

# TESTS OF NON LINEAR GAUSSIAN TERM STRUCTURE MODELS

Dr Marco Realdon

Newcastle Business School

Northumbria University

Newcastle Upon Tyne

City Campus East, NE1 8ST, UK

tel: +39/0429/3948

marco.realdon@gmail.com

First version 17/5/2011

This version 6/2/2016

## **Abstract**

Since the 2008 financial crisis Government bond yields in US, Europe and elsewhere have been historically low and challenged term structure models that cannot rule out negative yields. This paper uses US and German Government yields to test three factor Gaussian models that do and that do not rule out negative yields, namely affine models, quadratic

models, extensions of the Black and Black-Karasinski models. Quadratic models and a Vasicek-type model best fit observed yields when the stochastic factors driving the short rate are correlated. However the Black-Karasinski model for the US and the Black model for both US and Germany can best fit yields when interest rates are lowest, i.e. after 2008, despite the restriction of independent factors driving the short rate. A new linear-quadratic model whereby the central tendency of the short rate is a non-negative quadratic function of Gaussian factors performs particularly well for German yields. All models fit German yields better than US yields. All models fit the one year yield worse than longer term yields.

*Key words: quadratic model, Black model, Vasicek model, Black-Karasinski model, method of lines, Extended Kalman Filter.*

**JEL classification: G12; G13.**

## 1 Introduction

Since the 2008 financial crisis, due to strongly expansionary monetary policies, Government bond yields in US, Europe and elsewhere have been historically low almost to resemble Japanese Government bond yields. This setting challenges affine Gaussian term structure models (AGTSM) that do not rule out negative yields. Therefore this study uses US and German Government bond yields to test term structure models that do and do not rule out negative yields, and in particular models in which the instantaneous short interest rate is a non-negative function of Gaussian latent factors, such as quadratic term structure

models (QTSM) and extensions of the Black (1995) and Black-Karasinski (1991) models. Black and Black-Karasinski models are hereafter referred to as BBKM.

AGTSM, QTSM and BBKM are all driven by Gaussian latent factors. Gaussian factors are tractable and do not suffer from the admissibility restrictions that affect more general affine stochastic differential equations. As a result market prices of risk can be freely specified and the pricing models better fit observed yields as explained in Dai and Singleton (2002) for the case of AGTSM. Hence this paper concentrates on Gaussian latent factors driving the short rate.

In AGTSM and QTSM the short interest rate is either an affine or quadratic function of the latent Gaussian factors. Instead in BBKM the short rate is a more general non-negative function of the latent Gaussian factors. QTSM and BBKM can rule out a negative short interest rate and negative yields, unlike AGTSM. BBKM require numerical solutions for bond valuation, which become burdensome when the latent factors are not independent, therefore this paper concentrates on three independent factors driving the short interest rate for BBKM. AGTSM and QTSM have closed-form or quasi-closed form solutions for bond prices even when the latent factors are not independent. Unlike in AGTSM, in QTSM and BBKM the conditional variance of yields increases with the level of yields; in this sense yields are "hetero-schedastic".

The main empirical finding is that quadratic models and a Vasicek-type affine model with non-independent factors generally fit US and German bond yields better than BBKM with independent factors. The "non-independence" of the factors generally seems more important than the "non-negativity" of model

predicted yields over the sample period 1999-2011. However BBKM can best fit yields when central bank interest rates are lowest, especially after 2008, despite independence of the factors driving the short rate. When interest rates are lowest, a variant of the Black model fits US and German yields particularly well and the Black-Karasinski model fits US yields particularly well. The paper also tests a new linear-quadratic model whereby the short rate may turn negative, while the central tendency of the short rate is a quadratic non-negative function of Gaussian latent factors. This quadratic model fits German yields particularly well. All models fit German yields better than US yields. All models fit yields for short maturities of one or two years less well than yields of longer maturities.

The paper is organised as follows. The next section reviews the most relevant literature. Then two sections present the theoretical pricing models. Another section illustrates the empirical performance of the models. The conclusions follow.

## 2 Literature

The literature on dynamic term structure models is too vast to be summarised in this paper. A good survey is Dai and Singleton (2003). Here we refer only to Gaussian term structure models that simply rule out arbitrage and abstract from the macro-economy. Vasicek (1977) and Langetieg (1980) first studied affine Gaussian models. Babbs and Nowman (1999) used the Kalman filter to estimate affine Gaussian models. Dai and Singleton (2002) tested general

affine Gaussian models. Nowman (2010) successfully fitted a two factor affine Gaussian model to Euro and UK Sterling yield curves using the Kalman Filter. Joslin et al. (2011) studied affine Gaussian models whereby the factors are observable portfolios of yields. Duffee (2011) used affine Gaussian models to show that yields cannot detect variation in US Government bond risk-premia.

QTSM were studied already in the nineties and then in Ahn et al. (2002), Leippold and Wu (2002, 2003) and Chen et al. (2004). Ahn et al. (2002) illustrated the empirical advantage of general QTSM over affine models due to the unrestricted correlation between the factors driving the short interest rate. Gourieroux and Sufana (2005) and Realdon (2006, 2011) presented discrete time QTSM. Li and Zhao (2006) used a QTSM to provide evidence of un-spanned stochastic volatility in the pricing of interest rate derivatives. Jiang and Yan (2009) provided evidence of "jumps" in the short rate using a linear-quadratic model.

Other notable Gaussian models outside the families of affine or quadratic Gaussian models are those of Black and Karasinski (1991), Black (1995), Peterson et al. (2003). Gorovoi and Linetski (2004) provided an eigenfunction expansion for pricing a discount bond according to Black's model. Using the Japanese term structure of interest rates, Realdon (2009) tested a two factor version of the Black model. Kim and Singleton (2012) tested various non-affine Gaussian term structure models with two latent factors using Japanese Government bond yields. Instead this paper tests multifactor versions of the affine Gaussian model, Black model, Black-Karasinski model and discrete time

quadratic models using US and German Government bond yields.

### 3 Extended Black and Black-Karasinski Models (BBKM)

The paper tests 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

$$r_t = \sum_{i=1}^3 \max(x_{i,t}, 0)^q. \quad (1)$$

$q$  is a constant and will be set equal to 1 or 2. When  $q = 1$  we have a three factor generalisation of Black's (1995) model. We consider the case where  $q = 2$  because unreported tests show good empirical performance in comparison to other cases where  $q$  differs from 2. Given a filtered probability space with the usual properties, we assume

$$dx_{i,t} = \kappa_i (\mu_i - x_{i,t}) dt + \sigma_i dw_{i,t}^{\mathbb{Q}}$$

for  $i = 1, 2, 3$ .  $dx_{i,t}$  is the stochastic differential of the factor  $x_i$  and  $dw_{i,t}^{\mathbb{Q}}$  the stochastic differential of a Wiener process in the risk-neutral measure  $\mathbb{Q}$  over the infinitesimal time interval  $[t, t + dt]$ . The Wiener processes are independent unless otherwise stated, therefore  $dw_{1,t}^{\mathbb{Q}} dw_{2,t}^{\mathbb{Q}} = dw_{1,t}^{\mathbb{Q}} dw_{3,t}^{\mathbb{Q}} = dw_{2,t}^{\mathbb{Q}} dw_{3,t}^{\mathbb{Q}} = 0$ .  $\kappa_i, \sigma_i, \mu_i$  for  $i = 1, 2, 3$  are all constant parameters. Equation 1 implies that  $r_t$

cannot turn negative and bond yields for maturities longer than the instantaneous maturity are guaranteed to be positive, even when  $x_{i,t}$  are negative.

The paper also tests an extension of the Black-Karasinski (1991) model whereby  $r_t = \sum_{i=1}^3 \exp(x_{i,t})$  and a special case of the affine Gaussian model of Langetieg (1980) whereby  $r_t = \sum_{i=1}^3 x_{i,t}$ . We refer to this version of the Langetieg model as the three factor Vasicek model.

Let  $V$  denote the value at time  $t$  of a discount bond with face value 1.  $V(T) = 1$  is the terminal value of the discount bond at maturity  $T$ . Absent arbitrage and assuming for example equation 1, we obtain the pricing equation

$$V = V_1 \cdot V_2 \cdot V_3 \tag{2}$$

$$\frac{\partial V_i}{\partial t} + \frac{\partial^2 V_i}{\partial x_i^2} \frac{1}{2} \sigma_i^2 + \frac{\partial V_i}{\partial x_i} \kappa_i (\mu_i - x_i) - V_i \cdot \max(x_i, 0)^q = 0$$

$$\lim_{x_i \rightarrow -\infty} \frac{\partial^2 V_i}{\partial x_i^2} \rightarrow 0, \quad \lim_{x_i \rightarrow \infty} \frac{\partial^2 V_i}{\partial x_i^2} \rightarrow 0, \quad V_i(T) = 1 \quad \text{for } i = 1, 2, 3.$$

For  $i = 1, 2, 3$ ,  $V_i$  is a function of  $x_i$  and time  $t$ .  $V_i$  tends to be linear in the factor  $x_i$  as that factor tends to plus infinity, in which case  $V_i$  tends 0.  $V_i$  also tends to be linear in the factor  $x_i$  as that factor tends to minus infinity, in which case  $V_i$  tends to 1. Discount bond yields are computed as  $\frac{-\ln V}{T-t}$ , where again  $T$  is the bond maturity date and  $t$  the current date. The fact that the three factors are independent considerably simplifies the numerical solution to the bond pricing equation. Instead of solving for  $V$  on a grid with three "space" dimensions, we simply solve for  $V_1, V_2, V_3$  on three grids, each grid having one "space" dimension. In the empirical tests below, the partial

differential equations for  $V_1, V_2, V_3$  are each solved through the vertical method of lines.

### 3.1 Vertical method of lines (MOL)

Vertical MOL discretises the pricing equation in the "space" dimensions, but not in the "time" dimension. When the pricing equation is linear, vertical MOL reduces to a system of ordinary differential equations (ODE's), which can quickly be solved by computing a matrix exponential. Khaliq, Voss and Yousuf (2007) proposed vertical MOL for option pricing and explained its stability and accuracy. This paper uses vertical MOL as unreported simulations show that vertical MOL is quicker and more accurate than the implicit finite difference method, with no stability problems because the "time" dimension is not discretised. With vertical MOL we compute  $V_1$  on a grid of nodes in the space dimension  $x_1$ . Each node is  $x_{1,j} = j \cdot \delta x_1 + x_{1,0}$  for  $j = 1, 2, \dots, J$  and  $\delta x_1 = \frac{x_{1,J} - x_{1,0}}{J}$ . Therefore the pricing equation satisfied by  $V_1$  is solved on each node over the finite region  $[x_{1,0}, x_{1,J}]$  where  $x_{1,0}$  and  $x_{1,J}$  are respectively the lower and upper boundaries of the solution region in the  $x_1$  dimension. Vertical MOL discretises the pricing equation in the  $x_1$  dimension, but not in the time dimension. We define  $\tau = T - t$ , where again  $T$  is the bond maturity date and  $t$  the current date. Over the interval  $[0, \tau]$  at the grid points  $[x_{1,1}, \dots, x_{1,J}]$  we approximate  $V_1$  with the vector  $\mathbf{u} = (u_1, \dots, u_J)'$  and at the same grid points we



approximate the pricing equation for  $V_1$  with the system

$$\frac{\partial u_j}{\partial \tau} = \frac{u_{j+1} - 2u_j + u_{j-1}}{(\delta x_1)^2} \frac{\sigma_1^2}{2} + \frac{u_{j+1} - u_{j-1}}{2 \cdot \delta x_1} \kappa_1 (\mu_1 - x_{1,j}) - \max(x_{1,j}, 0)^q u_j \quad \text{for } j = 1, 2, \dots, J. \quad (3)$$

