Linear Gaussian Affine Term Structure Models with Unobservable Factors: Calibration and Yield Forecasting

Paresh Date[*] [1], Chieh Wang[*],[†]

[*] The Centre for the Analysis of Risk and Optimisation Modelling Applications (CARISMA), School of Information Systems, Computing and Mathematics, Brunel University, UB8 3PH, United Kingdom.

[†] Department of Public Finance and Taxation, Meiho Institute of Technology, Taiwan.

Abstract

This paper provides a significant numerical evidence for out-of-sample forecasting ability of linear Gaussian interest rate models with unobservable underlying factors. We calibrate one, two and three factor linear Gaussian models using the Kalman filter on two different bond yield data sets and compare their out-of-sample forecasting performance. One step ahead as well as four step ahead out-of-sample forecasts are analyzed based on the weekly data. When evaluating the one step ahead forecasts, it is shown that a one factor model may be adequate when only the short-dated or only the long-dated yields are considered, but two and three factor models performs significantly better when the entire yield spectrum is considered. Furthermore, the results demonstrate that the predictive ability of multi-factor models remains intact far ahead out-of-sample, with accurate predictions available up to one year after the last calibration for one data set and up to three months after the last calibration for the second, more volatile data set. The experimental

─────────────

[1]corresponding author. Email: paresh.date@brunel.ac.uk. Phone: +44 1895 265613, Fax: +44 1895 269732

data denotes two different periods with different yield volatilities, and the stability of model parameters after calibration in both the cases is deemed to be both significant and practically useful. When it comes to four step ahead predictions, the quality of forecasts deteriorates for all models, as can be expected, but the advantage of using a multi-factor model as compared to a one factor model is still significant.

In addition to the empirical study above, we also suggest a nonlinear filter based on linear programming for improving the term structure matching at a given point in time. This method, when used in place of a Kalman filter update, improves the term structure fit significantly with a minimal added computational overhead. The improvement achieved with the proposed method is illustrated for out-of-sample data for both the data sets. This method can be used to model a parameterized yield curve consistently with the underlying short rate dynamics.

Keywords Finance, forecasting, time series, filtering

1   Introduction

Exponential affine term structure models is one of the oldest and the most widely studied class of dynamic interest rate models. The main advantage of these models is the fact that the yields can be expressed as affine functions of the short rate. The exponential affine term structure models are often classified into three categories:

- Gaussian affine models. The single factor linear model proposed in Vasicek (1977) is a Gaussian affine model and was the first model for which closed-form formulae for bond prices were obtained. All the state variables in these types of models have constant volatilities. A multi-factor Gaussian affine model is discussed in Babbs and Nowman (1999). Extensions of the Gaussian affine models to match the current term structure are discussed in Hull and White (1990), Hull and White (1993) and Hull and White (1994). The Gaussian models have a high degree of tractability and a variety of products can be priced in closed-

form with these types of models. Recently, closed-form formulae for swaption pricing under a multi-factor Gaussian affine model have been reported in Schrager and Pelsser (2006).

- CIR affine models. Models of this type were first proposed in Cox et al. (1985) and were extended to multi-factor case in Beaglehole and Tenny (1991). All the state variables in these models have CIR-type square root volatilities. Unlike the Gaussian models, the interest rate is guaranteed to remain non-negative provided it starts from a non-negative value.

- A three-factor affine family. This family represents the models that mix Gaussian and CIR type state variables; see Balduzzi et al. (1996), Rhee (1999) and Longstaff and Schwartz (1992) for examples.

A general framework for multi-factor affine term structure models was proposed in Duffie and Kan (1996). An empirical comparison of several different short rate models appears in Chan et al. (1992).

In this paper, we model the behavior of government bond yields by using linear Gaussian term structure models. Cross-sectional as well as time-series data of gilt yields is used for calibration, *i.e.* each discrete measurement in time series consists of a cross-section of gilt yields. The short rate is assumed to be an affine function of unobservable state variables. Each yield is assumed to be the sum of the theoretical yield for the corresponding time to maturity and a zero mean stochastic disturbance. Since the theoretical yield is affine in the short rate, this set-up gives a linear state space system. The Kalman filter can then be used to calibrate the model using noisy yield measurements.

There are many accounts of Kalman filtering-based calibration and forecasting of a time series, including detailed treatments in Harvey (1989) and in Durbin and Koopman (2002). A brief overview of Kalman filtering-based calibration for interest rate models appears in (James and Webber (2000), chapter 18). In empirical research reports concerning this issue in detail, Babbs and Nowman (1999) and Lund (1997) estimate the generalized Vasicek models using Kalman filtering. Other filtering applications in

the area of interest rate modelling are reported in Rossi (2004), Jegadeesh and Pennacchi (1996), Ball and Torous (1996), Duan and Simonato (1995) and Gravelle and Morley (2005). The work presented here follows Babbs and Nowman (1999) in that we use linear Gaussian transition equations with constant prices of risk.

In contrast with the previously cited research, the emphasis of this paper is short and medium term out-of-sample forecasting of future yields. While a lot of empirical research has been carried out to predict the term structure and to test the expectations hypothesis (see, e.g. Lanne (2000) and references therein), relatively little empirical work has been done in testing the out-of-sample, short term forecasting ability of an affine term structure model with unobservable factors. We provide a significant numerical evidence of predictive ability of a simple linear Gaussian model with unobservable state variables. A forecasting model may be useful in a variety of situations, *e.*g. in predicting downside risk of future performance of a gilt portfolio or in generating scenarios for a stochastic programming based optimization. It is shown in Reisman and Zohar (2004) that using re-balancing based on a short term prediction of the term structure can significantly improve the returns from a bond portfolio. For multi-factor models, we demonstrate that the predictive ability remains unimpaired for a long period after calibration. Besides this empirical work based on the application of existing theoretical results, a new linear programming-based heuristic is suggested for estimating the unobservable states. This heuristic filter is aimed at improving the yield curve matching without increasing the calibration effort or the model complexity. This provides a very useful alternative to Nelson-Siegel type static yield curve models since it is consistent, by definition, to the assumed interest rate dynamics[2].

