Federal Reserve Bank of Chicago

Financial Signal Processing: A Self Calibrating Model

By: Robert J. Elliott, William C. Hunter and Barbara M. Jamieson

WP 2000-21

Financial Signal Processing:
A Self Calibrating Model

Robert J. Elliott
Department of Mathematical Sciences, University of Alberta,
Edmonton, Alberta, Canada T6G 2G1

William C. Hunter
Federal Reserve Bank of Chicago, 230 South LaSalle St.
Chicago, Illinois 60604 U.S.A.

and

Barbara M. Jamieson
Department of Finance and Management Science
University of Alberta
Edmonton, Alberta, Canada T6G 2R6

September 2000

The views expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System. Robert Elliott wishes to thank SSHRC for its support, and also the Federal Reserve Bank for its hospitality.

FINANCIAL SIGNAL PROCESSING: A SELF CALIBRATING MODEL

ROBERT J. ELLIOTT[1]

WILLIAM C. HUNTER[2]

and

BARBARA M. JAMIESON[3]

[1] Department of Mathematical Sciences, University of Alberta,
Edmonton, Alberta, Canada T6G 2G1

[2] Federal Reserve Bank, 230 South LaSalle St.
Chicago, Illinois 60604, U.S.A.

[3] Department of Finance and Management Science
University of Alberta,
Edmonton, Alberta, Canada T6G 2R6

Abstract. Previous work on multifactor term structure models has proposed that the short rate process is a function of some unobserved diffusion process. We consider a model in which the short rate process is a function of a Markov chain which represents the 'state of the world'. This enables us to obtain explicit expressions for the prices of zero-coupon bonds and other securities. Discretizing our model allows the use of signal processing techniques from Hidden Markov Models. This means we can estimate not only the unobserved Markov chain but also the parameters of the model, so the model is self-calibrating. The estimation procedure is tested on a selection of U.S. Treasury bills and bonds.

Keywords: Term structure, bond pricing, Hidden Markov Model, filtering.

Acknowledgments. Robert Elliott wishes to thank SSHRC for its support, and also the Federal Reserve Bank for its hospitality.

1. Introduction

The modelling of interest rates and term structure is a central problem in financial theory. A comprehensive survey can be found in the paper of Duffie and Kan (1993a). There, both single factor and multi-factor term structure models are described and a new model for the 'short rate' process r_t is proposed under which r_t is a function $r(X_t)$ of a 'state process' X_t which takes values in a subset $D \subset R^n$. In fact Duffie and Kan suppose that X_t is given by a stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dw_t$$

$$X_0 \in D \subset R^n, \quad t \geq 0,$$

where μ and $\sigma\sigma'$ are affine functions of X, (i.e. $\mu(X) = A + BX$). Duffie and Kan then show that the yield process also has the form

$$y_{t,\tau} = \alpha(\tau) + \beta(\tau) \cdot X_t.$$

It is known from the work of Harrison and Kreps (1979), and others, that only technical conditions are required for the equivalence between the absence of arbitrage and the existence of an equivalent martingale measure. Here we are assuming our processes are defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and a martingale measure is a probability measure Q equivalent to P such that the price processes of any security is a Q martingale after normalization at each time t by $\exp\left(\int_0^t r_s ds\right) = \exp\left(\int_0^t r(X_s)ds\right)$. Duffie and Kan (1993a,b) do not discuss the existence of an equivalent martingale measure, but assume such a measure Q exists; we shall follow their example and work under measure Q.

However, instead of assuming the short term rate r is a function of another diffusion process we shall suppose r is a function of a continuous time Markov chain. This is not unreasonable as any diffusion can be approximated by a Markov chain. Discrete time

Markov chain models for term structure have been discussed by Pye (1966) and Zipkin (1993). However, the novel feature of this work is the application of new results from Elliott, Aggoun and Moore (1995) which provide not only recursive estimates of the Markov chain but also formulae for re-estimating all parameters of the model, so that our model is 'self calibrating'.

Maximum likelihood estimation of the Cox, Ingersoll, Ross term structure model is carried out in the paper by Pearson and Sun (1994). The conclusion of the Pearson and Sun (1994) paper is that their data rejects the Cox, Ingersoll, Ross model. Filtering methods provide a continual, recursive up-date of optimal estimates in contrast to the static model-fitting of maximum likelihood. Consequently, our application of Hidden Markov filtering and estimation techniques appears new. We do not need to specify a priori the dynamics of the short rate process, other than to say it is a Markov chain.

2. Short Term Rate

Processes will be defined on a probability space $(\Omega, \mathcal{F}, Q)$ where, for pricing purposes, Q is an equivalent martingale measure.

Suppose $\{X_t\}$, $t \geq 0$, is a finite state Markov chain on $(\Omega, \mathcal{F}, Q)$ with state space $S = \{s_1, s_2, \dots, s_N\}$. Here the points s_i may be points in R^N, or any space whatsoever; however, without loss of generality we may identify the points in S with the unit vectors $\{e_1, e_2, \dots, e_N\}$. We suppose our vectors are column vectors in R^N, so with $'$ denoting transpose $e_i = (0, \dots, 0, 1, 0, \dots, 0)' \in R^N$. This representation of the state space of X will simplify the algebra. E will denote expectation under measure Q. The distribution of X_t is then $E[X_t] = p_t = (p_t^1, p_t^2, \dots, p_t^N)$, where $p_t^i = Q(X_t = e_i) = E[\langle X_t, e_i \rangle]$.

