+1}]$

as

$$\frac{\partial u_{i,j}^{k+1}}{\partial \tau} = \frac{\rho a_{x,i,j} a_{y,i,j}}{2} \frac{\partial^2 u_{i,j}^{k+1}}{\partial x \partial y} + \frac{u_{i+1,j}^{k+1} - 2u_{i,j}^{k+1} + u_{i-1,j}^{k+1}}{(\delta x)^2} \frac{(a_{x,i,j})^2}{2} + \frac{u_{i+1,j}^{k+1} - u_{i-1,j}^{k+1}}{2 \cdot \delta x} b_{x,i,j} + c_{x,i} u_{i,j}^{k+1}$$

$$\frac{\partial v_{i,j}^{k+1}}{\partial \tau} = \frac{\rho a_{x,i,j} a_{y,i,j}}{2} \frac{\partial^2 v_{i,j}^{k+1}}{\partial x \partial y} + \frac{v_{i,j+1}^{k+1} - 2v_{i,j}^{k+1} + v_{i,j-1}^{k+1}}{(\delta y)^2} \frac{(a_{y,i,j})^2}{2} + \frac{v_{i,j+1}^{k+1} - v_{i,j-1}^{k+1}}{2 \cdot \delta y} b_{y,i,j} + c_{y,j} v_{i,j}^{k+1}$$

$$\frac{\partial^2 u_{i,j}^{k+1}}{\partial x \partial y} \backsimeq \frac{u_{i+1,j+1}^{k+1}(\tau_k) - u_{i-1,j+1}^{k+1}(\tau_k) - u_{i+1,j-1}^{k+1}(\tau_k) + u_{i-1,j-1}^{k+1}(\tau_k)}{4\delta x \delta y}$$

$$\frac{\partial^2 v_{i,j}^{k+1}}{\partial x \partial y} \backsimeq \frac{v_{i+1,j+1}^{k+1}(\tau_k) - v_{i-1,j+1}^{k+1}(\tau_k) - v_{i+1,j-1}^{k+1}(\tau_k) + v_{i-1,j-1}^{k+1}(\tau_k)}{4\delta x \delta y}$$

$$u_{i,j}^{k+1}(\tau_k) = v_{i,j}^k(\tau_k)$$

$$v_{i,j}^{k+1}(\tau_k) = u_{i,j}^{k+1}(\tau_{k+1})$$

for $i = 1, 2, .., I$ and $j = 1, 2, .., I$.

We can rewrite this system of equations as

$$\frac{\partial \mathbf{u}_j^{k+1}}{\partial \tau} = \mathbf{M}_u(j) \cdot \mathbf{u}_j^{k+1} + \mathbf{q}_j^{k+1}(u) \tag{24}$$

$$\mathbf{u}_j^{k+1}(\tau_k) = \mathbf{v}_j^k(\tau_k)$$

$$\left(\frac{\partial \mathbf{v}^{k+1,i}}{\partial \tau}\right)' = \mathbf{M}_v(i) \cdot \left(\mathbf{v}^{k+1,i}\right)' + \left(\mathbf{q}^{k+1,i}(v)\right)'$$

$$\mathbf{v}^{k+1,i}(\tau_k) = \mathbf{u}^{k+1,i}(\tau_{k+1})$$

for $i = 1, 2, .., I$, $j = 1, 2, .., I$ and $k = 0, 1, 2, .., K$ with

$$\mathbf{u}_j^{k+1} = \begin{bmatrix} u_{1,j}^{k+1} \\ .. \\ u_{I,j}^{k+1} \end{bmatrix} \quad \mathbf{v}_j^{k+1} = \begin{bmatrix} v_{1,j}^{k+1} \\ .. \\ v_{I,j}^{k+1} \end{bmatrix} \quad \frac{\partial \mathbf{u}_j^{k+1}}{\partial \tau} = \begin{bmatrix} \frac{\partial u_{1,j}^{k+1}}{\partial \tau} \\ .. \\ \frac{\partial u_{I,j}^{k+1}}{\partial \tau} \end{bmatrix} \quad \frac{\partial \mathbf{v}_j^{k+1}}{\partial \tau} = \begin{bmatrix} \frac{\partial v_{1,j}^{k+1}}{\partial \tau} \\ .. \\ \frac{\partial v_{I,j}^{k+1}}{\partial \tau} \end{bmatrix}$$

$$\mathbf{M}_u(j) = \begin{bmatrix} 2\mathcal{A}_{x,1,j} + \mathcal{B}_{x,1,j} & \mathcal{C}_{x,1,j} - \mathcal{A}_{x,1,j} & 0 & .. & 0 & 0 & 0 \\ \mathcal{A}_{x,2,j} & \mathcal{B}_{x,2,j} & \mathcal{C}_{x,2,j} & .. & 0 & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & \mathcal{A}_{x,(I-1),j} & \mathcal{B}_{x,(I-1),j} & \mathcal{C}_{x,(I-1),j} \\ 0 & 0 & 0 & .. & 0 & \mathcal{A}_{x,I,j} - \mathcal{C}_{x,I,j} & \mathcal{B}_{x,I,j} + 2\mathcal{C}_{x,I,j} \end{bmatrix}$$

$$\mathbf{M}_v(i) = \begin{bmatrix} 2\mathcal{A}_{y,i,1} + \mathcal{B}_{y,i,1} & \mathcal{C}_{y,i,1} - \mathcal{A}_{y,i,1} & 0 & .. & 0 & 0 & 0 \\ \mathcal{A}_{y,i,2} & \mathcal{B}_{y,i,2} & \mathcal{C}_{y,i,2} & .. & 0 & 0 & 0 \\ .. & .. & .. & .. & .. & .. & .. \\ 0 & 0 & 0 & .. & \mathcal{A}_{y,i,(I-1)} & \mathcal{B}_{y,i,(I-1)} & \mathcal{C}_{y,i,(I-1)} \\ 0 & 0 & 0 & .. & 0 & \mathcal{A}_{y,i,I} - \mathcal{C}_{y,i,I} & \mathcal{B}_{y,i,I} + 2\mathcal{C}_{y,i,I} \end{bmatrix}$$

$$\mathcal{A}_{x,i,j} = \tfrac{1}{2}\left(\left(\tfrac{a_{x,i,j}}{\delta x}\right)^2 - \tfrac{b_{x,i,j}}{\delta x}\right), \quad \mathcal{A}_{y,i,j} = \tfrac{1}{2}\left(\left(\tfrac{a_{y,i,j}}{\delta y}\right)^2 - \tfrac{b_{y,i,j}}{\delta y}\right),$$

$$\mathcal{B}_{x,i,j} = c_{x,i} - \left(\tfrac{a_{x,i,j}}{\delta x}\right)^2, \qquad \mathcal{B}_{y,i,j} = c_{y,j} - \left(\tfrac{a_{y,i,j}}{\delta y}\right)^2,$$

$$\mathcal{C}_{x,i,j} = \tfrac{1}{2}\left(\left(\tfrac{a_{x,i,j}}{\delta x}\right)^2 + \tfrac{b_{x,i,j}}{\delta x}\right), \quad \mathcal{C}_{y,i,j} = \tfrac{1}{2}\left(\left(\tfrac{a_{y,i,j}}{\delta y}\right)^2 + \tfrac{b_{y,i,j}}{\delta y}\right),$$

$$\mathbf{Q}^{k+1}(u) = [..,\mathbf{q}_j^{k+1}(u),..] \quad \mathbf{Q}^{k+1}(v) = \begin{bmatrix} .. \\ \mathbf{q}^{k+1,i}(v) \\ .. \end{bmatrix} \quad \mathbf{U}^{k+1} = [..,\mathbf{u}_j^{k+1},..] \quad \mathbf{V}^{k+1} = \begin{bmatrix} .. \\ \mathbf{v}^{k+1,i} \\ .. \end{bmatrix}$$

$$\mathbf{q}_j^{k+1}(u) = \begin{bmatrix} 0 \\ \frac{\rho a_{x,2,j} a_{y,2,j}}{2} \frac{v_{3,j+1}^{k+1}(\tau_k) - v_{1,j+1}^{k+1}(\tau_k) - v_{3,j-1}^{k+1}(\tau_k) + v_{1,j-1}^{k+1}(\tau_k)}{4\delta x \delta y} \\ .. \\ \frac{\rho a_{x,i,j} a_{y,i,j}}{2} \frac{v_{i+1,j+1}^{k+1}(\tau_k) - v_{i-1,j+1}^{k+1}(\tau_k) - v_{i+1,j-1}^{k+1}(\tau_k) + v_{i-1,j-1}^{k+1}(\tau_k)}{4\delta x \delta y} \\ .. \\ \frac{\rho a_{x,I-1,j} a_{y,I-1,j}}{2} \frac{v_{I,j+1}^{k+1}(\tau_k) - v_{I-2,j+1}^{k+1}(\tau_k) - v_{I,j-1}^{k+1}(\tau_k) + v_{I-2,j-1}^{k+1}(\tau_k)}{4\delta x \delta y} \\ 0 \end{bmatrix}$$

$$\mathbf{q}^{k+1,i}(v) = \begin{bmatrix} 0 \\ \frac{\rho a_{x,i,2} a_{y,i,2}}{2} \frac{u_{i+1,3}^{k+1}(\tau_{k+1}) - u_{i-1,3}^{k+1}(\tau_{k+1}) - u_{i+1,1}^{k+1}(\tau_{k+1}) + u_{i-1,1}^{k+1}(\tau_{k+1})}{4\delta x \delta y} \\ .. \\ \frac{\rho a_{x,i,j} a_{y,i,j}}{2} \frac{u_{i+1,j+1}^{k+1}(\tau_{k+1}) - u_{i-1,j+1}^{k+1}(\tau_{k+1}) - u_{i+1,j-1}^{k+1}(\tau_{k+1}) + u_{i-1,j-1}^{k+1}(\tau_{k+1})}{4\delta x \delta y} \\ .. \\ \frac{\rho a_{x,i,I-1} a_{y,i,I-1}}{2} \frac{u_{i+1,I}^{k+1}(\tau_{k+1}) - u_{i-1,I}^{k+1}(\tau_{k+1}) - u_{i+1,I-2}^{k+1}(\tau_{k+1}) + u_{i-1,I-2}^{k+1}(\tau_{k+1})}{4\delta x \delta y} \\ 0 \end{bmatrix}.$$