The rest of the paper is organized as follows. Section 2 provides a brief overview of linear Gaussian affine term structure model used in this work. Section 3 presents a similarly brief overview of Kalman filtering-based calibration. Section 4 offers a detailed empirical analysis for calibration of a one,

---

[2]The consistency of a parameterized yield curve is defined in section 5.

two and three factor linear Gaussian models and yield prediction using all these models. Section 5 suggests a new linear programming heuristic for state estimation in a linear Gaussian model to improve the matching of the initial term structure. This heuristic is shown to out-perform the Kalman state update in term structure matching at a fairly small added computational cost. Finally, section 6 summarizes the contributions of this work.

2 Linear Gaussian Affine Term Structure Models

In the linear Gaussian model discussed in Babbs and Nowman (1999), the short rate at time t is described by

$$r_t = \mu - \Sigma_{i=1}^n \mathbf{x}_{t,i}, \tag{1}$$

where the states evolve as linear Gaussian processes with constant volatilities:

$$d\mathbf{x}_{t,i} = -\alpha_i \mathbf{x}_{t,i} dt + \Sigma_{j=1}^n \sigma_{i,j} d\mathbf{z}_{t,j}. \tag{2}$$

with $\mathbf{z}_{t,j}$ being independent Wiener processes. Suppose that each state process $\mathbf{x}_{t,i}$ has a constant price of risk, λ_i. Then the term structures at time t are of the form

$$r_t(\tau) = a_0(\tau) + \Sigma_{i=1}^n a_i(\tau)\mathbf{x}_{t,i} \tag{3}$$

with $a_i(\tau) = -H(\alpha_i \tau), a_0(\tau) = r_\infty - w(\tau)$, r_∞ and $w(\tau)$ are functions of constant parameters μ, λ_i, $\sigma_{i,j}$, α_j and $H(x) = (1 - e^{-x})/x$. The formulae of r_∞ and $w(\tau)$ can be expressed as

$$r_\infty = \mu + \Sigma_{i=1}^n \lambda_i \Sigma_{j=1}^n \frac{\sigma_{j,i}}{\alpha_j} - \frac{1}{2}\Sigma_{i=1}^n (\Sigma_{j=1}^n \frac{\sigma_{j,i}}{\alpha_j})^2, \tag{4}$$

$$w(\tau) = \Sigma_{i=1}^n H(\alpha_i \tau)\left(\Sigma_{j=1}^n \lambda_j \frac{\sigma_{i,j}}{\alpha_j} - \Sigma_{j=1}^n \Sigma_{k=1}^n \frac{\sigma_{k,j}\sigma_{i,j}}{\alpha_k \alpha_i}\right)$$
$$+ \frac{1}{2}\Sigma_{i=1}^n \Sigma_{j=1}^n H((\alpha_i + \alpha_j)\tau)\Sigma_{k=1}^n \frac{\sigma_{i,k}\sigma_{j,k}}{\alpha_i \alpha_j}. \tag{5}$$

Note that r_t in (1) is recovered from $r_t(\tau)$ as $\tau \to 0$. This model will be used later for numerical experiments with $n = 1, n = 2$ and $n = 3$ in section 4.

In addition, we use $\mathbf{x}_{t,i}$ as unobservable variables and estimate them from yield measurements using the Kalman filter. The basic set up of calibration of a linear time series model using the Kalman filter is briefly outlined in the next section.

3   Kalman filtering-based calibration

Consider a discrete time, linear state space system

$$\mathbf{x}_k = F\mathbf{x}_{k-1} + \epsilon_k,$$
$$\mathbf{b}_k = A\mathbf{x}_k + B + \mathbf{e}_k, \tag{6}$$

where $\epsilon_k, \mathbf{e}_k$ are zero mean, Gaussian and uncorrelated, the unknown $\mathbf{x}_k$ is the state vector at time k, $\mathbf{b}_k$ is measurement made at time k and A, B, F, $\mathbb{E}(\epsilon_k \epsilon_k^T) = \Sigma_\epsilon$, $\mathbb{E}(\mathbf{e}_k \mathbf{e}_k^T) = \Sigma_\mathbf{e}$ are constants or are known functions of time. Considering a three state model with M yield measurements for simplicity, a first order Euler discretisation of (1)-(2) will lead to

$$F = \begin{bmatrix} 1 - \alpha_1 & 0 & 0 \\ 0 & 1 - \alpha_2 & 0 \\ 0 & 0 & 1 - \alpha_3 \end{bmatrix}, B = \begin{bmatrix} a_0(\tau_1) \\ a_0(\tau_2) \\ \vdots \\ a_0(\tau_M) \end{bmatrix}, \tag{7}$$

$$A = \begin{bmatrix} a_1(\tau_1) & a_2(\tau_1) & a_3(\tau_1) \\ a_1(\tau_2) & a_2(\tau_2) & a_3(\tau_2) \\ \vdots & \vdots & \vdots \\ a_1(\tau_M) & a_2(\tau_M) & a_3(\tau_M) \end{bmatrix} \tag{8}$$

with $a_i(\tau_j)$ defined as in the last section. The time interval between two successive samples is assumed to be unity without loss of generality. Only $\mathbf{b}_{k,i} = r_{t_k}(\tau_i) + \mathbf{e}_{k,i}$, $i = 1, 2, \cdots, M$ is measured at discrete times t_k, where $r_{t_k}(\tau_i) = (A\mathbf{x}_k + B)_i$ is the theoretical yield for time to maturity τ_i and $(\mathbf{z})_i$ is i^{th} entry in vector $\mathbf{z}$. We wish to predict $\mathbf{x}_k$ based on measurement up to time t_{k-1}. When this prediction is carried out using the Kalman filter, the

6

joint density function of yield forecasting errors is available in closed-form and can be maximized to find the parameters A, B, F, Σ_ϵ and Σ_e.

The standard set of recursive equations for Kalman filtering is outlined below for reference.

