APPLICATION OF THE KALMAN FILTER FOR ESTIMATING CONTINUOUS TIME TERM STRUCTURE MODELS: THE CASE OF UK AND GERMANY

Somnath Chatterjee*
Department of Economics
University of Glasgow

January, 2005

Abstract

The purpose of this paper is to see how the term structure of interest rates has evolved in the sterling and euro treasury bond markets over the period 1999-2003. German bonds have been used as a proxy for euro-denominated bonds. A state-space representation for the single-factor Cox, Ingersoll and Ross (1985) model is employed to analyse the intertemporal dynamics of the term structure. Quasi-maximum likelihood estimates of the model parameters are obtained by using the Kalman filter to calculate the likelihood function. Results of the empirical analysis show that while the unobserved instantaneous interest rate exhibits mean reverting behaviour in both the UK and Germany, the mean reversion of the interest rate process has been relatively slower in the UK. The volatility component, which shocks the process at each step in time is also higher in the UK as compared to Germany.

Keywords: Kalman filter, Panel Data, Term Structure
JEL classification: G15

---

* Address for correspondence: University of Glasgow, Department of Economics, Adam Smith Building, Glasgow G12 8RT. E-mail: R.Chatterjee@socsci.gla.ac.uk

1 Introduction

Term structure modelling can explore two distinct, but related aspects. The first involves the fitting of a zero-coupon yield curve to a set of cross-sectional bond price observations on any given trading day. The second aspect, which is the focus of this paper, relates to the specification of the intertemporal dynamics of the term structure and addresses the issue of how bond yields evolve over time. Estimating the term structure is based on the premise that bonds with different maturities are traded at the same time. Bonds with long maturities are risky when held over short horizons and risk-averse investors demand compensation for bearing such risk. Arbitrage opportunities in these markets exist unless long-yields are risk-adjusted expectations of average future short rates. Restrictions are therefore imposed on inter-temporal interest rate behaviour by using the no-arbitrage argument. The absence of arbitrage, would ensure that movements of the term structure do not permit conditions to occur under which market participants may earn risk-free profits.

As interest rates are stochastic processes, models rely on the reduction of interest rate uncertainty and attempt to provide a parsimonious characterisation of the dynamics of the term structure. There exist various specifications that differ with respect to the number of underlying state variables and the type of the stochastic process. Affine term structure models are constructed by assuming that bond yields are a linear function of the underlying state variables that provide uncertainty to the model. Most modelling approaches are based on the concept that although interest rates change randomly over time, it is possible to divide the changes into two parts using a stochastic differential equation. The first part is a non-random, deterministic component, called the drift of the process, and the second is the random or noise part which entails the volatility component of the process. Examples are the one-factor Vasicek (1977) model with constant volatility, the Cox-Ingersoll-Ross (1985) model with square-root volatility and the two-factor model of Longstaff and Schwartz (1992). Stochastic differential equations have, in recent years, been increasingly used to model financial data. However, the process specified by a stochastic differential equation is defined in continuous time, while the observed data are sampled at discrete time intervals. As discussed in Durham and Gallant (2002) the resulting estimation problem turns out to be nontrivial, and research has focussed on developing computationally and statistically efficient estimation schemes. Although maximum likelihood is typically the estimator of choice, the transition density is generally unknown and has to be approximated.

The Vasicek (1977) model is a one-factor partial equilibrium model and starts out with the specification of a time series process for the instantaneous spot interest rate which is treated as the only factor of uncertainty. The no-arbitrage restriction then permits the derivation of a bond pricing formula whereby the bond price is a function of the unobserved instantaneous spot rate and the model's parameters. The approach was extended to include a second factor of uncertainty.

Cox, Ingersoll and Ross (1985, CIR hereafter) develop a general equilibrium asset pricing model that allows the derivation of the term structure of interest rates. The model is set up as a single-good, continuous time economy with a single state variable. A multivariate version was developed by Longstaff and Schwartz (1992) in which the two-factors were the short-term interest rate and the variance of changes in the short-term interest rate. Duffie and Kan (1996) define a general class of multifactor affine models of the term structure that allows for the nesting of some of the aforementioned term structure models such as Vasicek (1978), CIR (1985) and Longstaff and Schwartz (1992).

The literature would suggest that three state variables are adequate to explain most of the variability in bond yields. For example, Litterman and Scheinkman (1991) show that this can be captured by the level, the steepness and the curvature of the term structure. This paper focuses on the one-factor CIR model as the empirical estimation showed that the inclusion of additional factors did not increase the performance of the model for either country. A plausible explanation for this could be the limited period of observation. Most studies have concluded that the level is the most important factor in explaining interest variation over time. In fact, Litterman and Scheinkman (1991) have demonstrated that three factors notwithstanding, almost 90 percent of the variation in US Treasury rates is attributable to the variation in the first factor, which is considered to correspond to the level of interest rates. Thus from an empirical point of view a one-factor CIR model can be considered acceptable.

The purpose of this paper is to explore how the term structure has evolved in the sterling and euro treasury bond markets between January 1999 and January 2004. German bonds have been used to represent euro-denominated bonds as they are seen by market participants as the main component of the euro yield curve. Although there exists a considerable literature on empirically estimating the CIR model, most of the tests have been performed on US data. The few studies that have focussed on the UK and European markets relate to the pre-EMU period. Steeley (1997) has modelled the forward premium in the UK gilt-edged market over the period 1982-96 using a two-factor general equilibrium model of the term structure of interest rates. Nath and Nowman (2001) estimate multi-factor versions of the CIR model using the UK Gilt-edged market data over the period 1982-97.

It is believed that, this is the first study that estimates this model for the UK and Euro-denominated bond data since the launch of the single currency. By bringing together the empirical findings for the euro and sterling treasury bond markets an attempt is made to compare the dynamics of their respective term structures. This investigation into the intertemporal behaviour of the euro and sterling term structure may provide evidence on whether there exists any common factors.

