University of Alberta

The Term Structure of Interest Rates in a Continuous Markov Setting

by

Craig Armstrong Wilson ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of *Doctor of Philosophy*

in

Finance

Faculty of Business

Edmonton, Alberta

Spring 2004

Your file Votre référence
ISBN: 0-612-96336-5
Our file Notre référence
ISBN: 0-612-96336-5

When Fortune in her shift and change of mood
Spurns down her late beloved, all his dependants
Which labour'd after him to the mountain's top
Even on their knees and hands, let him slip down,
Not one accompanying his declining foot.

—*William Shakespeare, Timon of Athens, Act I, scene i*

As mem'ry slips through fickle Fortune's mane
So too it fades from noble Markov's chain

To my mother and father
Thank you for your support

Abstract

Two interest rate models are described, in which randomness in the short-term interest rate is due to a Markov chain. We model randomness through the mean-reverting level, which is supposed to switch from time to time, according to the state of a Markov chain. The short rate is modeled as a continuous stochastic process in continuous time. The first model we propose for the short rate has no diffusion term, so it changes smoothly through time with differentiable sample paths. The randomness comes from the drift term which depends on the Markov chain. The smooth sample path property is an attempt to model the central bank's desire to maintain a certain amount of predictability in the interest rate. We obtain several results about the term structure implied by this basic short rate model, including a technique for calibrating the model to a given initial yield curve. We extend the model to incorporate a diffusion term driven by a Brownian motion that is independent with the Markov chain. This change confounds the observation of the Markov chain with noise, and leads to a hidden Markov model. The Markov chain still controls the mean-reverting level of the short rate, but we also allow the possibility of a stochastic volatility parameter that modulates with the Markov chain, and governs the diffusion term. We again obtain results on the properties of longer-term rates that are implied by this short rate model. We then develop an

algorithm to filter the Markov chain's state and find maximum likelihood estimators for the parameters. This algorithm is based on the expectation maximisation algorithm, and it provides an exact, finite-dimensional, adaptive estimation method. Applying this algorithm to a time series of three-month Canadian Treasury-bill rates provides parameter estimates that are economically sensible.
Key Words: Interest Rate Modeling, Term Structure, Markov Chain, Hidden Markov Model

Contents

List of Tables

List of Figures

List of Symbols

$(\Omega, \mathcal{F}, P)$ The underlying probability space

$\{\mathcal{F}_t\}$ The filtration of sub-σ-fields

$\{X_t\}$ The Markov chain

S $= \{e_1, \ldots e_N\}$, the state space for the Markov chain

N The number of states

A The transition rate matrix

$\{M_t\}$ The martingale component of the Markov chain

$\{W_t\}$ A standard Brownian motion

$\{r_t\}$ The short-term interest rate process

a The speed of mean reversion parameter

$\bar{r}$ The level of mean reversion parameter

$B_t(T)$ The price at time t of a zero-coupon bond with time of maturity T

$R_t(\tau)$ The yield at time t of a zero-coupon bond with time until maturity τ

Chapter 1

Introduction

The risk-free interest rate is one of the most vital inputs in financial theory. There is still much debate on the appropriate term of the risk-free rate, (long, short, or medium), and about the relationship between rates for differing time horizons. Relationships between spot and forward rates and subsequently between long- and short-term rates were first explored by Irving Fisher [16], whose work was extended into the modern expectations hypothesis by Lutz [34]. This expectations hypothesis failed to account for greater interest rate risks associated with longer-term bonds, and Hicks [24] argued that the expectations obtained should be modified by a premium that increases with the term. This lead to the liquidity preference hypothesis. Modigliani and Sutch [35] argued that the relevant interest rate risk depended on the holding period of the borrower or lender, and thus the liquidity premium need not be increasing with the term. This resulted in the market segmentation or preferred habitat hypothesis. Of course, under the risk-neutral probability measure, no risk premiums are required, so the three hypotheses can be considered three formulations about how the measure is changed from the real to the risk-neutral world. Details of this can be found in Musiela [36] and Musiela and Rutkowski [37].

No matter which hypothesis is adopted, it does seem intuitive that longer-term rates, (and hence bond prices), should depend upon current and future short-term rates in some way. Similar reasoning also applies to other interest rate derivatives, such as options, futures and swaps. Given the magnitude of these markets and its importance to other aspects of the economy, it seems vital to continue working to improve our understanding of short-term interest rates and our ability to predict and control them.

Current models of the short-term interest rate often involve treating the short rate as a diffusion or jump diffusion process in which the drift term involves exponential decay toward some value. The basic models of this type are Vasiček [41] and Cox, Ingersoll and Ross [6], where the distinction between these two interest rate models rests with the diffusion term. The drift term, (of both models), tends to cause the short rate process to decay exponentially towards a *constant* level. This feature is responsible for the mean-reverting property exhibited by these processes.

An extension to these models has come in the form of allowing the drift to incorporate exponential decay toward a *manifold*, rather than a constant. This is known as the Hull and White [27] model, and it allows the short rate process the tendency to follow the initial term structure of interest rates. This is an important extension, because with a judicious choice of the manifold, the initial term structure predicted by the model can exactly match the existing term structure, and because of this feature, models of this class are called no arbitrage models. In general, this cannot be done with a constant mean-reverting level, and such models are often called equilibrium models, since they generate stationary interest rate processes. Although there are many other extensions to the basic models—incorporating stochastic volatility, non-linear drift (so decay is no longer exponential), and jumps, for example—the Hull-White extension is the most applicable to the bond pricing component of our study.

The Hull-White model has many advantages: it possesses a closed-form solution for the price of zero-coupon bonds, as well as for call options on such bonds, and it can also be calibrated to fit the initial yield curve exactly. However, one of the disadvantages of the model is that, because there is only one factor of randomness, it only allows parallel shifts in the yield curve through time. Bonds of all maturities are necessarily perfectly correlated with each other. This approach cannot explain the common phenomenon of yield curve twists. This motivates the need to incorporate an additional factor of randomness into the basic model.

The Hull-White model is described under the risk-neutral probability by the stochastic differential equation

$$dr_t = \alpha(t)\{\bar{r}(t) - r_t\}\, dt + \sigma(t) r_t^\rho dW_t, \tag{1.1}$$

where r_t represents the short-term, continuously compounded interest rate, and $\{W_t\}$ is a Brownian motion under the risk-neutral probability. The parameter ρ takes one of the two values 0 or 1/2, depending on whether it extends the Vasiček or Cox-Ingersoll-Ross model. The parameter functions $\alpha(t)$, $\bar{r}(t)$, and $\sigma(t)$ extend the basic models, in which these parameters are just constants. The randomness in this model comes from the Brownian motion, and for the extended Vasiček model when $\rho = 0$, it can be interpreted as adding white noise to the short rate. For the extended Cox-Ingersoll-Ross model the noise is multiplicative, but it is still applied directly to the short rate process.

The main problem with this model is in the way it handles the cyclical nature of interest rates. A time series of interest rates tends to appear cyclical because the supply and demand for money is closely related to income growth, which fluctuates with the business cycle. This has implications for real (adjusted for inflation) interest rates. For example, at a business cycle peak short-term rates should be rising and at a trough rates should be falling. This also has implications for the slope of the term structure—it should be steeper at a peak and flatter at a trough. Roma and Torous [38] find that this property of real interest rates cannot be explained by a simple additive noise type model, such as Vasiček. The Hull-White extension can provide a correction for this problem to a degree, but since the parameter functions are deterministic, it implies that the business cycle effects are known with certainty, which does not allow for the possible variation in length and intensity from what is expected. In addition, when the central bank targets a constant rate of inflation, this fluctuation is transferred to nominal interest rates, so the same characteristics could apply to them.

We approach this problem by modeling the mean-reverting level directly as a random process, and have the short rate chase the mean-reverting level in a linear drift type model. This is similar to the model proposed by Balduzzi, Das, and Foresi [2], except instead of a diffusion process, here the mean-reverting level is assumed to follow a finite-state, continuous-time Markov chain. The switching of the Markov chain to different levels produces a cyclical pattern in the short rate that is consistent with the above effect, and the randomness inherent in the Markov chain prevents the business cycle lengths and intensities from being completely predictable.

Current theory about the dynamics of the short-term, default-free interest rate suggests two alternative methods of modeling, an equilibrium approach and a no arbitrage approach. The later takes the current term structure as an input so as to force an exact fit to longer-term bond prices and other interest rate derivatives. Examples of this approach include the Ho-Lee [25], the Hull-White [27] and the Heath-Jarrow-Morton [21] models. On the other hand equilibrium models such as Vasiček [41] and Cox-Ingersoll-Ross [6], generally do not predict values that exactly match current term structures. In this sense, such models are not arbitrage free. However, this shortcoming is often made up for by the model's applicability to future time periods, as they are usually assumed stationary, where as no arbitrage models are non-stationary by their nature. Also, because of the limitations of financial data, the term structures used as inputs for no arbitrage models are finite, so in practice, arbitrage free predictions can often be achieved by equilibrium type models with a sufficiently large number of parameters.

In the particular case of the equilibrium type model used by Chan, Karolyi, Longstaff and Sanders [4], the interest rate is supposed to follow a mean reverting process of the form

$$dr_t = \alpha(\bar{r} - r_t)dt + \sigma r_t^{\gamma} dW_t, \tag{1.2}$$

where α, $\bar{r}$, and σ are positive and $\gamma \in [0, 1]$.

In unconstrained estimation by CKLS it is found that the variance elasticity, γ, is approximately 1.5, (for U.S. interest rates), and this causes the previous SDE to have an unstable solution, (r_t can go to infinity in a finite time), which is clearly undesirable. The above interest rate model has two important features: the drift term is linear and the volatility term is deterministic.

Relaxing one or both of these properties could resolve the problem. For instance if the drift term was non-linear so that it increased the mean reverting force as the interest rate became large, this could prevent the process from exploding; (this was examined by Aït-Sahalia [1]). Or if the volatility were allowed to be stochastic, as in Longstaff and Schwartz [33], the need for randomness implicit in the r_t^γ term could be reduced, requiring a smaller elasticity parameter, γ. These issues are addressed in a working paper by Licheng Sun [40], and he finds that the stochastic volatility was significant and the non-linear drift was not significant.

One small criticism of Sun's approach to addressing this issue is that the average interest rate, or mean reverting level, ($\bar{r}$ in the above formulation), is presumed constant, as it is in most equilibrium type models. However, it is well known that interest rates typically fluctuate with the business cycle, so it seems reasonable to allow the mean reverting level to change over time. (Taken to its extreme, this leads to a no arbitrage type model.) It seems intuitive that allowing the mean reverting level to change over time would reduce the deviations from said level and thus result in less volatility and possibly less need for randomness in the volatility. Allowing the parameters to switch according to a Markov chain introduces a certain non-linearity in the drift term. This allows us to address and test for the importance of stochastic volatility in this new context, and thus check the robustness of Sun's result. We anticipate that this form of non-linear drift will become significant at the expense of the stochastic volatility. It also allows us to investigate properties of the rate of mean reversion for different interest rate levels. For example, does the central bank prefer to lower interest rates more quickly than it raises them, (perhaps because of political pressure)?

Coupling a Markov chain with the Brownian motion in Equation 1.1 introduces a second factor of randomness into the model. The usual approach to incorporating more factors of randomness into the Vasiček model has been to include Brownian noise in the mean reversion or volatility parameters, (so that they are random functions of time). See, for example, Hull and White [26], Stein and Stein [39], Heston [22], or the thesis by Leblanc [32]. In contrast, we model this additional randomness factor as a finite-state, continuous-time Markov chain. This introduces random jump discontinuities in the mean reversion and volatility parameters, which could simulate discrete information arrival. As a consequence, this approach is well suited to a regime-switching application.

There are two main reasons that we choose to model the mean level, and volatility parameters as Markov chains. First of all, in keeping with the economic notion that participants of the short-term interest rate markets are rational, and thus forward looking, the principle of optimally choosing demand structures requires the Markov property (this is a basic property of any adaptive control process, see Bellman [3]). Intuitively, this means that if we are trying to predict a future value of the stochastic process, then the entire history of the process provides no more useful information than does the current value of the process. Indeed, if past information can be used to obtain a better estimate of return, then it will be used by rational investors and thus incorporated into the current value. The Markov chain is the simplest class of stochastic processes possessing this property so, by the Principle of Parsimony, it seems that this should be the next logical extension of the model in order to attempt to overcome the difficulties mentioned above. Secondly, the central bank authority governs the short-term interest rate, and the central bank adjusts the mean level of the short-term interest rate in discrete increments, (usually one quarter of one percent). This indicates that a Markov chain might also be the most natural tool for incorporating the required additional randomness for this application.

More information about stochastic volatility models, in which the volatility parameter is modeled as the solution to a stochastic differential equation, can be found in the references mentioned above, or for a very complete and rigorous treatment, see the thesis by Leblanc [32], or the sources cited within these papers. The paper by Elliott, Fischer, and Platen [13] examines the case when the mean level follows a Markov chain, but the volatility does not. Hansen and Poulsen [19] examine a special case of this, in which the mean level follows a two state process where switching is governed by a Poisson process. Landén [31] obtains a closed-form solution for a model with mean level depending on a finite state Markov chain. She also obtains a system of partial differential equations, whose solution applies for any affine term structure model, (which does describe our model). However, the approach we use to determine the bond price is simpler in the sense that we avoid requiring the solution of complicated partial differential equations, and instead require the solution to a homogeneous linear system of ordinary differential equations. As a final note on the motivation for this paper, there have been several empirical studies examining the issue of regime switching volatility for the short-term interest rate, and finding that the volatility does seem to be governed by a Markov chain. Some of these papers include Hamilton [18], Driffil [8], Gray [17], and Kalimipalli and Susmel [28]. Because of this, it seems important to explore the implications that Markov chain volatility has for the term structure.

This thesis is organised into three main components. First, in Chapter 2 we provide some background information on Markov chains and functions of Markov chains. This background is used throughout to value zero-coupon bonds and other quantities that are shown to be functions of the Markov chain. Next we explore implications from the basic interest rate model. This is a one-factor model that generates randomness exclusively through a Markov chain. In particular, it has no diffusion term. Its sample paths have several interesting properties, including smooth changes, and yet the process generates interest rates that are quite variable. Also, it generates interest rates that take values in a finite set. This is the first time that such a process has been used to describe short rate dynamics, and thus the resulting implication for the term structure are also new. In Chapter 3, we show how to value zero-coupon bonds, obtain bond and yield dynamics, and calibrate the model using the initial term structure as input. The remainder of the thesis explores and extension to the basic model, in which noise from an independent Brownian motion is introduced and modulated by a stochastic volatility depending on the Markov chain. In Chapter 4, we examine the properties if the term structure under this short rate model, Chapter 5 develops the methodology needed for filtering and estimating the parameters, and Chapter 6 applies the filtering method to historical Canadian 3-month T-Bill data. The remaining chapters and the appendix provide tables and figures, as well as detailed mathematical proofs.

This page is intentionally left blank.

Chapter 2

The Markov Chain

This chapter is devoted to describing the relevant features of the Markov chain used in modeling throughout most of this treatise.

2.1 Introduction

Randomness is modeled by a complete probability space denoted by $(\Omega, \mathcal{F}, P)$ and the revelation of information is modeled by the increasing filtration of sub-σ-fields $\{\mathcal{F}_t\}$ where $t \in [0, \infty)$ indexes time. We assume that the probability space is large enough to support the Markov chain defined below, and that the filtration satisfies the usual conditions of being right continuous and having $\mathcal{F}_0$, and hence all other members, contain all null events of $\mathcal{F}$.

Definition 1 A stochastic process $\{X_t\}$ is said to satisfy the *Markov property* (with respect to the filtration $\{\mathcal{F}_t\}$) if

$$P(X_{t+s} \in B|\mathcal{F}_t) = P(X_{t+s} \in B|X_t), \tag{2.1}$$

for all $s, t \geq 0$ and all Borel sets, B.

Taking $s = 0$, this definition implies that $\{X_t\}$ must be adapted to the filtration $\{\mathcal{F}_t\}$.

Definition 2 An adapted stochastic process that satisfies the Markov property and takes values in a countable set, (the state space), is called a (*continuous-time*) *Markov chain* (with respect to the filtration $\{\mathcal{F}_t\}$). The *transition matrix* of a Markov chain is a square matrix whose entries are functions of two time variables denoting the probability of going from state i at time t to state j at time $t+s$,

$$P_{ji}(s,t) = P(X_{t+s} = j | X_t = i). \tag{2.2}$$

Denoting the state space by S, the Markov property for a Markov chain is characterised by the *Chapman-Kolmogorov equation*

$$P_{ji}(s+t,u) = \sum_{k \in S} P_{ki}(t,u) P_{jk}(s,t+u), \tag{2.3}$$

for all $s,t,u \geq 0$ and all $i,j \in S$.

2.2 Semi-Martingale Representation

For our purposes, we consider a Markov chain—denoted by $\{X_t\}$—with a finite state space (with N states), which without loss of generality we can take to be the set of unit vectors in $\mathbf{R}^N$, $S = \{e_1, \ldots, e_N\}$, where e_i is a column vector with 1 in the i^{th} entry and 0 elsewhere. Furthermore, we assume that the transition matrix is such that the sample paths of the Markov chain are right continuous with left limits existing (so it is *progressively measurable*), and that there exists a *transition rate* (or *intensity*) *matrix*, $A(t)$, with non-negative off-diagonal entries and columns that sum to 0—a so-called (*conservative*) *Q-matrix*—that generates the transition matrix via the forward and backward Kolmogorov equations

$$\frac{\partial P_{ji}(s,t)}{\partial s} = \sum_{k \in S} A_{jk}(t+s) P_{ki}(s,t), \tag{2.4}$$

and

$$\frac{\partial P_{ji}(s,t)}{\partial t} = -\sum_{k \in S} P_{jk}(s,t) A_{ki}(t), \tag{2.5}$$

or, in matrix notation

$$\frac{\partial P(s,t)}{\partial s} = A(t+s) P(s,t), \tag{2.6}$$

and

$$\frac{\partial P(s,t)}{\partial t} = -P(s,t)A(t). \tag{2.7}$$

We will find it convenient to write the forward equation in integral form as

$$P(s,t) = I + \int_0^s A(t+u)P(u,t)\,du, \tag{2.8}$$

where I is the $(N \times N)$ identity matrix.

Furthermore, if the Markov chain is *homogeneous*, so that the transition matrix $P_{ji}(s,t)$ is independent of t, then the existence of a transition rate matrix A follows from the transition matrix being *standard* (i.e. $\lim_{s\downarrow 0} P(s) = I$). Also, the transition rate matrix is independent of time and is multiplicatively commutative with $P(s)$, i.e. $AP(s) = P(s)A$. A similar sufficient condition exists for the non-homogeneous case including continuity of $P(s,t)$, and such requirements are natural in our models. Details of these results can be found in Chung [5], Feller [15], Heyman and Sobel [23], and Karatzas and Shreve [29], for example.

If $p(t)$ is a vector representing the probability distribution of the state of the Markov chain at time t, then by the law of iterated projections $P(s,t)p(t)$ gives the probability distribution at time $t+s$. Post-multiplying by the vector $p(t)$ in the forward Kolmogorov equation gives

$$p(t+s) = p(t) + \int_0^s A(t+u)p(t+u)\,du = p(t) + \int_t^{t+s} A(u)p(u)\,du. \tag{2.9}$$

Notice that because of our choice of state space S, the probability distribution vector can be represented as the expectation $p(t) = E[X_t]$ and substituting gives

$$E[X_{t+s}] = E[X_t] + \int_t^{t+s} A(u)E[X_u]\,du. \tag{2.10}$$

Thus $E[X_t]$ satisfies an N-dimensional homogeneous linear system of differential equations, and can be represented by the fundamental matrix of the system, $E[X_{t+s}] = \Phi(t+s)\Phi^{-1}(t)E[X_t]$. (We assume throughout that a fundamental matrix satisfies the initial value problem associated with the differential equation such that at time $t = 0$ the fundamental matrix is simply the identity matrix, and uniqueness of IVP solutions permits us to discuss *the* fundamental matrix of a system.) The same argument holds when $p(t)$ is a conditional probability distribution given the events of some prior information. In particular, for $0 \leq s \leq t \leq u$,

$$E[X_u|\mathcal{F}_s] = E[X_t|\mathcal{F}_s] + \int_t^u A(v)E[X_v|\mathcal{F}_s]\,dv. \tag{2.11}$$

Note that Equation 2.11 satisfies the exact same system as the unconditional expectation, so the same fundamental matrix applies to this case, where for $s \leq t \leq u$

$$E[X_u|\mathcal{F}_s] = \Phi(u)\Phi^{-1}(t)E[X_t|\mathcal{F}_s]. \tag{2.12}$$

This relationship is used in the proof of the following lemma, which is given in the appendix. This lemma is adapted from Elliott [11].

Lemma 1 *If the transition rate matrix is finite in matrix norm for all $t \in [0, \infty)$, then the stochastic process, $\{X_t\}$, is a semi-martingale with respect to the filtration $\{\mathcal{F}_t\}$ with the decomposition*

$$X_t = X_0 + \int_0^t A(s)X_s\,ds + M_t, \tag{2.13}$$

where $\{M_t\}$ is a square-integrable, right-continuous, zero-mean martingale with respect to $\{\mathcal{F}_t\}$.

From now on we will assume that the transition rate matrix is bounded in matrix norm. This means that the dynamics of our Markov chain can be expressed as

$$dX_t = A(t)X_t\,dt + dM_t. \tag{2.14}$$

Another quantity of interest is the covariance matrix for the Markov chain, $\Sigma(t) = E[X_t X_t^\mathsf{T}] - E[X_t]E[X_t]^\mathsf{T}$. This can be easily determined by observing that $X_t X_t^\mathsf{T} = \text{diag}[X_t]$, so

$$\Sigma(t) = \text{diag}[\Phi(t)X_0] - \Phi(t)\text{diag}[X_0]\Phi(t)^\mathsf{T}. \tag{2.15}$$

The conditional covariance given $\mathcal{F}_s$ can be obtained by replacing X_0 by $\Phi^{-1}(s)X_s$. The auto-covariance matrix can be determined by first conditioning on the smaller σ-algebra and then taking expectation. This gives

$$\Sigma(t,u) = E[X_t X_u^\top] - E[X_t]E[X_u]^\top \tag{2.16}$$
$$= \begin{cases} \mathrm{diag}[\Phi(t)X_0]\Phi^{-1}(t)^\top\Phi(u)^\top - \Phi(t)\mathrm{diag}[X_0]\Phi(u)^\top & \text{if } t \le u \\ \Phi(t)\Phi^{-1}(u)\mathrm{diag}[\Phi(u)X_0] - \Phi(t)\mathrm{diag}[X_0]\Phi(u)^\top & \text{otherwise.} \end{cases}$$

2.3 Functions of the Markov Chain

Now consider a suitable, (measurable and bounded on compact subsets, for instance), vector-valued (or scalar-valued if $m = 1$) function $f : [0,\infty) \times S \times \Omega \to \mathbf{C}^m$ and the stochastic process $\{f_t\}$ where $f_t = f(t, X_t) = f(t)^\top X_t$ and $f(t) = (f(t,e_1),\ldots,f(t,e_N))^\top$ is an $N \times m$-matrix over $\mathbf{C}$. For now we assume that $f(t)$ is a deterministic function, in which case clearly $\{f_t\}$ is adapted to $\{\mathcal{F}_t\}$ and it is also a Markov process since for any bounded and measurable function g, $E[g(f_t)|\mathcal{F}_s] = g \circ f(t)^\top E[X_t|X_s]$. Also, the expected value is $E[f_t] = f(t)^\top E[X_t]$, where $E[X_t]$ can be determined using the forward Kolmogorov equation as outlined earlier. Neither of these observations are generally true if $f(t)$ is not deterministic. The following lemma, whose proof is in the appendix, describes the dynamics for functions of X_t.

Lemma 2 *If the matrix $f(t)$ is continuous, adapted to $\{\mathcal{F}_t\}$ and of finite variation, then the process f_t has the following semi-martingale decomposition*

$$f_t = f_0 + \int_0^t f(s)^\top A(s)X_s\,ds + \int_0^t \{df(s)^\top X_s\} + \int_0^t f(s)^\top\,dM_s. \tag{2.17}$$

Another quantity of interest is the time integral of f_t, $\int_0^t f_s\,ds$. This quantity depends on the entire history of X_t so we cannot expect it to be Markov; however for a deterministic $f(t)$, if we can apply Fubini's theorem or Tonelli's theorem, then the expectation can be calculated as $E[\int_0^t f_s\,ds] = \int_0^t f(s)^\top\Phi(s)X_0\,ds$, where $\Phi(s)$ is the fundamental matrix associated with the forward Kolmogorov equation.

We are particularly interested in the stochastic process $\exp(\int_0^t f_u\,du)$, and its expected value in particular. Quantities of this type will enter frequently in applications to interest rate modeling, so we summarise the approach to obtaining the expected value in the following lemma, which is proven in the appendix.

Lemma 3 *If $f(t)$ is deterministic and vector valued, $f_t = f(t)^\mathsf{T} X_t$ is scalar valued, and $\exp(\int_0^t f_s\,ds)$ is integrable for all t, then*

$$E\left[\exp\left(\int_0^t f_s\,ds\right)X_t\right] = \Phi_f(t)X_0, \tag{2.18}$$

and

$$E\left[\exp\left(\int_0^t f_s\,ds\right)\right] = \tilde{1}^\mathsf{T}\Phi_f(t)X_0, \tag{2.19}$$

where $\tilde{1}$ is a column vector in $\mathbf{R}^N$ with 1 in each entry, and $\Phi_f(t)$ is the fundamental matrix to the following N-dimensional homogeneous linear system of ordinary differential equations

$$y'(t) = \{A(t) + \mathrm{diag}[f(t)]\}y(t). \tag{2.20}$$

Remark 1 Note that if $f(t) = 0$ for all t, then the fundamental matrix is just that of the forward Kolmogorov ODE, $\Phi_0(t) = \Phi(t)$.

We will also encounter the need to find the expected value of the stochastic process $\exp(\int_0^t \{f(s)^\mathsf{T} X_s + g(s)^\mathsf{T} X_t\}\,ds)$ in our interest rate modeling. This is described in the following lemma, proven in the appendix.

Lemma 4 *If $f(t)$ and $g(t)$ are deterministic and N-dimensional vector valued, and $\exp(\int_0^t \{f(s)^\mathsf{T} X_s + g(s)^\mathsf{T} X_t\}\,ds)$ is integrable for all t, then*

$$E\left[\exp\left(\int_0^t \{f(s)^\mathsf{T} X_s + g(s)^\mathsf{T} X_t\}\,ds\right)X_t\right] = \Phi_{fg}(t)X_0, \tag{2.21}$$

and

$$E\left[\exp\left(\int_0^t \{f(s)^\mathsf{T} X_s + g(s)^\mathsf{T} X_t\}\,ds\right)\right] = \tilde{1}^\mathsf{T}\Phi_{fg}(t)X_0, \tag{2.22}$$

where $\tilde{1}$ is a column vector in $\mathbf{R}^N$ with 1 in each entry, and $\Phi_{fg}(t)$ is the fundamental matrix to the following N-dimensional homogeneous linear system of ordinary differential equations

$$y'(t) = \{\mathrm{diag}[h(t)]A(t)\mathrm{diag}[h(t)]^{-1} + \mathrm{diag}[f(t) + g(t)]\}y(t), \qquad (2.23)$$

where $h(t)$ is an N-dimensional vector with typical entry given by $h_i(t) = \exp\Big(\int_0^t \{f(s)^\top X_s + g(s)^\top e_i\}\, ds\Big)$.

These last two lemmas outline the types of solutions we can expect when incorporating a Markov chain into our interest rate models. Typically we obtain solutions in the form of a linear ODE. We will apply this solution technique to obtain bond prices, as well as Laplace and Fourier transforms of probability densities. These last are useful for obtaining moments and for pricing non-linear contingent claims. We now proceed to describe our basic interest rate model.

This page is intentionally left blank.

Chapter 3

The Basic Model

This chapter describes the equations used in our basic model of interest rates. We examine the qualitative features such a model imposes on short-term interest rates, as well as longer-term rates.

3.1 Introduction

The central bank usually implements its monetary policy by controlling the short-term interest rate of its particular currency. We model how short-term interest rates are changed by the central bank according to the following assumptions.

1. The short rate evolves continuously and smoothly through time.
2. At each instant there is an ideal short rate or *equilibrium value* such that if the short rate was at that value it would not be changed.
3. The rate of change experienced by the short rate depends on the difference between the current short rate value and the current equilibrium value.

These assumptions are an attempt to quantify the delicate balance that a central bank must strike between setting a rate that it considers ideal, the equilibrium value, and maintaining a certain amount of predictability in the short rate, smooth changes. "If we want things to stay as they are, things will have to change" (The Leopard, Giuseppe di Lampadusa).