The prediction of the state vector:

$$\hat{\mathbf{x}}_{k|k-1} = F\hat{\mathbf{x}}_{k-1|k-1}. \tag{9}$$

The prediction of the covariance matrix:

$$P_{k|k-1} = FP_{k-1|k-1}F^T + \Sigma_\epsilon. \tag{10}$$

The Kalman gain matrix:

$$K_k = P_{k|k-1}A^T(AP_{k|k-1}A^T + \Sigma_e)^{-1}. \tag{11}$$

The filtered state vector:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k\mathbf{v}_k \tag{12}$$

The filtered covariance matrix:

$$P_{k|k} = (I - K_kA)P_{k|k-1}. \tag{13}$$

The yield forecasting error:

$$\mathbf{v}_k = \mathbf{b}_k - (B + A\hat{\mathbf{x}}_{k|k-1}). \tag{14}$$

The variance of forecasting error:

$$\Sigma_k = AP_{k|k-1}A^T + \Sigma_{\mathbf{e}}. \tag{15}$$

The initial state $\mathbf{x}_0$ and the initial covariance matrix P_0 are parameterized in terms of A, B, F and Σ_ϵ. Let θ be the vector of unknown parameters from matrices A, B, F, $\Sigma_{\mathbf{e}}$ and Σ_ϵ. As mentioned earlier, the joint probability density function (also called the likelihood function) of observations is maximized over the parameter vector θ to get the estimate of θ. Since the forecast errors are Gaussian, the log likelihood function is expressed by:

$$L(\theta) = \sum_{k=1}^{T} \log p(\mathbf{b}_k|\mathcal{F}_{k-1}, \theta), \tag{16}$$

where T is number of samples. Since the forecast error is Gaussian, this reduces to minimizing

$$-L(\mathbf{b}_k, \theta) = \frac{MT}{2} \log 2\pi + \frac{1}{2} \sum_{k=1}^{T} \left(\log \det(\Sigma_k) + \mathbf{v}_k^T \Sigma_k^{-1} \mathbf{v}_k \right). \qquad (17)$$

This smooth nonlinear cost function can be minimized over the set of parameters using any standard nonlinear solver. We use MATLAB's "off-the-shelf" optimizer *fminsearch* which seemed to perform satisfactorily. Numerical difficulties may arise when the parameterized Σ_ϵ and $\Sigma_\mathbf{e}$ are not positive definite. We will avoid this issue by restricting the matrix $\Sigma_\mathbf{e}$ to be diagonal with positive entries:

$$\Sigma_e = \begin{bmatrix} h_1^2 & 0 & \dots & 0 \\ 0 & h_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & h_M^2 \end{bmatrix}$$

for vector-valued measurements $\mathbf{b}_k$ at each time t_k, and parameterizing Σ_ϵ for the 2-state case as

$$\Sigma_\epsilon = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \cos \phi \\ \sigma_1 \sigma_2 \cos \phi & \sigma_2^2 \end{bmatrix}$$

which requires no further constraint on ϕ for ensuring positive semi-definiteness. Irrespective of correlation between state noise variables, note that the yields themselves are necessarily correlated through the three state variables.

In the next section, we apply Kalman filtering-based calibration to linear Gaussian term structure models outlined in section 2.

4   Empirical Results

The aim of the empirical study presented here is to calibrate linear Gaussian term structure models using the Kalman filter for two different data sets (UK gilt yields and US treasury yields) and to examine their short and medium term out-of-sample forecasting performance. Our experiments use one, two and three factor linear Gaussian short rate models as defined in section 2.

4.1 UK gilt yield data

Gilt yields for 3 months, 6 months, 1 year, 2 years, 4 years, 8 years and 10 years from Datastream were used in the numerical experiments. Weekly data from January 2001 to June 2005 was considered. The data set contains 232 weekly observations with each observation consisting of 7 gilt yields of different maturities.

A principal component analysis of the bond yields shows that the nonzero eigenvalues of the correlation matrix of changes in yields are

$$\begin{bmatrix} 1 & 0.1071 & 0.0180 & 0.0038 & 0.0022 \end{bmatrix},$$

where all the eigenvalues are normalized with respect to the largest eigenvalue. We see that the first three eigenvalues account for 99.47% of total variation. This corroborates similar conclusions in Babbs and Nowman (1999). Figure 1 shows the first three principal components. The first component denoted by solid line clearly accounts for a parallel shift of the yield curve, the second component denoted by dashed line seems to account for twisting of the yield curve while the third component denoted by '$-\cdot$' seems to account for the slope of yield curve.

As mentioned previously, fminsearch in Matlab was used for non-convex optimization in calibration (see MATLAB (1995)). The parameters of one, two and three factor linear Gaussian models described earlier were estimated. In all the three cases, 1-180 data points were used to calibrate the model and the predictive ability of the model over the remaining period was tested, i.e. 181-232 data points were used for validation. The parameter estimation results are listed in tables 1-3 in the Appendix.

Let $L_i(\mathbf{b}_k, \theta_i)$ be the optimal value of negative log likelihood functions for a model with i factors, $i = 1, 2, 3$. Then a likelihood ratio test for the hypothesis that $j + 1$ factor model offers a better description of data than a j factor model is given by $LR_j = 2(L_{j+1}(\mathbf{b}_k, \theta_{j+1}) - L_j(\mathbf{b}_k, \theta_j)) \sim \chi^2_{1-\alpha}(d)$ where d is the number of parameter restrictions imposed to obtain a j factor model from a $j + 1$ factor model and $\chi^2_{1-\alpha}(d)$ denotes the $(1-\alpha)$ percentile of a χ^2 distribution with d degrees of freedom. In the present case, $LR_1 = 3807$

which rejects a one factor model in favor of a two factor model with more than 99% confidence. This is only to be expected, since a one factor model assumes that all the yields are perfectly correlated which is contrary to observations. Similarly, $LR_2 = 543$ which means a two factor model is rejected in favor of a three factor model more than 99% confidence.

It is customary in practice to use a short dated yield as a proxy for the short rate. In figure 2, we have plotted $\hat{r}_k = \mu - \sum_{i=1}^{3} \hat{\mathbf{x}}_{k|k-1,i}$ over time (solid line) and compared it with the observed three month yield (dotted line). It is seen that the two curves don't always have the same local slope or the even the same local level. This only serves to underline the importance of using unobservable factors instead of using proxy rates.