The rest of the paper is organised as follows. Section II provides the theoretical framework that discusses in detail the one-factor CIR model for the instantaneous interest rate. Section III provides an overview of the different estimation methods. In Section IV the state space representation of the CIR model is formulated and, in Section V the Kalman filter algorithm is employed. Section VI presents the data and results. Finally, Section VII concludes.

2 Theoretical Framework

Affine term-structure models are constructed by assuming that bond yields are a linear function of the underlying state variables that provide uncertainty in the model. Developing an affine term structure model involves a specification of a stochastic process for the state variables, or factors, that drive the dynamics of the term structure. In a one-factor term structure model, the factor is generally taken to be the instantaneous spot rate of interest, *r*. As mentioned in the introduction, it is possible to divide the change in its value (*dr*) into two parts, the first is a non-random deterministic component [μ(r,t)], called the drift of the process, and the second is a diffusion term or random part [σ(r,t)dW], which is the variance of the process. This involves the assumption, that the interest rate process is generated by a standard Brownian motion[1], also known as a Wiener process, and that its dynamics can be described by the following first-order stochastic differential equation:

$$dr = \mu(r,t)dt + \sigma(r,t)dW \tag{1}$$

where dW is a Wiener process.

The price of a pure discount bond, $P(t,T)$, in an affine term structure model would have the following functional form:

$$P(t,T) = \exp(A(\tau) - B(\tau)X) \tag{2}$$

[1] A Brownian Motion is a stochastic process where the change in a variable during each short period of time Δt has a normal distribution with mean equal to zero and a variance that is proportional to time.

where X is the state vector. The coefficients $A(\tau)$ and $B(\tau)$ are functions of the time to maturity, $\tau = T - t$, the parameters of the interest rate process and the market price of interest rate risk.

The set of prices of zero-coupon bonds as a function of time to maturity, $\tau = T - t$ will define the zero-coupon yield curve $R(t,T)$, where

$$R(t,T) = -\frac{1}{\tau}\ln[P(t,T)] = \frac{B(\tau)X - \ln A(\tau)}{\tau} \qquad (3)$$

The affine yield class property is displayed in equation (3). The zero-coupon yields are affine functions of the underlying factors, in this example the instantaneous short rate. For models where both the drift and volatility specifications are affine in r, it is possible to have closed form formulae for $A(\tau)$ and $B(\tau)$. Both the Vasicek (1977) and CIR (1985) models fulfil this criterion resulting in closed form solutions for the prices of pure discount bonds.

3 The Cox, Ingersoll, and Ross (1985) Model

The CIR model is characterised by one factor, the instantaneous interest rate r, that evolves in continuous time as described by the following first-order differential equation,

$$dr = k(\theta - r)dt + \sigma\sqrt{r}dw \qquad (4)$$

Interest rates appear to be pulled back to some long-term average level over time and this phenomenon is known as mean reversion. Therefore, the drift term includes a long-term mean parameter, defined as θ, and a mean reversion parameter denoted k. When the short rate deviates from its long-term mean, θ, it will revert back to this mean at a speed governed by the parameter k. This process is hampered in its ability to revert back to its mean level by the diffusion term, which essentially shocks the process at each step in time. This model is time homogeneous in the sense that neither the drift nor volatility terms are a function of time. By virtue of the square root process interest rates are prevented from becoming negative and are conditionally heteroskedastic i.e. the volatility of the short-term interest rates increases with an increase in the level of short-term interest rates. dw is a Wiener process. Gaussian processes like the Vasicek (1977) model and the square-root processes as proposed in the CIR (1985) model are the most popular versions of affine diffusions. While Gaussian processes have a constant variance matrix, square root processes introduce conditional heteroskedasticity by allowing σ to depend on the state. However, given the apparent stochastic properties of the volatility of interest rates, Gaussian or constant volatility models imply an element of simplification. In this study the movements in bond yields are estimated using the square root processes of the CIR model.

The absence of arbitrage would, intuitively, mean that assets which exhibit the same risk should earn exactly the same (excess) return. Thus in an arbitrage-free market, bonds of all maturities have the same market price of risk, which does not depend on maturity. Using risk-adjusted processes consistent with the absence of arbitrage, the effect of the market price of risk on the level of the short can be incorporated in the model. Therefore, the CIR process given by equation (4) can be represented as:

$$dr = (k(\theta - r) - \lambda r)dt + \sigma\sqrt{r}dw \tag{5}$$

where λ is the market value of risk. For the one-factor CIR model, the solution for the nominal price of a pure discount bond is given by

$$P(t,T) = A(t,T)e^{-B(t,T)r} \tag{6}$$

where, after incorporating the market value of risk, λ,

$$A(t,T) = [\frac{2\gamma e^{(k+\gamma+\lambda)(T-t)/2}}{(\gamma + k + \lambda)(e^{\gamma(T-t)} - 1) + 2\gamma}]^{2k\theta/\sigma^2} \tag{7}$$

$$B(t,T) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma + k + \lambda)(e^{\gamma(T-t)} - 1) + 2\gamma} \tag{8}$$

$$\gamma = \sqrt{(k+\lambda)^2 + 2\sigma^2} \tag{9}$$

The continuously compounded yield for discount bonds is given by:

$$R(t,T) = -\frac{\log P(t,T)}{T-t} \tag{10}$$

Using (6), this can be rewritten as:

$$R(t,T) = \frac{-\log A(t,T) + B(t,T)}{T-t} \tag{11}$$

4 Estimating the CIR model

A variety of methods have been developed in the finance literature for the estimation of CIR-type models. The two basic approaches may be characterised as the cross section approach and the time series approach.

In the cross-section approach, only information on the yields of bonds with different maturities at a point in time is used in the estimation process. This generates a different set of parameters for each time period. The state variable r_t, treated as an additional unknown parameter, is estimated jointly with the structural parameters. This solution is chosen when the purpose of the econometric analysis is to price derivative assets. The disadvantage of this approach is that the risk premium parameters cannot be identified because they are subsumed in the drift term.