We can rewrite this system of ordinary differential equations as

$$\frac{\partial \mathbf{u}}{\partial \tau} = \mathbb{M} \cdot \mathbf{u} \quad (4)$$

$$\mathbf{u}(\tau = 0) = \mathbf{1}_J$$

$$\frac{\partial \mathbf{u}}{\partial \tau} = \begin{pmatrix} \frac{\partial u_1}{\partial \tau} \\ \dots \\ \frac{\partial u_J}{\partial \tau} \end{pmatrix} \quad (5)$$

where  $\mathbf{1}_{J \times 1}$  is a  $J \times 1$  vector whose elements are all equal to 1 and

$$\mathbb{M} = \begin{pmatrix} 2\mathbb{A}_1 + \mathbb{B}_1 & \mathbb{C}_1 - \mathbb{A}_1 & 0 & \dots & 0 & 0 & 0 \\ \mathbb{A}_2 & \mathbb{B}_2 & \mathbb{C}_2 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbb{A}_{J-1} & \mathbb{B}_{J-1} & \mathbb{C}_{J-1} \\ 0 & 0 & 0 & \dots & 0 & \mathbb{A}_J - \mathbb{C}_J & \mathbb{B}_J + 2\mathbb{C}_J \end{pmatrix}$$

$$\mathbb{A}_j = \frac{1}{2} \left( \left( \frac{\sigma_1}{\delta x_1} \right)^2 - \frac{\kappa_1 (\mu_1 - x_{1,j})}{\delta x_1} \right)$$

$$\mathbb{B}_j = -\max(x_{1,j}, 0)^q - \left( \frac{\sigma_1}{\delta x_1} \right)^2$$

$$\mathbb{C}_j = \frac{1}{2} \left( \left( \frac{\sigma_1}{\delta x_1} \right)^2 + \frac{\kappa_1 (\mu_1 - x_{1,j})}{\delta x_1} \right).$$

The solution to system 4 is

$$\mathbf{u}(\tau) = \exp(\tau \cdot \mathbb{M}).$$

This matrix exponential can be computed very quickly, for example with Matlab, which employs the Padé approximation of Higham (2005). Unreported numerical exercises for the Black1 model with one stochastic factor showed that, using parameters similar to those estimated in the empirical tests, the implicit finite difference solution to the partial differential equation for Black1 approaches the vertical MOL solution as the number of time steps per year is increased. This implies that the error of the finite difference solution due to time discretisation is almost absent from the vertical MOL solution, as the Padé approximation in Matlab is extremely accurate to compute a matrix exponential. For example, as  $J = 200$ ,  $x_{1,0} = -1$ ,  $x_{1,J} = 1$ , the difference on the same nodes between the implicit finite difference solution with 2000 time steps per year and the vertical MOL solution is typically less than 1 basis point of a discount bond yield; such is the difference between the two solutions on most of the nodes and for most bond maturities up to 10 years. The implicit finite difference method and vertical MOL employ similar finite differences to approximate the derivatives in the "space" dimension.

### 3.2 Vasicek with correlated factors

The empirical tests compare BBKM with a three factor Vasicek-type model with correlated factors, such that

$$r_t = \sum_{i=1}^3 x_{i,t}$$

$$dx_{i,t} = \kappa_i (\mu_i - x_{i,t}) dt + \sigma_i dw_{i,t}^{\mathbb{Q}}$$

for  $i = 1, 2, 3$  and  $dw_{1,t}^{\mathbb{Q}} dw_{2,t}^{\mathbb{Q}} = \rho_{12} \cdot dt$ ,  $dw_{1,t}^{\mathbb{Q}} dw_{3,t}^{\mathbb{Q}} = \rho_{13} \cdot dt$ ,  $dw_{2,t}^{\mathbb{Q}} dw_{3,t}^{\mathbb{Q}} = \rho_{23} \cdot dt$ .  $\rho_{12}, \rho_{13}, \rho_{23}$  are correlation parameters. For this model the value of a discount bond is

$$V = e^{\mathfrak{A} + \sum_{i=1}^3 \mathfrak{B}_i x_{i,t}}$$

$$\mathfrak{B}_i = \frac{e^{-\kappa_i(T-t)} - 1}{\kappa_i}$$

and

$$\begin{aligned} \mathfrak{A} = & \frac{\rho_{12}\sigma_1\sigma_2}{2 \cdot \kappa_1\kappa_2} \left( \frac{1 - e^{-(\kappa_1+\kappa_2)(T-t)}}{\kappa_1 + \kappa_2} - \frac{1 - e^{-\kappa_2(T-t)}}{\kappa_2} - \frac{1 - e^{-\kappa_1(T-t)}}{\kappa_1} + T - t \right) + \\ & + \frac{\rho_{13}\sigma_1\sigma_3}{2 \cdot \kappa_1\kappa_3} \left( \frac{1 - e^{-(\kappa_1+\kappa_3)(T-t)}}{\kappa_1 + \kappa_3} - \frac{1 - e^{-\kappa_3(T-t)}}{\kappa_3} - \frac{1 - e^{-\kappa_1(T-t)}}{\kappa_1} + T - t \right) + \\ & + \frac{\rho_{23}\sigma_2\sigma_3}{2 \cdot \kappa_2\kappa_3} \left( \frac{1 - e^{-(\kappa_2+\kappa_3)(T-t)}}{\kappa_2 + \kappa_3} - \frac{1 - e^{-\kappa_3(T-t)}}{\kappa_3} - \frac{1 - e^{-\kappa_2(T-t)}}{\kappa_2} + T - t \right) + \\ & + \sum_{i=1}^3 (\mathfrak{B}_i + T - t) \left( \frac{\sigma_i^2}{2\kappa_i^2} - \mu_i \right) - \frac{\sigma_i^2}{4\kappa_i} \mathfrak{B}_i^2. \end{aligned}$$

Hereafter this model is referred to as "Vasicek correlated", while the special case whereby  $\rho_{12} = \rho_{13} = \rho_{23} = 0$  is referred to as "Vasicek".

### 3.3 Processes in the real measure and Kalman Filter

For all the above models we assume that in the real probability measure

$$dx_{i,t} = \kappa_i^* (\mu_i^* - x_{i,t}) dt + \sigma_i dw_{i,t} \quad (6)$$

for  $i = 1, 2, 3$ , where  $dw_{i,t}$  are differentials of Wiener processes in the real measure and  $dw_{1,t}dw_{2,t} = \rho_{12} \cdot dt$ ,  $dw_{1,t}dw_{3,t} = \rho_{13} \cdot dt$ ,  $dw_{2,t}dw_{3,t} = \rho_{23} \cdot dt$ . For all models except "Vasicek correlated" the correlation parameters are zero, i.e.  $\rho_{12} = \rho_{13} = \rho_{23} = 0$ .

Let  $t = 1, 2, \dots, M$ , denote the set of  $M$  dates on which we observe Government discount bond yields.  $x_{1,t}, x_{2,t}, x_{3,t}$  denote the values of the three latent factors on day  $t$ .  $\Delta$  is the time between consecutive observations and is approximately equal to one divided by the number trading days in one year. Therefore  $\Delta = \frac{1}{261}$  since we observe about 261 daily prices per year in the data. The number of trading days in one year varies between 260 and 262. Let  $l(\mathbf{x}_t | \mathbf{x}_{t-1})$  denote the real measure conditional transition density of  $\mathbf{x}_t = (x_{1,t}, x_{2,t}, x_{3,t})'$  given

$\mathbf{x}_{t-1} = (x_{1,t-1}, x_{2,t-1}, x_{3,t-1})'$ . It can be shown that

$$l(\mathbf{x}_t | \mathbf{x}_{t-1}) \sim N(\phi^* \mu^* + (\mathbf{I} - \phi^*) \mathbf{x}_{t-1}, \Sigma \Sigma') \quad (7)$$

$$\phi^* = \mathbf{I}_3 - \begin{pmatrix} e^{-\kappa_1^* \Delta} & 0 & 0 \\ 0 & e^{-\kappa_2^* \Delta} & 0 \\ 0 & 0 & e^{-\kappa_3^* \Delta} \end{pmatrix}, \quad \mu^* = \begin{pmatrix} \mu_1^* \\ \mu_2^* \\ \mu_3^* \end{pmatrix},$$

$$\Sigma \Sigma' = \begin{pmatrix} \frac{\sigma_1^2}{2\kappa_1^*} (1 - e^{-2\kappa_1^* \Delta}) & \frac{\rho_{12}\sigma_1\sigma_2}{\kappa_1^* + \kappa_2^*} (1 - e^{-(\kappa_1^* + \kappa_2^*)\Delta}) & \frac{\rho_{13}\sigma_1\sigma_3}{\kappa_1^* + \kappa_3^*} (1 - e^{-(\kappa_1^* + \kappa_3^*)\Delta}) \\ \frac{\rho_{12}\sigma_1\sigma_2}{\kappa_1^* + \kappa_2^*} (1 - e^{-(\kappa_1^* + \kappa_2^*)\Delta}) & \frac{\sigma_2^2}{2\kappa_2^*} (1 - e^{-2\kappa_2^* \Delta}) & \frac{\rho_{23}\sigma_2\sigma_3}{\kappa_2^* + \kappa_3^*} (1 - e^{-(\kappa_2^* + \kappa_3^*)\Delta}) \\ \frac{\rho_{13}\sigma_1\sigma_3}{\kappa_1^* + \kappa_3^*} (1 - e^{-(\kappa_1^* + \kappa_3^*)\Delta}) & \frac{\rho_{23}\sigma_2\sigma_3}{\kappa_2^* + \kappa_3^*} (1 - e^{-(\kappa_2^* + \kappa_3^*)\Delta}) & \frac{\sigma_3^2}{2\kappa_3^*} (1 - e^{-2\kappa_3^* \Delta}) \end{pmatrix}.$$

$\mathbf{I}_3$  is the  $3 \times 3$  identity matrix.  $N(\phi^* \mu^* + (\mathbf{I} - \phi^*) \mathbf{x}_{t-1}, \Sigma \Sigma')$  is the multivariate normal density with mean  $\phi^* \mu^* + (\mathbf{I} - \phi^*) \mathbf{x}_{t-1}$  and covariance  $\Sigma \Sigma'$ . The empirical tests use the Extended Kalman filter (EKF) to estimate the models on US and German yields. Two are the main reasons why EKF is used instead of Maximum likelihood (ML) estimation. The first reason is that ML requires "inferring" the latent factors from the observed yields on any observation date. This is complicated to do for BBKM and quadratic models with three latent factors and also requires the arbitrary assumption that some of the yields be observed without error. EKF estimation does not require this assumption, since it assumes that all yields are observed with error. The second reason to use EKF is that the Gaussian processes of the latent factors make EKF suitable for estimation. Details about EKF are provided in the Appendix. EKF is also used to estimate the linear quadratic models hereunder.