The experiments in one-step ahead prediction of the gilt yields using the models were performed. Instead of using standard deviation, mean relative absolute errors (MRAE) and maximum relative absolute errors (maximum RAE) are used as the error criterion for comparison of the models. MRAE corresponds to the percentage forecasting error and is hence seen as a sensible criterion for comparing the size of error from the point of view of forecasting a term structure accurately. MRAE (respectively, maximum RAE) is computed as the sample mean (respectively, maximum) of the relative absolute error:

$$\frac{|\text{observed yield-predicted yield}|}{\text{observed yield}}$$

over the relevant set of observations (either the in-sample data or the out-of sample data).

Tables 4-6 show the mean relative absolute errors for the three models when the entire yield spectrum is considered. The in-sample errors and the out-of-sample errors are computed separately. For a one factor model, the worst out-of-sample MRAE is seen to be 12.78%, which is reduced to 2.47% using a two factor model and further to 1.71% using a three factor model. To express this error in more traditional terms of basis points, the corresponding worst case error is 9 basis points in the relevant yield for a three factor model and is 72 basis points for a one factor model.

As the dimensions of observations was reduced, for example when only

the gilt yields with maturities 3 months, 6 months and 1 year were used instead of the yields with all the seven maturities, the forecasting ability of a one factor model shows a major improvement. A restricted yield spectrum may be useful when only the short dated or only the long dated yields are to be forecasted. The relative absolute errors in one-step ahead prediction when the yield spectrum is restricted to three short dated yields (with maturities 3 months, 6 months and 1 year) and when it is restricted to four long dated yields (with maturities 2 years, 4 years, 8 years and 10 years) are shown in tables 7 and 8 respectively. In both the cases, it is seen that restricting the yield spectrum makes the use of a single factor in short term prediction much more justifiable.

Tables 9 and 10 show the mean relative absolute 4-step ahead prediction errors for one and three factor models (a similar table for a two factor model is omitted for brevity). The n-step ahead prediction using Kalman filter is carried out using the formulae:

$$\hat{\mathbf{x}}_{k+n|k} = F^n \hat{\mathbf{x}}_{k|k}, \tag{18}$$

$$\hat{\mathbf{b}}_{k+n|k} = A\hat{\mathbf{x}}_{k+n|k} + B. \tag{19}$$

For a one factor model, the worst out-of sample error is seen to be 16.67% (which corresponds to an error of 87 basis points) while it is 6.46% using a two factor model and 4.33% (which corresponds to an error of 22 basis points) using a three factor model. Interestingly, the error for a one factor model doesn't show a significant increase as we move from one step ahead prediction to multi-step ahead prediction, while it increases significantly for a multi-factor model.

4.2   US treasury data

US treasury yields for 1 years, 2 years, 3 year, 5 years, 7 years, 10 years and 30 years from Datastream were used in the numerical experiments. Weekly data from December 1997 to August 2001 was considered. The data set contains 193 weekly observations with each observation consisting of 7 treasury yields of different maturities. While UK gilt data set represents a benign

11

economic environment, the US treasury data set is from a different and more volatile period. The purpose of using two different data sets during different periods and different countries is to assess whether the models under study and the yield curve matching heuristic proposed in the next section perform satisfactorily in different economic environments.

A principal component analysis of the bond yields shows that the eigenvalues of the correlation matrix of changes in yields are

$$\begin{bmatrix} 1 & 0.1443 & 0.025 & 0.0066 & 0.0026 & 0.0012 & 0.0006 \end{bmatrix},$$

where all the eigenvalues are normalized by the maximum eigenvalue. Similar to the previous case, we see that the first three eigenvalues account for more than 99% of total variation.

The parameters of one, two and three factor linear Gaussian models described earlier were estimated. In all the three cases, 1-180 data points were used to calibrate the model and the predictive ability of the model over the remaining period was tested, i.e. 181-193 data points were used for validation. The values of parameters obtained parameter estimation results are listed in tables 13-15. The results of one-step ahead prediction, in terms of mean relative absolute errors, are reported in tables 16-18. Again, it is seen that the the parameters of the models are quite stable and the out-of-sample prediction ability of the three factor model remains unaffected for a reasonably long period (up to three months) after calibration. Interestingly, the worst out-of-sample MRAE for a three factor model in this case is *worse* than the worst out-of-sample MRAE for a two factor model, as seen from tables 17 and 18, even though the likelihood ratio test seems to favor the three factor model. In a more volatile environment such as the one represented here, it appears that choosing a more parsimonious model may give better short term yield prediction. This, of course, is a single volatile data set and the conclusion has to be treated with caution.

The results of 4-step ahead prediction using one factor and multi-factor models and one step ahead prediction using a restricted spectrum one factor model are similar to the results with the UK gilt data in the previous section and are omitted for brevity.

5   A new method to improve the matching of initial term structure

While the main focus of the work presented so far has been short and medium term yield forecasting, it is worth considering what can be done to improve the term structure fit on a given day *after* observing the yields on that day. While standard methods (such as the Nelson-Siegel curves discussed in Nelson and Siegel (1985) and its variants, *e.g.* as discussed in chapter 15 of James and Webber (2000)) to interpolate a yield curve exist, it is known that most of these are not compatible with a Gaussian model, as shown in Björk and Christensen (1999). Compatibility issues are discussed later in section 5.1. Instead of looking for a parameterized yield curve consistent with a Gaussian model and then carrying out the necessary non-convex optimization, we take a simpler approach of using a nonlinear state correction filter based on linear programming. We consider the same state space system as in (6):

$$\mathbf{x}_k = F\mathbf{x}_{k-1} + \epsilon_k,$$
$$\mathbf{b}_k = A\mathbf{x}_k + B + \mathbf{e}_k, \tag{20}$$

where A, B and F are as defined in section 3. It is assumed that the system is already calibrated, *e.g.* using the Kalman filter as in the previous section. For a pre-calibrated model, we introduce a new assumption on the measurement noise in the *out-of-sample* data that the percentage error in the Kalman estimate is bounded:

$$|\mathbf{b}_{k,i} - (A\hat{\mathbf{x}}_{k|k} + B)_i| \leq \delta_k\, \mathbf{b}_{k,i},\, i = 1, 2, \ldots, M,$$