Denote the short-term interest rate by r and the equilibrium value by $\bar{r}$. The above assumptions imply that the short rate dynamics take the following form

$$\frac{dr}{dt} = f(\bar{r} - r); \quad \text{where } f(0) = 0. \tag{3.1}$$

Writing the function f as a Maclaurin series gives

$$f(x) = 0 + ax + bx^2 + o(x^2). \tag{3.2}$$

From this we obtain the most parsimonious short rate model by setting b and all higher order coefficients to zero,

$$\frac{dr}{dt} = a\{\bar{r} - r\}. \tag{3.3}$$

Of course the above assumptions do not preclude either the equilibrium value $\bar{r}$ or the function f from changing through time since both could presumably depend on the state of the economy. In this case both the equilibrium value $\bar{r}$ and the Maclaurin series coefficients a, b, etc. should be treated as functions of time. The above derivation shows that while it may not be the exact relationship, Equation 3.3 does provide a linear approximation to the true relationship, and is therefore worthy of consideration.

3.2 The Short Rate Model

We are now in a position to introduce randomness into the model. This is done by considering an underlying (complete) probability space $(\Omega, \mathcal{F}, P)$ with a filtration of sub-σ-fields $\{\mathcal{F}_t\}$, and letting the equilibrium short rate value $\bar{r}$ be an adapted stochastic process over the probability space.

One of the aims of this research is to examine the effect that a business cycle component in short-term interest rates would have on longer-term rates and interest rate derivatives. To this end we model the equilibrium value $\bar{r}$, which we will sometimes call the *mean-reverting level or manifold*, as a function of a Markov chain. In particular, we have the stochastic process $\bar{r}_t = \bar{r}(t, X_t)$, where $\{X_t\}$ is assumed to be a Markov chain of the type described in Chapter 2, on the filtered probability space. This will allow the short rate to trace out a cyclical sample path, but with varying cycle lengths and intensities. Incorporating these ideas into Equation 3.3 gives the following short rate dynamics,

$$dr_t = a(t)\{\bar{r}(t, X_t) - r_t\}\, dt. \tag{3.4}$$

Remark 2 It is clear that the short rate can be assured to be positive as long as it starts off positive and the minimum mean reverting level is non-negative, i.e. $\inf_{t,i} \bar{r}(t, e_i) \geq 0$.

Remark 3 Notice that Equation 3.4 implies that the resulting short rate process is a process adapted to $\{\mathcal{F}_t\}$ and of bounded variation, and thus a semi-martingale, albeit with the trivial martingale component 0 a.s. Furthermore, it is clear that by coupling Equation 3.4 with Equation 2.14, the processes $\{r_t\}$ and $\{X_t\}$ are jointly Markov with respect to $\{\mathcal{F}_t\}$.

Remark 4 Because the augmented filtration generated by the short rate process is equivalent to that generated by the Markov chain, it is inconsistent with the above dynamics to have observable short rates and a hidden Markov chain. For example, if market information could be modeled by a sub-filtration, $\{\mathcal{G}_t\}$, and the market observed the Markov chain projected onto this information, $Y_t = E[X_t|\mathcal{G}_t]$, which is a vector whose entries represent the conditional probability of the Markov chain being in each state, then for the short rate to be observable, we could replace X_t with Y_t in Equation 3.4. However, this would affect the methodology developed in Chapter 2, since functions of Y_t are not necessarily linear, as functions of X_t must be. Alternatively, we could model r_t as a hidden short rate and let $\hat{r}_t = E[r_t|\mathcal{G}_t]$ be the observed short rate. Unfortunately, this alternative leads to the same problem as above when applying the short rate model to interest rate derivative pricing. This indicates that the more realistic case of a hidden Markov chain requires a more detailed short rate model—something that we present in Chapter 4.

The quantity $\bar{r}(t, X_t)$ is the level that the process tends toward, and $a(t)$, which is assumed to be a positive valued function, is the rate at which the mean-reverting level is approached. In general, these dynamics do not describe a stationary process, because the parameters may vary with time; however, the process can be made stationary by requiring that the parameters depend on time only through the Markov chain. The time inhomogeneity is allowed for generality, and it happens that in this case the initial term structure can be matched exactly, as in Hull and White [27]. In particular, we can choose how $\bar{r}(t, e_i)$ follows time in each state in order to match the initial term structure of interest rates and $a(t)$ so that the term structure of volatilities is matched. This gives market estimates about how the equilibrium value and the Maclaurin coefficients will change through time. We assume that the parameter functions are suitably well behaved, continuous for instance, so that the various functions and integrals of them that follow are well defined and finite.

The main difference between the stochastic process described in Equation 3.4 and the Hull-White dynamics of Equation 1.1 is that instead of incorporating noise into the short rate using a diffusion term, randomness enters the short rate through the mean-reverting level. This is a substantial change because it means that the interest rate sample path is differentiable. This means that short rate changes smoothly through time, although with a discontinuous and random slope. We use a Markov chain to generate randomness rather than a Brownian motion because it seems intuitively better suited to deal with the business cycle effect.

The solution to Equation 3.4 is obtained by the variation of parameters method,

$$r_t = \exp\Big(-\int_0^t a(s)\,ds\Big)\Big\{r_0 + \int_0^t \exp\Big(\int_0^s a(u)\,du\Big)a(s)\bar{r}(s, X_s)\,ds\Big\}. \tag{3.5}$$

Writing $\bar{r}(s, X_s) = \bar{r}(s)^{\mathsf{T}} X_s$, where $\bar{r}(s) = (\bar{r}(s, e_1), \ldots, \bar{r}(s, e_N))^{\mathsf{T}}$, and letting

$$\alpha(s; t) = \exp\Big(-\int_s^t a(u)\,du\Big)a(s)\bar{r}(s) \tag{3.6}$$

we get

$$r_t = \exp\Big(-\int_0^t a(s)\,ds\Big)r_0 + \int_0^t \alpha(s;t)^{\mathsf{T}} X_s\,ds. \tag{3.7}$$

The expected future short rate can be easily determined from Equations 2.12 and 3.7. Applying Tonelli's theorem to take the expectation operator through the integral sign gives

$$E[r_t] = \exp\Big(-\int_0^t a(s)\,ds\Big)r_0 + \int_0^t \alpha(s;t)^{\mathsf{T}} \Phi(s) X_0\,ds. \tag{3.8}$$

Note that

$$r_t - E[r_t] = \int_0^t \alpha(s;t)^{\mathsf{T}}\{X_s - E[X_s]\}\,ds, \tag{3.9}$$

so we can write

$$(r_t - E[r_t])^2 = \int_0^t \int_0^t \alpha(s;t)^{\mathsf{T}}\{X_s - E[X_s]\}\{X_u - E[X_u]\}^{\mathsf{T}}\alpha(u;t)\,du\,ds. \tag{3.10}$$

Therefore applying Equation 2.16 gives the variance of r_t,

$$\mathrm{var}[r_t] = \int_0^t \int_0^t \alpha(s;t)^{\mathsf{T}}\Sigma(s,u)\alpha(u;t)\,du\,ds. \tag{3.11}$$

To obtain other moments of r_t we determine the characteristic function or Fourier transform $E[e^{i\theta r_t}]$. This is done by applying Lemma 3 to Equation 3.7. To do this, first fix t and θ and note that the first term in $i\theta r_t$ is deterministic, as is the vector $i\theta\alpha(s;t)$. By Lemma 3, for all $v \le t$ we have

$$E\Big[\exp\Big(\int_0^v i\theta\alpha(s;t)^{\mathsf{T}} X_s\,ds\Big)\Big] = \tilde{1}^{\mathsf{T}}\Phi_\alpha(v;t,\theta)X_0, \tag{3.12}$$

where $\Phi_\alpha(v;t,\theta)$ is the fundamental matrix for

$$y'(v;t,\theta) = \{A(v) + i\theta\,\mathrm{diag}[\alpha(v;t)]\}y(v;t,\theta). \tag{3.13}$$

Putting $v = t$ gives

$$E[e^{i\theta r_t}] = \exp\Big\{i\theta\exp\Big(-\int_0^t a(s)\,ds\Big)r_0\Big\}\tilde{1}^{\mathsf{T}}\Phi_\alpha(t;t,\theta)X_0. \tag{3.14}$$

$$+ \int_0^t a(s)\text{diag}[\bar{r}(s)]\Phi(s)X_0\, ds.$$

Differentiating gives the non-homogeneous linear ODE

$$y'(t) = \{A(s) + a(s)I\}y(t) + \varphi(t),$$

The n^{th} uncentred moment can then be calculated by differentiation as $E[r_t^n] = (-i)^n \frac{\partial^n}{\partial\theta^n} E[e^i\theta r_t]|_{\theta=0}$. Furthermore, the characteristic function characterises the probability density, which can be obtained through the inverse Fourier transform.

Another quantity of interest is the covariance between the short rate and the Markov chain. Since $\text{cov}[r_t, X_t] = E[r_tX_t] - E[r_t]E[X_t]$ and we have already determined the quantities $E[r_t]$ and $E[X_t]$, we need only determine $E[r_tX_t]$. Recalling that r_t is continuous and of bounded variation, and using Itô's integration by parts gives

$$\begin{aligned} r_tX_t - r_0X_0 &= \int_0^t r_s\, dX_s + \int_0^t X_s\, dr_s \qquad (3.15)\\ &= \int_0^t r_sA(s)X_s\, ds + \int_0^t r_s\, dM_s + \int_0^t X_s\, a(s)\{\bar{r}(s, X_s) - r_s\}\, ds \\ &= \int_0^t \{A(s) + a(s)I\}r_sX_s\, ds + \int_0^t a(s)\text{diag}[\bar{r}(s)]X_s\, ds + \int_0^t r_s\, dM_s. \end{aligned}$$

Taking expectation and applying Fubini's theorem yields

$$\begin{aligned} E[r_tX_t] &= r_0X_0 + \int_0^t \{A(s) + a(s)I\}E[r_sX_s]\, ds \\ &\quad + \int_0^t a(s)\text{diag}[\bar{r}(s)]\Phi(s)X_0\, ds. \end{aligned} \qquad (3.16)$$

Differentiating gives the non-homogeneous linear ODE

$$y'(t) = \{A(s) + a(s)I\}y(t) + \varphi(t), \qquad (3.17)$$

where $\varphi(t) = a(t)\text{diag}[\bar{r}(t)]\Phi(t)X_0$. Denoting the fundamental matrix of the homogeneous version of the above ODE by $\Phi_a(t)$ lets us write the solution as

$$E[r_tX_t] = \Phi_a(t)r_0X_0 + \Phi_a(t)\int_0^t \Phi_a^{-1}(s)a(s)\text{diag}[\bar{r}(s)]\Phi(s)X_0\, ds. \qquad (3.18)$$

Thus subtracting the product of the expectations from this quantity gives the covariance.

3.3 Zero-Coupon Bonds

This section provides the details of determining the value of a default free zero-coupon bond that pays \$1 with certainty at a fixed maturity time, T, with no other cash flows. To analyse this case, and throughout the remainder of this chapter, we assume that the probability measure, P, is a risk-neutral measure. Under such a measure, the value of all contingent claims, when discounted at the risk-free rate, follow a martingale process with respect to P and the filtration $\{\mathcal{F}_t\}$. The existence of such a measure follows from the absence of arbitrage and certain technical conditions, according to the fundamental theorem of asset pricing. Furthermore, if we assume that the economy is complete, so that any contingent claim can be tracked perfectly by a dynamic portfolio of existing assets, then the risk-neutral measure must be unique in the sense that all risk-neutral measures must coincide on all relevant events. Details of these claims can be found in Harrison and Kreps [20]. We denote the value of a zero-coupon bond maturing at time T by $B(T)$, and under the risk-neutral measure, its value is determined by solving

$$B(T) = E\left[\exp\left(-\int_0^T r_t\,dt\right)\right]. \tag{3.19}$$

To solve for the quantity in Equation 3.19, we must first determine the integral of the short rate process. This can be done by integrating Equation 3.7,

$$\int_0^T r_t\,dt = \int_0^T \exp\left(-\int_0^t a(s)\,ds\right) r_0\,dt + \int_0^T \int_0^t \alpha(s;t)^{\mathsf{T}} X_s\,ds\,dt. \tag{3.20}$$

However, it will be convenient to have the term X_s in the outer integral; (this is because it puts the integral in a functional form suitable for applying Lemma 3). We can accomplish this by interchanging the order of integration according to Tonelli's theorem. This gives

$$\int_0^T r_t\,dt = r_0 \int_0^T \exp\left(-\int_0^t a(s)\,ds\right) dt + \int_0^T \beta(t;T)^{\mathsf{T}} X_t\,dt, \tag{3.21}$$

where the vector

$$\beta(t;T) = \int_t^T \alpha(t;u)\,du = \left\{\int_t^T \exp\left(-\int_t^u a(s)\,ds\right) du\right\} a(t)\bar{r}(t). \tag{3.22}$$

The first term in Equation 3.21 is $\mathcal{F}_0$-measurable and it can be determined at least by numerical integration—it can be determined analytically for some special cases, in particular when $a(t)$ is constant. This means our task in solving Equation 3.19 is reduced to finding $E[\exp(\int_0^T\{-\beta(t;T)^\top X_t\}\,dt)]$. Since $\beta(t;T)$ is deterministic with positive entries, we can find this value by applying Lemma 3,

$$E\left[\exp\left(\int_0^T \{-\beta(t;T)^\top X_t\}\,dt\right)\right] = \tilde{1}^\top \Phi_\beta(T;T)X_0. \tag{3.23}$$

The solution we find is in terms of a linear system of ordinary differential equations, and we summarise this in the following theorem.

Theorem 1 *If P is a risk-neutral probability and the risk free short-term interest rate is characterised by Equation 3.4, then the value of a zero-coupon bond paying \$1 at time T is*

$$B(T) = \exp\left\{-r_0\int_0^T \exp\left(-\int_0^t a(s)\,ds\right)dt\right\}\tilde{1}^\top \Phi_\beta(T;T)X_0, \tag{3.24}$$

where $\tilde{1}$ is an N-dimensional column vector with 1 in each entry, and $\Phi_\beta(t;T)$ is the fundamental matrix for the N-dimensional homogeneous linear system of ordinary differential equations,

$$y'(t;T) = \{A(t) - \mathrm{diag}[\beta(t;T)]\}y(t;T), \tag{3.25}$$

where $\beta(t;T)$ is given in Equation 3.22.

In general, the above ODE must be solved numerically.

3.4 Zero-Coupon Bond Price Dynamics

This section describes how the value of a zero-coupon bond changes through time under the assumed short-term interest rate model. The flow property of ODEs allows us to write the solution to Equation 3.4 as

$$r_t = \exp\left(-\int_u^t a(s)\,ds\right)\left\{r_u + \int_u^t \exp\left(\int_u^s a(u)\,du\right)a(s)\bar{r}(s,X_s)\,ds\right\}, \tag{3.26}$$

or in the form of Equation 3.7,

$$r_t = \exp\Big(-\int_u^t a(s)\,ds\Big) r_u + \int_u^t \alpha(s;t)^\mathsf{T} X_s\,ds. \tag{3.27}$$

for any $u \le t$. Integrating from t to T gives

$$\int_t^T r_u\,du = r_t \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du + \int_t^T \beta(u;T)^\mathsf{T} X_u\,du, \tag{3.28}$$

and again the flow property of the ODE in Lemma 3 implies

$$E\Big[\exp\Big(\int_t^T \{-\beta(u;T)^\mathsf{T} X_u\}\,du\Big)\,\Big|\,\mathcal{F}_t\Big] = \tilde{1}^\mathsf{T} \Phi_\beta(T;T)\Phi_\beta^{-1}(t;T)X_t. \tag{3.29}$$

Therefore as a corollary to Theorem 1 we have the time t price of a bond maturing at time T is

$$B_t(T) = \exp\Big\{-r_t \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\}\tilde{1}^\mathsf{T}\Phi_\beta(T;T)\Phi_\beta^{-1}(t;T)X_t. \tag{3.30}$$

From this we can see that the time t bond price is a function of X_t.

Now write

$$B(t;T)^\mathsf{T} = \exp\Big\{-r_t \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\}\tilde{1}^\mathsf{T}\Phi_\beta(T;T)\Phi_\beta^{-1}(t;T), \tag{3.31}$$

so $B_t(T) = B(t;T)^\mathsf{T} X_t$. Note that $B(t;T)$ is continuous, adapted to $\mathcal{F}_t$, and of finite variation. Also note that

$$\begin{aligned}\frac{\partial B(t;T)^\mathsf{T}}{\partial t} &= \frac{\partial}{\partial t}\Big\{-r_t \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\} B(t;T)^\mathsf{T} \\ &\quad + \exp\Big\{-r_t \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\}\tilde{1}^\mathsf{T}\Phi_\beta(T;T)\frac{\partial \Phi_\beta^{-1}(t;T)}{\partial t}.\end{aligned} \tag{3.32}$$

Now

$$\begin{aligned}&\frac{\partial}{\partial t}\Big\{-r_t \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\} \\ &= -a(t)\{\bar{r}(t,X_t) - r_t\} \int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du \\ &\quad + r_t\Big\{1 - a(t)\int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\} \\ &= r_t - \beta(t;T)^\mathsf{T} X_t,\end{aligned} \tag{3.33}$$

and

$$\frac{\partial \Phi_\beta^{-1}(t;T)}{\partial t} = -\Phi_\beta^{-1}(t;T)\{A(t) - \text{diag}[\beta(t;T)]\}, \tag{3.34}$$

so

$$\frac{\partial B(t;T)^\mathsf{T}}{\partial t} = B(t;T)^\mathsf{T}\{[r_t - \beta(t;T)^\mathsf{T} X_t]I - A(t) + \text{diag}[\beta(t;T)]\}. \tag{3.35}$$

We are now in a position to apply Lemma 2 to obtain the dynamics of $B_t(T)$. This gives

$$\begin{aligned} B_t(T) &= B_0(T) + \int_0^t \Big\{B(s;T)^\mathsf{T} A(s) + \frac{\partial B(s;T)^\mathsf{T}}{\partial s}\Big\} X_s \, ds \\ &\quad + \int_0^t B(s;T)^\mathsf{T} \, dM_s. \end{aligned} \tag{3.36}$$

Simplifying yields the following theorem.

Theorem 2 *If P is a risk-neutral probability and the risk free short-term interest rate is characterised by Equation 3.4, then the dynamics at time t of a zero-coupon bond paying \$1 at time T are given by the following equation*

$$B_t(T) = B_0(T) + \int_0^t r_s B_s(T) \, ds + \int_0^t B(s;T)^\mathsf{T} \, dM_s, \tag{3.37}$$

where $B(s;T)^\mathsf{T}$ is given by Equation 3.31.

3.5 The Yield Curve

To determine longer-term interest rates we just solve for the yield of the zero-coupon bonds according to the formula $R(T) = -\ln\{B(T)\}/T$. By Theorem 1, this is

$$R(T) = r_0 \frac{1}{T} \int_0^T \exp\Big(-\int_0^t a(s) \, ds\Big) \, dt - \frac{\ln\{\tilde{1}^\mathsf{T} \Phi_\beta(T;T) X_0\}}{T}. \tag{3.38}$$

Remark 5 Note that we can write the second term in Equation 3.38 as $\ln\{\tilde{1}^\mathsf{T} \Phi_\beta(T;T)\}^\mathsf{T} X_0/T$. In this case, the term structure is jointly affine in the factors r and X, (in the sense of Duffie and Kan [9]).

For future longer-term rates, we denote the yield by

$$R_t(\tau) = -\ln\{B_t(t+\tau)\}/\tau. \tag{3.39}$$

This is the τ-term yield available at time t, and it is given by

$$R_t(\tau) = \frac{1}{\tau}\Big\{r_t \int_t^{t+\tau} \exp\Big(-\int_t^u a(s)\,ds\Big)\,du - \gamma(t;t+\tau)^{\mathsf{T}} X_t\Big\}, \tag{3.40}$$

where $\gamma(t;t+\tau)$ is a vector with typical entry

$$\gamma_i(t,t+\tau) = \ln\{\tilde{1}^{\mathsf{T}}\Phi_\beta(t+\tau;t+\tau)\Phi_\beta^{-1}(t;t+\tau)e_i\}. \tag{3.41}$$

We observe that $R_t(\tau)$ is linear in the two random variables r_t and X_t, so the expected future longer-term rate is the linear combination of $E[r_t]$ (Equation 3.8) and $E[X_t]$ (Equation 2.11). The variance of the future longer-term rate can be obtained from the quantities $\mathrm{var}[r_t]$ (Equation 3.11), $\Sigma(t,t)$ (Equation 2.16), and $\mathrm{cov}[r_t, X_t]$ (Equation 3.18).

The dynamics of the τ-term interest rate are calculated as follows

$$\begin{aligned} dR_t(\tau) = \frac{1}{\tau}\Big[&\Big\{\int_t^{t+\tau} \exp\Big(-\int_t^u a(s)\,ds\Big)du\Big\}dr_t \\ +\ & r_t\Big\{\exp\Big(-\int_t^{t+\tau} a(s)\,ds\Big) - 1 + a(t)\int_t^{t+\tau}\exp\Big(-\int_t^u a(s)\,ds\Big)du\Big\}dt \\ -\ & \gamma(t;t+\tau)^{\mathsf{T}}\,dX_t - \frac{\partial}{\partial t}\{\gamma(t;t+\tau)^{\mathsf{T}}\}X_t\,dt\Big]. \end{aligned} \tag{3.42}$$

The term we have yet to determine is the derivative of the vector $\gamma(t;t+\tau)$. Proceeding from the definition, we have

$$\frac{\partial}{\partial t}\gamma_i(t,t+\tau) = \frac{\tilde{1}^{\mathsf{T}}\frac{\partial}{\partial t}\{\Phi_\beta(t+\tau;t+\tau)\Phi_\beta^{-1}(t;t+\tau)\}e_i}{\tilde{1}^{\mathsf{T}}\Phi_\beta(t+\tau;t+\tau)\Phi_\beta^{-1}(t;t+\tau)e_i}. \tag{3.43}$$

Next, note that

$$\frac{\partial}{\partial t}\Phi_\beta(t;t) = \frac{\partial\Phi_\beta(s;t)}{\partial s}\Big|_{s=t} + \frac{\partial\Phi_\beta(s;t)}{\partial t}\Big|_{s=t}. \tag{3.44}$$

Since $\Phi_\beta(s;t)$ satisfies Equation 3.25, and $\beta(t;t)=0$, the first term is obviously

$$\frac{\partial\Phi_\beta(s;t)}{\partial s}\Big|_{s=t} = A(t)\Phi_\beta(t;t). \tag{3.45}$$

The second term is determined in the following way. Write

$$\Phi_\beta(s;t) = I + \int_0^s \{A(u) - \text{diag}[\beta(u;t)]\}\Phi_\beta(u;t)\,du, \tag{3.46}$$

so that

$$\begin{aligned}\frac{\partial \Phi_\beta(s;t)}{\partial t} &= \int_0^s -\text{diag}\left[\frac{\partial}{\partial t}\beta(u;t)\right]\Phi_\beta(u;t)\,du \\ &+ \int_0^s \{A(u) - \text{diag}[\beta(u;t)]\}\frac{\partial}{\partial t}\Phi_\beta(u;t)\,du,\end{aligned} \tag{3.47}$$

where

$$\frac{\partial \beta(u;t)}{\partial t} = \alpha(u;t). \tag{3.48}$$

Differentiating with respect to s gives

$$\begin{aligned}\frac{\partial}{\partial s}\left\{\frac{\partial \Phi_\beta(s;t)}{\partial t}\right\} &= \{A(s) - \text{diag}[\beta(s;t)]\}\left\{\frac{\partial \Phi_\beta(s;t)}{\partial t}\right\} \\ &- \text{diag}[\alpha(s;t)]\Phi_\beta(s;t),\end{aligned} \tag{3.49}$$

which is a non-homogeneous version of Equation 3.25. This means that

$$\frac{\partial \Phi_\beta(s;t)}{\partial t} = -\Phi_\beta(s;t)\int_0^s \Phi_\beta^{-1}(u;t)\text{diag}[\alpha(u;t)]\Phi_\beta(u;t)\,du, \tag{3.50}$$

and therefore

$$\frac{\partial \Phi_\beta(t;t)}{\partial t} = A(t)\Phi_\beta(t;t) - \Phi_\beta(t;t)\int_0^t \Phi_\beta^{-1}(u;t)\text{diag}[\alpha(u;t)]\Phi_\beta(u;t)\,du. \tag{3.51}$$

Replacing t by $t+\tau$ gives the derivative of the first term.

The next task is to find the derivative of $\Phi_\beta^{-1}(t;u)$. Now

$$\Phi_\beta^{-1}(t;u) = I - \int_0^t \Phi_\beta^{-1}(s;u)\{A(s) - \text{diag}[\beta(s;u)]\}\,ds, \tag{3.52}$$

so

$$\begin{aligned}\frac{\partial \Phi_\beta^{-1}(t;u)}{\partial u} &= -\int_0^t \frac{\partial \Phi_\beta^{-1}(s;u)}{\partial u}\{A(s) - \text{diag}[\beta(s;u)]\}\,ds \\ &+ \int_0^t \Phi_\beta^{-1}(s;u)\text{diag}[\alpha(s;u)]\,ds.\end{aligned} \tag{3.53}$$

This is the integral version of a non-homogeneous ODE that corresponds to the inverse to System 3.25. The solution is given by

$$\frac{\partial \Phi_\beta^{-1}(t;u)}{\partial u} = \left\{ \int_0^t \Phi_\beta^{-1}(s;u) \mathrm{diag}[\alpha(s;u)] \Phi_\beta(s;u)\, ds \right\} \Phi_\beta^{-1}(t;u). \quad (3.54)$$

Replacing u by $t+\tau$ gives the derivative of the second term.

We are now in a position to determine the derivative of $\gamma_i(t, t+\tau)$. We have

$$\begin{aligned} & \frac{\partial}{\partial t}\{\Phi_\beta(t+\tau;t+\tau)\Phi_\beta^{-1}(t;t+\tau)\} \quad (3.55) \\ = & \frac{\partial}{\partial t}\{\Phi_\beta(t+\tau;t+\tau)\}\Phi_\beta^{-1}(t;t+\tau) \\ & + \Phi_\beta(t+\tau;t+\tau)\frac{\partial}{\partial t}\Phi_\beta^{-1}(t;t+\tau) \\ = & \Big\{A(t+\tau)\Phi_\beta(t+\tau;t+\tau) - \Phi_\beta(t+\tau;t+\tau) \\ & \times \int_t^{t+\tau} \Phi_\beta^{-1}(u;t+\tau)\mathrm{diag}[\alpha(u;t+\tau)]\Phi_\beta(u;t+\tau)\, du\Big\} \\ & \times \Phi_\beta^{-1}(t;t+\tau). \end{aligned}$$

Substituting this quantity into Equation 3.43 gives the derivative of the vector $\gamma(t;t+\tau)$ and substituting that quantity in to Equation 3.42 gives the longer-term yield dynamics.

3.6 Matching the Initial Term Structure

Since the term structure of zero-coupon bond prices can be exactly matched by a one dimensional system, such as the Hull-White model, it seems clear that the general form of our model will possess too much freedom to have the functions $\bar{r}(t, e_i)$ be uniquely determined by the initial term structure. In fact, whenever the Markov chain and the associated function $\bar{r}(t, X_t)$ have a non-trivial state space, (so the Markov chain has more than one state and $\bar{r}(t, e_i) \neq \bar{r}(t, e_j)$ for some states e_i and e_j and for all t in some subset of time with positive Lebesgue measure), there is very little structure imposed on the mean-reverting manifold. Intuitively, this is because whatever one $\bar{r}(t, e_i)$ function is, we can choose another to "cancel" it out on average when calculating zero-coupon bond prices. Because of this freedom, in order to have the initial term structure be a useful input to the model, we must specify some relationship between mean-reverting functions associated with different states. One simple possibility for accomplishing this is to have the various mean-reverting functions shifted from each other by constant amounts, (or by some other known function of time such as exponential decay). This is the approach we present here.

Suppose that the mean-reverting level is the sum of two quantities,

$$\bar{r}(t, X_t) = \bar{r}_k(t)^{\mathsf{T}} X_t + \bar{r}_u(t). \tag{3.56}$$

Here the $\bar{r}_k(t)$ is a column vector of known functions, (or possibly constants), and the $\bar{r}_u(t)$ is an unknown real-valued function. We assume that all of the other parameters are known or estimated through time series analysis. The goal is to find $\bar{r}_u(t)$ as a function of the other parameters and the current term structure.

We can write the function β from Equation 3.22 as

$$\begin{aligned} \beta(t, X_t; T) &= \beta_k(t, X_t; T) + \beta_u(t; T) \qquad (3.57) \\ &= \Big\{\int_t^T \exp\Big(-\int_t^u a(s)\,ds\Big)\,du\Big\} a(t)\{\bar{r}_k(t, X_t) + \bar{r}_u(t)\}. \end{aligned}$$

This means that the fundamental matrix takes the form

$$\Phi_\beta(t; T) = e^{-\int_0^t \beta_u(s;T)\,ds}\Phi_{\beta_k}(t; T), \tag{3.58}$$

where $\Phi_{\beta_k}(t; T)$ is the fundamental matrix for the known case described in Theorem 1. Therefore the bond price in this situation can be written as

$$B_t(T) = A(t, T)B_t^k(T), \tag{3.59}$$

where

$$A(t, T) = \exp\Big\{-\int_t^T \beta_u(s; T)\,ds\Big\}, \tag{3.60}$$

and $B_t^k(T)$ is the bond price calculated from the known parameter case given by Theorem 1 and the discussion in Section 3.4.