## 4 Discrete time linear-quadratic pricing models

This section illustrates the linear-quadratic model in discrete time that will also be tested on US and German yields. We employ the discrete time version rather than the continuous time version of the linear-quadratic model because of fewer constraints to model parameters as explained below. Moreover in discrete time the conditional transition density of the factors, as opposed to the continuous time transition density, remains Gaussian, which simplifies EKF estimation. Again we set each time step equal to  $\Delta = \frac{1}{261}$  and employ the following definitions:

- $V_{n,t}$  is the time  $t$  value of a zero coupon bond with  $n$  trading days to maturity, thus the bond matures on trading day  $t + n$ ;
- $r_t$  is the continuously compounded risk-free interest rate for one trading day observed on day  $t$ , such that

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

The no-arbitrage risk-neutral valuation equation is

$$V_{n,t} = E_t^{\mathbb{Q}} [e^{-\Delta \cdot r_t} \cdot V_{n-1,t+1}] \quad (8)$$

where  $E_t^{\mathbb{Q}} [..]$  denotes conditional expectation on day  $t$  under the risk-neutral

measure  $\mathbb{Q}$ . Following Realdon (2011) we further assume that

$$\mathbf{r}_t = \mathbf{x}_t' \boldsymbol{\Psi} \mathbf{x}_t + \delta y_t \quad (9)$$

$$\mathbf{x}_t = (x_{1,t}, \dots, x_{m,t})' \quad (10)$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \phi(\boldsymbol{\mu} - \mathbf{x}_t) + \boldsymbol{\Sigma} \xi_{t+1}^{\mathbb{Q}} \quad (11)$$

$$\mathbf{x}_{t+1} - \mathbf{x}_t = \phi^*(\boldsymbol{\mu}^* - \mathbf{x}_t) + \boldsymbol{\Sigma} \xi_{t+1} \quad (12)$$

$$y_{t+1} - y_t = \phi_y(\mu_y + \mathbf{x}_t' \mathbf{L} \mathbf{x}_t - y_t) + (\boldsymbol{\Sigma}'_{yx}, \Sigma_y) \begin{pmatrix} \xi_{t+1}^{\mathbb{Q}} \\ \varepsilon_{y,t+1}^{\mathbb{Q}} \end{pmatrix} \quad (13)$$

$$y_{t+1} - y_t = \phi_y^*(\mu_y^* + \mathbf{x}_t' \mathbf{L} \mathbf{x}_t - y_t) + (\boldsymbol{\Sigma}'_{yx}, \Sigma_y) \begin{pmatrix} \xi_{t+1} \\ \varepsilon_{y,t+1} \end{pmatrix} \quad (14)$$

$$\left( \xi_{t+1}^{\mathbb{Q}}, \varepsilon_{y,t+1}^{\mathbb{Q}} \right)' \sim N(\mathbf{0}_{m+1}, \mathbf{I}_{m+1}) \quad (15)$$

$$\left( \xi_{t+1}', \varepsilon_{y,t+1}' \right)' \sim N(\mathbf{0}_{m+1}, \mathbf{I}_{m+1}) \quad (16)$$

$$\boldsymbol{\Sigma} = \mathbf{S} \sqrt{\Delta}, \quad \boldsymbol{\Sigma}_{yx} = \mathbf{S}_{yx} \sqrt{\Delta}, \quad \mathbf{S}_{yx} = (\sigma_{yx,1}, \dots, \sigma_{yx,m})', \quad \Sigma_y = s_y \sqrt{\Delta} \quad (17)$$

$$\phi = \Delta \cdot \kappa, \quad \phi_y = \Delta \cdot \kappa_y, \quad \phi = \Delta \cdot \kappa^*, \quad \phi_y = \Delta \cdot \kappa_y^*, \quad \mathbf{L} = \Delta \cdot \mathbf{1}_{m \times m} \quad (18)$$

$$V_{n,t} = \exp(A_n + \mathbf{B}'_n \mathbf{x}_t + \mathbf{x}_t' \mathbf{C}_n \mathbf{x}_t + D_n y_t). \quad (19)$$

$\mathbf{x}_t, \mu, \mu^*, \xi_{t+1}^{\mathbb{Q}}, \xi_{t+1}, \mathbf{B}_n, \boldsymbol{\Sigma}_{yx}, \mathbf{S}_{yx}$  are  $m \times 1$  vectors.  $\boldsymbol{\Psi}, \phi, \phi^*, \kappa, \kappa^*, \mathbf{C}_n, \boldsymbol{\Sigma}, \mathbf{S}, \mathbf{L}, \mathbf{1}_{m \times m}$  are  $m \times m$  matrices. In particular  $\mathbf{1}_{m \times m}$  is an  $m \times m$  matrix whose elements are all equal to 1.  $\mathbf{r}_t, A_n, A_0, y_t, \mu_y, \phi_y, \kappa_y, \kappa_y^*, \varepsilon_{y,t+1}^{\mathbb{Q}}, \varepsilon_{y,t+1}, D_n, \Sigma_y, s_y$  are scalars.  $N(\mathbf{0}_{m+1}, \mathbf{I}_{m+1})$  denotes the multivariate normal density with mean  $\mathbf{0}_{m+1}$  and covariance matrix  $\mathbf{I}_{m+1}$ .  $\mathbf{0}_{m+1}$  is a  $(m+1) \times 1$  vector of zeros.  $\mathbf{I}_{m+1}$  is the  $(m+1) \times (m+1)$  identity matrix, while  $\xi_{t+1}^{\mathbb{Q}} = \left( \varepsilon_{1,t+1}^{\mathbb{Q}}, \dots, \varepsilon_{m,t+1}^{\mathbb{Q}} \right)'$

and  $\xi_{t+1} = (\varepsilon_{1,t+1}, \dots, \varepsilon_{m,t+1})'$ .  $\varepsilon_{1,t+1}, \dots, \varepsilon_{m,t+1}$  and  $\varepsilon_{1,t+1}^{\mathbb{Q}}, \dots, \varepsilon_{m,t+1}^{\mathbb{Q}}$  are scalar Gaussian random shocks respectively in the real and risk-neutral measures.  $\mu, \mu^*, \phi, \phi^*, \kappa, \kappa^*, \Sigma, \mathbf{S}, \mu_y, \phi_y, \kappa_y, \kappa_y^*, \delta, \Sigma_{yx}, \mathbf{S}_{yx}, \Sigma_y, s_y$  are parameters. The factor processes  $\mathbf{x}$  and  $y$  are specified under both the real measure and the risk-neutral measure  $\mathbb{Q}$ .  $\mathbf{x}$  follows a Gaussian auto-regressive process, where the random terms  $\xi_{t+1}^{\mathbb{Q}}$  in the risk-neutral measure are normally distributed with mean of  $\mathbf{0}_{m \times 1}$  and covariance  $\mathbf{I}_m$ . The time  $t$  conditional covariance matrix of  $(\mathbf{x}_{t+1} - \mathbf{x}_t)$  is  $\Sigma \Sigma'$ . The discount bond value  $V_{n,t}$  is exponential linear in  $y_t$  since the short rate  $r_t$  is linear in  $y_t$ . The conditional mean of  $y_{t+1}$  is quadratic in  $\mathbf{x}_t$ , which causes the discount bond value  $V_{n,t}$  to be exponential quadratic in  $\mathbf{x}_t$ . The conditional covariance of  $(\mathbf{x}'_{t+1}, y'_{t+1})'$  is

$$E_t^{\mathbb{Q}} \left[ \begin{pmatrix} \mathbf{x}_{t+1} \\ y_{t+1} \end{pmatrix} \cdot (\mathbf{x}'_{t+1}, y_{t+1}) \right] = \begin{bmatrix} \Sigma \Sigma' & \Sigma \Sigma_{yx} \\ \Sigma'_{yx} \Sigma' & \Sigma_y \Sigma_y \end{bmatrix}. \quad (20)$$

This discrete time linear-quadratic model is a special case of Realdon (2011), who shows that we can recursively compute  $A_n, \mathbf{B}_n, \mathbf{C}_n, D_n$  appearing in 19 by solving the following system of Riccati difference equations:

$$A_n = A_{n-1} + \mathbf{B}'_{n-1} \phi \mu + D_{n-1} \phi_y \mu_y + (\phi \mu)' \mathbf{C}_{n-1} \phi \mu + \ln \frac{|\gamma|}{\text{abs}|\Sigma|} + \quad (21)$$

$$+ \frac{1}{2} (\mathbf{G}_{n-1} + D_{n-1} \Sigma'_{yx} \Sigma^{-1}) \gamma \gamma' (\mathbf{G}_{n-1} + D'_{n-1} \Sigma'_{yx} \Sigma^{-1})' + \frac{1}{2} D_{n-1} \Sigma_y \Sigma'_y D_{n-1}$$



$$\mathbf{B}'_n = \mathbf{B}'_{n-1}(\mathbf{I}_2 - \phi) + 2(\phi\mu)' \mathbf{C}_{n-1}(\mathbf{I}_2 - \phi) + 2(\mathbf{G}_{n-1} + D_{n-1}\boldsymbol{\Sigma}'_{yx}\boldsymbol{\Sigma}^{-1})\gamma\gamma' \mathbf{C}'_{n-1}(\mathbf{I} - \phi) \quad (22)$$

$$\mathbf{C}_n = -\Delta\boldsymbol{\Psi} + (\mathbf{I}_2 - \phi)' \mathbf{C}_{n-1}(\mathbf{I}_2 - \phi) + D_{n-1}\mathbf{L} + 2(\mathbf{I}_2 - \phi)' \mathbf{C}_{n-1}\gamma\gamma' \mathbf{C}'_{n-1}(\mathbf{I}_2 - \phi) \quad (23)$$

$$D_n = -\Delta\delta + D_{n-1}(1 - \phi_y) \quad (24)$$

with  $\mathbf{G}_{n-1} = \mathbf{B}'_{n-1} + 2(\phi\mu)' \mathbf{C}_{n-1}$  and  $\gamma = \left( (\boldsymbol{\Sigma}\boldsymbol{\Sigma}')^{-1} - 2\mathbf{C}_{n-1} \right)^{-1/2}$ . On day  $t$  the one day yield is

$$\tau_t = \frac{-\ln V_{1,t}}{\Delta} = -A_1 - \mathbf{B}'_1 \cdot \mathbf{x}_t - \mathbf{x}'_t \mathbf{C}_1 \mathbf{x}_t - D_1 y_t = \mathbf{x}'_t \boldsymbol{\Psi} \mathbf{x}_t + \delta y_t \quad (25)$$

since  $A_1 = 0$ ,  $B_1 = \mathbf{0}_{m \times 1}$ ,  $C_1 = -\Delta\boldsymbol{\Psi}$  and  $D_1 = -\Delta\delta$ .  $\tau_t \geq 0$  if  $\boldsymbol{\Psi}$  is symmetric and  $\delta = 0$ . When the stochastic factors are latent, parameter identification restrictions are needed. In this paper we:

- either set  $\delta = 1$  and  $\boldsymbol{\Psi} = \mathbf{0}_{m \times m}$ , where  $\mathbf{0}_{m \times m}$  is an  $m \times m$  matrix of zeros; hereafter we denote this model specification as DTQM1;
- or set  $\delta = 0$  and  $\boldsymbol{\Psi} = \mathbf{I}_m$ , which makes the model a pure quadratic model; hereafter this model specification is denoted as DTQM2 (or DTQM3 when all the latent factors are independent).

In DTQM1  $\tau_t$  may turn negative, unlike in DTQM2 and DTQM3.