$\mathbf{u}_j^{k+1}$ is the j-th column of $\mathbf{U}^{k+1}$. $\mathbf{v}^{k+1,i}$ is the i-th row of $\mathbf{V}^{k+1}$. The cross derivative term $\frac{\partial^2 V}{\partial x \partial y}\rho a_x a_y$ in equation 23 gives rise to the non-homogeneous terms $\mathbf{Q}^{k+1}(u)$ and $\mathbf{Q}^{k+1}(v)$ in ODE system 24. The solution to system 24 is

$$\mathbf{u}_j^{k+1} = -\mathbf{M}_u(j)^{-1}\left(\mathbf{I}_I - \exp\left(\delta\tau \cdot \mathbf{M}_u(j)\right)\right) \cdot \mathbf{q}_j^{k+1}(u) + \exp\left(\delta\tau \cdot \mathbf{M}_u(j)\right) \cdot \mathbf{v}_j^{k+1}(\tau_k) \tag{25}$$

$$\mathbf{v}^{k+1,i}(\tau_{k+1})' = -\mathbf{M}_v(i)^{-1}\left(\mathbf{I}_I - \exp\left(\delta\tau \cdot \mathbf{M}_v(i)\right)\right) \cdot \left(\mathbf{q}^{k+1,i}(v)\right)' + \exp\left(\delta\tau \cdot \mathbf{M}_v(i)\right) \cdot \left(\mathbf{u}^{k+1,i}\right)'$$

for $i = 1, 2, .., I$ and $j = 1, 2, .., I$. $\mathbf{I}_I$ is the $I \times I$ identity matrix. $\mathbf{q}_j^{k+1}(u)$ is known from the previous time step. We can assume that the cross derivative equals 0 on the boundaries, i.e. $\mathbf{q}_1^{k+1}(u) = \mathbf{q}_I^{k+1}(u) = [0, .., 0]'$ and

$\mathbf{q}^{k+1,1}\left(v\right) = \mathbf{q}^{k+1,I}\left(v\right) = \left[0,..,0\right]$. If $\rho = 0$, then $\mathbf{Q}^{k+1}\left(u\right) = \mathbf{Q}^{k+1}\left(v\right) = \mathbf{0}_{I \times I}$, where $\mathbf{0}_{I \times I}$ is an $I \times I$ matrix whose elements are all equal to 0, and solution 25 is unconditionally stable and consistent, which is a known property of sequential splitting (e.g. see Farago (2003) at page 8). If $\rho \neq 0$, solution 25 may be unstable, unless $\delta\tau \to 0$. The attraction of solution 25 is that the main computation it involves is a set of matrix exponentials, which can be computed very quickly and reliably.

$\mathbf{M}_u\left(j\right)$ depends on j and $\mathbf{M}_v\left(i\right)$ on i, but solution 25 is much quicker to compute when $\mathbf{M}_u$ does not depend on j and $\mathbf{M}_v$ does not depend on i. This is the case when in PDE 23 a_x and b_x only depend on x while a_y and b_y only depend on y. In this case computing solution 25 can become some 25 times faster, as shown by unreported simulations, and even valuations in the presence of three factors remain quite affordable. Other things equal, when $\rho = 0$ solution 25 reduces to

$$\mathbf{u}_j^{k+1}\left(\tau_{k+1}\right) = \exp\left(\delta\tau \cdot \mathbf{M}_u\left(j\right)\right) \cdot \mathbf{v}_j^k\left(\tau_k\right)$$
$$\mathbf{v}^{k+1,i}\left(\tau_{k+1}\right)' = \exp\left(\delta\tau \cdot \mathbf{M}_v\left(i\right)\right) \cdot \mathbf{u}^{k+1,i}\left(\tau_{k+1}\right)'.$$

When $\rho = 0$, Strang splitting may also be used rather than the sequential splitting just illustrated.

4.3 Comparison between vertical MOL "with sequential splitting" and the implicit finite difference method

Vertical MOL competes with implicit finite difference methods to solve PDE's with two stochastic factors. Fully implicit finite difference methods discretise PDE's also in the time dimension, are unconditionally stable and require solving a sparse system of linear equations at every time step. The system of linear equations is typically solved through an iterative numerical method such as successive over relaxation (SOR) or conjugate gradient (CG) or biconjugate gradient stabilised (BCGST) or quasi minimum residual (QMR), but such numerical methods may not always converge to the solution, i.e. they may "break down". Alternatively the mentioned system of linear equations can be solved using an analytic method such as Gaussian elimination, but according to Tavella and Randall (2000, page 98) when the pricing PDE involves multiple space dimensions Gaussian elimination "is never feasible in practice". With two "space" dimensions we could use the PDE splitting method of Yanenko (1971), which is a finite difference method similar to 24 but that discretises PDE 23 also in the time dimension. Yanenko's method involves solving a set of systems of linear equations where each system only involves just one space dimension and therefore could be solved through Gaussian elimination or other analytic methods. However Yanenko's method is guaranteed to be stable only when $b_x = b_y = c\left(x\right) + c\left(y\right) = 0$ in PDE 23. Instead vertical MOL "with sequential splitting" is guaranteed to be stable whenever $\rho = 0$.

Table 1 reports results from using the fully implicit finite difference method with SOR and vertical MOL "with sequential splitting". What is computed

is the price of a discount bond with face value 1 and time to maturity of one year. The computation solves for $W(x, y, \tau)$ assuming that $b_x = 0.5 \cdot (y - x)$, $b_y = 0.5 \cdot (0.1 - y)$, $c(x) = \max(x, 0)$, $c(y) = 0$, $\sigma_x = \sigma_y = 0.1$, $\rho = 0$, $\tau = 1$, 100 steps in both the x and y dimensions. x tends to its long term mean y, which is itself stochastic as y follows an Ornstein-Uhlenbeck process. y and x vary between -1 and 1. The two panels assume 10 and 100 time steps per year. MOL is clearly faster than the implicit finite difference method.

[TABLE 1 ABOUT HERE]

5 The quadratic models

For the quadratic and affine Gaussian models we adopt a discrete time setting. The choice of discrete time, as explained below, implies fewer restrictions on the parameters of quadratic models. $P_{n,t}$ is the price at time t of a default-free discount bond with n time periods to maturity; each time period is of length Δ, thus the bond matures at time $t + n \cdot \Delta$. r_t is the time t continuously compounded default-free interest rate during the time interval $[t, t + \Delta]$, such that

$$P_{1,t} = e^{-\Delta \cdot r_t}, \quad r_t = \frac{-\ln P_{1,t}}{\Delta}$$

with $\Delta = \frac{1}{261}$ as before. The no-arbitrage risk-neutral valuation equation is

$$P_{n,t} = E_t^Q \left[e^{-\Delta \cdot r_t} \cdot P_{n-1,t+1} \right] \tag{26}$$

where $E_t^Q [..]$ denotes conditional expectation at time t under the risk-neutral measure Q.

5.1 Quadratic model Q3.3

Here we introduce quadratic model Q3.3, which is a three factor model **where**

$$r_t = \alpha + \mathbf{x}_t' \mathbf{\Psi} \mathbf{x}_t \tag{27}$$

$$\mathbf{\Psi} = \mathbf{I}_3 \tag{28}$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \phi \boldsymbol{\mu} - \phi \mathbf{x}_t + \mathbf{\Sigma} \boldsymbol{\xi}_{t+1}^Q \tag{29}$$

$$\boldsymbol{\xi}_{t+1}^Q \backsim N\left(\mathbf{0}_{3 \times 1}, \mathbf{I}_3 \right). \tag{30}$$

$\mathbf{I}_3$ is the 3×3 identity matrix. $\boldsymbol{\mu}, \phi, \mathbf{\Sigma}$ are made up of constant parameters. $\phi, \mathbf{\Sigma}$ are 3×3 square matrixes. $\boldsymbol{\mu}$ is a 3×1 column vector. $\mathbf{x}_t = (x_{1,t}, x_{2,t}, x_{3,t})'$ is a vector of latent stochastic factors. α is a scalar parameter. Equation 29 states that $\mathbf{x}_t$ follows a Gaussian auto-regressive process. Equation 30 states that, in the risk-neutral measure Q, the random terms $\boldsymbol{\xi}_{t+1}^Q$ are normally distributed with mean $\mathbf{0}_{3 \times 1}$ and covariance $\mathbf{I}_3$. $\mathbf{0}_{3 \times 1}$ is a column vector of zeroes with 3

entries. $N(..,..)$ denotes the multivariate normal density. Under these assumptions, following Realdon (2006) we find that the time t value of the default-free zero coupon bond with n periods to maturity is

$$P_{n,t} = e^{A_n + B'_n \mathbf{x}_t + \mathbf{x}'_t C_n \mathbf{x}_t}$$

$$A_n = -\alpha \cdot \Delta + A_{n-1} + \left(B'_{n-1} + \boldsymbol{\mu}' C_{n-1}\right) \boldsymbol{\phi}\boldsymbol{\mu} + \ln \frac{|\gamma|}{abs\,|\boldsymbol{\Sigma}|} + \sum_{i=1}^{3} \frac{1}{2} \left(\left(B'_{n-1} + 2\left(\boldsymbol{\phi}\boldsymbol{\mu}\right)' C_{n-1}\right) \gamma_i\right)^2$$

(31)

$$B'_n = \left(B'_{n-1} + 2\boldsymbol{\mu}' C_{n-1}\right)\left(\mathbf{I}_3 - \boldsymbol{\phi}\right) + \sum_{i=1}^{3} 2\left(B'_{n-1} + 2\left(\boldsymbol{\phi}\boldsymbol{\mu}\right)' C_{n-1}\right)\gamma_i \gamma'_i C_{n-1}\left(\mathbf{I}_3 - \boldsymbol{\phi}\right)$$

(32)

$$C_n = -\Delta \cdot \boldsymbol{\Psi} + \left(\mathbf{I}_3 - \boldsymbol{\phi}\right)' C_{n-1}\left(\mathbf{I}_3 - \boldsymbol{\phi} + 2\sum_{i=1}^{3}\gamma_i \gamma'_i C_{n-1}\left(\mathbf{I}_3 - \boldsymbol{\phi}\right)\right) \quad (33)$$

where γ_i is the i-th column of the 3×3 matrix

$$\gamma = \left(\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\right)^{-1} - 2C_{n-1}\right)^{-1/2}.$$

The terminal conditions are

$$A_0 = 0, B_0 = \mathbf{0}_{3\times1}, C_0 = \mathbf{0}_{3\times3}.$$

$\mathbf{0}_{3\times3}$ is the 3×3 matrix made of zeroes. A_n, B_n, C_n only depend on n. The time t one-period yield $y_{1,t}$ is

$$y_{1,t} = \frac{-\ln P_{1,t}}{\Delta} = -A_1 - B'_1 \cdot \mathbf{x}_t - \mathbf{x}'_t C_1 \mathbf{x}_t = \alpha + \mathbf{0}'_{3\times1} \cdot \mathbf{x}_t + \mathbf{x}'_t \boldsymbol{\Psi} \mathbf{x}_t = r_t \quad (34)$$

since $A_1 = -\alpha \cdot \Delta$, $B_1 = \mathbf{0}_{3\times1}$ and $C_1 = -\boldsymbol{\Psi} \cdot \Delta$. The process of $\mathbf{x}_t$ under the physical measure is

$$\mathbf{x}_{t+1} = \left(\mathbf{I}_3 - \boldsymbol{\phi}^*\right)\mathbf{x}_t + \boldsymbol{\phi}^*\boldsymbol{\mu}^* + \boldsymbol{\Sigma}\boldsymbol{\xi}^*_{t+1} \quad (35)$$

where again $\boldsymbol{\xi}^*_{t+1} \backsim N\left(\mathbf{0}_{3\times1}, \mathbf{I}_3\right)$. $\boldsymbol{\Sigma}$ **is defined above and the parameters** $\rho_{12}, \rho_{13}, \rho_{32}$ **that appear in** Σ **now are arbitrary constants.** $\sigma_1\sqrt{\Delta}, \sigma_2\sqrt{\Delta}, \sigma_3\sqrt{\Delta}$ are respectively the conditional standard deviations of $x_{1,t+1}, x_{2,t+1}, x_{3,t+1}$. Assuming that $\boldsymbol{\Sigma}$ be lower triangular entails no loss in generality.