where $(\mathbf{z})_i$ represents the ith entry in vector $\mathbf{z}$ and δ_k is an unknown positive constant. This is a reasonable assumption from a practical point of view, provided the Kalman filter parameters are reasonably accurate. No further assumption is made about the statistical nature of the observation noise. Note that this does not change yield formulae (or A,B and F) since the mean and the variance of yield observations is relevant only to filtering and not to the actual yield relationships or to the assumptions on the underlying

linear Gaussian process. As such, using the properties of the observation noise as tuning parameters in a filtering procedure is certainly justified. The numerical value of δ_k need not be known. The purpose of this assumption is to establish a sensible feasible solution to our optimization problem described next. Assumptions on noise of this type are quite common in system identification and control literature; *e.g.* see Kacewicz (1999), Bravo et al. (2006) and references therein. With this assumption, an estimate of the unobserved state $\mathbf{x}_k$ may be obtained as the one which minimizes the *realized observation noise*, *i.e.* which solves the following linear programming problem at each time t_k:

$$\min_{\mathbf{x}_{k,1},\mathbf{x}_{k,2},\mathbf{x}_{k,3},\gamma} \gamma \text{ subject to}$$

$$\mathbf{b}_{k,i} - a_0(\tau_i) - \sum_{j=1}^{3} a_j(\tau_i)\mathbf{x}_{k,j} \leq \gamma\mathbf{b}_{k,i}$$

$$a_0(\tau_i) + \sum_{j=1}^{3} a_j(\tau_i)\mathbf{x}_{k,j} - \mathbf{b}_{k,i} \leq \gamma\mathbf{b}_{k,i}, \ i = 1, 2, \ldots, M \qquad (21)$$

where a_0, a_1 and a_2 are as defined in section 3, $\mathbf{b}_{k,i}$ is the possibly noisy yield measurement for time to maturity τ_i and M is the number of yield measurements at time t_k (at most 7 in our experiments). This minimization problem can be easily shown to be equivalent to minimizing the maximum relative absolute error

$$\frac{|\text{observed yield-predicted yield}|}{\text{observed yield}}$$

over yield measurements at time t_k. Both the objective function and the constraints in this optimization problem are linear in the decision variables *i.e.* it is a linear programming (LP) problem. Further, the following solution is always feasible: $\mathbf{x}_{k,j} = \left(\hat{\mathbf{x}}_{k|k}\right)_j$ and

$$\hat{\gamma}_k := \max_i \frac{\left|\mathbf{b}_{k,i} - \left(A\hat{\mathbf{x}}_{k|k} + B\right)_i\right|}{\mathbf{b}_{k,i}}$$

where $\hat{\mathbf{x}}_{k|k}$ is as defined in (12) and $(\mathbf{z})_i$ represents i^{th} entry in a vector $\mathbf{z}$, as before. Also, $\hat{\gamma}_k \leq \delta_k$, according to our assumption. Thus we have a

14

linear programming problem with a nonempty feasible set of solutions and its unique, bounded optimum can be found using efficient numerical techniques such as different variants of the simplex method; see, e.g. (Saigal (1995), chapter 4). With only 4 decision variables and $2M$ constraints, this problem can be solved extremely fast and yields an improvement in fitting the term structure.

With some abuse of notation, let $\hat{\mathbf{x}}_{k|k}$ be the arguments which minimize the above objective function. Then a prediction of state vector may be obtained by

$$\hat{\mathbf{x}}_{k|k-1} = F\hat{\mathbf{x}}_{k-1|k-1}.$$

However, note that $\hat{\mathbf{x}}_{k|k-1}$ is *not* used to find the updated estimate of state vector $\hat{\mathbf{x}}_{k|k}$.

It is difficult to give a justification for this method from a statistical point of view. The main heuristic justification is the fact that one finds the values of the state variables which achieve the smallest percentage errors in matching the term structure using this method. It is also worth mentioning that this method is consistent, by definition, to the underlying yield dynamics. The idea of consistency is elaborated upon in the next subsection.

5.1 Consistency with the underlying model

At time t_k, suppose that each yield $r_t(\tau)$ can be written as a specific parameterized function $f(\tau; \theta)$ for a given parameter vector θ. Further, suppose that the short rate is governed by a linear Gaussian process (1)-(2). It is natural to ask a question whether $r_t(\tau)$ can still be written in the same form $f(\tau, \theta)$ (perhaps for a different parameter vector θ) at time $t_{k+1} > t_k$. A parameterized yield curve which has this property is said to be *consistent* with the dynamics (1)-(2). More details and the formal definition of consistency may be found in Björk and Christensen (1999). Nelson-Siegel curves (Nelson and Siegel (1985)) are not consistent with the Gaussian dynamics as above. On the other hand, the proposed method for yield curve matching only modifies the unobservable factors and $r_t(\tau)$ is given by (3) at any time t (with $\mathbf{x}_{t,i}$ replaced by estimates $(\hat{\mathbf{x}}_{t|t})_i$). Hence the proposed yield curve matching is

15

consistent, by definition, to the underlying short rate dynamics (1)-(2). It is also worth noting that the parameters of the Gaussian model are quite stable and the proposed method requires only a simple linear program to be solved at each re-calibration to the observed yield curve. This is in contrast with the non-convex optimization needed to re-calibrate Nelson-Siegel type curves.

5.2   Empirical performance

Table 11 compares the out-of-sample mean relative absolute errors achieved using this linear programming step in place of Kalman update step, against a simple Kalman prediction and update, using a three factor model calibrated on UK gilt data in section 4.1 (with parameters listed in table 3). It is seen that the worst case error is reduced from 0.98% to 0.66% which is a very useful improvement at a negligible increase in computation cost. Table 12 shows out-of-sample mean relative absolute error when this modified method is used for one step ahead prediction. Compared to table 6, it is seen that one step ahead prediction using linear programming is better than Kalman filtering-based prediction, with the worst case error decreasing from 1.71% to 1.39%.