## 4.1 DTQM1

The empirical tests below consider DTQM1 where  $\delta = 1$ ,  $\Psi = \mathbf{0}_{m \times m}$  and  $m = 2$

so that  $\mathbf{r}_t = y_t$ , and where

$$\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \end{pmatrix} = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} + \begin{pmatrix} \kappa_{1,1} & 0 \\ \kappa_{2,1} & \kappa_{2,2} \end{pmatrix} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} \right) \Delta + \mathbf{S} \cdot \begin{pmatrix} \varepsilon_{1,t+1}^{\mathbb{Q}} \\ \varepsilon_{2,t+1}^{\mathbb{Q}} \end{pmatrix} \sqrt{\Delta}$$

$$\mathbf{S} = \begin{pmatrix} \sigma_1 & 0 \\ \rho_{12} \cdot \sigma_2 & \sqrt{1 - \rho_{12}^2} \cdot \sigma_2 \end{pmatrix}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \mu^* = \begin{pmatrix} \mu_1^* \\ \mu_2^* \end{pmatrix}, \kappa = \begin{pmatrix} \kappa_{1,1} & 0 \\ \kappa_{2,1} & \kappa_{2,2} \end{pmatrix}, \kappa^* = \begin{pmatrix} \kappa_{1,1}^* & 0 \\ \kappa_{2,1}^* & \kappa_{2,2}^* \end{pmatrix}$$

$$y_{t+1} = y_t + \kappa_y \left( \mu_y + (x_{1,t} + x_{2,t})^2 - y_t \right) \Delta + \begin{pmatrix} \sigma_{yx,1} & \sigma_{yx,2} & s_y \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t+1}^{\mathbb{Q}} \\ \varepsilon_{2,t+1}^{\mathbb{Q}} \\ \varepsilon_{y,t+1}^{\mathbb{Q}} \end{pmatrix} \sqrt{\Delta}.$$

$\mu_1, \mu_2, \mu_y, \kappa_{1,1}, \kappa_{2,1}, \kappa_{2,2}, \kappa_y$  and  $\sigma_{yx,1}, \sigma_{yx,2}, s_y$  are parameters and  $\varepsilon_{1,t}^{\mathbb{Q}}, \varepsilon_{2,t}^{\mathbb{Q}}, \varepsilon_{y,t}^{\mathbb{Q}}$  are the random shocks.  $\rho_{12}$  is the conditional correlation between  $x_{1,t+1}$  and  $x_{2,t+1}$ ,  $\rho_{1y}$  between  $x_{1,t+1}$  and  $y_{t+1}$ ,  $\rho_{2y}$  between  $x_{2,t+1}$  and  $y_{t+1}$ .  $\sigma_1, \sigma_2, \sigma_y$  are volatility parameters.  $\kappa_{2,1}$  is the element of the second row and first column of  $\kappa$ . The other indexed parameters have similar interpretation. DTQM1 is of interest since only the factor  $y_t$  drives the short interest rate  $\mathbf{r}_t$ , while  $x_{1,t}$  and  $x_{2,t}$  only drive the "drift" of the short rate. Thus the model can match very low and even negative short term yields and at the same time also the relatively

higher long term yields. Short term yields are mainly driven by  $y_t$  while longer term yields are also driven by  $x_{1,t}$  and  $x_{2,t}$ . Thus short term and long term yields can move quite independently according to DTQM1.  $\mathbf{r}_t$  tends to revert toward the level  $\mu_y + (x_{1,t} + x_{2,t})^2$ , therefore long term yields tend to be positive when  $\mu_y > 0$ . We also assume, without loss of generality, that

$$(\sigma_{yx,1}, \sigma_{yx,2}, s_y) = \left( \rho_{1y} \cdot \sigma_y, \frac{\rho_{2y} - \rho_{12} \cdot \rho_{1y}}{\sqrt{1 - \rho_{12}^2}} \cdot \sigma_y, \sqrt{1 - \rho_{1y}^2 - \frac{(\rho_{2y} - \rho_{12} \cdot \rho_{1y})^2}{1 - \rho_{12}^2}} \cdot \sigma_y \right).$$

The Appendix explains that the parameters of DTQM1 are identifiable.

## 4.2 DTQM2 and DTQM3

DTQM2 is a special case of the linear-quadratic model where  $\delta = 0$  and  $\Psi = \mathbf{I}_3$ , so that  $\mathbf{r}_t = x_{1,t}^2 + x_{2,t}^2 + x_{3,t}^2$ . Therefore DTQM2 is a three factor quadratic model that does not depend on  $y_t$ . DTQM2 also assumes

$$\begin{aligned}
\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \\ x_{3,t+1} \end{pmatrix} &= \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} + \begin{pmatrix} \kappa_{1,1} & 0 & 0 \\ \kappa_{2,1} & \kappa_{2,2} & 0 \\ \kappa_{3,1} & \kappa_{3,2} & \kappa_{3,3} \end{pmatrix} \left( \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} - \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix} \right) \Delta + \mathbf{S} \cdot \begin{pmatrix} \varepsilon_{1,t+1}^{\mathbb{Q}} \\ \varepsilon_{2,t+1}^{\mathbb{Q}} \\ \varepsilon_{3,t+1}^{\mathbb{Q}} \end{pmatrix} \sqrt{\Delta} \\
\mathbf{S} &= \begin{pmatrix} \sigma_1 & 0 & 0 \\ \rho_{12} \cdot \sigma_2 & \sqrt{1 - \rho_{12}^2} \cdot \sigma_2 & 0 \\ \rho_{13} \cdot \sigma_3 & \frac{\rho_{32} - \rho_{12} \cdot \rho_{13}}{\sqrt{1 - \rho_{12}^2}} \cdot \sigma_3 & \sqrt{1 - \rho_{13}^2 - \frac{(\rho_{32} - \rho_{12} \cdot \rho_{13})^2}{1 - \rho_{12}^2}} \cdot \sigma_3 \end{pmatrix} \\
\mu &= \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \quad \mu^* = \begin{pmatrix} \mu_1^* \\ \mu_2^* \\ \mu_3^* \end{pmatrix}, \quad \kappa = \begin{pmatrix} \kappa_{1,1} & 0 & 0 \\ \kappa_{2,1} & \kappa_{2,2} & 0 \\ \kappa_{3,1} & \kappa_{3,2} & \kappa_{3,3} \end{pmatrix}, \quad \kappa^* = \begin{pmatrix} \kappa_{1,1}^* & 0 & 0 \\ \kappa_{2,1}^* & \kappa_{2,2}^* & 0 \\ \kappa_{3,1}^* & \kappa_{3,2}^* & \kappa_{3,3}^* \end{pmatrix}.
\end{aligned}$$

$\rho_{12}$  is the conditional correlation between  $x_{1,t+1}$  and  $x_{2,t+1}$ ,  $\rho_{13}$  between  $x_{1,t+1}$  and  $x_{3,t+1}$ ,  $\rho_{23}$  between  $x_{2,t+1}$  and  $x_{3,t+1}$ .  $\sigma_1, \sigma_2, \sigma_3$  are volatility parameters. The value of a discount bond according to DTQM2 appeared in Redalton (2006) and is a special case of the above linear-quadratic model, i.e.  $V_{n,t} = \exp(A_n + \mathbf{B}'_n \mathbf{x}_t + \mathbf{x}'_t \mathbf{C}_n \mathbf{x}_t)$  with

$$A_n = A_{n-1} + \mathbf{B}'_{n-1} \phi \mu + (\phi \mu)' \mathbf{C}_{n-1} \phi \mu + \ln \frac{|\gamma|}{\text{abs}|\boldsymbol{\Sigma}|} + \frac{1}{2} \mathbf{G}_{n-1} \gamma \gamma' \mathbf{G}'_{n-1} \quad (26)$$

$$\mathbf{B}'_n = \mathbf{B}'_{n-1} (\mathbf{I}_3 - \phi) + 2 (\phi \mu)' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) + 2 \cdot \mathbf{G}_{n-1} \gamma \gamma' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) \quad (27)$$

$$\mathbf{C}_n = -\Delta\Psi + (\mathbf{I}_3 - \phi)' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) + 2 \cdot (\mathbf{I}_3 - \phi)' \mathbf{C}_{n-1} \gamma \gamma' \mathbf{C}'_{n-1} (\mathbf{I}_3 - \phi). \quad (28)$$

The quadratic model canonical form in Ahn, Dittmar and Gallant (2002) requires that  $\Psi = \mathbf{I}_3$ ,  $\kappa \cdot \mu \geq 0$ ,  $\kappa^* \cdot \mu^* \geq 0$ ,  $\mathbf{S}$  be diagonal and  $\kappa, \kappa^*$  be triangular. DTQM2 imposes these same conditions as in Ahn-Dittmar-Gallant, but with  $\mathbf{S}$  lower triangular. Therefore, unlike in Ahn-Dittmar-Gallant, in DTQM2 both  $\kappa$  and  $\mathbf{S}$  are lower triangular at the same time, rather than just  $\mathbf{S}$  (when  $\kappa$  is diagonal) or just  $\kappa$  (when  $\mathbf{S}$  is diagonal). This greater freedom is due to the fact that the quadratic model in this paper is in discrete time, rather than in continuous time as in Ahn-Dittmar-Gallant, and that the conditional covariance of  $\mathbf{x}_{t+1}$  in discrete time only depends on  $\Sigma$  and not on  $\phi$ . Other things as in DTQM2, DTQM3 assumes that the three factors are independent, so that

$$\kappa = \begin{pmatrix} \kappa_{1,1} & 0 & 0 \\ 0 & \kappa_{2,2} & 0 \\ 0 & 0 & \kappa_{3,3} \end{pmatrix}, \quad \kappa^* = \begin{pmatrix} \kappa_{1,1}^* & 0 & 0 \\ 0 & \kappa_{2,2}^* & 0 \\ 0 & 0 & \kappa_{3,3}^* \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}.$$

The Appendix proves that the parameters of DTQM2 and DTQM3 are identifiable in estimation.

### 4.3 The continuous time limit

As we observe about 261 trading days per year, when estimating the above discrete time model we set  $\Delta = \frac{1}{261}$ . Instead, if  $\Delta \rightarrow 0$  then  $\mathbf{x}_{t+\Delta} - \mathbf{x}_t$  tends to  $d\mathbf{x}_t$  and  $y_{t+\Delta} - y_t$  tends to  $dy_t$  such that

$$\begin{aligned} d\mathbf{x}_t &= \kappa \cdot (\mu - \mathbf{x}_t) \cdot dt + \mathbf{S} \cdot d\mathbf{w}_t^{\mathbb{Q}} \\ d\mathbf{x}_t &= \kappa^* \cdot (\mu^* - \mathbf{x}_t) \cdot dt + \mathbf{S} \cdot d\mathbf{w}_t \\ dy_t &= \kappa_y \cdot (\mu_y + \mathbf{x}'_t \mathbf{1}_{m \times m} \mathbf{x}_t - y_t) \cdot dt + (\mathbf{S}_{yx}, s_y) \cdot \begin{pmatrix} d\mathbf{w}_t^{\mathbb{Q}} \\ dw_{y,t}^{\mathbb{Q}} \end{pmatrix} \\ dy_t &= \kappa_y^* \cdot (\mu_y^* + \mathbf{x}'_t \mathbf{1}_{m \times m} \mathbf{x}_t - y_t) \cdot dt + (\mathbf{S}_{yx}, s_y) \cdot \begin{pmatrix} d\mathbf{w}_t \\ dw_{y,t} \end{pmatrix} \end{aligned}$$

where  $d\mathbf{x}_t$  is a  $m \times 1$  column vectors of stochastic differentials,  $d\mathbf{w}_t^{\mathbb{Q}}$  and  $d\mathbf{w}_t$  are  $m \times 1$  column vectors of stochastic differentials of independent Wiener processes under the risk-neutral measure  $\mathbb{Q}$  and under the real measure respectively. Similarly  $dy_t$  is a stochastic differential, while  $dw_{y,t}^{\mathbb{Q}}$  and  $dw_{y,t}$  are stochastic differentials of a Wiener process under the risk-neutral and real measures. Thus  $d\mathbf{x}_t$  follows a vector Ornstein-Uhlenbeck process. Continuous time quadratic models are special cases of discrete time quadratic models as  $\Delta \rightarrow 0$ . The Appendix shows that as  $\Delta \rightarrow 0$  the discrete time quadratic model of equations 26, 27, 28 tends to the continuous time model of Ahn, Dittmaer and Gallant (2002) whereby a system of Riccati ordinary differential equations needs solving.

## 5 Empirical tests

This section empirically tests the models illustrated above. The data consists of 3,188 daily observations of the discount functions derived from German Government bonds and US Government bonds over the period 1/1/1999-22/3/2011. The data is provided by Datastream. For any observation date the continuously compounded discount bond yields are computed from the discount function for the maturities of one, two, three, four, five, six, seven, eight, nine and ten years. Table 1 provides summary statistics of these yields in the sample period.