The value $B_t^k(T)$ is known, so to find the bond price, we need only determine $A(t,T)$. We do this by examining two quantities associated with the initial term structure—the ratio of two zero-coupon bonds and the slope of the term structure. First, we note that

$$\begin{aligned}\ln\left\{\frac{B_0(T)}{B_0(t)}\right\} &= -\int_0^T \beta_u(s;T)\,ds + \int_0^t \beta_u(s;t)\,ds + \ln\left\{\frac{B_0^k(T)}{B_0^k(t)}\right\} \\ &= -\int_t^T \beta_u(s;T)\,ds - \int_0^t \{\beta_u(s;T) - \beta_u(s;t)\}\,ds + \ln\left\{\frac{B_0^k(T)}{B_0^k(t)}\right\} \\ &= \ln\{A(t,T)\} - \int_0^t \left\{\int_t^T \exp\left(-\int_s^u a(v)\,dv\right) du\right\} a(s)\bar{r}_u(s)\,ds \\ &\quad + \ln\left\{\frac{B_0^k(T)}{B_0^k(t)}\right\}\end{aligned} \tag{3.61}$$

Next, differentiating with respect to the maturity gives

$$\begin{aligned}\frac{\partial \ln\{B_0(t)\}}{\partial t} &= -\int_0^t \frac{\partial \beta_u(s;t)}{\partial t}\,ds + \frac{\partial \ln\{B_0(t)\}}{\partial t} \\ &= -\int_0^t \exp\left(-\int_s^t a(v)\,dv\right) a(s)\bar{r}_u(s)\,ds + \frac{\partial \ln\{B_0(t)\}}{\partial t}.\end{aligned} \tag{3.62}$$

Finally, we observe that

$$\begin{aligned}&\int_t^T \exp\left(-\int_s^u a(v)\,dv\right) du \\ &= \exp\left(-\int_s^t a(v)\,dv\right) \int_t^T \exp\left(-\int_t^u a(v)\,dv\right) du.\end{aligned} \tag{3.63}$$

Combining the above equations and solving for $\ln\{A(t,T)\}$ yields

$$\begin{aligned}\ln\{A(t,T)\} &= \ln\left\{\frac{B_0(T)}{B_0(t)}\right\} - \ln\left\{\frac{B_0^k(T)}{B_0^k(t)}\right\} \\ &- \left\{\int_t^T \exp\left(-\int_t^u a(v)\,dv\right) du\right\}\left\{\frac{\partial \ln\{B_0(t)\}}{\partial t} - \frac{\partial \ln\{B_0^k(t)\}}{\partial t}\right\}.\end{aligned} \tag{3.64}$$

In theory this is sufficient to determine $A(t,T)$, since $B_0^k(t)$ is known for each t, and hence so is its derivative. However, because $B_0^k(t)$ must be found numerically, finding the derivative using limits could be quite tedious, even for a powerful computer, especially since we require this to be done for a large number of times t. To simplify this process, we outline an alternative procedure for obtaining $\partial B_0^k(t)/\partial t$.

Observe that

$$\begin{aligned}\frac{\partial B_0^k(t)}{\partial t} &= -r_0 \exp\Big(-\int_0^t a(s)\,ds\Big) B_0^k(t) \\ &\quad + \exp\Big\{-r_0 \int_0^t \exp\Big(-\int_0^s a(u)\,du\Big)\,ds\Big\} \tilde{1}^{\mathsf{T}} \frac{\partial}{\partial t} \Phi_{\beta_k}(t;t) X_0.\end{aligned} \tag{3.65}$$

Combining this with Equation 3.51 provides a method to obtain a theoretical bond price term structure slope using the predetermined fundamental matrix. We summarise this result in the following theorem.

Theorem 3 *If P is a risk-neutral probability and the risk free short-term interest rate is characterised by Equation 3.4, where $\bar{r}(t, X_t)$ is as given in Equation 3.56, then the value at time t of a zero-coupon bond paying \$1 at time T is given by Equation 3.59, where $B_t^k(T)$ is given by Theorem 1, $\ln\{A(t,T)\}$ is given by Equation 3.64, $\partial \ln\{B_0^k(t)\}/\partial t$ is given by Equation 3.65, and $\partial \Phi_{\beta_k}(t;t)/\partial t$ is given by Equation 3.51.*

From Equation 3.60 the function $\bar{r}(t)$ can be determined quite easily by differentiation

$$\bar{r}(T) = -\frac{1}{A(t,T)} \frac{\partial A(t,T)}{\partial T} - \frac{1}{a(T)} \frac{\partial}{\partial T} \Big\{ \frac{1}{A(t,T)} \frac{\partial A(t,T)}{\partial T} \Big\}, \tag{3.66}$$

for any $t \in [0, T]$, and where $A(t,T)$ is obtained from Theorem 3. As a caveat to this procedure, we should mention that the well-known problem of interpolating term structure observations, (see Vasiček and Fong [42] for an overview of the problem), can cause considerable error in estimating $\bar{r}(t)$ because of the instability of the differentiation operation.

3.7 Constant Parameters

This section examines the case when the parameters governing the basic model are constant in time. We discuss this special case of the model for two reasons. First, this allows many of the integrals that are left implicit in the solutions of various quantities from previous sections to be solved explicitly. These closed-form solutions can speed up algorithms for pricing interest rate derivatives, and simplify techniques for fitting and calibrating the model. Second, it allows the modeled short rate process to be stationary. This is an important feature for testing the model using a time series of short-term interest rate data. We proceed by evaluating some of the previously defined functions.

Recall that the Markov chain is being governed by a transition rate matrix, which we denoted by $A(t)$. This led to a transition probability matrix $\Phi(t)$, through the forward Kolmogorov equation. When the Markov chain is homogeneous, (and satisfying certain technical conditions for the existence of a transition rate matrix), so that the transition rate matrix is independent of t, then the transition probability matrix takes the form of the exponential matrix,

$$\Phi(t) = e^{At}. \tag{3.67}$$

The next quantity of interest is $\alpha(s;t)$ from Equation 3.6. Writing the short rate dynamics as

$$dr_t = a(\bar{r}(X_t) - r_t)\,dt, \tag{3.68}$$

and letting $\bar{r}$ denote the vector with typical entry $\bar{r}(e_i)$ allows us to put

$$\alpha(s;t) = e^{-a(t-s)}a\bar{r}. \tag{3.69}$$

Solving for the short rate gives

$$r_t = e^{-at}r_0 + ae^{-at}\bar{r}^{\mathsf{T}} \int_0^t e^{as} X_s\,ds. \tag{3.70}$$

To evaluate the integral in the second term, note that we can write

$$\begin{aligned} e^{at}X_t &= X_0 + \int_0^t e^{as}\,dX_s + \int_0^t ae^{as}X_s\,ds \qquad (3.71)\\ &= X_0 + \{A + aI\} \int_0^t e^{as}X_s\,ds + \int_0^t e^{as}\,dMs. \end{aligned}$$

Now, assuming that the matrix $A+aI$ is invertible, (which is generally true except when a takes N particular values that depend on the rate matrix), we have

$$\int_0^t e^{as} X_s\, ds = \{A+aI\}^{-1}\{e^{at}X_t - X_0\} - \{A+aI\}^{-1}\int_0^t e^{as}\, dMs. \quad (3.72)$$

This means that

$$\begin{aligned} r_t &= e^{-at}r_0 + ae^{-at}\bar{r}^{\mathsf{T}}\{A+aI\}^{-1}\{e^{at}X_t - X_0\} \\ &\quad - ae^{-at}\bar{r}^{\mathsf{T}}\{A+aI\}^{-1}\int_0^t e^{as}\, dMs. \end{aligned} \quad (3.73)$$

Taking expectation yields

$$E[r_t] = e^{-at}r_0 + ae^{-at}\bar{r}^{\mathsf{T}}\{A+aI\}^{-1}\{e^{at}e^{At} - I\}X_0, \quad (3.74)$$

and by homogeneity

$$E[r_t|\mathcal{F}_s] = e^{-a(t-s)}r_s + ae^{-a(t-s)}\bar{r}^{\mathsf{T}}\{A+aI\}^{-1}\{e^{a(t-s)}e^{A(t-s)} - I\}X_s. \quad (3.75)$$

Furthermore, if we assume that the Markov chain is recurrent and ergodic, then taking the limit gives

$$\lim_{t\to\infty} E[r_t|\mathcal{F}_s^T] = a\bar{r}^{\mathsf{T}}\{A+\alpha I\}^{-1}\pi, \quad (3.76)$$

where π is the limiting probability distribution vector for the Markov chain. Therefore in the constant parameter case, the expected interest rate converges to a constant that is independent of both the initial interest rate and the initial state of the Markov chain.

Remark 6 Note that π is the unique solution to the following system

$$\begin{aligned} A\pi &= 0 \\ \sum_{j=1}^{N}\pi_j &= 1 \\ \pi_j &\geq 0 \quad \text{for all } j = 1, ..., N. \end{aligned} \quad (3.77)$$

For the particular case where $N = 2$, we have

$$\begin{aligned} \pi_1 &= \frac{A_{12}}{A_{12}+A_{21}} \\ \pi_2 &= \frac{A_{21}}{A_{12}+A_{21}}. \end{aligned} \quad (3.78)$$

To find the variance of the short rate in this scenario, we must substitute Equation 3.69 into Equation 3.11, and perform the required integration. This leads to the slightly simplified equation

$$\mathrm{var}[r_t] = a^2 e^{-2at} \bar{r}^{\mathsf{T}} \Sigma^*(t) \bar{r}, \tag{3.79}$$

where

$$\Sigma^*(t) = \int_0^t \int_0^t e^{a(s+u)} \Sigma(s,u)\, ds\, du. \tag{3.80}$$

We are also interested in the quantity $\beta(t;T)$ from Equation 3.22. In the case of constant parameters, this simplifies to

$$\beta(t;T) = \{1 - e^{-a(T-t)}\}\bar{r}. \tag{3.81}$$

Using this fact, and applying Theorem 1 gives a bond price of

$$B_t(T) = \exp\left\{-r_t \frac{1 - e^{-a(T-t)}}{a}\right\} \tilde{1}^{\mathsf{T}} \Phi_c(T;T) \Phi_c^{-1}(t;T) X_t, \tag{3.82}$$

where $\Phi_c(u;T)$ solves

$$y'(u) = \{A - (1 - e^{-a(T-u)})\mathrm{diag}[\bar{r}]\} y(u) \tag{3.83}$$

If we write

$$\begin{aligned} &\Phi_c(T;T)\Phi_c^{-1}(t;T) \\ &= I + \int_t^T \{A - (1 - e^{-a(T-u)})\mathrm{diag}[\bar{r}]\} \Phi_c(u;T) \Phi_c^{-1}(t;T)\, du, \end{aligned} \tag{3.84}$$

then we can substitute $v = u - t$ to get

$$\begin{aligned} &\Phi_c(T-t+t;T)\Phi_c^{-1}(t;T) \\ &= I + \int_0^{T-t} \{A - (1 - e^{-a(T-t-v)})\mathrm{diag}[\bar{r}]\} \Phi_c(v+t;T) \Phi_c^{-1}(t;T)\, du. \end{aligned} \tag{3.85}$$

Writing $\Psi(v;t,T) = \Phi_c(v+t;T)\Phi_c^{-1}(t;T)$ makes it clear that this function satisfies the same initial value problem as $\Phi_c(v;T-t)$, and therefore we can replace $\Phi_c(T;T)\Phi_c^{-1}(t;T)$ by $\Phi_c(T-t;T-t)$ in Equation 3.82.

The vector of "hypothetical" bond prices $B(t;T)^{\mathsf{T}}$ mentioned in Equation 3.31 becomes

$$B(t;T)^{\mathsf{T}} = \exp\left\{-r_t \frac{1-e^{-a(T-t)}}{a}\right\} \tilde{1}^{\mathsf{T}} \Phi_c(T-t;T-t). \tag{3.86}$$

The bond price dynamics can still be obtained from Theorem 2, with this adjustment.

As a final consideration we look at the yield curve in this constant parameter example. From Equation 3.40 we have

$$R_t(\tau) = \frac{1}{\tau}\left\{r_t \frac{1-e^{-a\tau}}{a} - \gamma(\tau)^{\mathsf{T}} X_t\right\}, \tag{3.87}$$

where here $\gamma(\tau)$ is a vector with typical entry

$$\gamma_i(\tau) = \ln\{\tilde{1}^{\mathsf{T}} \Phi_c(\tau;\tau) e_i\}. \tag{3.88}$$

Chapter 4

The Hidden Markov Model

In this chapter we explore an extension of the basic model in which the information about the state of the Markov chain is confounded by observing the short-term interest rate in the presence of white noise.

4.1 The Model for the Short Rate

We base our analysis on the underlying probability space, $(\Omega, \mathcal{F}, P)$, where now the probability space is assumed to be large enough to support the Markov chain, $\{X_t\}$, described in Chapter 2 as well as a standard, 1-dimensional Brownian motion, $\{W_t\}$ that is stochastically independent with the Markov chain, $\{X_t\}$.

Next, we consider the filtrations of our two processes. We write $\{\mathcal{F}_t^X\}$ and $\{\mathcal{F}_t^W\}$ for the right-continuous and augmented filtrations generated by X and W respectively. These filtrations represent the information that is obtained by observing the respective processes. Because of the independence of X and W, having additional information about how one evolves in the future does not provide additional information about how the other will evolve. This is formalised in the following lemma, which is proven in the appendix.

Lemma 5 *For any $T > 0$, the stochastic process, W, is still a standard Brownian motion with respect to the filtration, $\{\mathcal{F}_T^X \vee \mathcal{F}_t^W\}$, where $\mathcal{F}_T^X \vee \mathcal{F}_t^W = \sigma(\mathcal{F}_T^X \cup \mathcal{F}_t^W)$. Also, the stochastic process, X, is still a*

Markov chain with respect to the filtration, $\{\mathcal{F}_t^X \vee \mathcal{F}_T^W\}$. In particular, this is true when $T > t$.

Remark 7 Because, by Lemma 5, X is a Markov chain with respect to the filtration, $\{\mathcal{F}_T^W \vee \mathcal{F}_t^X\}$, Lemma 1 implies that X is also a semimartingale with respect to $\{\mathcal{F}_T^W \vee \mathcal{F}_t^X\}$. In particular, M is a martingale with respect to this filtration.

We are now in a position to describe the model of the short-term interest rate. We model the short rate as a continuous stochastic process, $\{r_t\}$, whose dynamics are described by the following stochastic differential equation:

$$dr_t = a(t)(\bar{r}(t, X_t) - r_t)\,dt + \sigma(t, X_t)\,dW_t, \tag{4.1}$$

where $a(t)$, $\bar{r}(t, X_t)$, and $\sigma(t, X_t)$ are deterministic, bounded, Borel-measurable functions. Note that $\bar{r}$ and σ can both be thought of as finite-state, continuous-time Markov chains that take values in the appropriate sub-space of real-valued functions. Also, the fact that both $\bar{r}$, and σ depend on the same Markov chain, X, does not imply that they must be perfectly correlated with each other. In fact, any degree of correlation can be achieved by judicious choices of A, (and in particular N), $\bar{r}$, and σ. Similarly, when thought of as Markov chains, $\bar{r}$, and σ can have a state space with any, (possibly differing), number of states, (not greater than N). It follows that the above dynamics describe a situation that is more general than it may appear. On the other hand, we intentionally rule out the possibility of a depending on X to simplify the problem of solving for the bond price.

Remark 8 Because of the additive noise in this model, in general, (i.e. with a positive volatility), no parameter choice can assure us of strictly interest rates. However we find in Chapter 6 that the Markov chain drift significantly reduces the volatility estimate, and therefore it reduces the probability of obtaining negative rates.

The basic model of the short rate is a specialisation of this case when $\sigma(t, X_t) = 0$ for all t. As such, this general model can be derived similarly to the basic model, but with additional noise corrupting the outcome rate.

In light of Lemma 5, it is well known that the solution to Equation 4.1 is

$$\begin{aligned} r_t &= \exp\Big(-\int_0^t a(u)\,du\Big)\Big\{r_0 + \int_0^t \exp\Big(-\int_0^s a(u)\,du\Big)a(s)\bar{r}(s,X_s)\,ds \\ &\quad + \int_0^t \exp\Big(-\int_0^s a(u)\,du\Big)\sigma(s,X_s)\,dW_s\Big\}. \end{aligned} \tag{4.2}$$

Remark 9 It is often convenient for us to write the solution, for $u \geq t \geq 0$, as

$$r_u = e^{-K(u)}\Big(e^{K(t)}r_t + \int_t^u e^{K(s)}a(s)\bar{r}(s,X_s)\,ds + \int_t^u e^{K(s)}\sigma(s,X_s)\,dW_s\Big). \tag{4.3}$$

with $K(t) = \int_0^t a(s)\,ds$.

We will denote by $\mathcal{F}^r = \{\mathcal{F}_t^r; t \in [0,T]\}$, the right continuous augmentation of the filtration generated by r. From Equation 4.2, it is clear that $\mathcal{F}_t^r \subset \mathcal{F}_t^W \vee \mathcal{F}_t^X$.

Lemma 6 *The stochastic process, r, has the property where, for $0 \leq s \leq t \leq T$,*

$$E[r_t \mid \mathcal{F}_s^r \vee \mathcal{F}_T^X] = E[r_t \mid r_s \vee \mathcal{F}_t^X] \quad a.s., \tag{4.4}$$

and

$$E[r_t \mid \mathcal{F}_s^r \vee \mathcal{F}_s^X] = E[r_t \mid r_s \vee X_s] \quad a.s. \tag{4.5}$$

The next six lemmas describe some of the distributional properties of r, when conditioned on $\mathcal{F}_T^X$. None of these results are surprising, as the corresponding results for the Hull-White model, in which the parameter functions are deterministic, are well known, and conditioning on $\mathcal{F}_T^X$ allows us to think of the functions $\bar{r}$, and σ, as being determined. Because of this, all of the proofs are relegated to the appendix.

Lemma 7 *When conditioned on the σ-algebra, $\mathcal{F}_T^X$, r is a Gaussian process, with mean*

$$E[r_t \mid \mathcal{F}_T^X] = e^{-K(t)}\Big(r_0 + \int_0^t e^{K(s)}a(s)\bar{r}(s,X_s)\,ds\Big) \quad a.s., \tag{4.6}$$

and auto-covariance

$$\operatorname{cov}[r_s, r_t \mid \mathcal{F}_T^X] = e^{-(K(s)+K(t))} \int_0^{s\wedge t} e^{2K(u)} \sigma^2(u, X_u)\, du \quad a.s. \tag{4.7}$$

Lemma 8 *For any $t \in [0,T]$, the stochastic process, $\{r_u; u \in [t,T]\}$, is conditionally Gaussian given $\mathcal{F}_t^r \vee \mathcal{F}_T^X$, with mean*

$$E[r_u | \mathcal{F}_t^r \vee \mathcal{F}_T^X] = e^{-K(u)} \Big(e^{K(t)} r_t + \int_t^u e^{K(s)} a(s) \bar{r}(s, X_s)\, ds \Big) \quad a.s., \tag{4.8}$$

and auto-covariance

$$\operatorname{cov}[r_u, r_v | \mathcal{F}_t^r \vee \mathcal{F}_T^X] = e^{-K(u)} e^{-K(v)} \int_t^{u\wedge v} e^{2K(s)} \sigma^2(s, X_s)\, ds \quad a.s. \tag{4.9}$$

Lemma 9 *The random variable,*

$$R_{0,T} = \int_0^T r_u\, du, \tag{4.10}$$

is a conditionally normal random variable, given $\mathcal{F}_T^X$, with mean

$$E[R_{0,T} \mid \mathcal{F}_T^X] = \int_0^T e^{-K(u)} \Big(r_0 + \int_0^u e^{K(s)} a(s) \bar{r}(s, X_s)\, ds \Big)\, du \quad a.s., \tag{4.11}$$

and variance

$$\operatorname{var}[R_{0,T} \mid \mathcal{F}_T^X] = \int_0^T e^{2K(u)} \Big(\int_u^T e^{-K(s)}\, ds \Big)^2 \sigma^2(u, X_u)\, du \quad a.s. \tag{4.12}$$

Lemma 10 *The random variable,*

$$R_{t,T} = \int_t^T r_u\, du, \tag{4.13}$$

is a conditionally normal random variable, given $\mathcal{F}_t^r \vee \mathcal{F}_T^X$, with mean

$$E[R_{t,T} \mid \mathcal{F}_t^r \vee \mathcal{F}_T^X] = \int_t^T e^{-K(u)} \Big(e^{K(t)} r_t + \int_t^u e^{K(s)} a(s) \bar{r}(s, X_s)\, ds \Big)\, du, \tag{4.14}$$

and variance

$$\operatorname{var}[R_{t,T} \mid \mathcal{F}_t^r \vee \mathcal{F}_T^X] = \int_t^T e^{2K(u)} \Big(\int_u^T e^{-K(s)}\, ds \Big)^2 \sigma^2(u, X_u)\, du, \tag{4.15}$$

almost surely

Lemma 11 *The pair of random variables, r_T and $R_{0,T}$ are conditionally bi-variate normal, given $\mathcal{F}_T^X$, with covariance*

$$\operatorname{cov}[r_T, R_{0,T}|\mathcal{F}_T^X] = e^{-K(T)} \int_0^T e^{-K(u)} \int_0^u e^{2K(s)} \sigma^2(s, X_s)\, ds\, du \quad a.s. \tag{4.16}$$

Lemma 12 *For any $t \in [0,T]$, the pair of random variables r_T, and $R_{t,T}$ are conditionally bi-variate normal, given $\mathcal{F}_t^r \vee \mathcal{F}_T^X$, with covariance*

$$\operatorname{cov}[r_T, R_{t,T}|\mathcal{F}_t^r \vee \mathcal{F}_T^X] = e^{-K(T)} \int_t^T e^{-K(u)} \int_t^u e^{2K(s)} \sigma^2(s, X_s)\, ds\, du \quad a.s. \tag{4.17}$$

4.2 The Model for the Bond Price

We are interested in determining the price of a zero-coupon bond that pays \$1 at time T. (Here we are implicitly assuming that the probability of default is zero.) To do this, we suppose that the dynamics of the risk-free asset, S^0, which can be thought of as a money market account whose future value is subject to stochastic compounding, are as follows:

$$dS_t^0 = r_t S_t^0\, dt, \quad S_0^0 = 1. \tag{4.18}$$

The solution for the above equation is

$$S_t^0 = \exp\Big(\int_0^t r_s\, ds\Big) = e^{R_{0,t}}, \tag{4.19}$$

where $R_{0,T}$ is defined in Lemma 9.

We can think of the bond as a European contingent claim that expires at time T, and pays 1 in every state. As in the case for the basic model, we assume that P is a risk-neutral probability measure, so a no arbitrage price for the bond is the expected discounted value of 1 dollar,

$$B(0,T) = E[\tfrac{1}{S_T^0}] = E[e^{-R_{0,T}}]. \tag{4.20}$$

If the contingent claim paying 1 in every state can be replicated by an admissible hedging strategy, then this is the unique price that does not allow arbitrage. Otherwise we can obtain bounds for $B(0,T)$ by minimising and maximising the expectation over all risk-neutral probability measures.

The difficulty of finding the price from Equation 4.20 is that we do not know the unconditional distribution of $R_{0,T}$. However, since by Lemma 9, conditioned on the σ-algebra $\mathcal{F}_T^X$, $R_{0,T}$ is a normal random variable, applying Tonelli's theorem gives, almost surely,

$$\begin{aligned} E[e^{-R_{0,T}} \mid \mathcal{F}_T^X] &= \exp(-E[R_{0,T} \mid \mathcal{F}_T^X] + \tfrac{1}{2}\text{var}[R_{0,T} \mid \mathcal{F}_T^X]) \\ &= \exp\Big(-r_0 C(0,T) + \int_0^T \beta(t,T,X_t)\,dt\Big) \qquad (4.21) \end{aligned}$$

where

$$\beta(t,T,X_t) = \tfrac{1}{2}C^2(t,T)\sigma^2(t,X_t) - a(t)C(t,T)\bar{r}(t,X_t), \qquad (4.22)$$

and

$$C(t,T) = e^{K(t)} \int_t^T e^{-K(u)}\,du. \qquad (4.23)$$

The law of iterated projections gives

$$B(0,T) = E[E[e^{-R_{0,T}} \mid \mathcal{F}_T^X]] = \exp(-r_0 C(0,T))E[Z_{0,T}], \qquad (4.24)$$

where

$$Z_{0,T} = \exp\Big(\int_0^T \beta(t,T,X_t)\,dt\Big). \qquad (4.25)$$

The remainder of this section derives the price of zero-coupon bonds in considerable detail, and the methods employed here are repeated throughout the remainder of the chapter. The results are summarised in Theorem 4 at the end of this section.

The goal is to find the expected value of Z, and we will use the following approach. We determine the dynamics of Z, use Itô's rule to find dynamics for ZX, use the dynamics of X to solve for $E[ZX]$, and finally obtain $E[Z]$ from $E[ZX]$. It seems that since C is differentiable, so is Z, so the dynamics should be determined by differentiating with respect to T. However, because T enters the integral, (through C), it turns out that this approach is not very fruitful. To proceed, for T fixed, and $v \in [0,T]$, define

$$Z_{0,T}(v) = \exp\Big(\int_0^v \beta(u,T,X_u)\,du\Big). \tag{4.26}$$

Notice that T enters the above integral only through the deterministic function C, so $Z_{0,T}(v)$ is $\mathcal{F}_v^X$-measurable. Also $Z_{0,T} = Z_{0,T}(T)$ and

$$dZ_{0,T}(v) = \beta(v,T,X_v)Z_{0,T}(v)\,dv. \tag{4.27}$$

We can then apply integration by parts for general semi-martingales to get

$$d(Z_{0,T}(v)X_v) = dZ_{0,T}(v)\,X_{v^-} + Z_{0,T}(v^-)\,dX_v + d[Z_{0,T}(\cdot),X]_v. \tag{4.28}$$

Since $Z_{0,T}(\cdot)$ is differentiable, it has finite variation, so the square bracket process adds up the product of the jumps. However, since $Z_{0,T}(\cdot)$ is continuous, it has no jumps, and thus the bracket term is zero. Substituting from above and Lemma 2 gives

$$\begin{aligned} d(Z_{0,T}(v)X_v) &= \beta(v,T,X_v)Z_{0,T}(v)X_{v^-}\,dv \\ &\quad + A(v)Z_{0,T}(v)X_v\,dv + Z_{0,T}(v)\,dM_v. \end{aligned} \tag{4.29}$$

We can write this as

$$\begin{aligned} d(Z_{0,T}(v)X_v) &= \beta(v,T)^{\mathsf{T}}X_vZ_{0,T}(v)X_{v^-}\,dv \\ &\quad + A(v)Z_{0,T}(v)X_vdv + Z_{0,T}(v)\,dM_v. \end{aligned} \tag{4.30}$$

Now since X is right continuous with left limits existing, X has at most a countable number of (jump) discontinuities, so, almost surely, $X_{v^-} = X_v$ for Lebesgue-almost every $v \in [0,T]$. This means that

$$\begin{aligned} Z_{0,T}(v)X_v &= Z_{0,T}(0)X_0 + \int_0^v \beta(u,T)^{\mathsf{T}}X_uZ_{0,T}(u)X_u\,du \\ &\quad + \int_0^v A(u)Z_{0,T}(u)X_u\,du + \int_0^v Z_{0,T}(u)\,dM_u. \end{aligned} \tag{4.31}$$

Remark 10 For every vector $\mathbf{x} \in \mathbf{R}^n$ and every vector, e_j, of the canonical basis of $\mathbf{R}^n$,

$$(\mathbf{x}^{\mathsf{T}}e_j)e_j = \mathrm{diag}[\mathbf{x}]e_j, \tag{4.32}$$

where diag[$\mathbf{x}$] is the $n\times n$ matrix with the entries of $\mathbf{x}$ along the diagonal and zeros elsewhere.

We rewrite Equation 4.31, using Equation 4.32 and the fact that $Z_{0,T}(0) = 1$, to get

$$\begin{aligned} Z_{0,T}(v)X_v &= X_0 + \int_0^v (\text{diag}[\beta(u,T)] + A(u))Z_{0,T}(u)X_u\, du \\ &\quad + \int_0^v Z_{0,T}(u)\, dM_u. \end{aligned} \tag{4.33}$$

Now, since $Z_{0,T}(\cdot)$ is left continuous and adapted, hence predictable, and bounded on $[0,T]$, and M is square integrable, $\int_0^{\cdot} Z_{0,T}(u)\, dM_u$ is a zero-mean martingale. Taking expectation and applying Fubini's theorem gives

$$E[Z_{0,T}(v)X_v] = X_0 + \int_0^v (\text{diag}[\beta(u,T)] + A(u))E[Z_{0,T}(u)X_u]\, du. \tag{4.34}$$

Since a does not depend on X, $\beta(u,T)$ is deterministic. This is a homogeneous linear system of ordinary differential equations, and so it has the unique fundamental solution

$$E[Z_{0,T}(v)X_v] = \Phi_\beta(v;0,T)X_0, \tag{4.35}$$

where $\Phi_\beta(v;0,T)$ is the fundamental matrix for the N-dimensional linear system

$$y'(v) = (\text{diag}[\beta(v,T)] + A(v))y(v). \tag{4.36}$$

Note that requiring a to be independent of X allows us to easily find a matrix, (the term in parentheses), that is independent of y, and hence we obtain a linear system. Define $\Phi_\beta(0,T) = \Phi_\beta(T;0,T)$, so that

$$E[Z_{0,T}X_T] = \Phi_\beta(0,T)X_0. \tag{4.37}$$

Remark 11 By taking the derivative with respect to T on both sides of Equation 4.35, and noticing that C is bounded and continuously differentiable with respect to T, we can see that the matrix Φ_β is continuously differentiable with respect to T.