The results of yield curve matching and one step ahead prediction for a three factor model calibrated on US treasury data in section 4.2 (with parameters listed in table 15) are reported in tables 19 and 20 respectively. In this case, the improved worst out-of-sample MRAE for yield curve matching comes at the cost of deterioration of the worst out-of-sample MRAE for one-step ahead prediction. The latter fact can be seen by comparing tables 18 and 20. The proposed method may thus be seen as a trade-off between accuracy of one-step ahead prediction and that of yield curve matching.

6   Conclusions

An extensive empirical study were carried out for Kalman filtering-based calibration of linear Gaussian interest rate models using UK gilt yields data

and US treasury data. The following conclusions can be drawn from the evidence presented:

- A multi-factor model outperforms a one factor model in short term yield forecasting both in-sample and out-of-sample when the entire yield curve is considered. When only closely spaced (in logarithmic terms) yields are considered, even a one factor model is seen to be adequate for one-step ahead prediction.

- A three factor model performs better than a two factor model for UK gilt data and performs marginally worse for US treasury data. While this is a very limited evidence, it appears that a more parsimonious two factor model may actually be preferable in a volatile environment as compared to a three factor model. The short-term predictive ability of a linear Gaussian model is reasonably stable, with the mean relative error remaining at or below 2.47% for UK gilt data up to one year after calibration and at or below 3.50% for US treasury data up to three months after calibration, in the two factor case.

- In multi-step ahead prediction with four time-steps, the predictive ability of models deteriorates, as can be expected, although the performance of the multi-factor models may still be satisfactory from a portfolio management point of view. As far as one factor model is concerned, the error in multi-step ahead prediction is not much worse than the error in one-step ahead prediction.

- A valid criticism of Gaussian model is that they allow negative values of interest rate with positive probability. In forecasting terms, this did not pose a problem, since none of the conditional expectations computed (around 10000 in-sample and out-of-sample yield predictions in all) were close to zero. This may obviously be a problem in markets with very low interest rates (e.g. in modelling Japanese bonds at the time of writing).

- Further, a new linear programming state update heuristic was suggested to improve fit to a given term structure, while still operating in

"linear Gaussian model with unobservable states" framework. It was demonstrated that the modified update step offers a very useful trade-off between the accuracy in yield prediction and the accuracy in yield curve matching.

This study clearly provides a valuable evidence of the utility of a simple, linear Gaussian interest rate model when adequate yield measurements are available. Applications of the methodology presented include economic forecasting and scenario generation for a stochastic optimization of fixed income portfolios and downside risk prediction of gilt or treasury bond portfolios.

Acknowledgement

References

Babbs, S. H., Nowman, K. B., 1999. Kalman filtering of generalized vasicek term structure models. Journal of Financial and Quantitative Analysis 34, 115–130.

Balduzzi, S. F. P., Das, S. R., Sundaram, R., 1996. A simple approach to three factor affine term structure models. Journal of Fixed Income 6, 43–53.

Ball, C. A., Torous, W. N., 1996. Unit roots and estimation of interest rate dynamics. Journal of Empirical Finance 3, 215–238.

Beaglehole, D. R., Tenny, M. S., 1991. General solutions of some interest rate-contingent claim pricing equations. Journal of Fixed Income 1, 69–83.

Björk, T., Christensen, B. J., 1999. Interest rate dynamics and consistent forward rate curves. Mathematical Finance 9, 323–348.

Bravo, J., Alamo, T., Camancho, E., 2006. Bounded error identification of systems with time varying parameters. IEEE Transactions on Automatic Control 51, 1144–1150.

Chan, K. C., Karolyi, G. A., Longstaff, F. A., Sanders, A. B., 1992. An empirical comparison of alternative models of short-term interest rate. Journal of Finance 47, 1209–1227.

Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–407.

Duan, J. C., Simonato, J. G., 1995. Estimating and testing exponential-affine term structure models by Kalman filter. Working paper. Mcgill University, pp. 1–29.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

Durbin, J., Koopman, S., 2002. Time series analysis by state space methods. Oxford University Press.

Gravelle, T., Morley, J., 2005. A Kalman filter approach to characterizing the Canadian term structure of interest rates. Applied Financial Economics 15, 691–705.

Harvey, A. C., 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press.

Hull, J. C., White, A. D., 1990. Pricing interest rate derivative securities. Review of Financial Studies 3-4, 573–592.

Hull, J. C., White, A. D., 1993. One-factor interest-rate models and the valuation of interest-rate derivative securities. Journal of Financial and Quantitative Analysis 28, 235–254.

Hull, J. C., White, A. D., 1994. Numerical procedures for implementing term structure models 1: single-factor models. Journal of Derivatives 2, 7–16.

James, J., Webber, N., 2000. Interest Rate Modelling. John Wiley.

Jegadeesh, N., Pennacchi, G. G., 1996. The behaviour of interest rates implied by term structure of eurodollar futures. Journal of Money, Credit and Banking 28, 426–446.

Kacewicz, B., 1999. Worst case conditional system identification in a general class of norms. Automatica 35, 1049–1058.

Lanne, M., 2000. Testing the expectations hypothesis of the term structure of interest rates in the presence of a potential regime shift. Computing in Economics and Finance 2000, Society for Computational Economics, available at http://ideas.repec.org/p/sce/scecf0/294.html.

Longstaff, F. A., Schwartz, E. S., 1992. Interest rate volatility and the term structure: a two factor general equilibrium model. Journal of Finance 47, 1259–1282.

Lund, J., 1997. Non linear Kalman filtering techniques for term structure models. Working paper. University of Aarhus, pp. 1–34.

MATLAB, 1995. Optimization toolbox manual. Mathworks Inc.

Nelson, C., Siegel, A., 1985. Parsimonious modelling of yield curves. Journal of Business 60, 473–489.

Reisman, H., Zohar, G., 2004. Short-term predictability of the term structure. The Journal of Fixed Income 14, 7–15.

Rhee, J., 1999. Interest rate models. PhD thesis, University of Warwick, Warwick, UK.

Rossi, G. D., 2004. Kalman filtering of consistent forward rate curves: a tool to estimate and model dynamically the term structure. Journal of Empirical Finance 11, 277–308.

Saigal, R., 1995. Linear programming: a modern integrated analysis. Kluwer Academic Publishers.