Moreover, if the estimation is carried out sequentially at different points in time with different cross sections of rates, the estimated parameters can vary with sudden jumps when the observations have to contend with temporary shocks.

The time series approach, on the other hand, focuses on the dynamic implications of the model and ignores the cross-sectional information. A univariate time series approach is based on fitting equation (4) to estimate the parameters, using short-term observable data (e.g. the yield of one-month Treasury bills or money market rates) as an approximation of the unknown parameter estimates. In order to properly capture the information contained in the observed interest rates it would be necessary to use these rates across a range of maturities. However, if multivariate time series data are used it would give rise to an identification problem. The CIR model implies that any cross section of rates observed at time *t* is a function of the parameters (which are constant over time) and the value of the risk factors at time *t*. Therefore, using more interest rates than risk factors would result in the model becoming underidentified whereby its parameters cannot be consistently estimated. One solution is to allow for discrepancies between observed rates and the theoretical rates i.e. to introduce measurement errors in the relationship between observed rates and the state factors. These deviations can be explained by actual market features such as bid-ask spreads, rounding of prices, differences in the timing of observing financial variables and non-synchronous trading. In a modelling context this can be done by assuming that observed rates are affected by temporary shocks which are Gaussian white noise errors. Therefore equation (3) which is treated as an exact relationship between factors and yields would now read as:

$$R(\tau) = \frac{B(\tau)X}{\tau} - \frac{\ln A(\tau)}{\tau} + \varepsilon_t$$

Although the model is affine in the state vector *X*, the functions $A(\tau)$ and $B(\tau)$ are non-linear functions of the underlying parameters. So when this assumption about measurement errors is made, maximum likelihood estimation is no longer feasible, because the density of the yields is not available in closed form. Depending on the structure of the variance-covariance matrix of measurement errors, different estimation methods have been proposed using a panel-data approach.

A basic approach to resolving this estimation problem is to select as many different yields as factors and obtain the factors by inverting the model. Pearson and Sun (1994) followed this approach by formulating a likelihood function for a two-factor CIR model on the basis of the conditional density of the underlying factors. The model is estimated by replacing the two factors by two zero-coupon yields that are observed without error. Chen and Scott (1993) estimate a model with two factors and four maturities. In this case, the variance-covariance matrix of measurement errors has less than full rank. They assume that two yields are observed without error so that the model for these two maturities can be inverted directly to obtain the factors. The other yields are assumed to be measured with a normally distributed measurement error. The state variables can be uniquely determined and the inversion approach can be used to obtain the joint density function and therefore the log-likelihood function.

In the case where the variance-covariance matrix of measurement errors is assumed to be full rank, a quasi-maximum likelihood estimator based on the linear

Kalman filter is a common technique. The Kalman filter has been used in a series of papers dealing with the estimation of exponential affine term structure models. The Kalman filter is a linear estimation method and makes use of the assumption of an affine relationship between bond yields and state variables to subsequently estimate the parameter set. The main advantage of this technique stems from the fact that it allows the state variables to be unobserved magnitudes.

The nature of the application of the Kalman filter depends on whether the term structure model is Gaussian such as the Vasicek model or non-Gaussian such as the CIR model. A Gaussian distribution is fully characterised by its first two moments and the exact likelihood function is obtained as a by-product of the Kalman filter algorithm. An example of the Gaussian case is provided in Babbs and Nowman (1999), who estimated a two-factor generalised Vasicek model. Babbs and Nowman (1999) observed eight spot rates with maturities between one and ten years. When using non-Gaussian models the exact likelihood function is not available in closed-form, however a quasi-maximum likelihood estimator can be constructed from the first and second conditional moments of the state variables. Examples of the non-Gaussian CIR model, may be found in Duan and Simonato (1995), Lund (1997) and Geyer and Pichler (1999). De Jong (2000) provides an empirical analysis of the affine class of term structure models proposed by Duffie and Kan (1996) using a quasi-maximum likelihood estimator.

Markov chain Monte Carlo estimation is an alternative to the Quasi Maximum Likelihood approach and has recently been proposed by Lamoureax and Witte (2002). The main drawback of this approach is that it turns out to be computationally extremely time consuming because the state variables evolve very slowly. Lamoureaux and Witte (2002) report that it takes more than five days on a very sophisticated machine to obtain a sufficient number of iterations for a two-factor model.

In this paper, a panel-data estimation of the CIR model is presented from multivariate time series data. Combined use of time series and cross section data as entailed in the panel data approach allows for the identification of the market price of interest-rate risk, which is not identified from each dimension separately. Panel data estimation also provides an effective specification of the model. Its cross section dimension captures the restrictions imposed by the model on the parameters of the bond pricing equations and its time series dimension captures the dynamic model for the state variables.

The approach is based on a state-space representation of the term structure model where the underlying state variable(s) is treated as unobservable. This obviates the need to employ proxies for the unobserved factors. The yields are affine in the underlying state variables and the model explicitly allows for measurement errors. Quasi-maximum likelihood estimates of the model parameters are obtained by using an approximate Kalman filter to calculate the likelihood function.

5 The state space representation

This section demonstrates the reformulation of the CIR model given by equation (5) in the state space form and draws on the explanations provided in Harvey (1992). This formulation includes a *measurement equation* that relates the observable, or measurable bond yields to the unobservable state variables. The unobservable state variables are, in turn, assumed to follow a Markov process described by the *transition equation*.

