

# Cross sectional efficient estimation of stochastic volatility short rate models

Dmitri Danilov      and      Pranab K. Mandal<sup>1</sup>  
KUB, Tilburg                      EURANDOM, Eindhoven

## Abstract

We consider the problem of estimation of term structure of interest rates. Filtering theory approach is very natural here with the underlying setup being non-linear and non-Gaussian. Earlier works make use of Extended Kalman Filter (EKF). However, as indicated by de Jong (2000), the EKF in this situation leads to inconsistent estimation of parameters, though without high bias. One way to avoid this is to use methods like Efficient Method of Moments or Indirect Inference Method. These methods, however, are numerically very demanding. We use Kitagawa type scheme for nonlinear filtering problem, which solves the inconsistency problem without being numerically so demanding.

## 1 Introduction

Term structure of interest rates is a set of yields to maturity, at a given time, on bonds with different maturity dates. Typically, a model for the term structure of zero coupon bonds consists of a dynamic model for the evolution of the factors that influence the short term interest rate – *time series dimension* and a model for bond prices (or yields) as a function of the factors and the time to maturity – *cross sectional dimension*. One of the important parameters that links both models together is the market price of risk. The market price of risk is a key factor to price interest rate derivatives. It can be estimated from the yield data, provided the factors in time series dimension are known. On the other hand, by analyzing the time series data separately one can obtain the factor values. However, as we shall see in section 2, this method does not provide good estimate of the term structure for long maturities. The standard error for these estimates are very high. One of the possible reasons is that this method probably does not use all the information about the factor values contained in the cross-sectional dimension. An alternative approach is to use both the dimensions simultaneously for estimation. Recent literatures on this includes Babbs and Nowman (1999), Chen and Scott (1993), de Jong (2000), Frachot, Lesne and Renault (1995), Geyer and Pichler (1999), Pagan and Martin (1996), and Pearson and Sun (1994).

Application of filtering techniques in the estimation of term structure models using cross-sectional/time series data has been investigated by many. See, for example, the most recent works by Babbs and Nowman (1999), and de Jong (2000)

---

<sup>1</sup>Now in the University of Twente, Enschede

and the references therein. Babbs and Nowman (1999) considered general linear Gaussian model of the term structure while de Jong (2000) considered so called affine term structure model. Both articles used Kalman filtering technique, valid only for those models that are linear in nature and with Gaussian errors. However, one may have to consider nonlinear models and/or models with non-Gaussian error for better fit to the data. In fact, de Jong (2000) has such type of model and as pointed out in the article, the use of Kalman filtering techniques in this situation leads to inconsistent estimation of parameters though without high bias. It is also mentioned that the inconsistency problem can be avoided by using other estimation procedures such as indirect inference method of Gourieroux, Montfort and Renault (1992) or the efficient method of moment (EMM) of Gallant and Tauchen (1996), but these methods are numerically very demanding.

In this short note we want to explore the use of readily available filtering techniques suitable for nonlinear models with non-Gaussian errors. Namely, we shall use Kitagawa (1987) type filtering scheme used by Danilov and Mandal (2000) to estimate stochastic volatility in two factor short rate models. Similar technique is also used in Fridman and Harris (1998), and Hasbrouck (1999) in the time series dimension. The advantages are many-fold. Firstly, it will resolve the inconsistency problem faced by de Jong (2000). Secondly, this would not be as numerically demanding as the indirect inference or the EMM. Finally, this scheme still provides a deterministic way to evaluate the likelihood function and is numerically less intensive than simulated likelihood method or MCMC method. Even though our method is applicable to multifactor models, in order to reduce the amount of technicalities we shall illustrate the approach on two one-factor models: “affine” Cox, Ingersol, and Ross (CIR)(1985) model, and “nonlinear” Longstaff (1989) model. For the purpose of comparison we estimate one-factor Vasicek (1977) model as well.

Usually, the cross-sectional dimension of the term structure model gives a deterministic relation : yield  $Y$  as a function of the factors in the time-series model and some other parameters such as market price of risk. However, when using more bond values (of different maturities) than the number of factors, this exact relation may not (and for some models, such as affine ones, cannot) be satisfied by all the data in the yield vector. There could be several reasons for this. The exact no-arbitrage relation can be violated due to market imperfections. To some degree, prices are affected by seasonal and business oscillations. Most of all there could be measurement/human errors. We, therefore, as in de Jong (2000), assume that

$$Y^*(\tau, t) = Y(\tau, t) + \varepsilon_t(\tau),$$

where  $Y^*(\tau, t)$  is the observed yield at time  $t$  with maturity at  $\tau$ ,  $Y(\tau, t)$  is the theoretical yield given by the cross-sectional dimension of the model and  $\varepsilon_t(\tau)$  is the

observational error. We also assume that the model captures all time dependence via the factors at time  $t$  and, therefore, we assume that the errors of observation  $\varepsilon_t(\tau)$  are uncorrelated with respect to time  $t$  but, of course, may be correlated with respect to maturity  $\tau$ .

The article is organized as follows. In section 2 we treat the time series data and cross-sectional data separately and estimate the term structure for two one-factor models : Vasicek (1977), and CIR (1985), and a two-factor model introduced by Fong and Vasicek (1991). The general setup of the one-factor model on which Kitagawa filtering scheme will be applied and the properties of specific models considered are described in section 3. In section 4 we explain the Kitagawa scheme used and the implementation to specific models. Empirical results are discussed in section 5. Finally, some conclusions are offered in section 6.

## 2 Need for cross sectional estimation.

In this section we estimate the term structure by analyzing the time series and cross sectional data separately. The estimation is done in 4 steps. First, the short rate model is estimated using 3-month Treasury Bill as a proxy for the short rate. Second, given the estimated parameters of the model the (stochastic) volatility component  $v_t$  was estimated. Third, given  $\tilde{v}_t, r_t$  and set of time series parameters we estimate the market price of risk by minimizing

$$\sum_t \sum_{k=1}^N (Y^*(\tau_k, t) - Y(\tau_k, r_t, v_t))^2,$$

where  $\tau_k$  are maturities of the observed yields. Finally, we calculate the estimated yields at each point of time and with different maturities from the cross-sectional model.

We have considered two one-factor models : Vasicek (1977), and CIR (1985), and one two-factor model : Fong and Vasicek (1991) given by

$$\begin{cases} dr_t &= \kappa(\mu - r_t)dt + \sqrt{v_t}dW_t^{(1)}, \\ dv_t &= \lambda(\nu - v_t)dt + \tau\sqrt{v_t}dW_t^{(2)}. \end{cases} \quad (1)$$

For estimation of the time series dynamic the EMM of Gallant and Tauchen (1996) was employed. As an auxiliary model we have used the semi-nonparametric (SNP) model of type AR(L)-ARCH(M)-Hermite(K,0):

$$f(y_t|\theta) = C[P_K(z_t)]^2 \phi(y_t|\mu_{x_{t-1}}, \Sigma_{x_{t-1}}),$$

Table 1: EMM estimates of parameters

Parameter	FV	Vasicek	CIR
$\mu$	6.520(18.83)	6.09(17.64)	7.54(79.30)
$\kappa$	0.109(3.19)	0.064(7.45)	0.019(13.77)
$\sigma$		1.702(16.02)	0.270(28.62)
$\nu$	2.640(9.26)		
$\lambda$	1.482(1.67)		
$\tau$	1.934(4.34)		

where

$$\begin{aligned}
 C & \text{ is } && \text{the normalising constant,} \\
 P_k & \text{ is } && \text{the Hermite Polynomial of degree } K, \\
 x_{t-1} & \equiv && (y_{t-L}, \dots, y_{t-1}) \text{ is the lag vector so that the conditional} \\
 & && \text{distribution of } y_t \text{ given all the past depends only on } x_{t-1}, \\
 \mu_{x_{t-i}} & = && \psi_0 + \psi_1 y_{t-i} + \psi_2 y_{t-i-1} + \dots + \psi_L y_{t-i-L+1}, \\
 \Sigma_{x_{t-1}} & = && R_{x_{t-1}}^2, \\
 R_{x_{t-1}} & = && \tau_0 + \tau_1 |y_{t-M} - \mu_{x_{t-M-1}}| + \tau_2 |y_{t-M-1} - \mu_{x_{t-M-2}}| + \\
 & && + \dots + \tau_M |y_{t-1} - \mu_{x_{t-2}}|, \text{ and} \\
 z_t & = && (y_t - \mu_{x_{t-1}}) / R_{x_{t-1}}.
 \end{aligned}$$