We suppose this distribution evolves according to the Kolmogorov equation

$$\frac{dp_t}{dt} = Ap_t;$$

here A is a 'Q-matrix', that is, if $A = (a_{ji})$, $1 \le i$, $j \le N$, $\sum_{j=1}^N a_{ji} = 1$ and $a_{ji} \ge 0$ if $i \ne j$. We could take the components a_{ji} to be time varying, though this would complicate their estimation.

Recall that at any time t the state X_t of the Markov chain is one of the unit vectors, $e_1, e_2, \ldots, e_N$. Consequently, any real valued function of X_t, say $h(X_t)$, is just given by a vector $(h_1, h_2, \ldots, h_N) = h$, so $h(X_t) = \langle h, X_t \rangle$ where the brackets denote the scalar product in R^N.

HYPOTHESIS 1.1. We suppose the short rate process r_t is a function of X_t, that is, $r_t = r(X_t) = \langle r, X_t \rangle$ for some vector $r \in R^N$.

Again, we could take r to be time varying. This would still allow estimation of X but would complicate the parameter estimation of the model.

Write $\{\mathcal{F}_t\}$ for the right continuous, complete, filtration generated by X. Then, because we are working under the equivalent martingale measure Q, the price of a security paying u at time $T \ge t$ is given by

$$E[\exp\big(-\int_t^T r(X_s)ds\big)u|\mathcal{F}_t] = E[\exp\big(-\int_t^T r(X_s)ds\big)u|X_t],$$

as X is a Markov process. Taking $u = 1$ we obtain the price of a zero-coupon bond maturing at time T :

$$p_{t,T} = E[\exp\big(-\int_t^T r(X_s)ds\big)|X_t].$$

The yield for such a bond is then $y_{t,T} = -\frac{1}{(T-t)} \log p_{t,T}$. Note all expectations are under Q, so the dynamics and estimates derived below are also under measure Q.

3. Bond Dynamics

Suppose we have a security paying $u(X_T)$ at time $T \ge t$. As noted above, its price at time t is

$$F(X_t, t) = E[\exp\big(-\int_t^T r(X_s)ds\big)u(X_T)|X_t].$$

This is a function of X_t. Consequently there is a function $\phi_t = (\phi_t^1, \phi_t^2, \ldots, \phi_t^N)' \in R^N$, where $\phi_t^i = F(e_i, t)$, such that

$$F(X_t, t) = \langle \phi_t, X_t \rangle \ .$$

Now

$$\begin{aligned}
\exp\big(-\int_0^t r(X_s)ds\big)F(X_t, t) &= \exp\ \big(-\int_0^t r(X_s)ds\big)\ \langle \phi_t, X_t \rangle \\
&= E[\exp\big(-\int_0^T r(X_s)ds\big)u(X_T)|\mathcal{F}_t],
\end{aligned}$$

and so is a $(Q, \mathcal{F}_t)$ martingale.

Differentiating we have the following Itô representation:

$$\begin{aligned}
\exp\big(-\int_0^t r(X_s)ds\big)F(X_t, t) = F(X_0, 0) &+ \int_0^t \big(-r(X_s)\exp\big(-\int_0^s r(X_v)dv\big)\ \langle \phi_s, X_s \rangle\ ds \\
&+ \int_0^t \exp\big(-\int_0^s r(X_v)dv\big)\big[\langle \frac{d\phi_s}{ds}, X_s \rangle + \langle \phi_s, AX_s \rangle\ \big]ds \\
&+ \int_0^t \exp\big(-\int_0^s r(X_v)dv\big)\ \langle \phi_s, dM_s \rangle\ . \qquad (3.1)
\end{aligned}$$

We are using here the representation

$$F(X_s, s) = \langle \phi_s, X_s \rangle,$$

as above; consequently $dF(X_s, s) = \langle d\phi_s, X_s \rangle + \langle \phi_s, dX_s \rangle$. Further, the semimartingale form of the Markov chain is:

$$X_t = X_0 + \int_0^t AX_v dv + M_t \qquad (3.2)$$

where $\{M_t\}$ is a $(Q, \mathcal{F}_t)$ martingale. For the derivation of (3.2) see Elliott, Aggoun and Moore (1995). The left side of (3.1) is a martingale; therefore, the bounded variation terms on the right side of (3.1) must be the identically zero process. That is

$$\exp\big(-\int_0^s r(X_u)du\big)[-r(X_s)\langle\phi_s, X_s\rangle + \langle\frac{d\phi_s}{ds}, X_s\rangle + \langle\phi_s, AX_s\rangle] = 0. \tag{3.3}$$

Now $r(X_s) = \langle r, X_s\rangle$ where $r = (r_1, r_2, \dots, r_N)'$ and $r(X_s)\langle\phi_s, X_s\rangle = \langle \operatorname{diag} r \cdot \phi_s, X_s\rangle$, where $\operatorname{diag} r$ is the matrix with r on the diagonal. Therefore, from (3.3), with A^* denoting the transpose of matrix A,

$$\langle\frac{d\phi_s}{ds}, X_s\rangle + \langle A^*\phi_s, X_s\rangle - \langle \operatorname{diag} r \cdot \phi_s, X_s\rangle = 0 \quad \text{for all} \quad X_s.$$

Consequently, ϕ_s is given by the system of equations

$$\frac{d\phi_t}{dt} = (\operatorname{diag} r - A^*)\phi_t$$

with terminal condition

$$\phi_T = u = (u_1, u_2, \dots, u_N)',$$

where $u(X_T) = \langle u, X_T\rangle$.