Schrager, D. F., Pelsser, A. A. J., 2006. Pricing swaption and coupon bond options in affine term structure models. Mathematical Finance 6, 673–693.

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

Appendix: Tables and plots

UK gilt data

Figure 1: Principal components of yield correlation matrix
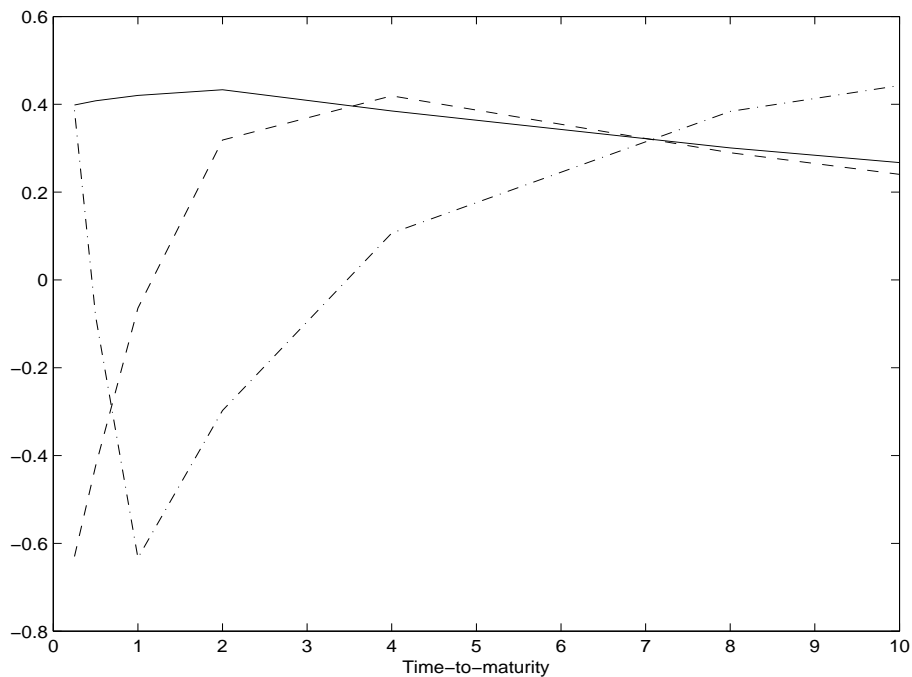
Table 1 : Parameter values for one factor model (computation time: 161 seconds)

Parameter values			
α_1	0.0448	σ_1	0.0356
λ_1	0.3405	μ	-0.4146
h_1	0.0044	h_2	0.0034
h_3	0.0022	h_4	0.0013
h_5	0.0012	h_6	0.0009
h_7	0.0013		

Figure 2: Theoretical short rate and three month yield
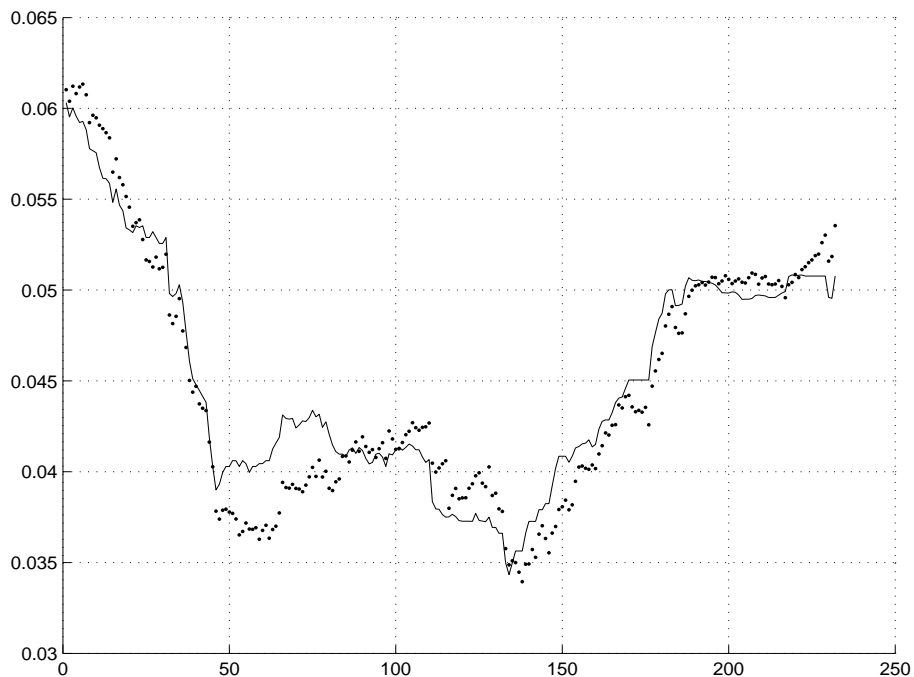
Table 2 : Parameter values for two factor model (computation time: 255 seconds)

Parameter values			
α_1	0.5854	α_2	0.2619
σ_1	0.0111	σ_2	0.2339
λ_1	−4.6918	λ_2	0.2162
μ	0.3084	h_1	0.0006
h_2	0.0006	h_3	0.0015
h_4	0.0015	h_5	0.0012
h_6	0.0009	h_7	0.0009
σ_{12}	0.0356		

Table 3 : Parameter values for three factor model (computation time: 366 seconds)

Parameter values					
α_1	0.1327	α_2	0.2260	α_3	2.0962
σ_1	0.0522	σ_2	0.0035	σ_3	0.0004
λ_1	0.0023	λ_2	-8.5213	λ_3	0.0010
σ_{12}	0.0002	μ	0.0023	h_1	0.0004
h_2	0.0010	h_3	0.0012	h_4	0.0008
h_5	0.0011	h_6	0.0007	h_7	0.0007

Table 4 : Relative absolute errors of 1-step ahead prediction for one factor model

$yields$	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
$3m$	0.0716	0.1278	0.2511
$6m$	0.0574	0.1210	0.2145
$1y$	0.0395	0.0993	0.1557
$2y$	0.0336	0.0589	0.1058
$4y$	0.0255	0.0105	0.0377
$8y$	0.0171	0.0346	0.0640
$10y$	0.0242	0.0272	0.0557