[Table 1 about here]

We estimate and test the following three factor Gaussian models:

- the first is the "Black-Karasinski" model (BK) whereby  $r_t = \exp(x_{1,t}) + \exp(x_{2,t}) + \exp(x_{3,t})$ ; for BK the MOL solution region in each space dimension is  $[x_{i,1}, x_{i,J}]$  for  $i = 1, 2, 3$ , with  $J = 200$ ,  $x_{i,1} = -(20 - 0.1)$  and  $x_{i,J} = 0$ ;

- the second is "Black1" and assumes  $r_t = \max(x_{1,t}, 0) + \max(x_{2,t}, 0) + \max(x_{3,t}, 0)$ ; the single factor version of this model was first proposed by Black (1995); for Black1 the MOL solution region in each space dimension is  $[x_{i,1}, x_{i,J}]$  for  $i = 1, 2, 3$ , with  $J = 200$ ,  $x_{i,1} = -(1 - 0.01)$  and  $x_{i,J} = 1$ ;

- the third is "Black2" and assumes  $r_t = \max(x_{1,t}, 0)^2 + \max(x_{2,t}, 0)^2 + \max(x_{3,t}, 0)^2$ ; for Black2 the MOL solution region in each space dimension is  $[x_{i,1}, x_{i,J}]$  for  $i = 1, 2, 3$ , with  $J = 200$ ,  $x_{i,1} = -(1 - 0.01)$  and  $x_{i,J} = 1$ ;

- the fourth is "Vasicek", with  $r_t = x_{1,t} + x_{2,t} + x_{3,t}$ ; for this model we have closed form solutions for bond prices and impose the restrictions  $\mu_1 = \mu_2 = \mu_3$ ,

$$\mu_1^* = \mu_2^* = \mu_3^*;$$

- the fifth is "Vasicek correlated", with  $r_t = x_{1,t} + x_{2,t} + x_{3,t}$ ; also for this model we have closed form solutions for bond prices and impose the restrictions

$$\mu_1 = \mu_2 = \mu_3, \mu_1^* = \mu_2^* = \mu_3^*;$$

- the sixth is a "Mixed 1" model whereby  $r_t = \max(x_{1,t}, 0) + x_{2,t} + x_{3,t}$  and  $\mu_2 = \mu_3 = \mu_2^* = \mu_3^* = 0$ ; for Mixed 1 the MOL solution region in the  $x_1$  space dimension is  $[x_{1,1}, x_{1,J}]$ , with  $J = 200$ ,  $x_{1,1} = -(1 - 0.01)$  and  $x_{1,J} = 1$ ; the solution for "Mixed 1" is partly numerical through MOL and partly exploits the Vasicek formulae;

- the seventh is a "Mixed 2" model whereby  $r_t = \max(x_{1,t}, 0) + \max(x_{2,t}, 0) + x_{3,t}$  and  $\mu_3 = \mu_3^* = 0$ ; for Mixed 2 the MOL solution region in the  $x_1$  and  $x_2$  space dimensions are  $[x_{1,1}, x_{1,J}]$  and  $[x_{2,1}, x_{2,J}]$ , with  $J = 200$ ,  $x_{1,1} = x_{2,1} = -(1 - 0.01)$  and  $x_{1,J} = x_{2,J} = 1$ ; the solution for "Mixed 2" is partly numerical through MOL and partly exploits the Vasicek formulae.

Only the fifth of these Gaussian models, i.e. "Vasicek correlated", has non-independent factors. We also test versions of the linear-quadratic model, namely DTQM1 with  $\mathbf{r}_t = y_t$  described above, DTQM2 and DTQM3 with  $\mathbf{r}_t = x_{1,t}^2 + x_{2,t}^2 + x_{3,t}^2$  described above. We compute  $V_{n,t} = e^{A_n + \mathbf{B}_n' \mathbf{x}_t + \mathbf{x}_t' \mathbf{C}_n \mathbf{x}_t + D_n y_t}$  where  $A_n, \mathbf{B}_n, \mathbf{C}_n, D_n$  solve equations 26, 27, 28 with 261 steps per year, since there are approximately 261 "trading days" in our sample period.

As in Kim and Singleton (2012), estimation of all models employs Quasi-Maximum-Likelihood estimation through the Extended Kalman Filter (EKF). Estimation uses yields of maturities from one year to ten years and requires the



maximisation of the log-likelihood function  $lk$ . Such maximisation is carried out with the Nelder-Mead simplex method.  $h_j$ , for  $j = 1, \dots, 10$ , denotes the standard deviation of the observation errors for the time series of the yield for maturity of  $j$  years. Setting the starting values of each  $h_j$  equal to few basis points ensures faster convergence to the optimal simplex solution. As the Kalman Filter is sensitive to the prior probability density of the starting values of the latent factors  $\mathbf{x}_0$  and  $y_0$  at time  $t = 0$ ,  $\mathbf{x}_0$  and  $y_0$  are treated as parameters to be estimated. This avoids arbitrary assumptions about the said prior probability density. Moreover unreported evidence shows that most models fit yields better when  $\mathbf{x}_0$  and  $y_0$  are treated as parameters rather than assuming that  $\mathbf{x}_0$  and  $y_0$  be distributed according to the unconditional densities of  $\mathbf{x}$  and  $y$ . Table 2 presents the estimation results for US yields and Table 3 for German yields (the Euro was introduced on 1/1/2002). The BHHH estimator provides the estimates of the standard deviations of the parameter estimates.

## 5.1 Results for US

Table 2 summarises the estimation results for the US. The columns headed "param" provide the parameter estimates and those headed "stdev" provide the corresponding standard deviations of the parameter estimates. Risk-premia drive the difference between the real measure and the risk-neutral measure, hence the difference between estimates of  $\kappa, \mu$  and of  $\kappa^*, \mu^*$  for all models.

All models fit short term yields less well than long term yields. This is highlighted by the standard deviation of the errors for the one year maturity,

which is the standard deviation of the daily difference between model predicted yields and observed yields for the one year maturity. Such standard deviation is measured by  $h_1$  and is 0.0019 and 0.0021 for the two Vasicek models (i.e. 19 and 21 basis points), between 0.0020 and 0.0023 for BK, Black1 and Black2, 0.0044 and 0.0036 for the "Mixed" models, 0.0045 for DTQM3, 0.0018 for DTQM2, 0.0020 for DTQM1. Therefore DTQM2 seems the best on this metric and DTQM3 the worst, which highlights the shortcoming of independent factors driving the short rate, as DTQM3 is the same as DTQM2 except that it assumes independent factors driving the short rate. According to  $h_1$  only DTQM2 beats "Vasicek correlated" in fitting one year yields, although "Vasicek correlated" allows the short rate  $r_t$  to turn negative. All models fit ten year yields much better than one year yields:  $h_{10} = 0.0001$  for all models except the Mixed models and DTQM3. Also the figures for "*Average h*", each of which is computed as  $\sum_{i=1}^{10} h_i / 10$ , show that DTQM2 and "Vasicek correlated" best fit observed US yields with "*Average h*" of 0.0003, closely followed by DTQM1. By contrast "*Average h*" is highest for DTQM3 and for the Mixed models. The Mixed models, which mix Black1 and Vasicek models, seem inferior to both Black1 and Vasicek models. The benefits of mixing models are not apparent. Note that the Vasicek models and the Mixed models have fewer parameters than the other models. All models have difficulty in fitting short term yields. As a term of comparison, Babbs and Nowman (1999) used the Kalman Filter to test a three-factor affine Gaussian model on US yields of similar maturities over a different period and estimated standard deviations of the observation errors

between 1 and 23 basis points.

The last row in Table 2, named AIC, shows the results of the Akaike information criterion (AIC) for each model. AIC compares the empirical performance of non-nested models with different numbers of parameters. The lowest AIC figure for DTQM2 ( $-444.410$ ) again suggests that DTQM2 is the "winner" of the race, closely followed by "Vasicek correlated" ( $-444.123$ ) and DTQM1 ( $-443.173$ ). AIC penalises models with more parameters. For each model also the starting values of the three latent factors are parameters.

[Table 2 about here]

## 5.2 Results for Germany

Table 3 presents the estimation results for Germany. All models fit German yields better than US yields. For each single model the value of the log-likelihood function  $lk$  is higher and "*Average h*" is lower for Germany than for the US, while the time window is the same for the two countries. The lowest  $h_1$  is 0.0009 for Black1, followed by 0.0010 for Black2, "Vasicek correlated", DTQM1 and DTQM3, while  $h_1$  is particularly high for the BK and Mixed models. As for the US, also for Germany all models fit one year yields less well than yields of longer maturities. As for the US, also for Germany "Vasicek correlated" fits observed yields well, even though it allows  $r$  to turn negative. For Germany the lowest "*Average h*" is 0.0019 for the Black models and for DTQM2, closely followed by "Vasicek correlated" and DTQM1 with "*Average h*" of 0.0020, while BK and Mixed models are the worst according to this metric. DTQM3 is less unsuitable

for Germany than it is for US.

The last row in Table 3, named AIC, shows the results of the Akaike information criterion (AIC) for each model for Germany. The lowest AIC figure for the DTQM2 model ( $-468.214$ ) makes it the "winner", closely followed by "Vasicek correlated" ( $-465.173$ ) and DTQM1 ( $-465.461$ ). This ranking for the top three models according to AIC almost mirrors the AIC ranking for the US. DTQM2, "Vasicek correlated" and DTQM1 have a common feature not shared by the other models: the factors driving the short rate are not independent, a feature that seems more important than the ruling out of negative yields, in order for a model to better fit observed yields.

[Table 3 about here]

### 5.3 More measures of empirical fit to observed yields

Table 4 for US and Table 5 for Germany display three other measures of how well model predicted yields match observed yields: the  $R^2$  measure, *MAPE* (mean absolute percentage errors) and *RMSE* (root mean squared errors).  $R^2$ , *MAPE* and *RMSE* are calculated using daily yield observations for all maturities, from one year to ten years, excluding the first observation date which is 1/1/1999. For any maturity  $R^2$  is the square of the correlation coefficient between observed yields and model predicted yields.  $R^2$  measures the fraction of the variation of the observed yields that is explained by the model. *MAPE* is the mean absolute value of the daily difference between model predicted yield and observed yield divided by the observed yield. *RMSE* is the standard devi-

ation of the daily difference between model predicted yield and observed yield. In the columns headed "Average" we compute the average  $R^2$ , average  $MAPE$  and average  $RMSE$  across all maturities of the corresponding row.

[Table 4 about here]

### 5.3.1 Results for US

According to  $R^2$  DTQM1, DTQM2 and the Vasicek models again best explain the variations of US yields of almost all maturities, but BK and Black1 follow closely. The average  $R^2$  across all maturities for DTQM2 and "Vasicek correlated" is 0.9969, for DTQM1 and Vasicek is 0.9968, for BK and Black1 is 0.9966. By contrast average  $R^2$  for the other models is clearly worse. For all models  $R^2$  for the one year maturity is lower than for longer maturities and signals the difficulty of models to fit short term yields. The one year maturity  $R^2$  for "Vasicek correlated", which is 0.9930, is the best, while the  $R^2$ 's for DTQM3 and for the Mixed models are the worst.

$MAPE$  and  $RMSE$  tell stories similar to that of  $R^2$ . The average  $MAPE$  of 1.9% for DTQM2 and of 1.96% for BK are the lowest, followed by 2.05% for Black1 and for "Vasicek correlated", and by 2.08% for DTQM1. According to average  $MAPE$  the Mixed models and DTQM3 are again the worst.  $MAPE$  for the one year maturity are lowest for DTQM2 (6,94%), for BK (7,34%) and for the Black models (7,72% for Black2 and 7,85% for Black1), while the Vasicek models are less competitive on this metric. For all models  $MAPE$  are highest for the one year maturity. This is due, beside the natural difficulty of models

in fitting short term yields, to the fact that percentage errors tend to be higher if observed yields are lower, and short term yields tend to be lower than long term yields.

DTQM1, DTQM2, BK and the Vasicek models also have the lowest average *RMSE* of 0.0008, followed by the average *RMSE* for Black1 at 0.0009. For one year yields DTQM2 has the lowest *RMSE* (0.0018), followed by "Vasicek correlated" (0.0019), DTQM1 and Black1 (0.0020).