5.2 Parameter identification conditions

As the factors $\mathbf{x}_t$ are not observable, we need to impose some parameter restrictions in order to be able to estimate the parameters. The parameter identification conditions for quadratic term structure models in continuous time, i.e. when $\Delta \to 0$, are provided by Ahn-Dittmar-Gallant (2002) and are:

- that $\boldsymbol{\Psi}$ and $\mathbf{C}_n$ be symmetric;
- the normalisation $\boldsymbol{\Psi} = \mathbf{I}_3$;
- that $\boldsymbol{\phi\mu} \geq \mathbf{0}_{3\times 1}$;
- that Σ be diagonal (triangular) and $\boldsymbol{\phi}$ be triangular (diagonal).

However an Appendix shows that for discrete time quadratic models the parameter identification conditions are the same as in Ahn-Dittmar-Gallant (2002) except that *both* $\boldsymbol{\phi}$ and $\boldsymbol{\Sigma}$ can be triangular matrixes at the same time, **which is an advantage of the discrete time version of the quadratic model over the continuous time version of Ahn-Dittmar-Gallant (2002).** The Appendix shows that this result holds even when $\boldsymbol{\Psi} \neq \mathbf{I}_3$ as long as $\boldsymbol{\Psi}$ is symmetric and has all diagonal entries equal to 1. As in Ahn-Dittmar-Gallant (2002) we concentrate on the case whereby the non-diagonal entries of $\boldsymbol{\Psi}$ are all equal to 0.

5.3 Specification of model Q3.1

Another key difference between this paper and Ahn-Dittmar-Gallant (2002) is that **$\boldsymbol{\Psi}$ may also be** a diagonal matrix with diagonal entries equal to 0 or 1 and therefore not all equal to 1. When some of the diagonal entries of $\boldsymbol{\Psi}$ are equal to 0, we have a "separation" between the factors that drive the short rate r_t, whose corresponding diagonal entries are equal to 1, and the factors that drive the central tendency of the short rate, whose corresponding diagonal entries are equal to 0. This "separation" is supported by the empirical analysis of Japanese yields by Kim and Singleton (2012). This "separation" implies that long term yields can be relatively high even when the short rate is extremely close to 0, and that short and long term yields move quite independently. This "separation" seems of interest especially when yields are very low, as has been be case in most Western countries since 2008. However this "separation" entails additional parameter identification conditions on the $\boldsymbol{\phi}$ matrix as shown in the Appendixes: for each 0 entry in the diagonal of $\boldsymbol{\Psi}$ we need to impose one restriction on the parameters of $\boldsymbol{\phi}$, while *both* $\boldsymbol{\Sigma}$ and $\boldsymbol{\phi}$ can be triangular matrixes at the same time.

In particular we test a subfamily of quadratic models, which we name Q3.1, which are three factor models in which only one factor drives r_t and the other factors drive the central tendency of r_t. We consider three such model variants,

namely Q3.1.1, Q3.1.2 and Q3.1.0. Q3.1.1 is such that

$$
r_t = \begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix}' \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix}
$$

$$
\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \begin{pmatrix} \phi_1 & 0 & 0 \\ -\phi_2 & \phi_2 & 0 \\ 0 & -\phi_3 & \phi_3 \end{pmatrix} \Delta \left(\begin{pmatrix} 0 \\ 0 \\ \mu_3 \end{pmatrix} - \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} \right) + \boldsymbol{\Sigma} \cdot \boldsymbol{\xi}_{t+1}^Q
$$

$$
\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \begin{pmatrix} \phi_1^* & 0 & 0 \\ -\phi_2^* & \phi_2^* & 0 \\ 0 & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta \left(\begin{pmatrix} 0 \\ 0 \\ \mu_3^* \end{pmatrix} - \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} \right) + \boldsymbol{\Sigma} \cdot \boldsymbol{\xi}_{t+1}^*
$$

so that

$$
\boldsymbol{\Psi} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{x}_{t+1} = \begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \\ \mu_3 \end{pmatrix},
$$

$$
\boldsymbol{\phi} = \begin{pmatrix} \phi_1 & 0 & 0 \\ -\phi_2 & \phi_2 & 0 \\ 0 & -\phi_3 & \phi_3 \end{pmatrix} \Delta, \quad \boldsymbol{\phi}^* = \begin{pmatrix} \phi_1^* & 0 & 0 \\ -\phi_2^* & \phi_2^* & 0 \\ 0 & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta
$$

$\phi_1, \phi_2, \phi_3, \phi_1^*, \phi_2^*, \phi_3^*$ are three scalar parameters. Q3.1.2 is the same as Q3.1.1 except for

$$
\boldsymbol{\phi} = \begin{pmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \\ -\phi_3 & -\phi_3 & \phi_3 \end{pmatrix} \Delta, \quad \boldsymbol{\phi}^* = \begin{pmatrix} \phi_1^* & 0 & 0 \\ 0 & \phi_2^* & 0 \\ -\phi_3^* & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta.
$$

Q3.1.0 is the same as Q3.1.1 except for

$$
\boldsymbol{\phi} = \begin{pmatrix} \phi_1 & 0 & 0 \\ -\phi_2 & \phi_2 & 0 \\ -\phi_3 & -\phi_3 & \phi_3 \end{pmatrix} \Delta, \quad \boldsymbol{\phi}^* = \begin{pmatrix} \phi_1^* & 0 & 0 \\ -\phi_2^* & \phi_2^* & 0 \\ -\phi_3^* & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta.
$$

The Appendix proves that the parameters in Q3.1.0, Q3.1.1, Q3.1.2 are identifiable. **Specifications Q3.1.0, Q3.1.1 and Q3.1.2 are new, since they are not included in previous specifications of quadratic models appeared in past literature, and are maximally flexible under the normalisation $r_t = x_{3,t}^2$. Instead Ahn-Dittmar-Gallant (2002) focused on the normalisation $r_t = x_{1,t}^2 + x_{2,t}^2 + x_{3,t}^2$.**

6 The affine Gaussian models

We also test affine Gaussian models where

$$r_t = \boldsymbol{\beta}' \cdot \mathbf{x}_t \tag{36}$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \boldsymbol{\eta} - \phi \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\xi}_{t+1}^Q + \mathbf{j}_u \cdot d\Pi(\lambda) \tag{37}$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \boldsymbol{\eta}^* - \phi^* \mathbf{x}_t + \boldsymbol{\Sigma} \boldsymbol{\xi}_{t+1}^* + \mathbf{j}_u \cdot d\Pi(\lambda^*) \tag{38}$$

$$\boldsymbol{\xi}_{t+1}^Q \backsim N(\mathbf{0}_{3\times 1}, \mathbf{I}_3), \quad \boldsymbol{\xi}_{t+1}^* \backsim N(\mathbf{0}_{3\times 1}, \mathbf{I}_3) \tag{39}$$

$\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\eta}^*, \mathbf{j}_u$ **are** 3×1 **vectors of constants.** $\mathbf{j}_u$ **is the jump size of** $\mathbf{x}_{t+1}$ **such that** $\mathbf{j}_u \sim N(\mathbf{0}_{3\times 1}, \mathbf{SS})$, $d\Pi(\lambda)$ **is the increment of a Poisson process with probability** $\left(1 - e^{-\Delta\lambda}\right)$ **in the risk-neutral measure and** $\mathbf{S} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_j \end{bmatrix}$.

Only one jump is possible during the period Δ. **In the only model with jumps we consider below,** $r_t = x_{3,t}$ **while only factor** $x_{3,t}$ **may jump, with jump size distributed according to a Gaussian density with mean** 0 **and standard deviation** σ_j. **Jump event risk is priced, while for simplicity jump size risk is not priced.** $d\Pi(\lambda^*)$ **is the increment of a Poisson process with probability** $\left(1 - e^{-\Delta\lambda^*}\right)$ **in the real measure, where** $\lambda^* \neq \lambda$. **Since the model is in discrete time and the tests rely on daily data, jumps can only take place at the end of each trading day, rather than during the day, and this seems an acceptable approximation. For such affine model discount bond prices can easily be computed, as shown below, and no ordinary differential equation needs solving numerically, unlike in the continuous time model of Baz and Das (1996). Possible jumps entail that the conditional probability density of the factors is a mixture of two Gaussian densities. Including jumps gives the affine Gaussian model one more chance to compete with the quadratic and Black models. Under these assumptions**

$$P_{n,t} = e^{\mathbb{A}_n + \mathbb{B}_n' \mathbf{x}_t}$$

$$\mathbb{A}_n = \mathbb{A}_{n-1} + \mathbb{B}_{n-1}' \cdot \boldsymbol{\eta} + \frac{1}{2} \cdot \mathbb{B}_{n-1}' \cdot \boldsymbol{\Sigma}\boldsymbol{\Sigma}' \cdot \mathbb{B}_{n-1} + \ln\left(e^{-\Delta\lambda} + \left(1 - e^{-\Delta\lambda}\right) \cdot e^{\frac{1}{2} \cdot \mathbb{B}_{n-1}' \cdot \mathbf{SS}' \cdot \mathbb{B}_{n-1}}\right) \tag{40}$$

$$\mathbb{B}_n = -\boldsymbol{\beta} \cdot \Delta + (\mathbf{I}_3 - \phi) \cdot \mathbb{B}_{n-1} \tag{41}$$

$$\mathbb{A}_0 = 0, \quad \mathbb{B}_0 = \mathbf{0}_{3\times 1}. \tag{42}$$

We consider three specifications of this affine model **where** $\lambda^* = \lambda = 0$, namely A3.1.1, A3.1.2 and A3.1.0 that correspond to and can be directly compared with Q3.1.1, Q3.1.2 and Q3.1.0 respectively. All these affine and quadratic models have the common feature that one factor drives r_t and the other factors drive

the central tendency of r_t. By model A3.1.1 we mean

$$r_t = \begin{pmatrix} 0, & 0, & 1 \end{pmatrix} \begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix}$$

$$\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \begin{pmatrix} \phi_1 \cdot \mu_1 \\ 0 \\ 0 \end{pmatrix} \Delta - \begin{pmatrix} \phi_1 & 0 & 0 \\ -\phi_2 & \phi_2 & 0 \\ 0 & -\phi_3 & \phi_3 \end{pmatrix} \Delta \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \boldsymbol{\Sigma} \cdot \boldsymbol{\xi}_{t+1}^Q$$

$$\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \begin{pmatrix} \phi_1^* \cdot \mu_1^* \\ 0 \\ 0 \end{pmatrix} \Delta - \begin{pmatrix} \phi_1^* & 0 & 0 \\ -\phi_2^* & \phi_2^* & 0 \\ 0 & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \boldsymbol{\Sigma} \cdot \boldsymbol{\xi}_{t+1}^*$$

so that

$$\boldsymbol{\beta} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} \phi_1 \cdot \mu_1 \\ 0 \\ 0 \end{pmatrix} \Delta, \quad \boldsymbol{\phi} = \begin{pmatrix} \phi_1 & 0 & 0 \\ -\phi_2 & \phi_2 & 0 \\ 0 & -\phi_3 & \phi_3 \end{pmatrix} \Delta, \quad \boldsymbol{\phi}^* = \begin{pmatrix} \phi_1^* & 0 & 0 \\ -\phi_2^* & \phi_2^* & 0 \\ 0 & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta.$$