Remark 12 For every scalar, $x \in \mathbf{R}$, and for every vector, e_j of the canonical basis of $\mathbf{R}^n$,

$$x = 1^{\mathsf{T}}(xe_j), \tag{4.38}$$

where $\tilde{1} \in \mathbf{R}^n$ is the vector with 1 in every entry.

We can use Equation 4.38 and the fact that X_T is a unit vector to get

$$E[Z_{0,T}] = \tilde{1}^\mathsf{T}\Phi_\beta(0,T)X_0, \tag{4.39}$$

so that, upon substituting this into Equation 4.24, the bond price at time $t = 0$ is

$$B(0,T) = \exp(-r_0 C(0,T))\tilde{1}^\mathsf{T}\Phi_\beta(0,T)X_0. \tag{4.40}$$

This gives the bond price in terms of the functions C, and Φ_β, both of which can be readily approximated using numerical methods, if not calculated explicitly, when the given data from Equation 4.1 has fairly weak restrictions. We will now find the price of the bond for intermediate values of t.

We will suppose that the Markov chain, X, is observable, so that we may use the information, $\mathcal{F}_t^X$, in calculating the bond price at time t. In this case the price of the bond at the intermediate time $t \in [0,T]$ is

$$B(t,T) = S_t^0 E\Big[\tfrac{1}{S_T^0}\ \Big|\ \mathcal{F}_t^r \vee \mathcal{F}_t^X\Big] = E[e^{-R_{t,T}}|\mathcal{F}_t^r \vee \mathcal{F}_t^X] \quad a.s. \tag{4.41}$$

To evaluate this conditional expectation we use the same trick as before by conditioning first on $\mathcal{F}_t^r \vee \mathcal{F}_T^X$ and then on the desired smaller σ-algebra. Now, by Lemma 10, we have, almost surely

$$\begin{aligned} E[\exp(-R_{t,T})|\mathcal{F}_t^r \vee \mathcal{F}_T^X] &= \exp\Big(-r_t C(t,T) + \int_t^T \beta(u,T,X_u)\,du\Big) \\ &= \exp(-r_t C(t,T))Z_{t,T}, \end{aligned} \tag{4.42}$$

where, C and β are as before and

$$Z_{t,T} = \exp\Big(\int_t^T \beta(u,T,X_u)\,du\Big) = \frac{Z_{0,T}}{Z_{0,T}(t)}, \tag{4.43}$$

and where $Z_{0,T}(t)$ is defined above. Now r_t and $Z_{0,T}(t)$ are both $\mathcal{F}_t^r \vee \mathcal{F}_t^X$-measurable, so the only difficulty in obtaining the time t bond price is determining $E[Z_{0,T}|\mathcal{F}_t^r \vee \mathcal{F}_t^X]$. Retaining the notation introduced previously, recall that, for $0 \leq t \leq v \leq T$,

$$\begin{aligned} Z_{0,T}(v)X_v &= Z_{0,T}(0)X_0 + \int_0^v (\text{diag}[\beta(u,T)] + A(u))Z_{0,T}(u)X_u\,du \\ &\quad + \int_0^v Z_{0,T}(u)\,dM_u \qquad (4.44) \\ &= Z_{0,T}(t)X_t + \int_t^v (\text{diag}[\beta(u,T)] + A(u))Z_{0,T}(u)X_u\,du \\ &\quad + \int_t^v Z_{0,T}(u)\,dM_u. \end{aligned}$$

Because, with the conditions observed above, a stochastic integral with respect to a martingale, is a martingale,

$$\begin{aligned} & E\Big[\int_t^v Z_{0,T}(u)\,dM_u \,\Big|\, \mathcal{F}_t^{\tau} \vee \mathcal{F}_t^X\Big] \qquad (4.45) \\ &= E\Big[E\Big[\int_t^v Z_{0,T}(u)\,dM_u \,\Big|\, \mathcal{F}_T^W \vee \mathcal{F}_t^X\Big] \,\Big|\, \mathcal{F}_t^{\tau} \vee \mathcal{F}_t^X\Big] \\ &= 0 \quad a.s., \end{aligned}$$

thus

$$\begin{aligned} & E[Z_{0,T}(v)X_v|\mathcal{F}_t^{\tau} \vee \mathcal{F}_t^X] - Z_{0,T}(t)X_t \qquad (4.46) \\ &= \int_t^v (\text{diag}[\beta(u,T)] + A(u))E[Z_{0,T}(u)X_u|\mathcal{F}_t^{\tau} \vee \mathcal{F}_t^X]\,du \quad a.s. \end{aligned}$$

For almost every $\omega \in \Omega$, with ω fixed, the above expression is a linear ordinary differential equation starting at time t. It has the unique solution

$$E[Z_{0,T}(v)X_v|\mathcal{F}_t^{\tau} \vee \mathcal{F}_t^X] = \Phi_\beta(v;t,T)Z_{0,T}(t)X_t \quad a.s., \qquad (4.47)$$

where $\Phi_\beta(v;t,T)$ is the fundamental matrix for Equation 4.36, starting at time t, hence

$$\Phi_\beta(v;t,T) = \Phi_\beta(v;0,T)\Phi_\beta^{-1}(t;0,T). \qquad (4.48)$$

Dividing both sides by $Z_{0,T}(t)$, (which is positive a.s.), and defining

$$\Phi_\beta(t,T) = \Phi_\beta(T;t,T) \qquad (4.49)$$

gives

$$E[Z_{t,T}X_T \mid \mathcal{F}_t^X \vee \mathcal{F}_t^r] = \Phi_\beta(t,T)X_t \quad a.s. \tag{4.50}$$

Finally, from this analysis, we obtain the bond price at time $t \in [0,T]$, when we can observe X and r:

$$B(t,T) = \exp(-r_t C(t,T))\tilde{1}^\mathsf{T}\Phi_\beta(t,T)X_t. \tag{4.51}$$

This gives an analytical solution for the price of a zero-coupon bond with maturity T for any time $t \in [0,T]$, depending on the solution to a linear ordinary differential equation. We have derived the following theorem.

Theorem 4 *If P is a risk-neutral probability and the risk free short-term interest rate is characterised by Equation 4.1, and the Markov chain X is observable, then the price of a zero-coupon bond that matures at time T, at any time $t \in [0,T]$ is given by*

$$B(t,T) = \exp(-r_t C(t,T))\tilde{1}^\mathsf{T}\Phi_\beta(t,T)X_t, \tag{4.52}$$

where $\Phi_\beta(t,T)$ is given by evaluating the fundamental matrix solution of the ODE 4.36 at two points as described.

Remark 13 Even if the existence of solutions for homogeneous linear systems of differential equations is guaranteed, in general they can't be solved explicitly, as they can in the one-dimensional case. In fact, even the simplest case of constant parameters seems to require numerical analysis.

Remark 14 We have assumed here, that the Markov chain, X, is observable, and thus the correct bond price is obtained by using the information $\mathcal{F}^X$ in addition to the information $\mathcal{F}^r$. It is more realistic to believe that the Markov chain is not observable, and thus the bond price should be obtained by using only the information contained in the filtration $\mathcal{F}^r$. In theory, this does not pose a serious problem since $\mathcal{F}_t^r \subset \mathcal{F}_t^r \vee \mathcal{F}_t^X$ for every $t \in [0,T]$ and we can use the law of iterated projections to obtain the bond price in the unobservable case from the price we obtained in the observable case,

$$B(t,T) = \exp(-r_t C(t,T))\tilde{1}^\mathsf{T}\Phi_\beta(t,T)E[X_t|\mathcal{F}_t^r]. \tag{4.53}$$

It should be noted that the above conditional expectation represents the vector of the conditional probability distribution for the Markov chain, (namely the entries are non-negative and they sum to 1). Methods for estimating the conditional expectation, (or filtering X), will be provided in Chapter 5. We will maintain the assumption that X is observable throughout remainder of this chapter; however, remarks similar to this can be applied to the other results we obtain in an obvious way.

4.3 Dynamics for the Bond Price

We know that under the Hull-White term structure model, zero-coupon bonds have lognormal type dynamics described by the following SDE: (for T fixed and $t \in [0, T]$)

$$dB(t,T) = r_t B(t,T)\,dt - C(t,T)\sigma(t)B(t,T)\,dW_t. \tag{4.54}$$

Knowledge about the dynamics is important for calibrating the model, since estimating bond volatility gives an estimate for the product function $C(t,T)\sigma(t)$. If σ can be estimated by the short rate model, then the bond volatility determines C, and hence $a(t)$, (by differentiation). This knowledge, and the knowledge of bond prices at every maturity T, allows us to form an ordinary differential equation for $\bar{r}(t)$ which can be solved numerically. This is how the model can be calibrated to fit any yield curve exactly. The dynamics of the bond price can also be used to model the whole term structure, and this can be used to explore relationships between interest rates of different maturities. The remainder of this section is devoted to obtaining the dynamics of the bond price for our model. The main result is stated in the following theorem.

Theorem 5 *If P is a risk-neutral probability and the risk free short-term interest rate is characterised by Equation 4.1, and the Markov chain X is observable, then the dynamics of the price of a zero-coupon bond with maturity T is given by, for $t \in [0, T]$,*

$$dB(t,T) = r_t B(t,T)\,dt - C(t,T)\sigma(t,X_t)B(t,T)\,dW_t + B(t,T)^{\mathsf{T}}\,dM_t. \tag{4.55}$$

Here $B(t,T)$ is the vector in $\mathbf{R}^N$ with typical entry

$$B_i(t,T) = \exp(-r_t C(t,T))\tilde{1}^{\mathsf{T}}\Phi_\beta(t,T)e_i, \tag{4.56}$$

or in other words it is the bond price that would have resulted if X_t was actually in state e_i.

Remark 15 It is interesting to note that this is similar to the dynamics of the bond price in the Hull-White model, except the dM term is added. This could be interpreted as additional volatility for the bond price, and could result in a distribution with fatter tails than normal. This would imply an implied volatility smile for option prices simulated using this model , but priced using a best fit Vasiček model.

In the previous section, we determined the price of a zero coupon bond at any intermediate time, in terms of the deterministic and continuously differentiable functions C and Φ_β. Writing this explicitly as a function of r and X, we have

$$B(t,T,r,X) = \exp(-rC(t,T))\tilde{1}^{\mathsf{T}}\Phi_\beta(t,T)X. \tag{4.57}$$

Because it depends on X, B is right continuous, but not continuous. To find the dynamics of B, we apply a generalised version of Itô's rule that takes the discontinuities of X into account. This involves the notion of random measure. We shall proceed to determine the dynamics indirectly.

Define the stochastic process, $V = \{V_t; t \in [0,T]\}$, by

$$V_t = E\Big[\exp\Big(-\int_0^T r_u\,du\Big)\;\Big|\;\mathcal{F}_t^r \vee \mathcal{F}_t^X\Big]. \tag{4.58}$$

It is clear that V is a martingale. Furthermore, by the definition of B,

$$V_t = \exp\Big(-\int_0^t r_u\,du\Big)B(t,T,r_t,X_t), \tag{4.59}$$

so we have the explicit representation $V_t = V(t,T,r_t,X_t)$. Applying the generalised Itô's rule to V gives

$$V(t,T,r_t,X_t) - V(0,T,r_0,X_0) \tag{4.60}$$

$$
\begin{aligned}
= & \int_0^t \tfrac{\partial}{\partial u} V(u,T,r_u,X_{u^-})\,du + \int_0^t \tfrac{\partial}{\partial r} V(u,T,r_u,X_{u^-})\,dr_u \\
& + \int_0^t \tfrac{\partial}{\partial X} V(u,T,r_u,X_{u^-})\,dX_u + \tfrac{1}{2}\int_0^t \tfrac{\partial^2}{\partial r^2} V(u,T,r_u,X_{u^-})\,d\langle r\rangle_u \\
& + \tfrac{1}{2}\int_0^t \tfrac{\partial^2}{\partial X^2} V(u,T,r_u,X_{u^-})\,d\langle X\rangle_u \\
& + \int_0^t \tfrac{\partial^2}{\partial r\,\partial X} V(u,T,r_u,X_{u^-})\,d\langle r,X\rangle_u \\
& + \sum_{0<u\le t}\Big(V(u,T,r_u,X_u) - V(u,T,r_u,X_{u^-}) \\
& - \tfrac{\partial}{\partial X} V(u,T,r_u,X_{u^-})\Delta X_u\Big),
\end{aligned}
$$

where $\Delta X_u = X_u - X_{u^-}$, and the summation is over all points of discontinuity. We use the representation $V(t,T,r_t,X_t) = V(t,T,r_t)^{\mathsf{T}} X_t$. From this representation, it is clear that the first partial derivative with respect to X is independent of X, and hence the second partial is zero. Also, because X is constant except for jump discontinuities that occur discretely in time, and r is continuous, the cross variation, $\langle r,X\rangle$ is zero, (incidentally, this also implies that $\langle X\rangle$ is zero). Finally, we can use this representation to show that the summation on the last line is zero:

$$
\begin{aligned}
& \sum_{0<u\le t}\Big(V(u,T,r_u,X_u) - V(u,T,r_u,X_{u^-}) - \tfrac{\partial}{\partial X} V(u,T,r_u,X_{u^-})\Delta X_u\Big) \\
= & \sum_{0<u\le t}\Big(V(u,T,r_u)^{\mathsf{T}} X_u - V(u,T,r_u)^{\mathsf{T}} X_{u^-} - V(u,T,r_u)^{\mathsf{T}}\Delta X_u\Big) \\
= & \sum_{0<u\le t}\Big(V(u,T,r_u)^{\mathsf{T}}\Delta X_u - V(u,T,r_u)^{\mathsf{T}}\Delta X_u\Big) \qquad (4.61) \\
= & \ 0.
\end{aligned}
$$

Taking this into account and expanding the dr and dX terms yields

$$
\begin{aligned}
V(t,T,r_t,X_t) = & \ V(0,T,r_0,X_0) + \int_0^t \tfrac{\partial}{\partial u} V(u,T,r_u,X_{u^-})\,du \\
& + \int_0^t \tfrac{\partial}{\partial r} V(u,T,r_u,X_{u^-})\,a(u)(\bar{r}(u,X_u) - r_u)\,du \qquad (4.62) \\
& + \int_0^t \tfrac{\partial}{\partial X} V(u,T,r_u,X_{u^-})\,dM_u + \int_0^t \tfrac{\partial}{\partial X} V(u,T,r_u,X_{u^-})\,A(u)X_u\,du
\end{aligned}
$$

$$+\int_0^t \tfrac{\partial}{\partial r}V(u,T,r_u,X_{u^-})\,\sigma(u,X_u)\,dW_u$$
$$+\tfrac{1}{2}\int_0^t \tfrac{\partial^2}{\partial r^2}V(u,T,r_u,X_{u^-})\sigma^2(u,X_u)\,du$$

Since V is a martingale, only the dW and dM integrals remain, while the du terms sum to zero,

$$\begin{aligned} V(t,T,r_t,X_t) &= V(0,T,r_0,X_0) + \int_0^t \tfrac{\partial}{\partial r}V(u,T,r_u,X_u)\,\sigma(u,X_u)\,dW_u \\ &\quad + \int_0^t \tfrac{\partial}{\partial X}V(u,T,r_u,X_u)\,dM_u. \end{aligned} \tag{4.63}$$

The partial derivatives of V are determined using Equation 4.59, as follows:

$$\begin{aligned} \tfrac{\partial}{\partial r}V(t,T,r_t,X_t) &= \exp\Big(-\int_0^t r_u\,du\Big)\tfrac{\partial}{\partial r}B(t,T,r_t,X_t) \qquad (4.64)\\ &= -\exp\Big(-\int_0^t r_u\,du\Big)C(t,T)B(t,T,r_t,X_t) \end{aligned}$$

and

$$\begin{aligned} \tfrac{\partial}{\partial X}V(u,T,r_u,X_u) &= \exp\Big(-\int_0^t r_u\,du\Big)\tfrac{\partial}{\partial X}B(t,T,r_t,X_t) \qquad (4.65)\\ &= \exp\Big(-\int_0^t r_u\,du\Big)\exp(-r_tC(t,T))\tilde{1}^{\mathsf{T}}\Phi_\beta(t,T). \end{aligned}$$

Note that, for the partial derivative with respect to r, the first term in Equation 4.59 seems to depend on r since it includes the integral of r_u with respect to u. Therefore one might expect to apply the product rule for differentiation. However, the r with respect to which we are differentiating represents the variable r_t. Since this variable, r_t, occurs with Lebesgue measure zero in the integral, the first term is independent of r_t and the partial derivative is as claimed. Substituting these expressions into Equation 4.63 gives

$$\begin{aligned} V(t,T,r_t,X_t) &= V(0,T,r_0,X_0) \\ &\quad + \int_0^t -C(u,T)\sigma(u,X_u)\exp\Big(-\int_0^u r_s\,ds\Big)B(u,T,r_u,X_u)\,dW_u \\ &\quad + \int_0^t \exp\Big(-\int_0^u r_s\,ds\Big)\exp(-r_uC(u,T))\tilde{1}^{\mathsf{T}}\Phi_\beta(u,T)\,dM_u. \end{aligned} \tag{4.66}$$

We proceed by rearranging Equation 4.59 to obtain

$$B(t,T,r_t,X_t) = \exp\Big(\int_0^t r_u\,du\Big)V(t,T,r_t,X_t). \tag{4.67}$$

Itô's integration by parts and Equation 4.66 give

$$\begin{aligned} B(t,T,r_t,X_t) &= B(0,T,r_0,X_0) + \int_0^t \exp\Big(\int_0^u r_s\,ds\Big)\,dV(u,T,r_u,X_u) \\ &\quad + \int_0^t V(u,T,r_u,X_u)\exp\Big(\int_0^u r_s\,ds\Big)r_u\,du \qquad (4.68) \\ &= B(0,T,r_0,X_0) - \int_0^t C(u,T)\sigma(u,X_u)B(u,T,r_u,X_u)\,dW_u \\ &\quad + \int_0^t \exp(-r_u C(u,T))\tilde{1}^\mathsf{T}\Phi_\beta(u,T)\,dM_u + \int_0^t r_u B(u,T,r_u,X_u)\,du \\ &= B(0,T,r_0,X_0) + \int_0^t r_u B(u,T,r_u,X_u)\,du \\ &\quad - \int_0^t C(u,T)\sigma(u,X_u)B(u,T,r_u,X_u)\,dW_u, + \int_0^t B(u,T,r_u)^\mathsf{T}\,dM_u \end{aligned}$$

where $B(t,T,r_t) = (B(t,T,r_t,e_1),\ldots,B(t,T,r_t,e_N))^\mathsf{T}$. This gives the dynamics for the bond price.

4.4 Pricing Derivatives of a Bond

In this section we wish to determine the price of a European contingent claim with expiration time, T_1, on a zero-coupon bond with maturity, $T_2 \geq T_1$. Since P is a risk neutral probability measure, a no arbitrage price for the contingent claim at any intermediate time, $t \in [0,T_1]$, is (with the typical abuse of notation)

$$\begin{aligned} h(t,T_1,T_2) &= E\Big[\exp\Big(-\int_t^{T_1} r_u\,du\Big)h(B(T_1,T_2))\;\Big|\;\mathcal{F}_t^r \vee \mathcal{F}_t^X\Big] \qquad (4.69) \\ &= E\Big[e^{-R_{t,T_1}}h\Big(\exp(-r_{T_1}C(T_1,T_2))\tilde{1}^\mathsf{T}\Phi_\beta(T_1,T_2)X_{T_1}\Big)\;\Big|\;\mathcal{F}_t^r \vee \mathcal{F}_t^X\Big]. \end{aligned}$$

To evaluate this expectation we need to find the conditional joint distribution of r_T, $R_{t,T}$, and X_T given $\mathcal{F}_t^r \vee \mathcal{F}_t^X$. Unfortunately, obtaining the joint distributions directly is quite complicated, so we content ourselves with obtaining their moment generating functions. For this we use a method similar to that used to obtain the bond price. Defining the conditional moment generating function as

$$\text{c.m.g.f.}(\theta_1,\theta_2,\theta_3) = E[\exp(\theta_1 r_T + \theta_2 R_{0,T} + \theta_3^\mathsf{T} X_T)|\mathcal{F}_t^r \vee \mathcal{F}_t^X], \quad (4.70)$$

where $R_{0,T}$ is defined in Lemma 9, we have the following theorem.

Theorem 6 *If P is a risk-neutral probability and the risk free short-term interest rate is characterised by Equation 4.1, then*

$$\text{c.m.g.f.}(\theta_1,\theta_2,\theta_3) = \exp\Big(r_t(\theta_1 e^{-K(T)}e^{K(t)} + \theta_2 C(t,T)\,du)\Big)\tilde{1}^\mathsf{T}\Theta(t,T)X_t, \quad (4.71)$$

where $\Theta(t,T)$, which implicitly depends on θ_1, θ_2, and θ_3, is given by evaluating the fundamental matrix solution of the ODE 4.86 (below) at two points as described below.

As is our custom, we begin by looking at the $t=0$ case first. From Lemma 11, we know that r_T, and $R_{0,T}$ are jointly conditionally bivariate normal given $\mathcal{F}_T^X$. Using the law of iterated projections, we calculate the unconditional moment generating function as

$$\begin{aligned}\text{m.g.f.}(\theta_1,\theta_2,\theta_3) &= E[\exp(\theta_1 r_T + \theta_2 R_{0,T} + \theta_3^\mathsf{T} X_T)] \quad (4.72)\\ &= \exp\Big(r_0\Big(\theta_1 e^{-K(T)} + \theta_2 \int_0^T e^{-K(u)}\,du\Big)\Big)E[Y_{0,T}],\end{aligned}$$

where

$$Y_{0,T} = \exp\Big(\int_0^T \Big(\eta(u,T)^\mathsf{T} X_u + e^{-K(u)}\int_0^u \xi(s,T)^\mathsf{T} X_s\,ds + \tfrac{1}{T}\theta_3^\mathsf{T} X_T\Big)\,du\Big), \quad (4.73)$$

with

$$\begin{aligned}\eta(u,T) &= \theta_1 e^{-K(T)}e^{K(u)}a(u)\bar{r}(u) + \tfrac{1}{2}\theta_1^2 e^{-2K(T)}e^{2K(u)}\sigma^2(u)\\ &\quad + \tfrac{1}{2}\theta_2^2 e^{2K(u)}\Big(\int_u^T e^{-2K(x)}\,dx\Big)^2\sigma^2(u), \quad (4.74)\end{aligned}$$

and

$$\xi(s,T) = \theta_2 e^{K(s)}a(s)\bar{r}(s) + \theta_1\theta_2 e^{-K(T)}e^{2K(s)}\sigma^2(s). \quad (4.75)$$

(We chose these representations for a reason that will become apparent momentarily.) The situation is now similar to the one we encountered when finding the bond price, except now we have a second integral of X inside the first integral. This poses a problem because we can't use the handy result from Equation 4.32 to obtain a solution in a similar way. One way to get around this problem is to rewrite the second term in the exponential using integration by parts, as follows:

$$\begin{aligned}\int_0^T e^{-K(u)} \int_0^u \xi(s,T)^\mathsf{T} X_s \, ds \, du &= \int_0^T \int_0^u \xi(s,T)^\mathsf{T} X_s \, ds \, dC(0,u) \\ &= C(0,T) \int_0^T \xi(u,T)^\mathsf{T} X_u \, du - \int_0^T C(0,u)\xi(u,T)^\mathsf{T} X_u \, du, \quad (4.76)\end{aligned}$$

where we recall that $C(0,T) = \int_0^T e^{-K(u)} \, du$. Substituting this into the expression for $Y_{0,T}$ gives

$$\begin{aligned}Y_{0,T} &= \exp\Big(\int_0^T \{\eta(u,T)^\mathsf{T} X_u + [C(0,T) - C(0,u)]\xi(u,T)^\mathsf{T} X_u \\ &\quad + \tfrac{1}{T}\theta_3^\mathsf{T} X_T\} \, du\Big). \qquad (4.77)\end{aligned}$$

To simplify notation, let

$$\lambda(u,T) = \eta(u,T) + [C(0,T) - C(0,u)]\xi(u,T), \qquad (4.78)$$

so that

$$Y_{0,T} = \exp\Big(\int_0^T \{\lambda(u,T)^\mathsf{T} X_u + \tfrac{1}{T}\theta_3^\mathsf{T} X_T\} \, du\Big). \qquad (4.79)$$

Note that the dependence of λ on θ_1 and θ_2 has been suppressed. Now define, for all $v \in [0,T]$,

$$Y_{0,T}(v) = \exp\Big(\int_0^v (\lambda(u,T)^\mathsf{T} X_u + \tfrac{1}{T}\theta_3^\mathsf{T} X_v) \, du\Big). \qquad (4.80)$$

Then $Y_{0,T} = Y_{0,T}(T)$, and $Y_{0,T}(\cdot)$ is adapted to the filtration, $\mathcal{F}^X$. Also,

$$\begin{aligned}dY_{0,T}(v) &= (\lambda(v,T)^\mathsf{T} X_v + \tfrac{1}{T}\theta_3^\mathsf{T} X_v) Y_{0,T}(v) \, dv + \tfrac{v}{T} Y_{0,T}(v)\theta_3^\mathsf{T} \, dX_v \\ &= \Big(\lambda(v,T)^\mathsf{T} X_v + \tfrac{1}{T}\theta_3^\mathsf{T} X_v + \tfrac{v}{T}\theta_3^\mathsf{T} A(v) X_v\Big) Y_{0,T}(v) \, dv \\ &\quad + \tfrac{v}{T} Y_{0,T}(v)\theta_3^\mathsf{T} \, dM_v. \qquad (4.81)\end{aligned}$$

This implies that, for Lebesgue-almost every $v \in [0, T]$, by Itô 's rule

$$d(Y_{0,T}(v)X_v) = X_{v^-}\, dY_{0,T}(v) + Y_{0,T}(v^-)\, dX_v \tag{4.82}$$
$$= \Big(\lambda(v,T)^{\mathsf{T}} X_v + \tfrac{1}{T}\theta_3^{\mathsf{T}} X_v + \tfrac{v}{T}\theta_3^{\mathsf{T}} A(v) X_v\Big) Y_{0,T}(v) X_{v^-}\, dv$$
$$+ \tfrac{v}{T} Y_{0,T}(v) X_{v^-}\theta_3^{\mathsf{T}}\, dM_v + A(v) Y_{0,T}(v) X_v\, dv + Y_{0,T}(v)\, dM_v$$
$$= \Big(\mathrm{diag}[\lambda(v,T)] + \tfrac{1}{T}\mathrm{diag}[\theta_3] + \tfrac{v}{T}\mathrm{diag}[\theta_3^{\mathsf{T}} A(v)] + A(v)\Big) Y_{0,T}(v) X_v\, dv$$
$$+ \Big(\tfrac{v}{T} X_v \theta_3^{\mathsf{T}} + I\Big) Y_{0,T}(v)\, dM_v.$$

Integrating gives

$$Y_{0,T}(v)X_v = X_0 + \int_0^v \Big(\tfrac{u}{T} X_u \theta_3^{\mathsf{T}} + I\Big) Y_{0,T}(u)\, dM_u, \tag{4.83}$$
$$+ \int_0^v \Big(\mathrm{diag}[\lambda(u,T)] + \tfrac{1}{T}\mathrm{diag}[\theta_3] + \tfrac{u}{T}\mathrm{diag}[\theta_3^{\mathsf{T}} A(u)] + A(u)\Big)$$
$$\times Y_{0,T}(u) X_u\, du$$

and taking expectation, (and using Fubini's theorem), yields

$$E[Y_{0,T}(v)X_v] = X_0 \tag{4.84}$$
$$+ \int_0^v \Big(\mathrm{diag}[\lambda(u,T)] + \tfrac{1}{T}\mathrm{diag}[\theta_3] + \tfrac{u}{T}\mathrm{diag}[\theta_3^{\mathsf{T}} A(u)] + A(u)\Big)$$
$$\times E[Y_{0,T}(u) X_u]\, du.$$