Overall, according to average  $R^2$ , *MAPE* and *RMSE* for the US, DTQM2 appears to "win", but is closely followed by "Vasicek correlated", DTQM1 and, to a lesser extent, by BK. The relatively good performance of BK may explain why the BK model has been popular in industry. The Mixed models and DTQM3 tend to be the worst models. The benefits of mixing Black1 and Vasicek models are not apparent. Black1 tends to prevail over Black2. DTQM1 performs slightly worse than DTQM2, but fares well, which supports the idea of a quadratic model where one factor drives the short rate and the other two factors drive the "drift" of the short rate, as explained above.

[Table 5 about here]

### 5.3.2 Results for Germany

The results for German yields are somewhat different from the US. The best average  $R^2$  is 0.9972 for "Vasicek correlated" and DTQM1, followed by 0.9970 for Vasicek and DTQM3, 0.9968 for DTQM2, while the Mixed models and Black1 are worst on this metric. The best  $R^2$ 's for the one year maturity are

0.9938 for DTQM1 and 0.9937 for "Vasicek correlated", both of which do not rule out negative yields.

The best average *MAPE* are 0.47% for DTQM1 and "Vasicek correlated", 0.51% for DTQM2, while the Mixed models perform worst with the highest average *MAPE*. The best *MAPE* for the one year maturity are 2.94% for DTQM3 and "Vasicek correlated", and 2.95% for DTQM1.

The average *RMSE* for DTQM1, DTQM2, DTQM3 and for the Vasicek models are the best at 0.0002. For the one year maturity, the *RMSE* of DTQM1, DTQM3, Black2 and "Vasicek correlated" are the best at 0.0010. Overall, according to average  $R^2$ , *MAPE* and *RMSE* for Germany, DTQM1 and "Vasicek correlated" seem the best models for German yields. Both models do not rule out negative yields and assume non-independent factors driving the short rate. DTQM1 performs better than DTQM2, therefore German yields provide even more support than US yields for a quadratic model where one factor drives the short rate and the other factors drive the "drift" of the short rate. DTQM2 performs well also on German yields, while the BK model seems more suitable to US yields than to German yields. Even according to average  $R^2$ , *MAPE* and *RMSE* for both US and Germany, the non-independence of the factors seems more important than the ruling out of negative yields.

#### **5.4 Sub-periods with unusually low interest rates**

As stated above, the sample covers the period 1/1/1999-22/3/2011, but now we consider the performance of models in two sub-periods when yields were

unusually low because of unusually expansionary monetary policies in US and in "Euroland".

#### 5.4.1 Results for US

From 06/11/2001 to 14/12/2004 (first sub-period) the US FED set the Federal Funds Target rate at or below 2% and again the FED set the Target rate at or below 2% since 13/04/2008 (second sub-period). Indeed the Target rate has not exceeded 0.25% since 16/12/2008 until 22/3/2011, which is the end of the sample period. Table 4 also presents  $R^2$ ,  $MAPE$  and  $RMSE$  for these two sub-periods with Target rate at or below 2%. For all models the  $R^2$  in the two sub-periods is lower than it is across the whole sample period, meaning that the models are less capable to explain observed yield changes during the two sub-periods.

DTQM2, Black1 and "Vasicek correlated" have the highest average  $R^2$ : respectively 0.9854, 0.9840 and 0.9840 for the first sub-period; respectively 0.9850, 0.9854 and 0.9850 (together with BK's 0.9855) for the second sub-period. For the one year maturity in the first sub-period again DTQM2 (0.9778), Black1 (0.9671) and "Vasicek correlated" (0.9638) have the highest  $R^2$ , while in the second sub-period Black1 (0.9711), BK (0.9699), "Vasicek correlated" (0.9675) and DTQM2 (0.9663) and have the highest  $R^2$ . During these sub-periods with lowest levels of the Target rate Black1 and BK can even beat DTQM2 and "Vasicek correlated", which seems due to the zero lower bound of the short rate in BK and Black1.



During the first sub-period the lowest average *MAPE* are 1.9% for DTQM2, 1.96% for BK and 2.05% for Black1 and "Vasicek correlated". During the second sub-period the lowest average *MAPE* are 3.49% for BK, 3.62% for DTQM2 and 3,77% for Black1. For the one year maturity during the first sub-period "Vasicek correlated" and the quadratic models (except DTQM3) have the lowest *MAPE*, while during the second sub-period BK, Black1 and DTQM2 have the lowest *MAPE*. These results confirm that BK and Black1 are more competitive when the Target rate is lowest. For all models and all sub-periods *MAPE* for the one year maturity tend to be the highest. *MAPE* are higher after 13/04/2008, as all models find it more difficult to fit the low yields after the 2008 financial crisis. The high *MAPE* for DTQM3 and for the Mixed models for the one year maturity after 13/04/2008 highlight that such models are particularly unsuitable for US yields.

During the first sub-period the quadratic models (except DTQM3), the Vasicek models and Black2 have the lowest average *RMSE* at 0.0008, while during the second sub-period DTQM2, "Vasicek correlated", BK, Black1 and Black2 have the lowest average *RMSE* at 0.0009. For the one year maturity, Black2 and "Vasicek correlated" have the lowest *RMSE* at 0.0014 in the first sub-period, followed by 0.0015 for the quadratic models (except DTQM3), while BK (0.0016), DTQM2 (0.0018) and Black1 (0.0019) have the lowest *RMSE* for the second sub-period, with "Vasicek correlated" (0.0021) not faring well after 2008 for the one year maturity. Again BK and Black1 seem very competitive after 2008.

Overall during the two sub-periods with lowest Target rate DTQM2, "Vasicek correlated", Black1 and BK best fit US yields, with no clear winner. When the Target rate is lowest, models that rule out negative yields can compete and even beat DTQM2 and "Vasicek correlated", which are the best performers across the entire US sample.

#### 5.4.2 Results for Germany

On all dates in the sample the ECB's Euro Main Refinancing Operations middle rate, or "main Refi rate", was above 2%, except for two sub-periods. The first sub-period is from 06/06/2003 to 05/12/2005, when the "main Refi rate" was at 2%, and the second from 21/01/2009 to the end of the sample period, when it was at or below 2%. Since 13/5/2009 until the end of the sample the "main Refi rate" was at 1%. Table 5 presents  $R^2$ ,  $MAPE$  and  $RMSE$  for all models for the two sub-periods in which the "main Refi rate" was at or below 2%.

According to  $R^2$  all models perform worse during the two mentioned sub-periods, particularly for one year yields. For the first sub-period "Vasicek correlated" and DTQM1 have the highest average  $R^2$ , respectively 0.9788 and 0.9785, and the highest  $R^2$  for the one year maturity, respectively 0.9300 and 0.9275. For the second sub-period the highest average  $R^2$ 's are 0.9718 for Black1, 0.9715 for DTQM1 and Black2, 0.9714 for "Vasicek correlated", while the highest  $R^2$ 's for the one year maturity are 0.8812 for Black1, 0.8771 for Black2, 0.8733 for DTQM1 and 0.8728 for "Vasicek correlated". According to  $R^2$  the Black models seem the best models after the financial crisis of 2008, as the merits of ruling

out negative yields become more apparent after the crisis.

According to *MAPE* all models perform worse after 2009 than during the first sub-period with low Refi rate (2003-2005). For the first sub-period the lowest average *MAPE* are 0.32% for DTQM1 and "Vasicek correlated" and 0.34% for DTQM2, while the lowest *MAPE* for the one year maturity are 1.75% for DTQM2, 1, 82% for "Vasicek correlated" and DTQM1. For the second sub-period the lowest average *MAPE* are 0.86% for DTQM2, 0.89% for Black1, 0.94% for "Vasicek correlated" and 0.95% for DTQM1, while for the one year maturity the lowest *MAPE* are 5.67% for Black1, 5.79% for DTQM2, 6.03% for DTQM3 and 6.38% for "Vasicek correlated".

Average *RMSE* for all models except the Mixed models are 0.0001 in the first sub-period. In the second sub-period the lowest average *RMSE* are 0.0002 for the Black, the Vasicek and the quadratic models (except DTQM3). *RMSE* for the one year maturity are better able to discriminate the models: for the first sub-period the quadratic models (except DTQM3) and "Vasicek correlated" have the lowest *RMSE* at 0.0005 followed by the Black models at 0.0007; for the second sub-period the lowest *RMSE* are for Black1, DTQM2 and DTQM3 at 0.0010, followed by Black2, DTQM1 and "Vasicek correlated" at 0.0011.

Overall for German yields DTQM1, DTQM2 and "Vasicek correlated" seem the best even in sub-periods with lowest main Refi rate, but after 2009 Black1 seems an equally good alternative.

## 5.5 Discussion of results

This section qualifies the above results. Unreported statistics document that the "errors", i.e. the daily differences between observed yields and model predicted yields of all maturities, do not follow a white noise process. There is overwhelming evidence of auto-correlation of daily "errors" and also of cross-correlation between the errors for the different maturities. These facts characterise all models and all maturities and inevitably point to the mis-specification of the models. To overcome mis-specification, future research may want to consider four factor models.

As the optimisation routine searches for the parameter values that maximise the likelihood function, the latent factors may occasionally be "pushed" by the Kalman Filter outside the finite solution region used by vertical MOL. Therefore latent factors were constrained to move within the solution region by placing a high penalty on the likelihood function whenever the factors were "pushed" by the Kalman Filter outside the solution region.

Mixed Black1-Vasicek models did not show advantages over the "pure" Black1 and "pure" Vasicek models. Mixed models perform worse than Black1 partly because of the restriction  $\mu_2 = \mu_3 = \mu_2^* = \mu_3^* = 0$ , which is absent in "pure" Black1. This restriction makes Mixed models estimation easier by preventing the coexistence of "very negative Vasicek factors" and "very positive Black 1 factors" even beyond the MOL grid upper boundary. Less clear is why Mixed models perform worse than Vasicek models, although this again confirms that ruling out negative yields may not always be of primary importance to fit

observed yields.

We could also mix the above models in many other ways, but this is left for future research.

## 6 Conclusion

Using US and German Government bond yields, this paper has tested affine and quadratic Gaussian models, multi-factor extensions of the Black model and of the Black-Karasinski model. All models fit the German yield curve better than the US yield curve. All models fit short term yields, such as one-year yields, less well than yields of longer maturities. The empirical fit of quadratic models, namely DTQM2 for US and DTQM1 or DTQM2 for Germany, seems the best together with that of a Vasicek-type affine model with correlated factors. The reason is that these quadratic and affine models are driven by non-independent stochastic factors. However when interest rates are lowest, a variant of the Black model (Black1) fits German and US yields particularly well and a variant of the Black-Karasinski model fits US yields well: although these models assume independent factors, they can best fit bond yields when central bank interest rates are lowest. DTQM1 is a relatively new linear-quadratic model, whereby the short rate may turn negative, while the central tendency of the short rate is a non-negative quadratic function of two Gaussian factors. It seems promising for future research to further test this model as well as extensions of the Black model.

