University of Alberta

Market Models of Interest Rate Dynamics with a Joint Short Rate/HJM Approach

by

Rogemar S. Mamon ©

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

Mathematical Finance

Department of Mathematical Sciences

Edmonton, Alberta

Fall 2000

0-612-59628-1

University of Alberta

Library Release Form

Name of Author: Rogemar Sombong Mamon

Title of Thesis: Market Models of Interest Rate Dynamics with a Joint Short Rate/HJM Approach

Degree: Doctor of Philosophy

Year this Degree Granted: 2000

(Signed)..

11555 80 Ave. N.W.

Edmonton, AB, T6G 0R7

Date: 29 August 2000

UNIVERSITY OF ALBERTA

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Market Models of Interest Rate Dynamics with a Joint Short Rate/HJM Approach** submitted by **Rogemar S. Mamon** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in Mathematical Finance.

Dr. R.J. Elliott (Supervisor)

Dr. B.A. Schmuland

Dr. A. Cadenillas

Dr. R.J. Karunamuni

Dr. Vikas Mehrotra (Faculty of Business)

Dr. Gordon A. Sick (University of Calgary)

Date: 25 August 2000

Abstract

There are two principal representations for modelling, understanding and computing the evolution of the term structure of interest rates: the spot rate paradigm and the Heath-Jarrow-Morton (HJM) Models. The spot rate models are most often specified under the equivalent martingale measure. The HJM models on the other hand provide better ways to match observed variances and covariances of changes in bond prices.

The relationship between short-term rates and forward rates has long been established in the famous Expectations theory of the yield curve. The economic argument of this theory, asserting that implied forward rate profile represents market's expectations of future short rates, is proved formally here via mathematical expectation under the forward measure.

Starting with different short rate models we determine their related HJM forms. In turn, these HJM dynamics are expressed in terms of a related short rate model. We present a class of models where, using stochastic flows the dynamics of the term structure under the two paradigms are investigated and reconciled in each interest rate model. In particular, a model in which the short rate is a function of a continuous time Markov chain is considered. We then move on to consider models that capture mean reversion phenomenon. The Vasicek, (Ornstein-Uhlenbeck process) and the Cox-Ingersoll-Ross (CIR), (Bessel process) models are examples of these.

A characterisation of generalised exponential affine models for bond prices is also presented and the form of the forward and short rates calculated, given that we have such bond price models. From a Discrete Markov Model to continuous Vasicek and CIR models, we investigate three mixture models. These are the Hull-White Model, whose mean reversion level is time varying; the two-factor Gaussian model, a Vasicek model whose mean reverting level is Vasicek by itself; and a Vasicek model with a Markovian mean reverting level. The

results show that the closed form solution for the bond price for these models are obtainable and all of them are exponential affine.

The bond price for the two-factor Gaussian Model is derived using Stochastic Flows and their Jacobians. The bond price for the Markovian mean reverting level has an Ordinary Differential Equation (ODE) component involving a fundamental matrix solution. A simulation study using auxiliary filters, which enables the parameters to be estimated via the Expectation Maximisation (EM) algorithm, demonstrates the feasibility of this Markovian Mean Reverting Interest Rate Model.

The results provide a wide class of discrete, continuous and mixture models for the interest rate market and for pricing other derivatives securities.

"Most people see things as they are and ask why. But, I dream of things that never were and ask why not." -Adopted from the book "The Road Less Traveled."

Acknowledgments

I wish to express my sincere gratitude to Professor Robert J. Elliott for his excellent supervision, the summer financial support and the first year National Science and Engineering Research Council (NSERC) research assistantship. Without those financial aids and superb guidance, this dissertation and the entire Ph.D. degree could not have been completed in three years.

It is also a great pleasure to thank Prof. Elliott for his generosity in paying my travel expenses from his Social Sciences and Humanities Research Council (SSHRC) grant, in coming to Australia during the finalisation of this thesis. In particular, my trip to Sydney visiting the Treasury and Commodities Group of the Macquarie Bank had given me the insight of "what are being done" in the banking industry and an idea of "how are they being done" as far as term structure theory is concerned.

Indeed, I am fortunate to have Prof. Elliott as a supervisor and mentor. For me, he shall always be remembered as a convivial fellow, a very rare specie of a first rate mathematician gifted with a cheerful human sensibility.

Many thanks are also due to the members of my supervisory and examining committee: Dr. B.A. Schmuland, Dr. A. Cadenillas, Dr. R. Karunamuni, Dr. V. Merohtra and Dr. G. Sick; whose careful reading of the thesis and valuable comments led to several improvements.

The moral support and encouragement of two dear friends, Bhushan Gopaluni and Abdulkadir Hussein are equally appreciated. Bhushan and Abdul had given me back-up support concerning computing and programming related matters in Matlab, cutting the time half through involved in the numerical implementation of the filtering algorithm. They also introduced me to the beautiful art of producing consistent TEX files. Bushan, a Ph.D. student from the Department of Chemical Engineering in particular, did an excellent job in debugging and assisted me in translating the algorithm to computer codes. A discussion with him on stochastic control models had also augmented and widened my perspectives of how to unify theoretical and applied modelling considerations. I am pleased to acknowledge his significant contribution.

Barbara Jamieson, a mentor in Finance had also given invaluable support and advice. My heartfelt thanks go to her for the fruitful discussion on Hidden Markov Models (HMM) algorithms and other related materials on Risk Management and Finance theory.

Marion Benedict's accommodating and ever patient way of responding to the graduate students' needs deserves special mention. Such an act of kindness and extending help in whatever capacity she can, are motherly gestures that have made my three-year stay in the Department of Mathematical Sciences a memorable one.

Further, I am grateful to the staff of the Department of Applied Mathematics at the University of Adelaide, Australia for their hospitality during a short visit in May, 2000.

To the Department of Mathematical Sciences, University of Alberta for the Teaching Assistantships and Sessional Lecturerships I received, I say thank you so much. Both appointments of course, carry monetary support which is much needed by a graduate student but more so, they provided a great deal of opportunities in honing my teaching skills.

I am also exceedingly delighted to convey my utmost appreciation to Mrs. Concepcion R. Llavore for her motherly words of wisdom, constant encouragement and the great motivation I derived from her to pursue and finish my Ph.D. degree. Moreover, it is from her whom I learn the beautiful art of effective communication. Indeed, I am deeply grateful for this very important and indispensable skill of a lifetime.

To my parents, thank you for all the prayers; for giving me your wholehearted support to whatever endeavours that tick to my mind. Thank you for believing in me.

Finally, to the Divine Providence, for His mercy during the difficult times of my life, for His compassion everytime I go astray and for His grace that gives me strength and fortitude to perform my daily responsibilities until I reach the goal of completing this degree, I say, all glory and honours belong to Him, now and forever.

Contents

List of Tables

List of Figures

Introduction and Summary of Results

The relationship among yields and maturities of otherwise identical securities is generally referred to as the term structure of interest rates. The term structure of interest rates has been the focus of many studies over a century now. In an effort to understand better the movements of interest rates, researchers have attempted to identify processes and rational investors' behaviours that will help explain fluctuations in yield curves over time.

An understanding of the yield curve has become increasingly important for a multitude of reasons. Together with theoretical explorations, research on interest rate models is undertaken to generate more insight in the following practical pursuits:

1. Management of interest rate risk exposures particularly, in hedging derivatives securities which are interest rate sensitive;

2. Forecasting of economic growth implied by short- and long-term interest rate behaviours. Statistical investigations in the *Journal of Financial Economics, Journal of Finance* and other academic publications have documented modest, but reliable positive correlations between the slope of the term structure, (differences between long term and short term interest rates), and future rates of economic growth, see [42] and [65]. This is also substantiated in [7];

3. Determining market expectations of future interest rates and inflation. Information derived from market expectation serves as primary indicator for central banks and federal reserve banks in their attempt to control and influence the levels of interest rates, or when adopting a certain monetary policy;

4. Investigating movements of key interest rates used as benchmark by investment banks; for instance, non-callable US Treasury securities are the benchmark for examining interest rates because they are very liquid and considered risk-free; and

5. Term structure of interest rates is important to corporate treasurers, who must decide whether to borrow by issuing long- or short-term debt, and to investors who must decide whether to buy long- or short-term bonds.

We are guided by the fact that, in addition to mathematical artifacts, there is a need to understand the economic forces underlying the successful and better functioning of capital markets.

To begin with, we present the related financial economics theory of interest. Most of the analytical tools and a brief formalisation of finance theory for contingent claim valuation are included in the Appendices. Based on these interrelated disciplines, we wish to build models describing the dynamics and structure of the interest rate market.

We have indicated that in the following chapters of this thesis we view the market mathematically and statistically. Of course, this does not contradict the view of the fundamental analysts that interest rates can be assessed by studying business, government and other macroeconomic analysis. However, any individual is bound by information constraints and personal biases and foibles. Paradoxically, it is the result of differences of opinions and purposes which make a market exists.

Mathematical modelling is expected to find trends or fit data even when there is noise.

Certainly, the thesis does not guarantee excellent predictive performance of these market models. We based our construction on the premise that the processes in the investigation are all random and trust available market prices, which is the essence of the efficient market hypothesis. Having, therefore, formulated a mathematical model based on some assumptions, we proceed to model and investigate the interest rate dynamics.

To attain this major objective, we give a brief outline of how we intend to proceed and what each chapter contains.

Chapter 1 reviews selected background material on the study of term structure of interest rates. We give an overview of interest rate models from their existence and uses, to recent economic theories concerning beliefs of how short rates and forward rates are related, to several practical considerations and modern studies of term structure theories.

Chapter 2 gives a formal treatment of the two paradigms of bond pricing. A mathematical demonstration showing the equivalence of the three descriptions of term structure is presented. Economic arguments regarding future expectations of short rates and bond prices are presented in four versions.

In Chapter 3, the Unbiased Expectation Hypothesis is given a mathematical proof via the powerful tool of changing measures. There, we introduce the construction of a forward measure by specifying its corresponding Radon-Nikodym derivative. The forward measure approach also provides an ingenious way of decomposing the expectation problem embedded in the valuation formula into products of two simpler expectations; thus, facilitating the valuation process of contingent claims in a more efficient way.

In Chapter 4, we model the short rate as a function of continuous time Markov chain. The bond price is calculated under this model. The dynamics of the instantaneous forward rate $f(t,T)$ are obtained and the short rate r_t and $f(t,t)$ are reconciled.

The Vasicek model is explored in Chapter 5. The dynamics of $f(t,T)$ and the short rate are derived. The short rate is related to the $f(t,T)$ dynamics.

In Chapter 6, similar methodology and objectives to those of Chapter 5 are pursued but with the emphasis on the Cox-Ingersoll-Ross (CIR) model.

Chapter 7 considers a Generalised Exponential Affine Model of the form $exp[A(t,T)-B(t,T)r_t]$. General relationships describing the dynamics of $f(t,T)$ and properties of the functions $A(t,T)$ and $B(t,T)$ are established and several characterisations of the model are also given.

Mean reverting models for interest rates are the topics discussed in Chapter 8. We study examples of these models and derive the corresponding bond prices. Again, the ultimate aim is to reconcile the short rate and the HJM forms. The purpose of this exercise is to validate the dynamics of $f(t, T)$ so they are consistent with the Stochastic Differential Equation (SDE) describing the evolution of r_t. In particular, we investigate the Hull-White Model, Two-Factor Gaussian Model and the interest rate model with Markovian mean reverting level.

Chapter 9 aims to test empirically the model for a mean reverting level following the Markov process. We carry out the simulation and assessment of the predictive performance of this model through Hidden Markov Model (HMM) and Filtering techniques. A review of HMM methodology, in conjunction with a unit-delay model, is presented and optimal filters are computed for this implementation. Current research in econometrics indicates that the state space model with filtering is the most robust method to use for modelling term structure. This is the reason we favour the filtering method in our empirical investigation, in addition to its consistency with the Efficient Market Hypothesis. A novel feature of the Hidden Markov filtering is that the parameters are updated as new information arrives, thus the model is self-calibrating.

Finally, we summarise in Chapter 10 results of this study. Several concluding remarks pertaining to future directions and other avenues of possible research are included.

Chapter 1

Preliminaries

This chapter gives a brief overview of the theory of interests starting from historical perspectives to a survey of several commonly known related economics and financial theories concerning the significance of term structure, the existence and significance of interest in particular; to the hypothesis describing relationship between short-term rates and forward rates and recent research in the field of modern term structure theory.

Most of the material prerequisite to the study of the results contained in the succeeding chapters are covered in Appendices A and B. Appendix A outlines a review of selected and related topics in measure-theoretic probability and stochastic calculus. A section on quantitative finance theory is covered in Appendix B that discusses recent developments of the field and presents the foundation of modern pricing theory and contingent claims. This is an attempt to systematise the evolution of these background material suitably needed in the study of term structure in the light of new scientific approaches of modern finance. In this regard, formalisation such as this, aims at advancing the field.

In the spirit of the methodology of this presentation, we wish to emphasize on the rigorous development of the basic results in modern finance theory which lie at the heart of the remarkable range of current applications of martingale theory and stochastic processes to financial markets.

1.1 Overview of Interest Rates

We recognise that mathematical tools alone cannot make things happen. It is the integration of sound economic theories and financial theories with the appropriate mathematical tools that forms the cornerstone of today's financial modelling. The spirit of integration such as this is essential to interface quantitative artifacts and the commonly used economic and financial theories.

This section is descriptive rather than analytical. We set aside the mathematics first and consider economic fundamentals. As we are probing a very practical matter which is of utmost concern to everyone, rates of interest, sometimes referred to as cost of capital, it would be naive to just focus on mathematics. We review the historical development of the study of interest rates and present selected basic economic theories in conjunction with the aim of this thesis.

The formal study of this subject can be traced back to the pioneering work of Irving Fisher in his two books entitled, "The Rate of Interest," [53], and "The Theory of Interest," [54].

1.2 An Economic Rationale of Interest Rates

A number of different theories have been advanced to explain the existence of interest. All of them, however, fall into two general categories, one for the supply side of the transaction and one for the demand side.

On the *supply side,* the primary issue is *time preference.* Most individuals and business firms exhibit a strong preference to have access to dollars today rather than an equal number of dollars tomorrow. Dollars tomorrow can only be used to meet deferred needs in an uncertain future. Interest is then the price that is sufficient to cause individuals and firms to overcome their time preference to defer consumption. Even individuals and firms with a strong recognition of the need for future dollars can easily move current dollars into the future by saving. Specifically, this argument is supported by monetary theories, popularised by Maynard Keynes, which hold that the level of the money supply determines the supply side of the interest market.

On the *demand side,* the primary issue is the *productivity of capital.* Virtually all business firms need capital with which to operate successfully. Some of this capital generally comes from borrowing. In the long run, the firm will be successful only if the return on capital employed is greater than the cost of borrowing. Of course, not all borrowing is done by firms; much is also done by individuals and government to finance current consumptions and for purpose of investments.

Although these two major theories are quite different, they are in no way incompatible. In fact, quite to the contrary, they serve to complement each other.

Though the above discussion barely scratches the surface of some economic, or even psychological and philosophical theories attempting to explain the existence of interest, it gives us insights regarding Fisher's theoretical framework on the determination of the rates by supply of debt (i.e, the demand for loan) and the demand for debt (i.e. supply of capital). Indeed, this is self-evident in the subtitle of his book "The Rate of Interest" *as determined by the impatience to spend income and the opportunity to invest it.*

1.3 Determinants of the Level of Interest

So far, we have seen that basic economic theory suggests that rates of interest, like other prices, are established by supply and demand. This sounds simple, but in practice there is a large number of factors that come together in complex ways to determine rates of interest. There are four most relevant factors that influence the shape of the yield curve and the general level of interest.

1. *Federal Reserve Policy.* Economic theory asserts that the money supply has a major effect on both the level of economic activity and the rate of inflation. The Federal Reserve Board as in the case of the United States,

or the Bank of Canada, in the case of Canada, controls the money supply.

If a monetary board like the Federal Reserve slows growth in the economy, it slows growth in the money supply. At the outset, this action would cause interest rates to increase and stabilise inflation. On the other hand, the reverse occurs whenever the Federal Reserve Board loosens the money supply by reducing interest rates.

As a case study, the Federal Reserve Board tightened up the money supply six times in 1994, to keep inflation in check, thereby controlling the growth of the existing economic recovery. The Fed's tools are primarily in the short-term rates. In effect, this tightening had the direct effect of pushing short-term interest rates up sharply. Long-term rates followed. The Fed's action to control the inflation had affected the investors' expectations about inflation. Thus, long-term rates leveled off and even dropped slightly in some financial markets.

Therefore, we see that when the Fed intervenes actively in the financial markets, there is a distortion of the yield curve. There is a temporary situation of interest rates being too low if the Fed is easing credit. On the other hand, interest rates would be temporarily too high if the Fed is tightening credit. Long-term rates however are not really that much affected when Fed takes intervention actions, except to the extent that market follows rational expectations.

2. *Foreign Trade Balance.* Individual and businesses in Canada buy and sell to people and firms in other countries. So, if Canada imports more than it exports, there would be a foreign trade deficit. If trade deficits occur, there is a need to finance it. The main source of funds for doing this is through debt.

   Logically, the larger the deficit, the more that must be borrowed. In effect, the interest rates go up as more borrowing is incurred. When the Bank of Canada attempts to make the interest rates lower, this causes Canadian interest rates to fall below rates abroad. Foreigners will sell Canadian bonds, thereby Canadian bond prices will be depressed. The result would then be higher Canadian rates.

In essence, if there is a trade deficit, this will hinder the Bank of Canada to intervene in its effort of lowering interest rates during recession.

3. *Federal Deficits.* There are occasions when the federal government spends more than it gets from tax revenues. If this happens, a fiscal deficit occurs. This should be covered through the means of either printing money or borrowing. If the government prints money, expectations for future inflation increase which in return can raise interest rates. If the government opts to borrow, this demand for funds can drive interest rates up. *Ceteris paribus,* the larger the federal deficit the higher the interest rates.

   Nevertheless, it is hard to determine whether long- or short-term rates are more affected by federal deficits. It depends on how the deficit is financed.

4. *Business Activity.* Business conditions also influence interest rates, especially in times of recession. During recessions, both the demand for money and rate of inflation tend to fall. At the same time, monetary boards like the Fed or Bank of Canada tends to increase money supply in an effort to stimulate the economy.

   Therefore, with more supply and less demand, the result is a tendency for interest rates to decline during recessions. In times of recession, short-term rates decline more sharply than long-term rates due to the following: (i) The Fed operates mainly in the short-term sector, and such intervention has the strongest effect in this area, and (ii) long-term rates are reflections of average expected inflation rate over the next 20 to 30 years. This expectation usually does not change much, even when the current rate of inflation is low because of a recession.

1.4 Interest Rate Levels and Stock Prices

We argue that interest rates have two effects on corporate profits: First, other things remaining the same, interest being a cost, the higher the rate of interest, the lower the firm's profit. Second, interest rates affect the level of economic activity, and economic activity affects corporate profits. Apparently, interest rates affect stock prices because of their effects on profits. However, its significance is more from the competition it creates in financial markets between stocks and bonds.

If real interest rates rise sharply, investors can obtain higher returns in the bond market, which induces them to sell stocks and transfer funds from the stock market to the bond market. If there is a massive sale of stocks in response to rising interest rates, this could cause stock prices to plummet. Similarly, if interest rates decline, the reverse situation holds.

As an illustration, the Dow Jones Industrial Index in December 1991 rose to 10% in less than a month. The bullish market during this time was due, almost entirely, to the sharp drop in long-term interest rates.

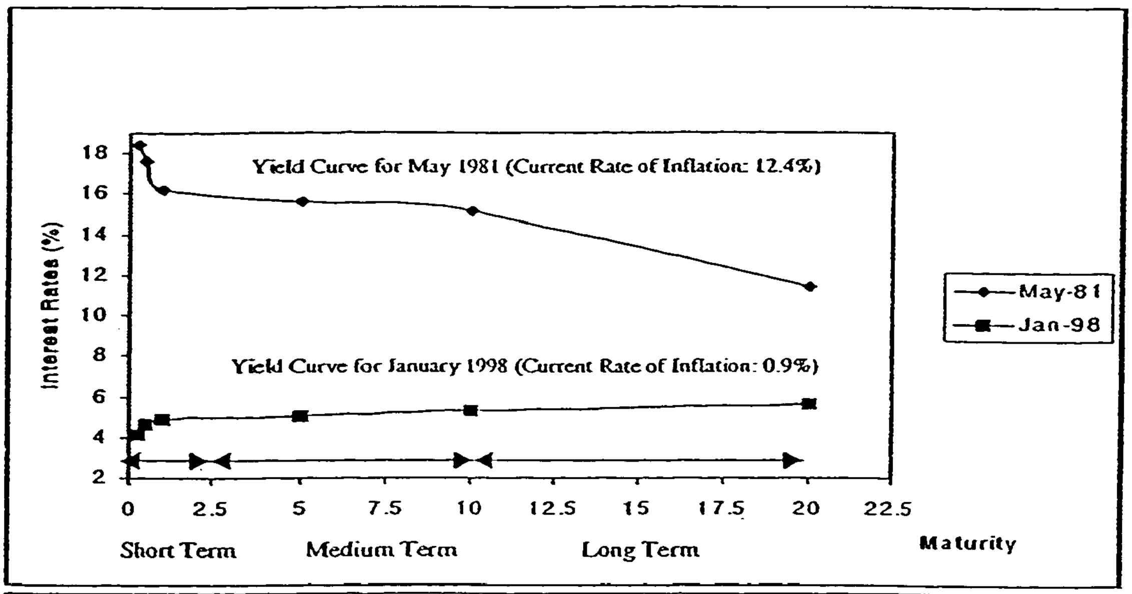
The bearish market of 1994, on the other hand, charactererised by declining common stock prices by more than 3% on average, was a result from a sharp increase in interest rates.

1.5 Interest Rates and Business Decisions

To describe the interaction between interest rates and business decisions, consider the yield curves for May 1981 and January 1998 on Canadian marketable bonds. From Figure 1.1, the January 1998 yield curve shows that short-term rates were lower than the long-term rates. Suppose at that time, a certain firm decided to (1) undertake a project (with positive net present value) having a 20-year life and will cost $10 million, and (2) raise the funds for the project by an issue of debt rather than by issuing stocks.

If the firm borrows on a short-term basis-say for one year-the rate on the loan might be 5%, so the interest cost for the year would be $500,000; if the firms uses long-term (20-year) financing, the rate might be 6%, and the cost of borrowing for the year would be $600,000. Therefore, at first glance, it would

Figure 1.1: Government of Canada Bond Interest Rates On Different Dates



Terms to Maturity	May 1981	January 1998
3 months	18.43%	4.18%
6 months	17.60	4.60
1 year	16.22	4.90
5 years	15.63	5.08
10 years	15.09	5.31
20 years	11.42	5.63

Source: Bank of Canada, Department of Monetary and Financial Analysis/ *Banque du Canada, Departement des Etudes monetaires et financieres.*
Inflation Rates taken from Statistics Canada.

seem that the firm should use short-term debt to finance the new project.

Nevertheless, this could prove to be a mistake. If short-term debt is used, the firm has to renew the loan every year, and the rate charged on each new loan will reflect the then-current short-term rate. Interest rates could return to their May 1981 levels, so by 1999 the firm could be paying 16.5%, or $1.65 million in interest per year. These high interest payments would cut and possibly eliminate the entire profits. The reduced profitability could easily increase the firm's risk to the point where its bond rating would be lowered, causing lenders to increase the risk premium built into the interest rates they charge, which in turn would force the firm to pay even higher rates. These very high interest rates would further reduce profitability, make lenders worry more, and causing them to be reluctant to renew the loan. If lenders refuse to renew the loan and demand payment, as they have the right to do so, the firm would have difficulty raising the cash. If the firm decided to make price cuts by converting physical assets to cash, heavy operating losses can ensue, and ultimately bankruptcy occurs.

If the firm used long-term financing, the cost of interest would remain constant at $600,000 per year. Thus, an increase in interest rates in the economy would not hurt the company. This would enable the firm to buy up some of its bankrupt competitors at bargain prices, as bankruptcies increase dramatically when interest rates rise, primarily because many firms do use short-term debt.

However, the above argument does not suggest that the firm should always avoid short-term debt. If inflation falls in the next few years, so will interest rates. If the company borrowed on a long-term basis for 6% in January 1998, it will be at a major disadvantage if competitors who used short-term debt in 1995 could borrow at a cost of only 4 or 5% in subsequent years. On the other hand, large federal deficits might drive inflation and interest rates up to a new record levels. In that case, the firm would wish to have had borrowed on a long-term basis in 1998.

Financing decisions would be easy, if accurate forecasts of future interest rates could be developed. However, predicting future interest rates with consistent accuracy is impossible. This work contributes to the modelling of interest rates and term structure.

1.6 More than Just One Interest Rate

Fisher stated "Instead of a single rate of interest, representing the rate of exchange between this year and next year, we now find a great variety of so-called interest rates." He continued, "These rates vary; because of *(1) risk,(2)* *the nature of security, (3) services in addition to the loan itself, (4) lack of free competition among lenders or borrowers, (5) length of time the loan has to run,* and other causes which most economists term *(6) economic frictions.*"

He added further that the long-term rates set a *rough norm* for the short rates which are much more variable. The stability of the economic system, call-terms (i.e., degree of liquidity), all influence the variability of rates.

This dissertation will focus only on the study of short rates and forward rates and explore their relationships.

We single out that at any particular point in time there is a vast array of interest rates being used in the myriad of financial transactions involving interest. However, a few key short term rates are widely watched as benchmarks of movements of interest rates. Three such key rates are:

1. *Prime rate*-the base rate used on high-grade corporate loans by major banks. Many loan rates are indexed to the prime rate.

2. *Federal funds rate*-the rate on reserves traded among commercial banks for overnight use. This rate changes daily and provides day-to-day information about interest rate movements.

3. *Discount rate*-the rate charged to member banks on loans by the Federal Reserve. Changes in this rate signal significant monetary policy adjustments by Federal Reserve Board which are likely to have widespread effect.

There is no single key indicator of long-term rates that is comparable to the above rates. The yields on Treasury bonds with a term of several years is a reasonable indicator of movements in long-term rates.

1.7 Term Structure Theories

We present here a brief summary of three different, but independent, theories concerning term structure.

The *expectations theory* is the simplest one and it is the one which has been the subject of empirical investigation for several decades. It had widespread appeal for the theoretical economists. It argues that a forward interest rate corresponding to a certain period is equal to the expected future spot interest rate for that preiod.

A second theory is known as *market segmentation theory.* By its name suggests, this theory hypothesizes that a relationship between short, medium and long-term interest rates does not necessarily exist. The reasoning behind this is as follows: The short-term interest rate is determined by supply and demand in the short-term bond market, the medium term interest rate is determined by supply and demand in the medium term bond market and so on. In other words, individuals and institutions investing in bonds of different maturity do not switch maturites. This means that the cross elasticity of demand is low, possibly zero; securities of different maturities are poor substitutes for one another.

The theory that conjectures that forward rates should always be higher than the expected future spot interest rates is known as *liquidity preference theory.* An inherent assumption is that investors prefer to have more liquidity and invest funds for short periods of time. This is a widely accepted theory, not necessarily inconsistent with the expectations hypothesis.

This theory is apparently the most appealing to the traders and investors. Borrowers usually prefer to borrow at fixed rates for long periods of time. If the forward rate equals expected future spot rate, interest rates would equal the average of expected future short-term interest rates. In the absence of any incentive to do otherwise, investors would tend to deposit their funds for short time periods and borrowers would tend to choose to borrow for long time periods. Thus, lending banks and financial intermediaries then find themselves financing susbstantial amounts of long-term fixed rate loans with short-term deposits. This entails excessive interest rate risk. It is therefore very necessary to match depositors with borrowers and avoid interest rate risk. Financial intermediaries in practice raise long-term interest rates relative to expected future short-term interest rates. This reduces the demand for long-term fixed rate borrowing and encourages investors to deposit their funds for long terms. This theory is consistent with the empirical evidence that yield curves tend to be upward sloping more often then they are downward sloping.

In *toto,* this theory of term structure, which often is treated as a modification of the expectations hypothesis, rests on the postulates that (1) the risks associated with holding long maturities are greater than those of holding short maturities, (2) the community prefers to avoid risk and (3) there are positive costs to society of obtaining the services of speculators.

The expectations hypothesis has been enunciated by Fisher, Keynes, Hicks, Lutz and others. It has a widespread appeal to the theoretical economists, primarily as a result of its consistency with the way similar phenomena in other markets, particularly future markets, are explained. In contrast, this hypothesis has been widely rejected by empirically minded economists and practical men of affairs. It was rejected by economists because investigators have been unable to produce evidence of a relationship between the term structure of interest rates and expectations of future short term rates. Nevertheless, Meiselman contends that previous investigators have not devised operational implications for the expectations hypothesis. Moreover, he contends that they have examined propositions which were mistakenly attributed to the expectations hypothesis, and when these propositions were found to be false, they rejected the expectations hypothesis.

Kessel's investigation, [81], showed that the term structure of interest rates can be explained better by a combination of the expectations and liquidity preference hypothesis than by either hypothesis alone. The two hypotheses can be viewed as complementary explanations of the same phenomenon. Support of this proposition was carried by the previous works of Macaulay, Culbertson, Meiselman, Walker and Hickman.

Available evidence shows that forward rates are higher estimates of future spot rates. This result is consistent with the Keynesian theory of "normal backwardation." The implications of this theory for the money and capital markets have been developed by Hicks in "Value and Capital." Hence, these findings support the Hicksian view that forward rates are equal to spot rate plus a liquidity premium.

In [81], Kessel concluded that correlations between forward and spot rates suggest that the market does have some power to foresee, up to a year in the future, spot rates from a month to a year to a maturity. This same conclusion is reached if forward rates, adjusted for liquidity premiums are used to predict subsequently observed spot rates and if the mean squared error is computed. Using either criterion, expectations theory seems to predict better than an inertia model.

So this classical theory predicts that in the absence of risk premia, forward rates will be equal to expected short rates. The modern theory however shows clearly that this will not generally be the case, even when risk premia are zero, unless the path of future interest rates is certain.

This can be shown by noting that the exponential function is convex. Therefore, by Jensen's inequality

$$P(t,T) = E_{\bar{P}}\left[exp - \int_t^T r_u du\right] > exp\left[E_{\bar{P}}\left(-\int_t^T r_u du\right)\right],$$

for some risk-adjusted probability measure $\bar{P}$.

In other words, the bond price is higher under the modern theory than if the forward rate curve were equal to the "risk-adjusted" path of the expected short rates.

This "price premium" or "yield discount," is a direct consequence of uncertainty in future interest rates and is largely unnoticed in traditional theory.

However, we shall show in this work that forward rates can still be equal to expected short rates. The evaluation of the expectation is with respect to a different measure, which we shall call a forward measure.

1.8 Studies on Modern Term Structure Modelling

The early work of "modern" term structure theories can be traced back to 1977 when Vasicek and Cox, Ingersoll and Ross (CIR) developed simultaneously similar models. Vasicek adopted a normality assumption while Cox, Ingersoll and Ross used a non-central chi-squared distribution. Recognising that single-factor models are over simplistic, multi-factor models were later developed to solve the curve fitting problem of the models by Vasicek and CIR. The early work on this area are the Brennan and Schwartz's model in 1978, Richard's model also in 1978 and Langetieg's model in 1980. While Longstaff and Schwartz's model and Chen and Scott's model in 1992 are representative of recent developments.

Longstaff and Schwartz (1992) developed two-factor term structure models where either an arbitrage-free or utility- based methodology is used. In the Richard and Brennan-Schwartz's model, factors are chosen arbitrarily.

In finance literature, models developed under a utility function are called *equilibrium models* while those which are formulated using the risk-neutral methodology are called *arbitrage-free models.*

Along with the developments of multi-factor models have been the estimation techniques for the parameters in these models. The regression method, generalised method of moments, maximum likelihood estimation, and most importantly the state space model have been used in estimating multi-factor Vasicek or CIR models.

Taking a slightly different view from the multi-factor models, researchers developed another series of models that take observed yield curve as given, so that fitting becomes never a problem. This goal is accomplished by making the parameters in the Vasicek or CIR model time dependent. The early work is the Ho and Lee model of 1986. They model the uncertainty by putting perturbation functions on the forward prices. The Black, Derman and Toy (1990), or BDT model is similar to the Ho-Lee model except that the distribution of the short rate is lognormal. The BDT model is richer than the Ho-Lee model because it also fits the volatility curve.

Dybvig, although he never published his work, gave an extension to the Ho-Lee model which is in spirit similar to a published work by Hull and White in 1990. Heath, Jarrow and Morton in 1992 provided a framework which relates forward and spot rates. Both the continuous time models of HJM and Hull and White let the parameters in the stochastic processes of the instantaneous rate be deterministic functions of time. All of these models are considered "time-dependent" models. These models can fit the yield curve but they do not have an easy form for the bond price.

An important influence of 1960's research on investment practice was the Samuelson-Fama efficient market hypothesis, which holds in a well-functioning and informed capital market. Asset-price dynamics are described by a sub-martingale in which the best estimate of an asset future price is the current price, adjusted for a "fair" expected rate of return. Under this hypothesis, attempts to use past price data or publicly available forecasts about future economic fundamentals to predict future security prices are doomed to failure. In essence, this theory supports why most models that we shall consider in the following chapters use processes which have the Markov property.

1.9 Term Structure and Monetary Regimes

Since we aim to shed light on issues relating term structure and the practice of monetary authorities in influencing the level of interest rates, we briefly review recent works on this area.

Along the lines of the analysis of the work mentioned in the succeeding paragraph, the mean reverting models that we propose in Chapter 8 could serve as alternative models. Such analysis will not be carried out here. Rather, we shall leave the emprical testing of these models with respect to the analysis of the work that we are about to describe to interested readers.

There is one empirical investigation concerning mean reverting models performed in Chapter 9. This, however, only intends to demonstrate the feasibility of applying the models to data, and to obtaining optimal estimates of parameters via filtering techniques.

Of particular interest to us are two papers on term structure models where the interest rate processes do not have continuous sample paths. These papers were presented at Isaac Newton Institute for Mathematical Sciences of the University of Cambridge in 1995.

In their paper entitled "Term Structure Modelling under Alternative Official Regimes," S.H. Babbs and N.J. Webber developed a class of term structure models when there are n state variables modelled as diffusion processes and an additional m state variables modelled as pure jump processes.

Under this study, the processes are not independent of each other. Furthermore, the spot rate is a specified function of the $n+m$ state variables. The authors presented an example where a jump process is a floor and another is a ceiling for the spot rate. This realistically reflects the actions of government monetary authorities especially when they exercise control in setting discount rates, Lombard rates and so forth over short-term rates.

The paper entitled "Interest Rate Distributions, Yield Curve Modelling and Monetary Policy" by L. El-Jahel, H. Lindberg and W. Perraudin treats a similar problem to that of the Babbs-Webber model. However, since the problem is a specific case, a closed form solution is obtained. El-Jahel, Lindberg and Perraudin explained two phenomena which influence the distribution of short-term interest rates: the practice of many monetary authorities of pegging an interest rate at the short end of the yield curve and periodically adjusting it in discrete jumps, and the attitude of monetary authorities in their reaction to inflationary shocks, either relaxed or stringent. The leptokurtosis of the short rate process is affected by the first of these phenomena. They also employed a variety of econometric techniques including among other things, non-parametric kernel estimates, unit root tests and simple autoregressions in examining the distributional properties of short-term interest rates in the U.K., the U.S. and Germany.

The techniques in their study were applied to the two commonly used single state variable yield curve models-those of Vasicek and Cox-Ingersoll-Ross. The parameters in these models were estimated with the aim of finding significant evidence in the above-mentioned data for misspecification of both models. They found that the mean reversion rate of adjustment of the short-term process tend to be overestimated. This empirical evidence prompted them to propose a Babbs-Webber model where the short-term interest rates are specified by a pure jump process whose jump rate is a function of a diffusion process. The assumptions they used are: an Ornstein-Uhlenbeck process for the diffusion variable and a quadratic function for the jump rate. They then obtained the power series representations of the conditional distribution of the diffusion state variable, given its past and bond yields by implementing the Karhunen-Loeve eigenfunction expansion techniques of physics. The expansion techniques may be applied to give similar representations of interest rate-based derivative values.

Chapter 2

The Term Structure of Interest Rates

2.1 The Bond Market Structure

Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ and a Brownian motion W_t, $0 \leq t \leq T$.

Suppose at time t, we have a riskless asset, a bond S_t^0 and a single risky asset S_t^1. Further, these assets have dynamics:

$$S_t^0 = exp \int_0^t r_u du \quad \text{and}$$

$$S_t^1 = S_0^1 + \int_0^t \alpha(u) S_u^1 du + \int_0^t \sigma(u) S_u^1 dW_u.$$

The functions r, α, and σ are all adapted stochastic processes. Let (H^0, H^1) denote a self-financing strategy. The wealth process for this trading strategy is given by

denote a self-financing strategy. The wealth process for this trading strategy is given by

$$X_t(H) = H_t^0 S_t^0 + H_t^1 S_t^1$$

and

$$\begin{aligned} dX_t(H) &= r H_t^0 S_t^0 dt + H_t^1 dS_t^1 \\ &= r(X_t - H_t^1 S_t^1)dt + H_t^1 dS_t^1. \end{aligned}$$

Suppose $\theta = \frac{\alpha(t)-r(t)}{\sigma(t)}$ and under the measure P^θ, the process W^θ is a Brownian motion, with

$$dW_t^\theta = \theta(t)dt + dW_t.$$

Under P^θ, the discounted wealth process $V_t = \frac{X_t}{S_t^0}$ has dynamics,

$$d\left[\frac{X_t}{S_t^0}\right] = H_t^1\sigma(t)\frac{S_t^1}{S_t^0}dW_t^\theta. \tag{2.1.1}$$

Since the drift of the SDE in (2.1.1) is zero, the discounted wealth process V_t is a P^θ−martingale.

Let $\phi \in L^2(\Omega, \mathcal{F}_T)$ be a contingent claim. Thus,

$$M_t := E^\theta\left[\frac{\phi}{S_T^0}\middle|\mathcal{F}_t\right] \quad \text{is a martingale}$$

and

$$M_t = M_0 + \int_0^t \varphi_u dW_u^\theta$$

for some adapted process φ, according to the Martingale Representation Theorem.

Suppose we take

$$H_t^1 = \frac{S_t^0\varphi_t}{\sigma(t)S_t^1}, \quad X(0) = M_0 = E^\theta\left[\frac{\phi}{S_T^0}\right]$$

and let

$$M_t = \frac{X_t}{S_t^0} = X(0) + \int_0^t H_u^1\sigma(u)\frac{S_u^1}{S_u^0}dW_u^\theta.$$

Then, with

$$H_t^0 = \frac{X_t}{S_t^0} - H_t^1\frac{S_t^1}{S_t^0},$$

(H^0, H^1) is a replicating strategy. That is,

$$X_T = H_T^0S_T^0 + H_T^1S_T^1 = \phi.$$

We see that the natural price for the claim at time 0 is $E^\theta\left[\frac{\phi}{S_T^0}\right]$.

Consequently, at time $t \in [0,T]$, the price of a claim ϕ is

$$\begin{aligned} X_t(\phi) &= X_t = S_t^0 E^\theta \left[\frac{\phi}{S_T^0} \middle| \mathcal{F}_t \right] \\ &= S_t^0 E^\theta \left[\frac{X_T}{S_T^0} \middle| \mathcal{F}_t \right], \end{aligned}$$

since $\frac{X_t}{S_t^0}$ is a martingale under P^θ.

We can extend this pricing formulation to a market with a bond price S_t^0 as the numéraire, with dynamics

$$dS_t^0 = r_t S_t^0 dt, \quad S_0^0 = 1$$

and d risky securities, $S_t^1, \cdots, S_t^d$ with dynamics

$$dS_t^i = S_t^i \left(\alpha_i(t)dt + \sum_{j=1}^{m} \sigma_{ij}(t) dW_t^j \right).$$

Again, $W_t = (W_t^1, \cdots, W_t^m)$ is an $m-$dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$.

Provided there exists a unique risk-neutral measure P^θ, this pricing valuation holds true.

The price for a claim $\phi \in L^2(\mathcal{F}_T)$ at time $t \leq T$ is therefore given by

$$X_t = S_t^0 E^\theta [\phi (S_T^0)^{-1} | \mathcal{F}_t].$$

Remarks: In the preceeding discussion, we have given the motivation for a "rational" pricing of contingent claims. Further, we have shown that martingales are tradables and non-martingales are non-tradables.

Now, in a market with d risky assets, we have

$$W_t^\theta = W_t + \left(\frac{\alpha_i(t) - r_t}{\sigma_i(t)} \right) t,$$

which is a $P^\theta-$ Brownian motion for $i = 1, \cdots, d$.

This can only happen if the two changes of the drift are the same. In other words,

$$\frac{\alpha_i(t) - r_t}{\sigma_i(t)} = \frac{\alpha_j(t) - r_t}{\sigma_j(t)},$$

for any pair (i, j). The term $\frac{\alpha_t - r_t}{\sigma_t}$ is called the **market price of risk,** where α represents the growth rate of the tradable, r as the growth rate of the bond and σ measures the risk of the asset.

Intuitively, the market price of risk can be thought of as the extra return above the risk-free rate per unit of risk. The market price of risk is defined using the drift and volatility functions of the stock price. We shall see in Chapter 7 (Section 7.4, Equation 7.4.12) that this terminology will be defined in terms of the drift and volatility structures of the spot rate and the two forms are apparently similar.

We therefore reach a conclusion - *all tradables in a market should have the same market price of risk.*

Definition 2.1.1 *A zero coupon bond maturing at time T is a claim that pays 1 at time T. The bond price $P(t,T)$ at time $t \in [0, T]$ is*

$$P(t, T) = S_t^0 E[(S_T^0)^{-1} | \mathcal{F}_t].$$

With the valuation formula, and $S_t^0 = exp(\int_0^t r_u du)$,

$$P(t, T) = E\left[exp\left(-\int_t^T r_u du \middle| \mathcal{F}_t\right)\right]$$

where E denotes the expectation under a martingale measure.

With this result, if $P(t, T)$ is a bond price at time t, a self-financing strategy (H_t^0, H_t^1) can be constructed such that the associated wealth process X_t at time t, $H_t^0 S_t^0 + H_t^1 S_t^1$, will have a value of 1 at time T.

Oftentimes, the rate r_t is deterministic. Thus, in this case, $P(t, T) = exp\left(-\int_t^T r_u du\right)$ and $dP(t, T) = r_t P(t, T)dt$ and therefore H_t^1 is identically 0.

In the interest-rate market, given a bond with a price $P(t,T)$, the forward rate $f(t,T)$ and the yield $Y(t,T)$ can be written in terms of the bond prices $P(t,T)$ as:

$$f(t,T) = -\frac{\partial}{\partial T} \log P(t,T) \tag{2.1.2}$$

and

$$Y(t,T) = -\frac{\log P(t,T)}{T-t}. \tag{2.1.3}$$

Conversely, using the above formulation in (2.1.2) and (2.1.3), the bond price can be given in terms of the forward rates or yields as:

$$P(t,T) = exp\Big(-\int_t^T f(t,u)du\Big)$$

and

$$P(t,T) = exp\Big(-(T-t)Y(t,T)\Big).$$

In other words, the three descriptions of the yield curve $f(t,T)$, $Y(t,T)$ and $P(t,T)$ are equivalent. The derivations of (2.1.2) and (2.1.3) are given in the next section.

For emphasis, $Y(t,T)$ is the average yield of the bond over its lifetime $T-t$ while $f(t,T)$ is the price at time t of instantaneous borrowing at time T.

If we consider $Y(t,t) = f(t,t) = r_t$, then all of these represent the instantaneous spot borrowing.

2.2 The Heath-Jarrow-Morton (HJM) Framework

The HJM model is built upon with an underlying motivation through the concept of forward rate agreement. Consider a time horizon, $0 \le t \le T <$

$T + \epsilon < T^*$. We wish to enter into a contract to borrow 1 at the future time T and repay it with interest at the time $T + \epsilon$. Whatever the rate of interest to be paid between T and $T + \epsilon$, it will be agreed today, and thus it must be $\mathcal{F}_t$−measurable.

Definition 2.2.1 *The instantaneous rate for the amount of dollars borrowed at time T, agreed upon time $t \leq T$, is the* **forward rate** $f(t, T)$.

We can replicate the above transaction by beginning to suppose that we have a portfolio consisting of:

Portfolio A: A zero-coupon bond $P(t, T)$ with a face value of 1 at time T and;

Portfolio B: A zero-coupon bond amounting to $\frac{P(t,T)}{P(t,T+\epsilon)}$ at maturity, $T+\epsilon$.

The value of these two portfolios are equal. This is because at time t, portfolio A is worth $P(t, T)$ while portfolio B has value $\frac{P(t,T)}{P(t,T+\epsilon)} \cdot P(t, T + \epsilon)$.