Again, we have a linear ordinary differential equation, which has a unique solution, say

$$E[Y_{0,T}(v)X_v] = \Theta(v; 0, T) X_0, \tag{4.85}$$

where $\Theta(v; 0, T)$ is the fundamental matrix for the N-dimensional ordinary differential equation

$$y'(v) = \Big(\mathrm{diag}[\lambda(v,T)] + \tfrac{1}{T}\mathrm{diag}[\theta_3] + \tfrac{v}{T}\mathrm{diag}[\theta_3^{\mathsf{T}} A(v)] + A(v)\Big) y(v). \tag{4.86}$$

Let $\Theta(0, T) = \Theta(0, T, T)$, so $E[Y_{0,T} X_T] = \Theta(0, T) X_0$, and hence

$$E[Y_{0,T}] = \tilde{1}^{\mathsf{T}} \Theta(0, T) X_0. \tag{4.87}$$

Therefore, substituting this into Equation 4.72 gives the multi-variate moment generating function,

$$\text{m.g.f.}(\theta_1, \theta_2, \theta_3) = \exp\Big(r_0(\theta_1 e^{-K(T)} + \theta_2 C(0,T))\Big) \tilde{1}^{\mathsf{T}} \Theta(0, T) X_0, \tag{4.88}$$

where we should point out that Θ depends on θ_1, θ_2, and θ_3. We can obtain the joint distribution of r_T, R_T, and X_T, which we call $F_{0,T}$, by obtaining the inverse of the Laplace transform. Returning to Equation 4.69, the value of the derivative is

$$h(0,T_1,T_2) = \int e^{-y} h\Big(\exp(-C(T_1,T_2)x)\tilde{1}^\top \Phi_\beta(T_1,T_2)z\Big)\, dF_{0,T_1}(x,y,z). \tag{4.89}$$

We now turn our attention to the case of an intermediate $t \in [0,T_1]$, where, for convenience, we drop the subscript 1, and we calculate the conditional joint moment generating function of r_T, $R_{t,T}$, and X_T given $\mathcal{F}_t^r \vee \mathcal{F}_t^X$. The derivation of this case is similar to the previous one; we again use the tower property, and first condition on the larger σ-algebra, $\mathcal{F}_t^r \vee \mathcal{F}_T^X$, which allows us to apply Lemma 10 to obtain a useful representation. We have,

$$\begin{aligned} \text{c.m.g.f.}(\theta_1,\theta_2,\theta_3) &= E[\exp(\theta_1 r_T + \theta_2 R_{0,T} + \theta_3^\top X_T)|\mathcal{F}_t^r \vee \mathcal{F}_t^X] \quad (4.90)\\ &= \exp\Big(e^{K(t)} r_t\Big(\theta_1 e^{-K(T)} + \theta_2 \int_t^T e^{-K(u)}\, du\Big)\Big) E[Y_{t,T}|\mathcal{F}_t^r \vee \mathcal{F}_t^X], \end{aligned}$$

where

$$Y_{t,T} = \exp\Big(\int_t^T \Big(\eta(u,T)^\top X_u + e^{-K(u)} \int_t^u \xi(s,T)^\top X_s\, ds + \tfrac{1}{T}\theta_3^\top X_T\Big)\, du\Big). \tag{4.91}$$

We may again use integration by parts on the second term to find that

$$Y_{t,T} = \exp\Big(\int_t^T \big(\lambda(u,T)^\top X_u\, du + \tfrac{1}{T}\theta_3^\top X_T\big)\, du\Big) = \frac{Y_{0,T}}{Y_{0,T}(t)}. \tag{4.92}$$

Since $Y_{0,T}(t)$ is $\mathcal{F}_t^r \vee \mathcal{F}_t^X$-measurable, the denominator can be taken out of the conditional expectation, and we may concentrate on determining $E[Y_{0,T}|\mathcal{F}_t^r \vee \mathcal{F}_t^X]$. By a simple extension of Equation 4.83, we have, for $0 \le t \le v \le T$,

$$\begin{aligned} Y_{0,T}(v)X_v &= Y_{0,T}(t)X_t + \int_t^v (\tfrac{u}{T} X_u \theta_3^\top + I) Y_{0,T}(u)\, dM_u \quad (4.93)\\ &\quad + \int_t^v \Big(\text{diag}[\lambda(u,T)] + \tfrac{1}{T}\text{diag}[\theta_3] + \tfrac{u}{T}\text{diag}[\theta_3^\top A(u)] + A(u)\Big)\\ &\quad \times Y_{0,T}(u) X_u\, du. \end{aligned}$$

Taking the conditional expectation causes the dM term to vanish and using Fubini's theorem gives

$$
\begin{aligned}
E[Y_{0,T}(v)X_v|\mathcal{F}_t^r \vee \mathcal{F}_t^X] &= Y_{0,T}(t)X_t \\
&+ \int_t^v \Big(\mathrm{diag}[\lambda(u,T)] + \tfrac{1}{T}\mathrm{diag}[\theta_3] + \tfrac{u}{T}\mathrm{diag}[\theta_3^\mathsf{T} A(u)] + A(u)\Big) \\
&\times E[Y_{0,T}(u)X_u|\mathcal{F}_t^r \vee \mathcal{F}_t^X]\, du.
\end{aligned}
\tag{4.94}
$$

This ordinary differential equation can be solved point wise for almost every $\omega \in \Omega$. Because the only randomness inside the integrand comes from the $E[Y_{0,T}(u)X_u|\mathcal{F}_t^r \vee \mathcal{F}_t^X]$ factor, each ω corresponds to the same fundamental matrix for the associated solution. This turns out to be the fundamental matrix corresponding to the same ordinary differential equation 4.86 as before, except now starting at time t. Since

$$
y(v) = \Theta(v;0,T)y(0) = \Theta(v;0,T)\Theta^{-1}(t;0,T)y(t), \tag{4.95}
$$

this fundamental matrix is $\Theta(0,v,T)\Theta^{-1}(0,t,T)$, so, letting

$$
\Theta(t,T) = \Theta(T;0,T)\Theta^{-1}(t;0,T), \tag{4.96}
$$

we have

$$
E[Y_{0,T}X_T|\mathcal{F}_t^r \vee \mathcal{F}_t^X] = \Theta(t,T)Y_{0,T}(t)X_t \quad a.s. \tag{4.97}
$$

Therefore,

$$
E[Y_{t,T}|\mathcal{F}_t^r \vee \mathcal{F}_t^X] = \frac{1}{Y_{0,T}(t)}\tilde{1}^\mathsf{T}\Theta(t,T)Y_{0,T}(t)X_t = \tilde{1}^\mathsf{T}\Theta(t,T)X_t \quad a.s. \tag{4.98}
$$

This means that the conditional joint moment generating function of r_T, $R_{t,T}$, and X_T given $\mathcal{F}_t^r \vee \mathcal{F}_t^X$ is

$$
\text{c.m.g.f.}(\theta_1,\theta_2,\theta_3) = \exp\Big(r_t(\theta_1 e^{-K(T)}e^{K(t)} + \theta_2 C(t,T)\,du)\Big)\tilde{1}^\mathsf{T}\Theta(t,T)X_t. \tag{4.99}
$$

In theory, we can obtain the conditional joint distribution, call it $F_{t,T}$, from the inverse Laplace transform, and use this to evaluate the price of the derivative at an intermediate time t. From Equation 4.69, this price can be expressed as

$$
h(t,T_1,T_2) = \int e^{-y} h\Big(\exp(-xC(T_1,T_2))\tilde{1}^\mathsf{T}\Phi_\beta(T_1,T_2)z\Big)\, dF_{t,T_1}(x,y,z). \tag{4.100}
$$

This page is intentionally left blank.

Chapter 5

Filtering the Hidden Markov Model

This chapter develops the methodological theory required to estimate the parameters of the hidden Markov model. The filtering method outlined here finds maximum likelihood estimates. We now assume that the stochastic processes evolve in discrete time, so we start by revising the required theory to pertain to a discrete time Markov chain.

It examines estimation techniques on a parametric model where the parameters themselves are prone to switch in accordance with a finite Markov chain. We consider the case of a discrete time autoregressive stochastic process and find exact adaptive filters. The technique employs the expectation maximisation (EM) algorithm to obtain, (or at least approach), true maximum likelihood parameter estimates. This contrasts with the usual estimation approach in this situation of quasi-maximum likelihood, despite the fact that the assumption of normal i.i.d. errors is plainly violated due to the dependence on the Markov chain. The approach we follow here is closely related to Elliott [12], and is in fact an extension of that model.

5.1 The Model

We begin by hypothesising the existence of a probability space $(\Omega, \mathcal{F}, P)$ on which the various stochastic processes reside. Here P represents the true probability of events, (as opposed to various constructed probability measures that follow). The first process is an N-dimensional, discrete time homogeneous Markov chain, $X = \{X_n; n \in \mathbf{N} = \{0, 1, 2, ...\}\}$, that takes values in the set of unit (column) vectors, $S = \{e_1, ..., e_N\}$, which is the canonical basis of $\mathbf{R}^N$. Denote by $\mathcal{F}^X = \{\mathcal{F}_n^X\}$ the filtration generated by the Markov chain X, and $P = [p_{ji}]; i, j \in \{1, ..., N\}$ its transition matrix, where $p_{ji} = P(X_n = e_j | X_{n-1} = e_i)$. It follows that $E[X_n | X_{n-1}] = PX_{n-1}$, where the conditional expectation gives the vector of conditional probabilities and the right hand side picks out the appropriate column of P. (Note that the entries of P must be non-negative and that the columns must sum to 1.) This motivates our choice of state space and stochastic matrix notation, which was done without any loss of generality. The following lemma is a direct application of this description of the probability distribution.

Lemma 13 *The Markov chain, X, has the following semi-martingale representation:*

$$X_n = PX_{n-1} + m_n, \tag{5.1}$$

where m is a martingale increment. I.e. $E[m_n | \mathcal{F}_{n-1}^X] = 0$.

We also have the following lemma, the proof of which can be found in Elliott [12].

Lemma 14 *The $N \times N$ matrix stochastic process,*

$$m_n m_n^{\mathsf{T}} - (\text{diag}[PX_{n-1}] - PX_{n-1}X_{n-1}^{\mathsf{T}}P^{\mathsf{T}}), \tag{5.2}$$

is a martingale increment (with respect to P and $\mathcal{F}^X$).

We presume that this Markov chain is not directly observable, (i.e. it is hidden), so that in the real world we do not have access to the information $\mathcal{F}^X$. However, we do observe a stochastic process $\{y_n; n \in \mathbf{N}^+ = \{1, 2, ...\}\}$, which has the form

$$y_n = \rho(X_{n-1})y_{n-1} + g(X_{n-1}) + \sigma(X_{n-1})\epsilon_n, \tag{5.3}$$

where $\{\epsilon_n\}$ is a sequence of i.i.d. standard normal random variables, independent with the Markov chain, X. It is clear that ρ, g and σ are

Markov chains, and by choosing a suitable number of states, N, for X, these Markov chains can have any degree of correlation with each other. Also notice that the function $\rho(X_{n-1})$ has the representation $\tilde{\rho}^{\mathsf{T}} X_{n-1} = \langle \tilde{\rho}, X_{n-1} \rangle$, where the vector $\tilde{\rho}$ has typical entry $\tilde{\rho}_i = \rho(e_i)$, and similarly for g and σ. We also make the typical assumption that each entry of $\tilde{\rho}$ is between -1 and 1, and that each entry of $\tilde{\sigma}$ is strictly positive.

Denote by $\mathcal{F}^y = \{\mathcal{F}_n^y\}$ the filtration generated by the observed process, y, $\mathcal{F}^\epsilon = \{\mathcal{F}_n^\epsilon\}$ the filtration generated by the noise, ϵ, and $\mathcal{G} = \{\mathcal{G}_n\} = \{\mathcal{F}_n^X \vee \mathcal{F}_n^y\} = \{\mathcal{F}_n^X \vee \mathcal{F}_n^\epsilon\}$ is the joint, (or global), filtration. The filtering problem will therefore involve the optimal use of the available information. We wish to make inferences about $\mathcal{G}$-adapted processes by conditioning on the filtration $\mathcal{F}^y$. This gives a best, (in mean square error sense), estimate of the unobservable processes.

5.2 The Reference Probability

The approach we use to obtain the required conditional expectations is to first solve the problem using a different probability measure, (so it's easy to solve), and then convert back to the original measure, P. Under the new probability measure, which is called the reference probability and denoted $\bar{P}$, the observations divided by the previous volatility is a sequence of i.i.d. standard normal random variables. We begin by constructing the reference probability.

Define two $\mathcal{G}$-adapted stochastic processes: $\{\lambda_n; n \in \mathbf{N}^+\}$ as

$$\begin{aligned} \lambda_n &= \exp\Big(-\frac{\langle \tilde{\rho}, X_{n-1} \rangle y_{n-1} + \langle \tilde{g}, X_{n-1} \rangle}{\langle \tilde{\sigma}, X_{n-1} \rangle} \epsilon_n \\ &\quad - \frac{1}{2} \frac{(\langle \tilde{\rho}, X_{n-1} \rangle y_{n-1} + \langle \tilde{g}, X_{n-1} \rangle)^2}{\langle \tilde{\sigma}, X_{n-1} \rangle^2} \Big), \end{aligned} \tag{5.4}$$

and $\{\Lambda_n; n \in \mathbf{N}\}$ as

$$\begin{aligned} \Lambda_0 &= 1 \\ \Lambda_n &= \prod_{k=1}^{n} \lambda_k. \end{aligned}$$

Several properties of $\{\Lambda_n\}$ are discussed in the following lemma, which is proven in the appendix. In the discussion that follows, we fix a terminating time $T \in \mathbf{N}$ and consider the case when $n \leq T$.

Lemma 15 *The stochastic process $\{\Lambda_n\}$ is a P-almost surely positive, martingale with respect to the measure P and the filtration $\mathcal{G}$ that has expectation 1 under P.*

We now define the reference measure as $\bar{P}(A) = \int_A \Lambda_T \, dP$ for all $A \in \mathcal{G}_T$. It follows from Lemma 15 that, for $A \in \mathcal{G}_n$, $\bar{P}(A) = \int_A \Lambda_n \, dP$, and since $\Lambda_n > 0$ a.s., $P(A) = \int_A 1/\Lambda_n \, d\bar{P}$. The earlier claim is now formalised in the following lemma, which is proven in the appendix.

Lemma 16 *Under the reference measure, $\bar{P}$, the stochastic process, X, is a Markov chain with the same transition matrix, P, and the stochastic process, $\{y_n/\langle\tilde{\sigma}, X_{n-1}\rangle\}$, is a sequence of i.i.d. standard normal random variables.*

The importance of this lemma will become clearer in the next section.

5.3 The Recursive Filter

The filtering problem is to find $E[H_n|\mathcal{F}_n^y]$ for any stochastic process H. We consider the case when the stochastic process, H, is adapted to the filtration $\mathcal{G}$. In this case, by the discussion following Lemma 3, we can use a version of Bayes' theorem to get

$$E[H_n|\mathcal{F}_n^y] = \frac{\bar{E}[H_n/\Lambda_n|\mathcal{F}_n^y]}{\bar{E}[1/\Lambda_n|\mathcal{F}_n^y]}. \tag{5.5}$$

For notational convenience, write $\sigma_n(H_n) = \bar{E}[H_n/\Lambda_n|\mathcal{F}_n^y]$, then we have $E[H_n|\mathcal{F}_n^y] = \sigma_n(H_n)/\sigma_n(1)$. Thus, to obtain the filter, it is sufficient to determine $\sigma_n(H_n)$ for $\mathcal{G}$-adapted processes, H, ($H_n \equiv 1$ being a special case). Our approach to solving this problem is to find a recursive relationship between $\sigma_n(H_n)$ and $\sigma_{n-1}(H_{n-1})$, where we take

$\sigma_0(H_0) = E[H_0]$ as the initial value. It turns out that if H is scalar valued, then the recursive relationship for $\sigma_n(H_n)$ involves terms containing $\sigma_{n-1}(H_{n-1}X_{n-1})$. (Details of this fact can be found in Elliott [12].) This motivates the notion that we may need to consider filtering the product, (vector-valued), stochastic process, $\{H_nX_n\}$. However, notice that $\langle \tilde{1}, X_n \rangle = 1$, so $\langle \tilde{1}, \sigma_n(H_nX_n)\rangle = \sigma_n(H_n\langle \tilde{1}, X_n\rangle) = \sigma_n(H_n)$. This means that it's also sufficient to consider this product stochastic process. Summarising, for scalar-valued, $\mathcal{G}$-adapted stochastic processes, H, we have

$$E[H_n|\mathcal{F}_n^y] = \frac{\langle \tilde{1}, \sigma_n(H_nX_n)\rangle}{\langle \tilde{1}, \sigma_n(X_n)\rangle}. \tag{5.6}$$

The following theorem provides the recursive relationship for $\sigma_n(H_nX_n)$ for a special class of processes, H, to which the process $H_n \equiv 1$ also belongs. The proof is given in the appendix.

Theorem 7 *Suppose H_0 is $\mathcal{F}_0^X$-measurable and $H_n = \alpha_n + \langle \beta_n, X_n\rangle + \gamma_n f(y_{n-1}, y_n)$, where α, β, and γ are $\mathcal{G}$-predictable (of the appropriate dimension), and f is scalar valued. Then a recursive relationship for $\sigma_n(H_nX_n)$ is given by*

$$\begin{aligned}\sigma_n(H_nX_n) = \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)[\langle e_i, \sigma_{n-1}(\alpha_n X_{n-1})\rangle p_i \\ + \operatorname{diag}[p_i]\sigma_{n-1}(\beta_n\langle e_i, X_{n-1}\rangle) + \sigma_{n-1}(\gamma_n\langle e_i, X_{n-1}\rangle) f(y_{n-1}, y_n) p_i],\end{aligned} \tag{5.7}$$

where $\Gamma^i(x_1, x_2) = \exp([(\tilde{\rho}_i x_1 + \tilde{g}_i)/\tilde{\sigma}_i^2]x_2 - \frac{1}{2}[(\tilde{\rho}_i x_1 + \tilde{g}_i)/\tilde{\sigma}_i]^2)$, and $p_i = Pe_i$ is the i^{th} column of the transition matrix, P.

We now look at some particular examples of processes, H, that are interesting in their own accord, but also useful for estimating the parameters of the model.

Example 1 $H_n = 1$. This provides a filter for the Markov chain (or equivalently the conditional probability distribution of the states). In the notation of the theorem, take $H_0 = 1$, $\alpha_n = 1$, $\beta_n = \tilde{0}$, and $\gamma_n = 0$, then we get

$$\sigma_n(X_n) = \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)\langle e_i, \sigma_{n-1}(X_{n-1})\rangle p_i. \tag{5.8}$$

Example 2 $H_n = J_n^{rs} = \sum_{k=1}^{n} \langle e_r, X_{k-1}\rangle \langle e_s, X_k\rangle$. The process J_n^{rs} measures the number of jumps made by the Markov chain from state e_r to state e_s up to time n. Here we take $H_0 = 0$, $\alpha_n = J_{n-1}^{rs}$, $\beta_n = \langle e_r, X_{n-1}\rangle e_s$, and $\gamma_n = 0$ to get

$$\begin{aligned} \sigma_n(J_n^{rs} X_n) &= \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)[\langle e_i, \sigma_{n-1}(J_{n-1}^{rs} X_{n-1})\rangle p_i \\ &\quad + \operatorname{diag}[p_i]\sigma_{n-1}(\langle e_r, X_{n-1}\rangle e_s \langle e_i, X_{n-1}\rangle)]. \end{aligned} \tag{5.9}$$

This turns out to take the following form, (the algebra is in the appendix):

$$\begin{aligned} \sigma_n(J_n^{rs} X_n) &= \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)\langle e_i, \sigma_{n-1}(J_{n-1}^{rs} X_{n-1})\rangle p_i \\ &\quad + \Gamma^r(y_{n-1}, y_n)\langle e_r, \sigma_{n-1}(X_{n-1})\rangle p_{sr} e_s. \end{aligned} \tag{5.10}$$

Example 3 $H_n = O_n^r = \sum_{k=1}^{n} \langle e_r, X_{k-1}\rangle$. This process measures the occupation time of state e_r up to (just prior to) time n, or more precisely, it measures the number of times $X_k = e_r$ for $k \in 0, ..., n-1$. Now let $H_0 = 0$, $\alpha_n = O_{n-1}^r + \langle e_r, X_{n-1}\rangle$, $\beta_n = \tilde{0}$, and $\gamma_n = 0$. Then Theorem 1 implies that

$$\begin{aligned} \sigma_n(O_n^r X_n) &= \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)[\langle e_i, \sigma_{n-1}(O_{n-1}^r X_{n-1})\rangle p_i \\ &\quad + \langle e_i, \sigma_{n-1}(\langle e_r, X_{n-1}\rangle X_{n-1})\rangle p_i], \end{aligned} \tag{5.11}$$

which, similarly to Example 2, can be re-written as

$$\begin{aligned} \sigma_n(O_n^r X_n) &= \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)\langle e_i, \sigma_{n-1}(O_{n-1}^r X_{n-1})\rangle p_i \\ &\quad + \Gamma^r(y_{n-1}, y_n)\langle e_r, \sigma_{n-1}(X_{n-1})\rangle p_r. \end{aligned} \tag{5.12}$$

Example 4 $H_n = G_n^r(f(\ell(y), y)) = \sum_{k=1}^{n} \langle e_r, X_{k-1}\rangle f(y_{k-1}, y_k)$, ($\ell(\cdot)$ is a lag operator). This process weights the occupation time by the function $f(\ell(y), y)$ at each time. It is useful for determining estimates for the autoregressive, drift and volatility vectors, $\tilde{\rho}$, $\tilde{g}$ and $\tilde{\sigma}$. Applying

Theorem 7 with $H_0 = 0$, $\alpha_n = G^r_{n-1}(f(\ell(y), y))$, $\beta_n = \tilde{0}$, and $\gamma_n = \langle e_r, X_{n-1}\rangle$ gives

$$\begin{aligned}
&\sigma_n(G^r_n(f(\ell(y), y))X_n) \\
&= \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)[\langle e_i, \sigma_{n-1}(G^r_{n-1}(f(\ell(y), y))X_{n-1})\rangle p_i \\
&\quad + \sigma_{n-1}(\langle e_r, X_{n-1}\rangle \langle e_i, X_{n-1}\rangle) f(y_{n-1}, y_n) p_i] \\
&= \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)\langle e_i, \sigma_{n-1}(G^r_{n-1}(f(\ell(y), y))X_{n-1})\rangle p_i \\
&\quad + \Gamma^r(y_{n-1}, y_n)\langle e_r, \sigma_{n-1}(X_{n-1})\rangle f(y_{n-1}, y_n) p_r .
\end{aligned} \tag{5.13}$$

5.4 Maximum Likelihood Estimation

The model requires estimates for the transition probabilities, p_{ji}, and the entries of the vectors $\tilde{\rho}$, $\tilde{g}$, and $\tilde{\sigma}$. It is well known that the class of maximum likelihood estimators, (MLE's), has several desirable properties such as consistency, efficiency, and robustness. We therefore attempt to find MLE's for the various parameters. The problem is that MLE's can often be difficult to calculate directly or explicitly. This can be solved in several ways: (1) it can be assumed away by considering a false, (but hopefully close), and easier to work with probability distribution, (2) we can obtain a sequence of estimates that converges to the MLE, or (3) we can simulate a family of empirical distributions and work numerically. We choose to use the EM algorithm to find the MLE by the second method. The reason for this is because the method is intuitive, and convergence is quite rapid so the method is also effective.

To begin with, we first characterize the maximum likelihood method in a way that relates to this chapter, (following Dembo and Zeitouni [7]). Consider a family of equivalent probability measures, $\{P_\theta; \theta \in \Theta\}$, defined on the measurable space $(\Omega, \mathcal{F})$, and a sub-sigma-field, $\mathcal{Y} \subset \mathcal{F}$, representing available information. For each $\alpha \in \Theta$, we define the likelihood function with respect to α, $L_\alpha : \Theta \rightarrow [0, \infty)$ as

$$L_\alpha(\theta) = E_\alpha\Big[\frac{dP_\theta}{dP_\alpha} \mid \mathcal{Y}\Big] = \frac{1}{E_\theta[\frac{dP_\alpha}{dP_\theta} \mid \mathcal{Y}]}, \tag{5.14}$$

where $\alpha \in \Theta$. The likelihood function does depend on the chosen index, α, however, any parameter, $\theta^* \in \Theta$, that maximises the likelihood function for (almost) every given outcome does not depend on the particular base, α. This follows easily from the equivalency of the family of measures: The Radon-Nikodym derivative is positive (a.s.), so maximising L over θ with some other index, β, is the same as maximising with α, but dividing by a positive constant as illustrated below,

$$L_\beta(\theta) = E_\beta\Big[\frac{dP_\theta}{dP_\beta} \Bigm| \mathcal{Y}\Big] = \frac{E_\alpha[\frac{dP_\theta}{dP_\alpha} \mid \mathcal{Y}]}{E_\alpha[\frac{dP_\beta}{dP_\alpha} \mid \mathcal{Y}]}. \tag{5.15}$$

This, of course, does not affect the maximising argument, (at least for almost every $\omega \in \Omega$). Clearly, the likelihood function and the maximising argument may depend on the information $\mathcal{Y}$ and on the realised state $\omega \in \Omega$.

Now, a maximum likelihood estimate (given information $\mathcal{Y}$) is defined as an element of the set $\{\arg\max_{\theta\in\Theta} L_\alpha(\theta)\}$, if it's not empty. It is often difficult to calculate a MLE directly, so instead, it is sometimes convenient to define a sequence that converges to a MLE, where the entries of the sequence are easier to obtain. One method for defining such a sequence is via the EM algorithm, which is described as follows: For any $\alpha \in \Theta$, define the function $Q_\alpha : \Theta \to \mathbf{R}$ by

$$Q_\alpha(\theta) = E_\alpha\Big[\log\Big(\frac{dP_\theta}{dP_\alpha}\Big) \Bigm| \mathcal{Y}\Big]. \tag{5.16}$$

Notice that, by Jensen's inequality, $\log(L_\alpha(\theta)) \geq Q_\alpha(\theta)$, with equality obtaining if and only if dP_θ/dP_α is $\mathcal{Y}$-measurable. Now, suppose that $\hat{\theta}_p$ is the p^{th} entry of the sequence, where any arbitrary $\alpha \in \Theta$ is chosen when $p = 0$. The next entry of the sequence is arbitrarily chosen from the set

$$\hat{\theta}_{p+1} \in \{\arg\max_{\theta\in\Theta} Q_{\hat{\theta}_p}(\theta)\}, \tag{5.17}$$

provided it's not empty. We further restrict the sequence constructed above so that we choose $\hat{\theta}_{p+1} = \hat{\theta}_p$ whenever $\hat{\theta}_p \in \{\arg\max_{\theta\in\Theta} Q_{\hat{\theta}_p}(\theta)\}$. We call this sequence an EM sequence. The following lemma, which is proven in the appendix, claims that the likelihood of subsequent terms in the sequence is non-decreasing.

Lemma 17 *For any $\alpha \in \Theta$, and any $p \in \mathbf{N}$ such that $\hat{\theta}_{p+1}$ as defined above exists, $L_\alpha(\hat{\theta}_{p+1}) \geq L_\alpha(\hat{\theta}_p)$, with equality if and only if $\hat{\theta}_{p+1} = \hat{\theta}_p$.*

It is clear from Lemma 17 that if $\hat{\theta}_{p+1} = \hat{\theta}_p$ for some p, then for all $n \geq p$, $\hat{\theta}_n$ exists and is equal to $\hat{\theta}_p$. Such an entry is called a fixed point of the EM sequence. A more general sufficient condition for the existence of an EM sequence is described in the next lemma. The proof is in the appendix.

Lemma 18 *If $Q_\alpha(\cdot)$ is continuous on Θ, and the non-empty set, $\{\theta \in \Theta; L_\alpha(\theta) \geq L_\alpha(\theta_0)\}$, is compact for all $\alpha, \theta_0 \in \Theta$, then an EM sequence exists from any starting point in the parameter space, Θ.*

More details on sufficient conditions for the EM sequence to converge to a maximum likelihood estimate can be found in Dembo and Zeitouni [7]. The EM sequences constructed in the next section all converge to a MLE.

5.5 Parameter Estimation

We apply the EM algorithm to get estimates for the parameters of the model, however, instead of updating the entire parameter set simultaneously, we update subsets of the parameters one at a time until the entire set has been updated. It is clear that, with a finite number of parameters, this approach causes no difficulties because of the continuity and compactness discussed in Lemma 18.

