Markov chains and the potential approach to modelling interest rates and exchange rates

L. C. G. Rogers & F. A. Yousaf[1]

University of Bath

Running title: The Markov chain potential approach

Department of Mathematical Sciences
University of Bath
Bath BA2 7AY, UK
phone = +44 1225 826224, fax = +44 1225 826492
e-mail = lcgr@maths.bath.ac.uk

[1]Supported by EPSRC studentship 97003358.

Markov chains and the potential approach to modelling interest rates and exchange rates

L. C. G. Rogers & F. A. Yousaf

University of Bath

Summary. The use of the state-price density as a modelling primitive in interest-rate modelling has been advocated by Constantinides (1992) and by Rogers (1997). Rogers shows how general concepts from the theory of Markov processes can be used to create many different interest-rate models, starting from a given underlying Markov process; this formulation has many advantages, conceptually and practically. In this paper, we investigate the calibration of potential models based on an underlying Markov chain. Such a simple structure offers further advantages, and appears well able to fit multi-currency yield curves and exchange rates.

JEL Classification:

AMS (1991) Subject Classification:

Keywords:

1 Introduction.

Within the mathematical finance literature, there have been several distinct classes of interest-rate model. The first historically was the family of *spot-rate models*, where one proposes a model for the evolution of the spot rate of interest under the pricing measure, and then attempts to find expressions for the prices of derivatives; the models of Vasicek [16], Cox, Ingersoll & Ross [7], Black, Derman & Toy [3] and Black & Karasinski [4] are well-known examples of this type. Next came the *whole-yield models*, starting with Ho & Lee [10] in a discrete setting, and then in the continuous setting by Babbs [1] and Heath, Jarrow & Morton [9]. Lately, there has been much interest in so-called *market models*, whose chief characteristic is the choice of some suitable numéraire process, relative to which the prices of various derivatives have some particularly tractable form; see Miltersen, Sandmann & Sondermann [12] and Brace, Gatarek & Musiela [5] for examples of such models. These three classes of models have been developed extensively; a thorough survey would be outside the aims of this paper, but we refer the reader to the excellent recent monograph of Musiela [11] for more details and references.

In amongst these, with elements in common but seemingly little noticed by the mathematical finance community at large, there was another approach, advocated by Constantinides [6] and by Rogers [14], named the *potential approach.* The key element of this approach is to view the state-price density process as the modelling primitive, and to express the prices of derivatives directly in terms of this. From one point of view, this method is based on the choice of a numéraire process, rather as in the market models, but the emphasis is very different; in the market model approach, the numéraire is taken to be something very concrete, closely related to some particular derivative of interest, and possibly to be chosen differently when dealing with another range of derivatives, whereas in the potential approach, the numéraire is something very abstract, and is viewed as something quite universal, to be used for pricing *every* interest-rate derivative. This leads to models which are typically harder to calibrate (and ease of calibration was a major reason for the development of market models), but the reward is a consistent interest-rate modelling system. As Rogers [14] emphasises, this consistency extends across many different currencies very simply; valuing cross-currency derivatives is only a little more difficult than valuing single-currency products.

To date, there has been very little work on fitting potential models to data (the paper of Rogers & Zane [15] appears to be the only study so far), and this paper is another contribution in that direction. Earlier references concentrated exclusively on the situation where the underlying Markov process was a diffusion, but in this paper we shall focus exclusively on the case where the underlying Markov process is a *finite Markov chain*. There are advantages and disadvantages to this modelling choice, which we shall discuss at length later. But for now, notice one clear advantage which comes when we are trying to price a very general derivative. European-style derivative prices are computed as an average over the statespace, so for a Markov chain, this is just a finite sum. Pricing an American-style derivative is just an optimal-stopping problem for a finite Markov chain, and provided the number of states of the chain is not too big, this will be a very simple numerical exercise. In fact, the number of states used in our calibrations was of the order of tens, so these pricing calculations are always going to be extremely fast, in contrast to many other methods.

The plan of the paper is as follows. In Section 2, we shall briefly summarise the main ideas of the potential approach, as a way of setting up our notation, and pointing out the special forms that some of the pricing expressions take in the Markov chain situation. Section 3 describes the dataset used, and discusses various issues to do with the calibration. In Section 4, we present and discuss the results of the calibration, and finally in Section 5 we draw conclusions.

2 The potential approach.

We begin by recalling the main elements of the potential approach, as set forth in Rogers [14], and making more explicit the forms they take when the underlying Markov process is a finite-statespace chain. Arbitrage-pricing theory gives the time-t

price of a contingent claim Y payable at time $T > t$ to be

$$Y_t = E\left[\exp\left(-\int_t^T r_s ds\right) Y | \mathcal{F}_t\right], \tag{2.1}$$

where $(r_t)_{t\geq 0}$ is the *spot rate of interest* process. The probability P used for the expectation is some fixed risk-neutral measure. By taking some equivalent reference measure $\tilde{P}$, we can express this price in terms of an expectation with respect to $\tilde{P}$ as

$$Y_t = \tilde{E}_t[\zeta_T Y]/\zeta_t, \tag{2.2}$$

where the *state-price density process* ζ is defined by

$$\zeta_t \equiv \exp\left(-\int_0^t r_s ds\right) \cdot \left.\frac{dP}{d\tilde{P}}\right|_{\mathcal{F}_t} \equiv \exp\left(-\int_0^t r_s ds\right) \cdot Z_t. \tag{2.3}$$

Assuming $r \geq 0$ (which we always shall), the process ζ_t is a positive supermartingale, and for *any* positive supermartingale ζ, (2.2) determines an arbitrage-free pricing system. The potential approach therefore seeks to model the state-price density process ζ with respect to the reference probability $\tilde{P}$, and computes prices using the characterisation (2.2).

One very natural way to build positive supermartingales is to take some Markov process $(X_t)_{t\geq 0}$ with resolvent $(R_\lambda)_{\lambda>0}$, fix some $\alpha > 0$, and some positive function g on the statespace of X and make an interest-rate model by setting

$$\zeta_t = e^{-\alpha t} R_\alpha g(X_t). \tag{2.4}$$

A particularly attractive feature of this modelling approach is that the spot rate process r can be expressed very simply as

$$r_t = \frac{g(X_t)}{R_\alpha g(X_t)}\ . \tag{2.5}$$

See Rogers [14], p.161 for the derivation.

In the context of a finite Markov chain X with finite statespace I and infinitesimal generator (or Q-matrix) Q, the resolvent has the simple expression

$$R_\lambda = (\lambda - Q)^{-1},$$

when we regard the transition semigroup $(P(t))_{t\geq 0}$ as a semigroup of matrices acting on the vector space $\mathbb{R}^I$, expressible in terms of Q as $P(t) = \exp(tQ)$. Thus, for example, the time-0 price of a zero-coupon bond delivering a riskless \$ 1 at time T is just

$$P(0,T) = \exp(-\alpha T) P(T)(\alpha - Q)^{-1}\mathbf{1}/R_\alpha g, \tag{2.6}$$

regarded as a function on I. Here, $\mathbf{1}$ is the vector all of whose entries are equal to 1.

A further feature of the potential approach is the ease with which yield curves in several countries can be modelled. Indeed, we can introduce another country j

without introducing any further sources of randomness, simply by taking a new positive function g^j and positive real α^j and defining the state-price density process ζ^j for country j by

$$\zeta_t^j = R_{\alpha^j} g^j(X_t).$$

As Rogers [14] shows, if Y_t^{ij} is the time-t price in currency i of one unit of currency j, then we have in general that

$$\zeta_t^i Y_t^{ij} / \zeta_t^j \equiv N_t^{ij} \tag{2.7}$$

is a $\tilde{P}$-martingale orthogonal to all the $\tilde{P}$-martingales of the form $\zeta_t^j S_t^j$, where S_t^j is a traded asset, valued in currency j. A special case of this (which we shall focus on exclusively below) is when the martingale N^{ij} is constant.

3 Discussion of the data and calibration methodology.

The data which is used in this study is daily yield curve data covering the period from 2nd January 1992 to 1st March 1996[2]

For each day we have values of the yield of bonds with maturity 1 month, 3 months, 6 months, 1 year, 2 years, 5 years, 7 years and 10 years. We shall use daily yield curve data for three currencies; these are sterling (GBP), the US dollar (USD) and the German Mark (DEM).

We also have daily exchange rate data between these three currencies, obtained from the United States Federal Reserve Data Exchange[3].

As a preliminary data-cleaning, any dates that were not common to every set were removed from all sets. This included public holidays and other days where one or more of the three markets was closed. In total we have 1029 days of data. Surface plots of the yield curve for each country, together with graphs of the exchange rates are shown in figure 1.

It is worth pointing out that the period under consideration in this study represented a turbulent time in the world markets. The years of 1992 and 1993 saw both the US and UK economies in the middle of deep recessions. Indeed, 1992 was a year of huge turmoil for the UK economy; it saw the surprise re-election of the Conservative party for a third consecutive term of office, and this was followed a few months later by the embarrassing débacle of 16th September 1992 - "Black Wednesday" - in which the UK was embarrassingly forced out of the ERM, losing 4 billion GBP trying to stop the pound devaluing. On this day, the UK government announced a 5% rise

[2]We are grateful to Dr Simon Babbs for supplying the GBP and DEM data. The USD data was taken from the website http://www.stls.frb.org/fred/index.html

[3]See http://www.federalreserve.gov/releases/H10/hist

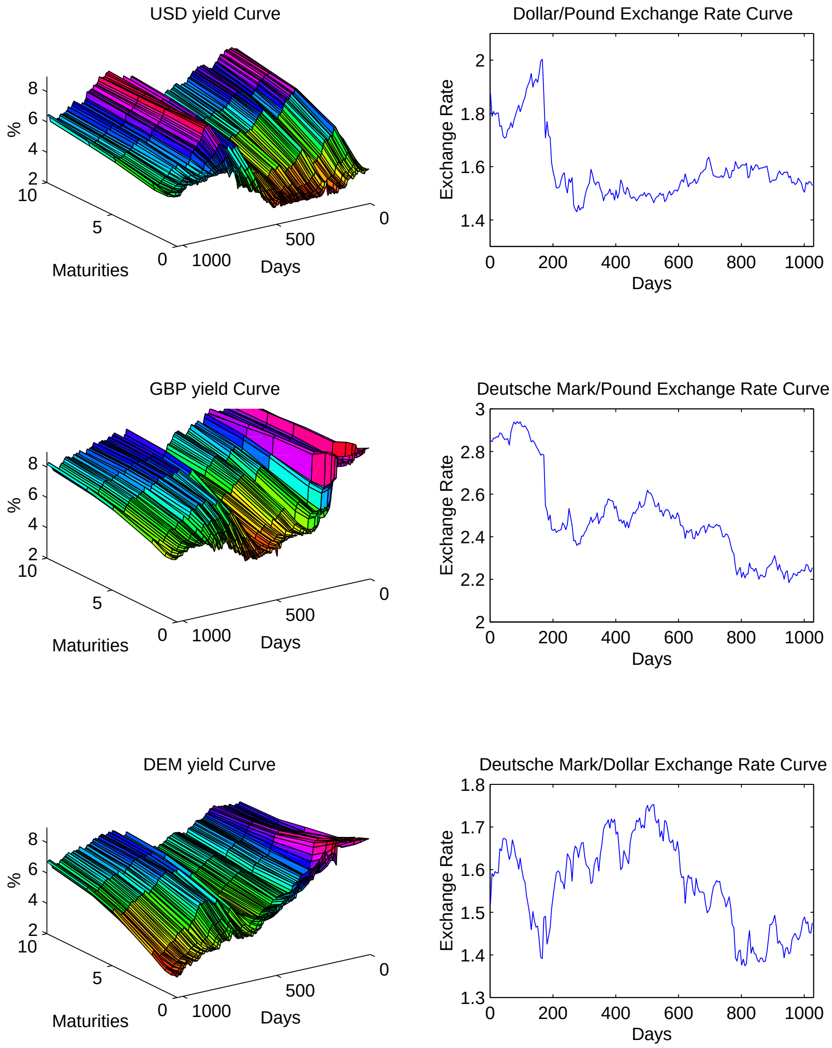

Figure 1: Yield and Exchange Rate Curves

in the base rate taking the rate to 15% in a desperate attempt to stop the pound's value sliding. The turmoil in the UK economy at this point was partly attributed (by many analysts) to the strength and dominance of the German Mark. In fact it can be seen that the German economy had a strong influence on most of the other major European economies at this time.