By model A3.1.2 we mean the same model as A3.1.1 but for

$$\boldsymbol{\phi} = \begin{pmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \\ -\phi_2 & -\phi_3 & \phi_3 \end{pmatrix} \Delta, \quad \boldsymbol{\phi}^* = \begin{pmatrix} \phi_1^* & 0 & 0 \\ 0 & \phi_2^* & 0 \\ -\phi_2^* & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta.$$

By model A3.1.0 we mean the same model as A3.1.1 but for

$$\boldsymbol{\phi} = \begin{pmatrix} \phi_1 & 0 & 0 \\ -\phi_2 & \phi_2 & 0 \\ -\phi_3 & -\phi_3 & \phi_3 \end{pmatrix} \Delta, \quad \boldsymbol{\phi}^* = \begin{pmatrix} \phi_1^* & 0 & 0 \\ -\phi_2^* & \phi_2^* & 0 \\ -\phi_3^* & -\phi_3^* & \phi_3^* \end{pmatrix} \Delta.$$

All parameters in models A3.1.1, A3.1.2 and A3.1.0 are identifiable. We also test model A3.3, which is the same as A3.1.1 except for $\boldsymbol{\beta} = \begin{pmatrix} 1, & 1, & 1 \end{pmatrix}'$. Model A3.3 is directly comparable with model Q3.3. **Finally we also test model AJ3.1.0, which is the same as A3.1.0 except for $\lambda^* \neq \lambda \neq 0$.**

6.1 Comparison with Cheng and Scaillet (2007)

The quadratic and affine models of this paper are in discrete time, but as the time step tends to zero, they converge to continuous time models that are special cases of the linear-quadratic jump-diffusion pricing model of Cheng and Scaillet (2007). The quadratic and affine models of this paper correspond respectively to the three factor models $\mathbb{LQ}_0^3(3)$ and $\mathbb{LQ}_0^0(3)$ according to the notation in Cheng and Scaillet. For these models no admissibility conditions are needed. As Cheng and Scaillet (2007) stated at page 583, the specification analysis of $\mathbb{LQ}_0^3(3)$ had already been studied in the quadratic Gaussian literature. Indeed Ahn-Dittmar-Gallant (2002) provided identification

18

conditions for continuous time quadratic models. The specification of $\mathbb{LQ}_0^0(3)$ was studied in Joslin, Singleton and Zhu (2011) and the above specification of the affine model follows Joslin, Singleton and Zhu (2011) with minor changes.

7   Empirical tests

We estimate and test the models illustrated above through Quasi Maximum Likelihood Extended Kalman filter (QML-EKF) estimation as well as through Quasi Maximum likelihood (QML) **and** maximum likelihood (ML) estimation. QML requires "inferring" the latent factors from the observed yields for any observation date. This is problematic to do for the extended Black model, but can be done quite easily for the three factor affine and quadratic models, if we assume that some of the yields are observed without error. QML-EKF estimation assumes that all yields are observed with errors. For the EKF we assume that the starting values for the latent factors are normally distributed as $N\left(\boldsymbol{\mu}^*, Cov\left(\mathbf{x}\right)\right)$ for quadratic models, as $N\left(\left(\mu_1^*, \mu_2^*, \mu_3^*\right)', Cov\left(\mathbf{x}\right)\right)$ for the Black model and as $N\left(\left(\mu_1^*, 0, 0\right)', Cov\left(\mathbf{x}\right)\right)$ for the affine models. $Cov\left(\mathbf{x}\right)$ is the unconditional covariance matrix of $\mathbf{x}$, which in vec form is

$$vec\left(Cov(\mathbf{x})\right) = \left(\mathbf{I}_{m^2} - \left(\mathbf{I}_m - \boldsymbol{\phi}\right) \otimes \left(\mathbf{I}_m - \boldsymbol{\phi}\right)\right)^{-1} \cdot vec\left(\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\right)$$

$\mathbf{I}_{m^2}$ is the identity matrix of size $m^2 \times m^2$. $\mathbf{I}_m$ has similar meaning. $m = 3$ is the number of factors.

7.1   Quasi Maximum Likelihood (QML) estimation

We also estimate quadratic models using QML, whereas for affine models we employ maximum likelihood estimation. QML poses the problem of inferring the latent factors on any observation date. We solve such problem as explained below. We employ the following notation:

   - $\widehat{\mathbf{x}}_t$ is the forecast of $\mathbf{x}_t$ conditional on information at time $t - 1$;

   - $E_{t-1}\left[..\right]$ is the expectation operator conditional on time $t - 1$ information under the physical probability measure;

   - $\mathbf{y}_t = \left(y_{1,t}, y_{2,t}, .., y_{30,t}\right)'$ are the discount bond yields observed in the market at time t for maturities of 1, 2, .., 30 years;

   - $\mathbf{z}\left(\mathbf{x}_t\right) = \left(z_1\left(\mathbf{x}_t\right), z_2\left(\mathbf{x}_t\right), .., z_{30}\left(\mathbf{x}_t\right)\right)'$ is the time t vector of discount bond yields computed using the model;

   - $\boldsymbol{\varepsilon}_t$ is the vector of observation errors at time t, which is normally distributed such that $\boldsymbol{\varepsilon}_t \sim N\left(\mathbf{0}_{30 \times 1}, \mathbf{H}\right)$; $\mathbf{0}_{30 \times 1}$ is a column vector of 30 zeroes; $\mathbf{H}$ is a 30×30 diagonal matrix;

   - the observation errors $\boldsymbol{\varepsilon}_t$ are not correlated with each other, with x_t, with lags of x_t and with lags of $\boldsymbol{\varepsilon}_t$;

   - $\mathbf{x}_0$ denotes the initial estimates of the latent factors and we set $\mathbf{x}_0 = \boldsymbol{\mu}$ for the quadratic models, $\mathbf{x}_0 = \left(\mu_1, 0, 0\right)'$ for the affine models and $\mathbf{x}_0 = \left(\mu_1^*, \mu_2^*, \mu_3^*\right)'$ for the Black model.

We use a first order Taylor's series approximation so that

$$\mathbf{y}_t = \mathbf{z}\left(\mathbf{x}_t\right) + \boldsymbol{\varepsilon}_t \backsimeq \mathbf{z}\left(\widehat{\mathbf{x}}_t\right) + \mathbf{D}_t\mathbf{x}_t - \mathbf{D}_t\widehat{\mathbf{x}}_t + \boldsymbol{\varepsilon}_t \qquad (43)$$

$$\mathbf{D}_t = \left[\frac{\partial \mathbf{z}\left(\mathbf{x}_t\right)}{\partial \mathbf{x}'_t}\right]_{\mathbf{x}_t = \widehat{\mathbf{x}}_t}. \qquad (44)$$

$\mathbf{D}_t$ is a 30×3 matrix, $\mathbf{z}\left(\mathbf{x}_t\right)$ is a 30×1 vector and $\mathbf{x}_t$ is a 3×1 vector. Then the expected value of $\mathbf{y}_t$ conditional on information at time $t - 1$ is

$$\widehat{\mathbf{y}}_t = E_{t-1}\left(\mathbf{y}_t\right) = \mathbf{z}\left(\widehat{\mathbf{x}}_t\right) + \mathbf{D}_t\widehat{\mathbf{x}}_t - \mathbf{D}_t\widehat{\mathbf{x}}_t = \mathbf{z}\left(\widehat{\mathbf{x}}_t\right). \qquad (45)$$

Then, assuming the yields for maturities of 1, 10 and 30 years are observed without error, we set

$$\widehat{\mathbf{x}}_t = E_{t-1}\left[\mathbf{x}_t\right] = \boldsymbol{\phi}\left(\boldsymbol{\mu} - \mathbf{x}_{t-1}\right) \text{ for quadratic models} \qquad (46)$$

$$\widehat{\mathbf{x}}_t = E_{t-1}\left[\mathbf{x}_t\right] = \boldsymbol{\eta} - \boldsymbol{\phi}\mathbf{x}_{t-1} \text{ for affine models} \qquad (47)$$

$$\mathbf{F}_t = \mathbf{D}_t\boldsymbol{\Sigma}\mathbf{D}'_t + \mathbf{H} \qquad (48)$$

$$\boldsymbol{\Xi}_t = \begin{bmatrix} \frac{\partial(z_1(\mathbf{x}_t))}{\partial x_{1,t}} & \frac{\partial(z_1(\mathbf{x}_t))}{\partial x_{2,t}} & \frac{\partial(z_1(\mathbf{x}_t))}{\partial x_{3,t}} \\ \frac{\partial(z_{10}(\mathbf{x}_t))}{\partial x_{1,t}} & \frac{\partial(z_{10}(\mathbf{x}_t))}{\partial x_{2,t}} & \frac{\partial(z_{10}(\mathbf{x}_t))}{\partial x_{3,t}} \\ \frac{\partial(z_{30}(\mathbf{x}_t))}{\partial x_{1,t}} & \frac{\partial(z_{30}(\mathbf{x}_t))}{\partial x_{2,t}} & \frac{\partial(z_{30}(\mathbf{x}_t))}{\partial x_{3,t}} \end{bmatrix}_{\mathbf{x}_t = \widehat{\mathbf{x}}_t} \qquad (49)$$

$$\begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} = \begin{pmatrix} \widehat{x}_{1,t} \\ \widehat{x}_{2,t} \\ \widehat{x}_{3,t} \end{pmatrix} + \boldsymbol{\Xi}_t^{-1} \cdot \begin{pmatrix} y_{1,t} & \widehat{y}_{1,t} \\ y_{10,t} & - & \widehat{y}_{10,t} \\ y_{30,t} & & \widehat{y}_{30,t} \end{pmatrix}. \qquad (50)$$

As the disturbances $\boldsymbol{\varepsilon}_t$ are all normally distributed, we can derive the conditional quasi-likelihood function of $\mathbf{y}_t$, which is

$$l\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right) \sim N\left(\widehat{\mathbf{y}}_t, \mathbf{F}_t\right) \qquad (51)$$

where $N\left(\widehat{\mathbf{y}}_t, \mathbf{F}_t\right)$ denotes the multivariate normal density with mean $\widehat{\mathbf{y}}_t$ and covariance matrix $\mathbf{F}_t$. As the vector $\mathbf{y}_t$ is made up 30 entries, we can write the log of $l\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right)$ as

$$\ln l\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right) = -\frac{30}{2}\ln\left(2\pi\right) - \frac{1}{2}\ln\left(abs\left(|\mathbf{F}_t|\right)\right) - \frac{1}{2}\left(\mathbf{y}_t - \widehat{\mathbf{y}}_t\right)'\mathbf{F}_t^{-1}\left(\mathbf{y}_t - \widehat{\mathbf{y}}_t\right). \qquad (52)$$

$abs\left(|\mathbf{F}_t|\right)$ denotes the absolute value of the determinant of $\mathbf{F}_t$. Then the quasi log-likelihood to be maximised in order to estimate the model parameters is

$$lk = \Sigma_{t=1}^M \ln l\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}\right). \qquad (53)$$

where M is the number of in-sample observation dates, which is $2,286$ as explained below. The time step Δ is the time between consecutive observations. We observe around 261 trading days in one year, so that $\Delta = 1/261$.