Thus,

$$P(t, T) - \frac{P(t, T)}{P(t, T + \epsilon)} \cdot P(t, T + \epsilon) = 0. \tag{2.2.4}$$

Equation (2.2.4) is equivalent to having a long position in portfolio A and simultaneously taking a short position in portfolio B at time t.

At time $T \geq t$, \$1 is received for portfolio A. An amount of $\frac{P(t,T)}{P(t,T+\epsilon)}$ must be paid for portfolio B at time $T + \epsilon$. Therefore, we have a transaction where, \$1 is borrowed at time T and \$ $\frac{P(t,T)}{P(t,T+\epsilon)}$ is paid at time $T + \epsilon$.

The interest paid on the dollar received at time T is $Y(t, T, T + \epsilon)$ and satisfies

$$\frac{P(t, T)}{P(t, T + \epsilon)} = exp[\epsilon \cdot Y(t, T, T + \epsilon)].$$

From which we obtain

$$Y(t, T, T + \epsilon) = -\frac{1}{\epsilon}[\log P(t, T + \epsilon) - \log P(t, T)]. \tag{2.2.5}$$

Apparently, it follows that

$$\begin{aligned} f(t,T) &= \lim_{\epsilon \to 0} \quad Y(t,T,T+\epsilon) \\ &= -\lim_{\epsilon \to 0} \left[\frac{\log P(t,T+\epsilon) - \log P(t,T)}{\epsilon} \right] \\ &= -\frac{\partial}{\partial T} \log P(t,T) \end{aligned} \tag{2.2.6}$$

Equation (2.2.6) implies further that

$$\begin{aligned} \log P(t,T) &= \int_t^T \frac{\partial}{\partial u} \log P(t,u) du \\ &= -\int_t^T f(t,u) du. \end{aligned}$$

Consequently,

$$P(t,T) = exp\left(-\int_t^T f(t,u)du\right).$$

In Section 2.1, we gave an argument on how to obtain the "rational" price of a contingent claim. Following that motivation, an alternative pricing model for the bond price in terms of the short rate is given as

$$P(t,T) = E\left[exp\left(-\int_t^T r_u du\right) \Big| \mathcal{F}_t \right].$$

The HJM model is specified by an SDE describing the dynamics of $f(t,T)$. In particular, for every $T \in (0, T^*]$, we consider

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW_t.$$

Both $\alpha(u,T)$ and $\sigma(u,T)$, for $0 \leq u \leq T$, are measurable in (u,ω) and adapted. Since, $P(t,T) = exp\left(-\int_t^T f(t,u)du\right)$, let us obtain the dynamics of $-\int_t^T f(t,u)du$.

$$\begin{aligned} d_t\left[-\int_t^T f(t,u)du\right] &= f(t,t)dt - \int_t^T (df(t,u))du \\ &= r(t)dt - \int_t^T [\alpha(t,u)dt + \sigma(t,u)dW_t]du. \end{aligned}$$

Write

$$\alpha^* = \int_t^T \alpha(t,u)du \quad \text{and}$$

$$\sigma^* = \int_t^T \sigma(t,u)du.$$

Define

$$X_t := -\int_t^T f(t,u)du.$$

Since $f(t,u)$ is an $\mathcal{F}_t$—adapted process, then X_t as well. We note that $dX_t = [r(t) - \alpha^*(t,T)]dt - \sigma^* dW_t$ and therefore X_t is an Itô process.

Furthermore, $P(t,T) = e^{X_t}$ so that

$$\begin{aligned} dP(t,T) &= e^{X_t}\big[r(t) - \alpha^*(t,T) + \frac{1}{2}\sigma^*(t,T)^2\big]dt \\ &\quad -e^{X_t}\sigma^*(t,T)dW_t. \\ &= P(t,T)\Big[\big(r(t) - \alpha^*(t,T) + \frac{1}{2}\sigma^*(t,T)^2\big)dt \\ &\quad -\sigma^*(t,T)dW_t\Big]. \end{aligned}$$

The discounted bond price $P(t,T)$ is a martingale under a risk neutral measure P, if for $0 \leq t \leq T \leq T^*$

$$\alpha^*(t,T) = \frac{1}{2}\big(\sigma^*(t,T)\big)^2.$$

Thus,

$$\int_t^T \alpha(t,u)du = \frac{1}{2}\left(\int_t^T \sigma(t,u)du\right)^2.$$

This implies that

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du.$$

If P itself is not a risk-neutral measure there may be a probability P^θ under which $\frac{P(t,T)}{S_t^0}$ is a martingale.

This is the result contained in the following theorem.

Theorem 2.2.1 (Heath, Jarrow and Morton) *For each $T \in (0, T^*]$ suppose $\alpha(u,T)$ and $\sigma(u,T) > 0$ for all u, T, and $f(0,T)$ is a deterministic function* *of T. The instantaneous forward rate $f(t,T)$ is defined by*

$$f(t,T) = f(0,T) + \int_0^t \alpha(u,T)du + \int_0^t \sigma(u,T)dW_u.$$

Then the term structure model determined by the processes $f(t,T)$ does not allow arbitrage if and only if there is an adapted process $\theta(t)$ such that

$$\alpha(t,T) = \sigma(t,T)\int_t^T \sigma(t,u)du + \sigma(t,T)\theta(t)$$

for all $0 \le t \le T^*$, and the process

$$\Lambda^\theta(t) := exp\left\{-\int_0^t \theta(u)dW_u - \frac{1}{2}\int_0^t \theta(u)^2du\right\}.$$

is an $(\mathcal{F}_t, P)$ martingale.

Proof: Consider an adapted process θ where $\Lambda^\theta(t)$ is an $(\mathcal{F}_t, P)$ martingale.

Let P^θ be a new probability measure such that

$$\left.\frac{dP^\theta}{dP}\right|_{\mathcal{F}_{T^*}} = \Lambda^\theta(T^*).$$

Now,

$$W_t^\theta = \int_0^t \theta(u)du + W_t$$

by Girsanov theorem and

$$\begin{aligned} dP(t,T) &= P(t,T)\Big[\big(r(t) - \alpha^*(t,T) \\ &\quad + \frac{1}{2}\sigma^*(t,T)^2 + \sigma^*(t,T)\theta(t)\big)dt \\ &\quad - \sigma^*(t,T)dW_t^\theta\Big]. \end{aligned}$$

For $P(t,T)$ to have rate of return $r(t)$ under P^θ, θ must satisfy

$$\alpha^*(t,T) = \frac{1}{2}\sigma^*(t,T)^2 + \sigma^*(t,T)\theta(t).$$

This must hold for all maturities T. Differentiating with respect to T, we obtain

$$\alpha(t,T) = \sigma(t,T)\sigma^*(t,T) + \sigma(t,T)\theta(t)$$

for $0 \leq t \leq T \leq T^*$.

■

We make the following observations in connection with the above theorem.

Remarks:

1. The process $\theta(t)$, if it exists, is independent of time T^*, the maturity of the bond $P(t,T)$ and

$$\theta(t) = -\left[\frac{-\alpha^*(t,T) + \frac{1}{2}\sigma^*(t,T)^2}{\sigma^*(t,T)}\right].$$

2. Under P, a "market" probability, the rate of return of the bond is $r(t) - \alpha^*(t,T) + \frac{1}{2}\sigma^*(t,T)^2$. Hence, the rate of return above the interest rate $r(t)$ is $-\alpha^*(t,T) + \frac{1}{2}\sigma^*(t,T)^2$ and the market price of risk is

$$\frac{-\alpha^*(t,T) + \frac{1}{2}\sigma^*(t,T)^2}{\sigma^*(t,T)} = -\theta(t).$$

It is very important to keep in mind that the market price of risk described above is defined in terms of the volatility and drift structures of the forward rate.

With regard to these Remarks, Theorem (2.2.1) requires that the market price of risk is independent of the maturity times T.

Under P^θ, we have

$$\begin{aligned} dP(t,T) &= P(t,T)\left[r(t)dt - \sigma^*(t,T)dW_t^\theta\right] \\ \text{and} \quad df(t,T) &= \sigma(t,T)\sigma^*(t,T)dt + \sigma(t,T)dW_t^\theta. \end{aligned} \qquad (2.2.7)$$

2.3 Single-Factor HJM Models

The Heath-Jarrow-Morton approach to term structure modelling is a powerful, technically rigorous interest rate model based on an exogenous specification of the dynamics of instantaneous continuously compounded forward rates $f(t,T)$.

For every fixed $T \leq T + \epsilon \leq T^*$, and given an initial forward rate curve $f(0,T)$, the dynamics of the instantaneous rate $f(t,T)$ are given by the integrated version

$$f(t,T) = f(0,T) + \int_0^t \alpha(u,T)du + \int_0^t \sigma(u,T)dW_u, \tag{2.3.8}$$

for $0 \leq t \leq T$.

The volatilities $\sigma(t,T)$ and the drifts $\alpha(t,T)$ can depend on the filtration generated by the Brownian motion W_t and the rates themselves up to time t. This means that for any fixed maturity T, the forward rate evolves according to its volatility $\sigma(t,T)$ and its own drift $\alpha(t,T)$.

We wish to formalise the descriptions of the properties of these volatility and drift functions. Following [5], we give technical conditions and constraints for α and σ for a general single-factor HJM Model. Further, we specify market completeness conditions which are the requirements validating the use of the HJM pricing paradigm.

On any Single-Factor HJM Model, we assume the following:

Conditions on the Volatility and Drift

1. The processes $\sigma(t,T)$ and $\alpha(t,T)$ depend only on the history of the Brownian motion up to time t. They are good integrands in the sense that $\int_0^T \sigma^2(t,T)dt < \infty$ and $\int_0^T |\alpha(t,T)|dt < \infty$.
2. The initial forward curve $f(0,T)$, is deterministic and satisfies the condition that $\int_0^T |f(0,u)|du < \infty$.

3. The drift α has finite integral given by $\int_0^T \left(\int_0^u |\alpha(t,u)|dt\right)du$.

4. The volatility σ satisfies the condition that

$$E\left[\int_0^T \left|\int_0^u \sigma(t,u)dW_t\right| du\right] < \infty.$$

The first two technical conditions see to it that the forward rates are well defined by their SDEs. While, the last two conditions are requirements for a Fubini-type result that the stochastic differential of the integral of $f(t,T)$ with respect to T is the integral of the stochastic differentials of f.

Market Completeness Conditions

1. It is required that there exists an $\mathcal{F}_t$−adapted process θ_t such that $\alpha(t,T) = \sigma(t,T)(\theta_t + \sigma^*(t,T))$, for all $t \leq T$ where $\sigma^*(t,T)$ is just the notation for $\int_t^T \sigma(t,u)du$ and the process $\Lambda^\theta(t) = exp\{-\int_0^t \theta(u)dW_u - \frac{1}{2}\int_0^t \theta(u)^2 du\}$ is a martingale.

2. The process $\sigma^*(t,T)$ is non-zero for almost all (t,ω), $t < T$ for every maturity T.

3. The expectation $E\left[exp\left(\frac{1}{2}\int_0^T (\theta_t + \sigma^*(t,T))^2 dt\right)\right]$ is finite.

4. The expectation $E\left[exp\left(\frac{1}{2}\int_0^T {\theta_t}^2 dt\right)\right]$ is finite.

The first condition ensures the absence of arbitrage. That is, it makes sure the existence of an equivalent martingale measure so that every single discounted bond price is a martingale.

On the other hand, the second condition states that the change of measure is unique. Or, plainly every risk can be hedged through martingale representation theorem.

And the last two conditions are assumptions needed before the Girsanov's theorem can be applied and technical requirements for a discounted price process to be a martingale under the new measure.

2.4 Multi-Factor HJM Model

We have been accumulating technical conditions as we have swept through. These summary of technical conditions in this section are intended for a discussion when the process is driven by n independent Brownian motions. We shall consider particular cases of this in chapter 8.

Single-factor model has the disadvantage that all the increments in the bond prices are perfectly correlated. The assumption of a single factor is also too simplified which may not be that realistic, especially when we are pricing a contingent claim which depends on the difference of two points on the yield curve.

Thus, it is worth considering multi-factor models. In an $n-$factor model, we shall be working with n independent Brownian motions. For each $T-$bond forward rate process, the volatility $\sigma_i(t,T)$ has a corresponding Brownian factor W_t^i. This formulation allows different bonds to depend on external "shocks" in different ways, and to have strong correlations with some bonds and weaker correlations with others. In its general form, the multi-factor HJM Model is given by

$$f(t,T) = f(0,T) + \sum_{i=1}^{n} \int_0^t \sigma_i(s,T)dW_s^i + \int_0^t \alpha(s,T)ds \tag{2.4.9}$$

for $0 \leq t \leq T$.

The generalised form for $f(t,T)$ in (2.4.9) merely tells us that the forward process starts with initial value $f(0,T)$ and is driven by various Brownian motion terms and a drift.

We see that the total instantaneous variance of $f(t,T)$ is $\sum_{i=1}^{n} {\sigma_i}^2(t,T)$. On the other hand, the covariance structure of the increments of the two forward rates $f(t,T)$ and $f(t,S)$ is given by

$$\sum_{i=1}^{n} \sigma_i(t,T)\sigma_i(t,S).$$

In particular, when $n = 1$, that is, if we have a single-factor model, the correlation of the changes in the forward rates of T—bond and S—bond is just one.

We can then give the expression for the instantaneous rate $r_t = f(t,t)$ analogous to (3.3.7). For this multi-factor model, we have

$$r_t = f(0,t) + \sum_{i=1}^{n} \int_0^t \sigma_i(s,t) dW_s^i + \int_0^t \alpha(s,t) ds.$$

Similar to what we outline in the summary of technical conditions for single-factor models in Section 2.3, we formalise volatility and drift conditions and also specify the required market completeness conditions.

On any Multi-Factor HJM Model, we assume the following:

Conditions on the Volatilities and Drifts

1. For each T, the process $\sigma_i(t,T)$ and $\alpha(t,T)$ are $\mathcal{F}_t$-adapted and their integrals $\int_0^T \sigma_i{}^2(t,T)dt$ and $\int_0^T |\alpha(t,T)|dt$ are finite.
2. The initial forward curve, $f(0,T)$, is deterministic and satisfies the condition that $\int_0^T |f(0,u)|du < \infty$.
3. The drift α satisfies $\int_0^T \left(\int_0^u |\alpha(t,u)|dt\right)du < \infty$.
4. Each volatility σ_i has finite expectation $E\left[\int_0^T | \int_0^u \sigma_i(t,u)dW_t^i|du\right]$.

Again here, the first two conditions ensure that the SDE for $f(t,T)$ is well-defined and the last two conditions are requirements of Fubini theorem for stochastic integrals.

Market Completeness Conditions

1. It is required that there exist $\mathcal{F}_t$—adapted processes $\theta_i(t)$, for $1 \leq i \leq n$, such that

$$\alpha(t,T) = \sum_{i=1}^{n} \sigma_i(t,T)\left(\theta_i(t) + \sigma_i^*(t,T)\right),$$

for all $t \leq T$.

2. $E\left[exp\left(\frac{1}{2}\sum_{i=1}^{n}\int_0^T \theta_i^2(t)dt\right)\right] < \infty$.

3. The matrix $\Sigma_t^* = \left(\sigma_i^*(t,T_j)\right)_{i,j=1}^{n}$ is non-singular for almost all (t,ω), $t < T_1$, for every set of maturities $T_1 < T_2 < ... < T_n$, and

4. $E\left[exp\left(\frac{1}{2}\sum_{i=1}^{n}\int_0^T (\theta_i(t) + \sigma_i^*(t,T))dt\right)^2\right] < \infty$.

The first two conditions are requirements we need to apply Girsanov's theorem for higher dimensions thus ensuring that the discounted bond prices are martingales. Unlike, the single-factor case, the drift is now allowed n "dimensions of freedom" away from its risk-neutral value. In other words, as a function of T, $\alpha(t,.)$ is allowed to deviate by any linear combination of the functions $\sigma_i(t,.)$. The second condition validates $\theta_i(t)$ to be a drift under an equivalent change of measure via the Girsanov theorem.

The nonzero volatility process $\sigma^*(t)$ in the single-factor model is replaced by a volatility matrix process Σ_t^* which has to be non-singular.

The last condition makes sure that the resulting driftless discounted bond price is a martingale (i.e., a multi-dimensional exponential martingale).

We can say therefore that validating the price obtained via the HJM framework amounts to checking the technical conditions described herein for the forward rate process.

2.5 Equivalence of Short Rate and HJM Models

Short rate models are based on the rate of instantaneous borrowing r_t and they are the ones commonly used in the market, in pricing derivative products which depend only on one underlying bond. As cited in [5], the short rates evolved from various historical starting points. Some emerged from discrete framework while others from equilibrium models and often are represented in a simple hierarchy with no apparent connection to any overarching model.

All of these are however HJM models and this is the reason why the HJM framework is considered in great details in this chapter. These two descriptions in valuing a contingent claim are equivalent via a mathematical transformation.

From equation (2.1.2), we have

$$\int_t^T f(t,u)du = -\log P(t,T) = h(r_t, t, T),$$

where $h(r_t, t, T)$ is the deterministic function

$$h(x,t,T) = -\log E\left[exp\left(-\int_t^T r_u du\right)\bigg| r_t = x\right]. \qquad (2.5.10)$$

As we are able to express the forward rate in terms of a function h which is also a function of r_t, equation (2.5.10) therefore specifies the HJM model in terms of the short rate.

On the other hand, the short rate can also be expressed in terms of the HJM model and this is the content of Theorem 6.2.1.

2.6 Versions of Expectations Hypothesis of the Yield Curve

For some time, the determination of the relationship between the yields and market expectations of future interest rates is one of the most popular and important research pursuit in the theory of interest. Such a relationship between yield rates and prices of the bonds (which contain information concerning term structure) is embodied in a unified framework called the *expectation hypothesis.*

We present four versions of the theory to reinforce us with more insights in our attempt to give a formal proof of the Expectations theory of the yield curve in Chapter 3. Each of these hypotheses is stated in its "pure" form, that is, with respect to the "true" stochastic process governing future short rates, or with respect to the physical measure, as opposed to the risk-adjusted process.

1. **The Unbiased Expectations Hypothesis (UEH)** This assumes that the forward rates $f(t,T)$, and the expected future short rates $E[r_T]$, are equal. That is,

$$f(t,T) = -\frac{\partial P(t,T)/\partial T}{P(t,T)} = E_{P^T}[r_T|\mathcal{F}_T]$$

for some probability measure P^T. The proof of this hypothesis via construction of the probability measure P^T is the goal of the next Chapter.

2. **The Local Expectation Hypothesis (LEH)** Under this hypothesis, the expected instantaneous return on any $(T-t)$–maturity bond is equal to the current short rate, r_t :

$$E_Q\left[\left.\frac{dP(t,T)/dt}{P(t,T)}\right|_{\mathcal{F}_t}\right] = r_t$$

for some probability measure Q.

3. **Returns to Maturity Expectations Hypothesis (RTM)** This theory supports the equality of expected returns, inclusive of capital invested from two alternative strategies (1) holding a discounting bond until maturity, or (2) rolling over a series of single-period bonds. The RTM Hypothesis mathematically states that

$$P(t,T)^{-1} = E_{P'}\left[\left(exp\int_t^T r(u)du\right)\middle|\,\mathcal{F}_t\right]$$

for some probability measure P'.

4. **Yields to Maturity Hypothesis (YTM)** This theory states that the yield from holding a bond equals the yield from rolling over a series of single period bonds. This means that,

$$-\frac{1}{T-t}\ln P(t,T) = E_{P'}\left[\frac{1}{T-t}\int_t^T r_s ds\middle|\,\mathcal{F}_t\right].$$

Cox, Ingersoll and Ross, [23], commented that since only one set of bond prices is observed, only one of the bond-pricing relationships described above can hold for a particular market. In other words, they concluded that the other three versions of the expectation hypothesis cannot be simultaneously responsible for generating bond price data from the set of expected future short rates.

Chapter 3

Forward Measures in the Interest Rate Market

3.1 The Valuation Problem

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and $\{\mathcal{F}_t\}$ be a standard filtration. Unless, otherwise stated, we assume that P is the risk-neutral probability measure in all the succeeding chapters. Suppose the short-term interest rate r_u can be given by $r(X_u)$, where $\{X_u, 0 \leq u \leq T\}$ is an $\mathbb{R}^n$-process defined on our complete probability space. Then at time $t \leq T$, the price of any contingent claim $\phi \in L^2(\Omega, \mathcal{F}_T, P)$ is given by

$$E\left[exp\left(-\int_t^T r_u du\right)\phi\middle|\mathcal{F}_t\right]. \tag{3.1.1}$$

The short rate r itself for example, can also be given by certain stochastic dynamics as in the Vasicek or Cox-Ingersoll-Ross (CIR) models.

The valuation of any contingent claim ϕ is central to Mathematical Finance and thus the evaluation of (3.1.1) is a fundamental problem. Direct calculation of the expectation in (3.1.1) seems to be difficult, especially when ϕ has a complex form. However, we first observe that, under certain technical conditions, we can use the forward measure so that the expectation in (3.1.1) is expressed as the product of two expectations. Having re-interpreted the above expectation, we discuss what happens when r follows certain stochastic dynamics.

3.2 The Forward Measure

In this section, we shall demonstrate how to express the valuation formula in (3.1.1) into product of two expectations. To go about this, the forward measure will be used as a tool. Furthermore, the forward measure approach enables us to establish with mathematical rigour the relationship between the forward rate and the spot rate of the interest rate process.

Following the discussion in [41], we introduce the concept of the forward measure P^T defined on $\mathcal{F}_T$ by setting

$$\left.\frac{dP^T}{dP}\right|_{\mathcal{F}_T} = \Lambda_{0,T} = \frac{exp(-\int_0^T r(X_{0,u}(x_0))du)}{P(0,T)}.$$

Here $X_{t,u}(x)$ denotes a stochastic process $X_{t,u}$ starting from $x \in \mathbb{R}^n$ at time t and r is the short rate. $P(t,T)$, the price of a zero coupon bond at time t, with maturity T is given by:

$$P(t,T) = E\left[exp\left(-\int_t^T r(X_{t,u}(x))du\right)\Big|\,\mathcal{F}_t\right].$$

Returning to the problem in (3.1.1), if ϕ is $\mathcal{F}_T-$ measurable and E^T denotes the expectation with respect to P^T, then from Bayes' Rule,

$$E_t^T[\phi] = E^T[\phi|\mathcal{F}_t] = \frac{E[\Lambda_{0,T}\phi|\mathcal{F}_t]}{E[\Lambda_{0,T}|\mathcal{F}_t]}.$$

Or,

$$\begin{aligned} E_t^T[\phi] &= \frac{E[(exp - \int_0^t r(X_{0,u}(x))du)(exp - \int_t^T r(X_{t,u}(x))du)\phi|\mathcal{F}_t]}{E[(exp - \int_0^t r(X_{0,u}(x))du)(exp - \int_t^T r(X_{t,u}(x))du)|\mathcal{F}_t]} \\ &= \frac{E[(exp - \int_t^T r(X_{t,u}(x))du)\phi|\mathcal{F}_t]}{P(t,T)}. \end{aligned}$$

Consequently,

$$E\left[\left(exp - \int_t^T r(X_{t,u}(x))du\right)\phi\middle|\,\mathcal{F}_t\right] = P(t,T)E^T[\phi|\mathcal{F}_t]. \tag{3.2.2}$$

It is the object of this thesis to relate the bond price, $P(t,T)$, in the short rate models given by

$$P(t,T) = E\left[exp\left(-\int_t^T r(u)du\right)\middle|\,\mathcal{F}_t\right] \tag{3.2.3}$$

and the bond price in the Heath-Jarrow-Morton (HJM) Model given by

$$P(t,T) = exp\left(-\int_t^T f(t,u)du\right), \tag{3.2.4}$$

where $f(t,T)$ is the forward rate at date $t \leq T$ for instantaneous risk-free borrowing or lending at date T.

In relating $P(t,T)$ in the short rate models to the HJM model, we shall look at different models for the short rate r. These will include the cases when r is a Markov process as in [37] and also when r_t follows the dynamics of the Vasicek and CIR models.

Then, we shall investigate the structure of the forward rate and short rate process when the price $P(t,T)$, of a zero-coupon bond is given by a generalised exponential affine model. This model is of particular interest because its form resembles to that of the HJM pricing framework as in (3.2.4).

3.3 A Proof of the Expectation Hypothesis

In terms of the short rate the bond price is given by (3.2.3).

On the other hand, the HJM model for term structure considers stochastic differential equations for the evolution of forward rates $f(t,T)$. For each $T \in (0,T^*]$, assuming that $0 \leq t \leq T < T+\epsilon \leq T^*$, suppose the dynamics of f are given by

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW_t. \tag{3.3.5}$$

The coefficients $\alpha(u,T)$ and $\sigma(u,T)$ for $0 \le u \le T$ are measurable in (u,ω) and adapted. The integral form of (3.3.5) is

$$f(t,T) = f(0,T) + \int_0^t \alpha(u,T)du + \int_0^t \sigma(u,T)dW_u. \tag{3.3.6}$$

Following the descriptions in [5], we can write down an integral equation for the instantaneous rate $r_t = f(t,t)$, namely:

$$r_t = f(t,t) = f(0,t) + \int_0^t \alpha(u,t)du + \int_0^t \sigma(u,t)dW_u. \tag{3.3.7}$$

Equation (3.3.7) is a classical result for r_t expressed as a stochastic integral of forward rates under the risk-neutral probability measure. We would like to establish the relationship between the forward rate and the short-term rate under the forward measure P^T. This relationship is stated as follows:

Lemma 3.3.1 *In terms of the short rate model, the forward rate is given by*

$$f(t,T) = E^T[r_{t,T}|\mathfrak{F}_t] \tag{3.3.8}$$

where E^T denotes the expectation under P^T.

Proof: From (3.2.3), we have the bond price in terms of the short rate:

$$P(t,T) = E\left[exp\left(-\int_t^T r(u)du\right)\middle|\mathfrak{F}_t\right].$$

Differentiating $P(t,T)$ with respect to T, we get

$$\begin{aligned}\frac{\partial P(t,T)}{\partial T} &= E\left[-r_{t,T}\,exp\left(-\int_t^T r_{t,u}du\right)\middle|\mathfrak{F}_t\right]\\ &= E^T[-r_{t,T}|\mathfrak{F}_t]E\left[exp\left(-\int_t^T r_{t,u}du\right)\middle|\mathfrak{F}_t\right]\\ &= -E^T[r_{t,T}|\mathfrak{F}_t]P(t,T).\end{aligned} \tag{3.3.9}$$

Also, from (3.2.4) the bond price in terms of the forward rate is given by

$$P(t,T) = exp\left(-\int_t^T f(t,u)du\right).$$

Differentiating $P(t,T)$ with respect to T, we obtain

$$\frac{\partial P(t,T)}{\partial T} = -P(t,T)f(t,T). \qquad (3.3.10)$$

Comparing (3.3.9) and (3.3.10), we see that

$$f(t,T) = E^T[r_{t,T}|\mathcal{F}_t].$$

■

The result of the above lemma tells us that

$$P(t,T) = exp\left(-\int_t^T E^u[r_{t,u}|\mathcal{F}_t]du\right). \qquad (3.3.11)$$

We shall investigate (3.3.11) with the aim of obtaining explicit solutions when r follows certain stochastic processes.

It is also worth noting that the result of the above lemma has laid down a mathematical foundation supporting the Expectation Hypothesis Theory. In its purest form, the Expectations theory of the yield curve states that the implied forward interest rate profile represents the market's expectations of the future short-term rates. Economics theory such as this is a reasonable argument however, it is loosely stated.

The key point to remember here is that the observed market forward rate curve provides the best forecast of future spot interest rates only if we are working on an expectation that is evaluated with respect to the forward measure.

Chapter 4

The Short Rate as a Function of a Continuous Time Markov Chain

4.1 Notation and Convention

Following [37], we shall assume that the short-term rate r is a function of a continuous time Markov chain. As noted in [37], this assumption is reasonable as any diffusion can be approximated by a Markov chain.

In this model, let X_t, $t \geq 0$, be a finite state Markov chain with state space $S = \{s_1, s_2, ..., s_n\}$. The points s_i can be points in $\mathbb{R}^n$ or any space whatsoever and can model factors of the economy. Without loss of generality, we may identify points in S with unit vectors $\{e_1, e_2, ..., e_n\}$ to simplify the algebra. In this representation of the state space of X, $e_i = (0, ..., 0, 1, 0, ...0)' \in \mathbb{R}^n$.

At any time t, we note that the state X_t of the Markov chain is one of the unit vectors, $e_1, e_2, ..., e_n$. Hence, for any real valued function of X_t, say $F(X_t)$, is just given by $F = (F_1, F_2, ..., F_n)$. And so $F(X_t) = \langle F, X_t \rangle$ where the brackets denote the scalar product in $\mathbb{R}^n$. We hypothesize that the short rate process r_t is a function of X_t. In other words, $r_t = r(X_t) = \langle r, X_t \rangle$ for some vector $r \in \mathbb{R}^n$. We could take r to be time varying but we would like to begin with the simplest case of the model.

It follows from the result of Lemma 3.3.1 that

$$f(t,T) = E^T[r(X_T)|\mathcal{F}_t] = \langle r, E^T[X_T|\mathcal{F}_t]\rangle. \tag{4.1.1}$$

Thus, the price at time $t \leq T$ for a zero-coupon bond is given by

$$\begin{aligned} P(t,T) &= exp\left(-\int_t^T \langle r, E^u[X_u|\mathcal{F}_t]\rangle du\right) \\ &= exp\left(-\int_t^T \langle r, E^u[X_u|X_t = x]\rangle du\right). \end{aligned} \tag{4.1.2}$$

To evaluate (4.1.2), we need to find the dynamics of X under P^T.

4.2 Evaluating $E^T[X_T|\mathcal{F}_t]$

Our concern at present is the evaluation of $E^T[X_T|\mathcal{F}_t]$ in (4.1.2).

Now, from (3.2.2) we have

$$E^T[X_T|\mathcal{F}_t] = \frac{E[exp(-\int_t^T \langle r, X_v\rangle dv)X_T|\mathcal{F}_t]}{P(t,T)}. \tag{4.2.3}$$

To evaluate (4.2.3), we need to know the structure and form of the Markov process X_t. This is the content of the following theorem and its corollary.

But, first we note that the unconditional distribution of X_t is the vector $E[X_t] = p_t = ({p_t}^1, {p_t}^2, ..., {p_t}^n)$, where

$${p_t}^i = P(X_t = e_i) = E[\langle e_i, X_t\rangle] = P(r_t = r_i).$$

Suppose this distribution evolves according to the Kolmogorov equation $\frac{dp_t}{dt} = Ap_t$.

Here A is a "Q-matrix," that is, if $A = (a_{ij})$, $1 \leq i,j \leq n$, $\sum_{j=1}^n a_{ij} = 0$ and $a_{ji} \geq 0$ if $i \neq j$. The components a_{ji} could be taken to be time varying, though this would complicate their estimation.

Theorem 4.2.1 *Let M be an $\mathbb{R}^n$–valued process given by*

$$M_t = X_t - X_0 - \int_0^t AX_u du.$$

Then, M is an $(\mathcal{F}_t, P)$ martingale.

Proof: Consider the matrix exponential $e^{A(t-u)}$. Then, by the Markov property,

$$E[X_t|X_u] = e^{A(t-u)} X_u$$

for $t \geq u$. This is to say, that one solves the Kolmogorov equation with initial condition X_u. For $t \geq u$, we have

$$\begin{aligned} E[M_t - M_u|\mathcal{F}_u] &= E[X_t - X_u|\mathcal{F}_u] - E\left[\int_u^t AX_v dv|\mathcal{F}_u\right] \\ &= e^{A(t-u)}X_u - X_u - \int_u^t Ae^{A(v-u)}X_u dv \\ &= \left[e^{A(t-u)} - I - \int_u^t Ae^{A(v-u)}dv\right] X_u \end{aligned}$$

where I is the $n \times n$ identity matrix. Henceforth,

$$E[M_t - M_u|\mathcal{F}_u] = \left[e^{A(t-u)} - I - [e^{A(v-u)}]_u^t\right] X_u = 0.$$

■

Corollary 4.2.1 *X is a semimartingale with representation*

$$X_t = X_0 + \int_0^t AX_u du + M_t. \tag{4.2.4}$$

Proof: X_0 is $\mathcal{F}_0$-measurable. $\int_0^t AX_u du$ is a process of finite variation. While M_t is a martingale from Theorem 4.2.1. Thus, X_t is a semimartingale.

■

Hence under the risk-neutral probability measure P, the semimartingale form of the Markov chain is:

$$X_t = X_0 + \int_0^t AX_u du + M_t \tag{4.2.5}$$

where $\{M_t\}$ is a $(P, \mathcal{F}_t)$ martingale, and $\mathcal{F}_t = \sigma\{X_u : u \leq t\}$.

Write $X_{0,t} := X_0 + \int_0^t AX_v dv + M_t$.

Define

$$\begin{aligned} \Lambda_{0,t} : &= exp\left(-\int_0^t r\big(X_{0,v}(x_o)\big)dv\right). \\ \Lambda_{t,u} : &= exp\left(-\int_t^u r\big(X_{t,v}(x_t)\big)dv\right). \end{aligned}$$

Thus, $d\Lambda_{0,t} = -r\big(X_{0,t}(x_0)\big)\Lambda_{0,t}dt$.

Further, we get

$$\begin{aligned} d(\Lambda_{t,v}X_{t,v}) &= \Lambda_{t,v}dX_{t,v} + X_{t,v}d\Lambda_{t,v} \\ &= \Lambda_{t,v}\big[AX_{t,v}dt + dM_t\big] + X_{t,v}\big[-r\big(X_{t,v}(x_t)\big)\Lambda_{t,v}dt\big]. \end{aligned}$$

Or, in integral form,

$$\begin{aligned} \Lambda_{t,T}X_{t,T} &= X_t + \int_t^T \Lambda_{t,v}AX_{t,v}dv + \int_t^T \Lambda_{t,v}dM_v \\ &\quad - \int_t^T r(v)\Lambda_{t,v}X_{t,v}dv. \end{aligned}$$

This implies that

$$\begin{aligned} E_t[\Lambda_{t,T}X_{t,T}] &= X_t + \int_t^T AE_t[\Lambda_{t,v}X_{t,v}]dv \\ &\quad - \int_t^T E_t[\langle r, X_v\rangle\Lambda_{t,v}X_{t,v}]dv. \end{aligned} \tag{4.2.6}$$

Equation (4.2.6) is the explicit expression for the numerator of (4.2.3). We note further that

$$\begin{aligned} \langle r, X_v\rangle \Lambda_{t,v} X_{t,v} &= \sum_{i=1}^{n} \langle X_{t,v}, e_i\rangle r_i e_i \\ &= R\Lambda_{t,v} X_{t,v} \end{aligned} \tag{4.2.7}$$

where R is the matrix with $r = (r_1, r_2, ..., r_n)'$ on the diagonal. In (4.2.6), we observe that we need to evaluate $E_t[\Lambda_{t,v} X_{t,v}]$.

Writing $\hat{z}_{t,T} := E_t[\Lambda_{t,v} X_{t,v}]$ we see from (4.2.6) that

$$\begin{aligned} \hat{z}_{t,T} : &= E_t[\Lambda_{t,v} X_{t,v}] \\ &= X_t + \int_t^T A\hat{z}_{t,v} dv - \int_t^T R\hat{z}_{t,v} dv. \end{aligned}$$

That is,

$$\hat{z}_{t,T} = X_t + \int_t^T (A - R)\hat{z}_{t,v} dv.$$

Therefore, the expression for the numerator in (4.2.3) simplifies to

$$\hat{z}_{t,T} = e^{(A-R)(T-t)} X_t = e^{B^*(T-t)} X_t, \tag{4.2.8}$$

where $B^* = (A - R)$.

Of course, $\hat{z}_{t,T}$ is a vector process. Next, we concern ourselves with an explicit expression for $P(t, T)$.

Write

$$\begin{aligned} P(t,T) := P(t,T,X_t) &= E\left[exp\left(- \int_t^T \langle r, X_u\rangle du \right) \middle| \mathcal{F}_t \right] \\ &= E\left[exp\left(- \int_t^T \langle r, X_u\rangle du \right) \middle| X_t \right] \end{aligned}$$

from the Markov property.

Now, $\langle X_T, \mathbf{1}\rangle = 1$ where $\mathbf{1} = (1, 1, .., 1)'$.

Therefore,

$$\begin{aligned} P(t,T,X_t) &= \left\langle E\left[exp\left(-\int_t^T \langle r, X_v\rangle dv\right) X_T \middle| \mathcal{F}_t\right], \mathbf{1}\right\rangle \\ &= E\left[exp\left(-\int_t^T \langle r, X_v\rangle dv\right) \langle X_T, \mathbf{1}\rangle \middle| \mathcal{F}_t\right] \\ &= E\left[exp\left(-\int_t^T \langle r, X_v\rangle dv\right) \middle| \mathcal{F}_t\right] \\ &= \langle e^{(A-R)(T-t)} X_t, \mathbf{1}\rangle \\ &= \langle X_t, e^{B(T-t)}\mathbf{1}\rangle \end{aligned} \tag{4.2.9}$$

where $B = (A-R)^*$, R =diag r, $r = (r_1, r_2, ..., r_n)'$.

With reference to (4.2.3), equations (4.2.8) and (4.2.9) are explicit expressions for the numerator and denominator, respectively. That is,

$$E^T[X_t|\mathcal{F}_t] = \frac{e^{(A-R)(T-t)}X_t}{\langle X_t, e^{B(T-t)}\mathbf{1}\rangle}. \tag{4.2.10}$$

Hence, invoking Lemma 3.3.1 together with (4.2.10) and (4.1.1) we have

$$\begin{aligned} f(t,T) &= \langle r, E^T[X_T|\mathcal{F}_t]\rangle \\ &= \left\langle r, \frac{e^{(A-R)(T-t)}X_t}{\langle X_t, e^{B(T-t)}\mathbf{1}\rangle}\right\rangle \end{aligned} \tag{4.2.11}$$

We can study the dynamics of $f(t,T)$ using expression (4.2.11). However, as we are having a quotient expression representing a vector in our scalar product, onerous efforts are entailed to do this direct computation. So, we shall not proceed in this way. Rather, we make use of the result we have in (2.1.2).

4.3 The Dynamics of $f(t,T)$

Restating (2.1.2), we have

$$f(t,T) = -\frac{\partial}{\partial T}\log P(t,T).$$

From (4.2.9) we obtain the zero-coupon bond price $P(t,T)$ given by $P(t,T) = P(t,T,X_t) = \langle X_t, e^{B(T-t)}\mathbf{1}\rangle$ and $X_t \in \{e_1, e_2, ..., e_n\}$.

Therefore, we get

$$\log P(t,T,X_t) = \langle X_t, \lambda_t\rangle \tag{4.3.12}$$

where $\lambda_t = ({\lambda_t}^1, {\lambda_t}^2, ..., {\lambda_t}^N)$ and ${\lambda_t}^i = \log\langle e_i, e^{B(T-t)}\mathbf{1}\rangle$.

Consequently, using (4.3.12)

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}\langle X_t, \lambda_t\rangle \\ &= -\langle X_t, \gamma_t\rangle \end{aligned} \tag{4.3.13}$$

where $\gamma_t = ({\gamma_t}^1, {\gamma_t}^2, ..., {\gamma_t}^n)$ and

$${\gamma_t}^i = \frac{\partial}{\partial T}{\lambda_t}^i = \frac{\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}. \tag{4.3.14}$$

Alternatively,

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}\langle X_t, \lambda_t\rangle = -\frac{\partial}{\partial T}\sum_{i=1}^{n}{\lambda_t}^i\langle X_t, e_i\rangle \\ &= -\sum_{i=1}^{n}\frac{\partial}{\partial T}{\lambda_t}^i\langle X_t, e_i\rangle \\ &= -\sum_{i=1}^{n}\frac{\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\langle X_t, e_i\rangle. \end{aligned} \tag{4.3.15}$$

Thus,

$$df(t,T) = \sum_{i=1}^{n}\frac{\partial}{\partial t}\left[\frac{-\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\langle X_t, e_i\rangle\right]. \tag{4.3.16}$$

We shall consider first $\frac{\partial}{\partial t}\left[\frac{-\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\langle X_t, e_i\rangle\right]$.

$$
\begin{aligned}
\frac{\partial}{\partial t}\left[\frac{-\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\langle X_t, e_i\rangle\right] &= \frac{\partial}{\partial t}\left[\langle e_i, -Be^{B(T-t)}\mathbf{1}\rangle\langle X_t, e_i\rangle\langle e_i, e^{B(T-t)}\mathbf{1}\rangle^{-1}\right] \\
&= \langle e_i, B^2e^{B(T-t)}\mathbf{1}dt\rangle\frac{\langle X_t, e_i\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle} \\
&\quad +\frac{\langle e_i, -Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, -Be^{B(T-t)}\mathbf{1}\rangle}\langle dX_t, e_i\rangle \\
&\quad +\langle e_i, -Be^{B(T-t)}\mathbf{1}\rangle\langle X_t, e_i\rangle(-1) \\
&\quad \langle e_i, e^{B(T-t)}\mathbf{1}\rangle^{-2}\langle e_i, -Be^{B(T-t)}\mathbf{1}dt\rangle
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial}{\partial t}\left[\frac{-\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\langle X_t, e_i\rangle\right] &= \frac{\langle X_t, e_i\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\Bigg(\langle e_i, B^2e^{B(T-t)}\mathbf{1}\rangle \\
&\quad -\frac{\langle e_i, -Be^{B(T-t)}\mathbf{1}\rangle^2}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\Bigg)\, dt \\
&\quad -\frac{\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\langle AX_t dt + dM_t, e_i\rangle \\
&= \frac{\langle X_t, e_i\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\Bigg(\langle e_i, B^2e^{B(T-t)}\mathbf{1}\rangle \\
&\quad -\frac{\langle e_i, -Be^{B(T-t)}\mathbf{1}\rangle^2}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\Bigg)\, dt \\
&\quad -{\gamma_t}^i\langle AX_t, e_i\rangle dt - {\gamma_t}^i\langle dM_t, e_i\rangle.
\end{aligned}
$$

Thus, from (4.3.16), we obtain the dynamics of $f(t,T)$ which is

$$\begin{aligned} df(t,T) &= \sum_{i=1}^{n}\Bigg[\frac{\langle X_t, e_i\rangle}{\langle e_i, e^{B(T-t)}\mathbf{1}\rangle}\Big(\langle e_i, B^2 e^{B(T-t)}\mathbf{1}\rangle \\ &\quad -{\gamma_t}^i\langle e_i, Be^{B(T-t)}\mathbf{1}\rangle - {\gamma_t}^i\langle AX_t, e_i\rangle\Big)\Bigg] \\ &\quad +\sum_{i=1}^{n}{\gamma_t}^i\langle dM_t, e_i\rangle. \end{aligned} \tag{4.3.17}$$

And the initial forward curve is given by

$$\begin{aligned} f(0,T) &= -\frac{\partial}{\partial T}\langle X_0, \gamma_0\rangle = -\langle X_0, \lambda_0\rangle \\ &= -\sum_{i=1}^{n}\frac{\langle e_i, Be^{BT}\mathbf{1}\rangle}{\langle e_i, e^{BT}\mathbf{1}\rangle}\langle X_0, e_i\rangle \end{aligned} \tag{4.3.18}$$

Or in integral form, the forward rate $f(t,T)$ associated with a Markov short rate is

$$\begin{aligned} f(t,T) &= -\sum_{i=1}^{n}\frac{\langle e_i, Be^{BT}\mathbf{1}\rangle}{\langle e_i, e^{BT}\mathbf{1}\rangle}\langle X_0, e_i\rangle \\ &\quad +\int_0^t\sum_{i=1}^{n}\Bigg[\frac{\langle X_u, e_i\rangle}{\langle e_i, e^{B(T-u)}\mathbf{1}\rangle}\Big(\langle e_i, B^2 e^{B(T-u)}\mathbf{1}\rangle \\ &\quad -{\gamma_u}^i\langle e_i, Be^{B(T-u)}\mathbf{1}\rangle - {\gamma_u}^i\langle AX_u, e_i\rangle\Big)\Bigg]dt \\ &\quad +\int_0^t\sum_{i=1}^{n}{\gamma_u}^i\langle dM_u, e_i\rangle. \end{aligned} \tag{4.3.19}$$

Equation (4.3.15) is an analytic expression for $f(t,T)$ while (4.3.19) gives the dynamics of $f(t,T)$ with drift and volatility terms. Further analysis of these drift and volatility functions will enable us to come up with a statement describing the implication of these models towards the financial market.

Finally, from (4.2.8), it is easy to see that $f(t,t) = r_t$. This is because

$$\begin{aligned} f(t,t) &= \left\langle r, \frac{e^0 X_t}{\langle X_t, 1\rangle} \right\rangle \\ &= \langle r, X_t \rangle \\ &= r_t. \end{aligned}$$

Chapter 5

The Short Rate Under the Vasicek's Model

5.1 Description of the Model

In this model, the short rate r under the risk-neutral measure P follows the stochastic dynamics given by

$$dr_t = a(b - r_t)dt + \sigma dW_t. \tag{5.1.1}$$

where a, b, and σ are strictly positive constants. The model proposed by Vasicek in (5.1.1) is a mean reverting version of the Ornstein-Uhlenbeck process. The SDE (5.1.1) has a Brownian part and a restoring drift which pushes it upwards when the process is below b and downwards when it is above. The magnitude of the drift is also proportional to the distance away from this mean.