## A Estimation with Extended Kalman filter (EKF)

This Appendix describes how EKF is implemented to estimate all the models in the text, except for DTQM1. For DTQM1 the details of how EKF is implemented are only slightly different from this Appendix. We introduce the following notation and assumptions:

- $\mathbf{x}_t = (x_{1,t}, x_{2,t}, x_{3,t})'$  for all models except for DTQM1;

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

- $E_{t-1}[\cdot]$  is the real measure expectation operator conditional on time  $t - 1$  information;  $\mathbf{P}_{t-} = E_{t-1}[(\mathbf{x}_t - \widehat{\mathbf{x}}_{t-})(\mathbf{x}_t - \widehat{\mathbf{x}}_{t-})']$ ;

- $\mathbf{o}_t = (o_{1,t}, \dots, o_{10,t})'$  are the discount bond yields observed in the market at time  $t$  for maturities of 1, ..., 10 years;

- $\mathbf{z}(\mathbf{x}_t) = (z_{1,t}(\mathbf{x}_t), \dots, z_{10,t}(\mathbf{x}_t))'$  is the time  $t$  vector of discount bond yields computed using a model; for example  $z_{2,t}(\mathbf{x}_t) = -\ln \frac{V}{2}$ ;

- $\epsilon_t$  is the vector of observation errors at time  $t$ , which is normally distributed such that  $\epsilon_t \sim N(\mathbf{0}_{10}, \mathbf{H}_{10})$ ;  $\mathbf{0}_{10}$  is a  $10 \times 1$  column vector of zeroes;  $\mathbf{H}_{10}$  is a  $10 \times 10$  diagonal matrix;

- the observation errors  $\epsilon_t$  are uncorrelated with the latent process  $\mathbf{x}_t$  and with all lags of  $\mathbf{x}_t$ ;  $\mathbf{x}_0$  denotes the initial values of the latent factors and are parameters to be estimated.

The measurement equation is

$$\mathbf{o}_t \simeq \mathbf{z}(\hat{\mathbf{x}}_{t-}) - \mathbf{D}_t \hat{\mathbf{x}}_{t-} + \mathbf{D}_t \mathbf{x}_t + \epsilon_t \quad (29)$$

$$\mathbf{D}_t = \left[ \frac{\partial \mathbf{z}(\mathbf{x}_t)}{\partial \mathbf{x}_t'} \right]_{\mathbf{x}_t = \hat{\mathbf{x}}_{t-}}. \quad (30)$$

$\mathbf{D}_t$  is a  $10 \times 3$  matrix,  $\mathbf{z}(\mathbf{x}_t)$  is a  $10 \times 1$  vector and  $\mathbf{x}_t$  is a  $3 \times 1$  vector. Then the EKF equations are

$$\hat{\mathbf{o}}_{t-} = E_{t-1}[\mathbf{o}_t] = \mathbf{z}(\hat{\mathbf{x}}_{t-}) \quad (31)$$

$$\hat{\mathbf{x}}_{t-} = \phi^* \mu^* + (\mathbf{I} - \phi^*) \mathbf{x}_{t-1} \quad (32)$$

$$\hat{\mathbf{P}}_{t-} = \phi^* \hat{\mathbf{P}}'_{t-1} \phi^{*'} + \Sigma \Sigma' \quad (33)$$

$$\mathbf{F}_t = \mathbf{D}_t \hat{\mathbf{P}}_{t-} \mathbf{D}_t' + \mathbf{H} \quad (34)$$

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t-} + \hat{\mathbf{P}}_{t-} \mathbf{D}_t' \mathbf{F}_t^{-1} (\mathbf{o}_t - \hat{\mathbf{o}}_{t-}) \quad (35)$$

$$\hat{\mathbf{P}}_t = \hat{\mathbf{P}}_{t-} - \hat{\mathbf{P}}_{t-} \mathbf{D}_t' \mathbf{F}_t^{-1} \mathbf{D}_t \hat{\mathbf{P}}_{t-}. \quad (36)$$

The approximate conditional likelihood function of  $\mathbf{o}_t$  is

$$l(\mathbf{o}_t | \mathbf{O}_{t-1}) \sim N(\hat{\mathbf{o}}_{t-}, \mathbf{F}_t) \quad (37)$$

where  $\mathbf{O}_{t-1} = \{\mathbf{o}_{t-1}, \mathbf{o}_{t-2}, \dots, \mathbf{o}_1\}$  and  $N(\hat{\mathbf{o}}_{t-}, \mathbf{F}_t)$  denotes the multivariate nor-

mal density with mean  $\widehat{\boldsymbol{o}}_{t-}$  and covariance  $\mathbf{F}_t$ . Then we can write

$$\ln(l(\boldsymbol{o}_t | \mathbf{O}_{t-1})) = -\frac{10}{2} \ln 2\pi - \frac{1}{2} \ln(\text{abs}(|\mathbf{F}_t|)) - \frac{1}{2} (\boldsymbol{o}_t - \widehat{\boldsymbol{o}}_{t-})' \mathbf{F}_t^{-1} (\boldsymbol{o}_t - \widehat{\boldsymbol{o}}_{t-}). \quad (38)$$

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

$$lk = \sum_{t=1}^M \ln l(\boldsymbol{o}_t | \mathbf{O}_{t-1}). \quad (39)$$

where  $M$  is the number of observation dates, which is 3,188. The time step  $\Delta$  is the time between consecutive observations. We observe about 261 daily prices per year in the data, therefore  $\Delta = 1/261$ .

## B Identification of parameters for the quadratic and linear quadratic models

### B.1 Identification of DTQM2

This section derives the parameter identification conditions for DTQM2. We consider the linear transformation  $\mathbf{x}_t = \boldsymbol{\Omega} \mathbf{f}_t + \boldsymbol{\Theta}$ , where  $\boldsymbol{\Theta}$ ,  $\mathbf{x}$  and  $\mathbf{f}_t$  are  $3 \times 1$  vectors and  $\boldsymbol{\Omega}$  is a  $3 \times 3$  matrix.  $\boldsymbol{\Omega}^{-1}$  is assumed to exist.  $\mathbf{x}_t$  are the latent factors and  $\mathbf{f}_t$  are "rotations" of  $\mathbf{x}_t$ . The linear transformation is invariant if



and only if  $\mathbf{\Omega} = \mathbf{I}_3$  and  $\mathbf{\Theta} = \mathbf{0}_3$ . Since in DTQM2

$$\begin{aligned}\mathbf{r}_t &= \mathbf{x}_t' \mathbf{\Psi} \mathbf{x}_t \\ \mathbf{x}_{t+1} &= (\mathbf{I}_3 - \phi) \mathbf{x}_t + \phi \mu + \mathbf{\Sigma} \xi_{t+1}\end{aligned}$$

we can employ the linear transformation to write

$$\begin{aligned}\mathbf{r}_t &= \mathbf{\Theta}' \mathbf{\Psi} \mathbf{\Theta} + \mathbf{f}_t' \mathbf{\Omega}' \mathbf{\Psi} \mathbf{\Theta} + \mathbf{\Theta}' \mathbf{\Psi} \mathbf{\Omega} \mathbf{f}_t + \mathbf{f}_t' \mathbf{\Omega}' \mathbf{\Psi} \mathbf{\Omega} \mathbf{f}_t \\ \mathbf{f}_{t+1} - \mathbf{f}_t &= \mathbf{\Omega}^{-1} \phi (\mu - \mathbf{\Theta} - \mathbf{\Omega} \mathbf{f}_t) + \mathbf{\Omega}^{-1} \mathbf{\Sigma} \xi_{t+1}.\end{aligned}$$

The said transformation is invariant only if  $\mathbf{\Theta}' \mathbf{\Psi} \mathbf{\Theta} + \mathbf{f}_t' \mathbf{\Omega}' \mathbf{\Psi} \mathbf{\Theta} + \mathbf{\Theta}' \mathbf{\Psi} \mathbf{\Omega} \mathbf{f}_t = 0$ , i.e. only if  $\mathbf{\Theta} = \mathbf{0}_3$ . The fact that  $\mathbf{\Theta} = \mathbf{0}_3$  entails that  $\mu$  can be uniquely identified.

Then we can re-write

$$\begin{aligned}\mathbf{r}_t &= \mathbf{f}_t' \mathbf{\Omega}' \mathbf{\Psi} \mathbf{\Omega} \mathbf{f}_t \\ \mathbf{f}_{t+1} - \mathbf{f}_t &= \mathbf{\Omega}^{-1} \phi (\mu - \mathbf{\Omega} \mathbf{f}_t) + \mathbf{\Omega}^{-1} \mathbf{\Sigma} \xi_{t+1}.\end{aligned}$$

Hereafter we need to impose conditions that imply that  $\mathbf{\Omega} = \mathbf{I}_3$ . As in DTQM2  $\mathbf{\Sigma}$  is lower triangular,  $\mathbf{\Omega}^{-1} \mathbf{\Sigma}$  needs to be lower triangular too, which implies that also  $\mathbf{\Omega}$  be lower triangular. Then as in DTQM2  $\mathbf{\Psi} = \mathbf{I}_3$  is a diagonal matrix,  $\mathbf{\Omega}' \mathbf{\Psi} \mathbf{\Omega}$  needs to be a diagonal matrix too, which implies that also  $\mathbf{\Omega}$  needs to be a diagonal matrix. In DTQM2  $\phi$  is lower triangular. Then  $\mathbf{\Omega}^{-1} \phi \mathbf{\Omega}$  and  $\mathbf{\Omega}' \mathbf{\Psi} \mathbf{\Omega}$  imply that the diagonal elements of  $\mathbf{\Omega}$  be all equal 1, in order for

the said transformation to be invariant.

## B.2 Identification of DTQM1

This section derives the parameter identification conditions for DTQM1. We consider the linear transformations  $y_t = \Omega_y f_{y,t} + \Theta_y$  and  $\mathbf{x}_t = \mathbf{\Omega} \mathbf{f}_t + \mathbf{\Theta}$ , where  $\mathbf{\Theta}$ ,  $\mathbf{x}_t$  and  $\mathbf{f}_t$  are  $2 \times 1$  vectors and  $\mathbf{\Omega}$  is a  $2 \times 2$  matrix.  $\mathbf{\Omega}^{-1}$  is assumed to exist.  $\mathbf{x}_t$  are the latent factors and  $\mathbf{f}_t$  are "rotations" of  $\mathbf{x}_t$ .  $\Theta_y, y_t, f_{y,t}$  and  $\Omega_y$  are scalars.  $y_t$  is a latent factor and  $f_{y,t}$  is a "rotation" of  $y_t$ . The linear transformations are invariant if and only if  $\mathbf{\Omega} = \mathbf{I}_2$ ,  $\mathbf{\Theta} = \mathbf{0}_2$ ,  $\Omega_y = 1$  and  $\Theta_y = 0$ . Then we can employ these linear transformation to rewrite DTQM1 as

$$\mathbf{r}_t = \Omega_y f_y + \Theta_y \quad (40)$$

$$\mathbf{f}_{t+1} - \mathbf{f}_t = \mathbf{\Omega}^{-1} \phi(\mu - \mathbf{\Theta} - \mathbf{\Omega} \mathbf{f}_t) + \mathbf{\Omega}^{-1} \mathbf{\Sigma} \xi_{t+1} \quad (41)$$

$$f_{y,t+1} - f_{y,t} = \Omega_y^{-1} \kappa_y \left( \mu_y + (\mathbf{\Omega} \mathbf{f}_t + \mathbf{\Theta})' \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} (\mathbf{\Omega} \mathbf{f}_t + \mathbf{\Theta}) - \Theta_y - \Omega_y f_{y,t} \right) \Delta + \quad (42)$$

$$+ \Omega_y^{-1} \begin{pmatrix} \sigma_{yx,1} & \sigma_{yx,2} & s_y \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t+1}^Q \\ \varepsilon_{2,t+1}^Q \\ \varepsilon_{y,t+1}^Q \end{pmatrix} \sqrt{\Delta}. \quad (43)$$

Since  $\mathbf{r}_t = y_t$  invariance of the transformation implies that  $\Theta_y = 0$  and  $\Omega_y = 1$ .