Table 5 : Relative absolute errors of 1-step ahead prediction for two factor model

$yields$	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
$3m$	0.0112	0.0082	0.0331
$6m$	0.0154	0.0077	0.0268
$1y$	0.0331	0.0164	0.0516
$2y$	0.0346	0.0152	0.0609
$4y$	0.0244	0.0145	0.0354
$8y$	0.0277	0.0247	0.0430
$10y$	0.0174	0.0213	0.0465

Table 6 : Relative absolute errors of 1-step ahead prediction for three factor model

yields	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
3m	0.0186	0.0109	0.0381
6m	0.0221	0.0171	0.0373
1y	0.0306	0.0167	0.0439
2y	0.0299	0.0108	0.0404
4y	0.0258	0.0122	0.0297
8y	0.0209	0.0113	0.0280
10y	0.0209	0.0090	0.0339

Table 7 : Relative absolute errors of 1-step ahead prediction for a one factor, three yields model

yields	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
3m	0.0191	0.0095	0.0437
6m	0.0145	0.0058	0.0222
1y	0.0436	0.0209	0.0562

Table 8 : Relative absolute errors of 1-step ahead prediction for a one factor, four yields model

yields	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
2y	0.0272	0.0420	0.0910
4y	0.0206	0.0100	0.0366
8y	0.0157	0.0312	0.0626
10y	0.0162	0.0376	0.0713

Table 9 : Relative absolute errors of 4-step ahead prediction for one factor model

yields	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
3m	0.0865	0.1667	0.2625
6m	0.0733	0.1568	0.2303
1y	0.0649	0.1311	0.2086
2y	0.0667	0.0867	0.1679
4y	0.0561	0.0313	0.1061
8y	0.0403	0.0231	0.0712
10y	0.0502	0.0202	0.0627

Table 10 : Mean relative absolute errors of 4-step ahead prediction for three factor model

yields	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
3m	0.0432	0.0414	0.0683
6m	0.0434	0.0433	0.0852
1y	0.0506	0.0359	0.0989
2y	0.0505	0.0255	0.0938
4y	0.0415	0.0244	0.0785
8y	0.0313	0.0224	0.0744
10y	0.0297	0.0199	0.0768

Table 11 : Out-of-sample mean relative absolute errors of matching a given term structure for three factor model with and without linear programming update

maturity	Kalman state update	Linear programming state update
3m	0.0015	0.0066
6m	0.0076	0.0025
1y	0.0098	0.0063
2y	0.0029	0.0022
4y	0.0080	0.0066
8y	0.0042	0.0015
10y	0.0032	0.0066

Table 12 : Out-of-sample mean relative absolute errors of one step ahead prediction for three factor model with linear programming update

maturity	Linear programming state update
3m	0.0064
6m	0.0117
1y	0.0139
2y	0.0104
4y	0.0118
8y	0.0114
10y	0.0115

US treasuries data

Table 13 : Parameter values for one factor model (computation time: 231 seconds)

Parameter values			
α_1	0.1237	σ_1	0.042
λ_1	-0.1047	μ	0.0044
h_1	0.0011	h_2	0.0007
h_3	0.0004	h_4	0.0006
h_5	0.0007	h_6	0.0006
h_7	0.0003		

Table 14 : Parameter values for two factor model (computation time: 320 seconds)

Parameter values			
α_1	0.5832	α_2	0.1804
σ_1	0.2191	σ_2	0.0107
λ_1	0.2394	λ_2	-0.1025
μ	0.0567	h_1	0.0007
h_2	0.0013	h_3	0.0009
h_4	0.0010	h_5	0.0010
h_6	0.0016	h_7	0.00214
σ_{12}	0.0136		

Table 15 : Parameter values for three factor model (computation time: 416 seconds)

Parameter values					
α_1	0.2756	α_2	1.729	α_3	0.0096
σ_1	0.0027	σ_2	0.0051	σ_3	0.0036
λ_1	-0.7098	λ_2	0.7314	λ_3	0.0049
σ_{12}	0.0000	μ	0.08	h_1	0.0008
h_2	0.0009	h_3	0.0009	h_4	0.0010
h_5	0.0014	h_6	0.0007	h_7	0.0006

Table 16 : Relative absolute errors of 1-step ahead prediction for one factor model

$yields$	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
$1y$	0.0171	0.0291	0.0634
$2y$	0.0374	0.0373	0.1009
$3y$	0.0402	0.0466	0.1053
$5y$	0.0478	0.0762	0.1179
$7y$	0.0499	0.0843	0.1269
$10y$	0.0524	0.0655	0.0888
$30y$	0.0474	0.0141	0.0275

Table 17 : Relative absolute errors of 1-step ahead prediction for two factor model

$yields$	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
$1y$	0.0161	0.0214	0.0461
$2y$	0.0235	0.0293	0.0645
$3y$	0.0198	0.0350	0.0647
$5y$	0.0202	0.0236	0.0542
$7y$	0.0198	0.0167	0.0423
$10y$	0.0178	0.0079	0.0195
$30y$	0.0251	0.0144	0.0291

Table 18 : Relative absolute errors of 1-step ahead prediction for three factor model

yields	in-sample MRAE	out-of-sample MRAE	Maximum out-of-sample RAE
1y	0.0221	0.0226	0.0545
2y	0.0247	0.0398	0.0748
3y	0.0254	0.0410	0.0704
5y	0.0257	0.0184	0.0369
7y	0.0299	0.0182	0.0560
10y	0.0170	0.0132	0.0306
30y	0.0177	0.0147	0.0317

Table 19 : Out-of-sample mean relative absolute errors of matching a given term structure for three factor model with and without linear programming update

maturity	Kalman state update	Linear programming state update
1y	0.0129	0.0217
2y	0.0169	0.0111
3y	0.0228	0.0215
5y	0.0075	0.0056
7y	0.0244	0.0190
10y	0.0197	0.0133
30y	0.0089	0.0221

Table 20 : Out-of-sample mean relative absolute errors of one step ahead prediction for three factor model with linear programming update

maturity	Linear programming state update
1y	0.0412
2y	0.0628
3y	0.0563
5y	0.0321
7y	0.0310
10y	0.0183
30y	0.0334