The Vasicek model can alternatively be expressed as

$$dr_t = (a - br_t)dt + \sigma dW_t.$$

The mean reverting level in this model is $\frac{a}{b}$ and the rate of adjustment to this level is b. Occasionally, we shall use this alternative form to derive important results in Chapter 8.

Applying Itô's Lemma, we can check that the solution to the SDE in (5.1.1) starting r at r_0, is

$$r_t = e^{-at}\left(r_0 + b(e^{at} - 1) + \sigma \int_0^t e^{au} dW_u\right). \tag{5.1.2}$$

Given the dynamics of r_t described in (5.1.2), we shall find the equivalent stochastic dynamics of $f(t, T)$.

5.2 The Dynamics of $f(t, T)$

Vasicek solved equation (3.2.3) and obtain analytic expression for $P(t, T)$ when r_t follows a mean reverting Ornstein-Uhlenbeck process. The analytic formula for $P(t, T)$ is given by

$$P(t, T) = A^*(t, T) exp[-B(t, T) r_t], \tag{5.2.3}$$

provided, when $a \neq 0$,

$$B(t, T) = \frac{1 - e^{-a(T-t)}}{a} \tag{5.2.4}$$

and

$$A^*(t, T) = exp\left[\frac{(B(t, T) - T + t)(a^2 b - \frac{\sigma^2}{2})}{a^2} - \frac{\sigma^2 B(t, T)^2}{4a}\right]. \tag{5.2.5}$$

Alternatively, the price $P(t, T)$ in (5.2.3) can be expressed as

$$P(t, T) = exp[A(t, T) - B(t, T) r_t] \tag{5.2.6}$$

where

$$\begin{aligned} A(t, T) &= \log A^*(t, T) \\ &= \frac{(B(t, T) - T + t)(a^2 b - \frac{\sigma^2}{2})}{a^2} - \frac{\sigma^2 B(t, T)^2}{4a} \end{aligned} \tag{5.2.7}$$

and the function $B(t, T)$ is the same as given in (5.2.4).

Now from (2.1.2),

$$f(t,T) = -\frac{\partial}{\partial T}\log P(t,T).$$

Thus,

$$\begin{aligned}
f(t,T) &= -\frac{\partial}{\partial T}\log exp[A(t,T) - B(t,T)r_t] \\
&= -\frac{\partial}{\partial T}[A(t,T) - B(t,T)r_t] \\
&= \frac{\partial}{\partial T}[B(t,T)r_t] - \frac{\partial}{\partial T}A(t,T) \\
&= r_t\frac{\partial}{\partial T}B(t,T) - \frac{\partial}{\partial T}A(t,T) \\
&= r_t\frac{\partial}{\partial T}B(t,T) + \frac{\partial}{\partial T}\frac{\sigma^2}{4a}B(t,T)^2 \\
&\quad -\frac{\partial}{\partial T}[B(t,T) - T + t]\left(\frac{a^2 b - \frac{\sigma^2}{2}}{a^2}\right).
\end{aligned}$$

We therefore have

$$\begin{aligned}
f(t,T) &= r_t e^{-a(T-t)} + \frac{\sigma^2}{4a}2B(t,T)e^{-a(T-t)} \\
&\quad -\left(\frac{a^2 b - \frac{\sigma^2}{2}}{a^2}\right)(e^{-a(T-t)} - 1). \qquad (5.2.8)
\end{aligned}$$

The initial curve $f(0,T)$ can now be written by evaluating (5.2.8) at $t = 0$. Hence,

$$\begin{aligned}
f(0,T) &= r_0 e^{-aT} + \frac{\sigma^2}{2a^2}e^{-aT} - \frac{\sigma^2 e^{-2aT}}{2a^2} \\
&\quad -\left(b - \frac{\sigma^2}{2a^2}\right)(e^{-aT} - 1) \\
&= r_0 e^{-aT} + \frac{\sigma^2}{2a^2}e^{-aT} - \frac{\sigma^2 e^{-2aT}}{2a^2} \\
&\quad -\left[be^{-aT} - b - \frac{\sigma^2}{2a^2}e^{-aT} + \frac{\sigma^2}{2a^2}\right] \\
&= (r_0 - b)e^{-aT} + b - \frac{\sigma^2}{2a^2}(e^{-2aT} - 2e^{-aT} + 1).
\end{aligned}$$

Finally, we get in a more compact form:

$$f(0,T) = b + e^{-aT}(r_0 - b) - \frac{\sigma^2}{2a^2}(1 - e^{-aT})^2 \tag{5.2.9}$$

We differentiate $f(t,T)$ in (5.2.8) with respect to t, to obtain its dynamics. We get

$$\begin{aligned} df(t,T) &= r_t a e^{-a(T-t)}dt + e^{-a(T-t)}dr_t \\ &+\frac{\sigma^2}{2a}\left[B(t,T)ae^{-a(T-t)} + e^{-a(T-t)}(-e^{-a(T-t)})\right]dt \\ &-\left(\frac{a^2b - \frac{\sigma^2}{2}}{a}\right) e^{-a(T-t)}dt. \\ &= r_t a e^{-a(T-t)}dt + abe^{-a(T-t)}dt \\ &-r_t a e^{-a(T-t)}dt + \sigma e^{-a(T-t)}dW_t \\ &+\left(\frac{\sigma^2}{2a}e^{-a(T-t)} - \frac{\sigma^2}{2a}e^{-2a(T-t)} - \frac{\sigma^2}{2a}e^{-a(T-t)}\right)dt \\ &-abe^{-a(T-t)}dt + \frac{\sigma^2}{2a}e^{-2a(T-t)}dt \\ &= \sigma e^{-a(T-t)}dW_t - \frac{\sigma^2}{a}[e^{-a(T-t)} - e^{-2a(T-t)}]dt. \end{aligned} \tag{5.2.10}$$

Using (5.2.10), the forward rate $f(t,T)$ in integral form with its drift and volatility terms is given by

$$\begin{aligned} f(t,T) &= f(0,T) + \int_0^t \alpha(u,T)du + \int_0^t \sigma(u,T)dW_u \\ &= (r_0 - b)e^{-aT} + b - \frac{\sigma^2}{2a^2}(e^{-2aT} - 2e^{-aT} + 1) \\ &+\sigma \int_0^t e^{-a(T-u)}dW_u \\ &-\frac{\sigma^2}{a}\int_0^t [e^{-a(T-u)} - e^{-2a(T-u)}]du. \end{aligned} \tag{5.2.11}$$

Equations (5.2.8) and (5.2.11) describe the same dynamics, only that (5.2.11) can tell us the effects or significance of the model by investigating the drift and volatility functions.

Now, if we integrate the original equation for the forward rates in (5.2.11) we then have the bond price $P(t,T)$ equal to

$$
\begin{aligned}
& exp\left[-\left\{\int_0^t\left(\int_t^T \sigma(s,u)du\right)dW_s + \int_t^T f(0,u)du + \int_0^t\int_t^T \alpha(s,u)duds\right\}\right] \\
= & \; exp\left[-\left\{\int_t^T \left[(r_0-b)e^{-au} + b - \frac{\sigma^2}{2a^2}(e^{-2au} - 2e^{-au} + 1)\right] du\right.\right. \\
& +\sigma\int_0^t\left(\int_t^T e^{-a(s-u)}du\right)dW_s \\
& \left.\left.-\frac{\sigma^2}{a}\int_0^t\left(\int_t^T e^{-a(s-u)} - e^{-2a(s-u)}du\right)ds\right\}\right]. \qquad (5.2.12)
\end{aligned}
$$

Using (5.2.11), the short rate is also given by

$$
\begin{aligned}
r_t &= f(0,t) + \int_0^t \sigma(s,t)dW_s + \int_0^t \alpha(s,t)ds \\
&= (r_0-b)e^{-at} + b - \frac{\sigma^2}{2a^2}(e^{-2at} - 2e^{-at} + 1) \\
&\quad +\sigma\int_0^t e^{-a(t-s)}dW_s + \int_0^t \frac{\sigma^2}{2}[e^{-a(t-s)} - e^{-2a(t-s)}]ds \\
&= (r_0-b)e^{-at} + b - \frac{\sigma^2}{2a^2}(1-e^{-at})^2 \\
&\quad +\sigma\int_0^t e^{-a(t-s)}dW_s + \frac{\sigma^2}{2}\int_0^t [e^{-a(t-s)} - e^{-2a(t-s)}]ds \qquad (5.2.13)
\end{aligned}
$$

And simplying by evaluating the integrals, equation (5.2.13) becomes equation (5.1.2). In conclusion, we have therefore shown that under this model, $r_t = f(t,t)$.

Chapter 6

The Short Rate Under the Cox-Ingersoll-Ross Model

6.1 Description of the Model

In the Vasicek's model, there is a positive probability that interest rates can become negative. Cox, Ingersoll and Ross proposed an alternative model that overcome this disadvantage. In rectifying this situation, let us consider how the model originates.

Let $W_t = (W_t^1, W_t^2, ..., W_t^n)$ be an n-dimensional Brownian motion. Suppose further that, $\alpha > 0$ and $\sigma > 0$ are constants. For $j = 1, .., n$, let $X_0^j \in \mathbb{R}^n$ be given so that

$$(X_0^1)^2 + (X_0^2)^2 + ... + (X_0^n)^2 \geq 0,$$

and let X_t^j be the solution to the SDE

$$dX_t^j = \frac{-1}{2}\alpha X_t^j dt + \frac{1}{2}\sigma dW_t^j. \tag{6.1.1}$$

Now, X_t^j is called the Ornstein-Uhlenbeck process. It always has a drift toward the origin.

The solution to the SDE in (6.1.1) is

$$X_t^j = e^{-\frac{1}{2}\alpha t}\left[X_0^j + \frac{1}{2}\sigma\int_0^t e^{\frac{1}{2}\alpha u}dW_u^j\right],$$

using Itô's Lemma.

We note that the solution is a Gaussian process with mean function

$$m_t^j = e^{-\frac{1}{2}\alpha t}X_0^j$$

and covariance function

$$\rho(s,t) = \frac{1}{4}\sigma^2 e^{-\frac{1}{2}\alpha(s+t)}\int_0^{s\wedge t} e^{\alpha u}du.$$

Write: $r_t := (X_t^1)^2 + (X_t^2)^2 + ... + (X_t^n)^2$.

We make the following observations:

1. If $n = 1$, we have $r_t = (X_t^1)^2$ and for each t, $P[r_t > 0] = 1$. But P[There are infinitely many values of $t > 0$ for which $r_t = 0$]$= 1$. See Figure 6.1.

2. If $n \geq 2$, P[There is at least one value of $t > 0$ for which $r_t = 0$]$= 0$. See Figure 6.2.

Let $f(x_1, x_2, ..., x_n) = {x_1}^2 + {x_2}^2 + ... + {x_n}^2$. Then,

$$\frac{\partial f}{\partial x_i} = 2x_i$$

and

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \begin{cases} 2, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Figure 6.1: A One-Dimensional Brownian Motion

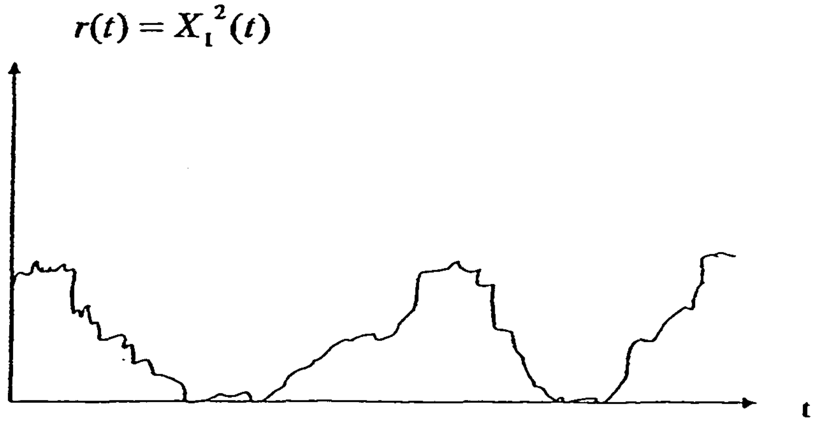
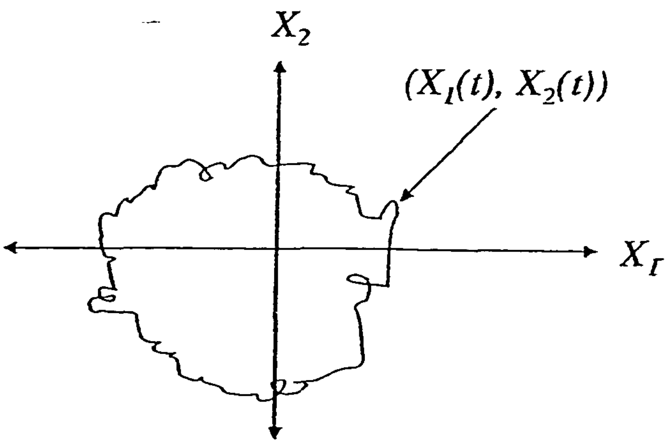


Figure 6.2: A Two-Dimensional Brownian Motion



From the Generalised Itô's Formula, we obtain

$$\begin{aligned}
dr_t &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dX_t^i + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 f}{\partial {x_i}^2} (dX_t^i)^2 \\
&= \sum_{i=1}^{n} 2X_t^i \left(-\frac{1}{2} \alpha X_t^i dt + \frac{1}{2} \sigma dW_t^i \right) \\
&\quad + \sum_{i=1}^{n} \frac{1}{4} \sigma^2 (dW_t^i)^2 \\
&= -\alpha r_t dt + \sigma \sum_{i=1}^{n} X_t^i dW_t^i + \frac{n\sigma^2}{4} dt \\
&= \left(\frac{n\sigma^2}{4} - \alpha r_t \right) dt + \sigma \sqrt{r_t} \sum_{i=1}^{n} \frac{X_t^i}{\sqrt{r_t}} dW_t^i.
\end{aligned}$$

Write $W_t := \sum_{i=1}^{n} \int_0^t \frac{X_u^i}{\sqrt{r_u}} dW_u^i$.

Then, W_t is a martingale,

$$dW_t = \sum_{i=1}^{n} \frac{X_t^i}{\sqrt{r_t}} dW_t^i$$

and

$$dW_t dW_t = \sum_{i=1}^{n} \frac{(X_t^i)^2}{r_t} dt = dt.$$

Hence, W_t is a Brownian motion. We have

$$dr_t = \left(\frac{n\sigma^2}{4} - \alpha r_t \right) dt + \sigma \sqrt{r_t} dW_t.$$

The risk-neutral Cox-Ingersoll-Ross (CIR) process is given by

$$dr_t = a(b - r_t)dt + \sigma \sqrt{r_t} dW_t. \tag{6.1.2}$$

Define $n := \frac{4a}{\sigma^2} > 0$.

If $n \in \mathbb{Z}^+$, then we have the representation

$$r_t = \sum_{i=1}^{n} (X_t^i)^2. \tag{6.1.3}$$

However, we do not require n to be an integer.

Remarks:

1. If $n < 2$ (i.e., $a < \frac{1}{2}\sigma^2$), then P[There are infinitely many values of $t > 0$ for which $r(t) = 0] = 1$. Therefore this is not a good parameter choice.

2. If $n \geq 2$ (i.e., $a \geq \frac{1}{2}\sigma^2$), then P[There is at least one value of $t > 0$ for which $r(t) = 0] = 0$.

We conclude that with the CIR processes, one can derive formulas under the assumption that $n = \frac{4a}{\sigma^2}$ is a positive integer, and they are still correct even when n is not an integer.

The SDE in (6.1.2) has the same mean reverting drift as Vasicek, but the noise term has a volatility proportional to $\sqrt{r_t}$. This tells us that as the short-term interest rate increases, its standard deviation also increases.

Under this model, Cox, Ross and Ingersoll show that the bond prices have the same general form with the Vasicek's model:

$$P(t,T) = A^*(t,T)e^{-B(t,T)r_t}. \tag{6.1.4}$$

However, the functions $B(t,T)$ and $A^*(t,T)$ are different. Here,

$$B(t,T) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma + a)(e^{\gamma(T-t)} - 1) + 2\gamma} \tag{6.1.5}$$

and

$$A^*(t,T) = \left[\frac{2\gamma e^{(a+\gamma)(\frac{T-t}{2})}}{(\gamma + a)(e^{\gamma(T-t)-1)} + 2\gamma)}\right]^{\frac{2ab}{\sigma^2}} \tag{6.1.6}$$

where $\gamma = \sqrt{a^2 + 2\sigma^2}$.

Alternatively, we can re-express $P(t,T)$ in (6.1.4) as

$$P(t,T) = exp\big[A(t,T) - B(t,T)r_t\big] \tag{6.1.7}$$

where

$$\begin{aligned} A(t,T) &= \ln A^*(t,T) \\ &= \frac{2ab}{\sigma^2}\Big[\ln 2\gamma + \frac{(a+\gamma)(T-t)}{2} \\ &\quad - \ln((\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma)\Big] \end{aligned} \tag{6.1.8}$$

and the function $B(t,T)$ is the same given in (6.1.5).

6.2 The Dynamics of $f(t,T)$

Again, invoking (2.1.2), we have the result

$$f(t,T) = -\frac{\partial}{\partial T}\log P(t,T).$$

Therefore,

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}\log exp[A(t,T) - B(t,T)r_t] \\ &= \frac{\partial}{\partial T}B(t,T)r_t - \frac{\partial}{\partial T}A(t,T) \\ &= r_t\frac{\partial}{\partial T}B(t,T) - \frac{\partial}{\partial T}A(t,T). \end{aligned} \tag{6.2.9}$$

We shall evaluate $\frac{\partial B(t,T)}{\partial T}$ and $\frac{\partial A(t,T)}{\partial T}$.

$$\begin{aligned} \frac{\partial B(t,T)}{\partial T} &= \frac{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]2\gamma e^{\gamma(T-t)}}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^2} \\ &\quad -\frac{[2(e^{\gamma(T-t)}-1)][\gamma(\gamma+a)e^{\gamma(T-t)}]}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^2} \end{aligned}$$

$$
\begin{aligned}
\frac{\partial B(t,T)}{\partial T} &= \frac{[(\gamma+a)e^{\gamma(T-t)} - (\gamma+a) + 2\gamma](2\gamma e^{\gamma(T-t)})}{[(\gamma+a)(e^{\gamma(T-t)} - 1) + 2\gamma]^2} \\
&\quad - \frac{[2e^{\gamma(T-t)} - 2][\gamma(\gamma+a)e^{\gamma(T-t)}]}{[(\gamma+a)(e^{\gamma(T-t)} - 1) + 2\gamma]^2} \\
&= \frac{(\gamma+a)e^{\gamma(T-t)} + (\gamma-a)2\gamma + e^{\gamma(T-t)}}{[(\gamma+a)e^{\gamma(T-t)} + \gamma - a]^2} \\
&\quad - \frac{-2\gamma(\gamma+a)e^{2\gamma(T-t)} + 2\gamma(\gamma+a)e^{\gamma(T-t)}}{[(\gamma+a)e^{\gamma(T-t)} + \gamma - a]^2} \\
&= \frac{2\gamma(\gamma+a)e^{2\gamma(T-t)} + 2\gamma(\gamma-a)e^{\gamma(T-t)}}{[(\gamma+a)e^{\gamma(T-t)} + \gamma - a]^2} \\
&\quad - \frac{2\gamma(\gamma+a)e^{2\gamma(T-t)} - 2\gamma(\gamma+a)e^{\gamma(T-t)}}{[(\gamma+a)e^{\gamma(T-t)} + \gamma - a]^2} \\
&= \frac{4\gamma^2 e^{\gamma(T-t)}}{[(\gamma+a)^2 e^{\gamma(T-t)} + \gamma - a]^2}. \qquad (6.2.10)
\end{aligned}
$$

For $\frac{\partial}{\partial T}A(t,T)$ we have from (6.1.8),

$$
\begin{aligned}
\frac{\partial}{\partial T}A(t,T) &= \frac{\partial}{\partial T}\ln A^*(t,T) \\
&= 2ab\left[\left(\frac{a+\gamma}{2}\right) - \frac{(\gamma+a)\gamma e^{\gamma(T-t)}}{(\gamma+a)(e^{\gamma(T-t)} - 1) + 2\gamma}\right]. \qquad (6.2.11)
\end{aligned}
$$

Thus, combining results (6.2.9), (6.2.10) and (6.2.11) we can write down the initial curve:

$$
\begin{aligned}
f(0,T) &= -\frac{\partial}{\partial T}\log P(0,T) \\
&= r_0 \frac{4\gamma^2 e^{\gamma T}}{[(\gamma+a)e^{\gamma T} + \gamma - a]^2} \\
&\quad - \frac{2ab}{\sigma^2}\left[\frac{a+\gamma}{2} - \frac{(\gamma+a)\gamma e^{\gamma T}}{(\gamma+a)(e^{\gamma T} - 1) + 2\gamma}\right]. \qquad (6.2.12)
\end{aligned}
$$

Finally, the forward rate is given by

$$\begin{aligned} f(t,T) &= r_t \frac{\partial}{\partial T} B(t,T) - \frac{\partial}{\partial T} A(t,T) \\ &= r_t \left[\frac{4\gamma^2 e^{\gamma(T-t)}}{[(\gamma+a)e^{\gamma(T-t)} + \gamma - a]^2} \right] \\ &\quad -ab(a+\gamma) + \frac{2ab(\gamma+a)\gamma e^{\gamma(T-t)}}{(\gamma+a)(e^{\gamma(T-t)} - 1) + 2\gamma}. \end{aligned} \tag{6.2.13}$$

To calculate the drift and volatility of $f(t,T)$, let us begin by assuming that r_t is a Markov diffusion though not necessarily time homogeneous with drift $\theta(t,T)$ and volatility $\beta(t,T)$. That is,

$$dr_t = \theta(r_t,t)dt + \beta(r_t,t)dW_t \tag{6.2.14}$$

where $\theta(r_t,t)$ and $\beta(r_t,t)$ are deterministic functions of space and time.

We have seen previously from (2.1.2) that,

$$\int_t^T f(t,u)du = -\log P(t,T) = h(r_t,t,T)$$

where $h(x,t,T)$ is the deterministic function

$$h(x,t,T) = -\log E\left[exp\left(-\int_t^T r_u du \right) \middle| r_t = x \right].$$

Theorem 6.2.1 *The required volatility and drift structures for $f(t,T)$ are respectively, given by*

$$\sigma(t,T) = \beta(r_t,t)\frac{\partial^2 h(r_t,t,T)}{\partial x \partial T}$$

and

$$\alpha(t,T) = \frac{\partial^2 h(r_t,t,T)}{\partial x \partial T}\theta(r_t,t) + \frac{\partial^2 h(r_t,t,T)}{\partial t \partial T}.$$

Proof: Recall that

$$\int_t^T f(t,u)du = -\log P(t,T) = h(r_t,t,T).$$

We therefore see that

$$f(t,T) = \frac{\partial h(r_t,t,T)}{\partial T}.$$

Thus, by Itô's Lemma,

$$\begin{aligned} df(t,T) &= \frac{\partial^2 h}{\partial x \partial T}(\beta(r_t,t)dW_t + \theta(r_t,t)dt) \\ &\quad + \frac{\partial^2 h}{\partial t \partial T}dt + \frac{1}{2}\frac{\partial^3 h}{\partial x^2 \partial T}\beta^2(r_t,t)dt \end{aligned} \tag{6.2.15}$$

Equation (6.2.15) has a volatility term which matches $\sigma(t,T)$. The forward rate $f(t,T)$ is linear in x. Thus, the term $\frac{1}{2}\frac{\partial^3 h}{\partial x^2 \partial T}\beta^2(r_t,t)dt$ is essentially 0.

The two nonzero coefficients of dt give us the drift structure of $f(t,T)$.

Furthermore, the initial curve $f(0,T)$ is given by

$$f(0,T) = \frac{\partial h}{\partial T}(r_0,0,T).$$

An HJM model with the same short rate under P is then identified by these volatility and drift structures and initial curve.

■

Applying Theorem 6.2.1, we obtain the volatility structure of $f(t,T)$ under the CIR model:

$$\begin{aligned} \sigma(t,T) &= \beta(r_t,t)\frac{\partial^2 h}{\partial x \partial T}(r_t,t,T) \\ &= \sigma\sqrt{r_t}\frac{\partial^2 h}{\partial x \partial T}(x,t,T) \\ &= \sigma\sqrt{r_t}\frac{\partial}{\partial T}B(t,T) \\ &= \frac{\sigma\sqrt{r_t}(4\gamma^2 e^{\gamma(T-t)})}{[(\gamma+a)e^{\gamma(T-t)}+\gamma-a]^2}. \end{aligned} \tag{6.2.16}$$

On the other hand, the drift structure is given by

$$
\begin{aligned}
\alpha(t,T) &= \frac{\partial^2 h(r_t,t,T)}{\partial x \partial T}\theta(r_t,t) + \frac{\partial^2 h(r_t,t,T)}{\partial t \partial T} \\
&= \frac{\partial}{\partial T}B(t,T)\alpha(t,r_t) + \frac{\partial}{\partial t}\frac{\partial B(t,T)}{\partial T}r_t \\
&\quad -\frac{\partial}{\partial t}\frac{\partial}{\partial T}A(t,T) \qquad (6.2.17) \\
&= \frac{4\gamma^2 e^{\gamma(T-t)}}{[(\gamma+a)e^{\gamma(T-t)}+\gamma-a]^2}a(b-r_t) \\
&\quad +\frac{\partial}{\partial t}\left[\frac{4\gamma^2 e^{\gamma(T-t)}}{[(\gamma+a)(e^{\gamma(T-t)}+\gamma-a)]^2}\right] \\
&\quad -2ab\frac{\partial}{\partial t}\left[\frac{(\gamma+a)\gamma e^{\gamma(T-t)}}{(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma}\right].
\end{aligned}
$$

Hence, we can write the stochastic dynamics of $f(t,T)$ in SDE form using results in (6.2.16) and (6.2.17). And we get the following:

$$
\begin{aligned}
df(t,T) &= \left[\frac{\partial}{\partial t}\frac{\partial B(t,T)}{\partial T}r_t + \frac{\partial}{\partial T}B(t,T)\theta(t,r_t)\right. \\
&\quad \left. -\frac{\partial}{\partial t}\frac{\partial}{\partial T}A(t,T)\right]dt \\
&\quad +\left[\frac{\partial}{\partial T}B(t,T)\beta(t,r_t)\right]dW_t. \qquad (6.2.18)
\end{aligned}
$$

6.3 Reconciling the HJM and the Short Rate Forms

We concluded the last section with equation (6.2.18) giving the dynamics $df(t,T)$. We have characterised the structures of the drift and volatility.

We shall demonstrate here the validity of such characterisation via the equivalence of the dynamics $d[f(t,T)|_{T=t}] := df(t,t)$ and that of dr_t.

Using equation (6.2.18), we have the following:

$$
\begin{aligned}
d[f(t,T)|_{T=t}] &= df(t,t) \\
&= \left[\frac{\partial}{\partial t}\frac{\partial B(t,T)}{\partial T}\bigg|_{T=t} r_t + \frac{\partial}{\partial T}B(t,T)\bigg|_{T=t} \cdot \theta(t,r_t)\right. \\
&\quad \left. - \frac{\partial}{\partial t}\frac{\partial}{\partial T}A(t,\ T)\bigg|_{T=t}\right] dt \\
&\quad \frac{\partial}{\partial T}B(t,T)\bigg|_{T=t} \beta(t,r_t)dW_t \\
&= \left[\left(\frac{\partial}{\partial t}1\right) \cdot r_t + 1 \cdot \theta(t,r_t) - \frac{\partial}{\partial t}0\right] dt \\
&\quad +1\beta(t,r_t)dW_t \\
&= (0 \cdot r_t + \theta(t,r_t) - 0)dt + \beta(t,r_t)dW_t \\
&= a(b-r_t)dt + \sigma\sqrt{r_t}dW_t \\
&= dr_t.
\end{aligned}
$$

Therefore, under the CIR Model, $df(t,t) = dr_t$.

Chapter 7

Characterisation of a Generalised Exponential Affine Bond Price

7.1 The Exponential Affine Model

Let $W_t = (W_t^1, W_t^2, ..., W_t^n)$ be a standard Brownian motion in $\mathbb{R}^n$ for some $n \geq 1$ restricted to some interval $[0, T]$, over a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$.

We suppose that we are given an adapted short rate process r such that $\int_0^T |r_t| dt < \infty$.

We consider one-factor term structure models for the short rate r given by the SDE of the form

$$dr_t = \alpha(r_t, t)dt + \sigma(r_t, t)dW_t \tag{7.1.1}$$

where $\alpha : \mathbb{R} \times [0, T] \to \mathbb{R}$ and $\sigma : \mathbb{R} \times [0, T] \to \mathbb{R}^n$, and α and σ satisfy technical conditions for the existence of a solution to (7.1.1) for all $T \geq t$ as discussed in Section 2.3 of Chapter 2.

As stated in [33], the one-factor models are so named because the Markov property of the solution of r to (7.1.1) implies from the price of the zero-coupon bond that the short rate is the only state variable, or "factor," on which the current yield curve depends.

That is, for all t and $T \geq t$, we can write

$$P(t,T) = F(t,T,r_t),$$

for some fixed $F : [0,T] \times [0,T] \times \mathbb{R} \to \mathbb{R}$.

In [33], parametric examples of one-factor models were given and each of these models is a special case of the SDE

$$dr_t = [\alpha_t^1 + \alpha_t^2 r_t + \alpha_t^3 r_t \log r_t]dt + [\sigma_t^1 + \sigma_t^2 r_t]^q dW_t \qquad (7.1.2)$$

for continuous functions α_t^1, α_t^2, α_t^3, σ_t^1, and σ_t^2 on $[0,T]$ into $\mathbb{R}$ and for some exponent $q \in [0.5, 1.5]$.

A subset of the models considered in (7.1.2), those with $\alpha_3 = \sigma_2 = 0$ are Gaussian in the short rates $\{r_{t_1}, ..., r_{t_k}\}$ at any finite set $\{t_1, ..., t_k\}$ of times and have a joint normal distribution under P. This follows from the properties of linear stochastic differential equations (see for example Appendix E of [33]).

In the Gaussian case, we can view a negative coefficient function α_t^2 as a mean reversion parameter, in that a higher short rate generates a lower drift, and vice versa. Empirically, mean reversion is widely believed to be a useful attribute to include in single-factor short rate models.

For the Gaussian model, the bond price processes are lognormal, see [33]. This is shown by defining a new process y that satisfies the relation $dy_t = -r_t dt$.

Since (r, z) is the solution of a 2-dimensional linear SDE, (see [78]) for any t, and $T \geq t$, the random variable

$$z_T - z_t = -\int_t^T r_u du$$

is normally distributed.

Under P, the mean μ and variance ϑ of $-\int_t^T r_u du$, conditional on $\mathcal{F}_t$, are computed in terms of r_t, α_t^1, α_t^2, and σ_t^1.

Consequently, we have

$$\begin{aligned} P(t,T) &= E\left[exp\left(-\int_t^T r_u du\right)\middle|\mathcal{F}_t\right] \\ &= exp[\mu + \frac{\vartheta}{2}] \\ &= exp[A(t,T) - B(t,T)r_t] \end{aligned}$$

for some coefficients $A(t,T)$ and $B(t,T)$ that depend only on t and T.

Gaussian models are special cases of single-fact :· models with the property that the bond price (which is a description of the term structure model) is given by

$$P(t,T,r_t) = exp[A(t,T) - B(t,T)r_t] \tag{7.1.3}$$

for some A and B which are continuously differentiable.

Since for all t, the yield

$$-\frac{\log P(t,T,r_t)}{T-t}$$

obtained from (7.1.3) is affine in r_t, we call (7.1.3) an **affine term structure model** or an **exponential affine bond price.** We say that a function $f : \mathbb{R} \to \mathbb{R}$ is **affine** if there are constants α and β such that for all x, $f(x) = \alpha + \beta x$.

A demonstration on how to calculate these coefficients is given in [41]. Further, Elliott and Van der Hoek use stochastic flows and their Jacobians to show why, when the short rate process is described by Gaussian dynamics (as in the Vasicek or Hull-White Model), or square root, affine Bessel processes, (as in the CIR or Duffie-Kan Models), the bond price is an exponential affine function.

Theorem 7.1.1 *Suppose the short rate process is given by Gaussian or affine-square root dynamics. Then its Jacobian has a conditional expectation under the risk-neutral (resp. forward) measure which is deterministic.*

Proof: See [41].

The above theorem implies that the bond price has the exponential affine form. The bond price is determined in the two-factor Hull-White and CIR models by integrating the ordinary differential equation. Hull-White model will be discussed in Chapter 8.

Conversely, if the bond price is exponential affine then we should be able to recover drifts and volatility which are affine in r_t under some technical conditions.

7.2 The Dynamics of the Forward Rate

Suppose the short rate dynamics is given by (7.1.1). And suppose further, the bond price is given by an affine term structure model as in (7.1.3). We wish to evaluate the dynamics of $f(t,T)$.

In other words, if we are given the dynamics

$$dr_t = \alpha(t,r_t)dt + \sigma(t,r_t)dW_t$$

and

$$P(t,T,r_t) = exp\big(A(t,T) - B(t,T)r_t\big),$$

our objective is to find $df(t,T)$.

In terms of the short rates,

$$P(t,T,r_t) = exp\left(-\int_t^T f(t,u)du\right).$$

So,

$$f(t,T) = \frac{\partial}{\partial T}B(t,T)r_t - \frac{\partial}{\partial T}A(t,T). \tag{7.2.4}$$

This implies that

$$\begin{aligned} r_t &= f(t,t) \\ & \quad \frac{\partial}{\partial T}B(t,T)r_t\Big|_{T=t} - \frac{\partial}{\partial T}A(t,T)\Big|_{T=t} \\ &= \frac{\partial}{\partial T}B(t,T)\Big|_{T=t} r_t - \frac{\partial}{\partial T}A(t,T)\Big|_{T=t}. \end{aligned} \tag{7.2.5}$$

Equation (7.2.5) therefore tells us that

$$\frac{\partial}{\partial T}B(t,T)\bigg|_{T=t} = 1 \tag{7.2.6}$$

and

$$\frac{\partial}{\partial T}A(t,T)\bigg|_{T=t} = 0. \tag{7.2.7}$$

Also, from (7.2.4) we obtain

$$\begin{aligned} d_t f(t,T) &= \left(\frac{\partial}{\partial t}\frac{\partial}{\partial T}B(t,T)\right) r_t dt \\ &\quad + \frac{\partial B(t,T)}{\partial T} dr_t - \left(\frac{\partial}{\partial t}\frac{\partial}{\partial T}A(t,T)\right) dt \\ &= \left(\frac{\partial^2}{\partial t \partial T}B(t,T)\right) r_t dt + \frac{\partial}{\partial T}B(t,T)\Big[\alpha(t,r_t)dt \\ &\quad + \sigma(t,r_t)dW_t\Big] - \left(\frac{\partial^2 A(t,T)}{\partial t \partial T}\right) dt. \end{aligned}$$

Thus,

$$\begin{aligned} df(t,T) &= \left[\frac{\partial^2 B(t,T)}{\partial t \partial T} r_t + \frac{\partial}{\partial T}B(t,T)\alpha(t,r_t) \right. \\ &\quad \left. - \frac{\partial^2 A(t,T)}{\partial t \partial T}\right] dt + \frac{\partial}{\partial T}B(t,T)\sigma(t,r_t)dW_t. \end{aligned} \tag{7.2.8}$$

7.3 Examples and Counterexamples

Example 7.3.1 Under the CIR model, the forward rate has dynamics given in (6.2.18) as

$$\begin{aligned}
df(t,T) &= \left[\frac{\partial}{\partial t}\frac{\partial B(t,T)}{\partial T}r_t + \frac{\partial}{\partial T}B(t,T)\theta(t,r_t) - \frac{\partial}{\partial t}\frac{\partial}{\partial T}A(t,T)\right] dt \\
&\quad + \left[\frac{\partial}{\partial T}B(t,T)\beta(t,r_t)\right] dW_t. \\
&= \left[\frac{\gamma(3\gamma+a)e^{\gamma(T-t)}ab}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^2} - \frac{\gamma^2 e^{\gamma(T-t)} - a\gamma e^{\gamma(T-t)}}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^2}\right. \\
&\quad + \left(\frac{[(\gamma+a)^2 - 2\gamma^2(3\gamma+a)(\gamma+a)]e^{2\gamma(T-t)} + 2(\gamma^2-a^2)e^{\gamma(T-t)} + (\gamma+a)^2}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^4}\right. \\
&\quad \left.\left. - \frac{\gamma(3\gamma+a)e^{\gamma(T-t)}a}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^2}\right) r_t\right] dt \\
&\quad + \frac{\gamma(3\gamma+a)e^{\gamma(T-t)}\sigma}{[(\gamma+a)(e^{\gamma(T-t)}-1)+2\gamma]^2} dW_t.
\end{aligned}$$

Clearly, we see that the drift is affine in r_t. We say that under the CIR model which is an exponential affine term structure model, the short rate and the forward rate processes have similar drift structures.

We might be tempted to believe that this is always the case. However, the second example illustrates that this is not true in general. That is, if we start with an affine drift function and a certain volatility structure for the short rate, we may not end up with a forward rate dynamics that have similar drift structures.

Example 7.3.2. Equation (5.2.10) gives the dynamics of the forward rate for the Vasicek's model. We have

$$df(t,T) = \sigma e^{-a(T-t)}dW_t - \frac{\sigma^2}{a}[e^{-a(T-t)} - e^{-2a(T-t)}]dt.$$

Notice that the drift for the forward rate does not have the same form as the form of the stochastic dynamics of the drift for r_t.

7.4 Further Results

In the previous section, we have seen that although we started with an exponential affine term structure bond price, the corresponding forward rate dynamics may not necessarily have affine drift and volatility structures in r_t. We pursue in this section some interesting properties of a generalised exponential affine interest model and prove the converse of Theorem 7.1.1.

7.4.1 The Fundamental PDE of the Bond Price

First, we develop the basics that are used to model interest rates. Portfolios of two (or more) securities are formed that are instantaneously riskless. A no-arbitrage argument requires that these portfolios earn only the risk-free rate. This allows a "market price of risk" to be specified. We shall use this market price of risk in the fundamental PDE, the solution to which gives the price of the claim.

Under this framework, we make the following assumptions: (i) markets are frictionless; (ii) all securities are infinitely divisible; and (iii) markets are efficient.

Suppose that the instantaneous return on the bond is given by $\frac{dP(t,T)}{P(t,T)}$. Further, let this return be given by

$$\frac{dP(t,T)}{P(t,T)} = \mu(t,T)dt + \sigma(t,T)dW_t, \tag{7.4.9}$$

where W_t is again a standard Wiener process. In Equation (7.4.9), the first term on the right hand side gives the expected return and the second term the random part of the return. By assuming the above form for the return, we are assuming that the randomness is generated by a diffusion process.

We begin with a portfolio, Π, of two bonds of different maturities, T_1 and T_2 such that the return on the portfolio is instantaneously riskless. This is done in the following way. Let the portfolio be such that a proportion ω_1 of the total value is invested in bonds of maturity T_1 and proportion $1-\omega_1$ is invested in bonds of maturity T_2. Then the return of the portfolio is given as

$$\frac{d\Pi}{\Pi} = [\omega_1\mu(t,T_1) + (1-\omega_1)\mu(t,T_2)]dt + [\omega_1\sigma(t,T_1) + (1-\omega_1)\sigma(t,T_2)]dW_t.$$

We choose the proportion ω_1 in such a way that it eliminates the second term on the right hand side. That is,

$$\omega_1 = \frac{\sigma(t,T_2)}{\sigma(t,T_2) - \sigma(t,T_1)}.$$

By no-arbitrage argument, the instantaneous return on the portfolio is then riskless. We then obtain

$$\frac{\mu(t,T_1) - r}{\sigma(t,T_1)} = \frac{\mu(t,T_2) - r}{\sigma(t,T_2)}. \tag{7.4.10}$$

Equation (7.4.10) relates the return on bonds of different maturities. Since, any two maturities could have been used to derive (7.4.10), the ratio in equation (7.4.10) is independent of the maturity of the bond.

Write $\lambda(r,t) := \frac{\mu(t,T)-r}{\sigma(r,t)}$.

This term $\lambda(r,t)$ is called the market price of risk. As in Section 2.1, equation (7.4.10) says that the expected excess return earned (return in excess of the risk-free rate) by holding a bond divided by the standard deviation of the return, that is, the excess per unit risk is independent of the maturity of the bond.

Using Equation (7.4.10), the return on the bond maturing at time T can be given as

$$\frac{dP(t,T)}{P(t,T)} = [r_t + \sigma(t,T)\lambda(t,T)]dt + \sigma(t,T)dW_t. \tag{7.4.11}$$

The bond price can be obtained as the solution to the above stochastic differential equation subject to the boundary condition $P(T,T) = 1$.

In addition to the assumptions stated at the outset of this section, we have the following assumptions for the models we are about to consider: (iv) the short-term interest rate follows a diffusion process; and (v) the price of the discount bond depends only on the short-term rate over its term.

We assume that the general form of the evolution of the short-term interest rate has dynamics

$$dr_t = \theta(r_t, t)dt + \beta(r_t, t)dW_t. \tag{7.4.12}$$

We note that the choice of the functional forms for $\theta(r, t)$ and $\beta(r, t)$ are driven by the trade-off between the need to make the model realistic and to maintain analytical tractability. Using the fact that the short-term rates are the only source of uncertainty in the model, Itô's Lemma applied to the bond price $P(t, T)$ gives

$$dP(t,T) = \frac{\partial}{\partial t}P(t,T)dt + \frac{\partial}{\partial r}P(t,T)dr_t + \frac{1}{2}\frac{\partial^2}{\partial r^2}P(t,T)(dr_t)^2. \tag{7.4.13}$$

Theorem 7.4.1 *The bond price $P(t, T)$ satisfies the partial differential equation given by*

$$\begin{aligned}&\frac{\partial}{\partial t}P(t,T) + [\theta(r_t,t) - \beta(r_t,t)\lambda(r_t,t)]\frac{\partial}{\partial r}P(t,T)\\&+\frac{1}{2}\beta(r_t,t)^2\frac{\partial^2}{\partial r^2}P(t,T) - r_t\frac{\partial}{\partial r}P(t,T) = 0.\end{aligned}$$

Proof: Substitute dr_t from (7.4.12) into (7.4.13) to obtain

$$\begin{aligned}dP(t,T) &= \frac{\partial}{\partial t}P(t,T)dt + \frac{\partial}{\partial r}P(t,T)[\theta(r_t,t)dt + \beta(r_t,t)dW_t]\\&\quad+\frac{1}{2}\frac{\partial^2}{\partial r^2}P(t,T)\beta(r_t,t)^2dt.\end{aligned} \tag{7.4.14}$$

From Equation (7.4.11),

$$dP(t,T) = P(t,T)[(r_t + \sigma(t,T)\lambda(t,T))dt + \sigma(t,T)dW_t]. \tag{7.4.15}$$

Now, comparing the drift and volatility parts of Equations (7.4.14) and (7.4.15), we get

$$P(t,T)[r_t+\sigma(t,T)\lambda(r_t,t)] = \frac{\partial}{\partial t}P(t,T)+\frac{\partial}{\partial r}P(t,T)\theta(r_t,t)+\frac{1}{2}\frac{\partial^2}{\partial r^2}P(t,T)\beta^2(r_t,t)$$

and

$$\sigma(t,T) = \frac{\frac{\partial}{\partial r}P(t,T)\beta(r_t,t)}{P(t,T)}.$$

Consequently, we have

$$\begin{aligned} rP(t,T) + \frac{\partial}{\partial r}P(t,T)\beta(r_t,t)\lambda(r_t,t) &= \frac{\partial}{\partial t}P(t,T) + \frac{\partial}{\partial r}P(t,T)\theta(r_t,t) \\ &+\frac{1}{2}\frac{\partial^2}{\partial r^2}P(t,T)\beta^2(r_t,t) \qquad (7.4.16) \end{aligned}$$

Rearranging (7.4.16), the result follows.

■

Write $\alpha := \theta(r_t,t) - \beta(r_t,t)\lambda(r_t,t)$. The fundamental PDE now becomes

$$\frac{\partial}{\partial t}P(t,T) + \alpha(r_t,t)\frac{\partial}{\partial r}P(t,T) + \frac{1}{2}\beta^2(r_t,t)\frac{\partial^2}{\partial r^2}P(t,T) - rP(t,T) = 0.$$

Notice that $\alpha(r_t,t)$ is equal to $\theta(r_t,t)$ minus a term which incorporates the market price of risk. Here, we call $\alpha(r_t,t)$ the "risk adjusted" drift of the short rate process.