Estimation of the SNP model is done by maximum likelihood, and Schwarz's Bayes information criterion (BIC) (see Schwarz (1978) ) was used to determine the correct order of the model. Moment generating conditions in EMM were estimated by Monte-Carlo, averaging the estimated scores of the AR(2)-ARCH(4)-Hermite(6,0) on a series of 200000 weekly observations generated by application of the Euler discretisation scheme with 20 intervals per week to the system of SDE (1). The estimation results are reported in Table 1. For details see Danilov and Drost (2000)<sup>2</sup>.

Vasicek (1977), and CIR (1985) models do not require volatility estimation. For Fong and Vasicek (1991) model the Nonparametric Method of Conditional Moments (see Danilov and Mandal (2000) ) was employed for estimation of  $v_t$ .

<sup>2</sup>These parameter estimations are obtained when the data are expressed in percentages. Since in our analysis we use data in decimal points (divided by 100), the parameter values were renormalised appropriately.

The yield functions for all three models under consideration belong to affine class and well known (for one factor models they are also given in section 3). The market price of risk is defined canonically. Since we were not particularly interested in economical implications of the models, we did not calculate the standard errors.

As we have seen, steps 1 and 2 do not require a set of yields with different maturities, but step 3 does, where yields with 10 different maturities, from 1 year to 10 years were used.

Figures 1-3 show the quality of the fit for these three models. In the figures “stars” denote observed average term structure, i.e., the bond yields averaged over observed time points:  $\frac{1}{n} \sum_{t=1}^n Y^*(\tau_k, t)$ . The solid line connects the average fitted term structure points, i.e.,  $\frac{1}{n} \sum_{t=1}^n \hat{Y}(\tau_k, r_t, v_t)$ . Dotted lines represents RMSE error bounds:

$$\frac{1}{n} \sum_t \hat{Y}(\tau_k, r_t, v_t) \pm \sqrt{\frac{1}{n} \sum_{t=1}^n (Y^*(\tau_k, t) - \hat{Y}(\tau_k, r_t, v_t))^2}.$$

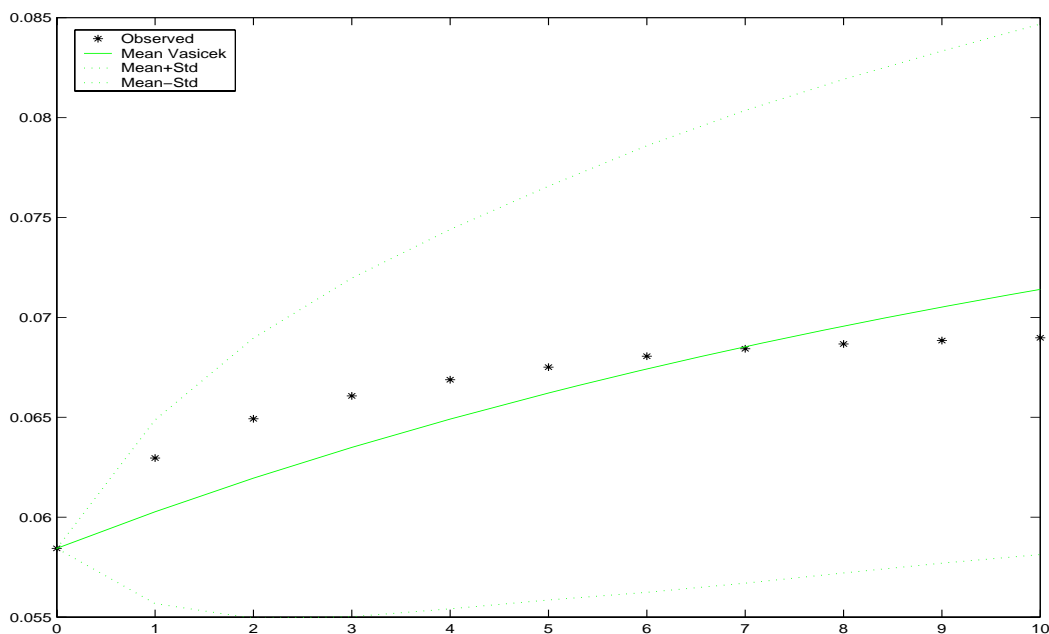


Figure 1. Fitted term structure of one factor Vasicek model (time series estimation).

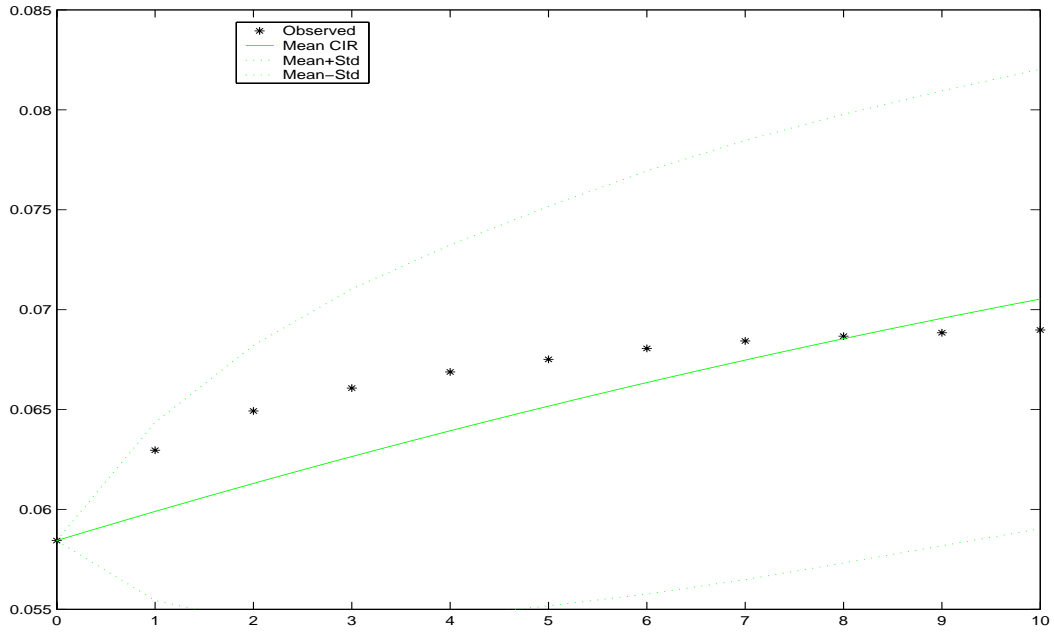


Figure 2. Fitted term structure of one factor CIR model (time series estimation).