Let the state vector X be a Markov process with $X_0 \approx p(X_0)$ and $X_t \mid X_{t-1} \approx p(X_t \mid X_{t-1})$. $P(X_0)$ is the density of the initial state and $P(X_t \mid X_{t-1})$ is the transition density. The exact transition density of the state variable for the CIR model is a non-central chi-square, $\chi^2[2cX_t; 2q+2, 2u]$, with $2q+2$ degrees of freedom and noncentrality parameter, 2*u*. (CIR 1985). Estimation of the unobservable state variables by the Kalman filter coupled with a quasi-maximum likelihood estimation of the model parameters can be accomplished by substituting the exact transition density by a Gaussian or normal density. Therefore, the probability density of the state vector at time *t*, conditional on its value at time, *t*-1, should be distributed in a manner such that:

$$X_t \mid X_{t-1} \approx N(\mu_t, Q_t)$$

where μ_t and Q_t are distributed in such a way that the two moments of the approximate normal and exact transition density are equal. The elements of a j x 1 vector μ_t would be defined as

$$\mu_{t,j} = \theta_j[1 - e^{-k\Delta t}] + X_{t-1,j}e^{-k_j\Delta t} \qquad (12)$$

where $\Delta t =$ the time interval between *t* and *t*-1.

The matrix Q_t is diagonal and is dependent on the state of the process. For a three-factor model, the conditional variance of the transition system would have the following form:

$$Q_{t_i} = \begin{bmatrix} \xi_1 00 \\ 0\xi_2 0 \\ 00\xi_3 \end{bmatrix},$$

where $\xi_j = \dfrac{\theta_j \sigma_j^2}{2k_j}(1 - e^{-k_j\Delta t})^2 + \dfrac{\sigma_j^2}{k_j}(e^{-k_j\Delta t} - e^{-2k_j\Delta t})X_j(t_{i-1}) \qquad (13)$

for $j = 1, 2, 3.$

Yields on zero-coupon bonds are the inputs to the estimation process. Eight maturities have been chosen that span the yield curve from 2 years to 25 years in order to incorporate information affecting trading at the short, medium and long ends of the yield curve.

In the CIR model, the measurement equation represents the affine relationship between zero coupon bond yields and the state variables. Under the assumption that measurement errors in bond yields are additive and normally distributed, the measurement equation for observed yields is given by:

$$R_t = Z(\psi)X_t + d(\psi) + \varepsilon_t, \qquad \varepsilon_t \approx N(0, H) \tag{14}$$

where $\psi = (\theta, \kappa, \sigma, \lambda, h_{1,\ldots N})$ is a vector of hyperparameters which contains the unknown parameters of the model including the parameters from the distribution of measurement errors. R_t is the *n* x 1 vector of observations, X_t is the unobservable j x 1 state vector at time *t*, Z is an n x j matrix, d is an n x 1 vector, ε_t is an *n* x 1 vector of measurement errors. H is the variance-covariance matrix of ε_t. In this estimation the number of observed bonds and the associated maturities do not change over time. Therefore, *H* has a constant dimension of *n* x *n* and is assumed to be a diagonal matrix. As 8 different maturities are considered in this estimation, the variance-covariance matrix of the measurement errors, H, is an 8 x 8 diagonal matrix.

$$H = \begin{bmatrix} h_1^2 0 \cdots 0 \\ 0 h_2^2 \cdots 0 \\ \vdots \vdots \ddots \vdots \\ 0 0 \ldots h_8^2 \end{bmatrix}$$

The values in the diagonal would differ implying that the variance of measurement errors will depend on the maturities under consideration. This can be justified on the grounds that trading activity and, therefore, bid-ask spreads are not equally distributed across maturities. In the case of a one-factor affine term structure model, equation (14) would read as:

$$R_t = \frac{B(t,T)}{T-t} X_t - \frac{\log A(t,T)}{T-t} + \varepsilon_t \qquad \varepsilon_t \approx N(0,H)$$

The stochastic differential equation (5) represents the dynamics of the state variable as specified in continuous time. As the transition equation captures the discrete dynamics of the state variable, it corresponds to the discrete time version of equation (5). This, along with a first order autoregression model, is used to formulate the transition equation,

$$X_t = \phi(\psi) X_{t-1} + c(\psi) + \eta_t, \;\; E(\eta_t \,|\Im_{t-1}) = 0, \quad \mathrm{var}(\eta_t \,|\Im_{t-1}) = Q_t \tag{15}$$

where

$c = \theta_j(1 - e^{-k_j \Delta t})$ is j x 1 vector and $\phi = e^{-k_j \Delta t}$ is a j x j diagonal matrix

$\Delta t =$ the time interval in the discrete sample (here 1 week)

and so the discretisation step $\Delta t = 1/52$ for weekly data.

η_t is g x 1 vector of disturbance terms with mean zero and variance-covariance matrix Q_t and where $\Im_{t-1}$ represents the information available at time t - 1.

It is further assumed that the error terms of the measurement (ε_t) and transition equations (η_t) are not correlated.

6 The Kalman Filter

Now that the model in (5) has been put in state space form, as defined in equations (14) and (15) and summarised below, the Kalman filter can be used to obtain information about X_t from the observed zero coupon yields.

Measurement Equation:

$$R_t = Z(\psi)X_t + d(\psi) + \varepsilon_t, \qquad \varepsilon_t \approx N(0, H)$$

Transition Equation:

$$X_t = \phi(\psi)X_{t-1} + c(\psi) + \eta_t, \quad E(\eta_t \,|\, \Im_{t-1}) = 0, \quad \text{var}(\eta_t \,|\, \Im_{t-1}) = Q_t$$

where $\psi = (\theta, \kappa, \sigma, \lambda, h_{1,\ldots N})$ is a vector of hyperparameters which contains the unknown parameters of the model.

A detailed explanation of the Kalman filter can be found in Harvey (1992) and Lutkepohl (1991). The Kalman filter recursion is a set of equations which allows an estimator to be updated once a new observation becomes available. It first forms an optimal predictor of the unobserved state variable vector given its previously estimated value. This prediction is obtained using the distribution of the unobserved state variables, conditional on the previous estimated values. These estimates for the unobserved state variables are then updated using the information provided by the observed variables. Although the Kalman filter relies on the normality assumption of the measurement error and initial state vector, it can calculate the likelihood function by decomposing the prediction error.