7.2 Empirical results

The sample is made up of daily yields for AAA-rated Euro area Government bonds for yearly maturities from one year up to thirty years. The yields are provided by the European Central Bank's web-site and are the result of a Svensson-type daily interpolation to the market prices of the AAA-rated Government bonds **in the Euro area**. The "in-sample" period is from 6-9-2004 to 2-8-2013 and is made up of 2286 trading days with 30 yields for each day. This is the sample used to estimate the parameters for each model, either though EKF-QML or though QML. The "out-of-sample" period is from 5-8-2013 to 8-10-2014 and is made up 301 trading days with 30 yields for each day. This sample is used to test each model, whose parameters have already been estimated "in-sample".

Table 2 presents summary statistics for the yield data in the sample. For every yield maturity the mean, standard deviation, minimum and maximum are computed, both in sample and out of sample. For all yield maturities the average and the standard deviation are much lower during the out-of-sample period. This fact is important to interpret the empirical fit of the models. For the shortest maturities the minimum observed yields are negative both in sample and out of sample. Therefore in QML estimation quadratic models must be adjusted so as to predict slightly negative one year yields.

[TABLE 2 ABOUT HERE]

Table 3 presents the results of EKF-QML estimation and Table 4 the results of QML estimation. The columns headed "param" report the paremeter estimates and the columns headed "stdev" report the respective standard deviations of the parameter estimates computed with the BHHH estimator. All optimisation of the quasi likelihood functions if carried out with the Nelder-Mead Simplex Method. The row *Average h* provides overall measures of the goodness of fit of the different models under EFK-QML. h is the estimated standard deviation of observation errors in the measurement equation of the Kalman Filter for every single maturity. *Average h* is the simple average of h across all maturities for a single model. *Average h* in the range $0.0013 - 0.0015$ in Table 3 for the affine models reveal that affine models worst fit observed yields. Quadratic models of the type Q3.1 perform better with *Average h* of 0.0008, while the Black and Q3.3 models perform best with *Average h* of 0.0005 *and 0.0006 respectively*. The rows AIC and AICc respectively compute the Akaike information criterion and the corrected Akaike information criterion. According to AIC **and to the log-likelihood function** lk, model Q3.1.1 is the best and the differences with the AIC of other models are significant.

[TABLE 3 ABOUT HERE]

Table 4 presents the results of QML estimation. The postscript "ne" next to the name of each model means "no error" and reminds us that such models under QML are estimated while assuming no error in our observations of the one year, ten year and thirty year maturities. According to QML estimation *Average*

h is in the range $0.0008 - 0.0009$ for affine models and 0.0005 for quadratic models, except for Q3.3 whose *Average h* is 0.0007. **Thus in Table 3** *Average* **h is higher for Q3.1 than for Q3.3, while the opposite is true in Table 4.** Also QML estimation suggests that quadratic models perform better than affine ones. As explained below, QML estimation results are not reported for the Black model. According to AIC **and to the log-likelihood function** lk, model Q3.1.1ne is the best and the differences with the AIC of other models are significant. This result confirms that in Table 3.

[TABLE 4 ABOUT HERE]

Table 5 presents RMSE (Root Mean Squared Errors) for all models and maturities, both under EKF-QML and under QML, both in sample and out of sample. The columns headed "in sam" report in sample RMSE and the columns headed "out of sam" report out of sample RMSE. RMSE **tend to be higher than** the corresponding standard deviations of errors h, since the errors are not "white noise" and are not distributed according to Gaussian densities. For example under EKF-QML estimation **for the Q3.1.1 model** $h_1 = 0.0055$ **and the in sample RMSE for the one year maturity is** 0.0056. Average RMSE tend to higher than the corresponding *Average h* for the various models, but also Average RMSE confirm that quadratic models perform better than affine ones both under EKF-QML and under QML estimation, while the Black model seems to perform best. In Table 5 RMSE are positive even for the perfectly observed maturities of one year, ten years and thirty years under QML estimation. The reason is that RMSE refer to errors that are differences between observed yields and model predicted yields, where the prediction is based on previous day information. In other words even for perfectly observed maturities there are prediction errors, while observation errors are absent.

Table 5 shows that for the affine models RMSE as per EKF-QML are higher out sample than in sample for the shorter maturities, while for longer maturities RMSE are similar in sample and out of sample. This seems due to the difficulty of affine model in matching the very low short term yields of the "out of sample" period. A similar consideration is also applicable to quadratic models, but to a lesser extent.

[TABLE 5 ABOUT HERE]

The tables show that all models have at least some factors that are negatively correlated. It is a strength of quadratic Gaussian models and of the extended Black model the fact that the latent factors may be negatively correlated and yet pose no admissibility problems, while yields are guaranteed to be non negative (or only slightly negative). Affine models in which factor volatility depends on "square roots" of the factors can accommodate negative correlation if only admissibility is sacrificed or if the short rate is allowed to turn negative.

7.3 Affine Gaussian vs quadratic Gaussian models

The conclusion that affine **Gaussian** models perform worse than quadratic and Black models is supported by Tables 3, 4 and 5. In Table 5 average RMSE (Root Mean Square Errors) for EKF-QML across all maturities confirm these conclusions both in sample and out of sample. In Table 5 all affine Gaussian models have higher RMSE than their corresponding quadratic counterparts both in sample and out of sample. For example Average RMSE for A3.1.1 are 0.0021 both in sample and out of sample, significantly higher than *0.0011* and *0.0012* respectively in sample and out of sample for Q3.1.1. We draw similar conclusions comparing A3.1.2 with Q3.1.2, A3.1.0 with Q3.1.0 and A3.3 with Q3.3. Moreover we draw similar conclusions **both** under QML estimation and EKF-QML estimation. A3.1.1ne has average RMSE of 0.0009 in sample and 0.0014 out of sample, while Q3.1.1ne has average RMSE of 0.0007 in sample *and out* of sample. **Similar are the conclusions when** comparing A3.1.2ne with Q3.1.2ne and A3.1.0ne with Q3.1.0ne. One exception is Q3.3, which out of sample performs much worse than A3.3 under QML estimation. The reason is explained below and is due to problems of model Q3.3 rather than **to** merits of model A3.3.

Overall Table 5 shows that quadratic models beat affine ones both in sample and out of sample, both under EKF-QML estimation and QML estimation (with the exception of model Q3.3). The superior performance of quadratic models is partly due to the fact that yields are very close to zero during the sample period. In such setting yields are likely to turn negative according to affine Gaussian models, but not according to quadratic Gaussian models. Low yields expose this notorious drawback of affine Gaussian models and this notorious merit of quadratic Gaussian models. However during the sample period we do observe negative yields for short maturities and for relatively protracted periods. German and French Government bond yields were negative for some time. In spite of negative yields, affine **Gaussian** models clearly under-perform quadratic models.

Quadratic models are more complicated than **affine Gaussian** models, but such complication seems worthwhile because of superior empirical fit to observed yields.

7.4 Specifications for affine Gaussian models

The results shed light on the performance of alternative specifications of affine models, namely A3.3, A3.1.0, A3.1.1, A3.1.2 and AJ3.1.0. According to *Average h* as per EKF-QML and QML estimation in Tables 3 and **4**, all affine **Gaussian** models, except for A3.3, fit yields similarly well. Both EKF-QML and QML show that the standard deviations of errors h across the different affine models are very similar for any given maturity. For **the** affine **Gaussian models** there seems to be no specification that clearly dominates the others. Table 5 confirms this. **AJ3.1.0 is the same as A3.1.0 except that r_t may "jump". Jumps improve the empirical performance of the model, as confirmed in Ta-**

bles 4 and 5. For example in Table 4 AIC is $-819,332$ for **AJ3.1.0** and $-815,125$ for **A3.1.0: a significant difference in favour of "jumps". However Tables 4 and 5 also confirm that quadratic models still perform clearly better than all the affine Gaussian models, even with "jumps".**

7.5 Specifications for quadratic models

Under both EKF-QML and QML estimation the different versions of Q3.1 have similar empirical fit. **According to AIC in Tables 3 and 4 model Q3.1.1 is the best and models Q3.1 beat Q3.3 both under EKF-QML estimation and QML estimation. AIC penalises Q3.3 for its greater number of parameters than models Q3.1.**

EKF-QML estimation shows that Q3.3 fits shorter term yields better and longer term yields worse than do the versions of Q3.1. This is the effect in models Q3.1 of "separating" the variables into the one that drives the short rate r_t and those that only drive the central tendency of the short rate. Q3.3 "lays the burden" of matching both short term and long term yields on all three factors, while Q3.1 models "lay the burden" of matching shorter yields mainly on the one factor that drives r_t and leave the other factors "more free" to match long term yields. Testing the models on maturities up to 30 years highlights this effect of "separating" the variables in Q3.1 models. Furthermore we recall that Q3.3 has more parameters than the Q3.1 models.

According to QML estimation in Table 4 Q3.3 performs worse than do the Q3.1 models. This is even more apparent in Table 5 where Average RMSE out of sample are *0.0860* for Q3.3: the reason is that Q3.3 finds it more difficult to match the very low short term yields out-of-sample**; since in Q3.3** all three factors drive the short rate r_t**, all three factors** must be close to 0 at the same time in order to match low short term yields**, which makes it difficult for Q3.3 to match the relatively higher long term yields at the same time**. This difficulty of Q3.3 is more apparent under **QML estimation out of sample, to the point that out of sample Q3.3 cannot match** the one year yield, which is one of three maturities assumed to be observed without error. Hence RMSE out of sample "explode" to *0.0860* for Q3.3 under QML estimation. In sample RMSE do not explode because the parameter $\alpha = -0.0038$ enables Q3.3 to match very low or even negative yields, but out of sample even such a low value of α does not suffice to enable Q3.3 **to match one year yields, because** during the out of sample period short term yields reached record low levels. Unlike the Q3.3 model, Q3.1 models under QML do not "suffer" when yields are very low, precisely because **only** one factor drives r_t, so that only that factor needs to be close to zero when the one year yield is close to zero.