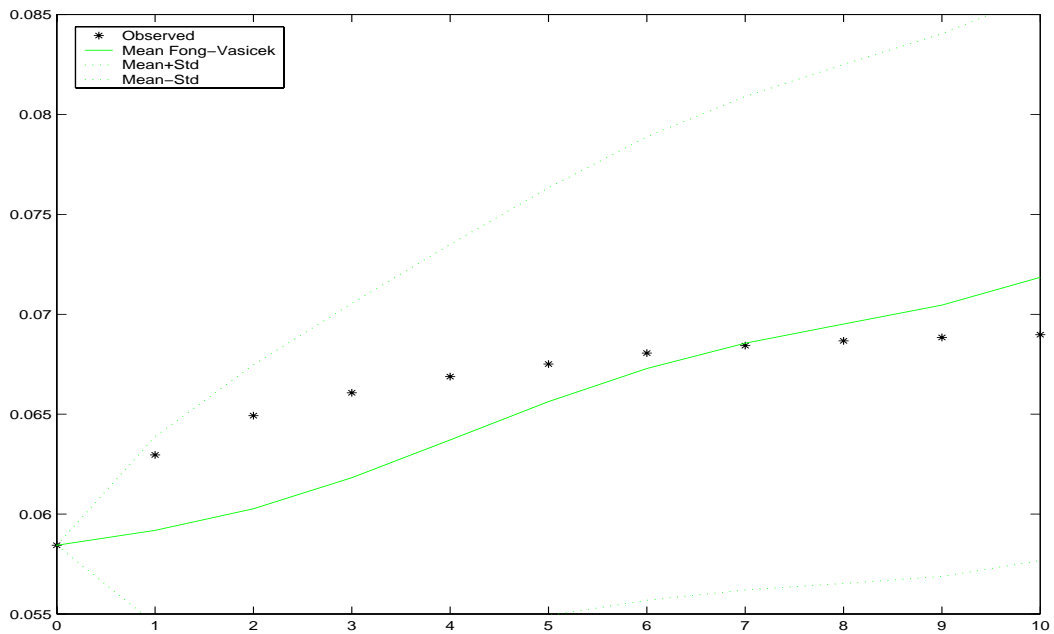


Figure 3. Fitted term structure of two-factor Fong-Vasicek model (time series estimation).

We see that the RMSE bounds get bigger and bigger as the maturity period increases. As mentioned earlier, this could be due to the non-usage of all the information about the parameters of the model hidden in cross sectional data. Also, extending the one-factor Vasicek (1977) model to two-factor Fong and Vasicek (1991) model by adding a stochastic volatility component does not improve the fit by much.

This leads to the approach of estimating all parameters of the financial model (including the market price of risk) simultaneously using cross-sectional data as well. Also, at the same time, we estimate unobserved factors such as stochastic volatility component. Idea of joint estimation of cross sectional dynamics of financial models is in line with earlier works such as de Jong (2000), Bams and Schotman (1998), Munnik and Schotman (1994).

### 3 Cross sectional properties of specific models.

#### 3.1 General Setup.

We consider the following (one-factor) setup. The short rate model is of the form

$$dr_t = \mu(r_t, \theta)dt + \sigma(r_t, \theta)dW_t, \quad (2)$$

where  $r_t$  is the short rate,  $\theta$  is a set of time series parameters and  $\mu, \sigma$  are the infinitesimal drift and variance functions, respectively. Contrary to the approach used in section 2, here we treat  $r_t$  as an unobserved stochastic process and it is subject to estimation. Under standard assumptions (see e.g. Rebonatto (1996), Cochran (2001) ) the Girsanov theorem applies, and thereby, one can obtain the risk neutral measure under which the Wiener process  $W_t$  is transformed as

$$d\widetilde{W}_t = dW_t + \lambda_t dt. \quad (3)$$

In what follows the random process  $\lambda_t$  will be restricted to some specific form depending on the model being considered. In all the cases, however,  $\lambda_t$  involves only one new real valued parameter  $\lambda$ , interpreted as the market price of risk. Thus, under the risk neutral measure, we obtain the dynamics

$$dr_t = \nu(r_t, \theta, \lambda)dt + \sigma(r_t, \theta)d\widetilde{W}_t.$$

This, in turn, leads to the yield formula

$$Y(\tau, t) = Y(\tau, r_t, \theta, \lambda).$$

As mentioned earlier in the introduction, to accommodate factors like market imperfections, human error we use the following as the observational model:

$$Y^*(\tau, t) = Y(\tau, r_t, \theta, \lambda) + \varepsilon_t(\tau), \quad (4)$$

where the (multivariate) observation error  $\varepsilon_t(\tau)$  are i.i.d. sequence of random vectors with unknown variance-covariance matrix. We shall apply the Kitagawa like algorithm on the observation equation (4) with the unobserved component model to be given by (2) and thus resulting in estimators for  $\theta, \lambda$  and filtered  $r_t$ .

### 3.2 Specific models.

First we consider the CIR (1985) model :

$$dr_t = \kappa(\mu - r_t)dt + \sigma\sqrt{r_t}dW_t.$$

Under the risk neutral measure (taking  $\lambda_t = \frac{\lambda}{\sigma}\sqrt{r_t}$  in (3)) the model becomes

$$dr_t = \kappa(\mu - r_t)dt - \lambda r_t dt + \sigma\sqrt{r_t}d\widetilde{W}_t,$$

resulting in the bond pricing formula of “*affine*” form

$$Y(\tau, t) = A(\tau, t) + B(\tau, t)r_t,$$

where

$$\begin{aligned} B(\tau, t) &= \frac{2(e^{\gamma(\tau-t)} - 1)}{(\gamma + \kappa + \lambda)(e^{\gamma(\tau-t)} - 1) + 2\gamma}, \\ A(\tau, t) &= \frac{2\kappa\mu}{\sigma^2} \ln\left(\frac{2\gamma e^{(\gamma+\kappa+\lambda)(\tau-t)/2}}{(\gamma + \kappa + \lambda)(e^{\gamma(\tau-t)} - 1) + 2\gamma}\right), \\ \gamma &= ((\kappa + \lambda)^2 + 2\sigma^2)^{1/2}. \end{aligned}$$

Therefore, we have 4 unknown parameters to estimate.

For Vasicek (1977) model we have

$$dr_t = \kappa(\mu - r_t)dt + \sigma dW_t$$

as real dynamics and under the risk neutral measure (by taking  $\lambda_t = \frac{\lambda}{\sigma}$  in (3)) the model becomes

$$dr_t = \kappa(\mu - r_t)dt - \lambda dt + \sigma d\widetilde{W}_t,$$

resulting in an affine bond pricing formula, as in CIR, with

$$\begin{aligned} B(\tau, t) &= \frac{1 - \exp(-\kappa\tau)}{\kappa}, \\ A(\tau, t) &= \theta(\tau - B(\tau, t)) + \frac{\sigma}{4\kappa}B(\tau, t)^2, \\ \theta &= \mu - \frac{\sigma}{2\kappa^2} - \frac{\lambda}{\kappa}, \end{aligned}$$



where  $\theta$  is the yield on infinite maturity bond.