Write $B = \operatorname{diag} r - A^*$, so $\phi_t = e^{-B(T-t)}u$ and the price at time $t \le T$ of a security paying $u(X_T)$ at time T is

$$\langle\phi_t, X_t\rangle = \langle e^{-B(T-t)}u, X_t\rangle.$$

A zero-coupon bond corresponds to taking $u = \mathbf{1} = (1, 1, \dots, 1)'$ so its price at time $t \le T$ is $F(X_t, t) = \langle e^{-B(T-t)}\mathbf{1}, X_t\rangle$.

4. Filtering and Model Estimation

We have shown the zero-coupon bond price which expires at time $t+\tau_i$, $i = 1, \dots, m$, is

$$\begin{aligned} F^i(X_t,t) &= E[\exp\big(-\int_t^{t+\tau_i} r(X_s)ds|X_t] \\ &= \langle e^{-B\tau_i}\mathbf{1}, X_t\rangle. \end{aligned}$$

Our hypothesis now is that the X process represents unobserved factors which would give rise to bond prices $F^i(X_t,t)$ in some 'ideal' world. That is, we suppose the state process X_t and the corresponding prices $F^i(X_t,t)$ are not observed directly. Rather we observe these prices $F^i(X_t,t)$ in noise for the different times to maturity $\tau_1,\dots,\tau_m$, at discrete times $t_1,t_2,\dots,t_k,\dots$. As $F^i(X_t,t)>0$ we suppose the noise is multiplicative, that is, we suppose that what we actually observe are the quantities $F^i(X_t,t)e^{\langle\widehat{\sigma}^i,X_t\rangle b_t^i}$, $i=1,\dots,m$, where $b_t^i\sim N(0,1)$, $t=t_j$, $j=1,2,\dots,k,\dots$ and $\widehat{\sigma}^i = \big(\widehat{\sigma}^i(1),\widehat{\sigma}^i(2),\dots,\widehat{\sigma}^i(N)\big)'$. (Note that theoretically this model would allow bond prices greater than one; in practice these are not observed.) In the estimation theory developed below the Gaussian random variables b, representing noise, could be replaced by other random variables, such as those with 'long tails'.

Equivalently, we suppose we observe the yield values in additive noise:

$$\begin{gathered} y_t^i = -\frac{1}{\tau_i}\,\log\,F^i(X_t,t) - \frac{\langle\widehat{\sigma}^i,X_t\rangle}{\tau_i}\,b_t^i, \\ i=1,\dots,m, \quad t=t_1,t_2,\dots,t_k,\dots\ . \end{gathered}$$

Write $\sigma^i(j) = -\widehat{\sigma}^i(j)/\tau_i$ and

$$\sigma^i = \big(\sigma^i(1),\sigma^i(2),\dots,\sigma^i(N)\big)'.$$

Also,

$$\begin{aligned} -\frac{1}{\tau_i}\,\log\,F^i(X_t,t) &= -\frac{1}{\tau_i}\,\log\langle e^{-B\tau_i}\mathbf{1},X_t\rangle \\ &= \langle g^i,X_t\rangle \end{aligned}$$

for $g^i \in R^N,\ i = 1, \ldots, m$. Here $g^i = \big(g^i(1), g^i(2), \ldots, g^i(N)\big)$ and $g^i(j) = -\frac{1}{\tau_i} \log \langle e^{-B\tau_i}\mathbf{1}, e_j\rangle$. Suppose the observation times $t_1 \leq t_2 \leq t_3 \leq \ldots$ are equally spaced, that is, $t_{j+1} - t_j = s > 0$, and write $\ell = t_\ell$,

$$e^{As} = P.$$

Then we have a discrete time version of the state process $X_\ell = X_{t_\ell}$, $\ell = 1, 2, \ldots, k, \ldots$ with

$$X_\ell = PX_{\ell-1} + M_\ell$$

where M_ℓ is an $(\mathcal{F}_\ell, Q)$ martingale increment. The multivariate observation process y has dynamics

$$y^i_\ell = \langle g^i, X_\ell\rangle + \langle \sigma^i, X_\ell\rangle\, b^i_\ell, \quad 1 \leq i \leq m, \quad \ell = 1, 2, \ldots, k, \ldots\ .$$

The filtering and estimation algorithms for the parameters $P = (p_{ij})$, g^i, σ^i are given in the Appendix.

5. Application

The results of the paper were applied in an example using data on the yields of 3-month and 6-month U.S. Treasury bills and 10-year and 30-year U.S. bonds. In what follows, the choice of the value of parameter N is discussed, the data are described, and the results of parameter estimation and yield prediction are presented and evaluated.