The results in Tables 3 and 4 confirm the merits of the "new" specification of quadratic models, namely **Q3.1,** as opposed to **Q3.3 according to** Ahn-Dittmar-Gallant (2002).

For the maturities observed without errors, i.e. $1, 10, 30$ **years, and**

for some "neighboring" maturities, RMSE as per QML estimation in Table 5 tend to be higher in sample than out of sample. This is not surprising since those maturities are observed without errors both in sample and out of sample. Moreover yields tend to be lower out of sample, implying lower RMSE out of sample.

7.6   The Black model

Discount bond prices according to the Black model are computed using MOL with "sequential splitting" on a grid of 100 steps in each space dimension (the solution region is $[-1, 1]$ in each "space" dimension) and with 20 time steps per year. According to EKF-QML estimation in Table 3 and Table 5 the Black model achieves the best overall empirical fit to observed yields. What is striking for the Black model in Table 5 are the extremely low RMSE out of sample, which are much lower than the out of sample RMSE of any other model, even of model Q3.3. The Black model performs very well, indeed better than all other models, when yields are very close to zero and for short term yields, as is apparent from the RMSE in Table 5 for the out-of-sample period. A relative weakness of the Black model is in matching the longest maturities. For these maturities in Table 3 all other models have lower h's, i.e. lower standard deviations of observation errors.

The disappointing performance of the Black model for long term yields is due to the fact that the short rate is non-linear in the factors, so that the higher factor volatility is, the lower bond prices are and the higher bond yields are; this effect is stronger for longer maturities. To match low long term yields, the estimated long term mean of factors is lowered, but lowering such long term mean below zero will not reduce model predicted long term yields by much, because of the zero lower bound for the short rate. For these two reasons the Black model tends to overestimate long term yields, when these are particularly low as is the case in our sample. This is a shortcoming of the Black model.

The RMSE of the Black model for almost all maturities are lower out of sample than in sample. This **outstanding** result is observed only for the Black model and **can be** explained by the fact that observed yields are lower out of sample than in sample, **as well as by good model performance**. **Lower yields tend to be associated with lower RMSE.**

The estimated parameters p, p^* and ρ of the Black model in Table 3 are significant and indicate that the latent factors x_1 and x_2 are not independent. Therefore it seems worthwhile to use vertical MOL with sequential splitting, as explained above, in order to compute W, where W reflects the fact that x_1 and x_2 are not independent. The parameter q_1 is estimated at $58, 15$ and this means that, for example, when $x_1 = 0.5$, $r \geq 0.5^{58,15} = 3 \cdot 10^{-18} \simeq 0$. The Black model is capable to match extremely low yields.

7.7   EKF-QML vs QML estimation

Quadratic models can predict negative yields if only $\alpha < 0$. α is like the "floor" of the short rate r_t. $\alpha < 0$ is necessary in QML estimation of quadratic models because of negative yields in the sample, otherwise quadratic models could not match negative one year yields, which are assumed to be perfectly observed.

To gauge the accuracy of equations 49 and 50 in inferring the latent factors for quadratic models, Table 6 is provided. Table 6 displays the mean absolute value of errors (MAE) across both the in sample and out of sample periods, which total 2.587 trading days. For example for the one year maturity MAE is

$$\Sigma_{t=1}^{2587} \frac{|y_{1,t} - z_1\left(\mathbf{x}_t\right)|}{2587}$$

where $\mathbf{x}_t$ is computed as in equations 49 and 50 for every trading day. MAE reveal the disappointing performance of model Q3.3, which has problems in matching the record low one year yields during the out of sample period, as explained before. Instead MAE show that for models Q3.1 equations 49 and 50 provide an accurate way to infer the latent factors. The yields predicted on the basis of the latent factors almost coincide with the perfectly observed yields used to infer the latent factors themselves.

[TABLE 6 ABOUT HERE]

All models have smaller *average* RMSE, both in sample and out of sample, when estimated through QML rather than through EKF-QML. This is not surprising as QML estimation is such that each model every day perfectly matches the one year, ten year and thirty year maturities. RMSE tend to higher (lower) for maturities that are further from (nearer to) the one year, ten year and thirty year maturities.

The results **of** QML estimation for the Black model are not displayed, because such results were unreliable. One problem is that the QML algorithm tends to "push" the latent factors outside the bounded solution region, which is $[-1, 1]$ for the MOL algorithm for the Black model. Another problem is that MOL only provides approximations to the partial derivatives of model predicted yields with respect to the latent factors. Such approximations seem too imprecise to infer the latent factors through equations 49 and 50 of the QML algorithm. This is a drawback of the Black model.

7.8   The term structures of errors

We test all models on yields up to thirty years, not up to ten years. This is a tough test that better highlights the deficiencies of models. EKF-QML estimation shows that all models have more difficulty in matching the extreme ends of the yield curve, i.e. the shortest or the longest maturities. For all models h's and RMSE are highest for the shortest maturities, such as one year, and gradually decrease for the longer maturities. For some models, such as Black or the quadratic models, h's and RMSE tend to rise again for the longest

maturities. Affine models do not fit the longest maturities worse than the other models.

Under EKF-QML the Black model has the lowest h's for the shortest maturities, such as one year, followed by Q3.3, whereas the affine models have the highest h's for such maturities. This fact reflects the merits of the models that do not permit the short rate to turn negative. For the longest maturities, such as thirty years, affine and Q3.1 models perform similarly and the Black model performs worst.

7.9   Calibration and unreported results

We also carry out some model calibration for affine and quadratic models, which is the same as QML estimation described above, i.e. still uses equations 49 and 50 for every trading date, except for the fact that model parameters are found by minimising the sum of squared pricing errors

$$SSE = \Sigma_{t=1}^{M} \left(\mathbf{y}_t - \mathbf{z} \left(\mathbf{x}_t \right) \right)' \left(\mathbf{y}_t - \mathbf{z} \left(\mathbf{x}_t \right) \right) \tag{54}$$

rather than the log-likelihood lk, and except for the fact that for all models we set $\widehat{\mathbf{x}}_t = \mathbf{x}_{t-1}$. Calibration is of interest because it does not depend on the physical probability measure of the factors. Notice that $\widehat{\mathbf{y}}_t = \mathbf{z} \left(\widehat{\mathbf{x}}_t \right) \neq \mathbf{z} \left(\mathbf{x}_t \right)$. Therefore in equation 54 the errors we mimimise for any date t are the differences between observed yields and the cross section of model predicted yields, where yields are predicted given that we already know the value of the latent factors $\mathbf{x}_t$ for date t. Calibration makes no assumption about the distribution of the errors and no assumption about the processes for the latent factors under the real probability measure. Through calibration the pricing model predicts only the cross section and not the time series of yields. For simplicity we do not report the calibration results, but the gist of the results is that Q3.1 models have lower SSE than affine models, except for model A3.1.0. Again the Q3.1 quadratic models beat affine models, except for affine model A3.1.0.

Finally, for simplicity and because of their disappointing empirical performance with respect to the other models, we do not report the results of EKF-QML estimation of Q3.2, which is a new specification of the quadratic model described in the Appendix whereby two factors drive the short rate and the third factor only drives the central tendency of the short rate.

7.10   Residual considerations

The low yields in the sample affect the performance of the tested models. Low yields highlight the merits of specification Q3.1, whereby long term yields can be relative high even while short term yields are almost zero, and the merits of the Black model, which can match very low short term yields. Low yields also stress the shortcomings of affine Gaussian models, which cannot rule out negative yields. In a sample with higher yields the merits of Q3.1 and Black model may

become less apparent, while affine Gaussian models may well become more competitive.

Low Government bond yields have been observed for a number of countries beyond the Eurozone after 2008. Also for these countries we could expect better empirical performance of quadratic Gaussian models and Black models in comparison to affine Gaussian models. Also for these countries Q3.3 can be expected to "beat" Q3.1 on short term yields and to "lose" on long term yields, for the same reasons given above for the Eurozone. Less obvious for non-Euro countries is the comparative empirical performance of quadratic Gaussian models and Black models, however these models fit US Government yields similarly well according to preliminary evidence not reported in the paper.

8   Conclusion

This paper has tested affine Gaussian, quadratic Gaussian and Black-type **Gaussian** models on the term structure of Euro area triple A Government bond yields for maturities up to 30 years. Quadratic Gaussian models "beat" affine Gaussian models both in sample and out of sample. Quadratic models better fit observed yields for the shortest maturities. An extended version of the Black (1995) model beats all other Gaussian models out-of-sample and seems the best model to match the extremely low Euro yields observed since 2013.

Estimation through Quasi Maximum Likelihood (QML), which assumes that some yields are observed without error, is feasible even for quadratic models.

This paper has tested promising new specifications of quadratic term structure model, whereby only one factor drives the short rate and the other factors drive the central tendency of the short rate. These new specifications are more suitable to QML estimation than the "classic" specification of quadratic models as per Ahn-Dittmar-Gallant (2002). However the "classic" specification fits short term yields slightly better **than the new specifications**.

This paper has also introduced the vertical method of lines (MOL) to solve **bond pricing equations** for the Black model. MOL seems preferable to implicit finite difference methods for bond pricing. For pricing bonds driven by **two or more** stochastic factors, vertical MOL involves computing exponentials of large matrixes, which can be **prohibitive**. However equation "splitting" can dramatically alleviate the burden of computations. "Sequential splitting" provides quick and unconditionally stable numerical solutions as long as the Wiener processes of the two factors are not correlated: when they are correlated, sequential splitting still provides quick solutions, but these no longer are unconditionally stable. This seems the main unsolved problem of vertical MOL.

Overall, the effort to pass from affine Gaussian term structure models to quadratic ones seems worthwhile, at least for Euro area yields. The effort to pass from quadratic Gaussian models to extensions of the Black model may also be worthwhile on the grounds of superior empirical fit to very low short

term yields. However in practice the tractability of quadratic models remains an important advantage over the Black model.