7.5 Further Characterisation of the Affine Yield Model

7.5.1 One-Factor Exponential Affine Model

As a recap, we are considering a zero-coupon bond with a particular form for the bond price given by

$$P(t,T,r_t) = exp[A(t,T) - B(t,T)r_t].$$

In the succeeding discussion, all models considered are assumed to be under a time homogeneous framework. We now consider the converse of Theorem 7.1.1.

Theorem 7.5.1 *Suppose a model belongs to an exponential affine class, i.e., $P(t,T) = exp[A(t,T) - B(t,T)r_t]$. Then, both $\alpha(r_t,t)$ and $\beta^2(r_t,t)$ are affine in r_t.*

Proof: We begin with the bond price in exponential affine form:

$$P(t,T) = exp[A(t,T) - B(t,T)r_t].$$

We therefore have

$$\frac{\partial}{\partial r}P(t,T) = -P(t,T)B(t,T) \qquad (7.5.17)$$

and

$$\frac{\partial^2}{\partial r^2}P(t,T) = P(t,T)B^2(t,T). \qquad (7.5.18)$$

The fundamental PDE is

$$\frac{\partial}{\partial t}P(t,T) + \alpha(r_t,t)\frac{\partial}{\partial r}P(t,T) + \frac{1}{2}\beta^2(r_t,t)\frac{\partial^2}{\partial r^2}P(t,T) - rP(t,T) = 0.$$

Substitute (7.5.17) and (7.5.18) to the fundamental PDE to obtain

$$\alpha(r_t,t)[-B(t,T)P(t,T)] + P(t,T)\left[\frac{\partial}{\partial t}A(t,T) - \frac{\partial}{\partial t}B(t,T)r_t\right] + \frac{1}{2}\beta^2(r_t,t)P(t,T)B(t,T) - rP(t,T) = 0.$$

Consequently, we get

$$-\alpha(r_t,t)B(t,T) = -\frac{\partial}{\partial t}A(t,T) + \frac{\partial}{\partial t}B(t,T)r_t - \frac{1}{2}\beta^2(r_t,t)B^2(t,T) + r_t.$$

Or,

$$\frac{1}{2}\beta^2B^2 - \alpha B = \frac{\partial}{\partial t}Br - \frac{\partial}{\partial t}A + r \qquad (7.5.19)$$

Differentiating (7.5.19) with respect to T, we get

$$\beta^2 B\frac{\partial B}{\partial T} - \alpha\frac{\partial B}{\partial T} = \frac{\partial^2 B}{\partial T \partial t}r - \frac{\partial^2 A}{\partial T \partial t} \tag{7.5.20}$$

Differentiating (7.5.20) once more with respect to T, we further obtain

$$\beta^2 B\frac{\partial^2 B}{\partial T^2} + \beta^2\left(\frac{\partial B}{\partial T}\right)^2 - \alpha\frac{\partial^2 B}{\partial T^2} = \frac{\partial^3 B}{\partial t \partial T^2}r - \frac{\partial^3 A}{\partial t \partial T^2}. \tag{7.5.21}$$

We evaluate (7.5.21) at $T = t$.

$$\beta^2 B(t,t)\left.\frac{\partial^2 B}{\partial T^2}\right|_{T=t} + \beta^2\left.\left(\frac{\partial B}{\partial T}\right)^2\right|_{T=t} - \alpha\left.\frac{\partial^2 B}{\partial T^2}\right|_{T=t}$$
$$= \left.\frac{\partial^3 B}{\partial t \partial T^2}r\right|_{T=t} - \left.\frac{\partial^3 A}{\partial t \partial T^2}\right|_{T=t}$$

We note that $B(t,t) = 0$, $A(t,t) = 0$ and $\frac{\partial B}{\partial T}|_{T=t} = 1$, and so we have

$$0 + \beta^2 - \alpha\left.\frac{\partial^2 B}{\partial T^2}\right|_{T=t} = \left.\frac{\partial^3 B}{\partial t \partial T^2}r\right|_{T=t} + \left.\frac{\partial^2 A}{\partial T \partial t}\right|_{T=t} \tag{7.5.22}$$

Therefore,

$$\alpha u(t) = b_1(t)r + b_2(t),$$

where

$$u(t) = -\left.\frac{\partial^2 B}{\partial T^2}\right|_{T=t} \neq 0$$
$$b_1(t) = \left.\frac{\partial^3 B}{\partial t \partial T^2}\right|_{T=t} \neq 0$$
$$b_2(t) = \left.\frac{\partial^2 A}{\partial T \partial t}\right|_{T=t} - \beta^2.$$

Or,

$$\alpha = c_1(t)r + c_2(t),$$

where

$$c_1(t) = \frac{b_1(t)}{u(t)}$$

and

$$c_2(t) = \frac{b_2(t)}{u(t)}.$$

Clearly, α is an affine function of r.

Restating (7.5.22) on the other hand, we have or

$$\begin{aligned}\beta^2 &= \left.\frac{\partial^3 B}{\partial t \partial T^2}\right|_{T=t} r + \left.\frac{\partial^2 A}{\partial T \partial t}\right|_{T=t} + \alpha \left.\frac{\partial^2 B}{\partial T^2}\right|_{t=T} \\ &= d_1(t) r + d_2(t),\end{aligned}$$

where

$$d_1(t) = \left.\frac{\partial^3 B}{\partial t \partial T^2}\right|_{T=t} \neq 0$$

and

$$d_2(t) = \alpha \left.\frac{\partial^2 B}{\partial T^2}\right|_{T=t}.$$

Clearly, β^2 is affine in r.

■

7.5.2 Multi-Factor Exponential Affine Model

We extend the result of the last section to n factors or sources of uncertainty for the interest rate r.

Suppose the dynamics of r is given by $dr_t = \alpha(r_t, t)dt + \beta(r_t, t)dW_t$ where $\alpha(r_t, t)$ is the risk-adjusted drift and W_t is a Brownian motion. Further, suppose $r \in \mathbb{R}^n$, $\alpha \in \mathbb{R}^n$ and $\beta \in R^n$.

Consider a function $\rho : [0, \infty) \to \mathbb{R}^n$ and assume that the short rate is modelled by $R(t) = \langle \rho(t), r_t \rangle$.

Thus,

$$P(t, T, r) = E\left[exp\left(-\int_t^T R(r_u)du\right)\middle| r_t = r\right]$$

and suppose

$$P(t, T, r) = exp[A(t, T) - B_1(t, T)r_1 - B_2(t, T)r_2 - \cdots - B_n(t, T)r_n].$$

Observe that

$$
\begin{aligned}
& E\left[exp\left(-\int_0^T R(r_u)du\right)\Big|\mathcal{F}_t\right] \\
&= E\left[exp\left(-\int_0^t R(r_u)du-\int_t^T R(r_u)du\right)\Big|\mathcal{F}_t\right] \\
&= exp\left(-\int_0^t R(r_u)du\right) E\left[exp\left(-\int_t^T R(r_u)du\right)\Big|\mathcal{F}_t\right] \\
&= exp\left(-\int_0^t R(r_u)du\right) P(t,T,r_t).
\end{aligned}
$$

With,

$$
exp\left(-\int_0^t R(r_u)du\right) P(t,T,r_t) = E\left[exp\left(-\int_0^T R(r_u)du\right)\Big|\mathcal{F}_t\right],
$$

we see that the random variable

$$
exp\left(-\int_0^T R(r_u)du\right)
$$

is a martingale.

Define

$$
V(t,r_t) = exp\left(-\int_0^t R(r_u)du\right) P(t,T,r_t)
$$

and $V(t,r_t)$ is a martingale.

By Itô's Lemma,

$$
\begin{aligned}
V(t,r_t) &= V(0,r_0) + \int_0^t \frac{\partial V}{\partial u}du + \int_0^t \frac{\partial V}{\partial r}\cdot dr \\
&\quad + \frac{1}{2}\int_0^t \sum_{i=1}^n \frac{\partial^2 V}{\partial r_i^2}\beta_i^2 du
\end{aligned}
$$

where $\frac{\partial V}{\partial r}\cdot dr$ represents the dot product of $\frac{\partial V}{\partial r}$ and dr, $r=(r_1, r_2, \cdots, r_n)$.

Equivalently,

$$
\begin{aligned}
V(t,r_t) &= V(0,r_0) + \int_0^t \left(\frac{\partial V}{\partial u} + \frac{\partial V}{\partial r}\cdot\alpha + \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 V}{\partial r_i^2}\beta_i^2\right) du \\
&\quad + \int_0^t \frac{\partial V}{\partial r}\cdot\beta dW.
\end{aligned}
$$

Since $V(t, r_t)$ is a martingale, the drift term must be zero; that is,

$$\int_0^t \left(\frac{\partial V}{\partial u} + \frac{\partial V}{\partial r} \cdot \alpha + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 V}{\partial r_i^2} \beta_i^2 \right) du = 0.$$

In differential form,

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial r} \cdot \alpha + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 V}{\partial r_i^2} \beta_i^2 = 0.$$

With

$$V = \left(exp - \int_0^t R(r_u) du \right) P(t, T, r_t)$$

we get the following:

$$\frac{\partial V}{\partial t} = -R(r_t)V + exp\left(- \int_0^t R(r_u) du \right) \frac{\partial}{\partial t} P(t, T)$$

$$\frac{\partial V}{\partial r_i} = exp\left(- \int_0^t R(r_u) du \right) \frac{\partial}{\partial r_i} P(t, T, r_t)$$

$$\frac{\partial^2 V}{\partial r_i^2} = exp\left(- \int_0^t R(r_u) du \right) \frac{\partial^2}{\partial r_i^2} P(t, T, r_t).$$

Therefore the SDE satisfied by the bond price is:

$$-R(r)P + \frac{\partial P}{\partial t} + \frac{\partial P}{\partial r} \cdot \alpha + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 P}{\partial r_i^2} \beta_i^2 = 0, \qquad (7.5.23)$$

where $R(r) = \langle \rho, r \rangle$.

We started to suppose that

$$P(t, T, r) = exp[A(t, T) - B_1(t, T)r_1 - B_2(t, T)r_2 - \cdots - B_n(t, T)r_n].$$

Hence, we also obtain the following:

$$\frac{\partial P}{\partial t} = P \left[\frac{\partial}{\partial t} A(t, T) - \sum_{i=1}^{n} \frac{\partial}{\partial t} B_i(t, T) r_i \right] \qquad (7.5.24)$$

$$\frac{\partial P}{\partial r_i} = -B_i(t,T)P \tag{7.5.25}$$

$$\frac{\partial^2 P}{\partial r_i^2} = B_i(t,T)^2 P \tag{7.5.26}$$

We substitute (7.5.24), (7.5.25) and (7.5.26) into (7.5.23) and obtain

$$\begin{aligned} -\langle \rho_t, r\rangle + \left[\frac{\partial}{\partial t}A(t,T) - \sum_{i=1}^{n} \frac{\partial}{\partial t}B_i(t,T)r_i\right] \\ -\sum_{i=1}^{n} \alpha_1(r,t)B_i(t,T) + \frac{1}{2}\sum_{i=1}^{n} B_i^2(t,T)\beta_i^2(r,t) = 0. \end{aligned} \tag{7.5.27}$$

For $1 \le j \le 2n$: denote by

$$\frac{\partial^j}{\partial T^j}\left(\frac{\partial A(t,T)}{\partial t}\right) = \phi_j(t,T).$$

$$\frac{\partial^j}{\partial T^j}B_i(t,T) = u_{ji}(t,T)$$

$$\frac{\partial^j}{\partial T^j}\left[\frac{1}{2}B_i^2(t,T)\right] = v_{ji}(t,T)$$

$$\frac{\partial^j}{\partial T^j}\left[\frac{\partial B_i(t,T)}{\partial t}\right] = \psi_{ji}(t,T).$$

We differentiate (7.5.27) with respect to T $2n$ times and evaluate each equation at $t = T$. We shall obtain the system of equations given by

$$\begin{aligned} &\sum_{i,j} \alpha_i(r,t)u_{ji}(t) + \sum_{i,j} \beta_i^2(r,t)v_{ji}(t) \\ = \; &\phi_j(t) + \psi_{j_1}(t)r_1 + \psi_{j_2}(t)r_2 + \cdots + \psi_{j_n}(t)r_n, \end{aligned} \tag{7.5.28}$$

for $1 \le i \le n$ and $1 \le j \le 2n$.

Now, let

$$X = \begin{pmatrix} \alpha \\ \beta^2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta^2 \end{pmatrix}'$$

be a $2n \times 1$–vector where $\alpha = \begin{pmatrix} \alpha_1, & \alpha_2, & \cdots, & \alpha_n \end{pmatrix}'$, $\beta^2 = \begin{pmatrix} \beta_1^2, & \beta_2^2, & \cdots, & \beta_n^2 \end{pmatrix}'$.

Write

$$u(t) := u_{ji}(t) \quad (2n \times n) - \text{matrix}$$

$$v(t) := v_{ji}(t) \quad (2n \times n) - \text{matrix}$$

$$\phi(t) := \phi_j(t) \quad (2n \times 1) - \text{vector}$$

$$\psi(t) := \psi_{ji}(t) \quad (2n \times n) - \text{matrix}$$

$$r(t) := [r_1(t), r_2(t), \cdots, r_n(t)]' \quad (n \times 1) - \text{vector}$$

$$Y(t) := [u(t), v(t)] \quad (2n \times 2n) - \text{matrix}$$

$$\Gamma(t) := \phi(t) + \psi(t)r(t) \quad (2n \times 1) - \text{matrix}$$

Thus, the system of equations in (7.5.28) is reduced to

$$Y(t)X = \Gamma(t)$$

or

$$X = Y(t)^{-1}\Gamma(t).$$

That is,

$$\begin{pmatrix} \alpha \\ \beta^2 \end{pmatrix} = [u(t), v(t)]^{-1}[\phi(t) + \psi(t)r(t)]$$
$$= [u(t), v(t)]^{-1}\phi(t) + [u(t), v(t)]^{-1}\psi(t)r(t).$$

Clearly, each component of X, i.e., the α_i and β_i^2 are affine in r with coefficients which are functions of t.

7.6 The Meaning of Neutral Risk and Market Price of Risk

Consider the PDE satisfied by the bond price

$$\frac{\partial P}{\partial t} + \frac{1}{2}\beta^2\frac{\partial^2 P}{\partial r^2} + (\theta - \lambda\beta)\frac{\partial P}{\partial r} - rP = 0. \quad (7.6.29)$$

The bond pricing PDE contains references to the functions $\theta - \lambda\beta$ and β, which are coefficients of the first derivative with respect to the spot rate and of the diffusive, second-order derivative, respectively.

The four terms in (7.6.29) could denote the following in order: time decay, diffusion, drift and discounting.

One interpretation of the solution of the above PDE is that it represents the expected present value of all cashflows. Analogous to equity options, this expectation is taken with respect to the risk-neutral variable and not with the real variable.

This is apparent because the drift term in the equation is not the drift of the real spot rate but the drift of another rate, called the risk-neutral spot rate. This rate has a drift of $\theta - \lambda\beta$. When pricing interest rate derivatives (including bonds of finite maturity), it is important to model the price using risk-neutral rate. This rate satisfies

$$dr_t = [\theta(r_t, t) - \lambda(r_t, t)\beta(r_t, t)]dt + \beta(r_t, t)dW_t.$$

The new market price of risk term is needed because the modelled variable r, is not traded. Thus, if λ is set to zero, then any results are applicable to the real world. In particular, if the distribution of the spot rate at some time is required then we would solve a Fokker-Planck equation with the real drift and not the risk-neutral drift.

The function λ is not however observed (except possibly via the whole yield curve), thus this could be a mechanism under which pricing becomes a straightforward calculation in the setting of no-arbitrage argument.

Chapter 8

Models for Mean Reversion Levels

We have mentioned the inadequacy of single-factor models in Chapter 2. This chapter considers formulation of multi-factor models. In particular, we shall be considering two-factor models. The factors would be interest rates themselves and the level of mean reversion process. In the study of interest rates, mean reversion appears to be the most relevant to include in the model and hence two factors apparently appear sufficient. This may sound simple, however we note that the most influential new studies on connections between real and financial variables use fewer, but carefully selected, explanatory variables, by contrast to Vector Autoregressions with numerous, sometimes over 100, explanatory variables that had come to dominate an earlier style of econometrics.

8.1 Mean Reversion

Interest rates appear to be pulled back to some long-run average level as time passes by. This phenomenon is known as *mean reversion*. When r is low, mean reversion tends to cause it to have a positive drift and when r is high, mean reversion tends to cause it to have a negative drift.

The supply and demand analysis backs up the very support of mean reversion. We observe that when interest rates are low, there is a high demand for funds on the part of the borrowers. This causes interest rates to rise. During the time when interest rates are high, the economy tends to slow down because there is less requirement on the part of the borrowers. Consequently, the rates decline.

One effect of mean reversion is that the volatility of interest rates becomes a decreasing function of maturity. For instance, the 5-year spot interest rate tends to have a lower volatility than the 2-year interest rate, the 2-year spot interest rate tends to have a lower volatility than the 1-year spot interest rate, and so on.

Furthermore, with mean reversion, the volatility of the 3-month forward interest rate starting in 3 months is greater than the volatility of the 3-month forward interest rate starting in 2 years; this in turn is greater than the volatility of the 3-month forward rate starting in 5 years; and so on. In other words, mean reversion also causes the forward rate volatility to decline as the maturity of the forward contract increases.

In addition, mean reversion has some impact on bond price volatilities. It is responsible for the fact that the curvature describing the relationship between the bond price volatility versus maturity is increasing and concave downward. Such interest rate behaviour is consistent with time preference theory that serves a significant portion espoused by Irving Fisher in his two volumes.

Both Vasicek's and CIR models capture this mean reversion property. However, the mean reverting levels in those models are constant. We shall investigate cases where the mean reverting level is (1) time varying, or (2) follows a certain stochastic dynamics or (3) satisfies a Markov process in continuous time with a finite discrete state space.

8.2 The Hull-White Model

This model is a generalisation of the Vasicek model using deterministic, time varying coefficients. In particular, the short rate process is supposed given by the SDE

$$dr_t = \big(\alpha(t) - \beta(t)r_t\big)dt + \sigma(t)dW_t \tag{8.2.1}$$

for $r_0 > 0$. Here α, β, and σ are deterministic functions of t.

Write

$$b(t) := \int_0^t \beta(u)du.$$

Thus, b is also non-random.

The solution of (8.2.1) can be obtained by the method of variation of constants and given by

$$r_t = e^{-b(t)}\left(r_0 + \int_0^t e^{b(u)}\alpha(u)du + \int_0^t e^{b(u)}\sigma(u)dW_u\right).$$

Consequently, r is a Gaussian, Markov process with mean

$$E[r_t] = m(t) = e^{-b(t)}\left[r_0 + \int_0^t e^{b(u)}\alpha(u)du\right]$$

and its covariance is

$$Cov(r_t, r_s) = e^{-b(s)-b(t)}\int_0^{s\wedge t} e^{2b(u)}\sigma^2(u)du.$$

It can be shown under this term structure model that the price of a zero-coupon bond is given by

$$P(t,T) = exp(-r_tC(t,T) - A(t,T)) \tag{8.2.2}$$

(see [38] for example) where,

$$C(t,T) = e^{b(t)}\int_t^T e^{-b(u)}du = e^{b(t)}\gamma(t), \quad \gamma(t) := \int_t^T e^{-b(u)}du$$

and

$$A(t,T) = \int_t^T \left[e^{b(u)}\alpha(u)\gamma(u) - \frac{1}{2}e^{2b(u)}\sigma^2(u)\gamma^2(u) \right] du.$$

We aim to get the dynamics of $f(t,T)$.

Now,

$$f(t,T) = -\frac{\partial}{\partial T}\log P(t,T) = \frac{\partial}{\partial T}\left[r_t C(t,T) + A(t,T)\right],$$

where $C(t,T)$ and $A(t,T)$ are defined as above and again,

$$\gamma(t) = \int_t^T e^{-b(u)}du$$

and

$$b(t) := \int_0^t \beta(u)du.$$

Further,

$$\begin{aligned} f(t,T) &= r_t\frac{\partial}{\partial T}C(t,T) + \frac{\partial}{\partial T}A(t,T) \\ &= r_t e^{b(t)}e^{-b(T)} + e^{b(T)}\alpha(T)\gamma(T) \\ &\quad -\frac{1}{2}e^{2b(T)}\sigma^2(T)\gamma^2(T) \\ &= r_t e^{b(t)-b(T)} + e^{b(T)}\alpha(T)\cdot 0 \\ &\quad -\frac{1}{2}e^{2b(T)}\sigma^2(T)(0)^2 \\ &= r_t e^{b(t)-b(T)}. \end{aligned} \tag{8.2.3}$$

With equation (8.2.3), we obtain

$$df(t,T) = r_t d[e^{b(t)-b(T)}] + e^{b(t)-b(T)}dr_t. \tag{8.2.4}$$

Note that $dr_t = \big(\alpha(t) - \beta(t)r_t\big)dt + \sigma(t)dW_t$.

Thus, equation (8.2.4) becomes

$$
\begin{aligned}
df(t,T) &= r_t[e^{b(t)-b(T)}\beta(t)]dt \\
&\quad + e^{b(t)-b(T)}[(\alpha(t)-\beta(t)r_t)dt + \sigma(t)dW_t] \\
&= r_t e^{b(t)-b(T)}\beta(t)dt + \alpha(t)e^{b(t)-b(T)}dt \\
&\quad - r_t e^{b(t)-b(T)}\beta(t)dt + e^{b(t)-b(T)}\sigma(t)dW_t \\
&= \alpha(t)e^{b(t)-b(T)}dt + e^{b(t)-b(T)}\sigma(t)dW_t.
\end{aligned}
$$

In other words, $f(t,T)$ has a drift $u(t,T)$ equal to $\alpha(t)e^{b(t)-b(T)}$ and volatility $v(t,T)$ equal to $e^{b(t)-b(T)}\sigma(t)$.

Therefore,

$$
\begin{aligned}
f(t,t) &= f(0,t) + \int_0^t u(s,t)ds + \int_0^t v(s,t)dW_s \\
&= f(0,t) + \int_0^t \alpha(s)e^{b(s)-b(t)}ds \\
&\quad + \int_0^t \sigma(s)e^{b(s)-b(t)}dW_s \\
&= r_0 e^{b(0)-b(t)} + \int_0^t \alpha(s)e^{b(s)-b(t)}ds \\
&\quad + \int_0^t \sigma(s)e^{b(s)-b(t)}dW_s.
\end{aligned}
$$

However, $e^{b(0)} = e^0 = 1$ since $b(0) = \int_0^0 \beta(u)du = 0$.

Consequently,

$$
\begin{aligned}
f(t,t) &= r_0 e^{-b(t)} + e^{-b(t)}\int_0^t \alpha(s)e^{b(s)}ds \\
&\quad + e^{-b(t)}\int_0^t \sigma(s)e^{b(s)}ds.
\end{aligned}
$$

Finally,

$$
f(t,t) = e^{-b(t)}\left[r_0 + \int_0^t \alpha(s)e^{b(s)}ds + \int_0^t \sigma(s)e^{b(s)}dW_s\right]
$$

which is the solution to the SDE described in (8.2.1).

Hence, we have shown that $f(t,t) = r_t$ under the Hull-White Model.

8.3 The Two-Factor Gaussian Model

We explore extension of the Vasicek model. Suppose the interest rate process is given by

$$dr_t = (\alpha_t - ar_t)dt + \sigma_1 dW_t^1 \tag{8.3.5}$$

where

$$d\alpha_t = (c - b\alpha_t)dt + \sigma_2 dW_t^2. \tag{8.3.6}$$

The model in equation (8.3.5) specifies a mean reversion level at a rate a. It is a Vasicek model where the mean reversion level α_t is Vasicek by itself with mean reversion level c at the rate of b described in (8.3.6). Alternatively, (8.3.5) can be characterised as the Vasicek model with a mean-reversion level which is time dependent and with stochastic component.

Here, W_t^1 and W_t^2 are independent Brownian motions on $(\Omega, \mathcal{F}, P)$ and $\{\mathcal{F}_t\}$ is the filtration generated by $W = (W_t^1, W_t^2)$. This particular model under current investigation is a special case of a generalised Gaussian model.

We shall consider first the formulation of the generalised Gaussian model then we can just treat the two-factor model as a particular case.

8.3.1 The Generalised Gaussian Model

Definition 8.3.1 *The* **Jacobian** *of the transformation T given by $x = g(u,v)$ and $y = h(u,v)$ is*

$$\begin{aligned}\frac{\partial(x,y)}{\partial(u,v)} &= \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \\ &= \frac{\partial x}{\partial u}\cdot\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\cdot\frac{\partial y}{\partial u}.\end{aligned}$$

Let P be the risk-neutral probability measure. Consider the process $x \in \mathbb{R}^n$. Suppose the process x follows the SDE given by

$$dx_t = \gamma_t dt + (Ax_t + b) + \sigma dW_t. \tag{8.3.7}$$

In this case, W is an $n-$dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$, $A \in \mathbb{R}^{n \times n}$ is a matrix, $\gamma_t \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}^{n \times n}$.

Write $\xi_{s,t}$ for the solution of (8.3.7) such that $\xi_{s,s} = x$. That is,

$$\begin{aligned} \xi_{s,t}(x) &= x + \int_s^t \gamma(u)du + \int_s^t (A\xi_{s,u}(x) + b)du \\ &\quad +\sigma(W_t - W_s). \end{aligned}$$

We suppose further that the map $x \to \xi_{st}(x)$ is differentiable a.s. Then, writing

$$D_{s,t} = \frac{\partial \xi_{s,t}(x)}{\partial x},$$

the Jacobian satisfies the equation

$$D_{s,t} = I + A \int_s^t D_{s,u} du.$$

That is, $D_{s,t}$ is the deterministic matrix $e^{A(t-s)}$.

8.3.2 The Bond Price Under the Two-Factor Model

Applied to the current investigation, we have a special case of (8.3.7) based on the SDE's (8.3.5) and (8.3.6); that is,

$$\xi_{s,t} = \begin{pmatrix} r_{s,t} \\ \alpha_{s,t} \end{pmatrix}, \qquad \gamma = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad A = \begin{pmatrix} -a & 1 \\ 0 & -b \end{pmatrix}$$

$$b = \begin{pmatrix} 0 \\ c \end{pmatrix}, \qquad W = \begin{pmatrix} W_t^1 \\ W_t^2 \end{pmatrix}, \qquad \sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$$

In matrix form, equation(8.3.5) and (8.3.6) can be expressed as

$$\begin{pmatrix} dr_t \\ d\alpha_t \end{pmatrix} = \begin{pmatrix} -a & 1 \\ 0 & -b \end{pmatrix} \begin{pmatrix} r_t \\ \alpha_t \end{pmatrix} dt + \begin{pmatrix} 0 \\ c \end{pmatrix} dt + \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} dW_t^1 \\ dW_t^2 \end{pmatrix}.$$

The components of the Jacobian can be calculated from the dynamics

$$\begin{aligned} r_{s,t}(r,\alpha) &= r + \int_s^t \Big(\alpha_{s,v}(v) - a r_{s,v}(r,\alpha) \Big) dv \\ & \quad + \sigma_1 (W_t^1 - W_s^1). \end{aligned}$$

and

$$\alpha_{s,t}(\alpha) = \alpha + c(t-s) - b \int_s^t \alpha_{s,v} dv + \sigma_2 (W_t^2 - W_s^2).$$

Then, differentiating these equations with respect to their initial conditions gives the following expressions for the components of the Jacobians.

$$\frac{\partial \alpha_{s,t}(\alpha)}{\partial \alpha} = 1 - b \int_s^t \frac{\partial \alpha_{s,v}(\alpha)}{\partial \alpha} dv.$$

So,

$$\frac{\partial \alpha_{s,t}(\alpha)}{\partial \alpha} = e^{-b(t-s)}.$$

Also,

$$\frac{\partial r_{s,t}(r,\alpha)}{\partial r} = 1 - a \int_s^t \frac{\partial r_{s,v}(\alpha)}{\partial r} dv.$$

So,

$$\frac{\partial r_{s,t}(r,\alpha)}{\partial r} = e^{-a(t-s)}.$$

Furthermore,

$$\frac{\partial r_{s,t}(r,\alpha)}{\partial \alpha} = \int_s^t \frac{\partial \alpha_{s,v}(r,\alpha)}{\partial \alpha} dv - a \int_s^t \frac{\partial r_{s,v}(r,\alpha)}{\partial \alpha} dv.$$

So,

$$\frac{\partial r_{s,t}(r,\alpha)}{\partial \alpha} = \frac{1}{(a-b)} [e^{-b(t-s)} - e^{-a(t-s)}].$$

Hence,

$$\int_t^T \frac{\partial r_{t,v}(r,\alpha)}{\partial r} dv = \frac{1}{a}(1 - e^{-a(T-t)}).$$

Write $B(t,T) := \frac{1}{a}(1 - e^{-a(T-t)})$.

In addition,

$$\int_t^T \frac{\partial r_{t,v}(r,\alpha)}{\partial \alpha} dv = \frac{-1}{a(a-b)} e^{-a(T-t)} + \frac{1}{b(a-b)} e^{-b(T-t)} - \frac{1}{ab}.$$

Write

$$C(t,T) := \frac{-1}{a(a-b)}e^{-a(T-t)} + \frac{1}{b(a-b)}e^{-b(T-t)} - \frac{1}{ab}.$$

Therefore, the price of a zero coupon bond in this model is

$$P(t,T,(r,\alpha)) = E\left[exp\left(-\int_t^T r_{t,v}(r,\alpha)dv\right)\middle| r_t = r, \alpha_t = \alpha\right].$$

Consequently,

$$\begin{aligned}\frac{\partial P}{\partial r} &= E\left[\left(-\int_t^T \frac{\partial r_{t,v}(r,\alpha)}{\partial r}dv\right) exp\left(-\int_t^T r_{t,v}(r,\alpha)dv\middle|\mathcal{F}_t\right)\right] \\ &= -B(t,T)P(t,T). \end{aligned} \quad (8.3.8)$$

Similarly,

$$\begin{aligned}\frac{\partial P}{\partial \alpha} &= E\left[\left(-\int_t^T \frac{\partial r_{t,v}(r,\alpha)}{\partial \alpha}dv\right) exp\left(-\int_t^T r_{tv}(r,\alpha)dv\middle|\mathcal{F}_t\right)\right] \\ &= -C(t,T)P(t,T). \end{aligned} \quad (8.3.9)$$

Integrating (8.3.8) in r :

$$P(t,T,(r,\alpha)) = exp(-B(t,T)r)\phi(\alpha,t) \quad (8.3.10)$$

where ϕ is independent of r. Therefore,

$$\begin{aligned}\frac{\partial P(t,T,(r,\alpha))}{\partial \alpha} &= exp\left(-B(t,T)r\frac{\partial \phi(\alpha,t)}{\partial \alpha}\right) \\ &= P(t,T,(r,\alpha))\frac{\partial \phi(\alpha,t)}{\partial \alpha}\frac{1}{\phi(\alpha,t)}, \qquad \text{from (8.3.10)} \\ &= -C(t,T)P(t,T,(r,\alpha)) \qquad \text{from (8.3.9)}.\end{aligned}$$

Hence,

$$\frac{\partial \phi}{\partial \alpha} = -C(t,T)\phi(\alpha,t)$$

and

$$\phi(\alpha,t) = exp(A(t,T) - C(t,T)\alpha)$$

where A is independent of α and r.

Thus,

$$P(t,T,(r,\alpha)) = exp(A(t,T) - B(t,T)r - C(t,T)\alpha). \tag{8.3.11}$$

This is the exponential-affine form for the price of a zero coupon bond in the two factor Gaussian model, with factors r and α.

8.3.3 Finding $A(t,T)$

In the last subsection, (8.3.11) gives the bond price where the exponential-affine form involves $A(t,T)$. We have specified that A is independent of α and r.

However, we did not explicitly state the form of $A(t,T)$.

The objective of this subsection is to obtain $A(t,T)$. We model the short rate process by $r(\xi_{s,t}(x))$. This is so if $r(x) = R'x + k$ for $R \in \mathbb{R}^n$ and $k \geq 0$.

The price of a zero coupon bond is then

$$P(t,T,x) = E\left[\left(-\int_t^T r(\xi_{t,u}(x))du\right)\middle| x_t = x\right].$$

Hence,

$$\frac{\partial P}{\partial x} = E\left[\left(-R\int_t^T D_{s,u}du\right) exp\left(-\int_t^T r(\xi_{t,u}(x))du\right)\right].$$

Write $R'\int_t^T D_{s,u}du := B(t,T) \in \mathbb{R}^n$.

We then have

$$\frac{\partial P}{\partial x} = -B(t,T)P$$

where $B(t,T)$ is deterministic.

Therefore, $P(t,T,x) = exp(A(t,T) - B(t,T)x)$ for some deterministic function $A(t,T)$.

Noting that

$$P(t,T,x) = E\left[exp\left(-\int_0^t r(\xi_{t,v}(x))dv\right)\bigg|\mathcal{F}_t\right]$$

we can express the valuation formula as

$$V(t,T,x) = exp\left(-\int_0^t r(\xi_{t,v}(x))dv\right) P(t,T,x)$$

or

$$V(t,T,x) = E\left[exp\left(-\int_0^T r(\xi_{t,v}(x))dv\right)\bigg|\mathcal{F}_t\right]. \qquad (8.3.12)$$

Equation (8.3.12) implies that $V(t,T,x)$ is a martingale.

Since the V process is a martingale the drift term must be zero. Using Itô's rule, we therefore have

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}(\gamma + A\xi + b) + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 V}{\partial x_i \partial x_j}\sum_{k=1}^{n}\sigma_{ik}\sigma_{jk} = 0$$

using (8.3.12).

By equation (8.3.12), we obtain

$$exp\left(-\int_0^t r(\xi_{t,v}(x))dv\right)\left[\frac{\partial P}{\partial t} - rP + \frac{\partial P}{\partial x}(\gamma + A\xi + b) + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 P}{\partial x_i \partial x_j}\sum_{k=1}^{n}\sigma_{ik}\sigma_{jk}\right] = 0 \qquad (8.3.13)$$

Therefore, the second term of the product in (8.3.13) above must be zero. In other words, the bond price satisfies a partial differential equation given by

$$\frac{\partial P}{\partial t} - rP + \frac{\partial P}{\partial x}(\gamma + A\xi + b) + \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2 P}{\partial x_i \partial x_j}\sum_{k=1}^{n}\sigma_{ik}\sigma_{jk} = 0,$$

with terminal condition $P(T,T,x) = 1$.

We have seen that $P(t,T,x)$ has the form $exp(A(t,T) - B(t,T)x)$. With $x = 0$, we see that $\lambda(t) = exp(A(t,T))$ satisfies the ordinary differential equation

$$\frac{\partial A}{\partial t} - k - B(\gamma + b) + \frac{1}{2}\sum_{i,j=1}^{n} B_i B_j \sum_{k=1}^{n} \sigma_{ik}\sigma_{jk} = 0, \tag{8.3.14}$$

and $A(T, T) = 0$.

Applying (8.3.14) to our two-factor Gaussian model, we obtain

$$\frac{\partial A}{\partial t} - B\theta_t + \frac{1}{2}({\sigma_1}^2 B^2 + {\sigma_2}^2 C^2 + 2\sigma_1\sigma_2 BC) = 0, \tag{8.3.15}$$

with $A(T, T) = 0$.

Solving the ODE given by (8.3.15) gives an expression for $A(t, T)$.

8.3.4 Reconciling the Short-Term Rate and HJM Forms

We aim to get the dynamics of the forward rate $f(t, T)$ and show that $f(t, t)$ and r_t are equal under this model.

Now,

$$\begin{aligned} f(t,T) &= \frac{-\partial}{\partial T}\ln P(t,T,x) \\ &= \frac{\partial}{\partial T}[B(t,T)r + C(t,T)\alpha - A(t,T)]. \end{aligned} \tag{8.3.16}$$

The stochastic dynamics of $f(t, T)$ is therefore

$$\begin{aligned} d_t f(t,T) &= \frac{\partial}{\partial t}\left(\frac{\partial}{\partial T}B(t,T)\right)r_t + \frac{\partial}{\partial T}B(t,T)dr_t \\ &+ \frac{\partial}{\partial t}\left(\frac{\partial}{\partial T}C(t,T)\right)\alpha_t + \frac{\partial}{\partial T}C(t,T)d\alpha_t \\ &- \frac{\partial}{\partial t}A(t,T) + \frac{1}{2}\left[\frac{\partial}{\partial T}B(t,T)\frac{\partial^2}{\partial r^2}r_t \right. \\ &\left. + \frac{\partial}{\partial T}C(t,T)\frac{\partial^2}{\partial \alpha^2}\alpha\right], \text{using Itô's Formula.} \end{aligned}$$

Separating the drift and volatility terms, we have

$$
\begin{aligned}
d_t f(t,T) &= \left[\frac{\partial}{\partial t}\left(\frac{\partial}{\partial T}B(t,T)\right) r_t + \frac{\partial}{\partial t}\left(\frac{\partial}{\partial T}C(t,T)\right)\alpha_t \right.\\
&\quad -\frac{\partial}{\partial t}\left(\frac{\partial}{\partial T}A(t,T)\right) + (\alpha_t - a r_t)\frac{\partial}{\partial T}B(t,T)\\
&\quad \left. +\frac{\partial}{\partial T}C(t,T)(c - b\alpha_t)\right] dt\\
&\quad \left[\frac{\partial}{\partial T}B(t,T)\sigma_1 dW_t^1 + \frac{\partial}{\partial T}C(t,T)\sigma_2 dW_t^2\right]. \qquad (8.3.17)
\end{aligned}
$$

where the deterministic functions $A(t,T)$, $B(t,T)$ and $C(t,T)$ are given in the previous subsections.

At this point, we consider the derivatives of the functions A, B and C with respect to T and evaluate these partial derivatives at $T = t$.

$$
\begin{aligned}
& B(t,T) = \frac{1}{a}(1 - e^{-a(T-t)})\\
& C(t,T) = \frac{-1}{a(a-b)}e^{-a(T-t)} + \frac{1}{b(a-b)}e^{-b(T-t)} - \frac{1}{ab}\\
& \frac{\partial}{\partial T}B(t,T) = e^{-a(T-t)}\\
\Rightarrow \quad & \left.\frac{\partial}{\partial T}B(t,T)\right|_{T=t} = 1. \qquad (8.3.18)
\end{aligned}
$$

Also,

$$
\begin{aligned}
\frac{\partial}{\partial T}C(t,T) &= \frac{e^{-a(T-t)}}{a-b} - \frac{e^{-b(T-t)}}{a-b}\\
\Rightarrow \left.\frac{\partial}{\partial T}C(t,T)\right|_{T=t} &= \frac{1}{a-b} - \frac{1}{a-b} = 0. \qquad (8.3.19)
\end{aligned}
$$

Equation (8.3.15) gives us the ODE

$$\frac{\partial A}{\partial t} + \frac{1}{2}(\sigma_1^2 B^2 + \sigma_2^2 C^2) = 0,$$

with $A(T,T) = 0$. Or,

$$\frac{\partial A}{\partial t} = -\frac{1}{2}(\sigma_1^2 B^2 + \sigma_2^2 C^2). \qquad (8.3.20)$$

Let $v = T - t$. Then

$$\frac{\partial A}{\partial t} = -\frac{\partial A}{\partial v}. \tag{8.3.21}$$

Also, $\frac{\partial A}{\partial T} = \frac{\partial A}{\partial v}$. But from (8.3.21), $\frac{\partial A}{\partial v} = -\frac{\partial A}{\partial t}$. Therefore, $\frac{\partial A}{\partial T} = -\frac{\partial A}{\partial t}$. Going back to (8.3.20) and noting that $\frac{\partial A}{\partial T} = -\frac{\partial A}{\partial t}$, we have

$$\frac{\partial A}{\partial T} = \frac{1}{2}(\sigma_1^2 B^2(t,T) + \sigma_2^2 C^2(t,T)).$$

Thus,

$$\begin{aligned}
\left.\frac{\partial A}{\partial T}(t,T)\right|_{T=t} &= \frac{1}{2}\left[\sigma_1^2 \left(B(t,T)|_{T=t}\right)^2 + \sigma_2^2 \left(C(t,T)|_{T=t}\right)^2\right] \qquad (8.3.22) \\
\left(B(t,T)|_{T=t}\right)^2 &= \left[\frac{1}{a}(1 - e^{-a(T-t)})\right]^2 = \left[\frac{1}{a}0\right]^2 = 0 \\
\left(C(t,T)|_{T=t}\right)^2 &= \left(\frac{-1}{a(a-b)} + \frac{1}{b(a-b)} - \frac{1}{ab}\right)^2 \\
&= \left(\frac{-b+a}{ab(a-b)} - \frac{1}{ab}\right)^2 = 0.
\end{aligned}$$

Equation (8.3.22) therefore becomes

$$\left.\frac{\partial}{\partial T} A(t,T)\right|_{T=t} = 0.$$

Using Equation (8.3.17), we have

$$
\begin{aligned}
df(t,T)|_{T=t} &= df(t,t) = \left[\frac{\partial}{\partial t}\left(\frac{\partial}{\partial T} B(t,T)\Big|_{T=t} \right) r_t \right. \\
&\quad + \frac{\partial}{\partial t}\left(\frac{\partial}{\partial T} C(t,T)\Big|_{T=t} \right) \alpha_t - \frac{\partial}{\partial t}\left(\frac{\partial}{\partial T}\Big|_{T=t} A(t,T) \right) \\
&\quad + (\alpha_t - a r_t)\left(\frac{\partial}{\partial T} B(t,T)\Big|_{T=t} \right) \\
&\quad \left. + \left(\frac{\partial}{\partial T} C(t,T)\Big|_{T=t} (c - b\alpha_t) \right) \right] dt \\
&\quad + \left[\frac{\partial}{\partial T} B(t,T)\Big|_{T=t} \sigma_1 dW_t^1 + \frac{\partial}{\partial T} C(t,T)\Big|_{T=t} \sigma_2 dW_t^2 \right] \\
&= \left[\frac{\partial}{\partial t}(1) \cdot r_t + \frac{\partial}{\partial t}(0) \cdot \alpha_t \right. \\
&\quad \left. - \frac{\partial}{\partial t}(0) + (\alpha_t - a r_t) \cdot (1) + 0 \cdot (c - b\alpha_t) \right] dt \\
&\quad + [1 \cdot \sigma_1 dW_t^1 + 0\sigma_2 \cdot dW_t^2], \\
&\quad \text{using Equations (8.3.18), (8.3.19) and (8.3.22).} \\
&= (\alpha_t - a r_t) dt + \sigma_1 dW_t^1 = dr_t.
\end{aligned}
$$

Hence, we have shown that $df(t,t) = dr_t$.

8.4 A Model with a Markovian Mean Reverting Level

We follow the motivation in [39], in which a Hidden Markov Model with mean reverting characteristics is considered as a model for financial time series, particularly interest rates. Hidden Markov filtering offers a powerful methodology to estimate efficiently the parameters for such a model.

Our first objective here however, is to derive the zero-coupon bond price under the Vasicek's model with the additional assumption that the mean reversion level α changes according to a continuous time finite state Markov chain.

This could be a model for the logarithm of an asset price, or in our case, an interest rate where a central bank provides a reference rate that changes from time to time. A closed form solution for $P(t,T)$ under this model would enable us to specify the dynamics of the forward rate which is the ultimate aim of this thesis.

8.4.1 The Markov Model for the Reference Level

Suppose that the reference level for the interest rate $\alpha = \{\alpha_t : 0 \leq t \leq T\}$ is a finite state continuous time Markov chain, where $T > 0$ is a finite time horizon.

Modifying the Vasicek Model, let the interest $r = \{r_t : 0 \leq t \leq T\}$ be described by the stochastic differential equation,

$$dr_t = \big(\alpha(t) - \beta(t) r_t\big) dt + \sigma d\tilde{W}_t. \tag{8.4.23}$$

Here, $W = \{W_t : 0 \leq t \leq T\}$ is a Wiener process independent of α_t, and $beta(t)$ and σ are positive constants. We consider the situation where the process r is observed and inferences are to be made about the process α and other parameters.

We assume that there is an underlying probability space $(\Omega, \mathcal{F}, P)$. As an adaptation to Chapter 4, we consider an $n-$state continuous time Markov chain $X = \{X_t : 0 \leq t \leq T\}$ that is identical to α after a transformation of the state space.

Choose the state space for X the set $\{e_1, ..., e_n\}$ of unit vectors in $\mathbb{R}^n$. That is, $e_i = \{0, ..., 1, ..., 0\}$, or the $i-$th component is 1, and zero otherwise. Then we can write

$$\alpha_t = \alpha(X_t) = \langle \alpha, X_t \rangle \tag{8.4.24}$$

for an appropriate vector $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$, where $\langle \alpha, X_t \rangle$ denotes the inner product of the vector α and X_t.

As in Chapter 4, we have the vector of probabilitie $p_t = E[X_t]$. Let $A = \{A_t : 0 \leq t \leq T\}$ be the family of transition intensity matrices associated with the continuous time Markov chain X, so that p_t satisfies the forward equation $\frac{dp_t}{dt} = A_t p_t$, with given initial probability vector p_0.