For the Longstaff (1989) model the short rate equation takes the form

$$dr_t = \kappa(\mu - \sqrt{r_t})dt + \sigma\sqrt{r_t}dW_t,$$

where  $\mu = \frac{\sigma^2}{4\kappa}$ , and under the risk neutral measure (with  $\lambda_t = \frac{2\lambda}{\sigma}\sqrt{r_t}$  in (3)) it satisfies

$$dr_t = \kappa(\mu - \sqrt{r_t})dt - 2\lambda r_t dt + \sigma\sqrt{r_t}d\tilde{W}_t.$$

Now, however, we do not have an affine form of the yield function. The pricing formula can be called “*semiaffine*” in the sense that

$$Y(\tau, t) = A(\tau, t) + B(\tau, t)r_t + C(\tau, t)\sqrt{r_t},$$

$$A(\tau, t) = \left( \frac{1 - c_0}{1 - c_0 e^{\gamma(\tau-t)}} \right)^{1/2} \exp \left( c_1 + c_2 \tau + \frac{c_3 + c_4 e^{\gamma(\tau-t)/2}}{1 - c_0 e^{\gamma(\tau-t)}} \right),$$

$$B(\tau, t) = \frac{2\lambda - \gamma}{\sigma^2} + \frac{2\gamma}{\sigma^2(1 - c_0 e^{\gamma(\tau-t)})},$$

$$C(\tau, t) = \frac{2\kappa(2\lambda + \gamma)(1 + e^{\gamma(\tau-t)/2})^2}{\gamma\sigma^2(1 - c_0 e^{\gamma(\tau-t)})},$$

$$\gamma = (4\lambda^2 + 2\sigma^2)^{1/2},$$

$$c_0 = (2\lambda + \gamma)/(2\lambda - \gamma),$$

$$c_1 = -\frac{\kappa^2}{\gamma^3\sigma^2}(4\lambda + \gamma)(2\lambda - \gamma),$$

$$c_2 = (2\lambda + \gamma)/4 - \kappa^2/\gamma^2,$$

$$c_3 = \frac{4\kappa^2}{\gamma^3\sigma^2}(2\lambda^2 - \sigma^2),$$

$$c_4 = -\frac{8\lambda\kappa^2}{\gamma^3\sigma^2}(2\lambda + \gamma).$$

As we can see, in this case, we have 3 parameters to estimate.

## 4 Kitagawa maximum-likelihood approach and unobserved component estimation.

A state-space model with one unobserved factor is given (suppressing the notation for parameters) by

$$\begin{aligned} y_t &= \tilde{Y}(r_t) + \varepsilon_t, & t = 0, 1, \dots \\ r_{t+1}|r_t &\sim p_{t+1|t}(r_{t+1}|r_t), \end{aligned}$$

where  $\{\varepsilon_t\}$  is an independent sequence of random vectors with unknown variance-covariance matrix,  $\tilde{Y}(r_t)$  is a (known) function (in our case yield), and  $p_{t+1|t}(r_{t+1}|r_t)$  is a (known) conditional density of the unobserved variable (short rate, in our case). Exploiting the Markovian property of  $\{r_t\}$  and denoting the observations  $(y_1, y_2, \dots, y_n)$  by  $Y_n$ , one has the following recursive filtering scheme **(A)** :

One-step-ahead prediction density :

$$f_{n|n-1}(r_n|Y_{n-1}) = \int_{-\infty}^{\infty} p_{n|n-1}(r_n|r_{n-1})f_{n-1}(r_{n-1}|Y_{n-1})dr_{n-1}.$$

$$\text{Filtering density : } f_n(r_n|Y_n) = \frac{p_{y|r}(y_n|r_n)f_{n|n-1}(r_n|Y_{n-1})}{p(y_n|Y_{n-1})},$$

$$\text{where } p(y_n|Y_{n-1}) = \int_{-\infty}^{\infty} p_{y|r}(y_n|r_n)f_{n|n-1}(r_n|Y_{n-1})dr_n.$$

$$\text{Estimate : } \hat{r}_n = \int_{-\infty}^{\infty} r_n f_n(r_n|Y_n)dr_n.$$

Smoothing density :

$$f_{n|N}(r_n|Y_N) = f_n(r_n|Y_n) \int_{-\infty}^{\infty} \frac{f_{n+1|N}(r_{n+1}|Y_N)p_{n|n-1}(r_{n+1}|r_n)}{f_{n+1|n}(r_{n+1}|Y_n)} dr_{n+1}.$$

$$\text{Smoothed estimate : } \hat{r}_n^{(S)} = \int_{-\infty}^{\infty} r_n f_{n|N}(r_n|Y_N)dr_n.$$

Set of recurrent formulas **(A)** allow us to calculate the estimator of unobservable component i.e.  $\hat{r}_n$ . In general, it would not be possible to evaluate these integrals analytically. However, one can approximate them by the standard Riemann sums of appropriate dimensions. For example, with node points at  $z_0, z_1, \dots, z_L$  approximation looks like

$$\int f(r)dr \approx \sum_{i=1}^L f(z_{i-1})\Delta Z_i, \quad \text{where } \Delta Z_i = (z_i - z_{i-1}). \quad (5)$$

In this case, though, one needs to prove that the estimator obtained by using numerical approximation indeed converges to  $\hat{r}_n$  as the partition becomes finer and finer. By induction it follows that

$$\begin{aligned}
f_{n|n-1}(r_n|Y_{n-1}) &= \frac{\int \cdots \int \prod_{i=1}^{n-1} p_{y|r}(y_i|r_i) \prod_{i=1}^n p_{i|i-1}(r_i|r_{i-1}) f_0(r_0) dr_0 \cdots dr_{n-1}}{\int \cdots \int \prod_{i=1}^{n-1} p_{y|r}(y_i|r_i) \prod_{i=1}^{n-1} p_{i|i-1}(r_i|r_{i-1}) f_0(r_0) dr_0 \cdots dr_{n-1}}, \\
f_n(r_n|Y_n) &= \frac{\int \cdots \int \prod_{i=1}^n p_{y|r}(y_i|r_i) \prod_{i=1}^n p_{i|i-1}(r_i|r_{i-1}) f_0(r_0) dr_0 \cdots dr_{n-1}}{\int \cdots \int \prod_{i=1}^n p_{y|r}(y_i|r_i) \prod_{i=1}^n p_{i|i-1}(r_i|r_{i-1}) f_0(r_0) dr_0 \cdots dr_n}, \\
\text{and } \hat{r}_n &= \frac{\int \cdots \int r_n \prod_{i=1}^n p_{y|r}(y_i|r_i) \prod_{i=1}^n p_{i|i-1}(r_i|r_{i-1}) f_0(r_0) dr_0 \cdots dr_n}{\int \cdots \int \prod_{i=1}^n p_{y|r}(y_i|r_i) \prod_{i=1}^n p_{i|i-1}(r_i|r_{i-1}) f_0(r_0) dr_0 \cdots dr_n}.
\end{aligned}$$

Using the Riemann sums of appropriate dimensions we obtain the following approximations to the above three quantities.

$$\begin{aligned}
&\tilde{f}_{n|n-1}(z_{i_{n-1}}|Y_{n-1}) \\
&\quad \frac{\sum_{i_{n-1}} \cdots \sum_{i_0} \prod_{s=1}^{n-1} p_{y|r}(y_s|z_{i_s-1}) \prod_{s=1}^n p_{s|s-1}(z_{i_s-1}|z_{i_{s-1}-1})}{\sum_{i_{n-1}} \cdots \sum_{i_0} \prod_{s=1}^{n-1} p_{y|r}(y_s|z_{i_s-1}) \prod_{s=1}^{n-1} p_{s|s-1}(z_{i_s-1}|z_{i_{s-1}-1})} \\
&= \frac{\times f_0(z_{i_0-1}) \Delta_0 \cdots \Delta_{n-1}}{\times f_0(z_{i_0-1}) \Delta_0 \cdots \Delta_{n-1}},
\end{aligned}$$

$$\begin{aligned}
&\tilde{f}_n(z_{i_{n-1}}|Y_n) \\
&\quad \frac{\sum_{i_{n-1}} \cdots \sum_{i_0} \prod_{s=1}^n p_{y|r}(y_s|z_{i_s-1}) \prod_{s=1}^n p_{s|s-1}(z_{i_s-1}|z_{i_{s-1}-1})}{\sum_{i_n} \cdots \sum_{i_0} \prod_{s=1}^n p_{y|r}(y_s|z_{i_s-1}) \prod_{s=1}^n p_{s|s-1}(z_{i_s-1}|z_{i_{s-1}-1})} \\
&= \frac{\times f_0(z_{i_0-1}) \Delta_0 \cdots \Delta_{n-1}}{\times f_0(z_{i_0-1}) \Delta_0 \cdots \Delta_n},
\end{aligned}$$