A   Derivation of identification conditions of the Ahn-Dittmar-Gallant (ADG) model and of model Q3.3

Ahn-Dittmar-Gallant (2002) derive as follows the conditions for the identification of the model parameters. They consider linear invariant transformations of $\mathbf{x}$, as $\mathbf{x}$ has Gaussian distribution and only linear transformations of $\mathbf{x}$ will retain a Gaussian distribution. A transformation is invariant if it does not alter model predicted yields. Denote the generic invariant linear transformation as $\mathbf{x} = \Omega \mathbf{y} + \mathbf{\Theta}$, where $\mathbf{\Theta}$ and $\mathbf{y}$ are 3×1 vectors and Ω is an 3×3 matrix. Ω^{-1} is assumed to exist. α is a scalar constant. Since $r_t = \alpha + \beta' \mathbf{x}_t + \mathbf{x}_t' \mathbf{\Psi} \mathbf{x}_t$ and $\mathbf{x}_{t+1} = (\mathbf{I}_3 - \boldsymbol{\phi}) \mathbf{x}_t + \boldsymbol{\phi} \boldsymbol{\mu} + \mathbf{\Sigma} \boldsymbol{\xi}_{t+1}$, it follows that

$$r_t = \alpha + \beta' \mathbf{\Theta} + \mathbf{\Theta}' \mathbf{\Psi} \mathbf{\Theta} + \beta' \Omega \mathbf{y}_t + \mathbf{y}_t' \Omega' \mathbf{\Psi} \mathbf{\Theta} + \mathbf{\Theta}' \mathbf{\Psi} \Omega \mathbf{y}_t + \mathbf{y}_t' \Omega' \mathbf{\Psi} \Omega \mathbf{y}_t \tag{55}$$

$$\mathbf{y}_{t+1} - \mathbf{y}_t = \Omega^{-1} \left(\boldsymbol{\phi} \boldsymbol{\mu} - \boldsymbol{\phi} \left(\mathbf{\Theta} + \Omega \mathbf{y}_t \right) \right) + \Omega^{-1} \mathbf{\Sigma} \boldsymbol{\xi}_{t+1}. \tag{56}$$

Then only if $\boldsymbol{\beta} = \mathbf{0}_{3 \times 1}$ do we need to impose $\mathbf{\Theta} = \mathbf{0}_{3 \times 1}$ in order for the transformation to be invariant. Only if $\mathbf{\Theta} = \mathbf{0}_{3 \times 1}$ can $\boldsymbol{\mu}$ be uniquely identified. Since $\mathbf{\Psi}$ is symmetric and $\boldsymbol{\beta} = \mathbf{\Theta} = \mathbf{0}_{3 \times 1}$, it follows that

$$r_t = \alpha + \mathbf{y}_t' \Omega' \mathbf{\Psi} \Omega \mathbf{y}_t \tag{57}$$

$$\mathbf{y}_{t+1} - \mathbf{y}_t = \Omega^{-1} \left(\boldsymbol{\phi} \boldsymbol{\mu} - \boldsymbol{\phi} \Omega \mathbf{y}_t \right) + \Omega^{-1} \mathbf{\Sigma} \boldsymbol{\xi}_{t+1}. \tag{58}$$

Then, in order for the transformation to be invariant, we need to constrain Ω to be equal to the identity matrix $\mathbf{I}_3$. Notice that Ω must be diagonal either when $\mathbf{\Sigma}$ is diagonal and $\boldsymbol{\phi}$ is triangular or when $\mathbf{\Sigma}$ is triangular and $\boldsymbol{\phi}$ is diagonal. Of our interest is only the case where $\mathbf{\Sigma}$ is triangular. If $\mathbf{\Sigma}$ is triangular, so is $\Omega^{-1} \mathbf{\Sigma}$ provided Ω is triangular. Indeed in order for the transformation to be invariant, $\Omega^{-1} \mathbf{\Sigma}$ needs to be triangular if so is $\mathbf{\Sigma}$. Then if Ω is triangular and if $\boldsymbol{\phi}$ is diagonal, $\Omega^{-1} \boldsymbol{\phi} \Omega$ will be diagonal and indeed, in order for the transformation to be invariant, $\Omega^{-1} \boldsymbol{\phi} \Omega$ needs to be diagonal if so is $\boldsymbol{\phi}$. Finally $\Omega' \mathbf{\Psi} \Omega$ must have all diagonal entries equal to 1 and $\Omega^{-1} \boldsymbol{\phi} \boldsymbol{\mu} \geq \mathbf{0}_{3 \times 1}$ in order for the transformation to be invariant. Since in the Ahn-Dittmar-Gallant (ADG) model $\mathbf{\Psi}$ has all diagonal entries equal to 1 and $\boldsymbol{\phi} \boldsymbol{\mu} \geq \mathbf{0}_{3 \times 1}$, then $\Omega = \mathbf{I}_3$.

However, unlike in the Ahn-Dittmar-Gallant (2002) continuous time setting, in discrete time $\boldsymbol{\phi}$ need not be diagonal. Notice that if Ω is lower triangular, then

$$
\boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{1,1} & \Omega_{2,1} & \Omega_{3,1} \\ 0 & \Omega_{2,2} & \Omega_{3,2} \\ 0 & 0 & \Omega_{3,3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ \Omega_{3,1} & \Omega_{3,2} & \Omega_{3,3} \end{pmatrix}
$$

$$
= \begin{pmatrix} \Omega_{1,1}^2 + \Omega_{2,1}^2 + \Omega_{3,1}^2 & \Omega_{2,1}\Omega_{2,2} + \Omega_{3,1}\Omega_{3,2} & \Omega_{3,1}\Omega_{3,3} \\ \Omega_{2,1}\Omega_{2,2} + \Omega_{3,1}\Omega_{3,2} & \Omega_{2,2}^2 + \Omega_{3,2}^2 & \Omega_{3,2}\Omega_{3,3} \\ \Omega_{3,1}\Omega_{3,3} & \Omega_{3,2}\Omega_{3,3} & \Omega_{3,3}^2 \end{pmatrix}
$$

$$
= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
$$

which implies that $\Omega_{3,3} = \Omega_{1,1} = \Omega_{2,2} = \pm 1$ and $\Omega_{3,1} = \Omega_{3,2} = \Omega_{2,1} = 0$. **$\Omega_{2,1}$ is the element of Ω in the second row and first column. The other elements of Ω have similar interpretation.** Finally if $\boldsymbol{\phi\mu} \geq \mathbf{0}_{3\times 1}$, then $\boldsymbol{\Omega}^{-1}\boldsymbol{\phi\mu} \geq \mathbf{0}_{3\times 1}$ if $\Omega_{3,3} = \Omega_{1,1} = \Omega_{2,2} = 1$. **Thus the transformation is invariant without additional restrictions on ϕ and $\boldsymbol{\Sigma}$,** therefore in the discrete time both ϕ and $\boldsymbol{\Sigma}$ can be triangular, unlike in continuous time. The key reason is that in discrete time estimation depends on the conditional covariance matrix of $\mathbf{x}_{t+1}$, which is $\boldsymbol{\Sigma\Sigma}'$ and does not depend on ϕ. Instead in **continuous** time the **factors** conditional covariance also depends on ϕ so that parameter identification entails that ϕ be diagonal if $\boldsymbol{\Sigma}$ is triangular.

More generally in Ahn-Dittmar-Gallant (2002) continuous time setting $\boldsymbol{\Psi}$ is symmetric with all diagonal entries equal to 1 so that $\boldsymbol{\Psi} = \begin{pmatrix} 1 & g_1 & g_2 \\ g_1 & 1 & g_3 \\ g_2 & g_3 & 1 \end{pmatrix}$. g_1, g_2, g_3 **are constants.** Then it can be shown that imposing this definition of $\boldsymbol{\Psi}$ in model Q3.3 gives

$$
\begin{pmatrix} \Omega_{1,1} & \Omega_{2,1} & \Omega_{3,1} \\ 0 & \Omega_{2,2} & \Omega_{3,2} \\ 0 & 0 & \Omega_{3,3} \end{pmatrix} \begin{pmatrix} 1 & g_1 & g_2 \\ g_1 & 1 & g_3 \\ g_2 & g_3 & 1 \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ \Omega_{3,1} & \Omega_{3,2} & \Omega_{3,3} \end{pmatrix} = \begin{pmatrix} 1 & g_1 & g_2 \\ g_1 & 1 & g_3 \\ g_2 & g_3 & 1 \end{pmatrix}.
$$

which implies that $\boldsymbol{\Omega} = I$. Therefore for the Q3.3 model in discrete time parameter identification **allows** *both* ϕ and $\boldsymbol{\Sigma}$ **to be** triangular whatever the values of g_1, g_2, g_3.

B Appendix: conditions for the identification of parameters of models Q3.1

We consider the linear transformation $\mathbf{x}_t = \boldsymbol{\Omega}\mathbf{y}_t + \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a 3×1 vector of parameters and $\boldsymbol{\Omega}$ is a 3×3 matrix of parameters. $\boldsymbol{\Omega}^{-1}$ is assumed to exist. $\mathbf{x}$ and $\mathbf{y}$ are 3×1 vectors of variables that change over time. The transformation is invariant if and only if $\boldsymbol{\Omega} = \mathbf{I}_3$ and $\boldsymbol{\Theta} = \mathbf{0}_{3\times 1}$. $\mathbf{I}_3$ is the 3×3 identity matrix.

$\mathbf{0}_{3\times1}$ is the 3×1 vector of zeros. Then, since

$$r_t = \alpha + \mathbf{x}_t'\boldsymbol{\Psi}\mathbf{x}_t$$
$$\mathbf{x}_{t+1} = (\mathbf{I}_3 - \boldsymbol{\phi})\,\mathbf{x}_t + \boldsymbol{\phi\mu} + \boldsymbol{\Sigma}\boldsymbol{\xi}_{t+1}$$

we can employ the linear transformation to write

$$r_t = \alpha + \boldsymbol{\Theta}'\boldsymbol{\Psi}\boldsymbol{\Theta} + \mathbf{y}_t'\boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Theta} + \boldsymbol{\Theta}'\boldsymbol{\Psi}\boldsymbol{\Omega}\mathbf{y}_t + \mathbf{y}_t'\boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Omega}\mathbf{y}_t \qquad (59)$$
$$\mathbf{y}_{t+1} - \mathbf{y}_t = \boldsymbol{\Omega}^{-1}\left(\boldsymbol{\phi\mu} - \boldsymbol{\phi}\boldsymbol{\Theta} - \boldsymbol{\phi}\boldsymbol{\Omega}\mathbf{y}_t\right) + \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\xi}_{t+1}. \qquad (60)$$

The said transformation is invariant only if $\boldsymbol{\Theta}'\boldsymbol{\Psi}\boldsymbol{\Theta} + \mathbf{y}_t'\boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Theta} + \boldsymbol{\Theta}'\boldsymbol{\Psi}\boldsymbol{\Omega}\mathbf{y}_t = 0$, i.e. if and only if $\boldsymbol{\Theta} = \mathbf{0}_{3\times1}$. The fact that $\boldsymbol{\Theta} = \mathbf{0}_{3\times1}$ entails that $\boldsymbol{\phi\mu}$ can be uniquely identified and that

$$r_t = \mathbf{y}_t'\boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Omega}\mathbf{y}_t \qquad (61)$$
$$\mathbf{y}_{t+1} - \mathbf{y}_t = \boldsymbol{\Omega}^{-1}\left(\boldsymbol{\phi\mu} - \boldsymbol{\phi}\boldsymbol{\Omega}\mathbf{y}_t\right) + \boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}\boldsymbol{\xi}_{t+1}. \qquad (62)$$

Hereafter we impose conditions that entail that $\boldsymbol{\Omega} = \mathbf{I}_3$ so that the transformation be invariant. As $\boldsymbol{\Sigma}$ is lower triangular, $\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}$ should also be triangular and $\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}$ can only be lower triangular if $\boldsymbol{\Omega}$ is also lower triangular, i.e. if $\Omega_{1,2} = \Omega_{1,3} = \Omega_{2,3} = 0$ so that