5.1. The Choice of N

In the estimation procedure proposed above, parameter N, which represents the size of the state space of the Markov chain, is the only parameter which is not estimated. Rather, a value is assigned to N which can be thought to represent the number of states of the world, e.g. ‘good’ and ‘bad’.

The determination of the optimal value of N, for a particular data set, is an important problem which has been considered in the literature. Although this problem cannot be resolved using the likelihood ratio test, a number of proposals have been advanced to address it.[1] In this paper, we do not explore this issue further, other than to compare the results obtained when N is assigned different values. However, it is interesting to note that in the regime-switching model, discussed in Hamilton (1988, 1994) for example, in which the state or regime of a time series process is modelled as a Markov chain, a state space of size two is typically assumed.

The model that we propose here requires the estimation of $N^2 + 2mN$ parameters, where m equals the number of securities being considered (four, in this case). Thus, the dimensionality of the model increases rapidly as N increases in size. From a numerical point of view, a smaller value of N is, therefore, preferable, unless our results are sensitive to the value assigned to N and superior for large N. Also, there is evidence, reported in what follows, that suggests that the model is overfitting the data; this provides further support for assigning N a smaller, rather than larger, value.

5.2. Data Description

We assembled a data set consisting of 270 weekly observations on the yields of 3-month and 6-month U.S. Treasury bills and 10-year and 30-year U.S. coupon bonds. The data were compiled by the Royal Bank of Canada and published in *The Financial Post.* The sample period ran from January 17, 1992 to June 21, 1997.

Table 1 provides descriptive statistics for the yield data, demonstrating that the yield curve was typically upward sloping during the sample period. If the entire period is considered, yield volatilities, as measured by standard deviations, appear to decline with maturity. However, if we consider the two subperiods, before mid-1994 and after mid-1994, a reverse pattern is observed; that is, yield volatilities increase with maturity. Also, the earlier period coincides with a relatively low short-term rate regime and the later period

[1] For more on this issue, see Hamilton (1994), pages 698-699.

with a relatively high short-term rate regime.

5.3. Analysis and Results

A computer program, written to implement the estimation procedure proposed in Section 4, was run on the data set. The data were processed in 18 groups of observations, the first group consisting of 100 yield vectors and subsequent groups consisting of 10 yield vectors each. At the end of each of the 18 passes through the data, parameter estimates were updated using the formulas given in the paper. As well, price estimates were obtained.

To start, a value of two was assigned to parameter N. Because the first group of observations consists of 100 yield vectors, a value of N in excess of 15 would be inappropriate since it would result in a situation where the number of parameters to be estimated exceeds the number of observations.

Table 2 gives the initial values that were assumed for the distribution of the state of the Markov chain, that is, for $E[X_0]$, and for the matrices g and σ. All entries in the transition matrix, A, were assigned an initial value of $1/N$. Table 2 also reports the re-estimated values of these parameters after the eighteenth pass through the data. Estimated prices are also reported. Note that the columns of the g and σ matrices and the price vector correspond, respectively, to the 3- and 6-month Treasury bills and the 10- and 30-year bonds.

To assess the predictive performance of the model, we calculated predicted yields using the formula:

$$E[y^i_{l+1} \mid y_1, \dots, y_l] = \langle g^i, A\hat{X}_l \rangle$$

where $\hat{X}_l = E[X_l \mid y_l]$, $i = 1, \dots, m$, and $l = 1, 2, \dots, k, \dots$. At the end of each pass through the data, predicted yields for the following week were obtained for each of the four securities considered. We then regressed actual yields on predicted yields for each of the securities in turn, using the model:

$$\text{Actual yield } = \alpha + \beta * \text{ Predicted yield } + \epsilon \ .$$

The regression results obtained were then assessed on the basis of the following criteria proposed by Fama and Gibbons (1984): (1) conditional unbiasedness, that is, an intercept, α, close to zero, and a regression coefficient, β, close to one; (2) serially uncorrelated residuals; and (3) a low residual standard error. Table 3 reports the results. For each of the four securities considered, the results suggest that the first criterion, conditional unbiasedness, is satisfied. However, the standard error of the estimate of α is relatively large in all four cases. Also, while the results for the 3- and 6-month Treasury bills indicate that the test for serially uncorrelated residuals is inconclusive, evidence for the 10- and 30-year bonds supports the hypothesis that residuals are not autocorrelated.

Figures 1 to 4 provide plots of actual yields and predicted yields for the 3- and 6-month Treasury bills and the 10- and 30-year bonds, respectively. They also give standard error bands for the predicted yields, derived using the method of cross-validation. To apply this method, we divided the data in half. The first 135 observations were processed in 18 groups, the first group consisting of 50 observations and subsequent groups consisting of 5 yield vectors each. At the end of each of the 18 passes through the data, a yield prediction for the following week was obtained. For each of the securities in turn, the predicted yields were then compared to the actual yields, residuals were recorded, and the standard deviation of the residuals was calculated. The second set of 135 observations was similarly processed and a set of 18 vectors of predicted yields obtained. These values were plotted, along with the corresponding actual yields, in Figures 1 to 4. Standard error bands around the predicted yields were derived by adding and subtracting to each predicted yield an amount equal to the product of a critical value, 1.96, and the standard deviation of the residuals derived using the first 135 observations.