Conversely, 1994 to 1996 saw a weakening of the German dominance and a recovery in the UK and US economies. These countries slowly came out of their long recession and this is reflected in the shape of the yield curve and exchange rate over this period. We have therefore chosen quite a varied and turbulent period for the calibration exercise.

We shall attempt to fit the data using a potential model based on an underlying Markov chain. In any of the calibration exercises, the first step is to fix the number N of states of the chain. This done, there are in total N^2 free parameters to be estimated: $N^2 - N$ off-diagonal entries of the Q-matrix Q, $N-1$ entries[4] of g, and the one value α. However, to make the problem somewhat easier, we restricted the fitting to *reversible* chains, where for some vector m of positive entries

$$m_i q_{ij} = m_j q_{ji} \quad \text{for all } i, j.$$

Thus the flux matrix $A \equiv (m_i q_{ij})_{i,j=1,\ldots,N}$ is symmetric with zero row sums. In choosing the reversible Q-matrix, we therefore have the choice of the $N(N-1)/2$ above-diagonal entries of A, and of $N-1$ of the entries of m; the diagonal entries of A are then determined by the zero-row-sum condition, and the last entry of m is fixed by the fact that the entries of m have to sum to 1. We therefore have in total $(N^2 + 3N - 2)/2$ free parameters to estimate. By restricting to reversible Markov chains, we have thus reduced the number of parameters by about half, but the principal reason for making this restriction is that by so doing we guarantee that all the eigenvalues of the Q-matrix are *real*, thereby avoiding the need to program with complex variables throughout.

Nevertheless, it is clear that our modelling assumptions involve a large number of parameters; in our examples, we took N in the range 10 to 25, so that the number of parameters to be estimated was of the order of hundreds. Using daily yield curve data for one week, the number of parameters is far in excess of the the number of data-points. Conventional statistical wisdom would frown on such a model, for a variety of reasons:

If you have more parameters than data points, you will be able to fit the data perfectly. This is clearly false. If, for example, you wish to model real-valued observations $y_1, \ldots, y_n$ taken at increasing times $t_1, \ldots, t_n$ as

$$y_i = \sum_{j=1}^{J} \alpha_j \exp(-\beta_j t_i) + \varepsilon_i$$

[4] One degree of freedom represents a redundant scaling of g.

for non-negative parameters α_j, β_j, then however large you take J you will be unable to fit the y_i if they are not decreasing. The same is true for our application; we are trying to fit a model with very strong structural properties, and there is no guarantee that we will be able to get a perfect fit (in fact, we don't). We *need* a highly-structured model because we do not simply wish to be able to fit yield curve data, we have to be able to price general derivatives; if we had simply done a principal-components analysis of yield curve data, we would have been unable to begin to value an American swaption.

Some of the parameters will be indeterminate. There are examples (such as a two-way analysis of variance) where this does indeed happen, but this can arise even when there are far more data points than parameters. Our estimation procedure looks for the minimum of a real-valued function of many variables, and there is no reason based on the number of data points why this minimum should not be unique.

The estimates of many of the parameters will be subject to large error. Though there is no general reason why this *must* happen, we do observe this. But if we find that a particular parameter cannot be estimated with high precision, this is because it has relatively little influence on the model values for the observables, so it really does not *matter* what value it takes! What matters is how well the fitted model fits the data.

In summary, we regard such conventional statistical wisdom in this case rather as the split infinitive (see Fowler [8], p 579, who distinguishes the meanings of 'to just have heard' and 'to have just heard'); we know the objections, and shall not hesitate to completely ignore them. Our methods will be justified by the quality of the fit that they achieve, and by the stability of the estimates we come up with. It should be remarked that the finance industry routinely works with models with time-dependent coefficients, in which the parameter space is every bit as large as those we shall be dealing with here, and in which problems of parameter stability are very hard to deal with in a satisfactory manner.

To introduce the estimation methods we shall use, we now explain carefully the modelling assumptions in use. Our model is parametrised by a vector[5] θ. The underlying Markov chain X takes values in a finite set I, and on day n we have a vector y_n of observations[6]. If the model were correct, the value of this observation vector y_n would be $Y(X_n, \theta)$, but we suppose that the observed values are the true values plus some independent Gaussian noise. We adopt a Bayesian standpoint, and suppose that the initial law of X is given by $\pi = (\pi_i)_{i=1}^N$, and the initial law of θ is given by density $f_0(\theta)$; conceptually, θ is unchanging with time, even though our knowledge of it varies[7].

[5] We can think of this as the above-diagonal entries of A, the first $N-1$ entries of m, the first $N-1$ entries of g, and the value of α stacked into a single vector if we wish.

[6] The observations happen to be the yields of the different maturities, though this is irrelevant for the present discussion.

[7] We shall later consider what happens if we modify this assumption.

We shall use the notation $\mathbf{z}_n \equiv (z_0, \ldots, z_n)$ in what follows to reduce the acreage of formulae. Based on the assumptions above, and ignoring irrelevant constants, the likelihood Λ_n of $(\mathbf{X}_n, \mathbf{y}_n, \theta)$ is

$$\begin{aligned} \Lambda_n &\equiv \Lambda_n(\mathbf{X}_n, \mathbf{y}_n, \theta) \\ &= f_0(\theta)\, \pi_{X_0} \prod_{j=1}^{n} p_{X_{j-1}X_j}(s_j; \theta) \exp[-b(y_j - Y(X_j; \theta))] \end{aligned} \tag{3.1}$$

where $p_{ij}(s;\theta) = P_\theta(X_s = j | X_0 = i)$, and $b(z) \equiv \frac{1}{2} z \cdot V^{-1} z$, where V is the covariance matrix of the Gaussian errors. We have also used the notation $s_j = t_j - t_{j-1}$ for the time between the $(j-1)$th and jth observations. We shall be more interested in the posterior distribution of (X_n, θ) given $\mathbf{y}_n$, so we introduce the notation

$$L_n(x, \mathbf{y}_n, \theta) = \sum_{\mathbf{X}_n : X_n = x} \Lambda_n(\mathbf{X}_n, \mathbf{y}_n, \theta), \tag{3.2}$$

and notice that directly from the definitions

$$L_n(x, \mathbf{y}_n, \theta) = \sum_{\xi} L_{n-1}(\xi, \mathbf{y}_{n-1}, \theta) p_{\xi x}(s_n; \theta) \exp[-b(y_n - Y(x; \theta))]. \tag{3.3}$$

It is clear that for the Markov chain model in mind this expression will be far too complicated to allow exact analysis, and we shall have to make simplifying assumptions in order to make progress. Here are the simplifications which we used.

Day-by-day calibration. In this case, we simply ignore all the 'earlier' information in (3.3) and, given the observations y_n on day n, we just compute

$$\min_\theta b(y_n - Y(x; \theta)), \tag{3.4}$$

where in the minimisation we make the arbitrary convention that x is some distinguished state (say, the first) in the statespace. The labelling of the states of the chain is clearly irrelevant under this simplifying assumption. This particular method can be expected to be simple to implement, but cannot be expected to be very stable. Nevertheless, it should furnish a lower bound for the fitting error; if the results of fitting under this assumption are disappointing, then the results will be disappointing under more realistic assumptions.

Rigid calibration. In this approach, we take some initial period of K days data, and then try to fit the model using an approximation to the likelihood (3.1). This calibration is more honest than the day-by-day fit, in that it requires the parameters to be the same for all days. The simplification used is based on the observation that the underlying state of the Markov chain does not change very frequently, so we replace the true likelihood (3.2) - which involves a sum over all possible paths of the chain during the K days of the calibration period - by the single term corresponding to a path which remains at its initial state throughout the calibration period. This is a reasonable thing to do when the length of the calibration period is up to a few tens of days, during which period a change of underlying state is comparatively unlikely. Since the particular state is not important, we may as well assume that it is the first one labelled, 1, say[8] The true calibration, involving a sum over all possible paths of the underlying chain during the K days, would be far too slow. So the calibration is achieved by minimising the expression

$$-\log f_0(\theta) + \sum_{j=1}^{K} [b(y_j - Y(1;\theta)) + q_1 s_j], \tag{3.5}$$

where $-q_1$ is the diagonal entry in the first position of the Q-matrix Q.

Having found our calibrated values θ^*, we can then check the model out-of-sample by taking the days after the calibration period and trying to fit the yield curves by allowing only changes in the (posterior) distribution of X.

Conditional-independence (CI) calibration. In this case, we imagine the situation where there has been a large amount of observed data, and we postulate that

$$L_n(x, \mathbf{y}_n, \theta) = \pi_n(x, \mathbf{y}_n)\, l_n(\theta, \mathbf{y}_n). \tag{3.6}$$

The motivation for this is that we have seen so much data that we have a pretty good idea what the values of the parameters must be; the values of θ will largely be determined by the long-run historical average behaviour of the system. On the other hand, the posterior distribution of X_n will be more influenced by recent history, because of the ergodicity of the Markov chain, and so some approximate conditional independence is reasonable; recent history tells us all we can know of X_n, distant history tells us all we can know of θ. We shall further assume that

$$l_n(\theta, \mathbf{y}_n) \propto \exp(-\frac{1}{2}(\theta - \hat{\theta}_n) \cdot S_n(\theta - \hat{\theta}_n)) \tag{3.7}$$

for some positive-definite symmetric matrix S_n. If we think that we have nearly identified the true value of θ, then such a quadratic approximation to the likelihood is quite natural.

The values $\hat{\theta}_n$, S_n, and $\pi_n(\cdot, \mathbf{y}_n)$ are computed recursively, using the assumed form (3.6) of the likelihood. Supposing that we know already $\hat{\theta}_{n-1}$, S_{n-1}, and $\pi_{n-1}(\cdot, \mathbf{y}_{n-1})$, returning to (3.3) and using (3.6) we see that

$$\begin{aligned} L_n(x, \mathbf{y}_n, \theta) &= \sum_{\xi} \pi_{n-1}(\xi, \mathbf{y}_{n-1})\, l_{n-1}(\theta, \mathbf{y}_{n-1}) p_{\xi x}(s_n; \theta) \exp[-b(y_n - Y(x;\theta))] \\ &\propto \sum_{\xi} \pi_{n-1}(\xi, \mathbf{y}_{n-1})\, p_{\xi x}(s_n; \theta) \exp[-b(y_n - Y(x;\theta))] \\ &\qquad . \exp[-\frac{1}{2}(\theta - \hat{\theta}_{n-1}) \cdot S_{n-1}(\theta - \hat{\theta}_{n-1}))] \end{aligned} \tag{3.8}$$

[8] An extension of this fitting would be to allow the chain just one jump during the K days.

We now sum this expression over x, and numerically pick θ to maximise; the maximising value is our new estimate $\hat{\theta}_n$ of θ. By computing the second derivative matrix with respect to θ at $\hat{\theta}_n$ we find the value of S_n,[9] and finally we compute π_n by

$$\pi_n(x, \mathbf{y}_n) \propto \sum_{\xi} \pi_{n-1}(\xi, \mathbf{y}_{n-1}) p_{\xi x}(s_n; \hat{\theta}_n) \exp[-b(y_n - Y(x; \hat{\theta}_n))].$$

Properly speaking, the posterior distribution π_n for X_n should be obtained by integrating the likelihood (3.8) with respect to θ, but we approximate this by assuming that the posterior distribution for θ can be replaced by the point mass at $\hat{\theta}_n$, to avoid the need to integrate over a large number of dimensions.