The transition intensity matrix A_t determines the dynamics of the reference level α as described in (8.4.24). Then, this is introduced into the interest rate model of (8.4.23) as a mean reversion level.

Hence, the interest rate is conditionally Gaussian, conditioned on the independent path of the Markov chain that describes the reference level.

8.4.2 Deriving the Bond Price $P(t,T)$

Consider the Hull-White Model for which the interest rate process is given by

$$dr_t = \big(\alpha(t) - \beta(t)r_t\big)dt + \sigma(t)dW_t \tag{8.4.25}$$

for $r_0 \geq 0$.

We take the case where $\beta(t) = a$, i.e., $\beta(t)$ is constant and $\sigma(t) = \sigma$. However, $\alpha(t)$ follows a Markov chain. That is,

$$\alpha_t = \alpha(X_t) = \langle \alpha, X_t \rangle$$

for an appropriate vector $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$ and X_t follows the stochastic dynamics $dX_t = AX_t dt + dM_t$ and $X_t \in \{e_1, ..., e_n\}$ as described in the preceding subsection.

Now, under this given model

$$P(t,T) = E\left[exp\left(-\int_t^T r_s ds\right)\middle| \mathcal{F}_t, \quad \alpha_u, \quad u \leq T\right].$$

So, if we know the trajectory of $\alpha_u, \quad u \leq T$, then (by the Hull-White Model)

$$P(t,T) = exp\big(-r_t C(t,T) - A(t,T)\big) \tag{8.4.26}$$

where r_t is the solution to the SDE described in (8.4.25) and

$$\begin{aligned} C(t,T) &= e^{b(t)} \int_t^T e^{-b(u)} du = e^{b(t)} \gamma(t) \\ b(t) &= \int_0^t \beta(u) du \end{aligned}$$

$$\begin{aligned} \gamma(t) &= \int_t^T e^{-b(u)} du \\ A(t,T) &= \int_t^T \left[e^{b(u)} \alpha(u) \gamma(u) - \frac{1}{2} e^{2b(u)} \sigma^2(u) \gamma^2(u) \right] du \\ r_t &= e^{-b(t)} \left(r_0 + \int_0^t e^{b(u)} \alpha(u) du + \int_0^t b(u) \sigma(u) dW_u \right). \quad (8.4.27) \end{aligned}$$

With $\beta(t) = a$ and $\sigma(t) = \sigma \quad \forall\, t$, equation (8.4.27) simplifies to

$$r_t = e^{-at} \left(r_0 + \int_0^t e^{au} \langle \alpha, X_u \rangle du + \int_0^t au\sigma dW_u \right).$$

Write $\xi_u := e^{au} \alpha$. Then,

$$r_t = e^{-at} \left(r_0 + \int_0^t \langle \xi_u, X_u \rangle du + \int_0^t a\sigma u dW_u \right).$$

In the succeeding discussion, we shall discuss the random variable $\int_0^t \langle \xi_u, X_u \rangle du$.

We first note that the deterministic function $C(t,T)$ can be obtained as follows:

$$b(t) = \int_0^t a du = at$$

$$\gamma(t) = \int_t^T e^{-au} du = \frac{1}{a}(e^{-at} - e^{-aT}).$$

Thus,

$$\begin{aligned} C(t,T) &= e^{b(t)} \int_t^T e^{-b(u)} du \\ &= e^{at} \int_t^T e^{-au} du \\ &= \frac{e^{at}}{a}(e^{-at} - e^{-aT}). \end{aligned}$$

For the function $A(t,T)$, we have

$$\begin{aligned}
A(t,T) &= \int_t^T \left[e^{b(u)}\alpha(u)\gamma(u) - \frac{1}{2}e^{2b(u)}\sigma^2(u)\gamma^2(u)\right] du \\
&= \int_t^T e^{au}\alpha(u)\frac{1}{a}(e^{-au} - e^{-aT})du \\
&\quad -\frac{\sigma^2}{2}\int_t^T e^{2au}\frac{1}{a^2}(e^{-au} - e^{-aT})^2 du \\
&= \int_t^T e^{au}\alpha(u)\frac{1}{a}(e^{-au} - e^{-aT})du \\
&\quad -\frac{1}{2}\left(\frac{\sigma}{a}\right)^2\left[(T-t) - \frac{3}{2a} + \frac{2}{a}e^{-a(T-t)}\right. \\
&\quad \left. - \frac{1}{2a}e^{-2a(T-t)}\right].
\end{aligned} \tag{8.4.28}$$

Let us evaluate the integral term of (8.4.28).

$$\begin{aligned}
&\int_t^T e^{au}\alpha_u\frac{1}{a}(e^{-au} - e^{-aT})du \\
&= \frac{1}{a}\left[\int_t^T \alpha_u du\right] - \frac{e^{-aT}}{a}\left[\int_t^T \alpha_u e^{au} du\right] \\
&= \frac{1}{a}\left(\int_t^T \alpha_u(1 - e^{-aT}e^{au})du\right) \\
&= \int_t^T \langle X_u, \alpha_u\rangle\left(\frac{1 - e^{-a(T-u)}}{a}\right) du \\
&= \int_t^T \langle X_u, \phi_u\rangle du,
\end{aligned} \tag{8.4.29}$$

where

$$\phi_u = \frac{1 - e^{-a(T-u)}}{a} \cdot \alpha_u.$$

We set aside (8.4.29) first and remember that what we aim to get is a closed form solution for (8.4.26).

So far, we have

$$\begin{aligned} P(t,T) &= exp(-r_t C(t,T) - A(t,T)) \\ &= exp(-r_t C(t,T)) \cdot exp\left(-\int_t^T \langle X_u, \phi_u\rangle du\right) \\ &\quad \cdot exp(G(t,T)) \end{aligned} \tag{8.4.30}$$

using equations (8.4.29) and (8.4.28) and where

$$\begin{aligned} G(t,T) &= \frac{1}{2}\left(\frac{\sigma}{a}\right)^2 \left[(T-t) - \frac{3}{2a} + \frac{2}{a}e^{-a(T-t)} \right. \\ &\quad \left. -\frac{1}{2}e^{-2a(T-t)}\right] \quad \text{and} \\ C(t,T) &= \frac{e^{at}}{a}(e^{-at} - e^{-aT}). \end{aligned}$$

Thus all that remains to be done is the evaluation of

$$e^{\left(-\int_t^T \langle X_u, \phi_u\rangle du\right)}$$

where ϕ_u is deterministic for $u \leq T$.

Define $\Lambda_{t,u} := exp\left(-\int_t^u \langle X_v, \phi_v\rangle dv\right)$.

Thus,

$$d\Lambda_{t,u} = -\langle X_u, \phi_u\rangle \Lambda_{t,u} du.$$

Further, if we consider the vector process ΛX : we obtain,

$$\begin{aligned} d(\Lambda_{t,v}X_{t,v}) &= \Lambda_{t,v}dX_{t,v} + X_{t,v}d\Lambda_{t,v} \\ &= \Lambda_{t,v}[AX_{t,v}dv + dM_v] \\ &\quad + X_{t,v}[-\langle X_v, \phi_u\rangle \Lambda_{t,v}dv]. \end{aligned}$$

Or in integral form,

$$\begin{aligned} \Lambda_{t,T}X_{t,T} &= X_t + \int_t^T \Lambda_{t,v}AX_{t,v}dv + \int_t^T \Lambda_{t,v}dM_v \\ &\quad - \int_t^T \langle X_v, \phi_v\rangle \Lambda_{t,v}X_{t,v}dv. \end{aligned}$$

Taking expectations:

$$\begin{aligned} E_t[\Lambda_{t,T}X_{t,T}] &= X_t + \int_t^T AE_t[\Lambda_{t,v}X_{t,v}]dv + \int_t^T \Lambda_{t,v}dM_v \\ &\quad - \int_t^T E_t[\langle X_v, \phi_v\rangle \Lambda_{t,v}X_{t,v}]dv. \end{aligned} \tag{8.4.31}$$

We note further that

$$\begin{aligned} \langle \phi_v, X_v\rangle \Lambda_{t,v}X_{t,v} &= \sum_{i=1}^{n} \langle X_{t,v}, e_i\rangle s_i(v) e_i \\ &= S(u)\Lambda_{t,v}X_{t,v}, \end{aligned}$$

where $S(u)$ is a time varying matrix with $s(u) = \big(s_1(u), s_2(u), ..., s_n(u)\big)'$ on the diagonal.

Write $\hat{z}_{t,T} := E_t[\Lambda_{t,v}X_{t,v}]$.

So,

$$\hat{z}_{t,T} = X_t + \int_t^T A\hat{z}_{t,v}dv - \int_t^T S(v)\hat{z}_{t,v}dv.$$

That is,

$$\hat{z}_{t,T} = X_t + \int_t^T \big(A - S(v)\big)\hat{z}_{t,v}dv.$$

We wish to find $\hat{z}_{t,T}$ such that

$$\hat{z}_{t,T} = X_t + \int_t^T H(v)\hat{z}_{t,v}dv \tag{8.4.32}$$

where $H(v) = A - S(v)$.

Equivalently, we would like to solve

$$\frac{d}{du}\hat{z}_{t,u} = H(u)\hat{z}_{t,u} \tag{8.4.33}$$

and $\hat{z}_{t,t} = X_t$.

Furthermore, in connection with (8.4.30), we see that

$$E\left[exp\left(-\int_t^T \langle X_v, \phi_v\rangle dv\right)\Big| \mathfrak{F}_t\right]$$

$$= E\left[exp\left(-\int_t^T \langle X_v, \phi_v\rangle dv\right)\langle X_T, \mathbf{1}\rangle \Big| \mathfrak{F}_t\right].$$

since $\langle X_T, \mathbf{1}\rangle = 1$

$$= \left\langle E\left[exp\left(-\int_t^T \langle X_v, \phi_v\rangle dv\right) X_T \Big| \mathfrak{F}_t\right], \mathbf{1}\right\rangle$$

$$= \langle \hat{z}_{t,T}, \mathbf{1}\rangle.$$

Therefore, we wish to obtain the solution of (8.4.33).

8.4.3 The Fundamental Matrix Solution

If $S(t)$ is a matrix satisfying certain conditions the matrix differential equation

$$\dot{\Phi}(t) = S(t)\Phi(t), \quad \Phi(0) = I$$

has a unique solution (see [57], for example) defined for $0 \leq t < \infty$.

Here I is the $n \times n$ matrix. For each $t \geq 0$, the matrix $\Phi(t)$ is nonsingular.

Suppose further we have the $n-$dimensional vector ξ, $n \times n$ matrix $S(t)$ and a deterministic equation

$$\dot{\xi}(t) = S(t)\xi, \quad \xi(0) = \xi. \tag{8.4.34}$$

In terms of Φ, the solution of the deterministic equation (8.4.34) is just

$$\xi(t) = \Phi(t)\xi(0).$$

Applied to our current investigation, $\xi(t,v) = \Phi(t,v)\xi(t)$, $S(t) = H(t)$ and $\xi(t) = \hat{z}_{t,t} = X_t$. And $\Phi(t,v)$ is the solution to $\dot{\Phi}(t,v) = H(v)\Phi(t,v)$.

Thus, $\hat{z}_{t,T} = \Phi(t,T)X_t$.

Finally,

$$E\left[exp\left(-\int_t^T \langle X_v, \phi_v\rangle dv\right)\bigg|\mathcal{F}_t\right] = \langle \hat{z}_{t,T}, \mathbf{1}\rangle = \langle \Phi(t,T)X_t, \mathbf{1}\rangle. \qquad (8.4.35)$$

The zero-coupon bond price is therefore

$$P(t,T) = exp[-r_t C(t,T) + G(t,T)]\langle \Phi(t,T)X_t, \mathbf{1}\rangle,$$

where

$$\begin{aligned} C(t,T) &= \frac{e^{at}}{a}(e^{-at} - e^{-aT}) \\ G(t,T) &= \frac{1}{2}\left(\frac{\sigma}{a}\right)^2 \left[(T-t) - \frac{3}{2a} + \frac{2}{a}e^{-a(T-t)} \right. \\ &\quad \left. -\frac{1}{2a}e^{-2a(T-t)}\right]. \end{aligned}$$

8.4.4 Reconciling the Short-Term Rate and $f(t,t)$

Under the current bond price, the forward rate is given by

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}\ln P(t,T) \\ &= \frac{\partial}{\partial T}\Big([r_t C(t,T) - G(t,T)] - \ln\langle \Phi(t,T)X_t, \mathbf{1}\rangle\Big). \end{aligned}$$

Thus,

$$\begin{aligned} f(t,T) &= r_t\frac{\partial}{\partial T}C(t,T) \\ &\quad -\frac{\partial}{\partial T}G(t,T) - \frac{\partial}{\partial T}\ln\langle \Phi(t,T)X_t, \mathbf{1}\rangle. \end{aligned}$$

We must therefore show that

$$\begin{aligned} &\frac{\partial}{\partial T}C(t,T)\bigg|_{T=t} = 1, \\ &\frac{\partial}{\partial T}G(t,T)\bigg|_{T=t} = 0 \quad \text{and} \\ &\frac{\partial}{\partial T}\ln\langle \Phi(t,T)X_t, \mathbf{1}\rangle\bigg|_{T=t} = 0. \end{aligned}$$

Now,

$$C(t,T) = \frac{1-e^{-a(T-t)}}{a} \quad \text{and}$$
$$\frac{\partial}{\partial T}C(t,T) = e^{-a(T-t)}.$$

Clearly,

$$\left[\frac{\partial}{\partial T}C(t,T)\right]_{T=t} = e^{-a(T-t)}|_{T=t} = e^{-a\cdot(0)} = 1.$$

Also,

$$\begin{aligned} G(t,T) &= \frac{1}{2}\left(\frac{\sigma^2}{a}\right)^2\left[(T-t) - \frac{3}{2a}\right. \\ &\quad \left. + \frac{2}{a}e^{-a(T-t)} - \frac{1}{2}e^{-2a(T-t)}\right]. \end{aligned}$$

Hence,

$$\frac{\partial}{\partial T}G(t,T) = \frac{1}{2}\left(\frac{\sigma}{a}\right)^2\left(1 - 2e^{-a(T-t)} + e^{-2a(T-t)}\right).$$

Again, we see that

$$\left[\frac{\partial}{\partial T}G(t,T)\right]_{T=t} = \frac{1}{2}\left(\frac{\sigma}{a}\right)^2(1-2+1) = 0. \tag{8.4.36}$$

Finally, we consider

$$\left.\frac{\partial}{\partial T}\ln\langle\Phi(t,T)X_t, \mathbf{1}\rangle\right|_{T=t}.$$

$$\begin{aligned} \frac{\partial}{\partial T}\ln\langle\Phi(t,T)X_t, \mathbf{1}\rangle &= \frac{\partial}{\partial T}\ln\left[exp\left(-\int_t^T\langle X_{t,v}, \phi_{t,v}\rangle dv\right)\right] \\ &= \frac{\partial}{\partial T}\left(-\int_t^T\langle X_{t,v}, \phi_{t,v}\rangle\right)dv \\ &= -\langle X_{t,T}, \phi_{t,T}\rangle \\ &\quad \text{where} \\ \phi_{t,T} &= \frac{1-e^{-a(T-t)}}{a}\cdot\alpha \quad \text{from Equation (8.4.29)} \end{aligned}$$

Therefore,

$$\frac{\partial}{\partial T}\ln\langle\Phi(t,T)X_t,\mathbf{1}\rangle\bigg|_{T=t} = \langle X_{t,T},\phi_{t,T}\rangle|_{T=t} = -\langle X_{t,t},\mathbf{0}\rangle = 0.$$

Consequently, we have proven that $f(t,t) = r_t$ under this mean reverting model.

Chapter 9

Empirical Test with Hidden Markov Models and Filtering

9.1 The Method of Filtering and the Efficient Market Hypothesis (EMH)

In the filtering method that we shall perform, historical and publicly current available information are used to calculate optimal filters in the estimation procedure of parameters.

We start with a unit-delay model to introduce the filtering problem of Hidden Markov Models (HMM). Under this setting, the filters at time k, are calculated based on the information available up to time $k-1$, hence the name unit-delay model. This is reasonable because asset prices do not react immediately but instead take a unit time step to adjust to whatever available information such as corporate announcements and governmental policies.

On the other hand, as the financial world adapts itself into electronic information networks, it is also worth considering zero-delay models. That is, the filters at time t are obtained using all available data up to time t. Again, we can argue that this is a sensible model because, we live in a technologically advanced world of computers where information can be transmitted in just a split of a second. Thus, prices adjust almost instantaneously.

Here, we observe that the filtering techniques employed to estimate optimal parameters of a model is consistent with the body of theory called Efficient Market Hypotheses (EMH). Financial theorists generally define three forms or levels[1] of capital market efficiency. These are described as (i) weak-form efficiency, (ii) semi-strong form efficiency and (iii) strong form.

The *weak form* of the EMH states that all information contained in past price movements is fully reflected in current market prices. *Semi-strong efficiency* argues that existing prices reflect all public information, good or bad. All the information currently known to the market is already impounded in current market prices. Except for the predictable upward drift which constitutes part of the normal return on a security, prices change only when new information arrives. The *strong form* of the EMH states that current market prices reflect all pertinent information, whether publicly available or privately held. If this form holds, even insiders would find it impossible to earn abnormal returns in the market.

Empirical studies conducted suggest that the market is indeed highly efficient in the weak form and reasonably efficient in the semistrong.[2] However, the strong-form EMH does not hold,[3] so abnormal profits can be made by those

[1] E.F. Fama, "The Behaviour of Stock Prices," *Journal of Business* 38 (January 1965), pp.34-105, "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* 25 (May 1970), pp.383-417; and *Foundations of Finance* (New York: Basic Books, 1976)

[2] E.F. Fama, L. Fischer, M. Jensen and R.Roll, "The Adjustment of Stock Prices to New Information," *International Economic Review* 10 (February 1969), pp. 1-21; M. Jensen, "The Performance of Mutual Funds in the Period 1954-64," *Journal of Finance* 23 (May 1968), pp. 389-416; R.S. Kaplan and R. Roll, "Investor Evaluation of Accounting Information: Some Empirical Evidence," *Journal of Business* 45 (April 1972), pp. 225-257; and M.S. Scholes, "Market for Securities: Substitution versus Price Pressure and the Effects of Information on Share Prices," *Journal of Business* 45 (April 1972), pp. 179-211.

[3] J.E. Finnerty, "Insiders and Market Efficiency," *Journal of Finance* 31 (September 1976), pp. 1141-1148; R.G. Ibbotson, "Price Performance of Common Stock New Issues," *Journal of Financial Economics* 2 (September 1975), pp. 235-272; and J.F. Jaffe, "The Effect of Regulation Changes on Insider Trading," *Bell Journal of Economics and Management Science*

who possess inside information.

The unit-delay model is explored here because of its consistency to the weak-form of EMH while the zero-delay model which will be implemented to the Markovian mean-reverting interest rate model is consistent with the semi-strong form of EMH.

9.2 The Markov Model and the Filtering Problem

In the process of testing empirically the model we proposed for the mean reversion level, we shall discuss the underlying assumptions, features and descriptions of a Markov Model. We begin with discrete time Markov chain and illustrate the basic idea of Hidden Markov Model (HMM) filtering. Calculation of recursive filters is demonstrated by considering continuous observations in a discrete time.

Having presented a theoretical framework of HMM filtering, we adopt its relevant theory and implement the techniques to the Markovian mean reverting model.

9.2.1 Discrete Time Markov Chains

We start with a process X with time parameter set $\{0, 1, 2, ...\}$ defined on $(\Omega, \mathcal{F}, P)$. As usual, X has a general finite state space $S = \{s_1, s_2, ..., s_n\}$.

As in Chapter 4, we can assume that $S = \{e_1, e_2, ..., e_n\}$. That is, the elements of S are identified with the standard unit vectors where

$$e_i = \{0, ..., 0, 1, 0, ..., 0\}' \in \mathbb{R}^n.$$

Write $\mathcal{F}_k = \sigma\{X_0, ..., X_k\}$ for the $\sigma-$ field generated by $X_0, ..., X_k$. Thus,

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset ... \subset \mathcal{F}_k$$

5 (Spring 1974), pp. 93-121

and $\{\mathcal{F}_k\}$ is a filtration which models all possible histories of X.

Since we assume that X_t is Markov, we have

$$P(X_{k+1} = e_i | \mathcal{F}_k) = P(X_{k+1} = e_i | X_k).$$

We recall that

$$a_{ji} = P(X_{k+1} = e_j | X_k = e_i) \quad \text{and}$$
$$A = (a_{ji}) \in \mathbb{R}^{n \times n},$$

a_{ji} is the one-step transition probability and A is called the transition matrix of the Markov chain X.

Theorem 9.2.1 *The expected value of X_{k+1} given X_k is completely defined by the transition matrix of the Markov chain and X_k.*

Proof:

$$\begin{aligned} E[X_{k+1}|X_k] &= \sum_{i=1}^{n} E[X_{k+1}|X_k = e_i]\langle X_k, e_i\rangle \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} E[\langle X_{k+1}, e_j\rangle | X_k = e_i]\langle X_k, e_i\rangle e_j \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}\langle X_k, e_i\rangle e_j = AX_k. \end{aligned}$$

■

Suppose further, we define a random variable V_k such that for each k :

$$V_{k+1} = X_{k+1} - AX_k \in \mathbb{R}^n.$$

Or,

$$X_{k+1} = AX_k + V_{k+1}.$$

Note that

$$\begin{aligned} E[V_{k+1}|\mathcal{F}_k] &= E[X_{k+1} - AX_k|X_k] \\ &= E[X_{k+1}|X_k] - AX_k \\ &= \mathbf{0} \in \mathbb{R}^n. \end{aligned}$$

We therefore have proven the following theorem which is a discrete analogue of Corollary 4.2.1.

Theorem 9.2.2 *The semimartingale representation of the Markov chain is*

$$X_{k+1} = AX_k + V_{k+1}.$$

Then, we shall consider X to be a very simple process where X is independently and uniformly distributed over its state space S at each time k.

On a measurable space $(\Omega, \mathcal{F})$, assume the existence of a probability measure $\bar{P}$ such that for every k :

$$\bar{P}(X_{k+1} = e_j | \mathcal{F}_k) = \bar{P}(X_{k+1} = e_j) = \frac{1}{n}.$$

So far, we have a simple process with its probability $\bar{P}$. With this, we wish to construct a new probability P where under this new measure P, X is a Markov chain with transition matrix A. To accomplish this goal, we state the following theorem which gives the form of the Radon-Nikodym derivative which allows the desired change of measure from $\bar{P}$ to P.

Theorem 9.2.3 *Let the new probability measure P be defined by putting*

$$\left.\frac{dP}{d\bar{P}}\right|_{\mathcal{F}_k} = \bar{\Lambda}_k$$

and

$$\bar{\Lambda}_k = \prod_{l=1}^{k} \bar{\lambda}_l$$

such that

$$\bar{\lambda}_l = n \sum_{j=1}^{n} (\langle AX_{l-1}, e_j \rangle \langle X_l, e_j \rangle).$$

Then under P, X is a Markov chain with transition matrix A.

Proof: First we claim that $\bar{E}[\bar{\lambda}_l|\mathcal{F}_{l-1}] = 1$. This is because,

$$\begin{aligned}\bar{E}[\bar{\lambda}_l|\mathcal{F}_{l-1}] &= n\bar{E}\left[\sum_{j=1}^{n}\langle AX_{l-1}, e_j\rangle\langle X_l, e_j\rangle \middle| \mathcal{F}_{l-1}\right] \\ &= n.\frac{1}{n}\sum_{j=1}^{n}\langle AX_{l-1}, e_j\rangle \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\langle X_{l-1}, e_i\rangle a_{ji} = 1.\end{aligned}$$

Now, General Bayes' theorem implies

$$\begin{aligned}P(X_{k+1} = e_j|\mathcal{F}_k) &= E[\langle X_{k+1}, e_j\rangle|\mathcal{F}_k] \\ &= \frac{\bar{E}[\bar{\Lambda}_{k+1}\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]}{\bar{E}[\bar{\Lambda}_{k+1}|\mathcal{F}_k]}.\end{aligned}$$

Since $\bar{\Lambda}_{k+1} = \bar{\Lambda}_k\bar{\lambda}_{k+1}$ and $\bar{\Lambda}_k$ is $\mathcal{F}_k$—measurable, we obtain

$$\begin{aligned}&\frac{\bar{E}[\bar{\lambda}_{k+1}\langle X_{k+1}, e_j\rangle|\mathcal{F}_k]}{\bar{E}[\bar{\lambda}_{k+1}|\mathcal{F}_k]} \\ &= n\cdot\bar{E}[\langle AX_k, e_j\rangle\langle X_{k+1}, e_j\rangle|\mathcal{F}_k] \\ &= \langle AX_k, e_j\rangle = P(X_{k+1} = e_j|X_k)\end{aligned}$$

as this depends only on X_k.

If $X_k = e_i$ we have $P(X_{k+1} = e_j|X_k = e_i) = a_{ji}$ and so, under P, X is a Markov chain with transition matrix A.

■

9.2.2 Hidden Markov Models

In this subsection, we discuss what are Hidden Markov models. Suppose, we do not observe X directly. Nevertheless, there is a function c with values in a finite set and we observe the values

$$Y_{k+1} = c(X_k, w_{k+1}), \quad k = 0, 1, 2, ...$$

Here, $\{w_k\}$ is a sequence of independent, identically distributed (IID) random variables which are independent of X. In other words, we have a situation where the Markov chain X is not observed directly but is hidden in the "noisy" observations Y.

Let the range of c consist of m points in an arbitrary set. These can be identified as before with unit vectors $\{f_1, f_2, ..., f_m\}$ where

$$f_j = (0, ..., 0, 1, 0, ..., 0)' \in \mathbb{R}^m.$$

Previously, we have $\mathcal{F}_k = \sigma\{X_0, ..., X_k\}$. Write

$$\mathcal{Y}_k = \sigma\{Y_1, ..., Y_k\}$$

and

$$\mathcal{G}_k = \sigma\{X_0, ..., X_k, Y_1, ..., Y_k\}.$$

We note that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset ...$ and therefore $\{\mathcal{F}_k\}, \{\mathcal{Y}_k\}, \{\mathcal{G}_k\}$ are increasing families of $\sigma-$fields.

These represent possible histories of the state process X, the observation process Y and the combined process (X, Y).

We also have

$$c_{ji} = P(Y_{k+1} = f_j | X_k = e_i) \quad 1 \leq j \leq m, \quad 1 \leq i \leq n.$$

As before,

$$E[Y_{k+1}|X_k] = CX_k, \quad C = (c_{ji}),$$

and if we define $W_{k+1} = Y_{k+1} - CX_k$, the semimartingale representation of Y is $Y_{k+1} = CX_k + W_{k+1}$ since W is a martingale increment.

Observe that there is a unit delay between X_k and its observation Y_{k+1}. This one-step delay is reasonable as Y may not react immediately to X. However, the alternative zero time-delay model can also be constructed. Such discussion is given in [75].

Similar to the motivation of constructing the Markov chain X, we can construct Y by changing probability measure.

Suppose under some probability measure $\bar{P}$, Y is a process such that

$$\bar{P}(Y_{k+1} = f_j | \mathcal{G}_k) = \bar{P}(Y_{k+1} = f_j) = \frac{1}{m}.$$

Further, under $\bar{P}$, X is a Markov chain which is independent of Y, with state space $S = \{e_1, e_2, ..., e_n\}$ and transition matrix $A = (a_{ji})$.

That is, $X_{k+1} = AX_k + V_{k+1}$ where

$$\begin{aligned} \bar{E}[V_{k+1}|\mathcal{G}_k] &= \bar{E}[V_{k+1}|\mathcal{F}_k] \\ &= \bar{E}[V_{k+1}|X_k] = \mathbf{0} \in \mathbb{R}^n. \end{aligned}$$

For, $1 \leq j \leq m$, $1 \leq i \leq n$ $C = (c_{ji})$ is a matrix with $c_{ji} \geq 0$ and $\sum_{j=1}^{m} c_{ji} = 1$; a similar calculation will yield the following result.

Theorem 9.2.4 *Define*

$$\bar{\lambda}_l = m \sum_{j=1}^{m} (\langle CX_{l-1}, f_j \rangle \langle Y_l, f_j \rangle)$$

and

$$\bar{\Lambda}_k = \prod_{l=1}^{k} \bar{\lambda}_l.$$

A new probability measure P can be defined by putting $\frac{dP}{d\bar{P}}\big|_{\mathcal{G}_k} = \bar{\Lambda}_k$. Then, under P, X remains a Markov chain with transition matrix A and $P(Y_{k+1} = f_j | X_k = e_i) = c_{ji}$. That is, under P

$$X_{k+1} = AX_k + V_{k+1}$$

and

$$Y_{k+1} = CX_k + W_{k+1}.$$

9.2.3 Filtering Problem

Let us suppose that we observe $Y_1, \cdots, Y_k$. We wish to estimate $X_0, X_1, \cdots, X_k$. The best (mean square) estimate of X_k given $\mathcal{Y}_k = \sigma\{Y_1, \cdots, Y_k\}$ is

$$E[X_k|\mathcal{Y}_k] \in \mathbb{R}^n.$$

By the Bayes' Theorem,

$$E[X_k|\mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k]}.$$

where $\bar{E}$ denotes the expectation under $\bar{P}$.

Write $q_k := \bar{E}[\bar{\Lambda}_k X_k|\mathcal{Y}_k] \in \mathbb{R}^n$. We see that q_k is an unnormalised conditional expectation of X_k given the observations $\mathcal{Y}_k$.

Since $X_k = e_i$ for only one i, $1 \leq i \leq n$, $\sum_{i=1}^n \langle X_k, e_i \rangle = 1$.

Consider $\langle q_k, \mathbf{1} \rangle$ with $\mathbf{1} = (1, 1, \cdots, 1)' \in \mathbb{R}^n$.

$$\begin{aligned}
\langle q_k, \mathbf{1} \rangle &= \bar{E}[\langle \bar{\Lambda}_k X_k, \mathbf{1} \rangle|\mathcal{Y}_k] \\
&= \bar{E}[\bar{\Lambda}_k \langle X_k, \mathbf{1} \rangle|\mathcal{Y}_k] \\
&= \bar{E}[\bar{\Lambda}_k \sum_{i=1}^{n} \langle X_k, e_i \rangle|\mathcal{Y}_k] \\
&= \bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k]
\end{aligned}$$

and so

$$E[X_k|\mathcal{Y}_k] = \frac{q_k}{\langle q_k, \mathbf{1} \rangle}.$$

We give a result concerning the dynamics of q and an algorithm of how q is updated as new observation Y_{k+1} arrives.

Theorem 9.2.5 *Write $B(Y_{k+1})$ for the diagonal matrix with entries*

$$m \left(\sum_{j=1}^{m} c_{ji} \langle Y_{k+1}, f_j \rangle \right).$$

Then,

$$q_{k+1} = AB(Y_{k+1}) q_k.$$

Proof: We write

$$\begin{aligned}
q_{k+1} &= \bar{E}[\bar{\Lambda}_{k+1} X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}\left[\bar{\Lambda}_k \left(m \sum_{j=1}^{m} (\langle C X_k, f_j \rangle \langle Y_{k+1}, f_j \rangle) \right) (A X_k + V_{k+1}) \Big| \mathcal{Y}_{k+1} \right].
\end{aligned}$$

The observations $Y_i's$ are IID and independent of X under $\bar{P}$, and X is a Markov chain with transition matrix A. Thus, V_{k+1} is a martingale increment, independent of $\mathcal{Y}_{k+1}$, and therefore

$$\begin{aligned} q_{k+1} &= \bar{E}\Big[\bar{\Lambda}_k\Big(m\sum_{j=1}^{m}\langle CX_k, f_j\rangle\langle Y_{k+1}, f_j\rangle\Big)AX_k|\mathcal{Y}_k\Big] \\ &= m\sum_{i=1}^{m}\bar{E}\Big[\langle X_k, e_i\rangle\bar{\Lambda}_k|\mathcal{Y}_k\Big]\left(\sum_{j=1}^{m}c_{ji}\langle Y_{k+1}, f_j\rangle\right)Ae_i \\ &= m\sum_{i=1}^{m}\langle\bar{E}[\bar{\Lambda}_kX_k|\mathcal{Y}_k], e_i\rangle\left(\sum_{j=1}^{m}c_{ji}\langle Y_{k+1}, f_j\rangle\right)Ae_i \\ &= m\sum_{i=1}^{m}\langle q_k, e_i\rangle\left(\sum_{j=1}^{m}c_{ji}\langle Y_{k+1}, f_j\rangle\right)Ae_i. \end{aligned}$$

If we write $B(Y_{k+1})$ for the diagonal matrix with entries $m\left(\sum_{j=1}^{m}c_{ji}\langle Y_{k+1}, f_j\rangle\right)$, then we see that $q_{k+1} = AB(Y_{k+1})q_k$.

■

9.3 Continuous Observations of a Markov Chain

9.3.1 The Model and its Characteristics

We shall assume continuously valued observations y of a finite Markov chain X with a discrete time parameter. The formulation of this model will serve as basis in the implementation of filtering techniques to the Markovian mean reverting level.

The Markov process X could represent the state of the economy: good, average, bad. In our case the observations are the interest rates, which we shall take as the returns from $T-$bills or $T-$bonds.

Under the real world probability P, the Markov chain X, has the dynamics

$$X_{k+1} = AX_k + V_{k+1}.$$

Again, we note that X is not observed directly. We suppose that there is a real valued process y such that

$$y_{k+1} = c(X_k) + \sigma(X_k)w_{k+1}.$$

The sequence $w_1, w_2, \cdots$, is an IID $N(0,1)$ random variable. Thus,

$$P(w_k \leq \alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{\frac{-x^2}{2}} dx.$$

Here we assume that there are vectors, $c = (c_1, \cdots, c_n)'$ and $\sigma = (\sigma_1, \cdots, \sigma_n)'$ such that

$$c(X_k) = \langle c, X_k \rangle \quad \text{and}$$

$$\sigma(X_k) = \langle \sigma, X_k \rangle, \quad \text{with} \quad \sigma_i > 0 \quad \text{for} \quad 1 \leq i \leq n.$$

9.3.2 Construction of a Reference Probability

We shall be working under a reference probability $\bar{P}$. Under $\bar{P}$, we choose X to be a Markov chain with transition matrix A. That is,

$$X_{k+1} = AX_k + V_{k+1}$$

where $\bar{E}[V_{k+1}|\mathcal{G}_k] = \mathbf{0} \in \mathbb{R}^n$. Also, under $\bar{P}$, the observed values $y_1, y_2, \cdots$, form a sequence of IID random variables each of which is $N(0,1)$. This is to say that,

$$\bar{P}(y_k \leq \alpha | \mathcal{G}_{k-1}) = \bar{P}(y_k \leq \alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{x^2}{2}} dx.$$

The following discussion outlines how to construct the real world probability P from $\bar{P}$.

Write $\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. That is, $X \sim N(0,1)$. For $l = 1, 2, \cdots$, write

$$\lambda_l = \frac{\phi(\langle \sigma, X_{l-1} \rangle^{-1}(y_l - \langle c, X_{l-1} \rangle))}{\langle \sigma, X_{l-1} \rangle \phi(y_l)},$$

$$\bar{\Lambda}_0 = 1 \quad \text{and} \quad \bar{\Lambda}_k = \prod_{l=1}^{k} \bar{\lambda}_l, \quad k \geq 1.$$

Definition 9.3.1 *A probability P is defined by setting $\frac{dP}{d\bar{P}}\big|_{\mathcal{G}_k} = \bar{\Lambda}_k$.*

Lemma 9.3.1 *Under P the random variables $w_1, w_2, \cdots$, form a sequence of IID $N(0,1)$ random variables where*

$$w_{k+1} := \langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle).$$

Proof: Let I be the indicator function. Then,

$$P(w_{k+1} \le \alpha | \mathcal{G}_k) = E[I(w_{k+1} \le \alpha)|\mathcal{G}_k].$$

The General Bayes' theorem can be employed to get

$$\begin{aligned} P(w_{k+1} \le \alpha | \mathcal{G}_k) &= \frac{\bar{E}[\bar{\Lambda}_{k+1} I(w_{k+1} \le \alpha)|\mathcal{G}_k]}{\bar{E}[\bar{\Lambda}_{k+1}|\mathcal{G}_k]} \\ &= \frac{\bar{E}[\bar{\lambda}_{k+1} I(w_{k+1} \le \alpha)|\mathcal{G}_k]}{\bar{E}[\bar{\lambda}_{k+1}|\mathcal{G}_k]}. \end{aligned}$$

Consider the denominator:

$$\begin{aligned} \bar{E}[\bar{\lambda}_{k+1}|\mathcal{G}_k] &= \bar{E}\left[\frac{\phi(\langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle))}{\langle \sigma, X_k \rangle \phi(y_{k+1})} \middle| \mathcal{G}_k \right] \\ &= \int_{-\infty}^{\infty} \frac{\phi(\langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle))}{\langle \sigma, X_k \rangle \phi(y_{k+1})} \phi(y_{k+1}) dy_{k+1}. \end{aligned}$$

Write $w := \langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle)$ which is equivalent to $\int_{-\infty}^{\infty} \phi(w)dw = 1$.

For the numerator,

$$\begin{aligned} \bar{E}[\bar{\lambda}_{k+1} I(w_{k+1} \le \alpha)|\mathcal{G}_k] &= \bar{E}\left[\frac{\phi(\langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle))}{\langle \sigma, X_k \rangle \phi(y_{k+1})} I(w_{k+1} \le \alpha) \middle| \mathcal{G}_k \right] \\ &= \int_{-\infty}^{\infty} \frac{\phi(\langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle))}{\langle \sigma, X_k \rangle \phi(y_{k+1})} \\ &\quad \times \phi(y_{k+1}) I(w_{k+1} \le \alpha) dy_{k+1} \\ &= \int_{-\infty}^{\alpha} \phi(w)dw. \end{aligned}$$

Henceforth,

$$P(w_{k+1} \leq \alpha | \mathcal{G}_k) = \int_{-\infty}^{\alpha} \phi(w) dw = P(w_{k+1} \leq \alpha)$$

and the result follows.

■

Now, we aim to estimate X, given the observations under the "real world" probability P. However, $\bar{P}$, is an easier measure to work with. Suppose

$$\begin{aligned} \hat{p}_k^i &= P[X_k = e_i | \mathcal{Y}_k] \\ &= E[\langle X_k, e_i \rangle | \mathcal{Y}_k] \\ &= E[I(X_k = e_i) | \mathcal{Y}_k] \end{aligned}$$

and $\hat{p}_k = (\hat{p}_k^1, \cdots, \hat{p}_k^n)$. Then $\hat{p}_k = E[X_k | \mathcal{Y}_k]$ is the conditional distribution of X_k given $\mathcal{Y}_k$ (under P). Thus,

$$E[X_k | \mathcal{Y}_k] = \frac{\bar{E}[\bar{\Lambda}_k X_k | \mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k | \mathcal{Y}_k]},$$

using Bayes' theorem.

We write $q_k := \bar{E}[\bar{\Lambda}_k X_k | \mathcal{Y}_k]$ for the unnormalised conditional contribution of X_k given $\mathcal{Y}_k$. Note again that $\sum_{i=1}^n \langle X_k, e_i \rangle = 1$. And therefore,

$$\sum_{i=1}^n \bar{E}\left[\langle \bar{\Lambda}_k X_k, e_i \rangle \middle| \mathcal{Y}_k \right] = \bar{E}\left[\bar{\Lambda}_k \sum_{i=1}^n \langle X_k, e_i \rangle \middle| \mathcal{Y}_k \right] = \bar{E}[\bar{\Lambda}_k | \mathcal{Y}_k] = \sum_{i=1}^n \langle q_k, e_i \rangle$$

and

$$\hat{p}_k = \frac{q_k}{\sum_{i=1}^n \langle q_k, e_i \rangle}.$$

9.3.3 Recursive Filters

We wish to derive the filter for q.

Lemma 9.3.2 *Write $B(y_{k+1})$ for the diagonal matrix with entries*

$$\frac{\phi(\sigma_i^{-1}(y_{k+1}-c_i))}{\sigma_i\phi(y_{k+1})}. \quad \textit{Then,}$$

$$q_{k+1} = AB(y_{k+1})q_k.$$

Proof: Write

$$\begin{aligned}
q_{k+1} &= \bar{E}[\bar{\Lambda}_{k+1}X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}(AX_k + V_{k+1})|\mathcal{Y}_{k+1}] \\
&= \bar{E}\left[\frac{\bar{\Lambda}_k\phi(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle))}{\langle\sigma, X_k\rangle\phi(y_{k+1})}(AX_k + V_{k+1})\middle|\mathcal{Y}_{k+1}\right] \\
&= \sum_{i=1}^{n}\bar{E}[\bar{\Lambda}_k\langle X_k, e_i\rangle|\mathcal{Y}_k]\frac{\phi(\sigma_i^{-1}(y_{k+1}-c_i))}{\sigma_i\phi(y_{k+1})}Ae_i \\
&= \sum_{i=1}^{n}\langle q_k, e_i\rangle\frac{\phi(\sigma_i^{-1}(y_{k+1}-c_i))Ae_i}{\sigma_i\phi(y_{k+1})} \\
&= AB(y_{k+1})q_k \quad \text{as desired.}
\end{aligned}$$

■

The parameters of this model are:

$$\begin{aligned}
A &= (a_{ji}) = \quad \text{transition matrix,} \\
c &= (c_i) = \quad \text{the function (vectors) and} \\
\sigma &= (\sigma_i) = \quad \text{the volatility vector.}
\end{aligned}$$

To estimate these we need estimates of the following processes:

$$\begin{aligned}
J_k^{rs} &= \sum_{u=1}^{k}\langle X_{u-1}, e_r\rangle\langle X_u, e_s\rangle \\
&= \quad \text{number of jumps from state } r \text{ to state } s \text{ in time } k; \\
O_k^r &= \sum_{u=1}^{k}\langle X_{u-1}, e_r\rangle \\
&= \quad \text{the amount of time } X \text{ has spent in state } r \text{ up to time } k; \text{ and} \\
T_k^r(f) &= \sum_{u=1}^{k}\langle X_{u-1}, e_r\rangle f(y_n) \\
& \quad \text{where} f(y) = y \quad \text{or} \quad y^2.
\end{aligned}$$

Consider the estimate $\hat{J}_k^{rs} = E[J_k^{rs}|\mathcal{Y}_k]$. General Bayes' theorem implies that

$$\begin{aligned}\hat{J}_k^{rs} &= E[J_k^{rs}|\mathcal{Y}_k] \\ &= \frac{\bar{E}[\bar{\Lambda}_k J_k^{rs}|\mathcal{Y}_k]}{\bar{E}[\bar{\Lambda}_k|\mathcal{Y}_k]}.\end{aligned}$$

It turns out however that there is no recursive expression for

$$\bar{E}[\bar{\Lambda}_k J_k^{rs}|\mathcal{Y}_k].$$

Consider the vector process

$$\bar{E}[\bar{\Lambda}_k J_k^{rs} X_k|\mathcal{Y}_k].$$

$$\bar{E}[\bar{\Lambda}_k J_k^{rs} X_k|\mathcal{Y}_k] = \sum_{i=1}^{n} \bar{E}[\bar{\Lambda}_k J_k^{rs}\langle X_k, e_i\rangle|\mathcal{Y}_k]$$

and

$$E[J_k^{rs}|\mathcal{Y}_k] = \hat{J}_k^{rs} = \frac{\bar{E}[\bar{\Lambda}_k J_k^{rs}|\mathcal{Y}_k]}{\langle q_k, \mathbf{1}\rangle}.$$

For any $\mathcal{G}$−adapted process Z, write

$$\begin{aligned}\hat{Z}_k : &= E[Z_k|\mathcal{Y}_k] \\ \sigma(Z_k) : &= \bar{E}[\bar{\Lambda}_k Z_k|\mathcal{Y}_k].\end{aligned}$$

We would like to derive a recursive formula for $\sigma(J^{rs}X)_k$.