The use of cross-validation to determine the standard error bands was deemed necessary because evidence suggested that the model was overfitting the data. This evidence included the observations that our data on Treasury bill and bond yields are not highly variable; our estimates of yield volatility, as reported in the σ matrix, are small in value; and standard error bands for predicted yields, calculated using the following formula:

$$\langle g^i, A\hat{X}_l \rangle \pm 1.96 \sqrt{g^{i\prime}(\operatorname{diag} A\hat{X}_l) g^i + \sigma^{i\prime}(\operatorname{diag} A\hat{X}_l)\sigma^i - [\langle g^i, A\hat{X}_l \rangle]^2} \ ,$$

$i = 1, \dots, m$ and $l = 1, 2, \dots, k, \dots$, are too narrow in width. Thus, it was deemed appropriate to derive the standard error bands using cross-validation methods.

Although possible overfitting of the data militates against increasing the value of $N,$ the analysis was repeated for some alternative N values to determine whether the results appear sensitive to changes in the value of $N.$ Tables 4 and 5 report results, of the type given in Table 2, for the cases $N = 4$ and $N = 6,$ respectively. The estimated values reported here are similar in magnitude to those given earlier in Table 2. Tables 6 to 8 report results, of the type given in Table 3, for the cases $N = 4,$ $N = 6,$ and $N = 9,$ respectively. These results lead to conclusions identical to those reached earlier for Table 3.

6. Conclusion

Our model of the short rate process gives rise to expressions for yields which incorporate two random components: a Markov chain X and a Gaussian noise term $b.$ Filtering techniques, using new results presented in Elliott, Aggoun and Moore (1995), enable us to estimate not only X but also the parameters of the model. Empirical work on bond prices show that a small state space for X is better and that our model predicts yields quite well.

References

D. Duffie and R. Kan (1994), *Multi-factor term structure models*, Phil. Trans. Royal Society, London, Series A, **347**, 577-586.

D. Duffie and R. Kan (1996), *A yield factor model of interest rates*, Math. Finance **6**, 379-406.

R.J. Elliott (1994), *Exact adaptive filters for Markov chains observed in gaussian noise*, Automatica **30**, 1399-1408.

R.J. Elliott, L. Aggoun and J.B. Moore (1995), *Hidden Markov Models: Estimation and Control*, Applications of Mathematics, Vol. **29**, Springer-Verlag, New York.

Eugene F. Fama and Michael R. Gibbons (1984), *A comparison of inflation forecasts*, Journal of Monetary Economics **13**, 327-348.

M. Harrison and D. Kreps (1979), *Martingales and arbitrage in multiperiod security markets*, Jour. of Econ. Theory **20**, 381-408.

J.D. Hamilton, *Rational expectations econometric analysis of changes in regime – An investigation of the term structure of interest rates*, Journal of Economic Dynamics and Control **12** (1988), 385-423.

J.D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, New Jersey, 1994.

N.D. Pearson and T-S. Sun (1994), *Exploiting the conditional density in estimating the term structure: an application to the Cox, Ingersoll and Ross model*, Jour. of Finance **49**, 1279-1304.

G. Pye (1966), *A Markov model of term structure*, Quarterly Jour. of Economics **25**, 60-72.

P. Zipkin (1993), *Mortgages and Markov chains: a simplified evaluation model*, Management Science **39**, 683-691.

Appendix

With a state process $X_\ell, \quad \ell = 1, 2, \dots,$ having dynamics

$$X_\ell = PX_{\ell-1} + M_\ell$$

and a multivariate observation process Y with components

$$Y_\ell^i = \langle g^i, X_\ell \rangle + \langle \sigma^i, X_\ell \rangle b_\ell^i, \quad 1 \le i \le M, \quad \ell = 1, 2, \dots$$

we are, therefore, in a situation analogous to the Hidden Markov Models discussed in

Elliott (1993), (see also the book by Elliott, Aggoun and Moore (1995)). The differences are that

a) the observation process is multidimensional, as discussed in Section 7 of Elliott (1993), and

b) the observations have the form mentioned in Section 8 of Elliott (1993).

Following Elliott (1993), we recall that the analysis takes place under a probability measure $\overline{Q}$ for which the $\{y_\ell^i\}$ are i.i.d. $N(0,1)$ random variables. (Note this change of measure is a mathematical artifact and is different to the equivalent martingale measure Q which gives rise to the prices.)

In fact we suppose we have a probability measure $\overline{Q}$ on $(\Omega, \mathcal{F})$ such that under $\overline{Q}$:

a) X_ℓ, $\ell = 1, 2, \ldots,$ is a Markov chain with transition matrix P, so that

$$X_n = PX_{n-1} + M_n,$$

where $\overline{E}[M_n | \mathcal{F}_{n-1}] = 0$, (here $\overline{E}$ denotes expectation under $\overline{Q}$), and

b) y_ℓ^i, $1 \le i \le m$, $\ell = 1, 2, \ldots$ is a sequence of $N(0,1)$ i.i.d. random variables.