The first step is to update the entries of the transition probability matrix, P. Let the previous estimates be denoted p_{sr} and the new estimates be denoted $\hat{p}_{sr}$. Now, note that $\Delta J_n^{rs} = \langle e_r, X_{n-1}\rangle\langle e_s, X_n\rangle$, and $E[\Delta J_n^{rs}|\mathcal{G}_{n-1}] = \langle e_r, X_{n-1}\rangle p_{sr}$, where J_n^{rs} is the number of jumps from e_r to e_s, by time n, as described in Example 2. The dependence on p_{sr} motivates the usefulness of J in updating the transition probabilities. Following Elliott and Yang [14], define $\hat{\Lambda}_0 = 1$, and

$$\hat{\Lambda}_n = \prod_{k=1}^{n} \prod_{\{r,s;p_{sr}>0\}} \left(\frac{\hat{p}_{sr}}{p_{sr}}\right)^{\Delta J_k^{rs}} = \prod_{k=1}^{n} \sum_{\{r,s;p_{sr}>0\}} \left(\frac{\hat{p}_{sr}}{p_{sr}}\right)\langle e_r, X_{k-1}\rangle\langle e_s, X_k\rangle. \tag{5.18}$$

Similarly to Lemma 15, it can be shown that $\hat{\Lambda}$ is a positive $(P, \mathcal{G})$-martingale with expectation 1. Therefore, we can define a probability measure, $\hat{P}$, which is equivalent to P, where, for $A \in \mathcal{G}_n$, $\hat{P}(A) = E[1_A \hat{\Lambda}_n]$. The following lemma, which is proven in the appendix, describes the transition probabilities under $\hat{P}$.

Lemma 19 *Under $\hat{P}$, the stochastic process, X, is a Markov chain with transition probabilities $\hat{p}_{sr}$.*

Using $\hat{\Lambda}_n$ as the Radon-Nikodym derivative described in the previous section, we see that our goal is to maximise $E[\log(\hat{\Lambda}_n)|\mathcal{F}^y_n]$ with respect to the parameters, $\hat{p}_{sr}$, in the allowable space of parameters. The space of parameters is restricted by the fact that p_{sr} is the probability of going from state e_r to state e_s, so $\hat{p}_{sr} \geq 0$, and $\sum_s p_{sr} = 1$. The second constraint is equivalent to

$$\sum_{k=1}^{n} \sum_{r,s=1}^{N} \langle e_r, X_{k-1} \rangle \hat{p}_{sr} = \sum_{r,s=1}^{N} O^r_n \hat{p}_{sr} = n, \tag{5.19}$$

and we will make the further restriction that once one estimate, p_{sr}, is zero, then all updated estimates, $\hat{p}_{sr}$, are zero. This gives the following constraint:

$$\sum_{\{r,s;p_{sr}>0\}} O^r_n \hat{p}_{sr} - n = 0. \tag{5.20}$$

The problem can be described as:

$$\begin{aligned} \text{Maximise} \quad & E\Big[\sum_{k=1}^{n} \sum_{\{r,s;p_{sr}>0\}} \Delta J^{rs}_k (\log \hat{p}_{sr} - \log p_{sr}) \;\Big|\; \mathcal{F}^y_n\Big] \quad (5.21) \\ & = \sum_{\{r,s;p_{sr}>0\}} E[J^{rs}_n|\mathcal{F}^y_n] \log \hat{p}_{sr} + \text{remainder} \\ \text{subject to} \quad & \sum_{\{r,s;p_{sr}>0\}} E[O^r_n|\mathcal{F}^y_n] \hat{p}_{sr} - n = 0, \end{aligned}$$

where the remainder term is independent of $\hat{p}_{sr}$. From this, it is clear that the function to be maximised is smooth and concave in $\hat{p}_{sr}$, so it is enough to show first order conditions. The Lagrangian is written as

$$\mathcal{L}(\hat{p}, \lambda) = \sum_{\{r,s;p_{sr}>0\}} (E[J^{rs}_n|\mathcal{F}^y_n] \log \hat{p}_{sr} - \lambda(E[O^r_n|\mathcal{F}^y_n]\hat{p}_{sr} - n)) + \text{remainder}. \tag{5.22}$$

Setting first derivatives to zero gives

$$\frac{E[J_n^{rs}|\mathcal{F}_n^y]}{\hat{p}_{sr}} = \lambda E[O_n^r|\mathcal{F}_n^y], \tag{5.23}$$

and the constraint. Multiplying by $\hat{p}_{sr}$ and summing over $\{r, s; p_{sr} > 0\}$ implies that the Lagrange multiplier, λ, is 1. This proves the following theorem.

Theorem 8 *The transition probability portion of the n^{th} element of the EM sequence is given by $\hat{p}^*_{sr} = 0$ if the $n-1^{th}$ element, $p_{sr} = 0$, and otherwise*

$$\hat{p}^*_{sr} = \frac{E[J_n^{rs}|\mathcal{F}_n^y]}{E[O_n^r|\mathcal{F}_n^y]} = \frac{\langle \tilde{1}, \sigma_n(J_n^{rs} X_n)\rangle}{\langle \tilde{1}, \sigma_n(O_n^r X_n)\rangle}. \tag{5.24}$$

We now proceed to determine the remaining parameter estimates. Unfortunately, trying to determine the entries of the three parameter vectors, $\tilde{\rho}$, $\tilde{g}$, and $\tilde{\sigma}$ simultaneously leads to a system of equations with an indeterminate solution. Instead, we will first update the parameters ρ and g, and then σ will be updated individually.

Following Elliott [12], and similarly to our construction of the reference probability, we begin by defining two $\mathcal{G}$-adapted stochastic processes

$$\begin{aligned}
\lambda_n^* = \exp\Big(&\frac{((\langle\hat{\tilde{\rho}}, X_{n-1}\rangle y_{n-1} + \langle\hat{\tilde{g}}, X_{n-1}\rangle) - (\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)}{\langle\tilde{\sigma}, X_{n-1}\rangle}\epsilon_n \\
&- \tfrac{1}{2}\Big(\frac{((\langle\hat{\tilde{\rho}}, X_{n-1}\rangle y_{n-1} + \langle\hat{\tilde{g}}, X_{n-1}\rangle) - (\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)}{\langle\tilde{\sigma}, X_{n-1}\rangle}\Big)^2\Big) \\
= \ & \exp\Big(\frac{1}{2\langle\tilde{\sigma}, X_{n-1}\rangle^2}\big((\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)^2 \\
&- (\langle\hat{\tilde{\rho}}, X_{n-1}\rangle y_{n-1} + \langle\hat{\tilde{g}}, X_{n-1}\rangle)^2 + 2y_n((\langle\hat{\tilde{\rho}}, X_{n-1}\rangle y_{n-1} + \langle\hat{\tilde{g}}, X_{n-1}\rangle) \\
&- (\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle))\big)\Big),
\end{aligned} \tag{5.25}$$

and

$$\begin{aligned}
\Lambda_0^* &= 1 \\
\Lambda_n^* &= \prod_{k=1}^{n} \lambda_k^*.
\end{aligned} \tag{5.26}$$

Again, similarly to Lemma 15, we have Λ^* is a positive $(P, \mathcal{G})$-martingale with expectation 1, so we can define a probability measure, P^*, equivalent to P, where $P^*(A) = E[1_A \Lambda_n^*]$ for all $A \in \mathcal{G}_n$. We proceed, as in the previous discussion, by constructing an EM sequence, where parameters without hats are taken to be the current estimates and the updated estimates are the maximising arguments over the parameters with hats for $E[\log \Lambda_n^* | \mathcal{F}_n^y]$. Now

$$\begin{aligned}
\log \Lambda_n^* &= \sum_{k=1}^{n} \frac{1}{2\langle \tilde{\sigma}, X_{k-1}\rangle^2}\Big(2y_k(\langle \hat{\tilde{\rho}}, X_{k-1}\rangle y_{k-1} + \langle \hat{\tilde{g}}, X_{k-1}\rangle) \\
&\quad -(\langle \hat{\tilde{\rho}}, X_{k-1}\rangle y_{k-1} + \langle \hat{\tilde{g}}, X_{k-1}\rangle)^2\Big) + \text{remainder} \\
&= \sum_{k=1}^{n}\sum_{i=1}^{N} \frac{\langle e_i, X_{k-1}\rangle}{2\tilde{\sigma}_i^2}\Big(2\hat{\tilde{\rho}}_i y_{k-1} y_k + 2\hat{\tilde{g}}_i y_k \\
&\quad -\hat{\tilde{\rho}}_i{}^2 y_{k-1}^2 - 2\hat{\tilde{\rho}}_i \hat{\tilde{g}}_i y_{k-1} - \hat{\tilde{g}}_i{}^2\Big) + \text{remainder} \\
&= \sum_{i=1}^{N} \frac{1}{2\tilde{\sigma}_i^2}\Big(2\hat{\tilde{\rho}}_i G_n^i(\ell(y) \times y) + 2\hat{\tilde{g}}_i G_n^i(y) \\
&\quad -\hat{\tilde{\rho}}_i{}^2 G_n^i(\ell^2(y)) - 2\hat{\tilde{\rho}}_i \hat{\tilde{g}}_i G_n^i(\ell(y)) - \hat{\tilde{g}}_i{}^2 O_n^i\Big) + \text{remainder},
\end{aligned} \tag{5.27}$$

so

$$\begin{aligned}
E[\log \Lambda_n^* | \mathcal{F}_n^y] &= \sum_{i=1}^{N} \frac{1}{2\sigma_i^2}\Big(2\hat{\tilde{\rho}}_i E[G_n^i(\ell(y) \times y) | \mathcal{F}_n^y] + 2\hat{\tilde{g}}_i E[G_n^i(y) | \mathcal{F}_n^y] \\
&\quad - \hat{\tilde{\rho}}_i{}^2 E[G_n^i(\ell^2(y)) \mid \mathcal{F}_n^y] - 2\hat{\tilde{\rho}}_i \hat{\tilde{g}}_i E[G_n^i(\ell(y)) \mid \mathcal{F}_n^y] - \hat{\tilde{g}}_i{}^2 E[O_n^i | \mathcal{F}_n^y]\Big) \\
&\quad + \text{remainder},
\end{aligned} \tag{5.28}$$

where the remainder term is independent of hat variables. As a function of $\hat{\rho}$, and $\hat{g}$, $E[\log \Lambda_n^* | \mathcal{F}_n^y]$ is smooth and concave, which can be checked by noticing that

$$E[G_n^i(\ell^2(y)) \mid \mathcal{F}_n^y] E[O_n^i | \mathcal{F}_n^y] - (E[G_n^i(\ell(y)) \mid \mathcal{F}_n^y])^2 \geq 0, \tag{5.29}$$

so the Hessian is negative semi-definite everywhere. Thus we can find a global maximum by solving the first-order conditions. Note that the above inequality holds with equality if and only if $E[O_n^i | \mathcal{F}_n^y] = 0$. The derivatives are

$$\begin{aligned}\frac{\partial}{\partial\hat{\tilde{\rho}}_i}E[\log\Lambda_n^*|\mathcal{F}_n^y] &= \frac{1}{\tilde{\sigma}_i^2}\Big(E[G_n^i(\ell(y)\times y)\mid\mathcal{F}_n^y]-\hat{\tilde{\rho}}_iE[G_n^i(\ell^2(y))\mid\mathcal{F}_n^y]\\ &\quad -\hat{\tilde{g}}_iE[G_n^i(\ell(y))\mid\mathcal{F}_n^y]\Big) \qquad (5.30)\\ \frac{\partial}{\partial\hat{\tilde{g}}_i}E[\log\Lambda_n^*|\mathcal{F}_n^y] &= \frac{1}{\tilde{\sigma}_i^2}\Big(E[G_n^i(y)|\mathcal{F}_n^y]-\hat{\tilde{\rho}}_iE[G_n^i(\ell(y))\mid\mathcal{F}_n^y]\\ &\quad -\hat{\tilde{g}}_iE[O_n^i|\mathcal{F}_n^y]\Big).\end{aligned}$$

Setting the derivatives to zero and solving simultaneously gives

$$\begin{aligned}\hat{\tilde{\rho}}_i^* &= \frac{E[G_n^i(\ell(y)\times y)\mid\mathcal{F}_n^y]-\hat{\tilde{g}}_i^*E[G_n^i(\ell(y))\mid\mathcal{F}_n^y]}{E[G_n^i(\ell^2(y))\mid\mathcal{F}_n^y]} \qquad (5.31)\\ \hat{\tilde{g}}_i^* &= \frac{E[G_n^i(y)|\mathcal{F}_n^y]-\hat{\tilde{\rho}}_i^*E[G_n^i(\ell(y))\mid\mathcal{F}_n^y]}{E[O_n^i|\mathcal{F}_n^y]}.\end{aligned}$$

This form is useful for the special cases when either ρ or g is restricted to be zero in the model. Continuing to solve for the optimal parameters proves the following theorem.

Theorem 9 *The autoregressive and drift components of the n^{th} entry of the EM sequence are given by*

$$\begin{aligned}\hat{\tilde{\rho}}_i^* &= \frac{\check{G}_n^i(\ell(y)\times y)\check{O}_n^i-\check{G}_n^i(y)\check{G}_n^i(\ell(y))}{\check{G}_n^i(\ell^2(y))\check{O}_n^i-[\check{G}_n^i(\ell(y))]^2} \qquad (5.32)\\ \hat{\tilde{g}}_i^* &= \frac{\check{G}_n^i(y)\check{G}_n^i(\ell^2(y))-\check{G}_n^i(\ell(y)\times y)\check{G}_n^i(\ell(y))}{\check{G}_n^i(\ell^2(y))\check{O}_n^i-[\check{G}_n^i(\ell(y))]^2},\end{aligned}$$

where $\check{G}_n = E[G_n|\mathcal{F}_n^y]$, and the various conditional expectations can be calculated via Examples 3 and 4. If the denominator is zero, then so are the numerators, and the updated parameters can be chosen the same as the previous parameters.

We continue with the final piece of the EM sequence, namely the volatility, σ. To begin, consider the stochastic processes

$$^*\lambda_n = \frac{\langle\tilde{\sigma},X_{n-1}\rangle}{\sqrt{\langle\hat{\tilde{\sigma}},X_{n-1}\rangle}}\exp\Big(\tfrac{1}{2}\Big(1-\frac{\langle\tilde{\sigma},X_{n-1}\rangle^2}{\langle\hat{\tilde{\sigma}},X_{n-1}\rangle}\Big)\epsilon_n^2\Big)$$

$$= \frac{\langle\tilde{\sigma}, X_{n-1}\rangle}{\sqrt{\langle\hat{\tilde{\sigma}}, X_{n-1}\rangle}} \exp\Big(\tfrac{1}{2}\Big(1 - \frac{\langle\tilde{\sigma}, X_{n-1}\rangle^2}{\langle\hat{\tilde{\sigma}}, X_{n-1}\rangle}\Big) \times \Big(\frac{y_n - \langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} - \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle}\Big)^2\Big), \tag{5.33}$$

and

$$\begin{aligned} {}^*\Lambda_0 &= 1 \\ {}^*\Lambda_n &= \prod_{k=1}^{n} {}^*\lambda_k. \end{aligned} \tag{5.34}$$

Again ${}^*\Lambda$ can be seen to be a positive $(P, \mathcal{G})$-martingale with expectation 1. (Note that $\{\epsilon_n^2\}$ is a sequence of i.i.d. central χ^2 random variables with 1 degree of freedom.) Therefore, we can define a probability measure, ${}^*P$, such that ${}^*P(A) = E[1_A {}^*\Lambda_n]$, for all $A \in \mathcal{G}_n$. Our goal is to find the argument that maximises $E[\log {}^*\Lambda_n | \mathcal{F}_n^y]$ with respect to $\hat{\tilde{\sigma}}$. Observe that

$$\begin{aligned} \log {}^*\Lambda_n &= -\tfrac{1}{2}\sum_{k=1}^{n}\Big(\log\langle\hat{\tilde{\sigma}}, X_{k-1}\rangle + \frac{(y_k - \langle\tilde{\rho}, X_{k-1}\rangle y_{k-1} - \langle\tilde{g}, X_{k-1}\rangle)^2}{\langle\hat{\tilde{\sigma}}, X_{k-1}\rangle}\Big) \\ &\quad + \text{remainder} \qquad (5.35) \\ &= -\tfrac{1}{2}\sum_{i=1}^{N}\Big[O_n^i \log\hat{\tilde{\sigma}}_i + \frac{1}{\hat{\tilde{\sigma}}_i}\Big(G_n^i(y^2) + \tilde{\rho}_i^2 G_n^i(\ell^2(y)) + \tilde{g}_i^2 O_n^i \\ &\quad - 2\tilde{\rho}_i G_n^i(\ell(y)\times y) - 2\tilde{g}_i G_n^i(y) + 2\tilde{\rho}_i\tilde{g}_i G_n^i(\ell(y))\Big)\Big] \\ &\quad + \text{remainder}. \end{aligned}$$

Taking conditional expectation gives

$$\begin{aligned} E[\log {}^*\Lambda_n | \mathcal{F}_n^y] &= -\tfrac{1}{2}\sum_{i=1}^{N}\Big[\check{O}_n^i \log\hat{\tilde{\sigma}}_i + \frac{1}{\hat{\tilde{\sigma}}_i}\Big(\check{G}_n^i(y^2) + \tilde{\rho}_i^2 \check{G}_n^i(\ell^2(y)) + \tilde{g}_i^2 \check{O}_n^i \\ &\quad - 2\tilde{\rho}_i \check{G}_n^i(\ell(y)\times y) - 2\tilde{g}_i \check{G}_n^i(y) + 2\tilde{\rho}_i\tilde{g}_i \check{G}_n^i(\ell(y))\Big)\Big] \\ &\quad + \text{remainder}, \end{aligned} \tag{5.36}$$

where we use the same notation as in Theorem 9. Recall that since it is a sum of squares, the portion of the second term that is in parentheses, and the occupation time, O_n^i, are non-negative, (and they are only zero simultaneously). Maximising over $\hat{\tilde{\sigma}}_i$ gives the following theorem.

Theorem 10 *The volatility estimate portion of the n^{th} entry of the EM sequence is given by*

$$\begin{aligned}\hat{\tilde{\sigma}}_i^* &= \frac{1}{\check{O}_n^i}\Big(\check{G}_n^i(y^2) + \tilde{\rho}_i^2 \check{G}_n^i(\ell^2(y)) + \tilde{g}_i^2 \check{O}_n^i - 2\tilde{\rho}_i \check{G}_n^i(\ell(y) \times y) \\ &\quad -2\tilde{g}_i \check{G}_n^i(y) + 2\tilde{\rho}_i \tilde{g}_i \check{G}_n^i(\ell(y))\Big), \end{aligned} \tag{5.37}$$

where $\check{G}_n = E[G_n|\mathcal{F}_n^y]$, and the various conditional expectations can be calculated via Examples 3 and 4. If the denominator is zero, then so is the numerator, and the updated parameter can be chosen the same as the previous parameter.

The proof can be completed by using the first order condition and noticing that $E[\log {}^*\Lambda_n|\mathcal{F}_n^y]$ is concave in $\tilde{\sigma}_i$ on $[0, 2\hat{\tilde{\sigma}}_i^*]$, and decreasing outside that region.

Thus, using Theorems 8, 9, and 10, we can uniquely update the EM sequence by a bootstrapping method to obtain a sequence that converges to a parameter inside the feasible set of parameters such that the likelihood is globally maximised, monotonically. Note that the procedure can be applied repeatedly at each step by cycling the new parameter estimates into the filtering equations to obtain new filters, which can be used to obtain new parameter estimates. In practice, a stopping criterion can be applied to each step, after which new data can be given, and the procedure repeated. The results of this and the previous section show that likelihood is increasing both within each step, and among steps.

This page is intentionally left blank.

Chapter 6

Estimating the Hidden Markov Model

6.1 The Interest Rate Model

To model the short-term interest rate, we begin with an extension of the Vasiček model developed in Chapter 4. The interest rate was supposed to follow a continuous time stochastic process defined by the SDE,

$$dr_t = a(X_t)(\bar{r}(X_t) - r_t)\,dt + \varsigma(X_t)\,dW_t, \tag{6.1}$$

where W is a standard Brownian motion independent with the continuous time Markov chain, X. The solution to this SDE is

$$r_u = e^{-K(u)}\Big(e^{K(t)}r_t + \int_t^u e^{K(s)}a(X_s)\bar{r}(X_s)\,ds + \int_t^u e^{K(s)}\varsigma(X_s)\,dW_s\Big), \tag{6.2}$$

where

$$K(t) = \int_0^t a(X_s)\,ds. \tag{6.3}$$

If we suppose $u - t$ is small and X is constant over that interval, then

$$r_u \approx e^{-a(X_t)(t-u)}r_t + \bar{r}(1 - e^{-a(X_t)(t-u)}) + \varsigma(X_t)\int_t^u e^{a(X_t)(s-u)}\,dW_s. \tag{6.4}$$

The stochastic integral has normal distribution with mean 0 and variance

$$\int_t^u e^{2a(X_t)(s-u)}\,ds = \frac{1-e^{-2a(X_t)(u-t)}}{2a(X_t)}, \tag{6.5}$$

and it is independent with similar stochastic integrals over other non-overlapping intervals.

This motivates the following discrete representation of the interest rate:

$$r_{n+1} = \rho(X_n)r_n + g(X_n) + \sigma(X_n)\epsilon_{n+1}, \tag{6.6}$$

where

$$\begin{aligned} \rho(X_n) &= e^{-a(X_n)\Delta} \\ g(X_n) &= \bar{r}(X_n)(1-e^{-a(X_n)\Delta}) \\ \sigma(X_n) &= \varsigma(X_n)\sqrt{\frac{1-e^{-2a(X_n)\Delta}}{2a(X_n)}} \end{aligned} \tag{6.7}$$

$\{X_n\}$ is a discrete time Markov chain, and $\{\epsilon_n\}$ are i.i.d. standard normal. The probability transition matrix is $P = e^{A\Delta}$, where A is the transition intensity matrix for the original continuous time Markov chain. Here Δ is the constant $t_{n+1} - t_n$. Since the columns of A sum to zero, so do the columns of $A\Delta$, and thus the columns of P sum to 1 as expected. Note that $P_{ji} = P(X_{n+1} = e_j | X_n = e_i)$. So we have the situation discussed in the earlier section. Note that the parameters estimated here will differ from those used in the bond pricing and other applications since here we are estimating under the true or physical probability, rather than the risk-neutral probability. However the estimations still provide a useful comparison between the hidden Markov model and the Vasiček model.

We now turn our attention to what this model implies about the behaviour of interest rates. First of all the short-term rates, following this model will be positively auto-correlated through time. The auto-correlation, (conditional on r_n, X_n, and X_{n+1}), is given by

$$\frac{\sigma(X_n)\rho(X_n)}{\sqrt{\rho(X_{n+1})^2\sigma(X_n)^2 + \sigma(X_{n+1})^2}}. \tag{6.8}$$

The auto-correlation coefficient, ρ, will be less than 1, whenever the mean reversion coefficient, a, is positive. This is the usual case, where a measures the rate that r is expected to approach the mean reverting level, $\bar{r}$.

Allowing the parameters to depend upon a Markov chain means that, (most importantly), the mean reverting level will change from time to time. When it does change, the interest rate will begin to converge to the new level, when it changes back, the interest rate will turn around and begin to converge back. It seems intuitive that such a data generating process would describe a cyclical pattern, but with a random cycle length.

6.2 Data

We use monthly observations of the short term interest rate implied by Government of Canada 3-month Treasury bills. The rates were obtained from the Bank of Canada website, excerpted from *Selected Canadian and International Interest Rates Including Bond Yields and Interest Arbitrage*. The data set includes bills from March 1934, when the first public tender occurred, until December 2002. There are several missing data points in 1934 and 1935, for months in which there was no tender. Since 1936, there was a tender in every month. The rates were converted from discount type rates to continuously compounded rates prior to any analysis.

A graph of this time series is included as Figure 9.1. From that graph we can see that interest rates remained quite flat until 1951, (actually October 1950). This is because the Bank of Canada had an easy money policy until that time, and simply targeted a low interest rate without regard to inflation, etc. Since 1951, interest rates do seem to exhibit some cyclical behaviour; however, the "cycles" seem to be quite random in length as well as intensity. This indicates that the hidden Markov model would be a good candidate for a data generating process, but only for the post 1951 portion.

The summary statistics of the time series from January 1951 to December 2002 are provided in Table 8.1. The important thing to notice here is the high positive skewness, indicating that an auto-regressive model might be appropriate. Also the standard deviation of the data is quite high at 3.80%, measuring the interest rate in units of %.

The results of an OLS regression of the simple model

$$r_{n+1} = \rho r_n + g + \sigma_n \tag{6.9}$$

are given in Table 8.2. From that we see that the auto-regressive coefficient, $\rho = 0.989$, is very significant and quite close to 1. Also, the standard error is 0.524% in units of %. This is quite a bit lower than the standard deviation observed in Table 8.1; however, it still means that fully around 1/3 of the observations are associated with an interest rate move of 0.5% or more that is only explained as noise, (given normal errors). This value seems exceedingly high considering we are dealing with monthly observations, and it likely explains why models incorporating stochastic volatility into interest rate models have been quite successful. Since this regression model can be considered a restricted version of our hidden Markov model, (i.e. when the number of states in the Markov chain is one, or when the parameters associated with each state are identical), it provides a natural comparison. Because of this we note that in case of i.i.d. normal errors, the model has three free parameters, and the sum of squared errors is approximately 171.

6.3 Results

We implement the model for a two-state Markov chain, for the case when the volatility is constant, or equivalently, when the volatility parameters associated with each state are equal. Because the columns of the Markov matrix must sum to 1, the matrix is associated with two free parameters, also both the auto-correlation and drift components of the model are associated with two free parameters each, and since we are restricting the volatility parameters to equal in each state, this is associated with 1 free parameter. This gives the model a total of seven parameters. The estimates obtained by the filtering algorithm are presented in Table 8.3, as are the corresponding parameters of the original model. Note that the parameters of the original model are in annual terms. The first panel of Table 8.3 give the parameters for the base case AR(1) model, which is a discrete version of Vasiček's model. This corresponds to our hidden Markov model for which the Markov chain has a degenerate single state. As such it provides a natural comparison for the more interesting two-state case. When the filtering algorithm is applied to the general two-state model in which all parameters, including the volatility σ, depend on the Markov chain, the parameters converge to the degenerate case with one of the states becoming reflective, immediately switching to the other state with probability 1, and the other state becoming absorbing and switching with probability 0. This provides parameter estimates for the absorbing state as in the first panel, whereas nonsense is obtained for the other state.

The second panel of Table 8.3 shows the results of the filtering algorithm when the volatility of the discrete model is independent of the state of the Markov chain. In this case the algorithm converges to meaningful numbers in each state. The nature of a homogeneous Markov chain is such that the time remaining until the chain switches does not depend on the amount of time it has been in the state. For our two-state Markov chain, the expected time remaining until switching is well known to be $1/(1 - P_{ii}) = 1/P_{ji}$ for $j \neq i$. This means that the expected remaining time until switching is around 2.88 months for state 1, and 6.37 months for state 2. It seems to switch too frequently on average to account for business cycles.

One major problem that is noticed about the coefficient estimates is that $\rho_2 > 1$. This implies that in state 2, (which is typically associated with rising rates on average), the short rate is diverging away from a fairly large negative rate, $\bar{r}_2 = -6.67$. Although this is mitigated by the mean reverting nature associated with state 1, it still leaves open the possibility of explosions in the short rate process. However, it does indicate that state 1 is associated with a falling interest rate and state 2 is associated with a rising interest rate, since it is diverging away from a low (negative) "target" rate. Combining this observation with the previous paragraph indicates that interest rates tend to increase slowly for longer time periods and fall more quickly for a shorter time period. This is consistent with the central bank preferring easy money policies to inflation control. Also note that even though there is a substantial difference in the other parameters for different states of the Markov chain, the volatility parameters for the continuous time model are quite similar. This indicates that the significance of stochastic volatility over non-linear drift found by Sun [40] may not be warranted under our non-linear drift model in which the drift switches according to a Markov chain.

Because of this possibility of explosions, we further restrict the model to keep ρ independent of the Markov chain's state. The third panel of Table 8.3 provides the maximum likelihood estimates in this case, which is most similar to the proposed term structure models. Here we find that the autoregressive parameter, ρ, falls slightly to 0.981 from 0.989, which increases the rate of mean reversion quite significantly from 0.128 to 0.235. Here the short rate switches back and forth between reasonable reversion rates of around 3.37% and 8.87%, and more importantly it can do this switching fairly infrequently, as would be anticipated if switching was due to business cycles. This can be seen by inverting the transition matrix parameters to find the expected time before switching from each state as 9.28 months and 9.61 months for states 1 and 2 respectively.

One criticism of the hidden Markov models is that they require at least twice as many parameters as the restricted base-case model, and don't substantially reduce the sum of squared errors between actual and predicted rates. One reason for this is that the maximum likelihood estimation process corresponds to minimising squared errors for the base case, but not for the general case. This would overstate the failure to reject the restricted model, or in other words it would lead to lower-powered test and a larger chance of making an error of the second type, (failing to reject an incorrect restricted model). The comparatively small value of volatility, (ς is 0.877 compared to 1.825 for the base case), suggests that most of the prediction error is due to the difficulty in filtering the state of the Markov chain. It seems that this difficulty could be alleviated by observing the short rate more frequently.

Therefore we conclude that the restricted hidden Markov model with constant mean reversion rate and constant volatility is at least economically significant. The maximum likelihood parameter estimates obtained from the filtering algorithm are also economically sensible. It seems that this model, derived in Chapter 4, is probably superior to the basic model derived in Chapter 3.

This page is intentionally left blank.