Random walk (RW) calibration. This method is very similar to the previous method, which can be seen as a special case. The theoretical justification is explained in Appendix A in more detail, and is based on the Kalman filter. The idea is that we shall now allow the value of θ to change from day to day according to a random walk. If the variance of the steps of the random walk is zero, then we arrive at the CI method, but if we allow the variance of the random walk step to be a fixed multiple of the posterior covariance of θ, then we obtain

$$\sum_{\xi} \pi_{n-1}(\xi, \mathbf{y}_{n-1})\, p_{\xi x}(s_n; \theta) \exp[-b(y_n - Y(x; \theta)) - \frac{\beta}{2}(\theta - \hat{\theta}_{n-1}) \cdot S_{n-1}(\theta - \hat{\theta}_{n-1}))], \tag{3.9}$$

where $\beta \in (0, 1)$ is fixed. The closer β is to 1, the closer we are to the CI fit. In the CI calibration, we expect that the matrices S_n will be growing approximately linearly with n, by analogy with the situation where we attempt to estimate the mean of a Gaussian distribution using a sequence of noisy observations of the mean; when we have seen $n-1$ observations, the nth receives weight $1/n$ in the estimation. The same thing happens with our CI calibration, so the most recent observations get relatively little weight in relation to the average over earlier times. On the other hand, we do not believe that there is *no* change in the interest-rate environment, and by introducing the parameter β, we allow the new day's observations to have the same importance in the estimation as yesterday's new observations did yesterday; the analogy is with the estimation of an underlying random walk process based on noisy observations of that process.

The last three approaches to calibration are (quasi-)Bayesian and produce estimates of the posterior distribution π_n of the underlying Markov chain X_n at time n, as well as point (ML) estimates $\hat{\theta}_n$ of the parameter θ. Thus to price a derivative on day n, we shall use the expression

$$\sum_{x} \pi_n(x, \mathbf{y}_n) F(x, \hat{\theta}_n), \tag{3.10}$$

where $F(x, \theta)$ is the price which the Markov chain potential model would produce if the starting state were x and the true parameter value were θ. This would apply,

[9] In practice, we compute only the diagonal terms of S_n

for example, to the pricing of zero-coupon bonds; so, in particular, we end up with a continuum of possible yield curves at any given time, even though the model with known θ could only produce one yield curve for each possible state of the Markov chain.

4 Numerical results.

The heart of the calibration procedure is a minimisation routine, and for this we chose the NAG routine E04JYF. Of several which we investigated, this one seemed to do the best job. Our first fitting attempt was a day-by-day calibration; we do not of course believe in this approach, but if the results of this fit were poor, then it would be impossible that a more realistic fitting procedure will produce anything other than poor results. For purposes of comparision, we split the dataset into 19 overlapping blocks of 100 days, and computed summary statistics, which we present in Table I. The data was GBP data, and we used 15 states in the Markov chain. Here we took the covariance matrix V, in (3.4), to be the identity matrix.

Perhaps the most interesting figures in this table are in the *Median* column. These present the median values of the sum of absolute errors in basis points for each day's fit. This sum consists of 8 terms, one for each maturity, so the basis-point error per maturity is 1/8 of the figure given in the *Median* column. The worst values are in the turbulent months of 1992, when the median error per maturity is 2bp, but for most of the periods under study, the error is 1bp or even a lot less. Even looking at the upper quartile, we find that only in three of the 19 periods did the error exceed 2bp per maturity. For more detailed analysis, we chose to use an 11 state Markov chain and focus on period 14 which contains two base rate changes occuring on 7th December 1994 (day 43) and 2nd February 1995 (day 79). The plots in Figure 2 refer to this period and show the stability of the parameters g_i and α, as well as the contributions of different maturities to the total residual error. We normalised the g values to sum to one, so as to remove the degree of indeterminacy and all maturities were weighted equally. The parameters exhibit no particular stability, which is not a surprise, but what is encouraging about these fits is that the errors are *small*; the median fit per maturity is consistently below 2 bp, and the upper quartile is below 4 bp, often a lot less. A model that is fitting yields to within a basis point is good enough to trade off, and we are here getting close to that degree of precision, without any particular effort, and with relatively few states.

The next fitting exercise we carried out was the rigid calibration, which one would expect to be quite poor in comparison with the day-by-day fit, and indeed it was. Working again with the GBP data, and taking a chain with 11 states, we used five consecutive days of data to calibrate the model, and then stepped ahead through the next 100 days (period 14) computing the fit each day. So at the end of the five-day calibration period, we have found a value θ^* for the parameter θ, and for subsequent days we hold this value fixed, but use the data to update the posterior distribution for X_n by the recipe

$$\pi_n(x, \mathbf{y}_n) \propto \sum_{\xi} \pi_{n-1}(\xi, \mathbf{y}_{n-1}) p_{\xi x}(s_n; \theta^*) \exp[-b(y_n - Y(x, \theta^*))].$$

The bond prices were then computed following (3.10). It is inconceivable that in practice one would fit a model to just five days' data and then run with that unaltered for the next 100 days, and the results of this fitting procedure, presented in Table III and in Figures 3 and 4, show why. These Figures and Table show also the results of variants of the rigid calibration, where we recalibrate the model every J days, using again the latest $K = 5$ days of data. The panels in Figures 3 and 4, correspond to $J = 100$, $J = 10$ and $J = 1$. For the case $J = 100$, we see from Table III that the median error in bp per maturity is of the order of 35, which is really quite useless. Notice that Figure 3 shows how the quality of fit deteriorates as we get further into the 100-day period, as one would expect. The fits for the case $J = 1$ are a lot better, but even here the median error is three times the worst that occurred in Table III, amounting to around 6bp per maturity.

This calibration is poor not only because of the rigidity imposed by the assumptions, but also because we have trained the model on just 5 consecutive days' data. Since this tiny calibration set cannot possibly represent the variety of yield curves that might arise, it is not surprising that as time rolls forward we encounter days where the yield curve is far from the possibilities of the 5 day calibration period, and so the fit is very poor. A better recipe might be to take the last 5 Mondays for our calibration set. The problem with this is that the assumption that the underlying state has not changed in this time becomes untenable, and we would have to evaluate a sum over all possible paths of the chain during this calibration period, and this would be slow and clumsy. We do need to have more influence of past data in our calibration method, but the obvious way to do this is via some recursive approach, and this was what we tried next.

The next fitting exercise was an implementation of the CI/RW fitting strategy (3.8) and (3.9); since the CI fitting is the special case $\beta = 1$ of the RW fitting, it makes sense to consider them all together. We started with S_0 equal to the identity, $\hat{\theta}_0$ equal to zero, and the prior distribution for X to be uniform over the 11 states. The data used was period 14 of the GBP data. Table IV shows summary statistics for the fits. Taking $\beta = 1$, we obtained a median error of just over 5bp per maturity, already better than even the one-day-ahead form of the rigid fit, and with $\beta = 0.2$ - allowing a random step with 4 times the posterior covariance - we obtained a median error of 2.5 bp per maturity, with the upper quartile at a little over 3 bp per maturity. Figures 5 and 6 display various results of the fitting procedure: notice the quite impressive stability of the g_i for the $\beta = 0.2$ case (compare with Figure 2). This justifies empirically the (at first sight) low value of β; although we have in principle allowed the random walk a lot of freedom to move, it turns out in practice that it is not moving very much.

It appears therefore that the CI/RW fitting methodology represents a good compromise between the unstable but close fitting day-by-day approach, and the very stable but poorly-fitting rigid approach. Moving on to the simultaneous fitting of yield curves in more than one country, and the exchange rate(s), we concentrated on the CI/RW calibration approach. The first fitting exercise we carried out was using USD and GBP data from the period 5th October 1994 to 6th March 1995, with 11 states in the Markov chain; we report summary statistics for these in Table V, with various diagnostics displayed in Figures 7 and 8. The fit was noticeably poorer than the single-country fit, as one would expect; for $\beta = 0.2$ we found a median fitting error of 3.5-4.5 bp per maturity We then moved on to fit three currencies, USD, GBP and DEM, summarising the results in Table VI, with diagnostics displayed in Figure 9. The inclusion of Germany worsens the fit of the US and UK very slightly, but with $\beta = 0.2$ we are still finding median errors of 3.5-4.5 bp per maturity.

The final fitting study we carried out was to include exchange rate data. Once again, we took USD and GBP data from 5th October 1994 to 6th March 1995, with 11 states in the Markov chain; we report summary statistics for these in Table VII, with various diagnostics displayed in Figures 10, 11, 12 and 13 By including the exchange rate in the calculation, we worsen the fit of the yield curves by about 1 bp per maturity at $\beta = 0.2$. The fit of the exchange rate is very good, mostly within about 0.5 bp. We tried to trade off the quality of the fit of the exchange rate and the fit of the yield curves, by attaching more weight to poorly fitting yields, but it seemed impossible to improve the fit of the yield curves very much by this. Rogers & Zane [15] found a similar behaviour. In view of the fact that we were fitting the exchange rate much better than the yield curves, it seems that the assumption made at (2.7) that the martingale N^{ij} is constant is relatively harmless; taking something more general would give greater flexibility to fit the exchange rate, but that is not where we appear to need the flexibility.

	Calendar Period	Day numbers	BRC	Statistics of day-by-day calibration, all values in bp						
				Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1	17th Feb 1992 - 15th July 1992	30-129	1	18.538	6.453	10.076	13.796	16.565	21.385	35.305
2	5th May 1992 - 25th Sept 1992	80-179	2	19.015	8.334	9.395	13.768	16.543	21.672	69.703
3	16th July 1992 - 7th Dec 1992	130-229	3	18.196	26.629	0.001	6.939	14.672	21.268	226.958
4	28th Sept 1992 - 18th Feb 1993	180-279	3	10.238	6.147	1.70	5.773	9.44	13.117	26.423
5	8th Dec 1992 - 7th May 1993	230-329	1	9.677	4.77	0.975	6.107	9.443	12.852	21.472
6	19th Feb 1993- 19th July 1993	280-379	0	6.678	4.613	0.001	2.94	6.422	8.992	23.569
7	10th May 1993 - 28th Sept 1993	330-429	0	5.261	3.382	0.128	2.874	4.5	7.631	15.704
8	20th July 1993 - 8th Dec 1993	380-479	1	5.397	3.488	0.055	2.692	4.463	8.144	15.510
9	29th Sept 1993 - 23rd Feb 1994	430-529	2	6.310	3.949	0.12	3.358	5.181	9.322	16.944
10	9th Dec 1993 - 11th May 1994	480-579	1	7.248	4.738	0.039	4.099	6.833	9.466	31.231
11	24th Feb 1994 - 22 July 1994	530-629	0	9.649	6.118	0.164	5.39	8.263	13.457	30.662
12	12th May 1994 - 4th Oct 1994	580-679	1	10.005	5.32	0.443	6.337	9.482	13.508	27.236
13	25th July 1994 - 16th Dec 1994	630-729	2	7.02	4.188	0.558	3.646	6.328	10.086	18.910
14	5th Oct 1994 - 6th Mar 1995	680-779	2	8.402	4.03	1.036	5.273	8.463	11.09	21.629
15	19th Dec 1994 - 23rd May 1995	730-829	1	8.437	2.608	0.77	6.666	8.243	10.081	15.573
16	7th Mar 1995 - 3rd Aug 1995	780-879	0	7.846	3.370	0.795	5.78	8.132	9.687	17.092
17	24th May 1995 - 16th Oct 1995	830-929	0	4.524	3.286	0.081	2.152	3.733	5.952	16.660
18	4th Aug 1995 - 28th Dec 1995	880-979	1	2.586	1.844	0.001	1.058	2.231	3.576	8.969
19	17th Oct 1995 - 8th Mar 1996	930-1029	3	5.503	3.436	0.001	2.387	4.956	8.552	14.726

Table I: Results of fits for the day-by-day calibration using the 19 sample periods of 100 days. Note that these results were obtained using a 15-state data imposed underlying Markov chain. The column marked 'mean' above refers to the average basis points (bp) error between model and observed values per day. The standard deviation is that of the daily basis point error in the period. Q1 and Q3 denote the first and third quartiles. Min, Median, Max again refer to the basis point error per day. BRC is used to denote the number of Bank of England base rate changes.