Write $\phi(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$ and set

$$\overline{\gamma}_\ell^i = \frac{\phi\big((y_\ell^i - \langle g^i, X_\ell\rangle)/\langle \sigma^i, X_\ell\rangle\big)}{\langle \sigma^i, X_\ell\rangle\, \phi(y_\ell^i)}$$

$$\overline{\gamma}_\ell = \prod_{i=1}^m \overline{\gamma}_\ell^i \quad \text{and}$$

$$\overline{\Lambda}_n = \prod_{\ell=1}^{n} \overline{\gamma}_\ell \quad \text{for} \quad n \geq 1.$$

Write $\{G_\ell\}$ for the complete filtration generated by the processes X and y; $\{Y_\ell\}$ will denote the complete filtration generated by y.

The probability measure Q can be defined by putting

$$\frac{dQ}{d\overline{Q}}|_{G_n} = \overline{\Lambda}_n.$$

If we define $b_\ell^i = \frac{(y_\ell^i - \langle g^i, X_\ell \rangle)}{\langle \sigma^i, X_\ell \rangle}$ it can be shown, as in Elliott (1994), that under Q the b_ℓ^i are i.i.d. $N(0,1)$ random variables, that is, under Q

$$y_\ell^i = \langle g^i, X_\ell \rangle + \langle \sigma^i, X_\ell \rangle \, b_\ell^i,$$

so the y^i give noisy observations of the state X. However, $\overline{Q}$ is an easier measure mathematically with which to work.

If $\{H_\ell\}$ is any $\{G_\ell\}$ adapted sequence we write

$$\sigma_n(H_n) = \overline{E}[\overline{\Lambda}_n H_n | Y_n].$$

A version of Bayes' Theorem (see [4]) implies

$$\begin{aligned} E[H_\ell | Y_\ell] &= \frac{\overline{E}[\overline{\Lambda}_\ell H_\ell | Y_\ell]}{\overline{E}[\overline{\Lambda}_\ell | Y_\ell]} \\ &= \frac{\sigma_\ell(H_\ell)}{\sigma_\ell(1)}\,, \quad \text{say,} \end{aligned} \tag{A.1}$$

where $\sigma_\ell(H_\ell) = \overline{E}[\overline{\Lambda}_\ell H_\ell | Y_\ell]$ is an unnormalized conditional expectation of H_ℓ given Y_ℓ.

Write $\Gamma^i(y_n) = \prod_{j=1}^{m} \frac{\phi\big((y_n^j - g^j(i)\big)/\sigma^j(i)\big)}{\sigma^j(i)\phi(y_n^j)}$. In particular consider

$$\begin{aligned}
\sigma_n(X_n) &= \overline{E}[\overline{\Lambda}_n X_n | Y_n] \\
&= \overline{E}\Big[\overline{\Lambda}_{n-1}\overline{\gamma}_n X_n \Big(\sum_{j=1}^{N} \langle X_n, e_j\rangle\Big)|Y_n\Big] \\
&\quad \Big(\text{because} \quad \sum_{j=1}^{N} \langle X_n, e_j\rangle = 1\Big) \\
&= \sum_{j=1}^{N} \overline{E}[\overline{\Lambda}_{n-1}\Gamma^j(y_n) e_j \langle PX_{n-1} + M_n, e_j\rangle | Y_n] \\
&= \sum_{j=1}^{N} \langle P\sigma_{n-1}(X_{n-1}), e_j\rangle \Gamma^j(y_n) e_j \\
&= \Gamma(y_n) P\sigma_{n-1}(X_{n-1})
\end{aligned} \tag{A.2}$$

where $\Gamma(y_n)$ is the diagonal matrix with $(\Gamma^1(y_n), \ldots, \Gamma^N(y_n))$ on the diagonal. Suppose $N_n^{rs} = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle \langle X_\ell, e_s\rangle$; then N_n^{rs} is the number of jumps from state e_r to e_s up to time n. Similar calculations give

$$\begin{aligned}
\sigma_n(N_n^{rs} X_n) = \sum_{i=1}^{N} &\langle P\sigma_{n-1}(N_{n-1}^{rs} X_{n-1}), e_i\rangle\, \Gamma^i(y_n) e_i \\
&+ \langle \sigma_{n-1}(X_{n-1}), e_r\rangle\, \Gamma^s(y_n) p_{sr} e_s.
\end{aligned}$$

With $J_n^r = \sum_{\ell=1}^{n} \langle X_{\ell-1}, e_r\rangle$, the occupation time in e_r,

$$\begin{aligned}
\sigma_n(J_n^r X_n) = \sum_{i=1}^{N} &\langle P\sigma_{n-1}(J_{n-1}^r X_{n-1}), e_i\rangle\, \Gamma^i(y_n) e_i \\
&+ \langle \sigma_{n-1}(X_{n-1}), e_r\rangle \sum_{i=1}^{N} \Gamma^i(y_n) p_{ir} e_i.
\end{aligned}$$