$$\begin{aligned}
&\text{and } \tilde{r}_n \\
&\quad \frac{\sum_{i_n} \cdots \sum_{i_0} z_{i_{n-1}} \prod_{s=1}^n p_{y|r}(y_s|z_{i_s-1}) \prod_{s=1}^n p_{s|s-1}(z_{i_s-1}|z_{i_{s-1}-1})}{\sum_{i_n} \cdots \sum_{i_0} \prod_{s=1}^n p_{y|r}(y_s|z_{i_s-1}) \prod_{s=1}^n p_{s|s-1}(z_{i_s-1}|z_{i_{s-1}-1})} \\
&= \frac{\times f_0(z_{i_0-1}) \Delta_0 \cdots \Delta_n}{\times f_0(z_{i_0-1}) \Delta_0 \cdots \Delta_n},
\end{aligned}$$

where  $\Delta_k = z_{i_k} - z_{i_{k-1}}$ . Note that these latter equations can also be arrived at following the recursive scheme **(A)** when at each recursive step the integrals are replaced by its one dimensional Riemann sum approximation.

Clearly,  $\tilde{r}_n$  converges to  $\hat{r}_n$  as the partition  $z_0, z_1, \dots, z_L$  becomes finer provided that all integrands are Riemann integrable. However, it is known that the convergence of the Riemann sums of type (5) is slow. On the other hand, if one considers

the following Riemann sum

$$\sum_{i=1}^L [f(z_{i-1}) + f(z_i)] \Delta Z_i / 2,$$

then the convergence to  $\int f(r)dr$  is faster<sup>3</sup>. This type of approximation is precisely what appears in Kitagawa (1987) method, and also used by us.

Let us again consider the recursive scheme **(A)**. Each density is approximated by piecewise linear functions, that is, it is specified by the number of segments, location of nodes and the value at each node. It is assumed that all the densities are supported on finite interval<sup>4</sup>. In the simplest case the nodes for all the densities are assumed same,  $z_0, z_1, \dots, z_L$ , say. Then the integration in the *one-step-ahead prediction* equation is evaluated as follows.

$$\begin{aligned} & \int_{-\infty}^{\infty} p_{n|n-1}(r_n|r_{n-1})f_{n-1}(r_{n-1}|Y_{n-1})dr_{n-1} \\ &= \int_{z_0}^{z_L} p_{n|n-1}(r_n|r_{n-1})f_{n-1}(r_{n-1}|Y_{n-1})dr_{n-1} \\ &= \sum_{i=1}^L \int_{z_{i-1}}^{z_i} p_{n|n-1}(r_n|r_{n-1})f_{n-1}(r_{n-1}|Y_{n-1})dr_{n-1}, \end{aligned}$$

where using the linearity of the functions in the interval  $(z_{i-1}, z_i)$ ,

$$\begin{aligned} & \int_{z_{i-1}}^{z_i} p_{n|n-1}(r_n|r_{n-1})f_{n-1}(r_{n-1}|Y_{n-1})dr_{n-1} \\ & \approx \left( p_{n|n-1}(r_n|z_{i-1})f_{n-1}(z_{i-1}|Y_{n-1}) + p_{n|n-1}(r_n|z_i)f_{n-1}(z_i|Y_{n-1}) \right) \\ & \quad \times \frac{(z_i - z_{i-1})}{2}. \end{aligned}$$

Note that, the quantity  $p(y_n|Y_{n-1})$  in the *filtering* equation can be evaluated as

$$\int_{-\infty}^{\infty} p_{y|r}(y_n|r_n)f_{n|n-1}(r_n|Y_{n-1})dr_n$$

and the integration is calculated as above. The integration in the *smoothing* equation is also evaluated similarly.

Note also that the recursion **(A)** allows us to calculate the likelihood function as

$$p(Y_n) = p(y_{n-1}|Y_{n-2})p(y_{n-2}|Y_{n-3}) \dots p(y_2|y_1)p(y_1),$$

<sup>3</sup>See, e. g. formulas (4.1.3) and (4.1.7) in Press (1992) for the rate of convergence.

<sup>4</sup>In case of infinite support, the end points of the grid are to be chosen in such a way that they cover the essential domain of the density.

where the quantities on the right hand side are already calculated during the filtering stage of the recursion. We will use this to obtain the Quasi-Maximum Likelihood (QML) estimate of the parameters.

Another interesting observation can be made here that we can rewrite the expression for the estimator of the unobserved component by

$$\hat{r}_n = \frac{E_{r_0, \dots, r_n} T_n \prod_{i=1}^n p_{y|r}(y_i | r_i)}{E_{r_0, \dots, r_n} \prod_{i=1}^n p_{y|r}(y_i | r_i)},$$

where the expectations in the numerator and the denominator are taken with respect to the trajectory of the state space process. Going by this probabilistic interpretation of  $\hat{r}_n$ , one can evaluate it by stochastic method (simulation) as well. We choose to evaluate it by the deterministic (grid) method because of its advantage in that it is computationally less intensive, especially so when the dimension of the unobserved vector is small. On the other hand, there are works on comparative complexity of the Monte-Carlo methods (see, e.g., Curtiss (1956), Bahvalov (1964), Danilov, Ermakov and Halton (2000)), that indicate significant superiority of the simulation based techniques over deterministic ones when the dimension of the problem is large.

#### 4.1 Implementation for specific models.

Successful application of the Kitagawa algorithm relies on the knowledge of two kind of distributions:

- Distribution,  $p_{y|r}(y_n | r_n)$ , of the observed process given the state process at current moment of time.
- Transition distribution,  $p_{n|n-1}(r_{n+1} | r_n)$ , of the state at moment  $n$  given the state at moment  $n - 1$ .

In our analysis we shall use a set of 4 bonds to estimate the model (see section 5.1). So  $\varepsilon_t(\tau)$  will be 4 dimensional normal distribution with zero mean and unknown variance covariance matrix  $C$  that defines  $p_{y|r}(y_n | r_n)$ . The unknown matrix  $C$  contains (due to symmetry)  $4(4+1)/2 = 10$  different elements, that is, in the case of CIR (1985) and Vasicek (1977) the number of parameters to be estimated is  $10 + 3 + 1 = 14$ , and that in case of Longstaff (1989) :  $10 + 2 + 1 = 13$ . In fact, we shall decompose  $C$  as

$$C = \Lambda^{1/2} R \Lambda^{1/2},$$

where  $\Lambda$  is the diagonal matrix consisting of the square roots of the diagonal elements of  $C$  i.e. variances and  $R$  is a correlation matrix. That will allow for more interpretable estimation results.

Situation with  $p_{n|n-1}(r_{n+1}|r_n)$  is a bit different. For Vasicek (1977) model it is given by a normal density. For the other two, however, the expressions are more complicated. In the case of CIR (1985) the transitional density at time  $t+h$  conditional on time  $t$  is defined as

$$p(r_{t+h}|r_t) = ce^{-u-v} \left(\frac{v}{u}\right)^{q/2} I_q(2(uv)^{1/2}),$$

$$c = \frac{2\kappa}{\sigma^2(1 - e^{-\kappa h})}, u = cr_{t+h}e^{-\kappa h}, v = cr_t, q = \frac{2\kappa\theta}{\sigma^2} - 1,$$

where  $I_q(\cdot)$  is the modified Bessel function of the first kind of order  $q$ .