Consider the conditional distribution of the state vector X_t given information at time *s*. The mean and covariance matrix of this distribution can be defined as

$$\hat{X}_{t/s} = E_s(X_t) \tag{16}$$

$$P_{t|s} = E_s([(X_t - \hat{X}_{t|s})(X_t - \hat{X}_{t|s})']) \tag{17}$$

where the expectations operator indicates that expectations are formed using the conditional distribution for that period.

To obtain the one-step ahead mean, $\hat{X}_{t|t-1}$ and covariance, $P_{t|t-1}$ of X_t we use the conditional distribution implied by setting $s = t-1$. This yields the following *prediction equations*

$$\hat{X}_{t|t-1} = E_{t-1}(X_t) = \phi(\psi)X_{t-1} + c(\psi) \tag{18}$$

where $\hat{X}_{t|t-1} = E_{t-1}(X_t)$

$$P_{t|t-1} = \phi(\psi)P_{t-1}\phi(\psi)' + Q_t \tag{19}$$

where $P_{t|t-1} = E_{t-1}[(X_t - \hat{X}_{t|t-1})(X_t - \hat{X}_{t|t-1})']$

To calculate the prediction equations we need to assume initial values for the elements of the state vector in the previous period, $\hat{X}_{t-1}$ and the system matrices $\phi(\psi)$, $c(\psi)$ and $Q(\psi)$. Starting values of X_0 and P_0 are provided.

The second step in calculating the Kalman filter is to revise the estimation from step-one using the updating equations that are actual observations which are based on actual observations of R available at time t. The updating equations are given by

$R_{t\|t-1} = Z\hat{X}_{t\|t-1} + d$;	estimation of R_t	(20)
$v_t = R_t - R_{t\|t-1}$;	observation vector estimation error	(21)
$F_t = ZP_{t\|t-1}Z_t' + H$;	covariance matrix of $R_{t\|t-1}$	(22)
$K_t = P_{t\|t-1}Z_t'F_t^{-1}$;	Kalman gain	(23)
$\hat{X}_t = \hat{X}_{t\|t-1} + K_t v_t$;	updating of the state vector	(24)
$P_t = P_{t\|t-1} - K_t Z_t P_{t\|t-1}$;	updating of state covariance matrix	(25)

The prediction and update steps must be repeated for each discrete-time step in the data sample. For the analysis in this chapter , weekly observations over a period of five years were used.

The intuition underlying the Kalman filter is that $\hat{X}_t$ is the best linear approximation of the true state vector X_t, if the state vector estimation error, $(X_t - \hat{X}_t)$ is independent of past and present observations R_t, i.e.

$$Cov[(X_t - \hat{X}_t), R_s] = 0 ; \quad s = 1, \ldots, t. \qquad (26)$$

The Kalman gain, K_t defined in equation (23) is derived to ensure that the above condition holds. In order to elaborate on this, one starts by assuming that the state vector estimation error, $(X_t - \hat{X}_t)$ is equal to the difference between the true state vector, X_t and the prediction of the state vector based on information in the previous period, $\hat{X}_{t|t-1}$ net of a proportion, K_t of the observation vector estimation error, $(R_t - R_{t|t-1})$, i.e.

$$(X_t - \hat{X}_t) = (X_t - \hat{X}_{t|t-1}) - K_t(R_t - R_{t|t-1}) . \qquad (27)$$

Equation (26) implies the state updating equation given by equation (24) i.e.

$$\hat{X}_t = \hat{X}_{t|t-1} + K_t v_t$$

where $v_t = (R_t - R_{t|t-1})$ which was defined in equation (21).

The above discussion implies that for the observations R_s, s = 1, …, t-1 and any arbitrary matrix K_t the following condition must hold

$$[Cov[X_t - \hat{X}_t, R_s]] = Cov[\{(X_t - \hat{X}_{t|t-1}) - K_t(R_t - R_{t|t-1})\}, R_s] = 0$$

$$= Cov[(X_t - X_{t|t-1}), R_s] - K_t Cov[(R_t - R_{t|t-1}), R_s] = 0 \qquad (28)$$

s = 1,…., t - 1.

As discussed in Duan and Simonato (1998), when the state space model is Gaussian, the Kalman filter provides an optimal solution to predicting, updating and evaluating the likelihood function. When the state-space model is non-Gaussian, the Kalman filter can still be applied to obtain approximate first and second moments of the model and the resulting filter is quasi-optimal. The use of this quasi-optimal filter yields an approximate quasi-likelihood function with which the parameter estimation can be carried out.

Quasi-Maximum likelihood estimation

In the state space form described above it is not possible to write the density of the observations $R_1,...,R_n$ directly, because the conditional density is assumed. The joint density function of the *n* x 1 vector of observations is given by

$$\ln L(R_1,...R_n;\psi)=\prod_{t=1}^{n} p(R_t \left|\Im_{t-1}\right.),$$

where ψ is a vector of hyperparameters and $p(R_t \left|\Im_{t-1}\right.)$ is the distribution of R_t conditional on the information set, $\Im$ at time *t* - 1. Given the information set $\Im_{t-1}$, the true state vector is normally distributed with mean $\hat{X}_{t|t-1}$ and covariance matrix P_t. Hence, R_t is also normally distributed with mean $R_{t|t-1}=Z_t\hat{X}_{t|t-1}+d_t$ and error variance-covariance matrix F_t.

Assuming that the prediction errors are normally distributed, the log-likelihood function is given by,

$$\log L(R_1,...,R_n;\psi)=-\frac{n}{2}\log 2\pi-\frac{1}{2}\sum_{t=1}^{n}\log\left|F_t\right|-\frac{1}{2}\sum_{t=1}^{T}v_t^{'}F_t^{-1}v_t \quad \ldots (29)$$

Since the prediction error is Gaussian, equation (29) is the quasi maximum likelihood estimator which best explains the observed values of R_t. Both F_t and v_t depend upon the parameter set given by ψ. Therefore, ψ is chosen so as to maximise the likelihood function $\log L$.