Diagnostic plots for day-by-day calibration

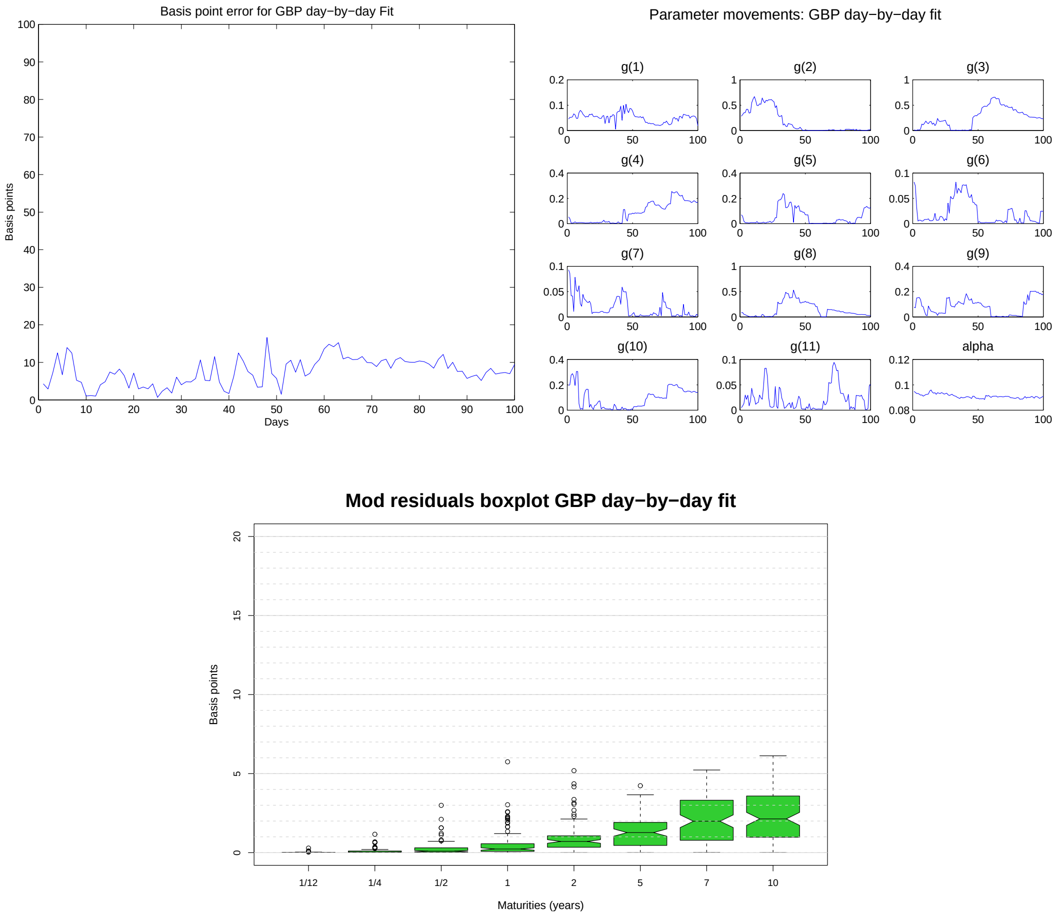

Figure 2: Diagnostic plots for the day-by-day calibration using period 14 GBP data and an 11-state Markov chain. The basis point error plot (top left) shows the total error, given in basis points, between the market and model yield curves for each fitted day. The evolution of the parameters **g** and α over the whole fitting period are given in the top right plot. Finally we give a series of boxplots showing the mean and quartiles of the mod residuals for each maturity.

Day-by-day calibration statistics (all values are in basis points)							
Data	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
Period 14	7.585	3.588	0.693	4.842	7.302	10.400	16.619

Table II: Summary statistics for the day-by-day calibration using an 11-state Markov chain on period 14 GBP data.

Rigid calibration statistics (all values are in basis points) 5 day calibration period							
Re-calibrate After	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
100 days	257.09	101.87	42.24	175.79	270.46	349.67	416.87
50 days	140.04	68.08	34.50	90.40	124.44	180.32	313.13
25 days	105.78	54.04	21.64	63.58	95.44	136.31	246.78
10 days	94.12	53.45	19.73	49.82	82.93	128.65	232.33
5 days	76.17	45.65	19.71	41.97	63.46	99.50	200.14
2 days	60.11	38.63	7.34	30.51	48.62	74.47	192.79
1 day	55.34	33.99	7.22	29.46	47.81	66.60	162.49

Table III: Summary statistics for the rigid calibration procedure. The results are for fits over a 100 day period using different re-calibration intervals. All results are for an 11-state chain using GBP data.

Daily basis point error plots for the rigid calibration

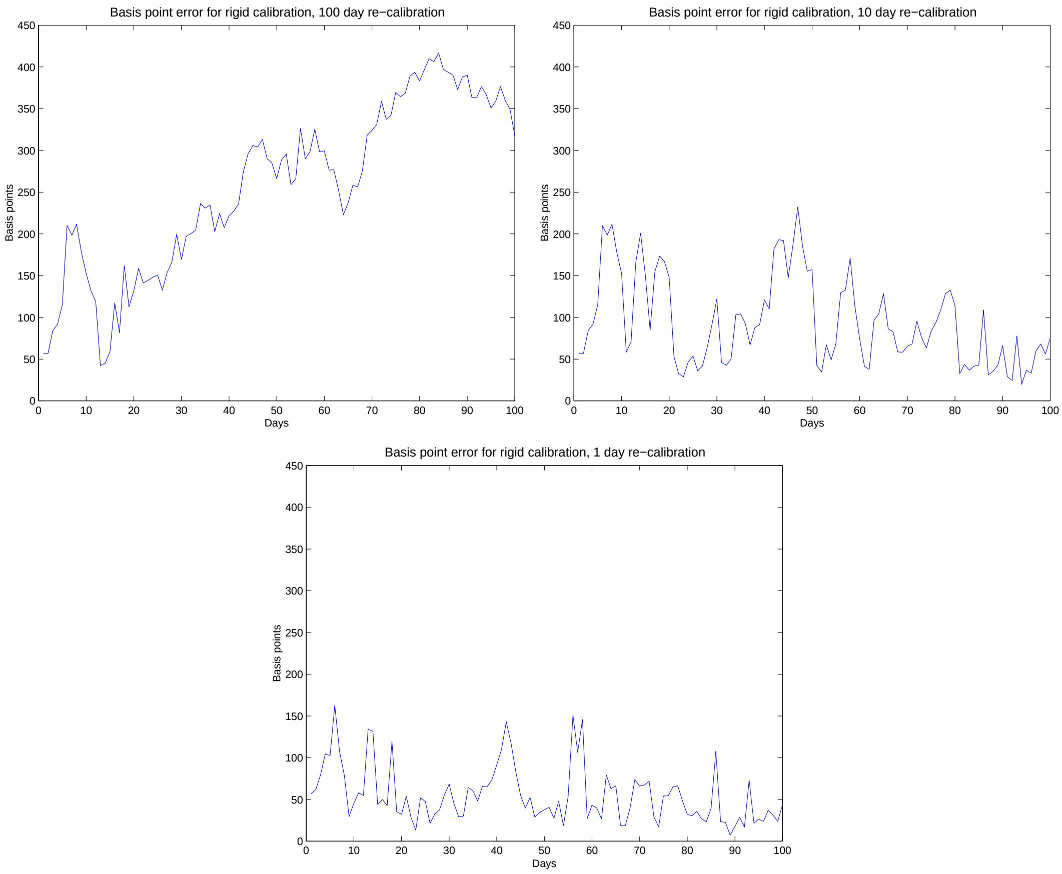

Figure 3: 'Basis point error' plots showing the cumulative error in basis points for each day over the 100 day fitting period, recalibrating after 100 days, 10 days, and 1 day.

Boxplots of mod residuals for the rigid calibration

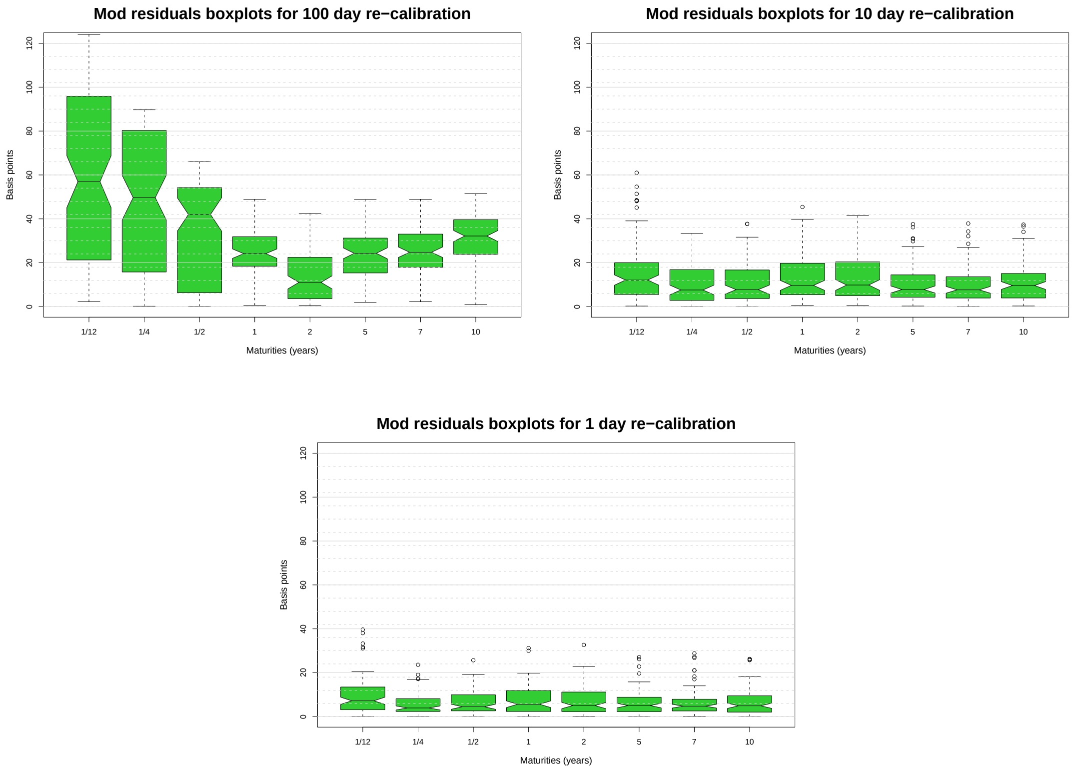

Figure 4: Three boxplots of the mod residuals for each maturity for the 100 day, 10 day and 1 day recalibration.

CI/RW GBP calibration statistics (all values are in basis points)							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	47.936	22.253	8.441	30.155	41.755	65.786	107.383
0.8	34.995	14.942	8.432	23.944	34.272	45.493	78.581
0.6	26.398	11.243	7.734	18.004	25.72	34.134	62.514
0.4	23.957	10.172	5.751	15.95	23.216	30.261	53.292
0.2	21.346	9.48	5.462	13.35	20.529	26.704	45.744
0.1	20.339	9.31	4.917	13.062	18.942	26.088	45.509

Table IV: This table contains summary statistics relating to the one country (GBP) CI/RW fits. Note that the case $\beta = 1.0$ corresponds to the CI calibration. These are daily fits on period 14 for varying values of the RW parameter β.