In the case of Longstaff (1989) model corresponding density takes form

$$p(r_{t+h}|r_t) = \frac{1}{\sqrt{2\pi\sigma^2h}} \left[ e^{-2(\sqrt{r_{t+h}} - \sqrt{r_t + \kappa h/2})^2 / (\sigma^2 h)} + e^{4\kappa\sqrt{r_t}/\sigma^2} e^{-2(\sqrt{r_{t+h}} + \sqrt{r_t + \kappa h/2})^2 / (\sigma^2 h)} \right]$$

$$+ \frac{2}{\sigma^2\sqrt{r_t}} e^{-4\kappa\sqrt{r_t}/\sigma^2} \left( 1 - \phi \left( \frac{2(\sqrt{r_{t+h}} + \sqrt{r_t + \kappa h/2})^2}{\sqrt{\sigma^2 h}} \right) \right).$$

Both expressions are somewhat involved but do not deliver any difficulty to program.

In numerical implementation of the algorithm the simple uniform grid was used. The number of grid points was equal to 500. Lower bound of grid was set to be 0. Upper bound for the grid points was set to 50%, that is far above any historically observed rate for US.

## 5 Empirical results.

### 5.1 Data Description.

The data description and yields correlations are gathered in the Tables 2 and 3. We observe that yields with different maturities are highly correlated. Therefore, it is not necessary to use all available maturities for estimation. Also, in order to keep the number of parameters in observation equation reasonably small we should restrict ourselves not to use too many maturities. In our analysis, as in de Jong (2000), we use bonds with 4 maturities: 3 month (almost the short rate), 1 year (short yield), 5 year (middle yield) and 10 year (long yield). This restricts the number of parameters to 14 in CIR (1985) and Vasicek (1977) and to 13 in Longstaff (1989) case, as noted earlier in section 4.1.

Table 2: Summary of Data

Maturity	3mth	1yr	2yr	3yr	4yr	5yr
mean	5.84	6.30	6.49	6.61	6.69	6.75
st. dev.	3.10	3.11	3.05	3.01	2.99	2.98
maximum	16.00	16.34	16.15	15.83	15.85	15.70
minimum	0.62	0.85	1.15	1.41	1.60	1.77
first obs.	0.93	1.19	1.53	1.78	1.95	2.09
last obs.	6.18	6.43	6.95	7.19	7.43	7.63
first autocorr.	.9819	.9837	.9866	.9887	.9902	.9914

Maturity	6yr	7yr	8yr	9yr	10yr
mean	6.81	6.84	6.87	6.88	6.90
st. dev.	2.98	2.97	2.96	2.95	2.94
maximum	15.47	15.28	15.17	15.10	15.07
minimum	1.93	2.07	2.19	2.28	2.34
first obs.	2.20	2.29	2.37	2.43	2.49
last obs.	7.77	7.86	7.94	8.00	8.07
first autocorr.	.9924	.9932	.9937	.9940	.9942

Maturity	3mth	1yr	2yr	3yr	4yr	5yr	6yr	7yr	8yr	9yr	10yr
3mth	1										
1yr	0.9897	1									
2yr	0.9744	0.9939	1								
3yr	0.9617	0.9855	0.9979	1							
4yr	0.9522	0.9783	0.9944	0.9990	1						
5yr	0.9440	0.9716	0.9906	0.9971	0.9994	1					
6yr	0.9365	0.9654	0.9867	0.9946	0.9980	0.9995	1				
7yr	0.9308	0.9605	0.9834	0.9924	0.9965	0.9987	0.9998	1			
8yr	0.9269	0.9570	0.9809	0.9905	0.9951	0.9978	0.9993	0.9999	1		
9yr	0.9240	0.9542	0.9789	0.9889	0.9939	0.9968	0.9986	0.9995	0.9999	1	
10yr	0.9216	0.9518	0.9770	0.9874	0.9926	0.9959	0.9979	0.9990	0.9996	0.9999	1

Table 3. Empirical yields correlations.



## 5.2 Estimation.

In order to make the parameters of 3 different models to be comparable we write them all in a standardized form

$$dr_t = -\kappa(r^s - \mu^s)dt + \sqrt{\sigma_0 + \sigma_1 r}dW_t,$$

where  $\sigma_1 = 0$  for Vasicek model,  $\sigma_0 = 0$  for CIR and Longstaff models,  $s = 1/2$  for Logstaff and  $s = 1$  otherwise. Parameter estimation for short rate equations are gathered in the following Table 4. For each model and each parameter of the model the corresponding cells contain point estimate, its standard error and t-statistic.<sup>5</sup> From the likelihood values we conclude that among the three models considered, the CIR provides the best fit and Vasicek is the worst. As we can see, all estimated parameters are significant. Autoregression parameter  $\kappa$  is small for all models, indicating significant persistence in all of them, i.e., high first order autocorrelations<sup>6</sup>. This, in turn, implies flat term structure of these models, that is in good agreement with graphical analysis (see next section).

Model Likelihood	$\kappa$	$\mu$	$\sigma_0$	$\sigma_1$	$\lambda$
L89 3.946e+03	[0.067]	5.060e-02 2.703e-02 1.872e+00	[0]	3.629e-03 1.304e-03 2.782e+00	-1.588e-02 2.311e-03 -6.871e+00
CIR85 4.031e+03	4.258e-02 2.016e-03 2.113e+01	6.277e-02 1.272e-03 4.934e+01	[0]	2.172e-03 2.035e-04 1.068e+01	-3.153e-02 6.900e-04 -4.570e+01
V77 3.350e+03	0.105	7.3e-02	7.45e-04	[0]	3.51e-03

Table 4. Parameter estimations.

We can see that both the stochastic volatility model show significant and negative market price of risk. Vasicek model in contrast estimate  $\lambda$  to be positive. The mean reverting parameter is in reasonable range from 5% for Longstaff model up to approximately 7% for Vasicek. The value of volatility parameter  $\sigma_1$  in Longstaff case is approximately 50% smaller than corresponding value in CIR. The variability factor  $\sigma_1\mu$  is about  $1.8 \cdot 10^{-4}$  for Longstaff and  $1.4 \cdot 10^{-4}$  for CIR which are similar. However contrast them with  $\sigma_0 = 7.5 \cdot 10^{-4}$  for the model with constant volatility.

<sup>5</sup>For Vasicek model standard errors are currently not available.

<sup>6</sup>For Longstaff case  $\kappa$  was calculated according to  $\kappa = \sigma_1/(4\sqrt{\mu})$ .

### 5.3 Quality of Fit.

In order to evaluate the quality of fit for the models under consideration we plot implied term structures together with empirical term structures and residual variances. As we already have seen in Table 4, the CIR model provides best fit. At Fig 4–6 we can see that the width of the error bounds is the smallest for CIR model. However the Longstaff model provide fit that is very similar to CIR notwithstanding that it has only 3 parameters in the state equation. Not surprisingly the Vasicek model reveals inferior fit, especially for the longest maturities.