7 Data and Estimation Results

Data description

The data comprises 265 weekly observations of zero-coupon yields for UK and German Treasury bonds from January 6 1999 to January 28, 2003. These observations were sampled every Wednesday to take advantage of high liquidity and avoid beginning and end of week effects. The data sets have a panel data structure with a time dimension and a cross-sectional (maturity) dimension. For the UK, the data set used here are zero coupon yields available in the Bank of England public domain yield curve database. In the case of Germany, zero coupon yields on euro-denominated bonds have been sourced from Reuters. Eight different maturities that would broadly cover the maturity spectrum of the yield curve are considered; they are 2-, 3-, 5-, 7-, 10-, 15-, 20- and 25-year bonds. Table 1 provides the summary statistics for the estimated zero coupon yields.

Table 1
Summary statistics of zero coupon yields:
Germany and UK (Jan 1999 to Jan 2004)

Maturity	Mean Yield		Standard Deviation	
years	GER	UK	GER	UK
2	4.18	4.89	0.87	0.86
3	4.36	4.97	0.79	0.78
5	4.61	5.00	0.67	0.64
7	4.83	4.97	0.56	0.52
10	5.05	4.87	0.51	0.35
15	5.12	4.73	0.50	0.21
20	5.55	4.60	0.39	0.22
25	5.56	4.48	0.39	0.24

Figure 1 shows the dynamic path of the UK term structure between January, 1999 and January, 2004. Similarly, Figure 2 shows the dynamic path of the German term structure over the aforesaid period.

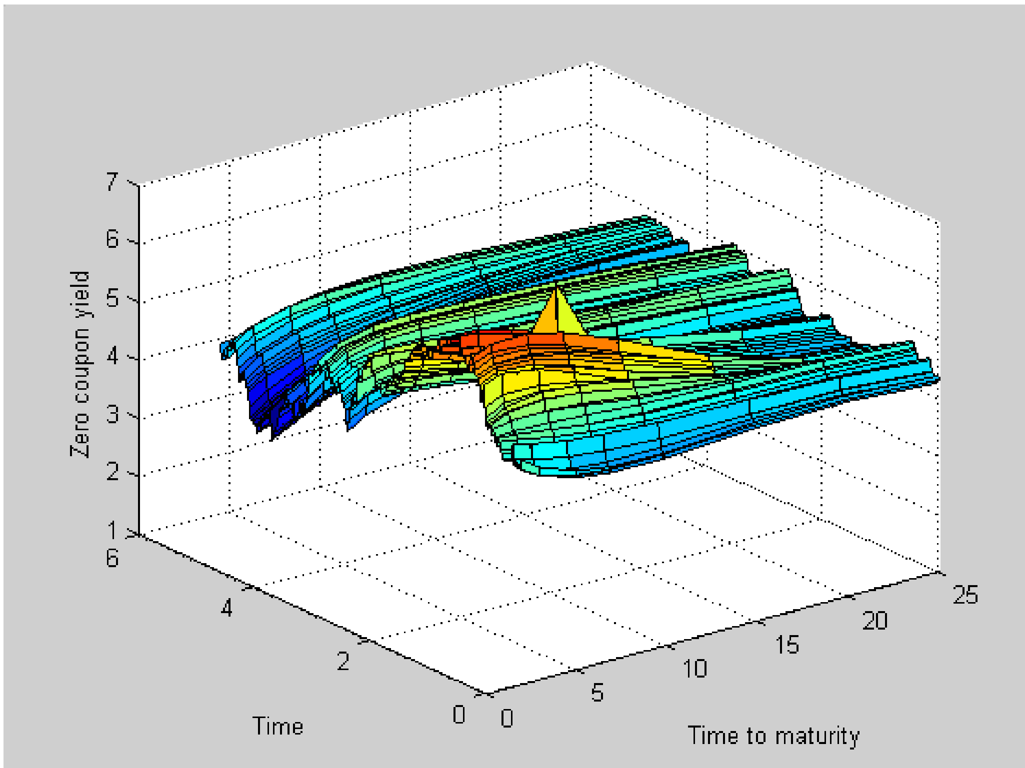

Figure 1 Dynamic path of the UK Term Structure (Jan'99 -Jan'04)

In contrast to the UK, the German term structure has evolved in a steady manner with no dramatic changes. It has maintained an upward slope during the five period since the launch of the euro.