Diagnostic plots for GBP CI calibration

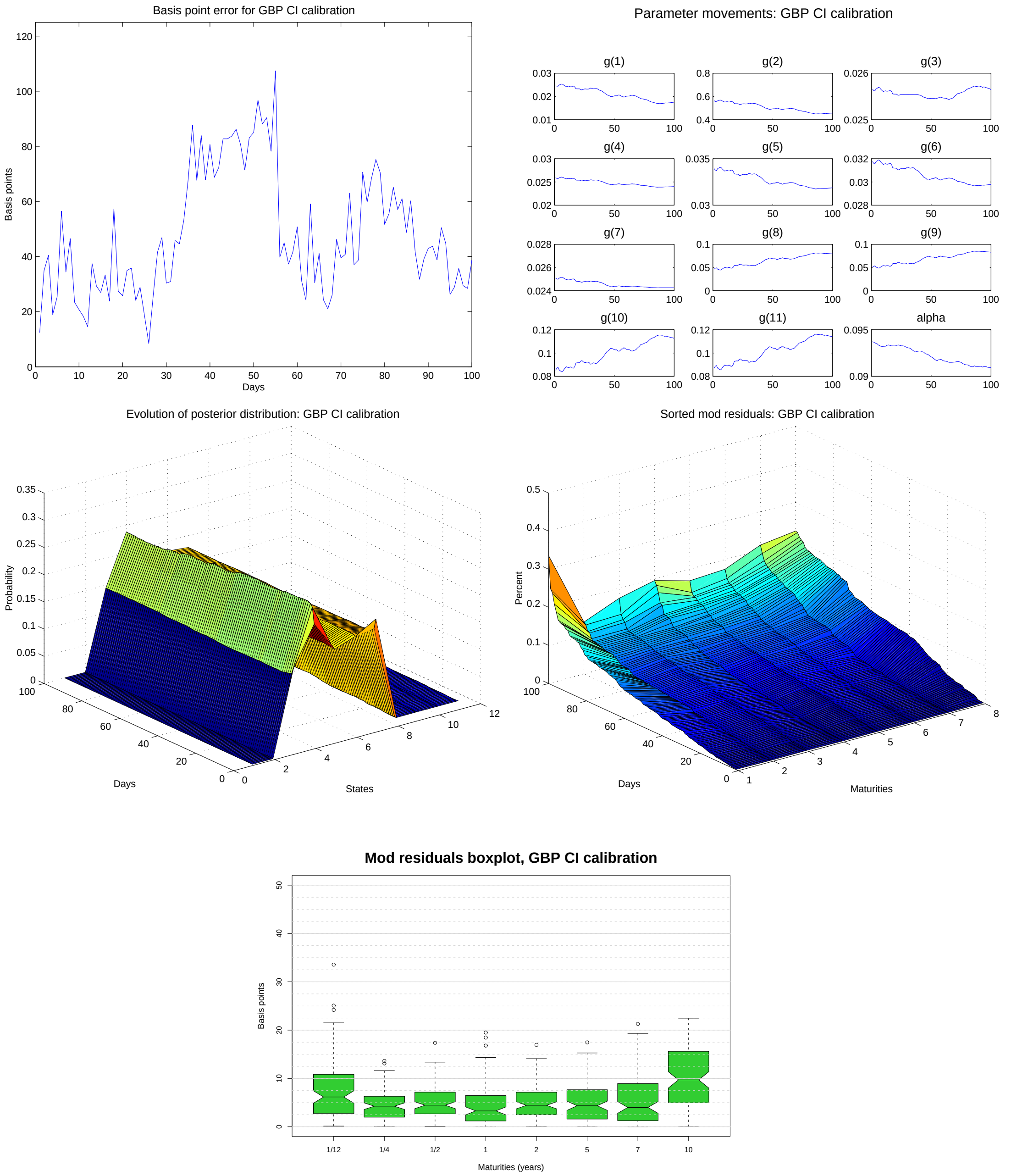

Figure 5: These plots relate to the one country (GBP) CI calibration for period 14. The ‘Parameter change’ plot shows how the **g** vector and α scalar change over the 100 day fitting period. We give a surface plot which shows the evolution of the posterior distribution over the 100 day fit. We also show the characteristics of the residuals in the ’Sorted mod residual’ and the boxplots.

Diagnostic plots for GBP RW calibration ($\beta = 0.2$)

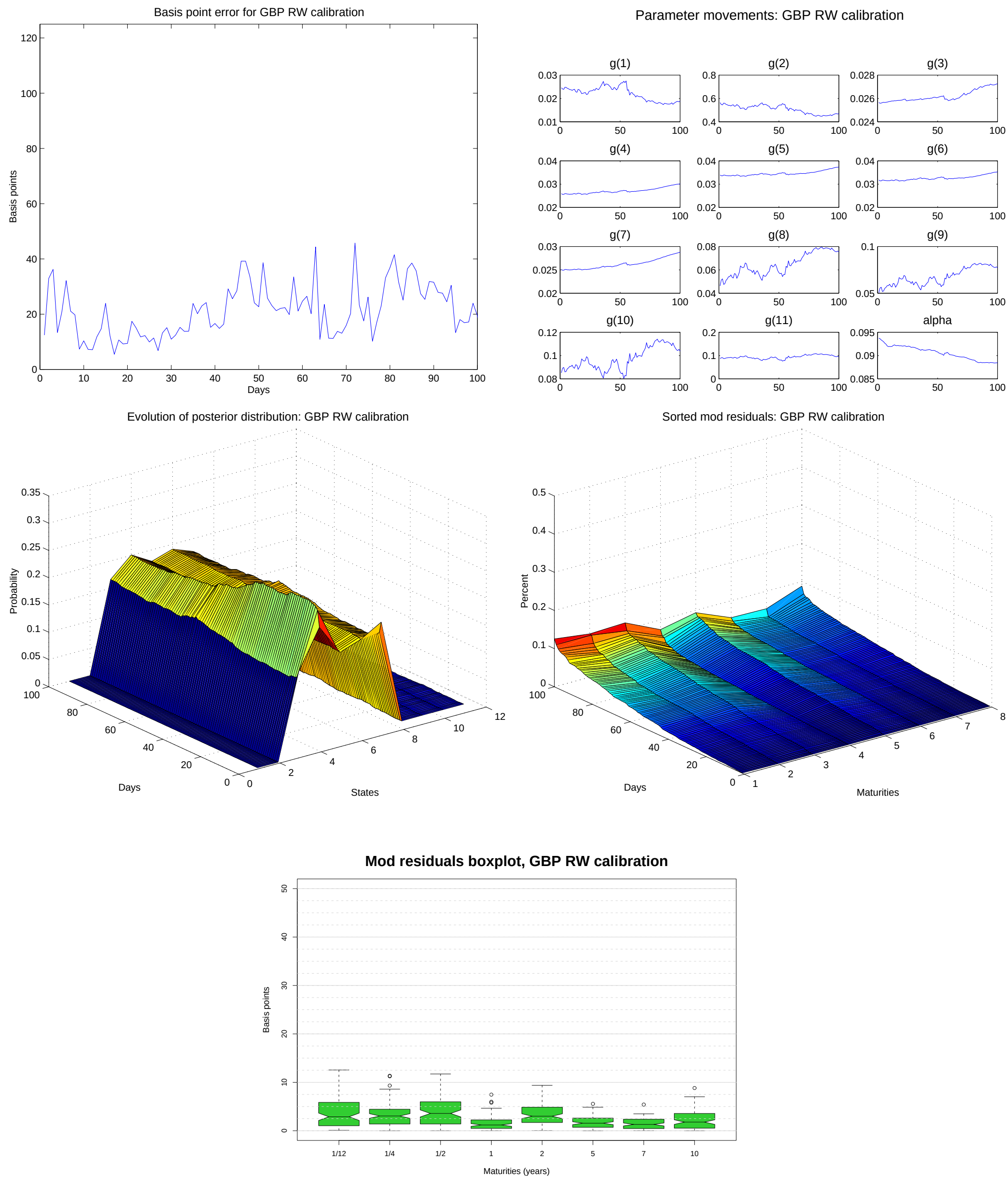

Figure 6: These plots relate to the one country CI calibration described in Case A. The 'Basis point error' plot shows the cumulative error in basis points for each day over the 100 day fitting period. The 'Parameter change' plot shows how the **g** vector and α scalar change over the 100 day fitting period. We give a surface plot which shows the evolution of the posterior distribution over the 100 day fit. We also show the characteristics of the residuals in the 'Sorted mod residual' and the boxplots.

CI/RW calibration statistics for two country fit
(all values are in basis points)

USD FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	93.082	36.671	28.606	60.205	87.108	127.584	168.642
0.8	59.183	19.086	16.899	45.524	55.416	71.192	123.935
0.6	47.947	13.160	18.402	38.282	47.458	54.798	91.380
0.4	43.583	11.195	16.978	36.625	42.664	50.295	74.498
0.2	38.153	11.564	15.619	30.589	37.258	45.501	73.653
0.1	36.775	11.586	15.335	29.052	35.939	44.924	73.549

GBP FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	62.968	26.815	9.552	41.592	62.022	81.679	128.058
0.8	37.816	13.280	9.292	30.283	37.566	46.950	76.318
0.6	33.551	12.923	8.444	24.170	33.806	41.459	76.651
0.4	31.264	12.261	12.330	22.219	28.115	38.620	74.929
0.2	29.184	12.738	6.061	19.819	27.085	38.287	73.427
0.1	28.338	12.902	7.086	19.283	26.448	37.447	72.702

Table V: Summary statistics for the two country CI/RW fits, period 14 of GBP and USD. This table gives breakdowns for the GBP and USD fits individually.

Diagnostic plots for two country (USD & GBP) CI calibration

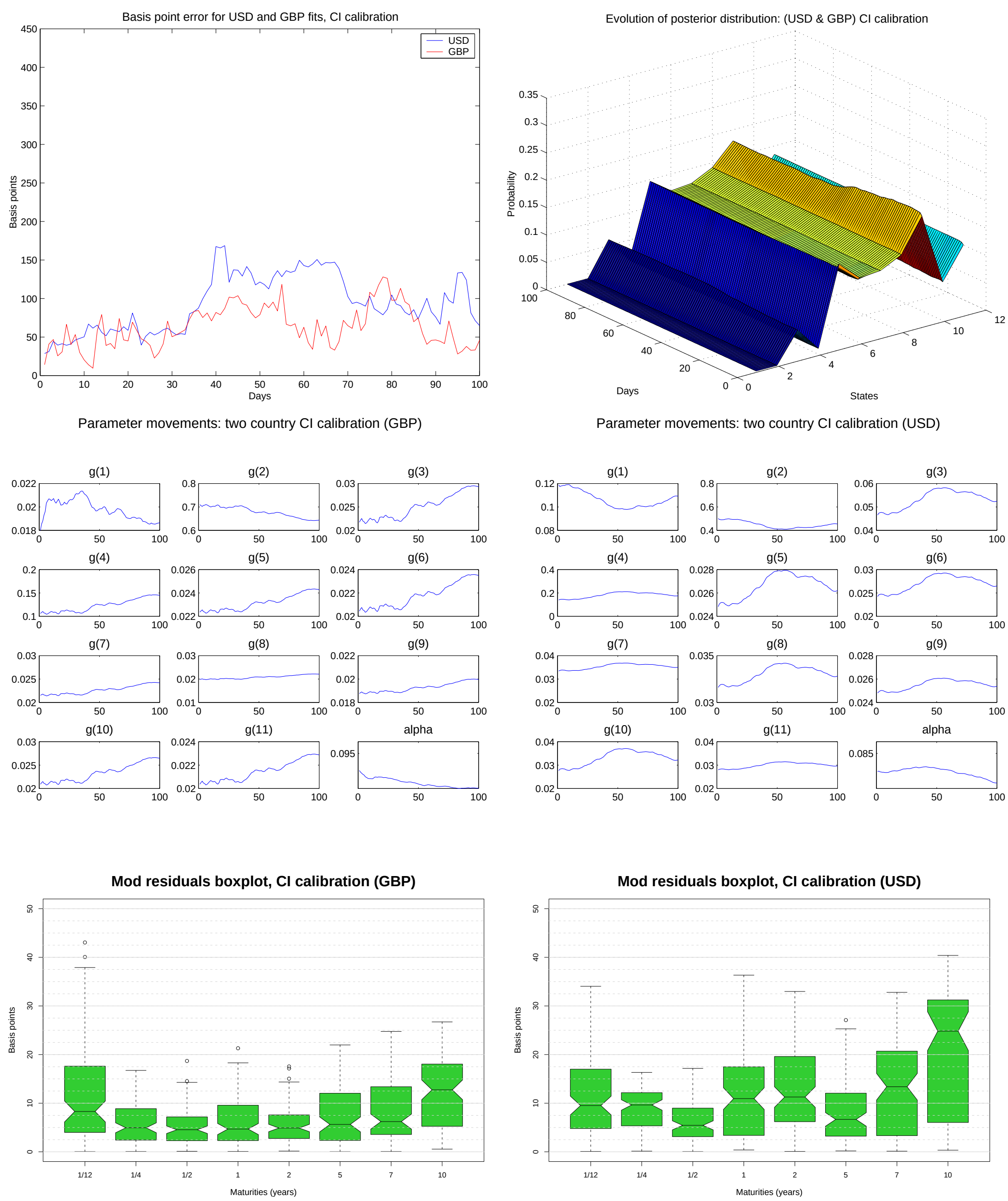

Figure 7: These plots refer to the two country fits for the CI calibration, USD and GBP, period 14. In this figure we show the basis point error plots (top left) for both the USD and the GBP. The worst fit (blue) is the USD.