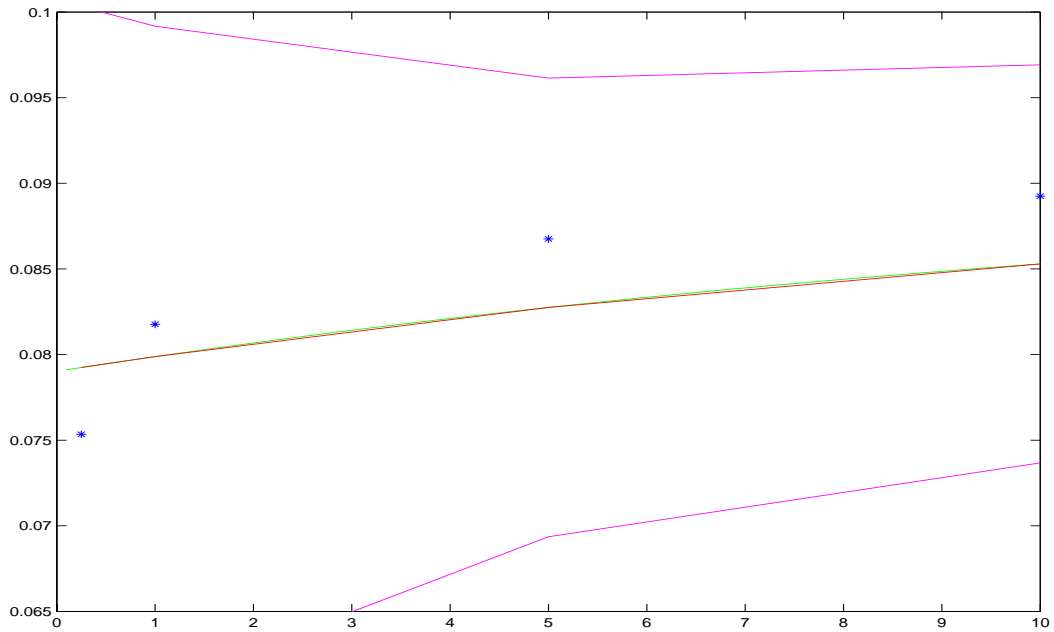


Figure 4. Fitted term structure of one factor Vasicek model.

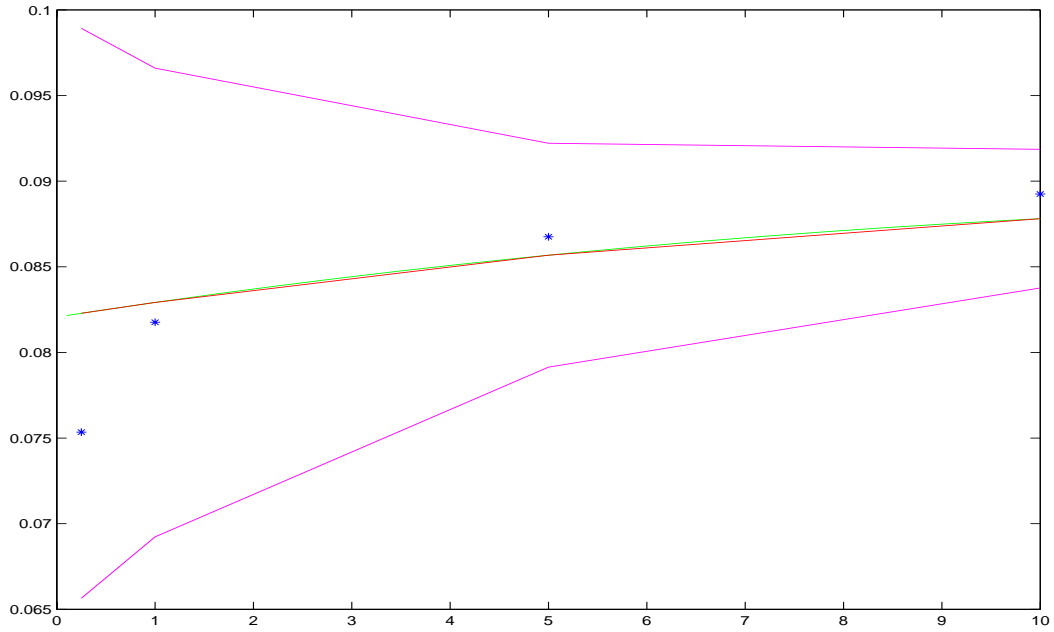


Figure 5. Fitted term structure of one factor CIR model.

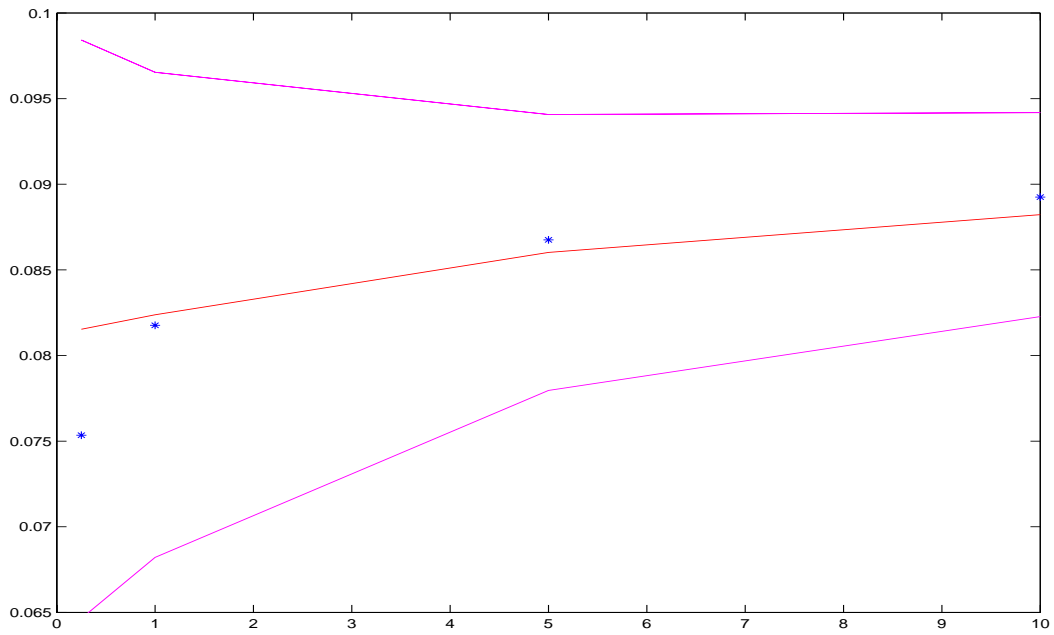


Figure 6. Fitted term structure of Longstaff model.

One of the important outputs of the Kitagawa method is the set of estimates for the unobserved component in the model. In the short rate models under consideration this component is  $r_t$ . The estimates for  $r_t$  in the best performing CIR model are plotted at figure 7, together with 3 month and 10 year yields (the shortest and the longest yields). We can see, that the estimated factor moves similarly to the long yield, this is confirmed also from the correlation coefficient between  $r_t$  and these yields, that reach 0.98 for the long yield against only 0.78 for the short yield. For the other two models as well the behavior of  $r_t$  is similar. Even though this is contradictory to the intuition that  $r_t$ , being the instantaneous rate, would be more similar to the short yield rather than the long yield, this finding is conforming to other earlier works. For example, for CIR, de Jong (2000) obtained similar term structure figure as in our figure 5. When he considered two factor models, the first factor was closely related to the long yield. This only shows the misspecification of the model and reaffirms the need to consider factor models with several factors.