Lemma 9.3.3 *With $B(y_{k+1})$ the diagonal matrix with entries $\frac{\phi\left(\sigma_i^{-1}(y_{k+1}-c_i)\right)}{\sigma_i\phi(y_{k+1})}$ and σ defined as above we have*

$$\sigma(J^{rs}X)_{k+1} = AB(y_{k+1})\sigma(J^{rs}X)_k + \langle q_k, e_r\rangle \frac{\phi(\sigma_r^{-1}(y_{k+1}-c_r))}{\sigma_r\phi(y_{k+1})} a_{sr} e_s.$$

Proof: Write $\sigma(J^{rs}X)_{k+1} = \bar{E}[\bar{\Lambda}_{k+1} J_{k+1}^{rs} X_{k+1}|\mathcal{Y}_{k+1}]$.

$$
\begin{aligned}
&\bar{E}[\bar{\Lambda}_{k+1}J^{rs}_{k+1}X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}[\bar{\Lambda}_k\bar{\lambda}_{k+1}(J^{rs}_k + \langle X_k, e_r\rangle\langle X_{k+1}, e_s\rangle)X_{k+1}|\mathcal{Y}_{k+1}] \\
&= \bar{E}\left[\bar{\Lambda}_k \frac{\phi(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle))}{\langle\sigma, X_k\rangle\phi(y_{k+1})} J^{rs}_k(AX_k + V_{k+1}) \middle| \mathcal{Y}_{k+1}\right] \\
&\quad + \bar{E}\left[\bar{\Lambda}_k \frac{\phi(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle))}{\langle\sigma, X_k\rangle\phi(y_{k+1})}\right. \\
&\quad \left. \times\langle X_k, e_r\rangle\langle AX_k + V_{k+1}, e_s\rangle e_s \middle| \mathcal{Y}_{k+1}\right] \\
&= \sum_{i=1}^{n} \bar{E}[\bar{\Lambda}_k\langle X_k, e_i\rangle J^{rs}_k|\mathcal{Y}_k] \frac{\phi(\sigma_i^{-1}(y_{k+1} - c_i))}{\sigma_i\phi(y_{k+1})} Ae_i \\
&\quad + \bar{E}[\bar{\Lambda}_k\langle X_k, e_r\rangle|\mathcal{Y}_k] \frac{\phi(\sigma_r^{-1}(y_{k+1} - c_r))}{\sigma_r\phi(y_{k+1})} a_{sr}e_s \\
&= \sum_{i=1}^{n} \langle\sigma(J^{rs}X)_k, e_i\rangle \frac{\phi(\sigma_i^{-1}(y_{k+1} - c_i))}{\sigma_i\phi(y_{k+1})} Ae_i \\
&\quad + \langle q_k, e_r\rangle \frac{\phi(\sigma_r^{-1}(y_{k+1} - c_r))}{\sigma_r\phi(y_{k+1})} a_{sr}e_s \\
&= AB(y_{k+1})\sigma(J^{rs}X)_k + \langle q_k, e_r\rangle \frac{\phi(\sigma_r^{-1}(y_{k+1} - c_r))}{\sigma_r\phi(y_{k+1})} a_{sr}e_s
\end{aligned}
$$

where $B(y_{k+1})$ is the matrix with terms $\frac{\phi(\sigma_i^{-1}(y_{k+1}-c_i))}{\sigma_i\phi(y_{k+1})}$ on its diagonal as desired.

∎

Similar calculations give the following results:

$$\sigma(O^rX)_{k+1} = AB(y_{k+1})\sigma(O^rX)_k + AB(y_{k+1})\langle q_k, e_r\rangle Ae_r.$$

$$\sigma\big(T^r(f)X\big)_{k+1} = AB(y_{k+1})\sigma(T^r(f)X)_k + \langle q_k, e_r\rangle \frac{\phi(\sigma_r^{-1}(y_{k+1} - c_r))}{\sigma_r\phi(y_{k+1})} f(y_{k+1})Ae_r.$$

9.3.4 Parameter Estimation

We suppose that $\{P_\theta, \theta \in \Theta\}$ is a family of probability measures on a measurable space $(\Omega, \mathcal{F})$, each of which is absolutely continuous with respect to some fixed probability measure P_0. Suppose further that $\mathcal{Y} \subset \mathcal{F}$.

The likelihood function for computing an estimate of θ based on the information given in $\mathcal{Y}$ is

$$L(\theta) = E_0 \left[\frac{dP_\theta}{dP_0} \middle| \mathcal{Y} \right].$$

The maximum likelihood estimate (MLE) of θ is then

$$\hat{\theta} \in \arg\max_{\theta \in \Theta} L(\theta).$$

The reasoning behind this choice is that the most likely value of θ is the one which maximises this conditional expectation of the density. However, the MLE is hard to compute.

The Expectation Maximisation (EM) algorithm is an alternative approximate method. The steps to perform in doing the EM algorithm are the following:

1. **Step 1:** Set $p = 0$ and choose $\hat{\theta}_0$.
2. **Step 2:** (E-Step): Set $\theta^* = \hat{\theta}_p$ and compute

$$Q(\theta, \theta^*) = E_{\theta^*} \left[\log \frac{dP_\theta}{dP_\theta^*} \middle| \mathcal{Y} \right].$$

3. **Step 3:** (M-Step): Find $\hat{\theta}_{p+1} \in \arg\max_{\theta \in \Theta} Q(\theta, \theta^*)$.
4. **Step 4:** Replace p by $p+1$ and repeat from Step 2 until some stopping criteron is satisfied.

The sequence $\{\hat{\theta}_p\}$ gives non-decreasing values of the likelihood function to a local maximum of the likelihood function.

From Jensen's inequality:

$$\log L(\hat{\theta}_{p+1}) - \log L(\hat{\theta}_p) \geq Q(\hat{\theta}_{p+1}, \hat{\theta}_p).$$

with equality only when $\hat{\theta}_{p+1} = \hat{\theta}_p$.

The model under current investigation is determined by the parameters

$$\theta := \{a_{ji}, c_i, \sigma_i, \quad 1 \leq i, j \leq n\}.$$

Further $a_{ji} \geq 0$, $\sum_{j=1}^n a_{ji} = 1$ and $\sigma_i > 0$. We wish to determine estimates of the afore-mentioned parameters given by a new set

$$\hat{\theta} = \{\hat{a}_{ji}, \hat{c}_i, \hat{\sigma}_i, \quad 1 \leq i, j \leq n\}$$

which maximises the analogues of the Q functions.

Consider first the parameter a_{ji}. We recall the form of the change of measure described in the preceeding section. Under P_θ, X is a Markov chain with transition matrix $A = (a_{ji})$. We wish to introduce a new probability measure $P_{\hat{\theta}}$, under which X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{ji})$. That is,

$$P_{\hat{\theta}}(X_{k+1} = e_j | X_k = e_i) = \hat{a}_{ji},$$

so $\hat{a}_{ji} \geq 0$ and $\sum_{j=1}^n \hat{a}_{ji} = 1$. Define

$$\begin{aligned} \Lambda_0 &= 1 \\ \Lambda_k &= \prod_{l=1}^k \left(\sum_{r,s=1}^n \frac{\hat{a}_{sr}}{a_{sr}} \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle \right). \end{aligned}$$

In case $a_{ji} = 0$, take $\hat{a}_{ji} = 0$ and $\frac{\hat{a}_{ji}}{a_{ji}} = 1$. Define $P_{\hat{\theta}}$ by setting

$$\left. \frac{dP_{\hat{\theta}}}{dP_\theta} \right|_{\mathcal{F}_k} = \Lambda_k.$$

Lemma 9.3.4 *Under $P_{\hat{\theta}}$ X is a Markov chain with transition matrix $\hat{A} = (\hat{a}_{ji})$.*

Proof:

$$E_{\hat{\theta}}[\langle X_{k+1}, e_s\rangle | \mathfrak{F}_k] = \frac{E_\theta[\Lambda_{k+1}\langle X_{k+1}, e_s\rangle | \mathfrak{F}_k]}{E_\theta[\Lambda_{k+1}|\mathfrak{F}_k]}$$

$$= \frac{E_\theta\left[\left(\sum_{r,s=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle\langle X_k, e_r\rangle\right)\langle X_{k+1}, e_s\rangle | \mathfrak{F}_k\right]}{E_\theta\left[\sum_{r,s=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle\langle X_k, e_r\rangle \middle| \mathfrak{F}_k\right]}.$$

Claim:

$$E_\theta\left[\sum_{r,s=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle\langle X_k, e_r\rangle \middle| \mathfrak{F}_k\right] = 1.$$

$$\begin{aligned}
& E_\theta\left[\sum_{r,s=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle\langle X_k, e_r\rangle \middle| \mathfrak{F}_k\right] \\
= & \sum_{r=1}^n E_\theta\left[\sum_{s=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle \middle| X_k = e_r\right]\langle X_k, e_r\rangle \\
= & \sum_{r=1}^n \left(\sum_{s=1}^n \frac{\hat{a}_{sr}}{a_{sr}} a_{sr}\right)\langle X_k, e_r\rangle \\
= & \sum_{r=1}^n \langle X_k, e_r\rangle = 1.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& E_\theta\left[\left(\sum_{r,s=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle\langle X_k, e_r\rangle\right)\langle X_{k+1}, e_s\rangle \middle| \mathfrak{F}_k\right] \\
= & E_\theta\left[\sum_{r=1}^n \frac{\hat{a}_{sr}}{a_{sr}}\langle X_{k+1}, e_s\rangle\langle X_k, e_r\rangle \middle| \mathfrak{F}_k\right] \\
= & \hat{a}_{sr}\langle X_k, e_r\rangle.
\end{aligned}$$

Consequently,

$$P_{\hat{\theta}}(X_{k+1} = e_s | X_k = e_r) = E_{\hat{\theta}}[\langle X_{k+1}, e_s\rangle | X_k = e_r] = \hat{a}_{sr}.$$

We therefore see that X is a Markov chain with transition matrix $\hat{A}$, under $P_{\hat{\theta}}$.

■

Theorem 9.3.1 *Given the observations up to time k, $\{y_1, \cdots, y_k\}$ and given the parameter set $\theta = \{a_{ji}, c_i, \sigma_i, \quad 1 \leq i, \quad j \leq n\}$, the EM estimates $\hat{a}_{ji}$ are given by*

$$\hat{a}_{ji} = \frac{\hat{J}_k^{ij}}{\hat{O}_k^i} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}.$$

Proof: We make an observation that

$$\left.\frac{dP_{\hat{\theta}}}{dP_\theta}\right|_{\mathcal{F}_k} = \prod_{l=1}^{k} \left(\sum_{r,s=1}^{n} \frac{\hat{a}_{sr}}{a_{sr}} \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle \right).$$

Equivalently,

$$\log \frac{dP_{\hat{\theta}}}{dP_\theta} = \sum_{l=1}^{k} \log \left(\sum_{r,s=1}^{n} \frac{\hat{a}_{sr}}{/a_{sr}} \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle \right)$$

and further

$$\begin{aligned} \log \frac{dP_{\hat{\theta}}}{dP_\theta} &= \sum_{l=1}^{k} \sum_{r,s=1}^{n} \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle (\log \hat{a}_{sr} - \log a_{sr}) \\ &= \sum_{r,s=1}^{n} J_k^{rs} \log \hat{a}_{sr} + R(a) \end{aligned}$$

where $R(a)$ is independent of the $\hat{a}_{sr}$. Hence,

$$L(\hat{\theta}) = E_\theta \left[\log \left. \frac{dP_{\hat{\theta}}}{dP_\theta} \right| \mathcal{Y}_k \right] = \sum_{r,s=1}^{n} \hat{J}_k^{rs} \log \hat{a}_{sr} + \hat{R}(a). \tag{9.3.1}$$

We know that

$$\sum_{s=1}^{n} \hat{a}_{sr} = 1. \tag{9.3.2}$$

Also,

$$\sum_{s=1}^{n} J_k^{rs} = O_k^r$$

and therefore

$$\sum_{s=1}^{n} \hat{J}_k^{rs} = \hat{O}_k^r. \tag{9.3.3}$$

The optimal estimate $\hat{a}_{ji}$ is the value which maximises the right hand side of (9.3.1) subject to (9.3.2).

Suppose λ is the Lagrange multiplier. Set

$$L(\hat{a},\lambda) = \sum_{r,s=1}^{n} \hat{J}_k^{rs} \log \hat{a}_{sr} + \hat{R}(a) + \lambda\left(\sum_{s=1}^{n} \hat{a}_{sr} - 1\right).$$

We differentiate L in $\hat{a}_{ji}$ and λ and equate the derivatives to 0. This gives us:

$$\frac{1}{\hat{a}_{ji}} \hat{J}_k^{ij} + \lambda = 0 \tag{9.3.4}$$

$$\sum_{s=1}^{n} \hat{a}_{si} = 1 \tag{9.3.5}$$

From (9.3.3), (9.3.4) and (9.3.5), $\lambda = -\hat{O}_k^r$ and therefore

$$\hat{a}_{ji} = \frac{\hat{J}_k^{ij}}{\hat{O}_k^i} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}$$

and the theorem is proved.

■

9.3.5 Updates of Parameters

Consider the parameter $c = (c_1, \cdots, c_n)' \in \mathbb{R}^n$. To change this parameter to $\hat{c} = (\hat{c}_1, \cdots, \hat{c}_n)'$, consider the factors

$$\lambda_{l+1}^* := exp[2^{-1}\langle\sigma, X_l\rangle^{-2}\{\langle c, X_l\rangle^2 - \langle\hat{c}, X_l\rangle^2 - 2y_{l+1}\langle c, X_l\rangle + 2y_{l+1}\langle\hat{c}, X_l\rangle\}].$$

Write $\Lambda_k^* = \prod_{l=1}^k \lambda_l^*$ and consider a new measure $P_{\hat{\theta}}$ defined by setting

$$\left.\frac{dP_{\hat{\theta}}}{dP_\theta}\right|_{\mathcal{G}_k} = \Lambda_k^*.$$

Thus,

$$L(\hat{\theta}) = E\left[\log \frac{dP_{\hat{\theta}}}{dP_\theta}\middle| \mathcal{Y}_k\right].$$

Further,

$$\begin{aligned} L(\hat{\theta}) &= E\left[\sum_{l=1}^{k}\Big(2\langle\sigma, X_{l-1}\rangle\Big)^{-2}\Big\{\langle c, X_{l-1}\rangle^2 - \langle\hat{c}, X_{l-1}\rangle^2 \right. \\ &\quad \left. -2y_l\langle c, X_{l-1}\rangle + 2y_l\langle\hat{c}, X_{l-1}\rangle\Big\}\middle|\mathcal{Y}_k\right] \\ &= E\left[\sum_{r=1}^{n}\frac{2T_k^r(y)\hat{c}_r - O_k^r\hat{c}_r^2}{2\sigma_r^2} + R(c)\middle| \mathcal{Y}_k\right] \end{aligned}$$

where $R(c)$ is independent of $\hat{c}$. That is,

$$L(\hat{\theta}) = \sum_{r=1}^{n}\frac{2\hat{T}_k^r(y)\hat{c}_r - \hat{O}_k^r\hat{c}_r^2}{2\sigma_r^2} + \hat{R}(c).$$

Differentiating $L(\hat{\theta})$ with respect to $\hat{c}_i$ and equating to zero, the optimal choice for $\hat{c}_i$, given the observations $y_1, \cdots, y_k$ is

$$\hat{c}_i = \frac{\hat{T}_k^i(y)}{\hat{O}_k^i} = \frac{\sigma\big(T^i(y)\big)_k}{\sigma(O^i)_k}.$$

Now, consider the parameters $\sigma_i, \quad 1 \le i \le n$. We shall change the parameters $\sigma = (\sigma_1, \cdots, \sigma_n)$ to $\hat{\sigma} = (\hat{\sigma}_1, \cdots, \hat{\sigma}_n)$.

We consider the factors

$$\lambda_l = \frac{\langle\sigma, X_l\rangle exp\left(-\frac{1}{2\langle\hat{\sigma}, X_l\rangle^2}(y_{l+1} - \langle c, X_l\rangle)^2\right)}{\langle\hat{\sigma}, X_l\rangle exp\left(-\frac{1}{2\langle\sigma, X_l\rangle^2}(y_{l+1} - \langle c, X_l\rangle)^2\right)}.$$

Write $\Lambda_k = \prod_{l=1}^k \lambda_l$ and define $P_{\hat{\theta}}$ so that

$$\left.\frac{dP_{\hat{\theta}}}{dP_\theta}\right|_{\mathcal{G}_k} = \Lambda_k.$$

Then,

$$\log \frac{dP_{\hat\theta}}{dP_\theta} = \sum_{l=1}^{k}\Big(-\log\langle\hat\sigma, X_{l-1}\rangle \\ -\frac{1}{2\langle\hat\sigma, X_{l-1}\rangle^2}(y_l - \langle c, X_{l-1}\rangle)^2 + R(c,\sigma)\Big),$$

where $R(c,\sigma)$ is independent of $\hat\theta$. Therefore,

$$E\left[\log\frac{dP_{\hat\theta}}{dP}\middle|\mathcal{F}_k\right] = E\Bigg[\sum_{l=1}^{k}\sum_{r=1}^{n}(-\langle X_{l-1}, e_r\rangle\log\hat\sigma_r \\ -\frac{\langle X_{l-1}, e_r\rangle}{2\hat\sigma_r^2}(y_l^2 - 2c_r y_l + c_r^2)\Bigg|\mathcal{Y}_k)\Bigg] + \hat R(c,\sigma) \\ = -\sum_{r=1}^{n}\left[\log\hat\sigma_r\hat O_k^r + \frac{1}{2\hat\sigma_r^2}(\hat T_k^r(y^2) - 2c_r\hat T_k^r(y) + c_r^2\hat O_k^r)\right] + \hat R(c,\sigma).$$

Differentiating in $\hat\sigma_i$ and putting the derivative to 0, we see that the optimal choice for $\hat\sigma_i$, given the observations y_k, is

$$\hat\sigma_i = \left(\frac{\hat T_k^i(y^2) - 2c_i\hat T_k^i(y) + c_i^2\hat O_k^i}{\hat O_k^i}\right)^{\frac{1}{2}} \\ = \left(\frac{\sigma\left(T^i(y^2)\right)_k - 2c_i\sigma\left(T^i(y)\right)_k}{\sigma(O^i)_k + c_i^2}\right)^{\frac{1}{2}}.$$

Note that these results provide not only estimates of the Markov chain but also of the parameters of the model.

9.4 Applications of Filtering Techniques to a Mean Reverting Interest Rate Model

9.4.1 The Model Revisited

In Chapter 8, section 8.4, we discussed a Markov model for the mean reverting level of interest rates. In this section, we aim to generate optimal filters for the state of the hidden Markov chain. Auxiliary filters will also be obtained to enable parameters of the model to be estimated using the EM algorithm. Then a simulation study will be conducted.

From the previous formulation of this model, we consider an $n-$state continuous time Markov chain $X = \{X_t; 0 \leq t \leq T\}$ that is identical to α after the transformation of the state space. As before, we choose the state space for X the set $(e_1, \cdots, e_n)$ of unit vectors.

We also write $\alpha = \langle \alpha, X_t \rangle$ and the vector of probabilities $p_t = E[X_t]$.

If $A = \{A_t; 0 \leq t \leq T\}$ is a family of transition intensity matrices associated with the continuous time Markov chain X, p_t satisfies the forward equation $\frac{dp_t}{dt} = A_t p_t$ with given initial probability vector p_0.

The interest rate $r = \{r_t; 0 \leq t \leq T\}$ is described by the SDE

$$dr_t = \gamma(\alpha_t - r_t)dt + \rho dW_t. \tag{9.4.6}$$

Here, $W = \{W_t; 0 \leq t \leq T\}$ is a Wiener process independent of α. The adjustment coefficient γ and volatility ρ are positive constants.

The transition intensity matrix A_t governs the dynamics of the reference level α that feeds into the interest rate model in (9.4.6).

We define the following filtrations for the processes involved in this model. Let $\mathcal{R}_t^0$ be the $\sigma-$field generated by r_u for $0 \leq u \leq t$, that is, $\mathcal{R}_t^0 = \sigma\{r_u; 0 \leq u \leq t\}$. Also, let $\mathcal{S}_t^0 = \sigma\{X_u; 0 \leq u \leq t\}$, $\mathcal{F}_t^0 = \sigma\{X_u, r_u; 0 \leq u \leq t\}$ and we write $\mathcal{R} = \{\mathcal{R}_t\}_{0 \leq t \leq T}$ and $\mathcal{S} = \{S_t\}_{0 \leq t \leq T}$ and $\mathcal{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}$ for the corresponding right-continuous, augmented filtrations.

As usual, the process $M = \{M_t; 0 \le t \le T\}$ defined by

$$M_t = X_t - X_0 - \int_0^t A_u X_u du$$

is an $\mathcal{S}-$martingale under P.

9.4.2 A Change of Measure

It is the intent of this section to obtain filters and estimators required to estimate the parameters of the model. To do this, we shall introduce a change of measure.

Let $\bar{P}$ be a probability measure on $(\Omega, \mathcal{F}_T)$ under which α is a finite state Markov chain with the transition intensity matrix family A as before. Further, let $\bar{W} = \{\bar{W}_t = \frac{r_t}{\rho}; 0 \le t \le T\}$ be a Wiener process, independent of α.

Suppose $\Gamma = \{\Gamma_t; 0 \le t \le T\}$ is a process defined by

$$\Gamma_t = exp\left[\int_0^t \eta(\langle X_u, \alpha\rangle - r_u)d\bar{W}_u - \frac{1}{2}\int_0^t \eta^2(\langle X_u, \alpha\rangle - r_u)^2 du\right]$$

where $\eta = \frac{\gamma}{\rho}$.

Consider a new measure, P, in $\mathcal{F}_T$ such that $P \sim \bar{P}$ and its Radon-Nikodym derivative with respect $\bar{P}$ is

$$\left.\frac{dP}{d\bar{P}}\right|_{\mathcal{F}_T} = \Gamma_T.$$

If we let W to be a process given by $W_0 = 0$ and

$$dW_t = \rho^{-1}[dr_t - \gamma(\langle X_t, \alpha\rangle - r_t)dt],$$

Girsanov's theorem tells us that the process W is a $P-$Wiener process, independent of α.

Under P, X and r follow respectively the following dynamics

$$X_t = X_0 + \int_0^t A_u X_u du + M_t$$

and

$$dr_t = \gamma(\alpha_t - r_t)dt + \rho dW_t.$$

The change of measure facilitates easier calculations in the sense that under $\bar{P}$ the observable process r is a Wiener process.

P which is the "real world" measure is a different measure. However, we can use a version of Bayes' theorem to convert calculations made under one measure into a corresponding quantity calculated under the other measure.

9.4.3 Calculation of Filters

Suppose we have an $\mathcal{F}-$adapted process given by $\psi = \{\psi_t; 0 \leq t \leq T\}$.

Let $\hat{\psi} = \{\hat{\psi}_t; 0 \leq t \leq T\}$ for the $\mathcal{R}-$optional projection of the process ψ.

Under P, we have $\hat{\psi}_t = E[\psi_t|\mathcal{R}_t]$ $P-$a.s.

Definition 9.4.1 *The $\mathcal{R}-$optional projection process $\hat{\psi}$ in the preceding discussion is called the* **filter** *of ψ.*

Denote by $\sigma(\psi) = \{\sigma(\psi_t); 0 \leq t \leq T\}$ the $\mathcal{R}-$optional projection of the process $\psi\Gamma$ under the measure $\bar{P}$.

From Theorem 2.3.2 of Elliott, Aggoun and Moore [36] we have

$$\hat{\psi}_t = \frac{\sigma(\psi_t)}{\sigma(1)} \quad \text{(P-a.s.)}.$$

Now let the process $J = \{J_t; 0 \leq t \leq T\}$ be described by

$$J_t = J_0 + \int_0^t \xi_u du + \int_0^t \langle \beta_u, dM_u \rangle + \int_0^t \delta_u dW_u,$$

where ξ and δ are $\mathcal{F}-$predictable, square integrable process; and β is an $\mathcal{F}-$predictable, square integrable, $n-$dimensional vector process.

With the Markov process $X_t = X_0 + \int_0^t AX_u du + M_t$ and Itô's Rule for semimartingales the process $J_t X_t$ has dynamics

$$\begin{aligned} J_t X_t &= J_0 X_0 + \int_0^t \xi_u X_{u-} du + \int_0^t X_{u-} \langle \beta_u, dM_u \rangle + \int_0^t \delta_u X_{u-} dW_u \\ &+ \int_0^t J_{u-} A_u X_u du + \int_0^t J_{u-} dM_u + \sum_{0 \leq u \leq t} \langle \beta_u, \Delta X_u \rangle \Delta X_u. \end{aligned}$$

Let $(a_{ji})_u$ be the j, ith element of the matrix A_u.

We shall derive the filter of $J_t X_t$.

Theorem 9.4.1 *The recursive equation for the evolution of $\sigma(JX)$ is given by*

$$\begin{aligned}\sigma(J_t X_t) &= \sigma(J_0 X_0) + \int_0^t \sigma(\xi_u X_{u-})du + \int_0^t A_u \sigma(J_{u-} X_u)du \\ &+ \sum_{i,j=1}^{n} \int_0^t \langle \sigma(\beta_u^j X_{u-} - \beta_u^i X_{u-}), e_i\rangle (a_{ji})_u du (e_j - e_i) \\ &+ \int_0^t \rho^{-1}(\eta B_u \sigma(J_{u-} X_u) + \sigma(\rho X_{u-}))dr_u\end{aligned}$$

for $0 \leq t \leq T$, where B_u is the $n \times n$ diagonal matrix with $(B_{ii})_u = \alpha_i - r_u$.

Proof: See Theorem 8.3.2 of Elliott, Aggoun and Moore [36].

Example 9.4.1 Suppose ξ, β and δ are all zero and we take $J_t = J_0 = 1$. Then,

$$\sigma(X_t) = E[X_0] + \int_0^t A_u \sigma(X_u)du + \int_0^t \rho^{-1}\eta B_u \sigma(X_u)dr_u.$$

With $\mathbf{1} = (1, \cdots, 1)$ and $\langle X_t, \mathbf{1}\rangle = 1$, we see that the recursive equation of the filter for X_t has the form

$$E[X_t|\mathcal{R}_t] = \frac{\sigma(X_t)}{\langle \sigma(X_t), \mathbf{1}\rangle}.$$

Define $\mathfrak{J}_t^{ij}$:=number of jumps that the process X makes from state e_i to e_j in the interval $[0, t]$ and hence,

$$\mathfrak{J}_t^{ij} = \int_0^t \langle X_{u-}, e_i\rangle a_{ji} du + \int_0^t \langle X_{u-}, e_i\rangle\langle e_j, dM_u\rangle. \tag{9.4.7}$$

The unnormalised filter for $\mathfrak{J}_t^{ij}$ is given by

$$\boxed{\sigma(\mathfrak{J}_t^{ij}) = \langle \sigma(J_t^{ij} X_t), \mathbf{1}\rangle},$$

and the normalised filter for $\mathcal{J}^{ij}$ is given by

$$\boxed{E[\mathcal{J}_t^{ij}|\mathcal{R}_t] = \frac{\sigma(\mathcal{J}_t^{ij})}{\langle\sigma(X_t), \mathbf{1}\rangle}}.$$

Corollary 9.4.1 *The recursive algorithm for the process* $\sigma(\mathcal{J}_t^{ij}X_t)$ *is*

$$\begin{aligned}\sigma(\mathcal{J}_t^{ij}X_t) &= \int_0^t a_{ji}\langle\sigma(X_u), e_i\rangle e_j du + \int_0^t A\sigma(\mathcal{J}_u^{ij}X_u)du \\ &+ \int_0^t \rho^{-1}\eta B_u\sigma(\mathcal{J}_u^{ij}X_u)dr_u. \end{aligned} \quad (9.4.8)$$

Proof: First note that $\langle X_{u-}, e_i\rangle X_{u-} = \langle X_{u-}, e_i\rangle e_i$ and therefore

$$\sum_{k,l=1}^{n}\langle(\beta_u^l X_{u-}, \beta_u^k X_{u-}), e_k\rangle a_{lk}(e_l - e_k) = \langle\langle X_{u-}, e_i\rangle e_i, e_i\rangle a_{ji}(e_j - e_i)$$
$$= \langle X_{u-}, e_i\rangle a_{ji}(e_j - e_i).$$

Then take $J_t = \mathcal{J}_t^{ij}$, $\quad J_0 = 0$,

$$\xi_u = \langle X_{u-}, e_i\rangle a_{ji}, \quad \xi_u = 0, \quad \text{and}$$

$$\beta_u = \langle X_{u-}, e_i\rangle e_j$$

and the result follows.

∎

Now, define the process $\mathcal{O}_t^i$ by

$$\begin{aligned}\mathcal{O}_t^i &:= \int_0^t \langle X_u, e_i\rangle du \\ &= \text{amount of time the process } X \text{ stays at state } i \text{ up to time } t \\ &= \int_0^t \langle X_u, e_i\rangle du. \end{aligned} \quad (9.4.9)$$

The unnormalised filter for $\mathcal{O}$ is given by

$$\sigma(\mathcal{O}_t^i) = \langle\sigma(\mathcal{O}_t^i X_t), \mathbf{1}\rangle$$

and the filter for $\mathcal{O}^i$ is given by

$$\boxed{E[\mathcal{O}_t^i|\mathcal{R}_t] = \frac{\sigma(\mathcal{O}_t^i)}{\langle\sigma(X_t), \mathbf{1}\rangle}}.$$

Corollary 9.4.2 *The recursive algorithm for the process $\sigma(\mathcal{O}_t^i X_t)$ is given by*

$$\begin{aligned}\sigma(\mathcal{O}_t^i X_t) &= \int_0^t \langle\sigma(X_u), e_i\rangle e_i du + \int_0^t A\sigma(\mathcal{O}_u^i X_u) du \\ &+ \int_0^t \rho^{-1}\eta B_u \sigma(\mathcal{O}_u^i X_u) dr_u.\end{aligned}$$

Proof: Observe that $\langle X_u, e_i\rangle X_u = \langle X_u, e_i\rangle e_i$ then apply Theorem 9.4.1 by taking

$$J_t = \mathcal{O}_t^i, \quad J_0 = 0, \quad \xi_u = \langle X_u, e_i\rangle \quad \text{and} \quad \beta_u = \delta_u = 0.$$

The result follows.

■

Define the process $\mathcal{K}_t^i := \int_0^t \langle X_u, e_i\rangle dr_u$.

Using the dynamics of r_u as a mean reverting process and the fact that $\langle X_u, e_i\rangle X_u = \langle X_u, e_i\rangle e_i$, we have

$$\mathcal{K}_t^i = \int_0^t \gamma(\alpha_i - r_u)\langle X_u, e_i\rangle du + \int_0^t \rho\langle X_u, e_i\rangle dW_u. \tag{9.4.10}$$

As usual, the unnormalised filter for $\mathcal{K}^i$ is given by

$$\boxed{\sigma(\mathcal{K}_t^i) = \langle\sigma(\mathcal{K}_t^i X_t), \mathbf{1}\rangle}$$

and the normalised filter for $\mathcal{K}^i$ is given by

$$\boxed{E[\mathcal{K}_t^i|\mathcal{R}_t] = \frac{\sigma(\mathcal{K}_t^i)}{\langle\sigma(X_t), \mathbf{1}\rangle}}.$$

Corollary 9.4.3 *The recursive algorithm for the process $\sigma(\mathcal{K}_t^i X_t)$ is given by*

$$\sigma(\mathcal{K}_t^i X_t) = \int_0^t \gamma(\alpha_i - r_u)\langle\sigma(X_u), e_i\rangle e_i du + \int_0^t A\sigma(\mathcal{K}_u^i X_u)du + \int_0^t (\rho^{-1}\eta B_u \sigma(\mathcal{K}_u^i X_u) + \langle\sigma(X_u), e_i\rangle e_i) dr_u$$

Proof: Apply Theorem 9.4.1 with

$$J_t = \mathcal{K}_t^i, \quad J_0 = 0, \quad \xi_u = \gamma(\alpha_i - r_u)\langle X_u, e_i\rangle$$
$$\rho_u = \eta\langle X_u, e_i\rangle \quad \text{and} \quad \beta_u = 0$$

and the above result is obtained.

■

Now, consider the process $\mathcal{J}_t^i$ defined by

$$\mathcal{J}_t^i = \int_0^t r_u \langle X_u, e_i\rangle du.$$

Again, the unnormalised filter for $\mathcal{J}^i$ is given by

$$\boxed{\sigma(\mathcal{J}_t^i) = \langle\sigma(\mathcal{J}_t^i X_t), \mathbf{1}\rangle}$$

and the normalised filter for $\mathcal{J}^i$ is given by

$$\boxed{E[\mathcal{J}_t^i | \mathcal{R}_t] = \frac{\sigma(\mathcal{J}_t^i)}{\langle\sigma(X_t), \mathbf{1}\rangle}}.$$

Corollary 9.4.4 *The recursive algorithm for the process $\sigma(\mathcal{J}_t^i X_t)$ is given by*

$$\sigma(\mathcal{J}_t^i X_t) = \int_0^t r_u\langle\sigma(X_u), e_i\rangle e_i du + \int_0^t A\sigma(\mathcal{J}_u^i X_u)du + \int_0^t \rho^{-1}\eta B_u \sigma(\mathcal{J}_u^i X_u) dr_u.$$

Proof: The result also follows from Theorem (9.4.1) by taking

$$J_t = \mathcal{J}_t^i, \quad Y_0 = 0, \quad \xi_u = r_u\langle X_u, e_i\rangle, \quad \text{and} \quad \beta_u = \delta_u = 0.$$

■

9.4.4 Parameter Estimation of the Model

We wish to estimate the transition intensity matrix $A = A_t$ which is constant and unknown and the vector α, the reference level values which are also unknown.

Since, the process r can be observed up to time t, we can use Expectation Maximisation (EM) algorithm to estimate the said unknown parameters.

Write

$$\begin{aligned} \theta : &= \text{set of parameters} \\ &= \{a_{ij}, \alpha_i; 1 \leq i, j \leq n\}. \end{aligned}$$

As a preliminary, an initial guess θ_0 for the parameter set is chosen. Then the EM algorithm is applied to obtain the first estimate θ_1 of the parameters. We repeat this procedure iteratively and hence generate a sequence of estimates $(\theta_k)_{k \in \mathbb{Z}^+}$. We note that in each iteration, there are two steps involved.

First Step: Expectation

Start with θ_k, the $k-$th iteration of the estimated parameters. Write

$$\begin{aligned} \hat{\theta} &= \text{set of possible parameter values} \\ P_{\hat{\theta}} &= \text{probability measure induced by the values } \hat{\theta} \text{ on } (\Omega, \mathcal{F}_t). \\ E_k &= \text{expectation under the measure } P_{\theta_k}. \end{aligned}$$

Under this Expectation Step, our objective is to calculate the quantity

$$Q(\hat{\theta}; \theta_k) := E_k \left[\log \left(\frac{dP_{\hat{\theta}}}{dP_{\theta_k}} \right) \middle| \mathcal{R}_t \right].$$

Second Step: Maximisation

This requires the maximisation of the quantity $Q(\hat{\theta}, \theta_k)$ with respect to $\hat{\theta}$ to obtain a new estimate θ_{k+1}.

Through this procedure, the EM algorithm ostensibly maximises iteratively the likelihood that the estimated parameters are indeed the true underlying parameters.

Lemma 9.4.1 *The $k+1-$parameter set estimate*

$$\theta_{k+1} = \{a_{k+1,ij}, \alpha_{k+1,i}; \quad 1 \leq i,j \leq n\}$$

generated by the EM algorithm is given by

$$a_{k+1,ji} = \frac{\sigma(\mathcal{J}_t^{ij})}{\sigma(\mathcal{O}_t^i)} \quad \textit{and} \quad \alpha_{k+1,i} = \frac{\gamma^{-1}\sigma(\mathcal{K}_t^i) + \sigma(\mathcal{T}_t^i)}{\sigma(\mathcal{O}_t^i)}.$$

Proof: We invoke Girsanov's theorem and Theorem T3 from Chapter 4 of Brémaud [9] to find the expression for $\frac{dP_{\bar{\theta}}}{dP_{\theta_k}}$, where

$$\begin{aligned} \bar{\theta} &= \{\bar{a}_{ij}, \quad \bar{\alpha}_i \quad 1 \leq i,j \leq n\} \qquad \text{and} \\ \theta_n &= \{a_{k,ij}, \quad \alpha_k, i \quad 1 \leq i,j \leq n\} \end{aligned}$$

are two possible sets.

Now,

$$\begin{aligned} \frac{dP_{\bar{\theta}}}{dP_{\theta_k}} &= exp\left[\int_0^t \eta\rho^{-1}\langle X_u, \bar{\alpha} - \alpha_k\rangle dr_u - \frac{1}{2}\int_0^t \eta^2\Big\{(\langle X_u, \bar{\alpha}\rangle - r_u)^2 \right. \\ &\quad \left. -(\langle X_u, \alpha_k\rangle - r_u)^2\Big\} du\right] \\ &\quad \times \prod_{i,j=1, i\neq j}^{n} exp\left[\int_0^t \log\left(\frac{\bar{a}_{ji}}{a_{k,ji}}\right) d\mathcal{J}_u^{ij} \right. \\ &\quad \left. - \int_0^t (\bar{a}_{ji} - a_{k,ji})\langle X_u, e_i\rangle du\right] \end{aligned} \tag{9.4.11}$$

where $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2, \cdots, \bar{\alpha}_n)$ and $\alpha_k = (\alpha_{k,1}, \alpha_{k,2}, \cdots, \alpha_{k,n})$

First, we observe that

$$\langle X_u, \bar{\alpha}\rangle = \sum_{i=1}^{n} \bar{\alpha}_i \langle X_u, e_i\rangle.$$

Therefore,

$$
\begin{aligned}
Q(\bar{\theta},\theta_k) &= E_k\left[\log\left(\frac{dP_{\bar{\theta}}}{dP_{\theta_k}}\right)\Bigg|\,\mathcal{R}_t\right] \\
&= \sum_{i=1}^{n}\eta\rho^{-1}\bar{\alpha}_i E_k\left[\int_0^t\langle X_u,e_i\rangle dr_u\Bigg|\,\mathcal{R}_t\right] \\
&\quad -\frac{1}{2}\eta^2 E_k\left[\int_0^t\left(\langle X_u,\bar{\alpha}\rangle^2-2r_u\langle X_u,\bar{\alpha}\rangle\right)du\Bigg|\,\mathcal{R}_t\right] \\
&\quad +\sum_{i,j=1,i\neq j}^{n}\Bigg(\log(\bar{a}_{ji})E_k[\mathcal{J}_t^{ij}|\mathcal{R}_t] \\
&\quad -\bar{a}_{ji}E_k\left[\int_0^t\langle X_u,e_i\rangle du\Bigg|\,\mathcal{R}_t\right]\Bigg)+R(\theta_k), \qquad (9.4.12)
\end{aligned}
$$

where $R(\theta_k)$ is independent of $\bar{\theta}$. Equation (9.4.12) can be further simplified using the fact that

$$\langle X_u,\bar{\alpha}\rangle^2=\sum_{i=1}^{n}\bar{\alpha}_i^2\langle X_u,e_i\rangle.$$

Thus,

$$
\begin{aligned}
Q(\bar{\theta},\theta_k) &= \sum_{i=1}^{n}\eta\rho^{-1}\bar{\alpha}_i E_k\left[\int_0^t\langle X_u,e_i\rangle dr_u\Bigg|\,\mathcal{R}_t\right] \\
&\quad -\sum_{i=1}^{n}\frac{1}{2}\eta^2\bar{\alpha}_i^2 E_k\left[\int_0^t\langle X_u,e_i\rangle du\Bigg|\,\mathcal{R}_t\right] \\
&\quad +\sum_{i=1}^{n}\eta^2\bar{\alpha}_i E_k\left[\int_0^t r_u\langle X_u,e_i\rangle du\Bigg|\,\mathcal{R}_t\right] \\
&\quad +\sum_{i,j=1,i\neq j}^{n}\Bigg(\log(\bar{a}_{ji})E_k[\mathcal{J}_t^{ij}|\mathcal{R}_t] \\
&\quad -\bar{a}_{ji}E_k\left[\int_0^t\langle X_u,e_i\rangle du\Bigg|\,\mathcal{R}_t\right]\Bigg)+R(\theta_k) \qquad (9.4.13)
\end{aligned}
$$

Then, we maximise $Q(\bar{\theta},\theta_k)$ with respect to $\bar{\theta}$. To do this, we equate the partial derivatives with respect to $\bar{a}_{ji}$ and $\bar{\alpha}_i$ of Equation (9.4.13) to zero.

The calculation shows that the next set of parameters

$$\theta_{k+1}=\{a_{k+1,ij},\quad \alpha_{k+1,i},\quad 1\le i,j\le n\}$$

generated by the EM algorithm is given by

$$a_{k+1,ji} = E_k[\mathcal{J}_t^{ij}|\mathcal{R}_t]\left(E_k\left[\int_0^t \langle X_u, e_i\rangle du|\mathcal{R}_t\right]\right)^{-1} \tag{9.4.14}$$

and

$$\alpha_{k+1,i} = \frac{\gamma^{-1}E_k\left[\int_0^t \langle X_u, e_i\rangle dr_u|\mathcal{R}_t\right] + E_k\left[\int_0^t r_u\langle X_u, e_i\rangle du\middle|\mathcal{R}_t\right]}{E_k\left[\int_0^t \langle X_u, e_i\rangle du|\mathcal{R}_t\right]}. \tag{9.4.15}$$

But,

$$E_k[\mathcal{J}_t^{ij}|\mathcal{R}_t] = \frac{\sigma(\mathcal{J}_t^{ij})}{\langle\sigma(X_t), \mathbf{1}\rangle}$$

and

$$E_k\left[\int_0^t \langle X_u, e_i\rangle du\middle|\mathcal{R}_t\right] = E_k[\mathcal{O}_t^i|\mathcal{R}_t] = \frac{\sigma(O_t^i)}{\langle\sigma(X_t), \mathbf{1}\rangle}.$$

Thus, (9.4.14) reduces to

$$a_{k+1}, ji = \frac{\sigma(\mathcal{J}_t^{ij})}{\sigma(O_t^i)}.$$

Furthermore,

$$E_k\left[\int_0^t \langle X_u, e_i\rangle dr_u\middle|\mathcal{R}_t\right] = E_k[\mathcal{K}_t^i|\mathcal{R}_t] = \frac{\sigma(\mathcal{K}_t^i)}{\langle\sigma(X_t), \mathbf{1}\rangle}$$

$$E_k\left[\int_0^t r_u\langle X_u, e_i\rangle du\right] = E_k[\mathcal{I}_t^i|\mathcal{R}_t] = \frac{\sigma(\mathcal{I}_t^i)}{\langle\sigma(X_t), \mathbf{1}\rangle}$$

Hence, Equation (9.4.15) becomes

$$\alpha_{k+1,i} = \frac{\gamma^{-1}\sigma(\mathcal{K}_t^i) + \sigma(\mathcal{I}_t^i)}{\sigma(\mathcal{O}_t^i)}.$$

■

The Case of Extended Parameters

Suppose we extend the parameter estimation that includes the speed of adjustment γ. The estimate $a_{ij}, \quad 1 \leq i, j \leq n$ is the same. However, if we wish to include γ as a parameter to be as well estimated, the estimation procedure will yield a different $\frac{dP_{\hat{\theta}}}{dP_{\theta_k}}$. We shall see this later.

Lemma 9.4.2 *The new estimate for γ is given by*

$$\gamma_{k+1} = \frac{C}{D} \quad \textit{where}$$

$$C = \int_0^t r_u dr_u - \sum_{i=1}^{n} \frac{\hat{\mathfrak{J}}_t^i \hat{\mathfrak{K}}_t^i}{\hat{\mathfrak{O}}_t^i}$$

and

$$D = \sum_{i=1}^{n} \frac{(\hat{\mathfrak{J}}_t^i)^2}{\hat{\mathfrak{O}}_t^i} - \int_0^t r_u^2 du.$$

The new estimate for α_i is

$$\alpha_{k+1,i} = \frac{\rho_{k+1}^{-1} \hat{\mathfrak{K}}_t^i + \hat{\mathfrak{J}}_t^i}{\hat{\mathfrak{O}}_t^i}.$$

Proof: Including γ as a parameter that we wish to estimate would lead (9.4.11) take a different form. In this case, our parameter sets are now

$$\bar{\theta} = \{\bar{\gamma}, \quad \bar{a}_{ij}, \bar{\alpha}_i; \quad 1 \leq i, j \leq n\} \quad \text{and}$$

$$\theta_k = \{\gamma_k, \quad a_{k,ij}, \alpha_{k,i}; \quad 1 \leq i, j \leq n\} \quad \text{and}$$

$$\begin{aligned} \frac{dP_{\bar{\theta}}}{dP_{\theta_k}} &= exp\Bigg[\int_0^t (\bar{\gamma}\rho^{-2}(\langle X_u, \bar{\alpha}\rangle - r_u) \\ &\quad -\gamma_k \rho^{-2}(\langle X_u, \alpha_k\rangle - r_u))dr_u \\ &\quad -\frac{1}{2}\int_0^t (\bar{\alpha}^2 \rho^{-2}(\langle X_u, \bar{\alpha}\rangle - r_u)^2 \\ &\quad -\gamma_k^2 \rho^{-2}(\langle X_u, \alpha_k\rangle - r_u)^2)du\Bigg] \times R_1, \end{aligned} \tag{9.4.16}$$

where $\bar{\alpha} = \{\bar{\alpha}_1, \bar{\alpha}_2, ..., \bar{\alpha}_k\}$, $\quad \alpha_k = \{\alpha_{k,1}, \alpha_{k,2}, ..., \alpha_{k,n}\}$ and R_1 is a term that is independent of $\bar{\gamma}$ and $\bar{\alpha}$.