Diagnostic plots for two country (USD & GBP) RW calibration ($\beta = 0.2$)

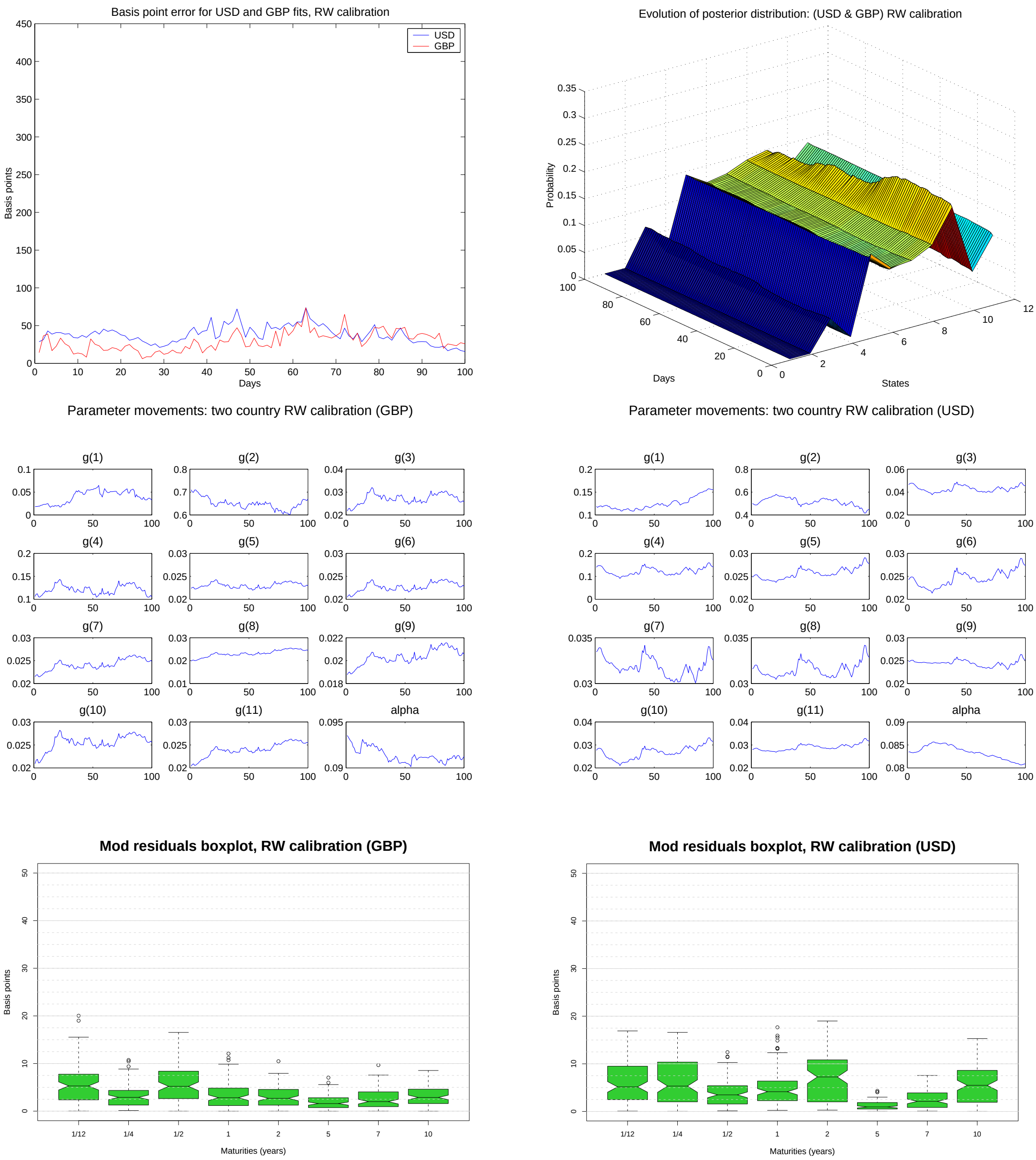

Figure 8: These plots refer to the two country fits for the RW calibration, USD and GBP, period 14. In this figure we show the basis point error plots (top left) for both the USD and the GBP. The worst fit (blue) is the USD.

CI/RW calibration statistics for the three country fit
(all values are in basis points)

USD FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	115.678	32.023	48.308	100.543	115.006	136.413	211.331
0.8	86.260	22.713	34.222	66.194	86.097	106.046	130.637
0.6	67.065	20.279	14.589	52.759	63.658	83.099	113.933
0.4	50.983	14.494	23.901	41.129	50.661	57.949	89.969
0.2	39.511	10.414	21.605	30.209	39.640	47.340	67.369
0.1	35.153	10.444	17.487	27.883	33.912	42.666	66.016

GBP FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	84.958	29.853	28.229	60.984	81.629	107.327	149.090
0.8	45.772	13.036	12.617	36.644	46.403	52.012	82.245
0.6	39.031	13.808	12.529	31.140	37.188	46.512	87.466
0.4	35.006	12.430	13.078	27.435	33.425	40.875	76.725
0.2	31.140	10.532	12.523	24.456	29.408	37.529	70.204
0.1	29.123	10.458	7.467	22.943	27.858	35.956	67.502

DEM FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	66.798	21.886	32.260	50.378	62.351	79.501	132.468
0.8	49.857	10.955	30.902	41.828	47.704	56.204	86.001
0.6	45.070	9.925	28.391	38.174	43.194	50.179	74.346
0.4	40.557	8.014	25.880	35.359	40.172	45.085	72.847
0.2	37.050	6.360	23.124	32.613	37.530	41.188	59.138
0.1	35.640	5.704	23.380	32.013	34.881	39.556	55.702

Table VI: Summary statistics for the 100 day, three country CI/RW fits on period 14. This table gives breakdowns for USD, GBP and DEM fits.

Diagnostic plots for three country (USD, GBP & DEM) RW calibration ($\beta = 0.2$)

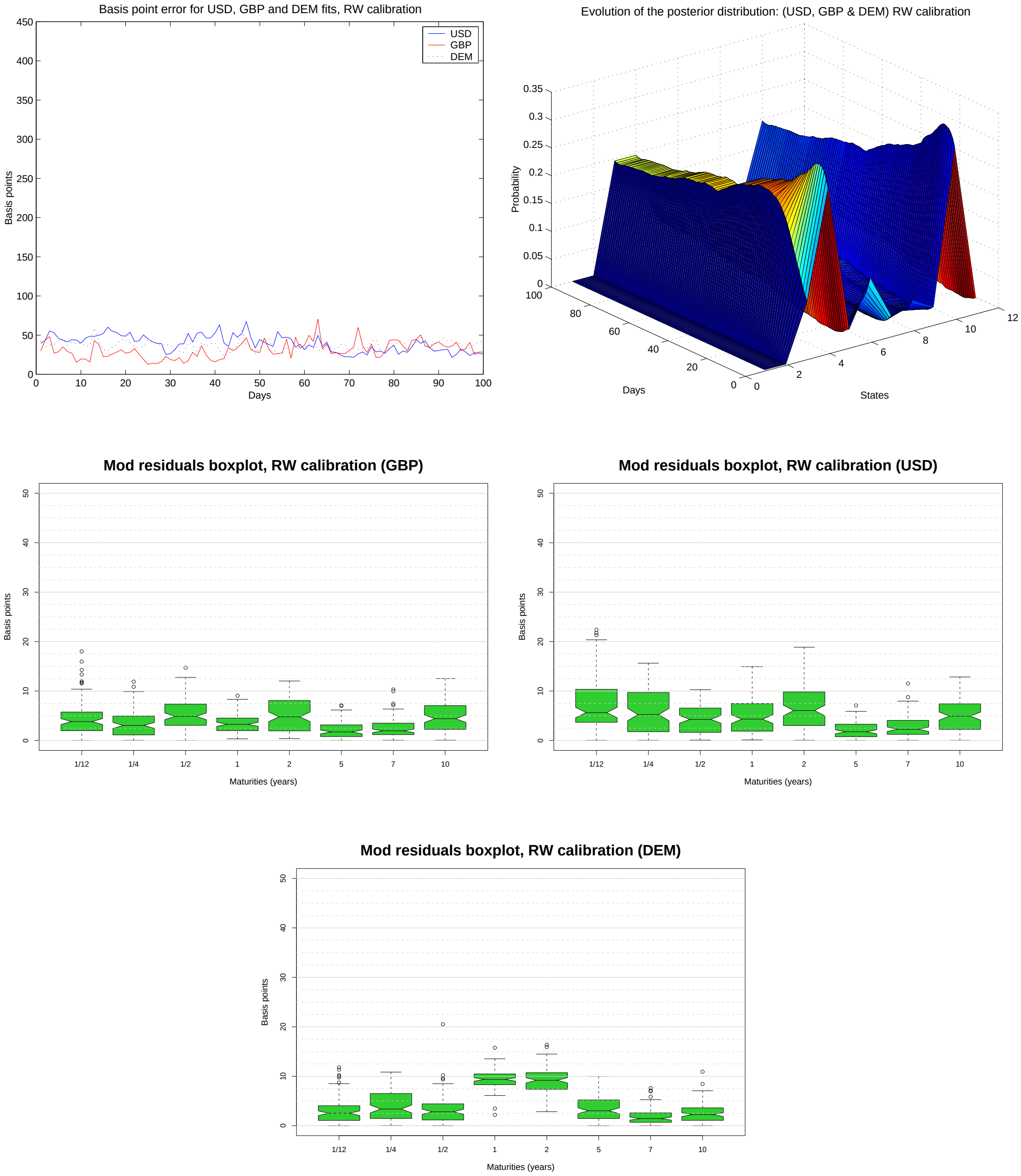

Figure 9: These plots are for a three country fit using the RW calibration method with $\beta = 0.2$. The first plot shows the basis point error for each of the three countries. The worst fit (blue line) is achieved by the USD and the the best fit (red line) is the GBP. The second plot (top right) shows the evolution in the posterior distribution during the fitting process. The boxplots are of the mod residuals for each maturity.

CI/RW calibration statistics for two country and exchange rate fit
(all values are in basis points)

USD FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	146.245	54.529	38.552	101.450	158.334	190.084	310.810
0.8	82.139	29.645	35.550	60.063	78.209	105.595	144.969
0.6	72.696	29.291	30.794	51.712	62.452	98.970	138.709
0.4	58.471	20.738	21.819	43.719	52.419	71.981	115.074
0.2	49.979	16.727	19.442	37.926	47.491	60.373	96.033
0.1	42.505	12.762	5.670	34.353	42.572	48.860	82.313

GBP FIT							
β	Mean	Std. Dev.	Min	Q1	Median	Q3	Max
1.0 (CI)	103.685	41.776	16.543	72.962	101.316	129.850	211.00
0.8	48.495	17.366	10.098	39.003	47.939	58.941	86.957
0.6	42.928	14.936	13.318	32.262	40.831	53.962	83.303
0.4	32.557	12.410	10.285	22.825	32.933	40.758	74.011
0.2	35.910	12.144	13.076	27.088	34.235	43.552	79.165
0.1	30.402	11.581	10.049	21.721	29.326	36.612	73.550

Table VII: Summary statistics for the two country and exchange rate CI/RW fits over 100 days (period 14). This table has breakdowns for the USD and GBP fitting errors.