With $f(y)$ a function of y and

$$G_n^r(f) = \sum_{\ell=1}^{n} \langle X_\ell, e_r\rangle\, f(y_\ell)$$

$$\sigma_n\big(G^r_n(f)X_n\big) = \sum_{i=1}^N \langle P\sigma_{n-1}\big(G^r_{n-1}(f)X_{n-1}\big), e_i\rangle\, \Gamma^i(y_n)e_i$$
$$+ \langle \sigma_{n-1}(X_{n-1}), e_r\rangle \sum_{i=1}^N \Gamma^i(y_n)p_{ir}f(y_n)e_i.$$

Now for any scalar process H_n

$$\begin{aligned}\sigma_n(H_n) &= \langle \sigma_n(H_nX_n), \mathbf{1}\rangle \\ &= \sigma_n(H_n\langle X_n, 1\rangle).\end{aligned}$$

Therefore, summing the components, the above formulae give expressions for $\sigma_n(N^{rs}_n)$, $\sigma_n(J^r_n)$ and $\sigma_n(G^r_n(f))$. As noted in Elliott (1994), we consider H_nX_n because, unlike H_n alone, closed form recursions are obtained. Also, the normalizing factor in (4.1) is

$$\sigma_n(1) = \langle \sigma_n(X_n), \mathbf{1}\rangle = \overline{E}[\overline{\Lambda}_n|Y_n].$$

Following Elliott (1994), the above estimates enable us to re-estimate the $g^i(j)$ and $\sigma^i(j)$ at time n as:

$$\begin{aligned}\widehat{g}^i(j) &= \langle g^i, e_j\rangle \\ &= -\frac{1}{\tau_i}\ \log\, \langle e^{-B\tau_i}, e_j\rangle \\ &= -\frac{1}{\tau_i}\ \log\ E\big[\exp\big(-\int_0^{\tau_i} r(X_s)ds\big)|X_0 = e_j\big] \\ &= \frac{\sigma_n\big(G^j_n(y^i)\big)}{\sigma_n(J^j_n)}\end{aligned} \tag{A.3}$$

$$\widehat{\sigma}^i(j) = \sigma_n(J^j_n)^{-1}\big[\sigma_n\big(G^j_n((y^i)^2)\big) - 2g^i(j)\sigma_n\big(G^j_n(y^i)\big) + g^i(j)^2\sigma_n(J^j_n)\big].$$

Also, the transition probabilities in the matrix $P = (p_{sr})$ can be re-estimated at time n by

$$\widehat{p}_{sr} = \sigma_n(N_n^{rs})/\sigma_n(J_n^r).$$

Consequently, if we accept the model is reasonable, (and almost any process can be approximated by a Markov chain), our algorithms give a recursive filter for the unobserved chain. This in turn allows the re-estimation of the model parameters, which include the yield prices for varying maturities. For example, with

$$\begin{aligned} F^i(X_t, t) &= E[\exp\big(-\int_t^{t+\tau_i} r(X_s)ds\big)|X_t] \\ &= \langle e^{-B\tau_i}\mathbf{1}, X_t\rangle, \end{aligned}$$

if $t = t_\ell$ so $X_{t_\ell} = X_\ell$

$$E[F^i(X_\ell, t_\ell)|Y_\ell] = \langle e^{-B\tau_i}\mathbf{1}, E[X_\ell|Y_\ell]\rangle$$

and

$$E[X_\ell|Y_\ell] = \frac{\sigma_\ell(X_\ell)}{\sigma_\ell(1)}\ .$$

Here, $\sigma_\ell(X_\ell)$ is given recursively by (A.2) and $\sigma_\ell(1) = \sigma_\ell(\langle X_\ell, \mathbf{1}\rangle)$ is the sum of the components of $\sigma_\ell(X_\ell)$. Furthermore,

$$-\frac{1}{\tau_i}\ \log\langle e^{-B\tau_i}\mathbf{1}, e_j\rangle = \widehat{g}^i(j),$$

so the j^{th} component of $e^{-B\tau_i}\mathbf{1}$ is $\exp\big(-\widehat{g}^i(j)\tau_i\big)$. The components $\widehat{g}^i(j)$ are re-estimated in our model by equation (A.3).

TABLE 1: **Descriptive Statistics for Treasury Yields**

	1/92-6/97		1/92-6/94		7/94-6/97	
Maturity	Mean	STD	Mean	STD	Mean	STD
3-month	4.30	1.01	3.27	0.50	5.12	0.35
6-month	4.46	1.02	3.42	0.47	5.29	0.37
10-year	6.61	0.70	6.48	0.79	6.71	0.61
30-year	7.02	0.57	7.07	0.61	6.97	0.54

Note: 'STD' denotes standard deviation.

TABLE 2: **Starting values and parameter estimates – 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 2$

$E[X_0]$ vector:	(0.50, 0.50)
A matrix:	$\begin{matrix} 0.50 & 0.50 \\ 0.50 & 0.50 \end{matrix}$
g matrix:	$\begin{matrix} 0.00 & 0.00 & 0.00 & 0.00 \\ -0.50 & -0.50 & -0.50 & -0.50 \end{matrix}$
σ matrix:	$\begin{matrix} 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & 0.50 & 0.50 & 0.50 \end{matrix}$
<u>After the eighteenth pass:</u>	
$E[X_{18} \mid \quad {}_{18}]$ vector:	(0.90, 0.10)
A matrix:	$\begin{matrix} 0.89 & 0.91 \\ 0.11 & 0.09 \end{matrix}$
g matrix:	$\begin{matrix} 0.05 & 0.05 & 0.07 & 0.07 \\ 0.05 & 0.05 & 0.07 & 0.07 \end{matrix}$
σ matrix:	$\begin{matrix} 0.97 & 2.13 & 5.51 & 4.24 \\ 0.89 & 1.94 & 5.01 & 3.58 \end{matrix}$
Estmated prices:	(0.99, 0.97, 0.51, 0.13)

Note: All entries in the σ matrix that was derived after the eighteenth pass are to be multiplied by 10^{-6}.