We get

$$
\begin{aligned}
Q(\bar{\theta}, \theta_k)\rho^2 &= \bar{\gamma}\left(E_k\left[\int_0^t \langle X_u, \bar{\alpha}\rangle dr_u \middle| \mathcal{R}_t\right]\right. \\
&\quad \left. -E_k\left[\int_0^t r_u dr_u \middle| \mathcal{R}_t\right]\right) \\
&\quad -\frac{1}{2}\bar{\gamma}^2\left(E_k\left[\int_0^t \langle X_u, \bar{\alpha}\rangle^2 du \middle| \mathcal{R}_t\right]\right. \\
&\quad -2E_k\left[\int_0^t r_u \langle X_u, \bar{\alpha}\rangle du \middle| \mathcal{R}_t\right] \\
&\quad \left. +E_k\left[\int_0^t r_u^2 du \middle| \mathcal{R}_t\right]\right) + R_2, \qquad (9.4.17)
\end{aligned}
$$

where again R_2 is independent of both $\bar{\gamma}$ and $\bar{\alpha}$.

Equation (9.4.17) could be written further as

$$
\begin{aligned}
Q(\bar{\theta}, \bar{\theta}_k)\rho^2 &= \bar{\gamma}\left[\sum_{i=1}^n \bar{\alpha}_i \hat{\mathcal{K}}_t^i - \int_0^t r_u dr_u\right] \\
&\quad -\frac{1}{2}\bar{\gamma}^2\left[\sum_{i=1}^n \bar{\alpha}_i^2 \mathcal{O}_t^i - 2\sum_{i=1}^n \bar{\alpha}_i \hat{\mathcal{J}}_t^i + \int_0^t r_u^2 du\right] + R_2 \qquad (9.4.18)
\end{aligned}
$$

In (9.4.18), we utilise the notation $\hat{\psi}$ as shorthand for the $\mathcal{R}$–optional projection of the process ψ under the measure P_{θ_k}. We also employ the notation for the processes $\mathcal{O}^i$, $\mathcal{K}^i$ and $\mathcal{J}^i$.

We maximise Q using (9.4.18) by equating the partial derivatives of Q with respect to $\bar{\alpha}_i$ and $\bar{\gamma}$ to 0. After simplification, the new estimate for γ is given by

$$\gamma_{k+1} = \frac{C}{D},$$

where

$$C = \int_0^t r_u dr_u - \sum_{i=1}^n \frac{\hat{\mathcal{J}}_t^i \hat{\mathcal{K}}_t^i}{\hat{\mathcal{O}}_t^i}$$

and

$$D = \sum_{i=1}^n \frac{(\mathcal{J}_t^i)^2}{\mathcal{O}_t^i} - \int_0^t r_u^2 du.$$

On the other hand, the new estimate for α_i is

$$\alpha_{k+1,i} = \frac{\gamma_{k+1}^{-1}\hat{\mathcal{K}}_t^i + \hat{\mathcal{J}}_t^i}{\hat{\mathcal{O}}_t^i}.$$

■

9.5 Application of the Self Calibrating Model

9.5.1 The Data and Estimation Procedure

In this section, we implement the theory and filtering techniques in the previous discussion for the Markovian mean reverting interest rate model. We analyse a data set consisting of 198 monthly observations on the yields of 3-month Canadian Treasury bills, 2-year and 10-year Canadian bonds. The sample period ran from June 1982 to December 1998. The data were compiled by the Bank of Canada, Department of Monetary and Financial Analysis. For further details on how the data were gathered and other related information, refer to Appendix D.

Parameter estimates were updated using the formulas of the previous section as soon as a new interest rate arrives.

We let n, the size of the state space of the Markov chain X, equal three; dt, the time step between observations be $\frac{30}{360}$ and the maturities of the various securities be $\frac{60}{360}$, 2 and 10.

9.5.2 The Choice of n

In the estimation procedure proposed above, parameter n, which represents the size of the state space of the Markov chain, is the only parameter which is not estimated. Rather, a value is assigned to n. In this application, we let $n = 3$.

The determination of the optimal value of n, for a particular data set, is an important problem which has been considered in the literature. Although this problem cannot be resolved using the likelihood ratio test, a number of proposals have been advanced to address it, see [61] p. 698-699.

We do not explore this issue further, other than to say that a comparison of results obtained can be made when n is assigned different values. However, it is interesting to note that in the regime-switching model, discussed in Hamilton, [60] and [61], in which the state or regime of a time series process is modelled as a Markov chain, a state space of size two is typically assumed.

The assignment of 3 states, is justified by our choice of designating the states of an economy either in a bad, medium or good situation.

9.5.3 Yield Estimation

At the end of each iteration through the data and with filters computed for each security, we shall calculate the yields based on the updated parameters. Figures 9.1, 9.2 and 9.3 show these results for the three securities considered.

9.5.4 The Unit Root Test

Macroeconomists have become aware of a new set of econometric difficulties that arise when one or more variables of interest may have unit roots in their time series representations. Standard asymptotic distribution theory often does not apply to regressions involving such variables, and the inference can go seriously astray if this is ignored. We are guided by this fact and any regression analysis, therefore, to be performed on the actual yields versus the estimated yields can only proceed after an evidence of stationarity of both series.

We shall use the standard Dickey-Fuller tests to detect nonstationarity. This approach uses formal statistical tests for unit roots. In carrying out these tests, we consider three autoregressive (AR) models. The first model is an

Figure 9.1 Average Returns on 3-month T-bill

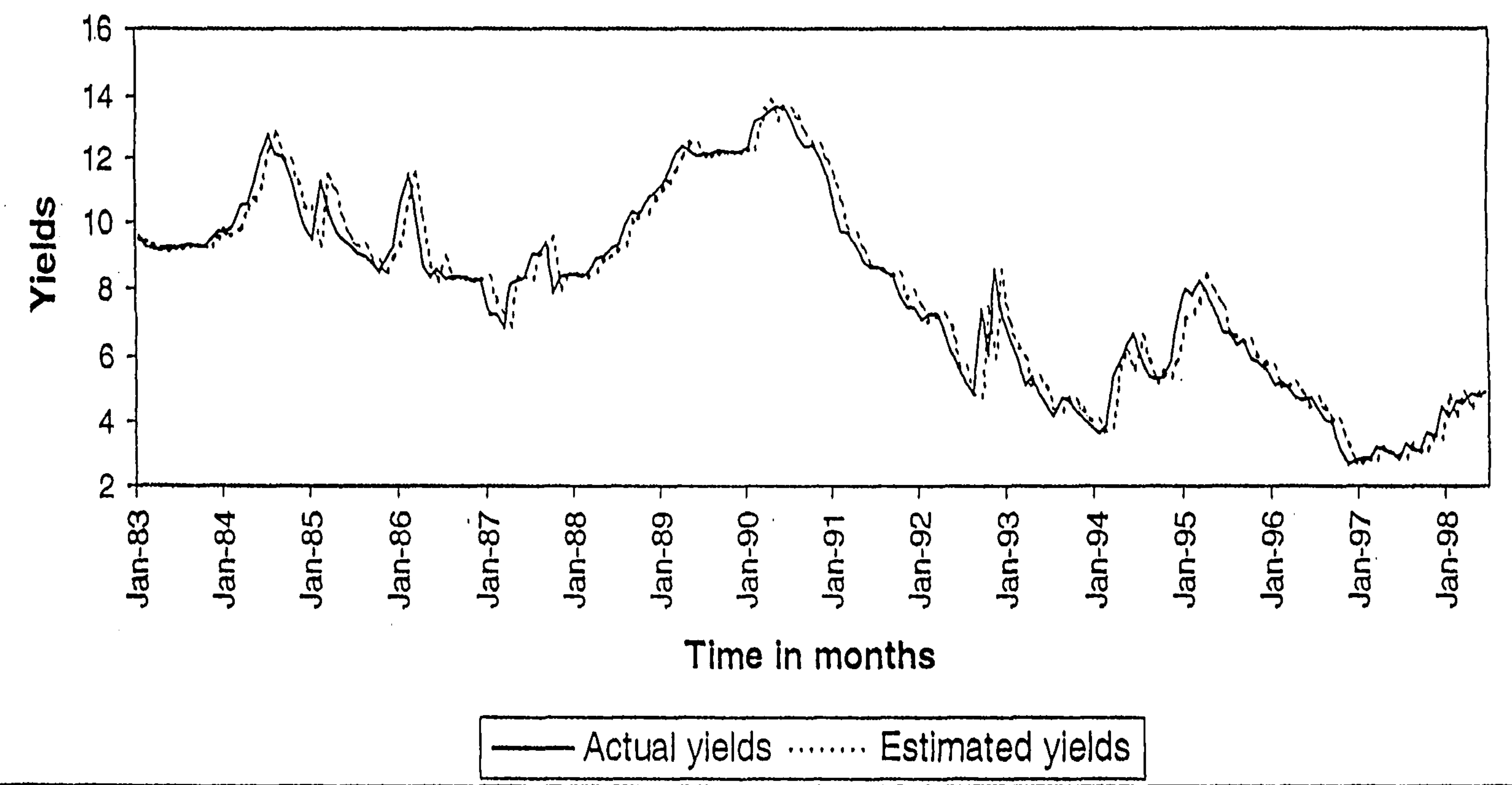


Figure 9.2 Average Returns on 2-year bond

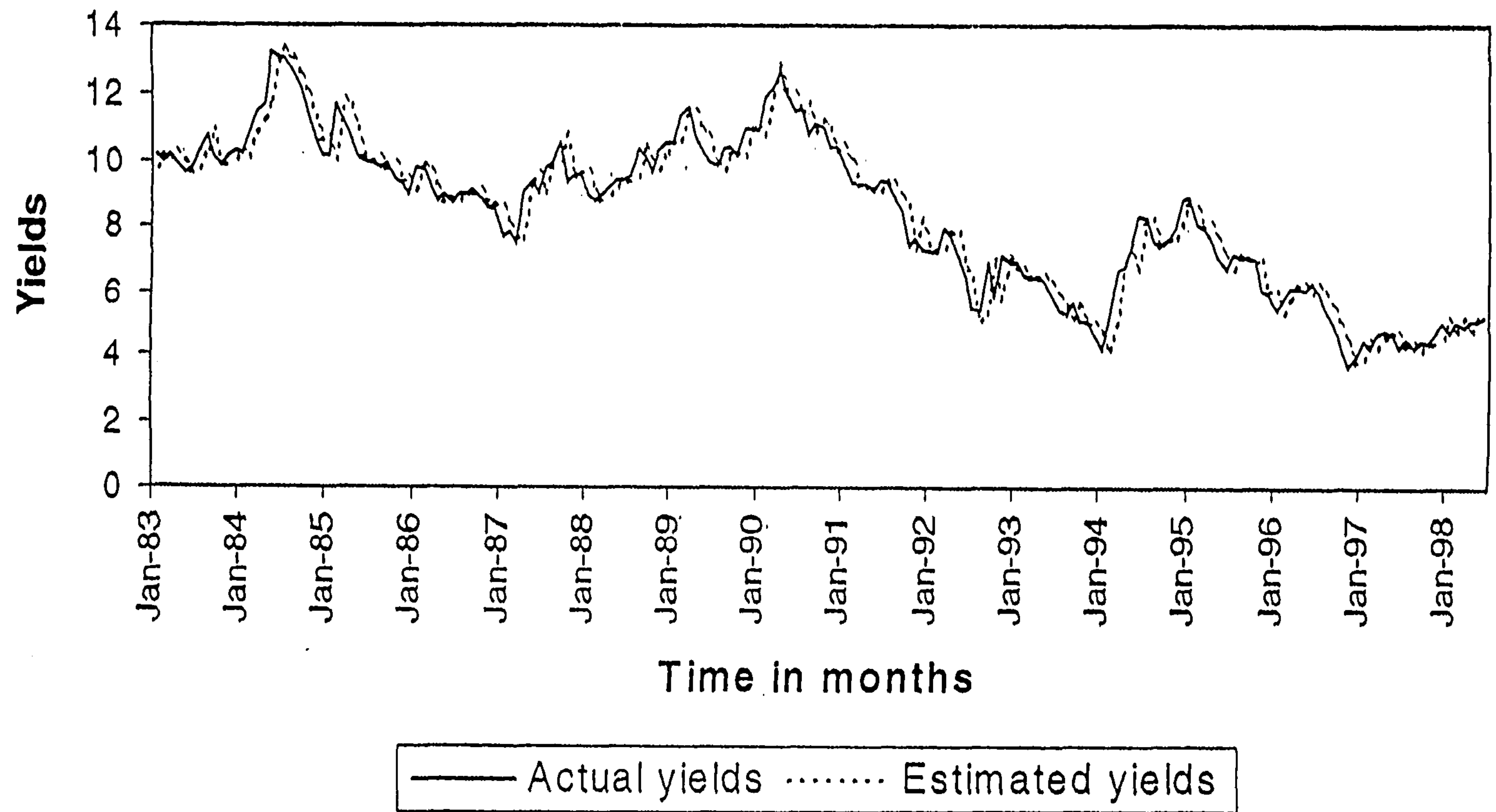


Figure 9.3 Average Returns on 10-year bond

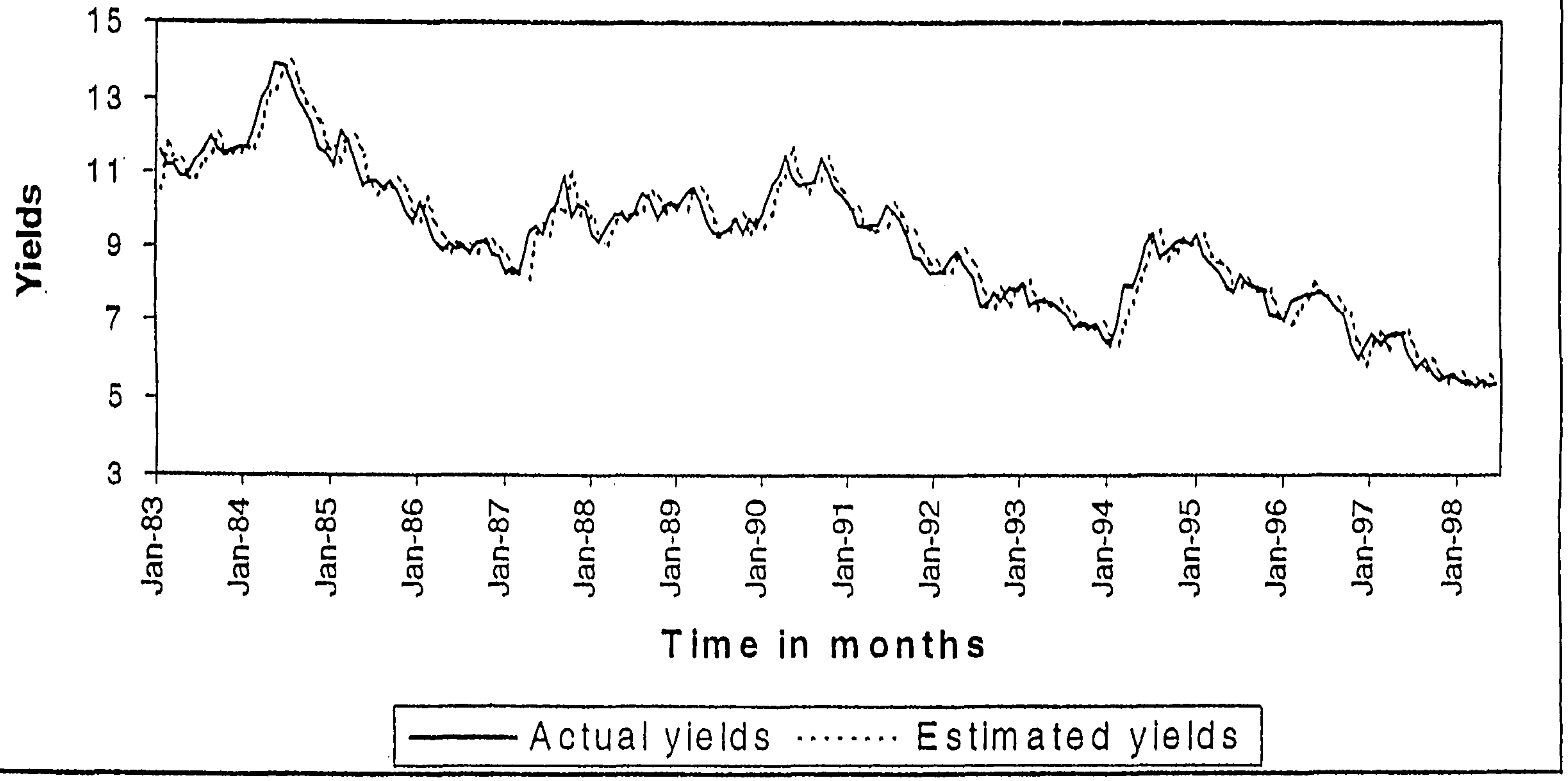


Figure 9.4: Residual Plot for 3-month T-bill returns

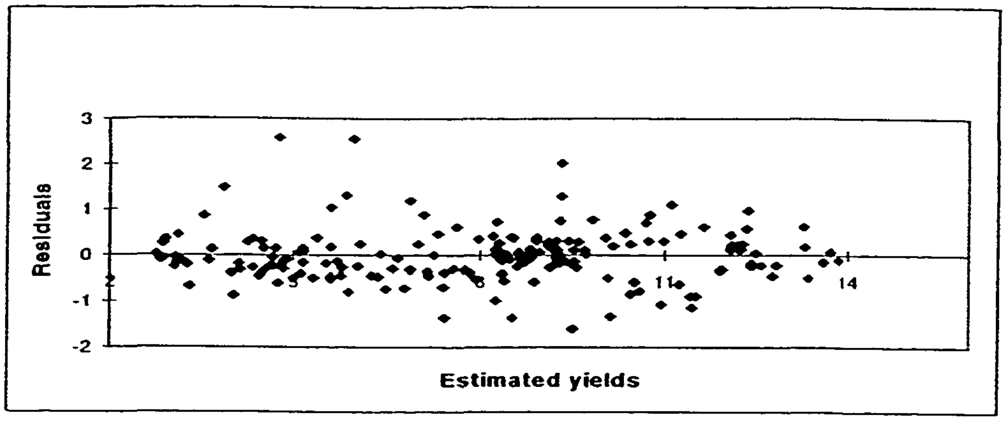


Figure 9.5: Residual Plot for 2-year T-bond returns

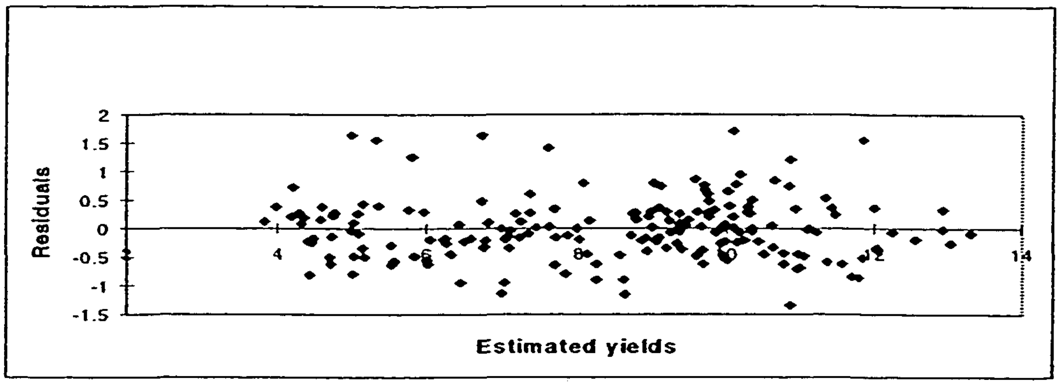
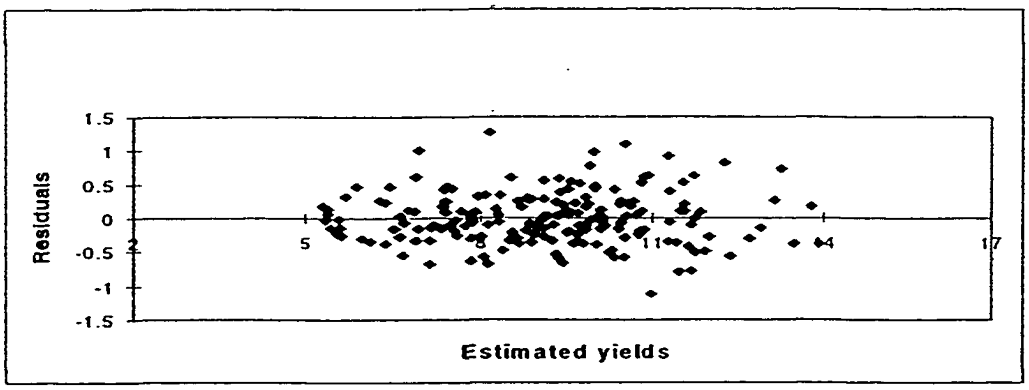


Figure 9.6: Residual Plot for 10-year T-bond returns



AR(1) process with no intercept (no constant). That is,

$$y_t = \alpha y_{t-1} + \epsilon_t, \tag{9.5.19}$$

where ϵ_t is white noise. We would like to test the hypothesis

$$\begin{aligned} H_0 : \alpha &= 1 \quad \textit{(model has a unit root, therefore nonstationary)} \\ &\text{versus} \\ H_1 : \alpha &< 1 \quad \textit{(model is stationary).} \end{aligned}$$

Model (9.5.19) is equivalent to

$$\Delta y_t = \gamma y_{t-1} + \epsilon_t \tag{9.5.20}$$

where $\gamma = \alpha - 1$.

The equivalent hypothesis test is now

$$\begin{aligned} H_0 : \gamma &= 0 \quad \textit{(model has a unit root, therefore nonstationary)} \\ &\text{versus} \\ H_1 : \gamma &< 0 \quad \textit{(model is stationary).} \end{aligned}$$

The result suggests running an OLS regression on Equation (9.5.20) and rejecting the null hypothesis if a significant negative values is found for $\hat{\gamma}$. Under the null hypothesis this reduces to $\Delta y_t = \epsilon_t$. So y_t is a random walk without drift and nonstationary.

The second model that we shall consider is an AR(1) model with a constant involved. That is,

$$\Delta y_t = b + \gamma y_{t-1} + \epsilon_t. \tag{9.5.21}$$

This has the same null and alternative hypothesis of the first model described in (9.5.19). When the null is true, Equation (9.5.21) reduces to $\Delta y_t = b + \epsilon_t$ so that y_t is a random walk with drift and thus nonstationary. The third model incorporates a constant and a time trend. Thus.

$$\Delta y_t = b + \gamma y_{t-1} + \xi t + \epsilon_t. \tag{9.5.22}$$

Table 9.1: Asymptotic critical values for unit root tests

Test statistic	1%	2.5%	5%	10%
τ_{nc}	-2.56	-2.23	-1.94	-1.62
τ_{c}	-3.43	-3.12	-2.86	-2.57
τ_{ct}	-3.96	-3.66	-3.41	-3.13

Adapted from Econometric Methods, by J. Johnston and J. DiNardo. The McGraw-Hill Companies, Inc., 4th edition, 1997.

Hence, there are three possible test regressions. Each has Δy as the regressand. In Equation (9.5.20) the only regressor is lagged y, in equation (9.5.21) a constant is included in the regressors and in Equation (9.5.22) there is a constant and a time trend in addition to lagged y.

We denote the three possible test statistics, $\frac{\hat{\gamma}}{s.e.(\hat{\gamma})}$, by τ_{nc}, τ_c, or τ_{ct} according to whether they come from Equation (9.5.20), Equation (9.5.21) or Equation (9.5.22). The relevant rows of Table 9.1 are indicated by these symbols.

For each of the above AR models we obtain estimates for $\hat{\gamma}$ and $s.e.(\hat{\gamma})$ derived from OLS.

The above results and analysis appear that we fail to reject the null hypothesis at 5% confidence level. In other words, given the Dickey-Fuller test applied to our series each consisting of 186 data points, there exist unit roots in each series. However, failure to reject a null hypothesis justifies at best only a cautious and provisional acceptance. We have to realise that low power in statistical tests is an often unavoidable fact of life with which one must live and not expect to be able to make definitive pronouncement.

Schwert (1987), Lo and MacKinlay (1989), Blough (1988) and others have documented that tests for unit roots or trend stationarity can have low power against some specific alternatives. Essentially, they show that tests for a unit root have low power in finite samples against the local alternative of a root

Table 9.2: Estimates of $\hat{\gamma}$ and $s.e.(\hat{\gamma})$ for each actual and estimated T-bills and T-bonds data series

Yield Data	Model 1	Model 2	Model 3
Actual 3-month T-bill returns	-0.00481 (0.00494)	-0.01744 (0.01463)	-0.04300 (0.25300)
Estimated 3-month T-bill returns	-0.00507 (0.00521)	-0.01926 (0.01548)	-0.04800 (0.22000)
Actual 2-year T-bond returns	-0.00447 (0.00431)	-0.02014 (0.01620)	-0.03100 (0.23800)
Estimated 2-year T-bond returns	-0.00451 (0.00474)	-0.02480 (0.01791)	-0.0300 (0.21700)
Actual 10-year T-bond returns	-0.00408 (0.00292)	-0.01300 (0.01462)	-0.0250 (0.20500)
Estimated 10-year T-bond returns	-0.00339 (0.00314)	-0.01424 (0.01586)	-0.0230 (0.20100)

Note: The numbers inside the parenthesis denote the standard error of the estimates.

Table 9.3: Estimated Values of Test Statistics for A Given Actual and Estimated Returns on T-Bills and T-Bonds

	$\hat{\tau}_{nc}$	$\hat{\tau}_{c}$	$\hat{\tau}_{ct}$	Decision
Actual 3-month T-bill returns	-0.97229	-1.19073	-0.16996	Accept H_0 at 5% confidence level
Estimated 3-month T-bill returns	-0.97224	-1.24419	-0.21818	Accept H_0 at 5% confidence level
Actual 2-year T-bond returns	-1.03760	-1.24313	-0.13025	Accept H_0 at 5% confidence level
Estimated 2-year T-bond returns	-0.95148	-1.38509	-0.13825	Accept H_0 at 5% confidence level
Actual 10-year T-bond returns	-1.39249	-0.88938	-0.12195	Accept H_0 at 5% confidence level
Estimated 10-year T-bond returns	-1.07927	-0.89814	-0.11443	Accept H_0 at 5% confidence level

close to but below unity.

Campbell and Perron [12] also pointed out the problem with the procedure that in finite samples, unit roots and stationary processes cannot be distinguished. For any unit root process, there are "arbitrarily close" stationary processes and vice versa. Conversely, take a stationary process and add to it a random walk with tiny innovation variance. That's a "close" unit root process. J.H. Cochrane [18] argued in the same spirit. Any test where a continuous parameter θ is equal to some value θ_0 has arbitrarily low power against alternatives $\theta_0 - \epsilon$ in finite samples. However, in most cases, the difference between θ_0 and $\theta_0 - \epsilon$ is not particularly important, from either a statistical or an economic perspective. What makes the unit root special is the impression that important statistical and economic issues hang on the difference between a root of precisely 1 and a root of $1 - \epsilon$ or between a random walk with component variance precisely 0 and a random walk component with innovation variance ϵ, in a way that say, an elasticity of demand of -1.0 is not importantly different from an elasticity of -0.99.

Empirical evidence shows that many or most aggregate economic time series contain a unit root, [28]. However, it is important to note that in this empirical work the unit root is the null hypothesis to be tested, and the way in which classical testing is carried out ensures that the null hypothesis is accepted unless there is strong evidence against it. An explanation therefore for the common failure to reject a unit root is simply that most economic time series are not informative about whether or not there is a unit root, or equivalently, the standard unit root tests are not very powerful against relevant alternatives.

A study of Rudebusch [106] shows that U.S. data on real GNP, which fails to reject the unit root hypothesis, also fails to reject a stationary hypothesis when the latter is set up as the null hypothesis. Several studies, [28], [98] and [30] suggest that, in trying to decide by classical methods whether economic data are stationary or integrated it would be useful to perform tests of the null hypothesis of stationarity as well as tests of the null hypothesis of a unit root. In light of the argument that the Dickey-Fuller statistics have low power in finite samples, we also look at the hypothesis wherein the null hypothesis is stationarity.

9.5.5 The Null Hypothesis of Stationarity

We shall use the results derived in the paper entitled "Testing the hypothesis of stationarity against the alternative of a unit root," by D. Kwiatkowski, P.C.B. Phillips, P. Schmidt and Y. Shin [82], to test the null hypothesis of stationarity. The authors basically propose a test of the null hypothesis that an observable series is stationary around a deterministic trend. Their assumption is that the series is expressed as the sum of deterministic trend, random walk and stationary error. The test is the Lagrange Multiplier (LM) or the Rao score test of the hypothesis that the random walk has zero variance. The asymptotic distribution of the statistic is derived under the null and under the alternative that the series is difference-stationary.

The one-sided LM statistic for the stationarity hypothesis was derived as a special case of the statistic developed by Nabeya and Tanaka, [91]. Let e_t, $t = 1, 2, \cdots, T$, be the residuals from the regression of y on an intercept and the time trend. Let $\hat{\sigma}_\epsilon^2$ be the estimate of the error variance from this regression (the sum of the squared residuals, divided by $T-2$). Define the partial sum process of the residuals:

$$S_t = \sum_{i=1}^{t} e_i, \quad t = 1, 2, \cdots, T.$$

Then the LM statistic is

$$LM = \sum_{t=1}^{T} S_t^2 / \hat{\sigma}_\epsilon^2.$$

Furthermore, in the event that we wish to test the null hypothesis of level stationarity instead of trend stationarity, we simply define e_t as the residual from the regression of y on an intercept only, that is, $e_t = y_t - \bar{y}$, instead of as above, and the rest of the construction of the test statistc is unaltered. The test is an upper tail test. For the test of both level-stationary and trend-stationary hypotheses, the denominator of the LM statistic is $\hat{\sigma}_\epsilon^2$, which converges in probability to σ_ϵ^2. However, when errors are no longer IID, the appropriate denominator of the test statistic is an estimate of σ^2 instead of σ_ϵ^2. To establish this, consider the numerator of the test statistic normalised by $(T-2)^{-2}$. That is,

$$\eta = (T-2)^{-2} \sum S_t^2. \tag{9.5.23}$$

This has an asymptotic distribution equal to σ^2 times the functional of a Brownian bridge. Let η_μ be defined as in (9.5.23), with subscript μ indicating that we have extracted a mean but not a trend from y. This implies that

$$\eta_\mu \to \sigma^2 \int_0^1 V(r)^2 dr.$$

Here, $V(r)$ is a standard Brownian bridge: $V(r) = W(r) - rW(1)$, where $W(r)$ is a Wiener process. The above convergence signifies weak convergence of the associated probability measures.

We divide η_μ by a consistent estimator of σ^2 to get the test statistic that we shall actually use. The test statistic, therefore, is

$$\hat{\eta}_\mu = \frac{\eta_\mu}{s^2(l)} = (T-2)^{-2} \sum S_t^2 / s^2(l).$$

The estimator $s^2(l)$ is of the form

$$s^2(l) = T^{-1} \sum_{t=1}^{T} e_t^2 + 2T^{-1} \sum_{s=1}^{l} w(s,l) \sum_{t=s+1}^{T} e_t e_{t-s}.$$

Here, $w(s,l)$ is an optional weighting function that corresponds to the choice of a spectral window. We shall use the Bartlett windows $w(s,l) = 1 - \frac{s}{l+1}$.

The trend-stationary case is similar to that of the level-stationary. We let η_τ be defined as in (9.5.23), where the subscript τ indicates that we have extracted a mean and a trend from y, and serves to distinguish the trend-stationary case from the level-stationary case. The authors showed that its asymptotic distribution is given by

$$\eta_\tau \to \sigma^2 \int_0^1 V_2(r)^2 dr.$$

Table 9.4: Upper critical values for $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$

	Upper tail percentiles of the distribution of $\int_0^1 V(r)^2 dr$			
Critical level	0.10	0.05	0.025	0.01
Critical value	0.347	0.463	0.574	0.739
	Upper tail percentiles of the distribution of $\int_0^1 V_2(r)^2 dr$			
Critical level	0.10	0.05	0.025	0.01
Critical value	0.119	0.146	0.176	0.216

Here $V_2(r)$ is the second level Brownian bridge. Further,

$$\hat{\eta}_\tau \to \int_0^1 V_2(r)dr.$$

Table 9.4 exhibits the upper critical values for the test statistics $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$.

We apply the above results and generate Table 9.5. All decisions are based on a 5% confidence level.

The above analysis shows that we cannot reject either the unit root hypothesis or the trend stationary hypothesis, and it is not clear what to conlude. The data are not sufficiently informative to distingush between these hypotheses. Presumably other alternatives, such as fractional integration or stationarity around a nonlinear trend (which is a reminiscent behaviour for interest rates with mean reverting properties) could be considered. In fact, Sims (1989), Campbell and Perron emphasized the fact that the unit roots are indistinguishable from nonlinear trends.

Table 9.5: Estimated test statistics for $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ with truncation parameters $l = 4$ and $l = 8$

	$\hat{\eta}_\mu$	$\hat{\eta}_\tau$	**Decision**
Actual 3-month T-bill returns	2.95356E-05 2.95343E-05	2.95400E-05 2.95432E-05	Accept H_0 at 5% level
Estimated 3-month T-bill returns	2.95356E-05 2.95344E-05	2.95400E-05 2.95435E-05	Accept H_0 at 5% level
Actual 2-year T-bond returns	2.95347E-05 2.95325E-05	2.95418E-05 2.95467E-05	Accept H_0 at 5% level
Estimated 2-year T-bond returns	2.95352E-05 2.95336E-05	2.95432E-05 2.95495E-05	Accept H_0 at 5% level
Actual 10-year T-bond returns	2.95391E-05 2.95413E-05	2.95472E-05 2.95576E-05	Accept H_0 at 5% level
Estimated 10-year T-bond returns	2.95343E-05 2.95317E-05	2.95341E-05 2.95314E-05	Accept H_0 at 5% level

Note: Numbers above are for lag l=4 and numbers below are for lag l=8.

9.5.6 Regression Analysis

In [12], J.H. Cochrane commented that so long as one does not get too creative with breaking trends and structural shifts, any test will show that interest rates have unit roots, and lag selection procedures indicates near random walk structure. The model does quite well for one-step ahead forecasting. Yet, interest rates are almost certainly stationary in levels. To quote Cochrane, "Interest rates were about 6% in ancient Babylon; they are about 6% now." The chances of a process with a random walk component displaying this behaviour are infinitesimal. One way to make this argument more formal, as pointed out by Stan Fischer [12] is to calculate

$$P(|r_{2000}| < 100\% | r_{4000BC} = 6\%).$$

This probability is infinitesimal if interest rates are or contain a random walk; it is near one if interest rates are an AR(1) with coefficient 0.99

We then propose to perform regression on the assumption that our data on the T-bills and T-bond returns are stationary in levels following the methodology of Pearson and Sun, [96]. We regress actual yields on estimated yields, using the model:

$$\boxed{\text{Actual Yields} = \alpha + \beta * \text{Estimated Yield} + \epsilon}$$

The regression results obtained were assessed on the basis of the following criteria proposed by Fama and Gibbons, [51]:

1. conditional unbiasedness, that is, an intercept, α, close to zero, and a regression coefficient, β, close to one;
2. serially uncorrelated residuals; and
3. a low residual standard error.

Table 9.6 reports these regression results.

9.5.7 Results and Analysis

For each of the securities, we can conclude that the intercept is zero and the slope is one. Based on Table 9.6 and on the basis of the Durbin-Watson test and the plot of residuals, we can also conclude that the residuals do not display first-order serial correlations. In the filtering procedure, the algorithm starts to stabilize after 5 time periods. This means that the algorithm learns to adapt quickly to a given dynamics of the process we are trying to model.

In principle the techniques used in this chapter can be applied to financial time series such as yield rates. We based this conclusion on the criteria mentioned above. The interest model is characterised by a finite state Markov chain in combination with a conditionally Gaussian observation process. With this, it is straightforward to compute bond prices and other interest rate derivatives as solutions of the corresponding integro-differential equations which is another area of further research.

Table 9.6: Regression of actual yields on estimated yields for 3-month T-bills, 2-year and 10-year bonds

Parameter	Term to Maturity of Security		
	3-month	2-year	10-year
α	0.133759605 (0.127169909)	0.181577777 (0.14747549)	0.088858565 (0.01492704)
β	0.97849351 (0.014935499)	0.973449048 (0.016878195)	0.987562393 (0.014926704)
R-Squared	0.958893383	0.94758432	0.959660174
Durbin-Watson D statistic	2.003373533	1.906014747	1.805502099

Chapter 10

Concluding Remarks

With the attempt of central banks to develop a set of indicators derived from government bond prices that could be used to guide monetary policy and along with the rapid growth in interest rate derivatives, the demand for a "good" term structure model has become the major task of both academics and practitioners. A "good" term structure model needs to be able to accomplish two major missions.

First, it needs to be able to fit the current market data. It is in this light that we carried out the implementation of a model via filtering. Empirical evidence in forecasting and modelling yield curve favours the use of filtering techniques over the time series analysis methodology because the parameters obtained by the former method are optimal and in turn, could capture new information efficiently as they arrive. See, for example [15] and [7].

Second, it needs to be able to reflect fundamental economic conditions. Afterall, interest rates are not exogenous financial variables and should be determined within the economy. This is the primary reason why a survey of related theory in economics and finance were presented. The comprehensive review aims to give an outline of the economic forces affecting the term structure of interest rates and the interaction of these forces.

In the course of this study, we derive the forward rate dynamics starting with a model described by its short rate. This enables us to determine the information structure inherent in the bond prices.

Several single-factor models were presented and studied. Two-factor models were also investigated.

The drawback about single-factor models is that they are oversimplified to perfectly fit the market data, no matter how one changes the combination of parameters. Then, why should we believe that prices calculated by these models should reflect the "right" price? The answer is *no,* we don't. All of these are just approximations. In fact, theoretical models are seldom used in pricing contingent claims. If this is so, then why do we study them here?

The answer is *hedging.* We know that in finance, we do not question market prices because they are determined by smartest people, the traders and portfolio managers. Again trusting the market prices is all that we can do and we rely on this as this is all the essence of Efficient Market Hypothesis. Models should be able to match these market prices. Parameters in the models are set so that they produce market prices. Once the market prices are matched, we look at the models and ask what hedge ratios these models tell us.

If a model, even though simple enough, can capture the most important characteristics of the underlying risks, hedges suggested by the model will be robust and reliable. If a model does not have a closed-form solution, then the hedge ratio needs to be computed numerically. Sometimes this is slow and may not be good enough for traders. That is why closed-form solutions are important, not because they are elegant, but because they are fast.

In term structure modelling, empirical evidence showed that single-factor models do not really fit the yield curve. The empirical studies that show one-factor models cannot fit the yield curve are Chen and Scott (1993) and Pearson and Sun (1990) for maximum likelihood; Heston (1989) and Gibbons and Ramaswamy (1993) for generalised method of moments; Litterman and Scheinkman (1991) for factor analysis.

Research since then has been on developing a term structure model that can well describe the curvature of the observed term structure. There are two approaches, one is to add some flexibility in the model so that fitting curves becomes no problem. Ho and Lee (1986), Hull and White (1990) and Heath, Jarrow and Morton (1992) follow this approach. Another approach is to allow more than one state variable for the term structure. This approach is used by Langtieg (1980), Chen and Scott (1992) and Longstaff and Schwartz (1992). This should create more flexibility and should improve the fitting.

We were guided by this approach in considering two-factor models where the fundamental characteristics of mean reversion is incorporated.

Having therefore justified the validity of this research pursuit in reference to the theoretical point of view and applicability considerations of the industry sector, we identify the valuable contributions embodied in this study.

10.1 Main Contributions

1. **Development of a model with a Markovian mean reversion level.** This model blends continuous and discrete processes. The model is a Vasicek model and the mean reverting level is described by the semimartingle form of a Markov chain. A closed-form solution for the bond price is obtained involving a fundamental matrix. This model could serve as a model for the logarithm of an asset price, or in our case, an interest rate where a central bank or regulatory board provides a reference rate that changes from time to time.

2. **Implementation of the model described in the preceding paragraph using HMM and filtering techniques.** We derived a finite dimensional filter for the unobservable state of the Markov chain based on observations of the mean reverting diffusion process. Various auxiliary filters are developed that allow us to estimate the parameters of the Markov chain.

   The filtering methods we used to perform the empirical test of the model provide a continual, recursive update of optimal estimates in contrast to the static model-fitting of maximum likelihood. The application of

Hidden Markov filtering and estimation techniques appears new and makes use of the results generated by Elliott, Aggoun and Moore, see [36]. Further, we do not need to specify a priori the dynamics of the mean-reversion level, other than to say it is Markov chain.

3. **Formulation of the mathematical framework for the n-factor Gaussian interest rate models.** In particular, we presented two examples when n=2 and employed methodology of stochastic flows and forward measures to derive the bond price.

4. **Dual Approach in studying term structure.** This joint approach of specifying the dynamics of the short rate and forward rate and reconciling the two forms in all the models studied is the distinct feature of this research. This unique way of dealing with term structure theory is guided by two principles in bond pricing; specifically, the bond valuation formula is either expressed in terms of the short rate employing a risk-neutral measure or in terms of the HJM pricing methodology.

   The joint short rate/HJM approach is performed starting from a model where the short rate is a function of a Markov chain with discrete state space in continuous time. Then single continuous models are explored such as the Vasicek's model which is a version of Ornstein-Uhlenbeck process and the CIR model which is a representative of the Bessel process.

5. **Investigation of a General class of Exponential Affine models.** The joint short rate/HJM approach is extended to generalised exponential affine models. We derived necessary and sufficient conditions in order for a model to be classified in the affine yield category.

   In addition, general conditions were obtained for the deterministic functions of the drift and volatility component in the light of reconciling the short rate and HJM forms. This result is exemplified by the Vasicek and CIR models.

6. **Presentation of Hull-White model with a dual objective.** On one hand, this model is considered to illustrate how the joint approach can be

performed when the model's parameters are no longer constants but time varying. In other words, this becomes an extension of the exercise done for Vasicek and CIR models. Though the application of the apporach does not culminate in this model, such application have given us insights on how to further extend the approach to other general models whose short rate dynamics are given and the parameters are no longer constant.

On the other hand, the Hull-White model provides an impetus to the development of more general mean reverting models. We consider a Vasicek model whose mean reversion level has dynamics of its own. One specific case is that the dynamics of the mean reversion level is Vasicek by itself and the other case, the mean reverting level is a Markov process.

7. **A mathematical proof of the Expectations theory via a forward measure approach.** This formalises the principal finding of Meiselman and other proponents of this theory that a relationship exists between expectations of future short-term rates and forward rates. Indeed, the fact that forward rates incorporate predictions of future short-term rates with an appreciable accuracy in a statistical sense, demonstrates, by a fortiori argument, that forward rates are functions of expected spot rates.

8. **Re-interpretation of the expectation involved in contingent claim valuation.** This is a remarkable result which was derived through the interplay of forward measure and application of Bayes' theorem. Thus, the expectation in the valuation formula can be expressed as a product of two simpler expectations. Forward measure approach is a valuable tool that can facilitate the computation of an expectation problem especially when the contingent claim has a complex form.

9. **A survey of economics theory regarding the relationship between short rates and forward rates.** We are taking into account that it is the integration of sound economic and financial theories with appropriate mathematical tools that forms the cornerstone of today's financial modelling. Afterall, our term structure models should be grounded on economic fundamentals and principles.

10. **Presentation of related literature on the theory of interest rates starting from classical theory to recent works on "modern" term structure theory.** This provides an assessment of how the field advances and what are the main problems researchers considered and remedied so far and have been looking at recently.

11. **Review of selected mathematical concepts and theories appropriate for the study of term structure models.** This review has laid down the groundwork in setting a mathematical framework for the analysis of bond structure and interest rate market dynamics. We undertake this effort to establish a solid foundation of mathematical modelling in direct response to the profound scientific challenges posed by this area of finance. For emphasis, we point out that the area has both stimulated and benefitted from advances in a range of mathematical sciences, most notably, probability, differential equations, statistics, optimisation and numerical analysis.

10.2 Future Directions

One of the main contributions of this thesis is the extensive study of mean reverting interest rate models. We have worked though on the framework of the Vasicek model and all results and analysis were deduced from this framework.

We wonder about the corresponding analysis and results that will be generated if we extend this work on mean reversion to the CIR framework where the interest rate process has dynamics which is a version of the Bessel process.

Having results from this research at our disposal and extending this research to CIR model, we could offer alternative models of term structure with focus on monetary aspects. Analysis described in the paper of S. Babbs and N. Webber could be based on the results of the mean reverting models of Chapter 8 and a determination of their impacts on monetary regimes could be carried out.

Throughout the entire study, we have concentrated on term structure explorations based on bond prices and made an empirical investigation of implementing interest rate models based on bond prices especially government indexed instruments.

As the market of derivatives securities is becoming huge, currently estimated to be fifteen trillion dollars, it is apparent that this market can implicitly tell us term structure information based on prices of interest rate sensitive contingent claims.