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ \Omega_{3,1} & \Omega_{3,2} & \Omega_{3,3} \end{pmatrix}.$$

Since $\boldsymbol{\Omega}$ is lower triangular and since the transformation is invariant if only if $\boldsymbol{\Omega}'\boldsymbol{\Psi}\boldsymbol{\Omega} = \boldsymbol{\Psi}$, it follows that when $\boldsymbol{\Psi} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ the transformation is invariant if and only if

$$\begin{pmatrix} \Omega_{1,1} & \Omega_{2,1} & \Omega_{3,1} \\ 0 & \Omega_{2,2} & \Omega_{3,2} \\ 0 & 0 & \Omega_{3,3} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ \Omega_{3,1} & \Omega_{3,2} & \Omega_{3,3} \end{pmatrix} = \begin{pmatrix} \Omega_{3,1}^2 & \Omega_{3,1}\Omega_{3,2} & \Omega_{3,1}\Omega_{3,3} \\ \Omega_{3,1}\Omega_{3,2} & \Omega_{3,2}^2 & \Omega_{3,2}\Omega_{3,3} \\ \Omega_{3,1}\Omega_{3,3} & \Omega_{3,2}\Omega_{3,3} & \Omega_{3,3}^2 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This condition is satisfied only when $\Omega_{3,3} = 1$, $\Omega_{3,1} = \Omega_{3,2} = 0$, i.e. when $\boldsymbol{\Omega}$ is such that

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then we notice that

31

$$\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}\boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{\Omega_{1,1}} & 0 & 0 \\ -\frac{1}{\Omega_{1,1}}\frac{\Omega_{2,1}}{\Omega_{2,2}} & \frac{1}{\Omega_{2,2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ -\phi_{2,1} & \phi_{2,2} & 0 \\ 0 & -\phi_{3,2} & \phi_{3,3} \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ \frac{\Omega_{2,1}}{\Omega_{2,2}}\phi_{2,2} - \left(\frac{1}{\Omega_{2,2}}\phi_{2,1} + \frac{1}{\Omega_{1,1}}\frac{\Omega_{2,1}}{\Omega_{2,2}}\phi_{1,1}\right)\Omega_{1,1} & \phi_{2,2} & 0 \\ -\Omega_{2,1}\phi_{3,2} & -\Omega_{2,2}\phi_{3,2} & \phi_{3,3} \end{pmatrix}$$

therefore $\Omega_{2,1} = 0$ in order for the transformation to be invariant, in which case

$$\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}\boldsymbol{\Omega} = \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ -\frac{\Omega_{1,1}}{\Omega_{2,2}}\phi_{2,1} & \phi_{2,2} & 0 \\ 0 & -\Omega_{2,2}\phi_{3,2} & \phi_{3,3} \end{pmatrix}.$$

Finally, only if $\phi_{2,1} = \phi_{2,2}$ and $\phi_{2,2} = \phi_{3,3}$ can we impose that $\Omega_{1,1} = \Omega_{2,2} = 1$ in order for the transformation to be invariant. Then $\boldsymbol{\Omega} = \mathbf{I}_3$, the transformation $\mathbf{x} = \boldsymbol{\Omega}\mathbf{y} + \boldsymbol{\Theta}$ does not alter the bond yields we observe (is observationally invariant), we can uniquely identity the model parameters and infer the values of the latent factors $\mathbf{x}$. We have considered the case of Q3.1.1. Other things equal, for model Q3.1.0 we impose that

$$\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}\boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{\Omega_{1,1}} & 0 & 0 \\ -\frac{1}{\Omega_{1,1}}\frac{\Omega_{2,1}}{\Omega_{2,2}} & \frac{1}{\Omega_{2,2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ -\phi_{2,1} & \phi_{2,2} & 0 \\ -\phi_{3,1} & -\phi_{3,2} & \phi_{3,3} \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ \frac{\Omega_{2,1}}{\Omega_{2,2}}\phi_{2,2} - \left(\frac{1}{\Omega_{2,2}}\phi_{2,1} + \frac{1}{\Omega_{1,1}}\frac{\Omega_{2,1}}{\Omega_{2,2}}\phi_{1,1}\right)\Omega_{1,1} & \phi_{2,2} & 0 \\ -\Omega_{1,1}\phi_{3,1} - \Omega_{2,1}\phi_{3,2} & -\Omega_{2,2}\phi_{3,2} & \phi_{3,3} \end{pmatrix}$$

so that the transformation is invariant if

$$\phi_{3,1} = \phi_{3,2} = \phi_{3,3}$$
$$\phi_{2,1} = \phi_{2,2}$$
$$\boldsymbol{\phi}\boldsymbol{\mu} \geq 0.$$

Notice that if $\phi_{2,1} = 0$ and $\phi_{3,1} = \phi_{3,2} = \phi_{3,3}$ still $\boldsymbol{\Omega} = \mathbf{I}_3$: this is the case of Q3.1.2. The case where $\phi_{2,1} = 0$ and $\phi_{3,1} = 0$ is of no interest.

C Specification of model Q3.2

In another subfamily of three factor quadratic models, namely Q3.2, two factors drive r_t and the third factor only drives the central tendency of r_t so that

$$r_t = \begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix}' \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix}$$

$$\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ -\phi_{2,1} & \phi_{2,2} & 0 \\ -\phi_{3,1} & -\phi_{3,2} & \phi_{3,3} \end{pmatrix} \Delta \left(\begin{pmatrix} 0 \\ 0 \\ \mu_3 \end{pmatrix} - \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} \right) + \boldsymbol{\Sigma} \cdot \begin{pmatrix} \xi_{1,t+1}^Q \\ \xi_{2,t+1}^Q \\ \xi_{3,t+1}^Q \end{pmatrix}$$

μ is the same as in model 3.1 above. In the text we do not consider this model because of its disappointing empirical performance. Q3.2 is here mentioned just to complete the theoretical analysis of the possible specifications of three factor quadratic models. What follows shows that the only condition needed for parameter identification is $\phi_{2,1} = \phi_{2,2}$. Under Q3.2, all else as in model Q3.1, we have

$$\begin{pmatrix} \Omega_{1,1} & \Omega_{2,1} & \Omega_{3,1} \\ 0 & \Omega_{2,2} & \Omega_{3,2} \\ 0 & 0 & \Omega_{3,3} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ \Omega_{2,1} & \Omega_{2,2} & 0 \\ \Omega_{3,1} & \Omega_{3,2} & \Omega_{3,3} \end{pmatrix} =$$

$$= \begin{pmatrix} \Omega_{2,1}^2 + \Omega_{3,1}^2 & \Omega_{2,1}\Omega_{2,2} + \Omega_{3,1}\Omega_{3,2} & \Omega_{3,1}\Omega_{3,3} \\ \Omega_{2,1}\Omega_{2,2} + \Omega_{3,1}\Omega_{3,2} & \Omega_{2,2}^2 + \Omega_{3,2}^2 & \Omega_{3,2}\Omega_{3,3} \\ \Omega_{3,1}\Omega_{3,3} & \Omega_{3,2}\Omega_{3,3} & \Omega_{3,3}^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

which implies that $\Omega_{2,2} = \Omega_{3,3} = 1, \Omega_{3,2} = \Omega_{3,1} = \Omega_{2,1} = 0$, so that

$$\boldsymbol{\Omega} = \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

What is left to determine is $\Omega_{1,1}$. Then we notice that

$$\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}\boldsymbol{\Omega} = \begin{pmatrix} \frac{1}{\Omega_{1,1}} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ -\phi_{2,1} & \phi_{2,2} & 0 \\ -\phi_{3,1} & -\phi_{3,2} & \phi_{3,3} \end{pmatrix} \begin{pmatrix} \Omega_{1,1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \phi_{1,1} & 0 & 0 \\ -\Omega_{1,1}\phi_{2,1} & \phi_{2,2} & 0 \\ -\Omega_{1,1}\phi_{3,1} & -\phi_{3,2} & \phi_{3,3} \end{pmatrix}.$$

Therefore $\Omega_{1,1} = 1$ in order for the transformation to be invariant and, to this effect, we can impose $\phi_{2,1} = \phi_{2,2}$.

References

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

[2] Babbs S.H. and Nowman B.K., 1999, "Kalman filtering of generalised Vasicek term structure models", Journal of Financial and Quantitative Analysis 34, 115-130.

[3] **Baz J and Das S.R., 1996, "Analytical Approximations of the Term Structure for Jump-Diffusion Processes: A Numerical Analysis," Journal of Fixed Income 6, n.1, 78-86.**

[4] Black F., 1995, "Interest rates as options", Journal of Finance 50, 1371-1376.

[5] Chen L., Filipovic D. and Poor V., 2004, "Quadratic term structure models for risk-free and defaultable rates", Mathematical Finance, 14, n.4, 515-536.

[6] **Cheng P. and Scaillet O., 2007, "Linear-quadratic jump-diffusion modeling", Mathematical Finance 17, n.4, 575-598.**

[7] Constantinides G., 1992, "A theory of the nominal term structure of interest rates", The Review of Financial Studies 5, n.4, 531-552.

[8] Dai Q. and Singleton K., 2003, "Term structure dynamics in theory and reality", The Review of Financial Studies 16, n.3, 631-678.

[9] Duffee G., 2011, "Information in (and not in) the Term Structure", Review of Financial Studies 24, n.9, 2895-2934.

[10] Duffie D. and Kan R., 1996, "A yield factor model of interest rates", Mathematical Finance 6, 379-406.

[11] Farago I., 2003, "Splitting method", August 2003 lecture notes at http://www.cs.elte.hu/~faragois/phdcourse/lecturefi2.pdf.

[12] Gourieroux C. and Sufana R., 2003, "Wishart quadratic term structure models", Working paper CREF HEC Montreal.

[13] Joslin S., Singleton K.J. and Zhu H., 2011, "A New Perspective on Gaussian Dynamic Term Structure Models", Review of Financial Studies 24, 926-970.

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

[15] Higham N. J., 2005, "The Scaling and Squaring Method for the Matrix Exponential Revisited," SIAM Journal Matrix Anal. Appl. 26, n.4, 1179-1193.

[16] Langetieg T.C., 1980, "A multivariate model of the term structure", Journal of Finance 35, 71-97.

[17] Leippold M. and Wu L., 2002, "Asset pricing under the quadratic class", Journal of Financial and Quantitative Analysis 37, n.2, 271-294.

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

[19] Rambeerich N., Tangman D.Y., Lollchund M.R., Bhuruth M. , 2013, "High-order computational methods for option valuation under multifactor models", European Journal of Operational Research 224, 219–226.

[20] Randall C. and Tavella D., 2000, Pricing financial instruments: The finite difference method, John Wiley & Sons (New York).

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

[22] Realdon M., 2009, "Extended Black" term structure models", International Review of Financial Analysis 18, 232-238.

[23] Yanenko, N.N., 1971, "The method of fractional steps", Springer-Verlag, Berlin.