TABLE 3: **Regressions of actual yields on predicted yields – 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 2$

	Term to maturity of security			
Parameter	3-month	6-month	10-year	30-year
α	0.21	0.34	0.73	1.11
	(0.22)	(0.33)	(0.96)	(1.08)
β	0.97	0.95	0.89	0.84
	(0.05)	(0.06)	(0.14)	(0.15)
R-squared	0.97	0.93	0.71	0.65
Durbin-Watson D statistic	1.17	1.21	1.94	1.88
s	0.14	0.19	0.33	0.32

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

TABLE 4: **Starting values and parameter estimates – 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 4$

Starting values:				
$E[X_0]$ vector:	(0.25, 0.25, 0.25, 0.25)			
g matrix:	0.00	0.00	0.00	0.00
	−0.25	−0.25	−0.25	−0.25
	0.00	0.00	0.00	0.00
	−0.25	−0.25	−0.25	−0.25
σ matrix:	0.25	0.25	0.25	0.25
	0.25	0.25	0.25	0.25
	0.50	0.50	0.50	0.50
	0.50	0.50	0.50	0.50
After the eighteenth pass:				
$E[X_{18} \mid \; _{18}]$ vector:	(0.89, 0.04, 0.04, 0.04)			
A matrix:	0.88	0.90	0.92	0.91
	0.04	0.04	0.04	0.03
	0.04	0.03	0.04	0.05
	0.04	0.03	0.01	0.01
g matrix:	0.05	0.05	0.07	0.07
	0.05	0.05	0.07	0.07
	0.05	0.05	0.07	0.07
	0.05	0.05	0.07	0.07
σ matrix:	0.96	2.13	5.51	4.28
	0.84	1.84	4.73	3.34
	0.84	1.85	4.76	3.37
	0.83	1.78	4.61	3.18
Estimated prices:	(0.99, 0.97, 0.51, 0.13)			

Note: All entries in the σ matrix that was derived after the eighteenth pass are to be multiplied by 10^{-6}.

TABLE 5: **Parameter estimates – 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 6$

After the eighteenth pass:				
	0.05	0.52	0.07	0.07
	0.05	0.05	0.07	0.07
g matrix:	0.05	0.05	0.07	0.07
	0.05	0.05	0.07	0.07
	0.05	0.05	0.07	0.07
	0.05	0.05	0.07	0.07
	0.96	2.12	5.49	4.26
	0.84	1.85	4.76	3.37
σ matrix:	0.95	2.10	5.42	4.01
	0.96	2.13	5.49	4.11
	0.96	2.12	5.48	4.09
	0.85	1.84	4.77	3.33
Estimated prices:	(0.99, 0.97, 0.51, 0.13)			

Note: All entries in the σ matrix that was derived after the eighteenth pass are to be multiplied by 10^{-6}.

TABLE 6: **Regressions of actual yields on predicted yields − 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 4$

Parameter	Term to maturity of security			
	3-month	6-month	10-year	30-year
α	0.21	0.35	0.75	1.12
	(0.22)	(0.32)	(0.96)	(1.08)
β	0.97	0.94	0.89	0.84
	(0.05)	(0.06)	(0.14)	(0.15)
R-squared	0.97	0.93	0.71	0.65
Durbin-Watson D statistic	1.18	1.22	1.95	1.89
s	0.14	0.19	0.35	0.32

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

TABLE 7: **Regressions of actual yields on predicted yields – 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 6$

Parameter	Term to maturity of security 3-month	6-month	10-year	30-year
α	0.22	0.35	0.75	1.12
	(0.22)	(0.32)	(0.96)	(1.08)
β	0.97	0.94	0.89	0.84
	(0.05)	(0.06)	(0.14)	(0.15)
R-squared	0.97	0.93	0.71	0.65
Durbin-Watson D statistic	1.19	1.23	1.96	1.89
s	0.14	0.19	0.35	0.32

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.

TABLE 8: **Regressions of actual yields on predicted yields – 3- and 6-month Treasury bills and 10- and 30-year bonds;** $N = 9$

	Term to maturity of security			
Parameter	3-month	6-month	10-year	30-year
α	0.20	0.32	0.74	1.12
	(0.22)	(0.32)	(0.95)	(1.07)
β	0.97	0.95	0.89	0.84
	(0.04)	(0.06)	(0.14)	(0.15)
R-squared	0.97	0.93	0.71	0.65
Durbin-Watson D statistic	1.18	1.21	1.95	1.89
s	0.14	0.19	0.34	0.32

Note: The numbers in parentheses are the standard errors of the corresponding parameter estimates. 's' denotes the residual standard error.