As in DTQM1  $\mathbf{\Sigma}$  is lower triangular,  $\mathbf{\Omega}^{-1} \mathbf{\Sigma}$  needs to be lower triangular too,

which implies that also  $\mathbf{\Omega}$  be lower triangular. Then the transformation  $\mathbf{x}_t = \mathbf{\Omega}\mathbf{f}_t + \mathbf{\Theta}$  is invariant if

$$\begin{aligned} \mu_y + (\mathbf{\Omega}\mathbf{f}_t + \mathbf{\Theta})' \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} (\mathbf{\Omega}\mathbf{f}_t + \mathbf{\Theta}) &= \mu_y + (x_1 + x_2)^2 \\ \mathbf{\Omega}^{-1} \begin{pmatrix} \kappa_{1,1} & 0 \\ \kappa_{2,1} & \kappa_{2,2} \end{pmatrix} \mathbf{\Omega} &= \begin{pmatrix} \kappa_{1,1} & 0 \\ \kappa_{2,1} & \kappa_{2,2} \end{pmatrix} \end{aligned}$$

which implies that  $\mathbf{\Theta} = \mathbf{0}_2$  and  $\mathbf{\Omega} = \mathbf{I}_2$  so that the parameters of DTQM1 are identifiable.

### B.3 Quadratic model in continuous time as special case of quadratic model in discrete time

This appendix shows that as the time step  $\Delta \rightarrow 0$ , the discrete time quadratic model of equations 26, 27 and 28 tends to the continuous time model of Ahn, Dittmaer and Gallant (2002), whereby

$$\begin{aligned} d\mathbf{x}_t &= \kappa(\mu - \mathbf{x}_t) dt + \mathbf{S}d\mathbf{w}_t^Q \\ r_t &= \alpha + \beta'\mathbf{x}_t + \mathbf{x}_t'\mathbf{\Psi}\mathbf{x}_t \end{aligned}$$

$$\frac{\partial V}{\partial \tau} = \frac{1}{2} \text{tr} \left[ \frac{\partial^2 V}{\partial \mathbf{x} \partial \mathbf{x}'} \mathbf{S}\mathbf{S}' \right] + \left( \frac{\partial V}{\partial \mathbf{x}} \right)' \kappa(\mu - \mathbf{x}) - (\alpha + \beta'\mathbf{x} + \mathbf{x}'\mathbf{\Psi}\mathbf{x}) V, \quad \text{s.t. } V(0) = 1$$

$$V(0) = \exp(A + \mathbf{B}'\mathbf{x} + \mathbf{x}'\mathbf{C}\mathbf{x})$$

$$\frac{\partial \mathbf{C}}{\partial \tau} = -\Psi - \mathbf{C}\kappa - \kappa'\mathbf{C} + 2\mathbf{C}'\mathbf{S}\mathbf{S}'\mathbf{C}. \quad (44)$$

$$\frac{\partial \mathbf{B}}{\partial \tau} = -\beta - \kappa'\mathbf{B} + 2\mathbf{C}\kappa\mu + 2\mathbf{C}\mathbf{S}\mathbf{S}'\mathbf{B} \quad (45)$$

$$\frac{\partial A}{\partial \tau} = -\alpha + \mathbf{B}'\kappa\mu + \text{tr}(\mathbf{C}\mathbf{S}\mathbf{S}') + \frac{1}{2}\mathbf{B}'\mathbf{S}\mathbf{S}'\mathbf{B} \quad (46)$$

$$A(0) = 0, \quad \mathbf{B}(0) = 0, \quad \mathbf{C}(0) = 0$$

with  $\alpha = 0, \beta' = \mathbf{0}'_3$ .  $\mathbf{0}_3$  is a  $3 \times 1$  vector of zeros.  $\text{tr}(\mathbf{C}\mathbf{S}\mathbf{S}')$  denotes the trace of  $\mathbf{C}\mathbf{S}\mathbf{S}'$ . First we note that as  $\Delta \rightarrow 0$

$$\mathbf{C}_n = -\Psi\Delta + (\mathbf{I}_3 - \phi)' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) + 2\gamma' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) (\mathbf{I}_3 - \phi)' \mathbf{C}_{n-1} \gamma \quad (47)$$

can be re-written as

$$\begin{aligned} \mathbf{C}_n - \mathbf{C}_{n-1} &= -\Psi\Delta - \phi' \mathbf{C}_{n-1} - \mathbf{C}_{n-1} \phi + 2\gamma' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) (\mathbf{I}_3 - \phi)' \mathbf{C}_{n-1} \gamma \\ &= -\Psi\Delta - \phi' \mathbf{C}_{n-1} - \mathbf{C}_{n-1} \phi + 2\mathbf{C}_{n-1} \gamma \gamma' \mathbf{C}_{n-1} \end{aligned}$$

since  $\lim_{\Delta \rightarrow 0} \frac{\phi' \phi}{\Delta} \rightarrow 0$ . Then since  $\gamma \gamma' = \left( (\Sigma \Sigma')^{-1} - 2\mathbf{C}_{n-1} \right)^{-1} = \frac{\Sigma \Sigma'}{\mathbf{I}_3 - 2\mathbf{C}_{n-1} \Sigma \Sigma'}$  and since  $\left( (\Sigma \Sigma')^{-1} - 2\mathbf{C}_{n-1} \right) \gamma \gamma' = \mathbf{I}_3$ , it follows that

$$\begin{aligned} \gamma \gamma' &= \gamma \gamma' \left( (\Sigma \Sigma')^{-1} - 2\mathbf{C}_{n-1} \right) \gamma \gamma' \\ &= \gamma \gamma' (\Sigma \Sigma')^{-1} \gamma \gamma' - 2\gamma \gamma' \mathbf{C}_{n-1} \gamma \gamma' \\ &= \frac{\Sigma \Sigma' (\Sigma \Sigma')^{-1} \Sigma \Sigma' - 2 \cdot \Sigma \Sigma' \mathbf{C}_{n-1} \Sigma \Sigma'}{(\mathbf{I}_3 - 2\mathbf{C}_{n-1} \Sigma \Sigma')^2} \\ &= \Sigma \Sigma' - \frac{2 \cdot \Sigma \Sigma' \mathbf{C}_{n-1} \Sigma \Sigma'}{(\mathbf{I}_3 - 2\mathbf{C}_{n-1} \Sigma \Sigma')^2} \end{aligned}$$

and it also follows that  $\lim_{\Delta \rightarrow 0} \gamma \gamma' \rightarrow \Sigma \Sigma'$  since  $\lim_{\Delta \rightarrow 0} \Sigma \Sigma' - \frac{2 \cdot \Sigma \Sigma' \mathbf{C}_{n-1} \Sigma \Sigma'}{(\mathbf{I}_3 - 2\mathbf{C}_{n-1} \Sigma \Sigma')^2} \rightarrow \Sigma \Sigma'$ .

By similar arguments

$$A_n = A_{n-1} + \mathbf{B}'_{n-1} \phi \mu + (\phi \mu)' \mathbf{C}_{n-1} \phi \mu + \ln \frac{|\gamma|}{\text{abs}|\Sigma|} + \frac{1}{2} (\mathbf{B}'_{n-1} + 2(\phi \mu)' \mathbf{C}_{n-1}) \gamma \gamma' (\mathbf{B}'_{n-1} + 2(\phi \mu)' \mathbf{C}_{n-1})'$$

$$\mathbf{B}'_n = \mathbf{B}'_{n-1} (\mathbf{I}_3 - \phi) + 2(\phi \mu)' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi) + 2 \cdot (\mathbf{B}'_{n-1} + 2(\phi \mu)' \mathbf{C}_{n-1}) \gamma \gamma' \mathbf{C}_{n-1} (\mathbf{I}_3 - \phi)$$

can be re-written as

$$\begin{aligned} \mathbf{B}'_n - \mathbf{B}'_{n-1} &= -\beta \cdot \Delta - \mathbf{B}'_{n-1} \phi + 2(\phi \mu)' \mathbf{C}_{n-1} + 2\mathbf{C}_{n-1} \Sigma \Sigma' \mathbf{B}_{n-1} \\ A_n - A_{n-1} &= -\alpha \cdot \Delta + \mathbf{B}'_{n-1} \phi \mu + \ln \frac{|\gamma|}{\text{abs}|\Sigma|} + \frac{1}{2} \mathbf{B}'_{n-1} \Sigma \Sigma' \mathbf{B}_{n-1}. \end{aligned}$$

Then as  $\Delta \rightarrow 0$

$$\begin{aligned}\frac{\mathbf{C}_n - \mathbf{C}_{n-1}}{\Delta} &= -\boldsymbol{\Psi} - \kappa' \mathbf{C}_{n-1} - \mathbf{C}_{n-1} \kappa + 2\mathbf{C}_{n-1} \mathbf{S} \mathbf{S}' \mathbf{C}_{n-1} \\ \frac{\mathbf{B}'_n - \mathbf{B}'_{n-1}}{\Delta} &= -\beta - \mathbf{B}'_{n-1} \kappa + 2(\kappa \mu)' \mathbf{C}_{n-1} + 2\mathbf{C}_{n-1} \mathbf{S} \mathbf{S}' \mathbf{B}_{n-1} \\ \frac{A_n - A_{n-1}}{\Delta} &= -\alpha + \mathbf{B}'_{n-1} \kappa \mu + \frac{1}{\Delta} \ln \frac{|\gamma|}{\text{abs}|\boldsymbol{\Sigma}|} + \frac{1}{2} \mathbf{B}'_{n-1} \mathbf{S} \mathbf{S}' \mathbf{B}_{n-1}\end{aligned}$$

become differential equations 44, 45 and 46.

$$\text{Notice that } \lim_{\Delta \rightarrow 0} \left( \ln \left( \left| \left( (\boldsymbol{\Sigma} \boldsymbol{\Sigma}')^{-1} - 2\mathbf{C}_{n-1} \right)^{-\frac{1}{2}} \right| \right) - \ln(\text{abs}|\boldsymbol{\Sigma}|) \right) \rightarrow \text{tr}(\mathbf{C} \boldsymbol{\Sigma} \boldsymbol{\Sigma}').$$

## References

- [1] Ahn D., Dittmar R., Gallant R., 2002, "Quadratic term structure models: theory and evidence", *Review of financial studies* 15, 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] Black F., 1995, "Interest rates as options", *Journal of Finance* 50, 1371-1376.
- [4] Chen L., Filipovic D. and Poor V., 2004, "Quadratic term structure models for risk-free and defaultable rates", *Mathematical Finance* 14, 515-536.
- [5] Dai Q. and Singleton K., 2003, "Term structure dynamics in theory and reality", *Review of Financial Studies* 16, 631-678.

- [6] Duffee G., 2011, "Information in (and not in) the Term Structure", *Review of Financial Studies* 24, n.9, 2895-2934.
- [7] Gouriéroux C. and Sufana R., 2005, "Wishart quadratic term structure models", Working paper University of Toronto.
- [8] 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.
- [9] Higham N. J., 2005, "The Scaling and Squaring Method for the Matrix Exponential Revisited," *SIAM Journal Matrix Anal. Appl.* 26, n.4, 1179-1193.
- [10] Khaliq A.Q.M., Voss D.A. and Yousuf M., 2007, "Pricing exotic options with L-stable Padé schemes", *Journal of Banking & Finance* 31, 3438-3461.
- [11] 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.
- [12] Langetieg T.C., 1980, "A multivariate model of the term structure", *Journal of Finance* 35, 71-97.
- [13] Leippold M. and Wu L., 2002, "Asset pricing under the quadratic class", *Journal of Financial and Quantitative Analysis* 37, 271- 294.
- [14] Leippold M. and Wu L., 2003, "Design and estimation of quadratic term structure models", *European Finance Review* 7, 47-73.

- [15] Nowman B.K., 2010, "Modelling the UK and Euro yield curves using the Generalized Vasicek model: Empirical results from panel data for one and two factor models", *International Review of Financial Analysis* 19 334-341.
- [16] Realdon M., 2006, "Quadratic term structure models in discrete time", *Finance Research Letters* 3, n.4, 277-289.
- [17] Realdon M., 2009, "Extended Black" term structure models", *International Review of Financial Analysis* 18, 232-238.
- [18] Realdon M., 2011, "Discrete time linear-quadratic pricing of bonds and options", *Applied Financial Economics* 21, n.7, 463-467.
- [19] Vasicek O.A., 1977, "An equilibrium characterization of the term structure", *Journal of Financial Economics* 5, 177-188.