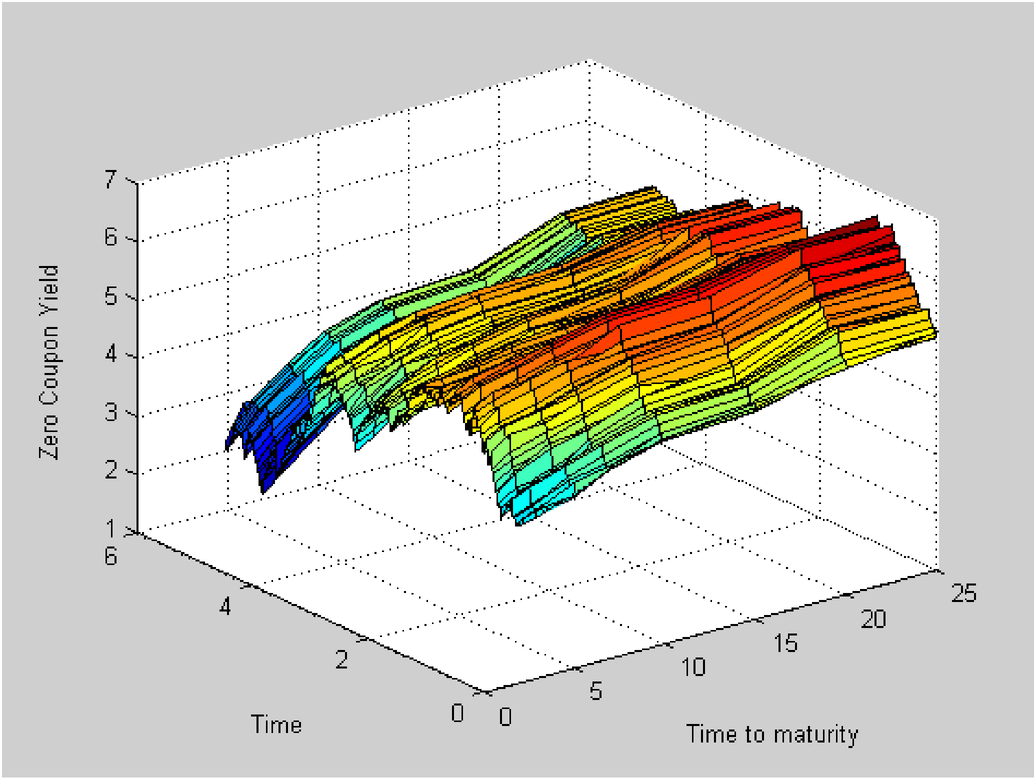

Figure 2 Dynamic path of the German term structure (Jan'99 -Jan'04)

Parameter Estimation

The Kalman filter was used to estimate the one-factor CIR model using data on the UK and German term structure of interest rates. The objective was to estimate the parameters of the processes that are posited to drive interest rate changes.

The standard errors of the parameter vector $\psi = (\kappa, \theta, \sigma, \lambda, h_1, \cdots, h_8)$ can be computed by using the result shown by White (1982). He showed that the covariance matrix for $\sqrt{n}(\hat{\psi} - \psi)$ converges to

$$\left[E\left(\frac{\partial^2 L}{\partial \psi_i \partial \psi_j} \right) \right]^{-1} E\left(\frac{\partial L}{\partial \psi_i} \frac{\partial L}{\partial \psi_i} \right) \left[E\left(\frac{\partial^2 L}{\partial \psi_i \partial \psi_j} \right) \right]^{-1}$$

where L is the log-likelihood function. The standard errors are given by the diagonals of the above matrix result. Thus for each observation, the partial derivatives of the likelihood with respect to the twelve parameters $\psi = (\kappa, \theta, \sigma, \lambda, h_1, \cdots, h_8)$ were numerically determined, evaluated at the maximum likelihood estimate $\hat{\psi}$.

The elements of $\frac{\partial L}{\partial \psi_i}$ can be computed by using the symmetric central difference method.

$$\frac{\partial L}{\partial \psi_i} = \frac{L(\psi_i + \delta_i) - L(\psi_i - \delta_i)}{2\delta_i}$$

Diagonal elements of $\dfrac{\partial^2 L}{\partial\psi_i \partial\psi_j}, (i = j)$

These are $\dfrac{\partial^2 L}{\partial\psi_i^2}$ and can be computed using the symmetric central difference method.

$$\frac{\partial^2 L}{\partial\psi_i^2} = \frac{\dfrac{L(\psi_i + \delta_i) - L(\psi_i)}{\delta_i} - \dfrac{L(\psi_i) - L(\psi_i - \delta_i)}{\delta_i}}{\delta_i}$$

or

$$\frac{\partial^2 L}{\partial\psi_i^2} = \frac{L(\psi_i + \delta_i) - 2L(\psi_i) + L(\psi_i - \delta_i)}{\delta_i^2}$$

Off-Diagonal elements of $\dfrac{\partial^2 L}{\partial\psi_i \partial\psi_j}, (i \neq j)$

These are $\dfrac{\partial^2 L}{\partial\psi_i \partial\psi_j}$ and can be computed along each axis (*i* or *j*) in turn so that

$$\frac{\partial^2 L}{\partial\psi_i \partial\psi_j} = \frac{\dfrac{L(\psi_i + \delta_i, \psi_j + \delta_j) - L(\psi_i - \delta_i, \psi_j + \delta_j)}{2\delta_i} - \dfrac{L(\psi_i + \delta_i, \psi_j - \delta_j) - L(\psi_i + \delta_i, \psi_j - \delta_j)}{2\delta_i}}{2\delta_j}$$

$$\frac{\partial^2 L}{\partial\psi_i \partial\psi_j} = \frac{L(\psi_i + \delta_i, \psi_j + \delta_j) - L(\psi_i - \delta_i, \psi_j + \delta_j) - L(\psi_i + \delta_i, \psi_j - \delta_j) + L(\psi_i - \delta_i, \psi_j - \delta_j)}{4\delta_i \delta_j}$$

Estimation Results

In keeping with the different dynamics of the term structure observed in the two markets different starting values are chosen. For the UK term structure, the true values or initial starting values chosen for the parameters were $\kappa = 0.15$, $\theta = 0.05$, $\sigma = 0.1$, $\lambda = -0.1$. Results of the parameter estimation using the Kalman filter over the entire observation period from January, 1999 to January, 2004 are shown in Table 2. Figures in parenthesis indicate t-values.

Table 2 The Kalman Filter estimates of the one-factor CIR model for UK Treasury bond yields from 06.01.1999 to 28.01.2004

κ	θ	σ	λ
0.1443	0.0879	0.0801	-0.1176
(3.45)	(3.46)	(3.76)	(2.53)

Significant parameter estimates were obtained for all the parameters at the 5% level The significant mean reversion parameter of 0.1443 implies mean reversion in the underlying interest rate. The estimate of 0.1443 indicates a mean half life of 4.8 years which is the expected time for the short rate to return halfway to its long-run average mean, θ.[2] Half-life gives the *slowness* of the mean reversion process and a value of 4.8 years would indicate slow mean reversion for interest rates. Accordingly, this process is also characterised by a low but significant volatility estimate (σ = 0.0801).