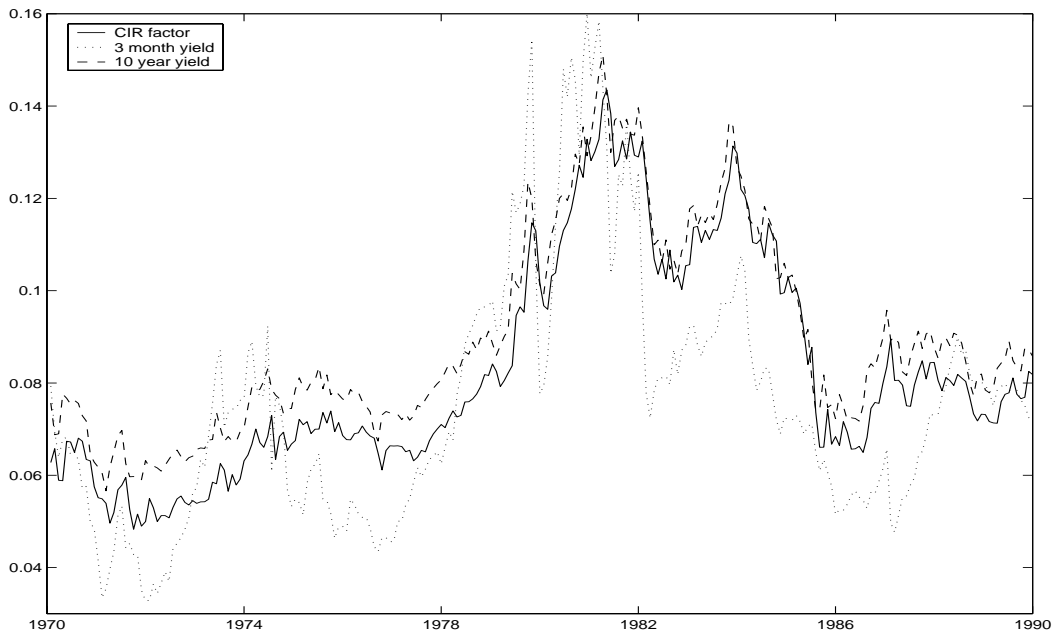


Figure 7. Estimated unobserved component in CIR model.

## 6 Conclusion

We develop a method for efficient estimation of the term structure model. The method is based on application of Kitagawa algorithm to nonlinear filtering setup. Performance of the method was checked on a number of one factor models. We

found that estimation techniques using unobservable component approach seems to fit better the long yield part of the empirical yield curve. Estimation techniques that treat short rate as observable random process and estimate time series parameters of this process directly, in contrast, fit better short maturities. Performance of the former at the long maturities is similar to performance of latest on the short maturities. Apparently the model misspecification affect differently those procedures. Therefore, we can suggest for the future research that the two-factor model, with an extra (volatility) factor like Fong and Vasicek (1991) or Longstaff and Schwartz (1992) is necessary. On the other hand, use of cross sectional data to estimate the volatility factor is necessary in order to obtain good fit for longer maturities. Further, a comparison between the Kalman filter method (used by de Jong (2000) and the Kitagawa method to estimate the term structures for CIR (1985) model does not show much difference. Possibly, for this particular model (CIR) use of Kalman filter is as good as the use of theoretically better Kitagawa-like method. Nevertheless, use of Kitagawa method removes the inconsistency problem (see de Jong (2000) that comes with the use of Kalman filter.

## References

- [1] Babbs, S., and Nowman, K. B. (1999), Kalman Filtering of Generalized Vasicek Term Structure Models. *Journal of Financial and Quantitative Analysis*, **34**, No. 1, pp. 115 – 130.
- [2] Bams, D., and Schotman, P. (1998), Risk Premia in the Term Structure of Interest Rates: A Panel Data Approach, Working Paper, University of Maastricht.
- [3] Bahvalov, N.S., (1964), On the optimal estimation of the speed of convergency of the quadrature processes and Monte-Carlo integration methods on the classes of functions, In “*Numerical methods for the solution of differential and integral equations and quadrature formulas*”, M., Nauka, pp. 5-63, (in Russian).
- [4] Chen, R., and Scott, L. (1993), ML Estimation for a Multifactor Equilibrium Model of the Term Structure. *Journal of Fixed Income*, **December**, pp. 14 – 31.
- [5] Cox, J.C., Ingersoll, J.E., and Ross, S.A. (1985), A theory of the term structure of interest rates, *Econometrica*, **53**, pp. 385 – 407.
- [6] Cochran, J.H. (2001), *Asset Pricing*, Princeton, N.J. : Princeton University Press.

- [7] Curtiss, J.H., (1956), A Theoretical comparison of the efficiencies of two classical methods and a Monte Carlo Method for computing one component of the solution of a set of linear algebraic equations, In: Meyer, H.A. (Ed.), *Symp. on Monte Carlo Methods*. Wiley, New York, pp. 191-233.
- [8] Danilov, D., and Drost, F. (2000), Term Structure Models with Stochastic Volatility: Risk Premia Specifications, Exact Pricing Formulas and Empirical Evaluation of Fit, manuscript, Tilburg University.
- [9] Danilov D.L., Ermakov S.M., and Halton J.H., (2000), Asymptotic complexity of the Monte-Carlo methods for solving linear systems, *Journal of Statistical Planning and Inference*, Vol. 85 (1-2) pp. 5-18.
- [10] Danilov, D., and Mandal, P. K. (2000), Estimation of the Volatility Component in Two-factor Stochastic Volatility Short Rate Models, *EURANDOM Preprint 2000-040*.
- [11] de Jong, F. (2000), Time Series and Cross-section Information in Affine Term-Structure Models, *Journal of Business and Economic Statistics*, **18**, No. 3, pp. 300 – 314.
- [12] Fong, G., and Vasicek, O. (1991), Fixed-income volatility management, *Journal of portfolio management*, **17**, No. 4, pp. 41 – 46.
- [13] Frachot, A., Lesne, J. P., and Renault, E. (1995), Indirect Inference Estimation of Factor Models of the Yield Curve, *working paper, CREST*.
- [14] Fridman, M., and Harris, L. (1998), A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models, *Journal of Business and Economic Statistics*, **16**, No. 3, pp. 284 – 291.
- [15] Gallant, A.R., and Tauchen, G. (1996), Which Moments to Match, *Econometric Theory* **12**, pp. 657 – 681.
- [16] Geyer, A., and Pichler, S. (1999), A State-space Approach to Estimate and Test Multi-factor Cox-Ingersoll-Ross Models of Term Structure, *Journal of Financial Research*, **XXII**, No. 1, pp. 107 – 130.
- [17] Gouriéroux, C., Monfort, A., and Renault, E. (1992), Indirect Inference, *Journal of Applied Econometrics*, **8**, S85 – S118.
- [18] Hasbrouck, J. (1999), The Dynamics of Discrete Bid and Ask Quotes, *Journal of Finance*, **54**, No. 6, pp. 2109 – 2142.

- [19] Kitagawa, G. (1987), Non-Gaussian State-Space Modeling of Nonstationary Time Series, *Journal of the American Statistical Association*, **82**, pp. 1032 – 1041.
- [20] Longstaff, F. (1989), A Nonlinear Equilibrium Model of the Term Structure of Interest Rates, *Journal of Financial Economics*, **23**, pp. 195 – 224.
- [21] Longstaff, F., and Schwartz, (1992), Interest Rate Volatility and the Term Structure: A Two Factor General Equilibrium Model, *Journal of Finance*, **47**, pp. 1259 – 1282
- [22] Munnik, J., and Schotman, P., (1994) Cross-sectional versus time series estimation of term structure models: Empirical results for the Dutch bond market, *Journal of Banking & Finance*, **18**, pp. 997 – 1025.
- [23] Pagan, A., and Martin, V. (1996), Simulation Based Estimation of Some Factor Models in Econometrics. *working paper, University of Melbourne, Dept of Economics*.
- [24] Pearson, N., and Sun, T. S. (1994), Exploiting the conditional Density in Estimating the Term Structure: An Application to the CIR Model, *Journal of Finance*, **46**, pp. 1279 – 1304.
- [25] Press, W.H. (1992), *Numerical recipes in FORTRAN : the art of scientific computing*, Cambridge, Cambridge University Press.
- [26] R. Rebonato (1996). *Interest-Rate Option Models: Understanding, Analysing and Using Models for Exotic Interest Rate Options*. Chichester, Wiley.
- [27] Schwarz, G. (1978), Estimating the Dimension of the model, *Annals of Statistics*, **6**, pp. 461 – 464.
- [28] Vasicek, O. (1977), An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, pp. 177 – 188.