Chapter 7

Conclusion

We analysed two short rate models, in which the short rate dynamics were controlled in part by a Markov chain. The models differed in the diffusion term of the dynamics: The basic model having no diffusion term and the hidden Markov model having additive noise that is modulated by a possibly stochastic volatility parameter. We obtained term structure results, including a technique to match the initial term structure. We then developed a filtering technique to obtain maximum likelihood estimates for the parameters of the hidden Markov model. We estimated the model and concluded that a constant volatility version of the model was economically the most valid, although we couldn't conclude that it was statistically superior.

This page is intentionally left blank.

Chapter 8

Tables

Mean	6.258062
Standard Error	0.152163
Median	5.193571
Mode	9.348395
Standard Deviation	3.801026
Sample Variance	14.44780
Kurtosis	0.948456
Skewness	0.993428
Range	20.75091
Minimum	0.630497
Maximum	20.75091
Sum	3905.031
Count	624
95% Confidence Level	0.298814

Table 8.1: Three-Month Canadian Treasury-Bill Yield (Continuously Compounded), 1951–2002 Monthly Time Series Summary Statistics

Regression Stats						
Mult R	0.9904					
R^2	0.9810					
Adj R^2	0.9810					
Std Error	0.5240					
Obs	623					
ANOVA						
	df	SS	MS	F	Signif F	
Regression	1	8798.7	8798.7	32042	0	
Residual	621	170.53	0.2746			
Total	622	8969.3				
	Coefs	Std Error	t Stat	P-Value	Lower 95%	Upper 95%
Intercept	0.0696	0.0405	1.7200	0.0859	−0.0099	0.1492
Lagged Rate	0.9894	0.0055	179.00	0	0.9786	1.0003

Table 8.2: OLS Estimates: AR(1) Base Case

ML Coefficients		HMM Parameters	
Base Case: AR(1) Model			
ρ	0.989407	a	0.127794
g	0.069641	$\bar{r}$	6.574247
σ	0.524023	ς	1.824943
SSE	170.5266		
R^2	0.980988		
Hidden Markov Model: Constant σ			
ρ_1	0.910129	a_1	1.130027
ρ_2	1.015129	a_2	−0.180188
g_1	0.084838	r_1	0.943998
g_2	0.100863	r_2	−6.666865
σ	0.151707	ς_1	0.550458
		ς_2	0.521588
P_{12}	0.156964	A_{12}	2.621451
P_{21}	0.347454	A_{21}	5.802818
SSE	153.3317		
R^2	0.982196		
Hidden Markov Model: Constant σ and ρ			
ρ	0.980599	a	0.235100
g_1	0.065504	r_1	3.376321
g_2	0.172408	r_2	8.886552
σ	0.250578	ς	0.876544
P_{12}	0.104078	A_{12}	1.403520
P_{21}	0.107802	A_{21}	1.453739
SSE	168.0105		
R^2	0.981268		

Table 8.3: Obtaining Parameter Estimates from the Filtering Algorithm

This page is intentionally left blank.

Chapter 9

Figures

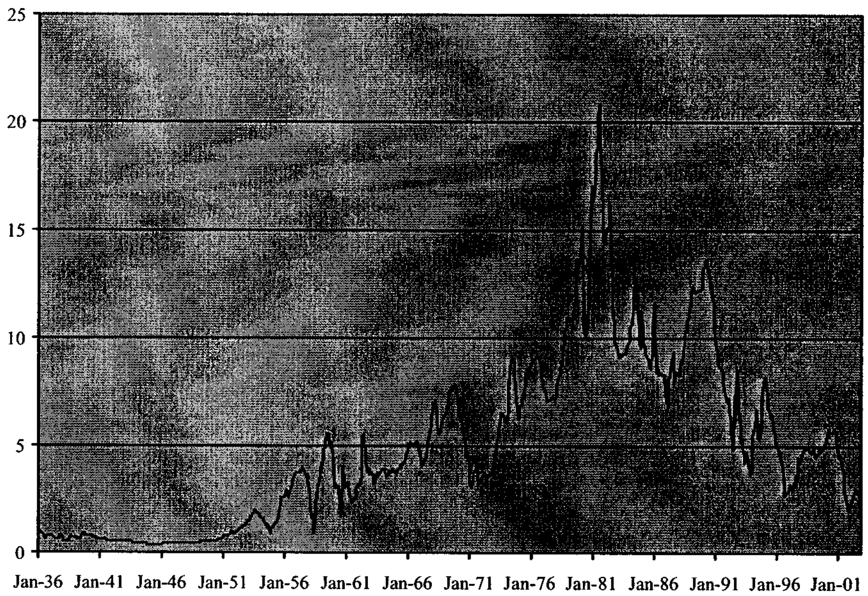

Figure 9.1: Three-Month Canadian Treasury-Bill Yield, 1936–2002

Bibliography

[1] Yacine Aït-Sahalia. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, 9(2), 1996, 385–426.

[2] Pierluigi Balduzzi, Sanjiv R. Das, and Silverio Foresi. The central tendency: A second factor in bond yields. *Review of Economics and Statistics*, 80(1), 1998, 62–72.

[3] Richard Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, New Jersey, 1961.

[4] K. C. Chan, G. Andrew Karolyi, Francis A. Longstaff, and Anthony B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47(3), 1992, 1209–1227.

[5] Kai Lai Chung, *Markov Chains With Stationary Transition Probabilities*, Springer-Verlag, New York, 1967.

[6] John C. Cox, Jonathon E. Ingersoll, Jr. and Stephen A. Ross, A theory of the term structure of interest rates, *Econometrica*, 53(2), 1985, 385–408.

[7] A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications*, Volume 23, pages 91–113, 1986.

[8] J. Driffill, Changes in regime and term structure, *Journal of Economic Dynamics and Control*, 16, 1992, 165–173.

[9] Darrell Duffie and Rui Kan, Multi-factor term structure models, *Philosophical Transactions: Physical Sciences and Engineering* 347(1684), Mathematical Models in Finance, The Royal Society, 1994, 577–586.

[10] Robert J. Elliott, *Stochastic Calculus and Applications*, Springer-Verlag, New York, 1982.

[11] Robert J. Elliott, New finite dimensional filters and smoothers for noisily observed Markov chains, *IEEE Transactions on Information Theory*, 39(1), 1993, 265–271.

[12] Robert J. Elliott. Exact adaptive filters for Markov chains observed in Gaussian noise. *Automatica*, Volume 30, pages 1399–1408, 1994.

[13] R. Elliott, P. Fischer, and E. Platen, Filtering and parameter estimation for a mean reverting interest rate model, *Canadian Applied Mathematics Quarterly*, 7(4), 1999.

[14] Robert J. Elliott and Hailaing Yang. How to count and guess well: Discrete adaptive filters. *Applied Mathematics & Optimization*, Volume 30, pages 51–78, 1994.

[15] Willy Feller, On the integro-differential equations of purely discontinuous Markoff processes, *Transactions of the American Mathematical Society*, 48(3), 1940, 488–515.

[16] Irving Fisher, Appreciation and interest, *Publications of the American Economic Association*, 11, 1896, 23–29, 91–92.

[17] S. Gray, Modeling the conditional distribution of interest rates as a regime switching process, *Journal of Financial Economics*, 42, 1996, 27–62.

[18] J. D. Hamilton, Rational expectations economic analysis of changes in regime, *Journal of Economic Dynamics and Control*, 12, 1988, 385–423.

[19] Asbjørn T. Hansen and Rolf Poulsen, A simple regime switching term structure model, *Finance and Stochastics*, Volume 4(4), 2000, 409–429.

[20] Michael J. Harrison and David M. Kreps, Martingales and arbitrage in multi-period security markets, *Journal of Economic Theory*, 20, 1979, 381–408.

[21] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1), 1992, 77–105.

[22] Steven L. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies*, 6(2), 1993, 327–343.

[23] Daniel P. Heyman and Matthew J. Sobel, *Stochastic Models in Operations Research, Volume I: Stochastic Processes and Operating Characteristics*, McGraw-Hill, New York, 1982.

[24] John R. Hicks, *Value and Capital*, 2nd ed. Oxford University Press, London, 1946.

[25] Thomas S. Y. Ho and Sang-Bin Lee, Term structure movements and pricing interest rate contingent claims, *Journal of Finance*, 41(5), 1986, 1011–1029.

[26] John Hull and Alan White, The pricing of options on assets with stochastic volatilities, *Journal of Finance*, 42(2), 1987, 281–300.

[27] John Hull and Alan White, Pricing interest-rate derivative securities, *Review of Financial Studies*, 3(4), 1990, 573–592.

[28] Madhu Kalimipalli and Raul Susmel, Regime-switching stochastic volatility and short-term interest rates, *Journal of Empirical Finance*, Forthcoming, 2004.

[29] Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus, Second Edition*, Springer-Verlag, New York, 1988, 1991.

[30] Guiseppe di Lampedusa, *The Leopard*, Pantheon, New York, 1960 (1957).

[31] Camilla Landén, Bond pricing in a hidden Markov model of the short rate, *Finance and Stochastics*, Volume 4(4), 2000, 371–389.

[32] B. Leblanc, *Modélisations de la Volitilité d'un Actif Financier et Applications*, Université de Paris VII, Thèse, 1997.

[33] Francis A. Longstaff and Eduardo S. Schwartz. Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance*, 47(4), 1992, 1259–1282.

[34] Friedrich A. Lutz, The structure of interest rates, *Quarterly Journal of Economics*, 55, 1940, 36–63.

[35] Franco Modigliani and Richard C. Sutch, Innovations in interest rate policy, *American Economic Review*, 56, 1966, 178–197.

[36] Marek Musiela, Stochastic PDEs and term structure models, *Journées Internationales de Finance, IGR-AFFI*, 1993.

[37] Marek Muliela and Marek Rutkowski, *Martingale Methods in Financial Modelling*, Springer-Verlag, New York, 1997.

[38] Antonio Roma and Walter Torous, The cyclical behavior of interest rates, *Journal of Finance*, 52(4), 1997, 1519–1542.

[39] Elias M. Stein and Jeremy C. Stein, Stock price distribution with stochastic volatility: an analytic approach, *Review of Financial Studies*, 4(4), 1991, 727–752.

[40] Licheng Sun. Nonlinearity and stochastic volatility: Dynamics of short term riskless interest rate. Working paper, University of Georgia, 2001.

[41] Oldrich Vasiček, An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 1977, 177–188

[42] Oldrich A. Vasiček and H. Gifford Fong, Term structure modeling using exponential splines, *Journal of Finance*, 37(2), 1982, 339–348.

Appendix A

Mathematical Proofs

Proof of Lemma 1 We have, for any $t \in [0,T]$,

$$M_t = X_t - X_0 - \int_0^t A(s)X_s\,ds. \tag{A.1}$$

Clearly $M_0 = 0$, (the zero vector in $\mathbf{R}^N$), and M is adapted to $\{\mathcal{F}_t\}$. Also,

$$\begin{aligned} E[M_t^\top M_t] &= E\Big[X_t^\top X_t - X_t^\top X_0 - \int_0^t X_t^\top A(s)X_s\,ds - X_0^\top X_t + X_0^\top X_0 \\ &\quad + \int_0^t X_0^\top A(s)X_s\,ds - \int_0^t X_s^\top A^\top(s)X_t\,ds \qquad \text{(A.2)} \\ &\quad + \int_0^t X_s^\top A^\top(s)X_0\,ds + \int_0^t\int_0^t X_s^\top A^\top(s)A(u)X_u\,ds\,du\Big] \\ &\le 2 + 4t\max_{s\in[0,t]}(\|A(s)\|) + t^2\max_{s\in[0,t]}(\|A(s)\|)^2 \\ &< \infty, \end{aligned}$$

where the $n\times n$ matrix norm is defined $\|A\| = \sup\{\|A\mathbf{x}\|; \mathbf{x}\in\mathbf{R}^n, \|\mathbf{x}\| = 1\}$. This shows that M is square integrable. Now, let $0 \le s \le t$, and apply Equation 2.11 to get

$$E[M_t - M_s \mid \mathcal{F}_s^X] = E\Big[X_t - X_s - \int_s^t A(u)X_u\,du \,\Big|\, \mathcal{F}_s^X\Big] = 0 \quad a.s. \tag{A.3}$$

Thus M is a martingale with respect to the filtration, $\{\mathcal{F}_t\}$. The function $t \mapsto \int_0^t A(s)X_s\,ds$ is differentiable, hence it has bounded variation on compact intervals. Therefore the semi-martingale decomposition of X is as claimed.

Proof of Lemma 2 Because $f_t = f(t)^\mathsf{T} X_t$, in light of Lemma 1 we can use Itô's integration by parts for general semi-martingales, (see Elliott [10] for details). This gives

$$f_t = f_0 + \int_0^t f(s^-)^\mathsf{T} \, dX_s + \int_0^t df(s)^\mathsf{T} X_{s^-} + [f(\cdot), X]_t \qquad \text{(A.4)}$$

where $f(s^-) = \lim_{t \uparrow s} f(t)$ is the left limit (similarly for X_{s^-}) and $[f(\cdot), X]_t$ is the general or *optional* quadratic covariation. We have, by continuity, $f(s^-) = f(s)$. Since $f(t)$ is of finite variation, the optional co-variation simply adds up the products of the jumps of the two processes, but $f(t)$ is also continuous, so it has no jumps. Thus the square bracket term is identically zero. Now since X_t is right continuous with left limits existing, it has a countable number of discontinuities, all of the jump type. Thus $X_{s^-} = X_s$ for Lebesgue almost every $s \in [0, t]$ and it can be replaced inside the Stieltjes integral, again because of the continuity of $f(t)$. The result follows by substituting the dynamics of X determined in Lemma 1 and noting that $f(s)$ is continuous and adapted, hence predictable and bounded on compact intervals, and M_s is a square-integrable martingale, so the stochastic integral of $f(s)$ with respect to dM_s is a zero-mean martingale.

Proof of Lemma 3 Let $g(t) = \exp(\int_0^t f_s \, ds) I$, and $g_t = g(t)^\mathsf{T} X_t$ a vector in $\mathbf{R}^N$. Clearly $g(t)$ is adapted and continuous, so from Lemma 2 we have

$$\begin{aligned} g_t &= g_0 + \int_0^t g(s)^\mathsf{T} A(s) X_s \, ds + \int_0^t \{dg(s)^\mathsf{T} X_s\} + \int_0^t g(s)^\mathsf{T} \, dM_s \\ &= g_0 + \int_0^t A(s) g_s \, ds + \int_0^t f_s g_s ds + \int_0^t g(s)^\mathsf{T} \, dM_s \qquad \text{(A.5)} \\ &= g_0 + \int_0^t \{A(s) + \text{diag}[f(s)]\} g_s \, ds + \int_0^t g(s)^\mathsf{T} \, dM_s. \end{aligned}$$

The last equality follows from

$$f_s g_s = (f(s)^\mathsf{T} X_s) X_s g(s) = \text{diag}[\tilde{f}(s)] X_s g(s) = \text{diag}[f(s)] g_s, \qquad \text{(A.6)}$$

because of our choice of the Markov chain's state space. Taking expectation and applying Fubini's theorem to take the expectation through the integral sign gives

$$E[g_t] = g_0 + \int_0^t \{A(s) + \mathrm{diag}[f(s)]\} E[g_s]\, ds. \tag{A.7}$$

This is the integral version of the homogeneous linear system of ordinary differential equations described in Equation 2.20, so it has the unique fundamental solution

$$E[g_t] = \Phi_f(t) g_0, \tag{A.8}$$

where $\Phi_f(t)$ is the fundamental matrix. Clearly $\Phi_f(t)$ is deterministic. Finally, because the state space consists of unit vectors, $g(t) = \tilde{1}^\mathsf{T} g_t$ and since $g_0 = X_0$,

$$E[g(t)] = \tilde{1}^\mathsf{T} \Phi_f(t) X_0. \tag{A.9}$$

Proof of Lemma 4 Denoting by $h(t)$, the vector with typical entry

$$h(t)_i = \exp\Big(\int_0^t \{f(s)^\mathsf{T} X_s + g(s)^\mathsf{T} e_i\}\, ds\Big), \tag{A.10}$$

we find that

$$dh(t)_i = \{f(t)^\mathsf{T} X_t + g(t)^\mathsf{T} e_i\} h(t)_i\, dt, \tag{A.11}$$

or

$$dh(t) = \{f(t)^\mathsf{T} X_t I + \mathrm{diag}[g(t)]\} h(t)\, dt. \tag{A.12}$$

Writing $h_t = h(t)^\mathsf{T} X_t$, we are interested in finding $E[h_t]$. Let $H_t = h_t X_t = \mathrm{diag}[h(t)] X_t$ and $H(t) = \mathrm{diag}[h(t)]$, then $h_t = \tilde{1}^\mathsf{T} H_t$. Since $H(t)$ is continuous, adapted and of bounded variation, we can apply Lemma 2 to get

$$\begin{aligned} H_t = H_0 &+ \int_0^t H(s)^\mathsf{T} A(s) X_s\, ds \qquad \text{(A.13)} \\ &+ \int_0^t [\{f(s)^\mathsf{T} X_s I + \mathrm{diag}[g(s)]\} H(s)]^\mathsf{T} X_s\, ds + \int_0^t H(s)^\mathsf{T}\, dM_s \\ = &\; H_0 + \int_0^t H(s)^\mathsf{T} \{A(s) + \mathrm{diag}[f(s) + f(s)]\} X_s\, ds + \int_0^t H(s)^\mathsf{T}\, dM_s \\ = &\; H_0 + \{\mathrm{diag}[h(s)] A(s) \mathrm{diag}[h(s)]^{-1} + \mathrm{diag}[f(s) + g(s)]\} H_s\, ds \\ &+ \int_0^t \mathrm{diag}[h(s)]\, dM_s. \end{aligned}$$

The matrix diag$[h(s)]$ is invertible since its entries are exponentials and thus strictly positive. Taking expectation gives the result.

Proof of Lemma 5 We know that W is continuous, and adapted to the filtration $\{\mathcal{F}_t^W \vee \mathcal{F}_T^X\}$. We wish to show that $\{W_t\}$ and $\{W_t^2 - t\}$ are both martingales with respect to this filtration. In this case, we can conclude by Lévy's characterisation that W is a standard Brownian motion with respect to $\{\mathcal{F}_t^W \vee \mathcal{F}_T^X\}$. To do this we fix $0 \le s \le t \le T$. Since $\mathcal{F}_s^W \vee \mathcal{F}_T^X$ is generated by sets of the type $A \cup B$, where $A \in \mathcal{F}_s^W$ and $B \in \mathcal{F}_T^X$, it is sufficient to restrict our attention to these sets when considering the conditional expectation. Let A and B be arbitrary sets of the above type, then

$$\begin{aligned} E[1_{A\cup B} W_t] &= E[(1_A + 1_B - 1_A 1_B) W_t] \qquad (A.14) \\ &= P(B)E[W_t] + E[1_A W_t] - P(B)E[1_A W_t], \end{aligned}$$

since $\mathcal{F}_T^X$ and $\mathcal{F}_s^W$ are independent. The first term is clearly zero, so

$$\begin{aligned} E[1_{A\cup B} W_t] &= (1 - P(B))E[1_A W_t] \\ &= (1 - P(B))E[1_A E[W_t|\mathcal{F}_s^W]] \qquad (A.15) \\ &= (1 - P(B))E[1_A W_s] \\ &= E[1_{A\cup B} W_s]. \end{aligned}$$

Thus, by definition of conditional expectation,

$$E[W_t|\mathcal{F}_s^W \vee \mathcal{F}_T^X] = E[W_s|\mathcal{F}_s^W \vee \mathcal{F}_T^X] = W_s \quad a.s., \qquad (A.16)$$

and W is a martingale with respect to the larger filtration. Similarly, for arbitrary A and B of the appropriate type described above,

$$E[1_{A\cup B}(W_t^2 - t)] = E[1_{A\cup B}(W_s^2 - s)]. \qquad (A.17)$$

This shows that $\{W_t^2 - t; t \in [0, T]\}$ is also a martingale with respect to the filtration $\mathcal{F}_T^X \vee \mathcal{F}^W$, and the result follows. We now turn to the second claim, and for this claim we require $A \in \mathcal{F}_T^W$, and $B \in \mathcal{F}_s^X$. We have, by the independence of $\mathcal{F}_s^X$ and $\mathcal{F}_T^W$,

$$\begin{aligned} E[1_{A\cup B} X_t] &= P(A)E[X_t] + (1 - P(A))E[1_B X_t] \qquad (A.18) \\ &= P(A)E[E[X_t|\mathcal{F}_s^X]] + (1 - P(A))E[1_B E[X_t|\mathcal{F}_s^X]] \\ &= P(A)E[E[X_t|X_s]] + (1 - P(A))E[1_B E[X_t|X_s]] \\ &= E[1_{A\cup B} E[X_t|X_s]]. \end{aligned}$$

Therefore

$$E[X_t|\mathcal{F}_T^W \vee \mathcal{F}_s^X] = E[E[X_t|X_s]|\mathcal{F}_T^W \vee \mathcal{F}_s^X] = E[X_t|X_s], \qquad \text{(A.19)}$$

so X retains the Markov property when conditioned on the larger filtration.

Proof of Lemma 6 Using Equation 4.3,

$$\begin{aligned} E[r_t \mid \mathcal{F}_s^r \vee \mathcal{F}_T^X] &= E\Big[e^{-K(t)}\Big(e^{K(s)}r_s + \int_s^t e^{K(u)}a(u)\bar{r}(u, X_u)\,du \\ &\quad + \int_s^t e^{K(u)}\sigma(u, X_u)\,dW_u\Big) \,\Big|\, \mathcal{F}_s^r \vee \mathcal{F}_T^X\Big] \\ &= e^{-K(t)}\Big(e^{K(s)}r_s + \int_s^t e^{K(u)}a(u)\bar{r}(u, X_u)\,du \qquad \text{(A.20)} \\ &\quad + E\Big[\int_s^t e^{K(u)}\sigma(u, X_u)\,dW_u \,\Big|\, \mathcal{F}_s^r \vee \mathcal{F}_T^X\Big]\Big) \\ &= e^{-K(t)}\Big(e^{K(s)}r_s + \int_s^t e^{K(u)}a(u)\bar{r}(u, X_u)\,du\Big) \quad a.s., \end{aligned}$$

where the last equality follows from Lemma 1 and the fact that $\mathcal{F}_s^r \vee \mathcal{F}_T^X \subset \mathcal{F}_s^W \vee \mathcal{F}_T^X$, and W is a Brownian motion with respect to that filtration. Since the integrand is bounded, the stochastic integral is a martingale. The last term depends on r_s and $\mathcal{F}_t^X$ only, so $E[r_t \mid \mathcal{F}_s^r \vee \mathcal{F}_T^X] = E[r_t \mid r_s \vee \mathcal{F}_t^X]$ almost surely. Proceeding to the second claim we have, by the tower property,

$$\begin{aligned} &E[r_t \mid \mathcal{F}_s^r \vee \mathcal{F}_s^X] = E[E[r_t \mid \mathcal{F}_s^r \vee \mathcal{F}_T^X]|\mathcal{F}_s^r \vee \mathcal{F}_s^X] \qquad \text{(A.21)} \\ &= E\Big[e^{-K(t)}\Big(e^{K(s)}r_s + \int_s^t e^{K(u)}a(u)\bar{r}(u, X_u)\,du\Big) \,\Big|\, \mathcal{F}_s^r \vee \mathcal{F}_s^X\Big] \\ &= e^{-K(t)}\Big(e^{K(s)}r_s + \int_s^t e^{K(u)}a(u)\bar{r}(u)^\top E[X_u \mid X_s]\,du\Big), \end{aligned}$$

by Tonelli's theorem, Lemma 5, and the Markov property for X.

Proof of Lemma 7 We consider the stochastic process $Y = \{Y_t; t \in [0, T]\}$, where

$$Y_t = \int_0^t e^{K(s)}\sigma(s, X_s)\,dW_s. \qquad \text{(A.22)}$$

We claim that, for any $n \in \mathbf{N}$, $t \in [0,T]$, and $\theta_1, \dots, \theta_n \in \mathbf{R}$, and any partition of $[0,t]$, $0 = t_0 \le t_1 \le \cdots \le t_n = t$, almost surely

$$E\Big[\exp\Big(\sum_{i=1}^{n} \theta_i Y_{t_i}\Big) \,\Big|\, \mathcal{F}_T^X\Big] = \exp\Big(\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \theta_i\theta_j \int_0^{t_i \wedge t_j} e^{2K(s)}\sigma^2(s, X_s)\,ds\Big). \tag{A.23}$$

Since

$$r_t = e^{-K(t)}\Big(r_0 + \int_0^t e^{K(s)} a(s)\bar{r}(s, X_s)\,ds + Y_t\Big), \tag{A.24}$$

this claim will be sufficient to prove the lemma. We use induction on n. Suppose $n = 1$. By Itô's rule,

$$\begin{aligned} e^{\theta Y_t} &= e^{\theta Y_0} + \int_0^t \theta e^{\theta Y_s}\,dY_s + \tfrac{1}{2}\int_0^t \theta^2 e^{\theta Y_s}\,d\langle Y\rangle_s \qquad \text{(A.25)} \\ &= 1 + \int_0^t \theta e^{\theta Y_s} e^{K(s)}\sigma^2(s, X_s)\,dW_s + \tfrac{1}{2}\theta^2 \int_0^t e^{\theta Y_s} e^{2K(s)}\sigma^2(s, X_s)\,ds. \end{aligned}$$

Thus, by Tonelli's theorem,

$$\begin{aligned} E[e^{\theta Y_t} \mid \mathcal{F}_T^X] &= 1 + E\Big[\int_0^t \theta e^{\theta Y_s} e^{K(s)}\sigma^2(s, X_s)\,dW_s \,\Big|\, \mathcal{F}_T^X\Big] \\ &\quad + \tfrac{1}{2}\theta^2 E\Big[\int_0^t e^{\theta Y_s} e^{2K(s)}\sigma^2(s, X_s)\,ds \,\Big|\, \mathcal{F}_T^X\Big] \qquad \text{(A.26)} \\ &= 1 + 0 + \tfrac{1}{2}\theta^2 \int_0^t E[e^{\theta Y_s} \mid \mathcal{F}_t^X] e^{2K(s)}\sigma^2(s, X_s)\,ds \quad a.s. \end{aligned}$$

The second term is zero by Lemma 5. This initial value problem has the unique solution

$$E[e^{\theta Y_t} \mid \mathcal{F}_T^X] = \exp\Big(\tfrac{1}{2}\theta^2 \int_0^t e^{2K(s)}\sigma^2(s, X_s)\,ds\Big) \quad a.s., \tag{A.27}$$

so the claim is true when $n = 1$. Similarly, for $0 \le s \le t \le T$,

$$e^{\theta Y_t} = e^{\theta Y_s} + \int_s^t \theta e^{\theta Y_u} e^{K(u)}\sigma^2(u, X_u)\,dW_u + \tfrac{1}{2}\theta^2 \int_s^t e^{\theta Y_u} e^{2K(u)}\sigma^2(u, X_u)\,du, \tag{A.28}$$

so, almost surely

$$E[e^{\theta Y_t} \mid \mathcal{F}_s^W \vee \mathcal{F}_T^X] = e^{\theta Y_s} + \tfrac{1}{2}\theta^2 \int_s^t E[e^{\theta Y_u} \mid \mathcal{F}_s^W \vee \mathcal{F}_T^X] e^{2K(u)}\sigma^2(u, X_u)\,du, \tag{A.29}$$

hence

$$E[e^{\theta Y_t} \mid \mathcal{F}_s^W \vee \mathcal{F}_T^X] = e^{\theta Y_s} \exp\Big(\tfrac{1}{2}\theta^2 \int_s^t e^{2K(u)}\sigma^2(u, X_u)\, du\Big) \quad a.s. \tag{A.30}$$

Now suppose the claim is true for $n = m$. Then

$$\begin{aligned}
E\Big[\exp\Big(\sum_{i=1}^{m+1} \theta_i Y_{t_i}\Big) \,\Big|\, \mathcal{F}_T^X\Big] &= E\Big[E\Big[\exp\Big(\sum_{i=1}^{m+1} \theta_i Y_{t_i}\Big) \,\Big|\, \mathcal{F}_{t_m}^W \vee \mathcal{F}_T^X\Big] \,\Big|\, \mathcal{F}_T^X\Big] \\
&= E\Big[\exp\Big(\sum_{i=1}^{m} \theta_i Y_{t_i}\Big) E[e^{\theta_{m+1} Y_{t_{m+1}}} \mid \mathcal{F}_{t_m}^W \vee \mathcal{F}_T^X] \,\Big|\, \mathcal{F}_T^X\Big] \qquad \text{(A.31)} \\
&= E\Big[\exp\Big(\sum_{i=1}^{m} \theta_i Y_{t_i} + \theta_{m+1} Y_{t_m}\Big) \,\Big|\, \mathcal{F}_T^X\Big] \\
&\quad \times \exp\Big(\tfrac{1}{2}\theta_{m+1}^2 \int_{t_m}^{t_{m+1}} e^{2K(s)}\sigma^2(s, X_s) ds\Big) \\
&= E\Big[\exp\Big(\sum_{i=1}^{m-1} \theta_i Y_{t_i} + (\theta_m + \theta_{m+1}) Y_{t_m}\Big) \,\Big|\, \mathcal{F}_T^X\Big] \\
&\quad \times \exp\Big(\tfrac{1}{2}\theta_{m+1}^2 \int_{t_m}^{t_{m+1}} e^{2K(s)}\sigma^2(s, X_s)\, ds\Big) \\
&= \exp\Big(\tfrac{1}{2} \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} \theta_i \theta_j \int_0^{t_i \wedge t_j} e^{2K(s)}\sigma^2(s, X_s)\, ds \\
&\quad + \sum_{i=1}^{m-1} \theta_i (\theta_m + \theta_{m+1}) \int_0^{t_m} e^{2K(s)}\sigma^2(s, X_s)\, ds \\
&\quad + \tfrac{1}{2}(\theta_m + \theta_{m+1})^2 \int_0^{t_m} e^{2K(s)}\sigma^2(s, X_s)\, ds\Big) \\
&\quad \times \exp\Big(\tfrac{1}{2}\theta_m^2 \int_{t_m}^{t_{m+1}} e^{2K(s)}\sigma^2(s, X_s)\, ds\Big) \\
&= \exp\Big(\tfrac{1}{2} \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \theta_i \theta_j \int_0^{t_i \wedge t_j} e^{2K(s)}\sigma^2(s, X_s)\, ds\Big) \quad a.s.
\end{aligned}$$

Therefore, by induction the claim is true for all $n \in \mathbf{N}$. The lemma follows.