The market price of risk (λ = -0.1176) is negative, a necessary condition for positive risk premia. The result implies that the risk premium for holding long term bonds is positive.

In the case of the German term structure, the initial starting values chosen for the parameters were $\kappa = 0.15$, $\theta = 0.04$, $\sigma = 0.05$, $\lambda = -0.1$. Results of the parameter estimation using the Kalman filter are shown in Table 3. Figures in parenthesis indicate t-values.

Table 3 The Kalman Filter estimates of the one-factor CIR model for German Treasury bond yields from 06.01.1999 to 28.01.2004

κ	θ	σ	λ
0.1579	0.0646	0.0556	-0.00095
(20.83)	(15.1)	(2.37)	(0.12)

Parameter estimates are significant for all the parameters except the market price of risk. This would suggest that this variable has not been priced by the market. In accordance with the lower level of short-term yields for German Treasury bonds, the long-term mean parameter is 6.46 per cent as compared to 8.79 per cent for the UK Treasury. The mean reversion of 0.1579 implies a mean half-life of 4.38 years and this is somewhat smaller than that obtained for the UK term structure. However, the

[2] The half life is given by $e^{-kt} = 0.5$. This implies $t = -\ln(0.5)/k$

volatility parameter given by 0.0556 is significantly smaller than that obtained for the UK term structure.

6 Conclusion

In this chapter a single-factor CIR model has been estimated for the UK and German term structure for the period January, 1999 to January, 2004. Modelling continuous time term structure models, started with the specification of a time series process for the instantaneous spot interest rate. The no-arbitrage condition then permits the derivation of a bond pricing formula whereby the bond price is a function of the unobserved instantaneous spot rate and the model's parameters. These parameters are the long-run mean, the speed of adjustment towards the long-run mean, the volatility of the short-term interest rate and the market price of risk. The model was estimated for a single factor using a quasi maximum likelihood approach based on the Kalman filter. The Kalman filter algorithm uses observable data on bonds to extract values for the unobserved state variables. It combines both the cross section and time series information in the term structure.

Yields on zero-coupon bonds were used as inputs for the estimation process. The empirical analysis was based on weekly observations of UK and German Treasury zero coupon bonds over the period January 1999 to January 2004. Eight maturities were chosen that spanned the yield curve from 2 years to 25 years and were expected to incorporate influences on the short, medium and long end of the term structure. The parameters of the model and their standard errors were estimated.

Results of the empirical analysis showed that the unobserved instantaneous interest rate exhibits mean reverting behaviour in both the UK and German term structure. However, the mean reversion of the interest rate process has been relatively slower in the UK as compared to Germany since the introduction of the euro. Accordingly, the volatility component, which shocks the process at each step in time was also higher in the UK as compared to Germany. The results indicated that the one-factor CIR model provides a good representation of the UK Gilt-Edged market. However, its inability to meaningfully account for the market price of risk has impinged on its efficacy in capturing the dynamics of the German term structure.

References

Babbs, S.H. and K.B. Nowman (1999). "Kalman Filtering of Generalised Vasicek Term Structure Models" . *Journal of Financial and Quantitative Analysis,* pp 115-130.

Bolder, D.J. (2001), "Affine Term-Structure Models: Theory and Implementation." *Bank of Canada Working Paper 2001-15*

Chen, R., and L. Scott (1993). "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates." *Journal of Fixed Income,* pp. 14 - 31

Cox J.C., J.E. Ingersoll and S.A. Ross, (1985)." A Theory of the Term Structure of Interest Rates." *Econometrica,* 53 pp. 385-407.

De Jong, F. (2000). "Time-series and cross-section information in affine term structure models." *Journal of Business and Economic Statistics* 18, 300-314.

Duan J.C. and J.G. Simonato (1998). "Estimating and testing exponential affine term structure models by the Kalman Filter." *Review of Quantitative Finance and Accounting*.

Duffie, D. and R. Kan (1996). "A Yield-Factor Model of Interest Rates," *Mathematical Finance,* 6, 379-406.

Durham, G.B., and A.R. Gallant (2002). "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes", *Journal of Business and Economic Statistics,* 20, 325-327.

Geyer, A.L.J., and Pichler, S., (1999). "A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure of Interest Rates," *Journal of Financial Research,* Vol 22(1), 107-130.

Harvey, A.C. (1989). "Forecasting, Structural Time Series Models and the Kalman Filter." (Cambridge University Press)

Lamoureaux, C., and H.D. Witte (2002). "Empirical analysis of the yield curve: the information in the data viewed through the window of Cox, Ingersoll and Ross," *Journal of Finance*, 57, 1479-1520.

Litterman, R. and Scheinkman, J. (1991). "Common factors affecting bond returns," *Journal of Fixed Income* (June): 54-61

Longstaff, F.A. and E.S. Schwartz (1992a): "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance,* XLVII, 1259-1282.

Lund, J. (1997). "Econometric Analysis of Continuous-Time Arbitrage-Free models of the Term Structure of Interest Rates", Department of Finance, The Aarhus School of Business

Lutkepohl, H. (1991). "Introduction to Multiple Time Series Analysis", *Springer*

Nath, P. and K. Ben Nowman, (2001). "Estimates of the continuous time Cox-Ingersoll-Ross term structure model: further results for the UK gilt-edged market", *Applied Economic Letters*, 2001, 8, 85-88.

Pearson, N.D., and T.S. Sun (1994). "Explaining the Conditional Density in Estimating the Term Structure: An Application to the Cox-Ingersoll-Ross Model," *Journal of Finance*, XLIX, 1279-1304.

Steeley, J.M. (1997). "A two-factor model of the UK yield curve," *The Manchester School*, Vol LXV, 32-58.

Vasicek, O., (1977). "An Equilibrium Characterisation of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

White, H. (1982), “Maximum Likelihood Estimation of Misspecified Models”, *Econometrica*, 50, 1-25.