Diagnostic plots for two-country (USD & GBP) CI calibration with exchange rates

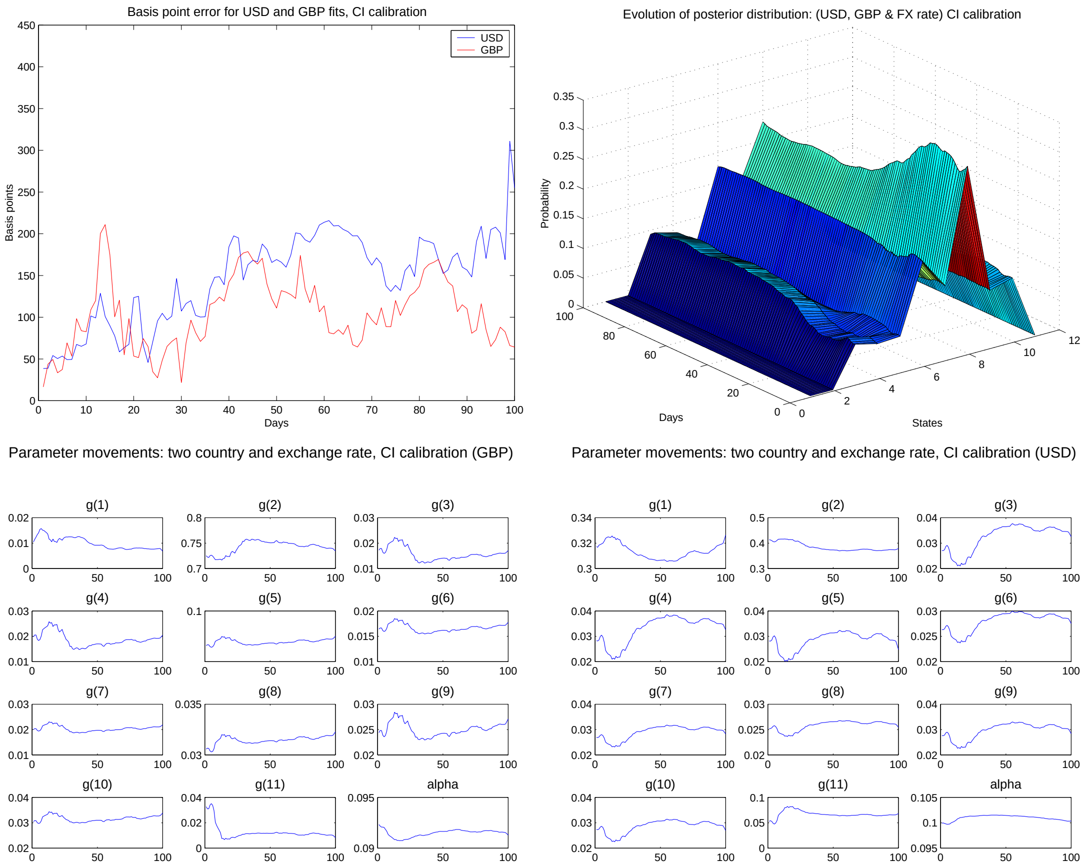

Figure 10: These plots refer to the two country and exchange rate fits for the CI calibration, period 14, over a 100 day period using USD and GBP data. In this figure we show the basis point error plots (top left) for both the USD and the GBP, the worst fit (blue) is the USD.

Diagnostic plots for two-country (USD & GBP) CI calibration with exchange rates

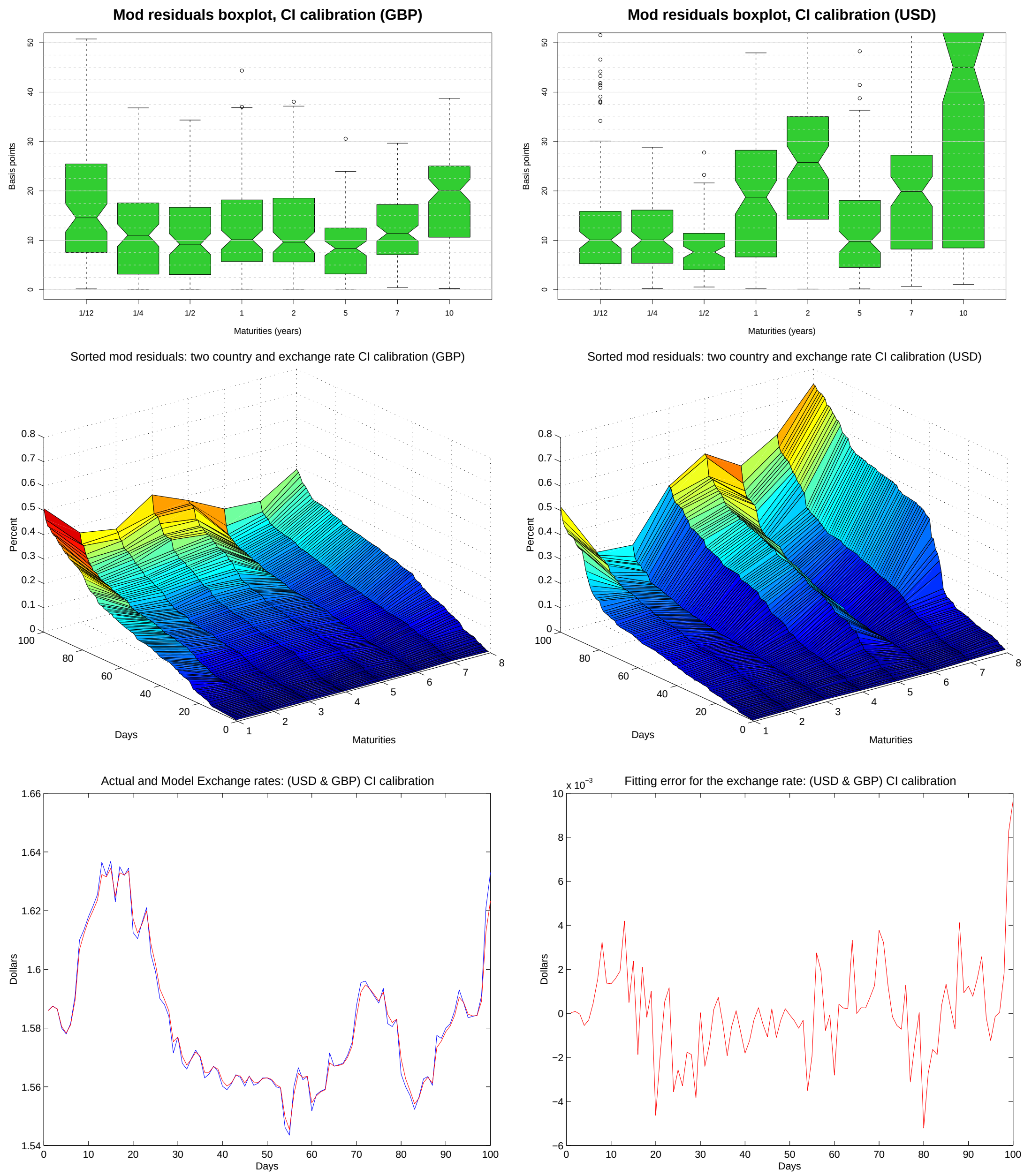

Figure 11: These plots refer to the two country and exchange rate fits for the CI calibration of the USD and GBP, period 14 data. The penultimate plot in this figure shows the observed data and the fitted curve for the exchange rates (there are two curves in this picture). The final plot is of the fitting error in the exchange rate.

Diagnostic plots for two-country (USD & GBP) RW calibration with exchange rates ($\beta = 0.2$)

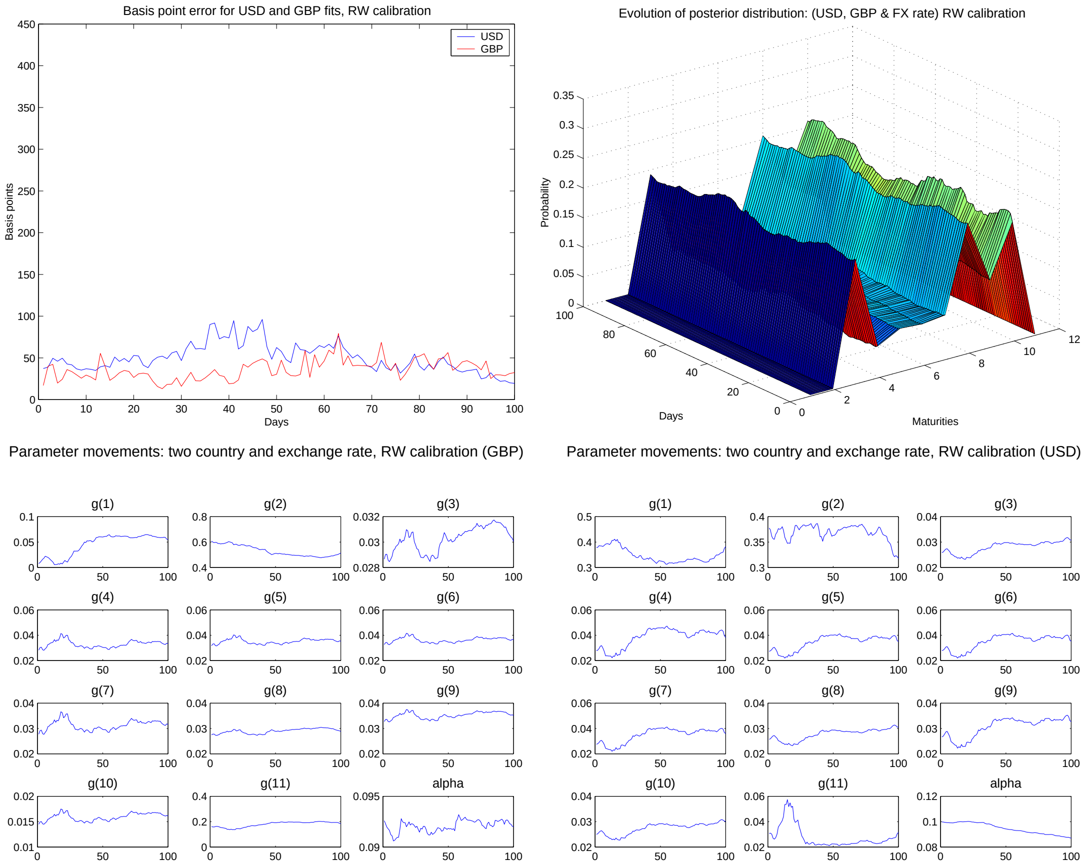

Figure 12: These plots refer to the two country and exchange rate fits for the CI calibration, period 14, over a 100 day period using USD and GBP data. In this figure we show the basis point error plots (top left) for both the USD and the GBP, the worst fit (blue) is the USD.