Proof of Lemma 8 This follows from Lemma 5, Lemma 7, and the Markov property of Brownian motion. In particular, for $t \leq u_1 \leq \cdots \leq$

$u_n \le u$,

$$\begin{aligned} & E\Big[\exp\Big(\sum_{i=1}^{n}\theta_i r_{u_i}\Big)\ \Big|\ \mathcal{F}_t^r \vee \mathcal{F}_T^X\Big] \qquad (A.32)\\ &= E\Big[\exp\Big(\sum_{i=1}^{n}\theta_i e^{-K(u_i)}\Big(e^{K(t)}r_t + \int_t^{u_i} e^{K(s)}a(s)\bar{r}(s,X_s)\,ds \\ &\quad + \int_t^{u_i} e^{K(s)}\sigma(s,X_s)\,dW_s\Big)\Big)\ \Big|\ \mathcal{F}_t^r \vee \mathcal{F}_T^X\Big] \\ &= \exp\Big(\sum_{i=1}^{n}\theta_i e^{-K(u_i)}\Big(e^{K(t)}r_t + \int_t^{u_i} e^{K(s)}a(s)\bar{r}(s,X_s)\,ds\Big)\Big) \\ &\quad + E\Big[\exp\Big(\sum_{i=1}^{n}\theta_i e^{-K(u_i)}\int_t^{u_i} e^{K(s)}\sigma(s,X_s)\,dW_s\Big)\ \Big|\ \mathcal{F}_t^r \vee \mathcal{F}_T^X\Big]. \end{aligned}$$

Renaming the variable, $v = s + t$, Lemma 5, and the Markov property of Brownian motion reduce this case to that of Lemma 7.

Proof of Lemma 9 This result can be seen by approximating the Riemann integral with Riemann sums in such a way that they change monotonically as the partition mesh decreases. We have, with $\Pi = \{0 = t_0 \le t_1 \le \cdots \le t_n = T\}$ a partition on $[0,T]$,

$$\begin{aligned} R_{0,T} &= \lim_{\|\Pi\|\to 0}\sum_{i=1}^{n} r_{t_i}(t_i - t_{i-1}) \\ &= \lim_{\|\Pi\|\downarrow 0}\sum_{i=1}^{n} r_i^*(t_i - t_{i-1}) \qquad (A.33)\\ &= \lim_{\|\Pi\|\downarrow 0}\sum_{i=1}^{n} r_i^{**}(t_i - t_{i-1}), \end{aligned}$$

where $\|\Pi\| \downarrow 0$ indicates that each successive partition is a refinement of the previous partition, $r_i^* = \min(r_t; t \in [t_{i-1}, t_i])$, and $r_i^{**} = \max(r_t; t \in [t_{i-1}, t_i])$. By monotone convergence,

$$\begin{aligned} E[\exp(\theta R_{0,T})|\mathcal{F}_T^X] &= E\Big[\lim_{\|\Pi\|\to 0}\exp\Big(\sum_{i=1}^{n} r_{t_i}\theta(t_i - t_{i-1})\Big)\ \Big|\ \mathcal{F}_T^X\Big] \\ &= \lim_{\|\Pi\|\downarrow 0} E\Big[\exp\Big(\theta\sum_{i=1}^{n} r_i^*(t_i - t_{i-1})\Big)\ \Big|\ \mathcal{F}_T^X\Big] \qquad (A.34)\\ &= \lim_{\|\Pi\|\downarrow 0} E\Big[\exp\Big(\theta\sum_{i=1}^{n} r_i^{**}(t_i - t_{i-1})\Big)\ \Big|\ \mathcal{F}_T^X\Big], \end{aligned}$$

so, by the squeezing theorem,

$$E[\exp(\theta R_{0,T})|\mathcal{F}_T^X] = \lim_{\|\Pi\|\to 0} E\Big[\exp\Big(\sum_{i=1}^n r_{t_i}\theta(t_i - t_{i-1})\Big) \Bigm| \mathcal{F}_T^X\Big]. \quad \text{(A.35)}$$

Since $\{r_t\}$ is a Gaussian process when conditioned on the σ-algebra $\mathcal{F}_T^X$,

$$\begin{aligned} E[\exp(\theta R_{0,T})|\mathcal{F}_T^X] &= \lim_{\|\Pi\|\to 0} \exp\Big(\sum_{i=1}^n E[r_{t_i}|\mathcal{F}_T^X]\theta(t_i - t_{i-1}) \\ &\quad + \tfrac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \operatorname{cov}[r_{t_i}, r_{t_j}|\mathcal{F}_T^X]\theta^2(t_i - t_{i-1})(t_j - t_{j-1})\Big) \quad \text{(A.36)} \\ &= \exp\Big(\theta\int_0^T E[r_t|\mathcal{F}_T^X]\,dt + \tfrac{1}{2}\theta^2\int_0^T\int_0^T \operatorname{cov}[r_t, r_s|\mathcal{F}_T^X]\,ds\,dt\Big). \end{aligned}$$

Thus $R_{0,T}$ is conditionally normal, with

$$E[R_T|\mathcal{F}_T^X] = \int_0^T e^{-K(u)}\Big(r_0 + \int_0^u e^{K(s)}a(s)\bar{r}(s, X_s)\,ds\Big)\,du. \quad \text{(A.37)}$$

Rather than calculate the conditional variance from the conditional moment generating function above, we may calculate it directly from the definition:

$$\begin{aligned} \operatorname{var}[R_{0,T}|\mathcal{F}_T^X] &= E[(R_{0,T} - E[R_{0,T}|\mathcal{F}_T^X])^2|\mathcal{F}_T^X] \quad \text{(A.38)} \\ &= E\Big[\Big(\int_0^T\int_0^u e^{-K(u)}e^{K(s)}\sigma(s, X_s)\,dW_s\,du\Big)^2 \Bigm| \mathcal{F}_T^X\Big] \\ &= E\Big[\Big(\int_0^T\int_u^T e^{-K(s)}e^{K(u)}\sigma(u, X_u)\,ds\,dW_u\Big)^2 \Bigm| \mathcal{F}_T^X\Big] \\ &= E\Big[\int_0^T\Big(\int_u^T e^{-K(s)}\,ds\Big)^2 e^{2K(u)}\sigma^2(u, X_u)\,du \Bigm| \mathcal{F}_T^X\Big] \\ &= \int_0^T e^{2K(u)}\Big(\int_u^T e^{-K(s)}\,ds\Big)^2\sigma^2(u, X_u)\,du. \end{aligned}$$

The second last equality follows from Lemma 5 and the isometry property, and the change in order of integration is justified by Tonelli's Theorem.

Proof of Lemma 10 This proof is similar to the proof of Lemma 9, except we use the fact that $\{r_u; u \in [t, T]\}$ is conditionally a Gaussian process given $\{\mathcal{F}_t^r \vee \mathcal{F}_T^X\}$.

Proof of Lemma 11 Again, the proof of this is similar to the proof of Lemma 9; the interested reader can refer to that proof to fill in the details. We have

$$\begin{aligned}
&E[\exp(\theta_1 r_T + \theta_2 R_{0,T})|\mathcal{F}_T^X] \\
&= \lim_{\|\Pi\|\to 0} E\Big[\exp\Big(\theta_1 r_T + \sum_{i=1}^{n}\theta_2(t_i - t_{i-1})r_{t_i}\Big) \Bigm| \mathcal{F}_T^X\Big] \qquad \text{(A.39)}\\
&= \lim_{\|\Pi\|\to 0} E\Big[\exp\Big(\sum_{i=1}^{n-1}\theta_2(t_i - t_{i-1})r_{t_i} \\
&\quad + (\theta_1 + \theta_2(t_n - t_{n-1}))r_{t_n}\Big) \Bigm| \mathcal{F}_T^X\Big] \\
&= \lim_{\|\Pi\|\to 0} \exp\Big(\sum_{i=1}^{n-1}\theta_2(t_i - t_{i-1})E[r_{t_i}|\mathcal{F}_T^X] \\
&\quad + (\theta_1 + \theta_2(t_n - t_{n-1}))E[r_{t_n}|\mathcal{F}_T^X] \\
&\quad + \tfrac{1}{2}\theta_2^2 \sum_{i=1}^{n-1}\sum_{j=1}^{n-1}(t_i - t_{i-1})(t_j - t_{j-1})\mathrm{cov}[r_{t_i}, r_{t_j}|\mathcal{F}_T^X] \\
&\quad + \sum_{i=1}^{n-1}\theta_2(t_i - t_{i-1})(\theta_1 + \theta_2(t_n - t_{n-1}))\mathrm{cov}[r_{t_i}, r_{t_n}|\mathcal{F}_T^X] \\
&\quad + \tfrac{1}{2}(\theta_1 + \theta_2(t_n - t_{n-1}))^2 \mathrm{var}[r_{t_n}|\mathcal{F}_T^X]\Big) \\
&= \lim_{\|\Pi\|\to 0} \exp\Big(\theta_1 E[r_T|\mathcal{F}_T^X] + \theta_2 \sum_{i=1}^{n}(t_i - t_{i-1})E[r_{t_i}|\mathcal{F}_T^X] \\
&\quad + \tfrac{1}{2}\theta_1^2 \mathrm{var}[r_T|\mathcal{F}_T^X] \\
&\quad + \tfrac{1}{2}\theta_2^2 \sum_{i=1}^{n}\sum_{j=1}^{n}(t_i - t_{i-1})(t_j - t_{j-1})\mathrm{cov}[r_{t_i}, r_{t_j}|\mathcal{F}_T^X] \\
&\quad + \theta_1\theta_2 \sum_{i=1}^{n}(t_i - t_{i-1})\mathrm{cov}[r_{t_i}, r_T|\mathcal{F}_T^X]\Big) \\
&= \exp\Big(\theta_1 E[r_T|\mathcal{F}_T^X] + \theta_2 \int_0^T E[r_t|\mathcal{F}_T^X]\,dt + \tfrac{1}{2}\theta_1^2 \mathrm{var}[r_T|\mathcal{F}_T^X] \\
&\quad + \tfrac{1}{2}\theta_2^2 \int_0^T\int_0^T \mathrm{cov}[r_t, r_s|\mathcal{F}_T^X]\,ds\,dt + \theta_1\theta_2 \int_0^T \mathrm{cov}[r_t, r_T|\mathcal{F}_T^X]\,dt\Big).
\end{aligned}$$

This shows that r_T, and $R_{0,T}$ have the appropriate joint moment generating function. The covariance is determined by substituting the auto-covariance obtained in Lemma 7 into the last term of the above equation.

Proof of Lemma 12 This proof is similar to the proof of Lemma 11.

Proof of Lemma 13 See Elliott [12].

Proof of Lemma 14 See Elliott [12].

Proof of Lemma 15 Non-negativeness is clear. Now consider $P(\Lambda_n = 0) \leq P(\exists k \in \mathbf{N}; \lambda_k = 0) \leq P(\exists k \in \mathbf{N}; \epsilon_k = \pm\infty) = 0$, so Λ_n is positive a.s. Also $E[\Lambda_n|\mathcal{G}_{n-1}] = \Lambda_{n-1}E[\exp(-a\epsilon_n - \frac{1}{2}a^2)|\mathcal{G}_{n-1}] = \Lambda_{n-1}$, since ϵ_n is normal and independent of $\mathcal{G}_{n-1}$, and $a = (\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)/\langle\tilde{\sigma}, X_{n-1}\rangle$ is $\mathcal{G}_{n-1}$-measurable. This also shows that $E[\Lambda_n] = 1$, so Λ is a martingale with expectation 1.

Proof of Lemma 16 We proceed with the first claim first.

$$\begin{aligned}
\bar{P}(X_n = e_j|\mathcal{F}^X_{n-1}) &= \bar{E}[1_{\{X_n=e_j\}}|\mathcal{F}^X_{n-1}] = \bar{E}[1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}] \\
&= \frac{E[\Lambda_n 1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}]}{E[\Lambda_n|\mathcal{G}_{n-1}]} = E[\lambda_n 1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}] \\
&= E\Big[\exp\Big(-\frac{\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle}\epsilon_n \qquad (A.40)\\
&\quad -\frac{1}{2}\frac{(\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)^2}{\langle\tilde{\sigma}, X_{n-1}\rangle^2}\Big) \;\Big|\; \mathcal{G}_{n-1}\Big] E[1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}] \\
&= E[1_{\{X_n=e_j\}}|\mathcal{F}^X_{n-1}] = P(X_n = e_j|X_{n-1}).
\end{aligned}$$

For the second claim, since $\sum_{k=1}^n((\langle\tilde{\rho}, X_{n-1}\rangle + \langle\tilde{g}, X_{k-1}\rangle)/\langle\tilde{\sigma}, X_{k-1}\rangle)^2$ is finite almost surely, for all n, and $\{\Lambda_n\}$ is a martingale, a discrete version of Girsanov's theorem implies that the sequence, $\{\epsilon_n + (\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)/\langle\tilde{\sigma}, X_{n-1}\rangle\} = \{y_n/\langle\tilde{\sigma}, X_{n-1}\rangle\}$, is a sequence of i.i.d. standard normal random variables.

Proof of Lemma 17 For the first part, it is sufficient to show that the log likelihood is non-decreasing in p:

$$\begin{aligned}
\log(L_\alpha(\hat{\theta}_{p+1})) - \log(L_\alpha(\hat{\theta}_p)) &= \log\Big(\frac{L_\alpha(\hat{\theta}_{p+1})}{L_\alpha(\hat{\theta}_p)}\Big) = \log(L_{\hat{\theta}_p}(\hat{\theta}_{p+1})) \\
&\geq Q_{\hat{\theta}_p}(\hat{\theta}_{p+1}) \geq Q_{\hat{\theta}_p}(\hat{\theta}_p) = 0. \qquad (A.41)
\end{aligned}$$

In the above equation, equality holds if and only if $dP_{\hat{\theta}_{p+1}}/dP_{\hat{\theta}_p}$ is $\mathcal{Y}$ measurable and $\max_{\theta\in\Theta} Q_{\hat{\theta}_p}(\theta) = 0$. Since $Q_{\hat{\theta}_p}(\hat{\theta}_p) = 0$, the result follows from our restriction on the choice of $\hat{\theta}_{p+1}$.

Proof of Lemma 18 From Lemma 17 we have,

$$\hat{\theta}_{p+1} \in \{\arg\max_{\{\theta\in\Theta; L_\alpha(\theta)\geq L_\alpha(\hat{\theta}_p)\}} Q_{\hat{\theta}_p}(\theta)\}, \tag{A.42}$$

which is non-empty by Weierstrass' theorem.

Proof of Lemma 19 This proof is similar to that of Lemma 16.

$$\begin{aligned}
\hat{P}(X_n = e_j|\mathcal{F}^X_{n-1}) &= \hat{E}[1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}] \qquad (A.43)\\
&= \frac{E[\hat{\Lambda}_n 1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}]}{E[\hat{\Lambda}_n|\mathcal{G}_{n-1}]}\\
&= \sum_{\{r,s;p_{sr}>0\}} \left(\frac{\hat{p}_{sr}}{p_{sr}}\right)\langle e_r, X_{n-1}\rangle E[\langle e_s, X_n\rangle 1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}]\\
&= \sum_{\{r;p_{jr}>0\}} \left(\frac{\hat{p}_{jr}}{p_{jr}}\right)\langle e_r, X_{n-1}\rangle E[1_{\{X_n=e_j\}}|\mathcal{G}_{n-1}]\\
&= \sum_{\{r;p_{jr}>0\}} \left(\frac{\hat{p}_{jr}}{p_{jr}}\right)\langle e_r, X_{n-1}\rangle P(X_n = e_j|X_{n-1}),
\end{aligned}$$

which depends only on X_{n-1}. In particular

$$\begin{aligned}
&\hat{P}(X_n = e_j|X_{n-1} = e_i) \qquad (A.44)\\
&= \sum_{\{r;p_{jr}>0\}} \left(\frac{\hat{p}_{jr}}{p_{jr}}\right)\langle e_r, X_{n-1}\rangle P(X_n = e_j|X_{n-1} = e_i) = \left(\frac{\hat{p}_{ji}}{p_{ji}}\right)p_{ji} = \hat{p}_{ji}.
\end{aligned}$$

Proof of Theorem 7 We have, using Lemma 13,

$$\begin{aligned}
\sigma_n(H_nX_n) &= \sigma_n(\alpha_nX_n) + \sigma_n(\langle\beta_n, X_n\rangle X_n) + \sigma_n(\gamma_n f(y_{n-1}, y_n)X_n)\\
&= \sigma_n(\alpha_nPX_{n-1}) + \sigma_n(\alpha_n m_n) + \sigma_n(\langle\beta_n, PX_{n-1}\rangle PX_{n-1}) \qquad (A.45)\\
&\quad + \sigma_n(\langle\beta_n, PX_{n-1}\rangle m_n) + \sigma_n(\langle\beta_n, m_n\rangle PX_{n-1}) + \sigma_n(\langle\beta_n, m_n\rangle m_n)\\
&\quad + \sigma_n(\gamma_n f(y_{n-1}, y_n)PX_{n-1}) + \sigma_n(\gamma_n f(y_{n-1}, y_n)m_n).
\end{aligned}$$

Now, since m is a martingale increment, the terms involving a single m are zero, as we show, using the independence of X and ϵ, for the case of α:

$$\begin{aligned}\sigma_n(\alpha_n m_n) &= \bar{E}[\alpha_n m_n/\Lambda_n|\mathcal{F}_n^y] = \frac{E[\frac{\Lambda_T}{\Lambda_n}\alpha_n m_n \mid \mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \frac{E[E[E[\frac{\Lambda_T}{E[\Lambda_T|\mathcal{G}_n]}\alpha_n m_n \mid \mathcal{G}_n] \mid \mathcal{G}_{n-1}\vee\mathcal{F}_n^\epsilon] \mid \mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \frac{E[\alpha_n E[m_n|\mathcal{F}_{n-1}^X]|\mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} = 0.\end{aligned} \tag{A.46}$$

The other cases are similar. Also, we can make use of Lemma 14 in order to simplify the term with two m's as follows:

$$\begin{aligned}\sigma_n(\langle\beta_n, m_n\rangle m_n) &= \sigma_n(m_n m_n^\mathsf{T}\beta_n) = \frac{E[\frac{\Lambda_T}{\Lambda_n}m_n m_n^\mathsf{T}\beta_n \mid \mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \frac{E[E[E[\frac{\Lambda_T}{E[\Lambda_T|\mathcal{G}_n]}m_n m_n^\mathsf{T}\beta_n \mid \mathcal{G}_n] \mid \mathcal{G}_{n-1}\vee\mathcal{F}_n^\epsilon] \mid \mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \frac{E[E[m_n m_n^\mathsf{T}|\mathcal{F}_{n-1}^X]\beta_n|\mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \frac{E[(\mathrm{diag}[PX_{n-1}] - PX_{n-1}X_{n-1}^\mathsf{T}P^\mathsf{T})\beta_n|\mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \frac{E[\frac{\Lambda_T}{\Lambda_n}(\mathrm{diag}[PX_{n-1}] - PX_{n-1}X_{n-1}^\mathsf{T}P^\mathsf{T})\beta_n \mid \mathcal{F}_n^y]}{E[\Lambda_T|\mathcal{F}_n^y]} \\ &= \sigma_n(\mathrm{diag}[PX_{n-1}]\beta_n) - \sigma_n(PX_{n-1}X_{n-1}^\mathsf{T}P^\mathsf{T}\beta_n).\end{aligned} \tag{A.47}$$

Substituting these facts into the previous equation for $\sigma_n(H_n X_n)$ gives

$$\begin{aligned}\sigma_n(H_n X_n) &= \sigma_n(\alpha_n PX_{n-1}) + \sigma_n(PX_{n-1}X_{n-1}^\mathsf{T}P^\mathsf{T}\beta_n) \\ &\quad + \sigma_n(\mathrm{diag}[PX_{n-1}]\beta_n) - \sigma_n(PX_{n-1}X_{n-1}^\mathsf{T}P^\mathsf{T}\beta_n) \\ &\quad + \sigma_n(\gamma_n f(y_{n-1}, y_n)PX_{n-1}) \\ &= \sigma_n(\alpha_n PX_{n-1}) + \sigma_n(\mathrm{diag}[PX_{n-1}]\beta_n) + \sigma_n(\gamma_n f(y_{n-1}, y_n)PX_{n-1}).\end{aligned} \tag{A.48}$$

We continue by using the more cumbersome notation:

$$\sigma_n(H_n X_n) = \bar{E}\Big[\frac{1}{\Lambda_n}(\alpha_n PX_{n-1} + \mathrm{diag}[PX_{n-1}]\beta_n \tag{A.49}$$

$$
\begin{aligned}
&+ \gamma_n f(y_{n-1}, y_n) P X_{n-1}) \Big| \mathcal{F}_n^y\Big] \\
= \;& \bar{E}\Big[\tfrac{1}{\Lambda_{n-1}} \exp\Big(\frac{\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle}\epsilon_n \\
&+ \tfrac{1}{2}\frac{(\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)^2}{\langle\tilde{\sigma}, X_{n-1}\rangle^2}\Big) \\
&\times (\alpha_n P X_{n-1} + \mathrm{diag}[P X_{n-1}]\beta_n + \gamma_n f(y_{n-1}, y_n) P X_{n-1}) \Big| \mathcal{F}_n^y\Big] \\
= \;& \bar{E}\Big[\tfrac{1}{\Lambda_{n-1}} \exp\Big(\frac{\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle} \\
&\times \Big(\frac{y_n}{\langle\tilde{\sigma}, X_{n-1}\rangle} - \frac{\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle}\Big) \\
&+ \tfrac{1}{2}\frac{(\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)^2}{\langle\tilde{\sigma}, X_{n-1}\rangle^2}\Big) \\
&\times (\alpha_n P X_{n-1} + \mathrm{diag}[P X_{n-1}]\beta_n + \gamma_n f(y_{n-1}, y_n) P X_{n-1}) \Big| \mathcal{F}_n^y\Big] \\
= \;& \bar{E}\Big[\tfrac{1}{\Lambda_{n-1}} \exp\Big(\frac{\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle^2} y_n \\
&- \tfrac{1}{2}\frac{(\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)^2}{\langle\tilde{\sigma}, X_{n-1}\rangle^2}\Big) \\
&\times (\alpha_n P X_{n-1} + \mathrm{diag}[P X_{n-1}]\beta_n + \gamma_n f(y_{n-1}, y_n) P X_{n-1}) \Big| \mathcal{F}_n^y\Big] \\
= \;& \bar{E}\Big[\tfrac{1}{\Lambda_{n-1}} \bar{E}\Big[\exp\Big(\frac{\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle}{\langle\tilde{\sigma}, X_{n-1}\rangle^2} y_n \\
&- \tfrac{1}{2}\frac{(\langle\tilde{\rho}, X_{n-1}\rangle y_{n-1} + \langle\tilde{g}, X_{n-1}\rangle)^2}{\langle\tilde{\sigma}, X_{n-1}\rangle^2}\Big) \\
&\times (\alpha_n P X_{n-1} + \mathrm{diag}[P X_{n-1}]\beta_n \\
&+ \gamma_n f(y_{n-1}, y_n) P X_{n-1}) \Big| \mathcal{F}_n^y \vee \mathcal{F}_{n-1}^X\Big] \Big| \mathcal{F}_n^y\Big] \\
= \;& \bar{E}\Big[\tfrac{1}{\Lambda_{n-1}} \sum_{i=1}^{N} \exp\Big(\frac{\langle\tilde{\rho}, e_i\rangle y_{n-1} + \langle\tilde{g}, e_i\rangle}{\langle\tilde{\sigma}, e_i\rangle^2} y_n - \tfrac{1}{2}\frac{(\langle\tilde{\rho}, e_i\rangle y_{n-1} + \langle\tilde{g}, e_i\rangle)^2}{\langle\tilde{\sigma}, e_i\rangle^2}\Big) \\
&\times (\alpha_n P e_i + \mathrm{diag}[P e_i]\beta_n + \gamma_n f(y_{n-1}, y_n) P e_i) \\
&\times \bar{P}(X_{n-1} = e_i | \mathcal{F}_{n-1}^X) \Big| \mathcal{F}_n^y\Big] \\
= \;& \bar{E}\Big[\tfrac{1}{\Lambda_{n-1}} \sum_{i=1}^{N} \Gamma^i(y_{n-1}, y_n)(\alpha_n p_i + \mathrm{diag}[p_i]\beta_n + \gamma_n f(y_{n-1}, y_n) p_i)
\end{aligned}
$$

$$
\begin{aligned}
&\times \langle e_i, X_{n-1}\rangle \Big| \mathcal{F}^y_n\Big] \\
= &\sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\Big[\langle e_i, \bar{E}[\tfrac{1}{\Lambda_{n-1}}\alpha_n X_{n-1} \mid \mathcal{F}^y_n]\rangle p_i \\
&+ \operatorname{diag}[p_i]\bar{E}[\tfrac{1}{\Lambda_{n-1}}\beta_n\langle e_i, X_{n-1}\rangle \mid \mathcal{F}^y_n] \\
&+ \bar{E}[\tfrac{1}{\Lambda_{n-1}}\gamma_n\langle e_i, X_{n-1}\rangle \mid \mathcal{F}^y_n] f(y_{n-1}, y_n) p_i\Big] \\
= &\sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\Big[\langle e_i, \bar{E}[\tfrac{1}{\Lambda_{n-1}}\alpha_n X_{n-1} \mid \mathcal{F}^y_{n-1}]\rangle p_i \\
&+ \operatorname{diag}[p_i]\bar{E}[\tfrac{1}{\Lambda_{n-1}}\beta_n\langle e_i, X_{n-1}\rangle \mid \mathcal{F}^y_{n-1}] \\
&+ \bar{E}[\tfrac{1}{\Lambda_{n-1}}\gamma_n\langle e_i, X_{n-1}\rangle \mid \mathcal{F}^y_{n-1}] f(y_{n-1}, y_n) p_i\Big] \\
= &\sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)[\langle e_i, \sigma_{n-1}(\alpha_n X_{n-1})\rangle p_i + \operatorname{diag}[p_i]\sigma_{n-1}(\beta_n\langle e_i, X_{n-1}\rangle) \\
&+ \sigma_{n-1}(\gamma_n\langle e_i, X_{n-1}\rangle) f(y_{n-1}, y_n) p_i].
\end{aligned}
$$

The second last equality follows from the independence of the noise, ϵ, and the Markov chain, X, and the $\mathcal{G}$-predictability of the various processes.

Proof of Example 2. Considering the second term of $\sigma_n(J^{rs}_n X_n)$ gives

$$
\begin{aligned}
&\sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\operatorname{diag}[p_i]\sigma_{n-1}(\langle e_r, X_{n-1}\rangle e_s\langle e_i, X_{n-1}\rangle) \qquad \text{(A.50)} \\
&\quad= \sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\sigma_{n-1}(\langle e_r, X_{n-1}\rangle\langle e_i, X_{n-1}\rangle)\operatorname{diag}[p_i] e_s \\
&\quad= \sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\langle e_i, \sigma_{n-1}(\langle e_r, X_{n-1}\rangle X_{n-1})\rangle p_{si} e_s \\
&\quad= \sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\langle e_i, \operatorname{diag}[e_r]\sigma_{n-1}(X_{n-1})\rangle p_{si} e_s \\
&\quad= \sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\langle e_i, \langle e_r, \sigma_{n-1}(X_{n-1})\rangle e_r\rangle p_{si} e_s \\
&\quad= \sum_{i=1}^N \Gamma^i(y_{n-1}, y_n)\langle e_r, \sigma_{n-1}(X_{n-1})\rangle\langle e_i, e_r\rangle p_{si} e_s \\
&\quad= \Gamma^r(y_{n-1}, y_n)\langle e_r, \sigma_{n-1}(X_{n-1})\rangle p_{sr} e_s.
\end{aligned}
$$