Diagnostic plots for two-country (USD & GBP) RW calibration with exchange rates ($\beta = 0.2$)

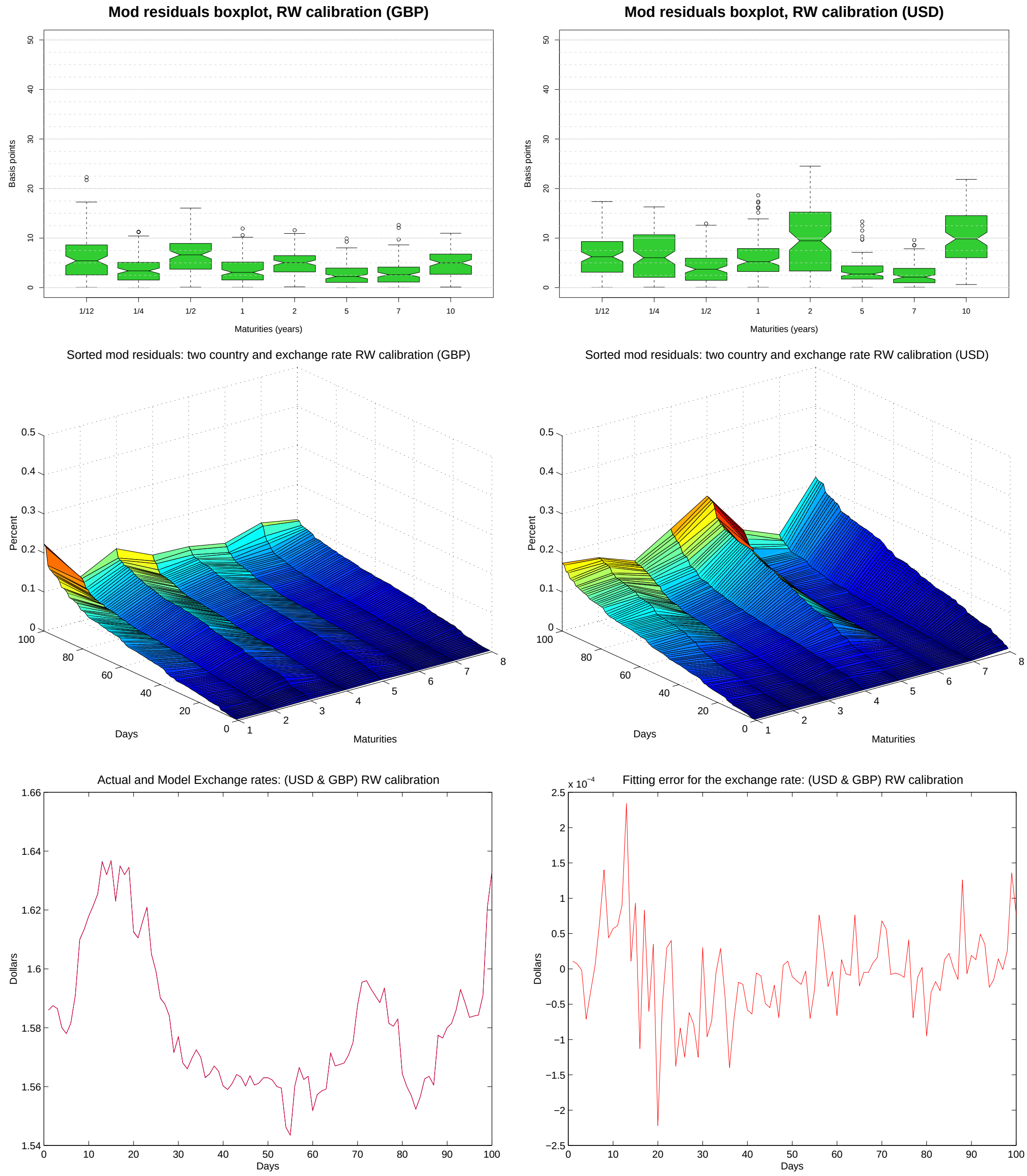

Figure 13: These plots refer to the two country and exchange rate fits for the RW calibration of the USD and GBP, period 14 data. The penultimate plot in this figure shows the observed data and the fitted curve for the exchange rates (there are two curves in this picture). The final plot is of the fitting error in the exchange rate.

5 Conclusions.

In this study, we have carried out a number of calibration exercises for potential models of interest rates based on an underlying Markov chain. At a theoretical level, such models offer persuasive advantages:

- the approach generates a model to account for all derivatives;

- pricing of a European-style derivative is simply a sum over a (typically small) finite number of states, and pricing of an American-style derivative is an optimal stopping problem for a Markov chain with (typically few) states;

- adding a new country can be done without complicating the underlying Markov process;

- exchange rates are modelled within the same modelling framework as interest rates.

What we have done here is by way of a pilot study, to investigate the feasibility of this approach. Most of the fitting runs were done using only 11 states of the Markov chain, and we were insisting on fitting a time-homogeneous model, both very stringent requirements which would undoubtedly be abandoned in practice. If we allowed a different model to be fitted each day, we were able to come up with fits of the yield curve in one country with median errors of the order of 1bp per maturity; sometimes more, sometimes less. At the other end of the scale, by calibrating to 5 days' data and then using the calibrated model to fit the next day, we were coming up with median errors of the order of 6 bp per maturity, which is too high to be much use. By taking a fitting methodology in between these two extremes, we were able to produce one-country fits with median errors of around 2.5 bp per maturity, with good parameter stability.

Incorporating more than one country inevitably worsened the fit; when we fitted USD and GBP data, we came up with median errors of the order of 3.5-4.5 bp per maturity, and including DEM as well increased the errors very slightly. However, including the exchange rate in the USD/GBP fitting exercise worsened the median fit by about 1 bp per maturity, which would lead to quite significant mispricing.

Given the restrictions to time-homomogeneous chains with no more than 11 states, the fits we have come up with are very encouraging. There are obvious extensions which could be carried out, and some will be the subject of a later study. For example, we could simply increase the number of states. Since the calibration procedure was quite lengthy on the machine[10] available to us (of the order of 200 CPU minutes to fit a single country, of the order of 300 CPU minutes to fit two countries), we preferred to investigate a larger number of relatively small problems, rather than try a few huge fits. Another obvious place where the modelling could be extended would

[10] A Sun Ultra E3500 with 400 MHz UltraSPARC II processor

be by dropping the reversibility criterion; this requires code which can cope with complex eigenvalues, but the principles are the same. Another extension would be to allow time-dependent Markov processes. In some sense, this is completely trivial; as Rogers [13] remarked (p 101), if we take a time-homogeneous model and apply a deterministic time-change, we can exactly fit any given initial yield curve! However, this is really far too easy, and we need to be aware of the model changing completely in a day, always a problem with time-inhomogeneous models.

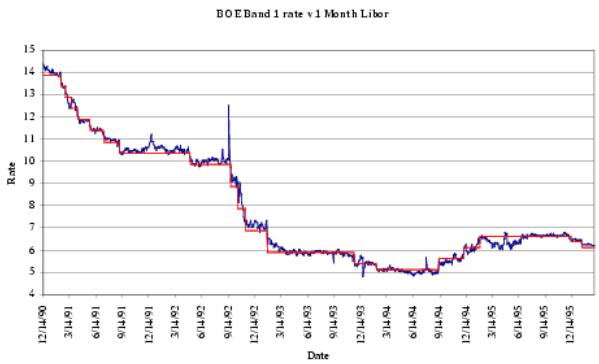

Figure 14: Base Rate against 1 Month LIBOR

The relative success of such a primitive probabilistic model is either to be expected, or something quite remarkable, depending on your point of view. On the grounds that there are many parameters, it might be thought to be expected; but our earlier comments show that large parameter spaces are not in themselves guarantors of a close fit. To be able to fit more than one yield curve reasonably closely using nothing more sophisticated than an 11 state chain does seem to us to be remarkable. In Figure 14, we present a plot of 1-month LIBOR and Bank of England band one stop rate. The agreement is evident, and the conclusion unavoidable: if we were able to model the band one rate, we would already have a good model for 1-month LIBOR! Now the band one rate is a jump process, taking relatively few values. It is not fanciful to imagine that *this* could be well modelled by a Markov chain with a small number of states. Indeed, looking at Figure 14, the interpretation of 1-month LIBOR as a noisy observation of the band one rate seems quite natural, and the very interesting paper of Babbs & Webber [2] uses elements of this interpretation in its modelling. In short, focusing on the volatilities of various yields and rates may actually be concentrating on the *noise* in the system, and overlooking the *signal!*

Appendix A

For ease of reference, we summarise here the Kalman filter argument which we used as the basis of the fitting procedures of the earlier parts of the paper. To begin with, suppose we have a pair of discrete-time vector processes θ and Y evolving according to the dynamic linear model

$$\begin{aligned} \theta_n &= \theta_{n-1} + \varepsilon_n, & \text{(A.1)} \\ Y_n &= C\theta_n + \eta_n, & \text{(A.2)} \end{aligned}$$

where the ε are independent $N(0,Q)$ and the η are independent $N(0,R)$ random variables. If $\mathcal{Y}_n$ denotes the σ-field generated by $\{Y_k : k \leq n\}$, and if we have that conditional on $\mathcal{Y}_n$ the law of θ_n is $N(\hat{\theta}_n, V_n)$, then

$$\left(\begin{matrix} \theta_{n+1} \\ Y_{n+1} \end{matrix} \middle| \, \mathcal{Y}_n \right) \sim N\left(\begin{pmatrix} \hat{\theta}_n \\ C\hat{\theta}_n \end{pmatrix}, \begin{pmatrix} Q+V_n & (Q+V_n)C^T \\ C(Q+V_n) & R+C(Q+V_n)C^T \end{pmatrix} \right), \tag{A.3}$$

and likewise

$$\left(\begin{matrix} \theta_{n+1} \\ Y_{n+1} - C\theta_{n+1} \end{matrix} \middle| \, \mathcal{Y}_n \right) \sim N\left(\begin{pmatrix} \hat{\theta}_n \\ 0 \end{pmatrix}, \begin{pmatrix} Q+V_n & 0 \\ 0 & R \end{pmatrix} \right). \tag{A.4}$$

It is an easy though tedious exercise to confirm from (A.3) that the law of θ_{n+1} given $\mathcal{Y}_{n+1}$ is $N(\hat{\theta}_{n+1}, V_{n+1})$, where $\hat{\theta}_{n+1}$ is the value of θ maximising the joint density of the distribution (A.3) or equivalently (A.4):

$$\exp[\, -\frac{1}{2}(\theta - \hat{\theta}_n)^T (Q+V_n)^{-1} (\theta - \hat{\theta}_n) - \frac{1}{2}(y - C\theta)^T R^{-1} (y - C\theta) \,], \tag{A.5}$$

and $-V_{n+1}^{-1}$ is the second derivative of the log-likelihood with respect to θ. The actual estimation problem we face has non-linear dynamics, but we shall suppose that a local linear approximation is adequate, so we replace (A.2) with

$$Y_n = Y(x_n, \theta_n) + \eta_n,$$

giving the analogue of (A.5) to be

$$\exp[\, -\frac{1}{2}(\theta - \hat{\theta}_n)^T (Q+V_n)^{-1} (\theta - \hat{\theta}_n) - \frac{1}{2}(y - Y(x_n,\theta))^T R^{-1} (y - Y(x_n,\theta)) \,]. \tag{A.6}$$

If we have $Q = 0$, the full CI fitting assumption, we find that (A.6) reduces to the exponential terms in (3.8), and if we take $Q = (\beta^{-1} - 1)V_n$, we obtain exactly the exponential terms in (3.9).

References

[1] S. H. Babbs, *A family of Itô process models for the term structure of interest rates*, University of Warwick, preprint 90/24, 1990.

[2] S. H. Babbs and N. Webber, *Term Structure Modelling Under Alternative Official Regimes*, Mathematics of Derivative Securities, Hardback ISBN 0521-5842-48 (1997), pp 394-422

[3] F. Black, E. Derman and W. Toy, *A One-factor Model of Interest Rates and its Application to Treasury Bond Options*, Financial Analysts Journal, **46** (1990), pp 33- 39

[4] F. Black and P. Karasinski, *Bond and Option Pricing when short rates are lognormal*, i Journal of Financial Analysts, **47** (1991), pp 52-59

[5] A. Brace and M. Musiela, *A Multifactor Gauss Markov Implementation of Heath, Jarrow and Morton*, Mathematical Finance, **4** (1994), pp 259-283

[6] G. M. Constantinides, *A Theory of the Nominal Term Structure of Interest Rates*, Review of Financial Studies, **5** (1992) pp 531-552

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

[8] H. W. Fowler, *A Dictionary of Modern English Usage*, 2nd Edition Oxford University Press (revised by E. Gowers), (1968)

[9] D. C. Heath, R. A. Jarrow, & A. Morton, *Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation*, Journal of Financial Quantitative Analysis, **25** (1990), pp 419-440

[10] T. S. Y. Ho and S. B. Lee, *Term Structure Movements and Pricing Interest Rate Contingent Claims*, Journal of Finance, **41** (1986), pp 1011-1029

[11] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer-Verlag Berlin, ISBN 354061477X (1997)

[12] K. R. Miltersen, K Sandmann and D. Sondermann, *Closed form solutions for term structure derivatives with log-normal interest rates*, Journal of Finance, xi **52** (1997), pp 409-430

[13] L. C. G. Rogers, *Which Model for the Term-Structure of Interest Rates Should One Use?*, Mathematical Finance, **65** (1995), pp 93-116

[14] L. C. G. Rogers, *The Potential Approach to the Term-Structure of Interest Rates and Foreign Exchange Rates*, Mathematical Finance, **7** (1997), pp 157-176

[15] L. C. G. Rogers and O. Zane, *Fitting Potential Models to Interest Rate and Foreign Exchange Data*, Vasicek and Beyond (book), ISBN 1899-3325-02 (1996), pp 327-342

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