In other words, instead of considering only the bond markets we are interested to know what happens when the market we are going to study includes interest rate derivative products. An example would be a characterisation of exchange-traded interest rate options on Treasury bond futures, Treasury note futures and Eurodollar futures.

Further, one of the fundamental determinants in the Black-Scholes pricing equation is the volatility of the contingent claim's underlying variable. For valuation of interest rate derivatives, we wonder how a dynamic volatility can be constructed especially based on the mean reverting models that we develop.

Finally, a new direction in interest rate modelling is to employ techniques commonly used in non-linear analysis. A simple example is a two-factor model where a certain parameter has a dynamical behaviour. This two-factor model extends to a three-factor non-linear model equivalent to the Lorenz system of differential equations, [115]. Perhaps, by using non-linear analysis in capturing the non-linear properties and features of the interest rate process, more insights and developments can be found in understanding and modelling interest rate dynamics.

Bibliography

[1] D.E. Allen, M. Manzur and N. Morkel-Kingsbury. *Interest Rate Term Premia and Purchasing Power Parity Deviations: The Missing Link?* Working paper, Curtin University of Technology and Edith Cowan University, Australia, 1999.

[2] N. Anderson, F. Breedon, M. Deacon, A. Derry, and G. Murphy. *Estimating and Interpreting the Yield Curve.* John Wiley & Sons, New York, N.Y., 1996.

[3] R.B. Ash. *Real Analysis and Probability.* Academic Press, New York and London, 1972.

[4] A. Bhargava. *On the theory of Testing for Unit Roots in Observed Time Series.* Review of Economic Studies, **53**: 369-384, 1986.

[5] M. Baxter and A. Rennie. *Financial Calculus. An Introduction to Derivative Pricing.* Cambridge University Press, Cambridge, 1996.

[6] T. Bjork, Y. Kabanov, and W. Runggaldier. *Bond Market Structure in the Presence of Marked Point Processes.* Math Finance, **7**: 211-239, 1997.

[7] D.J. Bomhoff. *Financial Forecasting.* Academic Press Inc., San Diego and London, 1994.

[8] A. Brace, D. Gatarek, and M. Musiela. *The market model of interest rate dynamics.* Math Finance, **7**: 127-156, 1997.

[9] P.Brémaud. *Point Processes and Queues.* Springer-Verlag, 1981.

[10] S.J. Brown and P.H. Dybvig. *The Empirical Implications of the Cox, Ingersoll, Ross theory on the Term structure of interest rates.* Journal of Finance, **66:** 617-632, 1986.

[11] J.Y. Campbell. *A Defence of Traditional Hypotheses about the term structure of interest rates.* Journal of Finance, **61:** 183-193, 1986.

[12] J.Y. Campbell and P. Perron. *Pitfalls and Opportunities: What Macroeconomist Should Know About Unit Roots.* NBER Macroecomics Annual, 141-200, 1991.

[13] J.Y. Campbell and R.J. Shiller. *A simple Account of the Behaviour of Long-Term Interest Rates.* American economic Review, **74:** 44-48, 1994.

[14] K.C. Chan, G.A. Karolyi, F.A. Longstaff and A. Sanders. *An Empirical Comparison of alternative models of the short-term interest rate.* Journal of Finance, **47:** 1200-1227, 1992.

[15] R. Chen. *Understanding and Managing Interest Rate Risks.* Series in Mathematical Finance, **Vol. 1.,** New Jersey, 1996.

[16] Citicorp. *Interest Rate Spreads and Market Analysis.* Irving Professional Inc., Chicago, London and Singapore, 1996.

[17] K. Clinton. *The Term Structure of Interest Rates as a leading indicator of economic activity.* Bank of Canada Review, Winter 1994-95.

[18] J.H. Cochrane. *A Critique of the Application of Unit Root Tests.* Journal of Economic Dynamics and Control, **15:** 275-284, 1991.

[19] J.W. Conard. *An Introduction to the Theory of Interest.* University of California Press, Part III, 1959.

[20] J.W. Conard. *The Behaviour of Interest Rates.* National Bureau of Economic Research: Columbia University Press, New York and London, 1966.

[21] G.M. Constantinides. *A Theory of the nominal term structure of interest rates.* Review of Financial Studies, **5:** 531-552, 1992.

[22] A. Cornyn, J. Lederman, and R. Klein. *Controlling and Managing Interest Rate Risk.* New York Institute of Finance, 1997.

[23] J. Cox, J. Ingersoll and S. Ross. *A re-examinatin of the traditional hypothesis about the term structure of interest rates.* Journal of Finance, **4,** 1981.

[24] J. Cox, J. Ingersoll and S. Ross. *A theory of the term structure of interest rates.* Econometrica, **53:** 385-407, 1985.

[25] J. Cox, J. Ingersoll and S. Ross. *An Intemporal General Equilibrium Model of Asset Prices.* Econometrica, **53, No.2:** 363-384, 1985.

[26] R.C. Dalang, A. Morton and W. Willinger. *Equivalent martingale measures and no-arbitrage in stochastic securities market models.* Stochastics Stochastic Rep., 29: 185-201, 1990.

[27] M.A. Dempster and S.R. Pliska. *Mathematics of Derivatives Securities.* Proceedings of Mathematical Finance Programme, Jan.-June 1995, Isaac Newton Institute for Mathematical Sciences, Cambridge University Press, 1997.

[28] D.N. DeJOng, J.C. Nankevris, N.E. Savin and C.H. Whiteman. *Integration versus trend stationarity time series.* 1989.

[29] D.A. Dickey, W.R. Bell and R.B. Miller. *Unit Roots in Time Series Models: Test and Implications.* American Statistician, **40:** 12-26, 1986.

[30] F.X. Diebold and G.D. Rudebusch. *On the power of Dickey-Fuller tests against fractional alternatives.* Economic Letters, **35:** 155-160, 1991.

[31] J.C. Dodds and J.L. Ford. *Expectations, Uncertainty and the Term Structure of Interest Rates.* London, 1974.

[32] L.U. Dothan. *On the term structure of interest rates.* Journal of Financial Economics, **6:** 59-69, 1978.

[33] D. Duffie. *Dynamic Asset Pricing Theory.* Princeton University Press, Princeton, 1992.

[34] D. Duffie and R. Kan. *A Yield-Factor Model of Interest Rates.* Math Finance **6:**379-406, 1996.

[35] R.J. Elliott. *Stochastic Calculus and Applications.* Springer-Verlag, Berlin and New York, 1982.

[36] R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov Models: Estimation and Control.* Applications of Mathematics **29.** Springer-Verlag, Berlin-Heidelberg-New York, December 1994.

[37] R.J. Elliott, W.C. Hunter and B.M. Jamieson. *Financial Signal Processing.* Math Finance (to appear).

[38] R.J. Elliott and P.E. Kopp. *Mathematics of Financial Markets.* Springer-Verlag, Berlin-Heidelberg-New York, May 1998.

[39] R.J. Elliott, P. Fisher, and E. Platen. *Filtering and parameter estimation for a mean reverting interest rate model.*

[40] R.J. Elliott and E. Platen. *Hidden Markov Chain Filtering for Generalised Bessel Processes.*

[41] R.J. Elliott and J. van der Hoek. *Stochastic Flows and the Forward Measure.*

[42] A. Estrella and G.A. Hardouvelis. *The Term Structure as a predictor of real economic activity.* Journal of Finance, **46:** 555, 576, 1991.

[43] G.B.A. Evans and N.E. Savin. *Testing for Unit Roots 1.* Econometrica, **49:** 753-779, 1984.

[44] G.B.A. Evans and N.E. Savin. *Testing for Unit Roots 2.* Econometrica, **53:** 241-1269, 1984.

[45] E.F. Fama. *Short-term interest rates as predictors of inflation.* American Economic Review, **65:** 269-286, 1975.

[46] E.F. Fama. *Foundations of Finance.* Book Inc, Publishers. New York, 1976.

[47] E.F. Fama. *Interest Rates and Inflation.* American Economic Review, **67:** 487-496, 1977.

[48] E.F. Fama. *Term Premiums in Bond Returns.* Journal of Financial Economics, **13:** 529-546, 1984.

[49] E.F. fama. *Inflation Uncertainty and Expected Returns on Treasury Bills.* Journal of political Economy, **84:** 427-448, 1976.

[50] E.F. Fama and R.R. Bliss. *The Information in Long-Maturity Forward Rates.* American Economic Review, **77:** 680-692, 1987.

[51] E.F. Fama and M.R. Gibbons. *A Comparison of inflation forecasts.* Journal of Monetary Economics, **13:** 327-348, 1948.

[52] L.S. Fan. *A Study in the maturity of interest rates.* Ph.D. thesis, University of Minnesota, 1965.

[53] I. Fisher. *The Rate of Interest.* The Macmillan Company, New York, 1907.

[54] I. Fisher. *The Theory of Interest.* The Macmillan Company, New York, 1930.

[55] K. Garbade and P. Wachtel. *Time Variation in the Relationship between Inflation and Interest Rates.* Journal of Monetary Economics, **4:** 755-765, 1978.

[56] M.R. Gibbons and K. Ramaswamy. *A test of the Cox, Ingersoll and Ross Model of the Term Structure.* Review of Financial Studies, **6:** 619-658, 1993.

[57] J.Hale. *Ordinary Differential Equations.* J. Wiley & Sons/Interscience, New York, 1969.

[58] H. Geman and M. Yor. *Bessel Processes, Asian Options and Perpetuities.* Math Finance, **4:**349-375, 1993.

[59] G. Grimmett and D. Stirzaker.*Probability and Random Processes.* Oxford University Press, 2nd edition, 1992.

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

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

[62] J.M. Harrison and D.M. Kreps. *Martingales and arbitrage in multiperiod securities markets.* Journal of Economic Theory, **20:**381-408, 1979.

[63] J.M. Harrison and S.R. Pliska. *Martingales and Stochastic Integrals in the Theory of continuous trading.* Stochastic Processes and their Applications, **11:** 215-260, 1981.

[64] J.M. Harrison and S.R. Pliska. *A Stochastic Calculus model of continuous trading: Complete Markets.* Stochastic Process and their Applications, **15:** 313-316, 1983.

[65] A. Harvey and R. Campbell. *The Term Structure and world economic growth.* Journal of Fixed Income, **1:** 7-19, 1991.

[66] D. Heath, R. Jarrow and A. Morton. *Bond pricing and the term structure of interest rates: A discrete time approximation.* Journal of Financial and Quantitative Analysis, **25:** 419-440, 1990.

[67] D. Heath, R. Jarrow and A. Morton. *Bond Pricing and the term structure of interest rates. A New Methodology for contingent claims valuation.* Econometrica, **60:** 77-105.

[68] T. Ho and S. Lee. *Term structure and Pricing Interest Rate Contingent Claims.* Journal of Finance, **41:** 1011-1029, 1986.

[69] R. Holmes. *Opportunities for mathematicians in finance and investments.* HIM&R, Inc., Massachusetts, 1997.

[70] S.D. Howison, F.P. Kelly and P. Wilmott. *Mathematical Models in Finance.* Proceedings of the Royal Society Meeting, Chapman and Hall, London, 1995.

[71] J. Hull. *Options, Futures and Other Derivative Securities.* Prentice Hall, Englewood Cliffs, 1991.

[72] J. Hull. *Introduction to Futures and Options Markets.* Prentice Hall, Englewood Cliffs, 1995.

[73] J. Ingersoll. *Theory of Financial Decision Making.* Rowman and Littlefield Publishers, Inc., U.S.A., 1987.

[74] J. Jacod and A.N. Shiryayev. *Limit Theorems for Stochastic Processes.* Grundlehren der Math. Wissenschaften **288.** Springer-Verlag, Berlin, Heidelberg, New York, 1987.

[75] B. Jamieson. *Applications in Finance of Hidden Markov Models.* M.Sc. Thesis, University of Alberta, 1996.

[76] J. Johnston and J. DiNardo. *Econometric Methods.* The McGraw-Hill Companies, Inc., New York, 4th edition, 1997.

[77] D.J. Jüttner. *Interest Rates.* Papers and Proceedings of a Symposium, Centre for Studies in Money, Banking and Finance, Macquarie University, Longman Cheshire Pty. Ltd., Melbourne, 1981.

[78] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, Berlin and New York, 1988.

[79] S.G. Kellison. *The Theory of Interest.* Richard D. Irwin Inc., Illinois, 2nd edition, 1991.

[80] D.P. Kennedy. *Characterising Gaussian models of the term structure of interest rates.* Math Finance, **7**: 107-118, 1997.

[81] R.A. Kessel. *The Cyclical Behaviour of the Term Structure of Interest Rates.* National Bureau of Economic Research: Columbia University Press, London and New York, 1965.

[82] D. Kwiatkowski, P.C.B. Phillips, P. Schmidt and Y. Shin. *Testing the Null Hypothesis of Stationarity against Alternatives of Unit Root: How*

Sure Are We that Economic Time Series Have a Unit Root? Journal of Econometrics, **54:** 159-178, 1992.

[83] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance.* Chapman & Hall, London, 1995.

[84] J.B. Long. *Stock, Prices, Inflation and the term structure of interest rates.* Journal of Financial Economics, **1:** 131-170, 1974.

[85] A.F. Lutz.*The Structure of Interest Rates.* American Economic Association: Readings in the Theory of Income Distribution, Philadelphia, 1946.

[86] J.H. McCulloch. *An Estimate of the Liquidity Premium.* Journal of Political Economy, **83:** 95-119, 1975

[87] D. Meiselman. *The Term Structure of Interest Rates.* Englewood Cliffs, New Jersey, 1962.

[88] R.C. Merton. *Continuous Time Finance.* Basil Blackwell Inc., London, 1990.

[89] J. Michaelson. *The Term Structure of Interest Rates.* Intext, Educational Publishers, New York and London, 1973.

[90] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling.* Applications of Mathematics **36**, Springer-Verlag, New York, 1997.

[91] S. Nabeya and K. Tanaka. *Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative.* Annals of Statistics, **16:** 218-235, 1988.

[92] C.R. Nelson. *The Term Structure of Interest Rates.* Basic Books Publishers Inc., New York and London, 1972.

[93] C.R. Nelson and C.I. Plosser. *Trends and Random walks in Macroecomic Time Series.* Journal of Monetary Economics, **10:** 139-162, 1982.

[94] B. Øksendal. *Stochastic Differential Equations.* Springer-Verlag, New York, 4th edition, 1995.

[95] G.S. Oldsfield and R.J. Rogalski. *The stochastic Properties of Term Structure Movements.* Journal of Monetary Economics, **19:** 229-254, 1987.

[96] N.D. Pearson and T.S. Sun. *Exploiting the Conditional Density in Estimating the Term Structure: An application to the Cox, Ingersoll and Ross Model.* Journal of Finance, **49:** 1279-1304, 1994.

[97] G. Pennachi. *Identifying the Dynamics of Real Interest Rates and Inflation: Evidence Using Survey Data.* Review of Financial Studies, **4:** 53-86, 1991.

[98] P.C.B. Phillips. *To criticize the critics: An Objective Bayesian analysis of stochastic trends.* Journal of Applied Econometrics, **6:** 333-364, 1991.

[99] P.C.B. Phillips and P. Perron. *Testing a unit root in time series regression.* Biometrika, **75:** 335-346, 1988.

[100] J.L. Pierce. *Interest Rates and their Prospect in the Recovery.* Brookings Papers on Economic Activity, **1:** 89-122, 1975.

[101] J. Price. *Derivatives and Financial Mathematics.* Nova Science Publishers Inc., New York, 1997.

[102] P. Protter. *Stochastic Integration and Differential Equations.* Applications of Mathematics **21.** Springer-Verlag, Berlin, Heidelberg, New York, 1990.

[103] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin, 2nd editon, 1996.

[104] L.C.G. Rogers. *Equivalent martingale measures and no-arbitrage.* Stochastics Stochastics Rep., **51:** 41-49, 1994.

[105] L.C.G. Rogers. *The Potential Approach to the term structure of interest rates and foreign exchange rates.* Math Finance, **7:** 157-176, 1997.

[106] G.D. Rudebush. *The Uncertain Root in Real GNP.* American Economic Review, **83:** 264-272, 1993.

[107] K. Sandman and D. Sondermann. *A Note on the Stability of Lognormal Interest Rate Models and the Pricing of Eurodollar Futures.* Math Finance, **7:** 119-126, 1997.

[108] W. Schachermayer. *A Hilbert-space proof of the fundamental theorem of asset pricing in discrete time.* Insurance, Mathematics and Economics, **11:** 249-257, 1992.

[109] E. Schwartz. *Theory and Applications of the Theory of Interest Rate.* Praeger Publishers, London, 1993.

[110] R.J. Shiller, J.Y. Campbell and K.L. Schoenholtz. *Forward Rates and Future Policy: Interpreting the term structure of interest rates.* Brookings Papers on Economic Activity, **1:** 173-223, 1983.

[111] A.N. Shiryayev. *Probability.* Graduate Texts in Mathematics **95.** Springer-Verlag, Berlin, Heidelberg, New York, 1984.

[112] R.F. Stambaugh. *The Information in Forward Rates: Implications for Models of the Term Structure.* Journal of Financial Economics, **21:** 41-70, 1988.

[113] R. Startz. *Do Forecast Errors or Term Premia Really make the difference between long and short rates?* Journal of Financial Economics, **10:** 323-329, 1982.

[114] K. Tanaka. *Time Series Analysis.* John Wiley and Sons, New York, 1996.

[115] J. Tice and N. Webber. *A Nonlinear model of the term structure of interest rates.* Math Finance, **7:** 177-210, 1997.

[116] O. Vasicek. *An Equilibrium Characterisation of the Term Stucture.* Journal of Financial Economics, **5:** 177-188, 1977.

Appendix A

Related Probability Concepts and Pertinent Results in Stochastic Calculus

A.1 Conditional Expectations and Martingales

Definition A.1.1 *On $(\Omega, \mathfrak{F})$, a probability measure $\bar{P}$ is* **absolutely continuous** *with respect to the probability measure P if for each set A in $\mathfrak{F}$, $P(A) = 0$ implies $\bar{P}(A) = 0$. If $\bar{P}$ is absolutely continuous with respect to P, we use the notation $\bar{P} \sim P$ or $\bar{P} << P$.*

Theorem A.1.1 *On $(\Omega, \mathfrak{F})$, $\bar{P}$ is absolutely continuous with respect to P if and only if there exists a non-negative random variable Λ such that, $\forall A \in \mathfrak{F}$,*

$$\bar{P}(A) = \int_A \Lambda dP.$$

Proof: See Lamberton and Lapeyre, [83].

Remarks:

1. Implication from left to right of Theorem A.1.1 is called the Radon-Nikodym theorem.
2. Λ is called both the Radon-Nikodym derivative and density of $\bar{P}$ with respect to P. It is sometimes denoted by $\frac{d\bar{P}}{dP}$.

Let $X \in L^1$ and $\mathcal{A}$ be a sub σ−field of $\mathcal{B}$. If X is non-negative and integrable we can use the Radon-Nikodym theorem to deduce the existence of an $\mathcal{A}$−measurable random variable denoted by $E[X|\mathcal{A}]$ and called the **expectation of X given $\mathcal{A}$**. This is uniquely determined except on an event of probability zero, such that

$$\int_A X dP = \int_A E[X|\mathcal{A}]dP \quad \text{for all} \quad A \in \mathcal{A}.$$

Classical Results Involving Conditional Expectations

Suppose $\mathcal{A}_1$ and $\mathcal{A}_2$ are two sub σ−fields of $\mathcal{F}$ such that $\mathcal{A}_1 \subset \mathcal{A}_2$. Then,

1. (Tower Property) $E[E[X|\mathcal{A}_2]|\mathcal{A}_1] = E[X|\mathcal{A}_1]$.
2. (Taking out what is known) If X, Y, $XY \in L^1$ and Y is $\mathcal{A}$−measurable then
$$E[XY|\mathcal{A}] = YE[X|\mathcal{A}].$$
3. If X and Y are independent, then $E[X|\sigma(Y)] = E[X]$.

Definition A.1.2 *Suppose $(\Omega, \mathcal{F}, P)$ is a probability space with a filtration $\{\mathcal{F}_t\}$, $t \in [0, \infty)$. A real-valued adapted stochastic process $\{M_t\}$ is said to be a* **supermartingale (resp. submartingale)** *with respect to the filtration $\{\mathcal{F}_t\}$ if*

1. *$E[|M_t|] < \infty$ for all t,*
2. *$E[M_t|\mathcal{F}_s] \leq M_s$ if $s \leq t$, (resp. $E[M_t|\mathcal{F}_s] \geq M_s$ if $s \leq t$).*

If $E[M_t|\mathcal{F}_s] = M_s$ for $s \leq t$ then $\{M_t\}$ is said to be a **martingale.**

Theorem A.1.2 *Suppose $\{W_t\}$ is a standard Brownian motion with respect to the filtration $\{\mathcal{F}_t\}$, $t \geq 0$. Then,*

1. *$\{W_t\}$ is an $\mathcal{F}_t$–martingale.*
2. *$\{W_t^2 - t\}$ is an $\mathcal{F}_t$–martingale.*
3. *$\{exp(\sigma W_t - \frac{\sigma^2}{2}t)\}$ is an $\mathcal{F}_t$–martingale.*

Proof: See Elliott, [38].

The converse of this theorem is also true and such converse due to Lévy gives a characterisation of Brownian motion using properties (1), (2) and (3). Furthermore, the above properties can be shown to imply other well-known properties of Brownian motion, including for example, that W is a Gaussian process with independent increments. The above results can be extended to local martingales and such discussion can be found in Proposition 3.3.8 of Jacod and Shiryayev [74].

A.2 Other Pertinent Results of Continuous Time Stochastic Calculus

Definition A.2.1 *A set K of random variables contained in $L^1(\Omega, \mathcal{F}, P)$ is said to be* **uniformly integrable** *if*

$$\int_{\{|X| \geq c\}} |X| dP$$

converges to zero uniformly in $X \in K$ as $c \to \infty$.

A martingale $\{M_t\}$, $t \in [0, \infty)$ (or $t \in [0, T]$) is said to be uniformly integrable if the set of random varibales $\{M_t\}$ is uniformly integrable.

Remark: A consequence of $\{M_t\}$ being a uniformly integrable martingale on $[0, \infty)$ is that $M_\infty = \lim M_t$ in the $L^1(\Omega, \mathcal{F}, P)$ norm, that is,

$$\lim_t \|M_t - M_\infty\|_1 = 0.$$

In this case, $\{M_t\}$ is a martingale on $[0, \infty]$ and $M_t = E[M_\infty | \mathcal{F}_t]$ a.s. $\forall\, t$.

Write $\mathcal{M} :=$ set of uniformly integrable martingales.

An important concept is that of "localisation." If $\mathcal{C}$ is a class of processes then $\mathcal{C}_{loc}$ is the set of processes defined as follows:

$X \in \mathcal{C}_{loc}$, if there is an increasing sequence $\{T_n\}$ of stopping times $T_1 \leq T_2 \leq T_3 \leq \cdots$ such that

$$\lim T_n = \infty \quad \text{a.s.} \quad \text{and} \quad X_{t \wedge T_n} \in \mathcal{C}.$$

For example, $\mathcal{C}$ might be the collection of bounded processes, or the processes of bounded variation. The **variation** of a real-valued function f over an interval $[a, b]$ is

$$\int_a^b |df| \doteq \sup_{\pi} \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|$$

where π is the set of all partitions of the interval $[a, b]$. If there exists a $D \in \mathbb{R}$ such that $\int_a^b |df| < D$ then f is said to be of **bounded variation.**

The right continuous with left limits adapted stochastic process X is of **integrable variation** if $E\left[\int_0^\infty |dX_s|\right] < \infty$. Given an adapted stochastic process X, if there exists a right continuous with left limits, predictable and with finite variation process A such that $X_t - A_t$ is a martingale, then A is called the **compensator** of X.

Definition A.2.2 *If $M \in \mathcal{C}_{loc}$ and M is a martingale then M is called a* **local martingale.**

Now, we can define the most general form of stochastic processes.

Definition A.2.3 *An adapted process* $X = \{X_t\}, \quad t \geq 0$, *is a* **semimartingale** *if it has a decomposition of the form*

$$X_t = X_0 + M_t + A_t,$$

where M_t *is a local martingale and* A_t *is a process which is continuous on the right with limits on the left (CORLOL),* $A_0 = 0$ *and if almost every sample path is of finite variation on each compact subset of* $[0, \infty)$.

CORLOL processes are also known as CADLAG processes in terms of its French equivalent terminology: continué à droite avec des limités à gauche.

Itô calculus will be described for a class of processes known as Itô processes.

Definition A.2.4 *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space with a filtration* $\{\mathcal{F}_t\}, \quad t \geq 0$, *and* $\{W_t\}$ *is a standard* $\mathcal{F}_t$*–Brownian motion. A real valued* **Itô process** $\{X_t\} \quad t \geq 0$ *is a process of the form*

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s$$

where

1. X_0 *is* $\mathcal{F}_0$*–measurable;*
2. K *and* H *are adapted to* $\mathcal{F}_t$*; and*
3. $\int_0^T |K_s| ds < \infty$ *a.s. and* $\int_0^T |H_s^2| ds < \infty$ *a.s.*

If X is an Itô process the differentiation rule has the following form:

Theorem A.2.1 (Itô's Lemma) *Suppose* $\{X_t, t \geq 0\}$ *is an Itô process of the form*

$$X_t = X_0 + \int_0^t K_s ds + \int_0^t H_s dW_s.$$

Suppose f *is twice differentiable. Then,*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s.$$

Proof: See Elliott, [35].

Here, by definition, $\langle X\rangle_t = \int_0^t H_s^2 ds$; that is, the (predictable) quadratic variation of X is the quadratic variation of its martingale component $\int_0^t H_s dW_s$.

Extension: If $F : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ is differentiable in the first component and twice differentiable in the second then

$$\begin{aligned} F(t, X_t) &= F(0, X_0) + \int_0^t \frac{\partial F}{\partial s}(s, X_s) ds + \int_0^t \frac{\partial F}{\partial X}(s, X_s) dX_s \\ &\quad + \frac{1}{2}\int_0^t \frac{\partial^2 F}{\partial X^2}(s, X_s) d\langle X\rangle_s. \end{aligned}$$

We can extend our definition of an Itô process to the situation where the (scalar) stochastic integral involves an m–dimensional Brownian motion.

Definition A.2.5 $\{X_t\}$, $0 \le t \le T$, *is an Itô process if*

$$X_t = X_0 + \int_0^t K_s ds + \sum_{i=1}^{m} \int_0^t H_s^i dW_s^i$$

where K and H^i are adapted to $\{\mathcal{F}_t\}$, $\int_0^T |K_s| ds < \infty$ a.s. and for all i, $1 \le i \le m$, $\int_0^T |H_s^i|^2 ds < \infty$ a.s.

An n–dimensional Itô process is then a process $X_t = (X_t^1, \cdots, X_t^n)$, each component of which is an Itô process, in the sense of Definition A.2.4.

The differentiation rule takes the form:

Theorem A.2.2 (Itô's Rule for Multi-Dimensional Processes) *Suppose $X_t = (X_t^1, \cdots, X_t^n)$ is an n–dimensional Itô process with*

$$X_t^i = X_0^i + \int_0^t K_s^i ds + \sum_{j=1}^{m} \int_0^t H_s^{ij} dW_s^j,$$

and suppose $f : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ is in $C^{1,2}$, (the space of functions continuously *differentiable in t and twice continuously differentiable in $x \in \mathbb{R}^n$). Then,*

$$
\begin{aligned}
f(t, X_t^1, \cdots, X_t^n) &= f(0, X_0^1, \cdots, X_0^n) + \int_0^t \frac{\partial f}{\partial s}(s, X_s^1, \cdots, X_s^n) ds \\
&+ \sum_0^n \int_0^t \frac{\partial f}{\partial x_i}(s, X_s^1, \cdots, X_s^n) dX_s^i \\
&+ \frac{1}{2} \sum_{j=1}^n \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(s, X_s^1, \cdots, X_s^n) d\langle X^i, X^j \rangle_s.
\end{aligned}
$$

Here,

$$dX_s^i = K_s^i ds + \sum_{j=1}^m H_s^{i,j} dW_s^j$$

and

$$d\langle X^i, X^j \rangle_s = \sum_{r=1}^m H_s^{i,r} H_s^{j,r} ds.$$

Proof: See Øksendal, [94].

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space with filtration $\{\mathcal{F}_t\}$, $0 \le t \le T$. Let $W_t = (W_t^1, \cdots, W_t^m)$ be an $m-$dimensional $\mathcal{F}_t-$Brownian motion and $f(x,t)$, $\sigma(x,t)$ be measurable functions of $x \in \mathbb{R}^n$ and $t \in [0,T]$ with values respectively, in $\mathbb{R}^n$ and $L(\mathbb{R}^m, \mathbb{R}^n)$, the space of $m \times n$ matrices. ξ is an $\mathbb{R}^n-$valued, $\mathcal{F}_0-$measurable random variable.

Definition A.2.6 *A process X_t, $0 \le t \le T$ is a* **solution of the stochatic differential equation (SDE)**

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dW_t$$

with initial condition $X_0 = \xi$ if for all t the integrals $\int_0^t f(X_s, s)ds$ and $\int_0^t \sigma(X_s, s)dW_s$ are well defined and

$$X_t = \xi + \int_0^t f(X_s, s)ds + \int_0^t \sigma(X_s, s)dW_s \quad a.s.$$

We present here a theorem that discusses how martingales in particular Brownian motion transform under a different probability measure.

Theorem A.2.3 (Girsanov) *Suppose* (θ_t), $0 \leq t \leq T$, *is an adapted, measurable process such that* $\int_0^T \theta_s^2 ds < \infty$ *a.s. and also so that the process*

$$\Lambda_t = exp\left(-\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right)$$

is an $(\mathcal{F}_t, P)$ *martingale. Define a new measure* P^θ *on* $\mathcal{F}_T$ *by putting*

$$\left.\frac{dP^\theta}{dP}\right|_{\mathcal{F}_T} = \Lambda_T.$$

Then the process

$$W_t^\theta = W_t + \int_0^t \theta_s ds$$

is a standard Brownian motion on $(\mathcal{F}_t, P^\theta)$.

Remark: A sufficient condition, known as **Novikov's condition**, for Λ to be a martingale is that

$$E\left[exp\left(\frac{1}{2}\int_0^T \theta_s^2 ds\right)\right] < \infty.$$

For details of this condition, see Elliott, [38].

Let $\{W_t\}$, $t \geq 0$, denote a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{F}_t^0 = \sigma\{W_s : s \leq t\}$ and $\mathcal{F}$ is the completion of $\mathcal{F}_t^0$, so that $\{\mathcal{F}_t\}$ $t \geq 0$, is the filtration generated by W which satisfies the usual conditions of right continuity and completeness. If (H_t), $0 \leq t \leq T$, is a measurable adapted process on $[0, T]$ such that $E[\int_0^T H_s^2 ds] < \infty$ then $\int_0^t H_s dW_s$ is a **square integrable martingale.** The representation result tells us that all square integrable martingales on $\{\mathcal{F}_t\}$, $0 \leq t \leq T$, are of this form.

Theorem A.2.4 (Martingale Representation Theorem) *Suppose* $\{M_t\}$, $0 \leq t \leq T$, *is a square integrable martingale on* $\{\mathcal{F}_t\}$, *where* $\{\mathcal{F}_t\}$, *is the completion of* $\sigma\{W_s : s \leq t\}$. *Then, there is a measurable, adapted process* (H_t), $0 \leq t \leq T$, *such that* $E[\int_0^T H_s^2 ds] < \infty$ *and for all* $t \in [0, T]$,

$$M_t = M_0 + \int_0^t H_s dW_s \quad a.s.$$

Proof: See [38].

Most fundamental results of modern finance theory that concerns hedging were generated through the use of this very powerful theorem. The next theorem relates conditional expectations under two different measures.

Theorem A.2.5 (Bayes' Rule) *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space and* $\mathcal{G} \subset \mathcal{F}$ *is a sub* $\sigma-$*field. Suppose* $\bar{P}$ *is another probability measure absolutely continuous with respect to* P *and with Radon-Nikodym derivative*

$$\frac{d\bar{P}}{dP} = \Lambda.$$

Then if ϕ *is any integrable* $\mathcal{F}-$*measurable random variable*

$$\bar{E}[\phi|\mathcal{G}] = \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}.$$

Proof: Suppose B is any set in $\mathcal{G}$. We must show

$$\int_B \bar{E}[\phi|\mathcal{G}]d\bar{P} = \int_B \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}d\bar{P}.$$

Now the right hand side is

$$\begin{aligned}
\bar{E}\left[I_B \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\right] &= E\left[\frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]} I_B \Lambda|\mathcal{G}\right] \\
&= E\left[I_B \left(\frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\right)\Lambda\right] \\
&= E\left[I_B \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]} E[\Lambda|\mathcal{G}]\right] = E\left[I_B E[\Lambda\phi|\mathcal{G}]\right] \\
&= E[I_B\Lambda\phi] = \bar{E}[I_B\phi] = \int_B \phi d\bar{P} \\
&= \int_B \bar{E}[\phi|\mathcal{G}]d\bar{P},
\end{aligned}$$

and the result is proven.

■

We conclude this section with Itô's Differentiation Rule for a general semimartingale. The proofs of these theorems can be found in Elliott, [35] or Jacod and Shiryayev, [74].

Theorem A.2.6 *Suppose X is a semimartingale and F a twice continuously differentiable function. Then $F(X)$ is a semimartingale, and,*

$$\begin{aligned} F(X_t) &= F(X_0) + \int_0^t F'(X_{s-})dX_s + \frac{1}{2}\int_0^t F''(X_s)d\langle X^c, X^c\rangle_s \\ &+ \sum_{0<s\leq t} [F(X_s) - F(X_{s-}) - F'(X_{s-})\Delta X_s]. \end{aligned}$$

We give the differentiation rule for a vector $\mathbb{R}^n$−valued semimartingale. One must be cautious here as the notation becomes very involved.

Theorem A.2.7 (Itô's Rule for an n-Dimensional Semimartingale) *Suppose X is a process with values in $\mathbb{R}^n$, each of whose components X^i is a semimartingale. Suppose F is a real valued twice continuously differentiable function on $\mathbb{R}^n$. Then $F(X_t)$ is a semimartingale and,*

$$\begin{aligned} F(X_t) &= F(X_0) + \sum_{i=1}^n \int_{]0,t]} \frac{\partial}{\partial x_i} F(X_{s-})dX_s^i \\ &+ \frac{1}{2}\sum_{i,j=1}^n \int_0^t \frac{\partial^2}{\partial x_i \partial x_j} F(X_{s-})d\langle X^{ic}, X^{jc}\rangle_s \\ &+ \sum_{0<s\leq t}\left(F(X_s) - F(X_{s-}) - \sum_{i=1}^n \frac{\partial}{\partial x_i} F(X_{s-})\Delta X_s^i\right). \end{aligned}$$

Appendix B

Related Quantitative Finance Theory

This appendix describes topics on quantitative finance theory most related to the area of term structure studies. The motivation for much of the theory discussed herein is to provide foundation for the pricing of contingent claims.

B.1 The Financial Market Model

In all the models that we consider, all processes are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$ where $t \in [0, T]$ and $0 < T < \infty$.

The following conditions are also assumed:

1. $\mathcal{F}_0 = \{A \subset \Omega | P(A) = 0\} \cup \Omega$.
2. $\{\mathcal{F}_t\}$ is right continuous, i.e., $\mathcal{F}_t = \cap_{s>t} \mathcal{F}_s$, $0 \leq t \leq T$; and
3. $\mathcal{F}_T = \mathcal{F}$.

Definition B.1.1 *A* **contingent claim** *is a positive* $\mathcal{F}_T$*−measurable random variable* X_T *defined on the probability space* $(\Omega, \mathcal{F}, P)$.

Further, the tuple $(\Omega, \mathcal{F}, P, \mathbb{T}, \mathbb{F}, S)$ is the **securities market model.** Here, $\mathbb{T}$ will be a time set which is usually taken as the **trading horizon** in practice and $\mathbb{F} = \{\mathcal{F}_{t\in\mathbb{T}}\}$ is the usual filtration of our probability space. S denotes a $(d+1)$−dimensional stochastic process $S = \{S_t^i; t \in \mathbb{T}, \quad 0 \leq i \leq d\}$ representing the time evolution of the **securities price process.**

The security labelled 0 is taken as riskless (i.e. non-random) *bond* (or *bank account*) with price process S^0 while the d risky (i.e., random) stocks labelled $1, \cdots, d$ have price processes $S^1, S^2, \cdots, S^d$. The process S is assumed to be adapted to the filtration $\mathbb{F}$, so that for each $i \leq d$, S_t^i is $\mathcal{F}_t$−measurable, i.e., the prices of the securities at all times up to time t are known. Most frequently, we shall take the filtration $\mathbb{F}$ as that generated by the price process $S = \{S^0, S^1, \cdots, S^d\}$. Then $\mathcal{F}_t = \sigma\{S_u; u \leq t\}$ is the smallest σ−field such that all the $\mathbb{R}^{d+1}$−valued random variable $S_u = (S_u^0, S_u^1, \cdots, S_u^d)$, $u \leq t$, are $\mathcal{F}_t$−measurable. In other words, at time t the investors know the values of the price vectors $(S_u; u \leq t)$, but they have no information about later values of S.

We also require at least one of the price processes to be strictly positive, that is, to act as the **numéraire,** in the model. As is customary we assign this role to the bond price S^0, although in principle any strictly positive S^i could be used for this purpose.

B.2 Self-Financing and Replicating Trading Strategies

Definition B.2.1 *A* **trading strategy** *is a measurable process $H_t = (H_t^0, H_t^1)$ with values in $\mathbb{R}^2$ which is adapted to the filtration $\{\mathcal{F}_t\}$, $t \geq 0$, where $\mathcal{F}_t = \sigma\{W_u; u \leq t\} = \sigma\{S_u; u \leq t\}$.*

Occasionally, as **hedging** (a term more commonly known to traders and portfolio managers) means taking positions against the risk of market movements, we shall use the term trading strategy and hedging interchangeably.

H_t^0 and H_t^1 denote respectively, the amount of the bond S_t^0 and the holdings of the risky asset S_t^1, at time t. Consequently, the **value** or **wealth,** of the portfolio at time t is

$$V_t(H) = H_t^0 S_t^0 + H_t^1 S_t^1.$$

The description (H_t^0, H_t^1) is a dynamic strategy detailing the amount of each component to be held at each instant. And one particularly interesting set of strategies or portfolios are those that are financially self-contained or self-financing.

Definition B.2.2 *A* **self-financing strategy** $H = (H_t)$, $0 \leq t \leq T$ *is given by two measurable adapted processes* (H_t^0), (H_t^1), *such that*

1. $\int_0^T |H_t^0| dt < \infty$. *a.s.*

2. $\int_0^T (H_t^1)^2 dt < \infty$ *a.s.*

3. *The processes* (H_t^0) *and* (H_t^1) *satisfy the SDE*

$$dV_t(H) = H_t^0 dS_t^0 + H_t^1 dS_t^1. \qquad \text{(B.2.1)}$$

The corresponding value of the process therefore for a self-financing strategy is

$$\begin{aligned} V_t(H) &= H_t^0 S_t^0 + H_t^1 S_t^1 \\ &= H_0^0 S_0^0 + H_0^1 S_0^1 + \int_0^t H_u^0 dS_u + \int_0^t H_u^1 dS_u^1 \quad \text{a.s.} \end{aligned}$$

for all $t \in [0, T]$.

Indeed, if H^0 and H^1 are of bounded variation then

$$dV_t(H) = H_t^0 dS_t^0 + H_t^1 dS_t^1 + S_t^0 dH_t^0 + S_t^1 dH_t^1.$$

and (B.2.1) is equivalent to saying that

$$S_t^0 dH_t^0 + S_t^1 dH_t^1 = 0 \qquad \text{(B.2.2)}$$

Intuitively, (B.2.2) means that the changes in the holdings of the bond, $S_t^0 dH_t^0$, can only take place due to corresponding changes in the holding of the stock $S_t^1 dH_t^1$, that is, there is no net inflow or outflow of capital.

Write $\tilde{S}_t^1 := e^{-rt} S_t^1$, for the discounted price of the risky asset.

Write $\tilde{V}_t(H) := e^{-rt} V_t(H)$, for the discounted wealth process.

Theorem B.2.1 *Suppose* $H = (H_t) = (H_t^0, H_t^1)$, $0 \le t \le T$, *is a pair of measurable adapted processes which satisfy (1) and (2) of Definition (B.2.2). Then* H *is a self-financing strategy if and only if*

$$\tilde{V}_t(H) = V_0(H) + \int_0^t H_u^1 d\tilde{S}_u \quad a.s.$$

Proof: ($\Longrightarrow$) Suppose $H = (H_t^0, H_t^1)$ is self-financing so that

$$dV_t(H) = H_t^0 dS_t^0 + H_t^1 dS_t^1.$$

Then,

$$\begin{aligned} d(\tilde{V}_t) &= d\big(e^{-rt} V_t(H)\big) \\ &= -r\tilde{V}_t(H)dt + e^{-rt} dV_t(H) \\ &= -re^{-rt}(H_t^0 e^{rt} + H_t^1 S_t^1)dt \\ &\quad + e^{-rt} H_t^0 d(e^{rt}) + e^{-rt} H_t^1 dS_t^1 \\ &= H_t^1(-re^{-rt} S_t^1 dt + e^{-rt} dS_t^1) \\ &= H_t^1 d\tilde{S}_t^1. \end{aligned}$$

($\Longleftarrow$) Consider $V_t(H) = e^{rt}\tilde{V}_t(H)$. We shall evaluate dV_t.

$$\begin{aligned} dV_t &= d\big(e^{rt}\tilde{V}_t(H)\big) \\ &= e^{rt} d\tilde{V}_t(H) + \tilde{V}_t(H)(re^{rt}dt) \end{aligned}$$

But $d\tilde{V}_t(H) = H_t^1 d\tilde{S}_t^1$ by hypothesis. Thus,

$$\begin{aligned}
dV_t(H) &= e^{rt}H_t^1 d\tilde{S}_t^1 + \tilde{V}_t(H)(re^{rt}dt) \\
&= e^{rt}H_t^1 d\tilde{S}_t^1 + e^{-rt}V_t(H)(re^{rt}dt) \\
&= e^{rt}H_t^1 d\tilde{S}_t^1 + rV_t(H)dt \\
&= e^{rt}H_t^1(-re^{-rt}S_t^1 dt + e^{-rt}dS_t^1) \\
&\quad +r(H_t^0 S_t^0 dt + H_t^1 S_t^1 dt) \\
&= -rH_t^1 S_t dt + H_t^1 dS_t^1 \\
&\quad +rH_t^0 S_t^0 dt + rH_t^1 S_t^1 dt \\
&= H_t^0 r S_t^0 dt + H_t^1 dS_t^1 \\
&= H_t^0 dS_t^0 + H_t^1 dS_t^1 \\
&\quad \text{since} \quad rS_t^0 dt = dS_t^0 \quad \text{and} S_t^0 = e^{rt}. \\
dV_t(H) &= H_t^0 dS_t^0 + H_t^1 dS_t^1.
\end{aligned}$$

Hence, $H = (H_t^0, H_t^1)$ is self-financing.

■

Suppose, $H = (H_t)$, $0 \le t \le T$ is a self-financing strategy so that H^0 and H^1 satisfy (B.2.1). If there are no contributions or withdrawals the corresponding wealth process is given by

$$V_t(H) = H_t^0 S_t^0 + H_t^1 S_t^1.$$

Now, assume that there are contributions to the wealth process (say, from dividends) or withdrawals (consumptions). Let these be modelled by the adapted right-continuous, increasing process D_t (for contributions) and C_t (for consumptions). Here, C_t is the accumulated consumption. Thus,

$$\begin{aligned}
V_t(H) &= H_t^0 S_t^0 + H_t^1 S_t^1 + D_t - C_t \\
&= V_0(H) + \int_0^t H_u^0 dS_u^0 + \int_0^t H_u^1 dS_u^1 + D_t - C_t.
\end{aligned}$$

The self-financing condition of (B.2.1) becomes

$$S_t^0 dH_t^0 + S_t^1 dH_t^1 = dD_t - dC_t.$$

Definition B.2.3 *Suppose σ_t is the volatility of the risky security S_t^1. A* **replicating strategy** *on a claim X is a self-financing portfolio (H_t^0, H_t^1) such that $\int_0^T \sigma_t^2 (H_t^1)^2 dt < \infty$ and $V_T = H_T^0 S_T^0 + H_T^1 S_T^1 = X$.*

Why should we care about replicating strategies? The claim X gives the value of some derivative which we need to pay off at time T. We want a price if there is one, as of now, given a model for S_t^1 and S_t^0.

If there is a replicating strategy (H_t^1, H_t^2), then the price of X at time t must be $V_t = H_t^0 S_t^0 + H_t^1 S_t^1$. And specifically, the price at time zero is $V_0 = H_0^0 S_0^0 + H_0^1 S_0^1$. If it were lower, a market player could buy one unit of the derivative at time t and sell H_t^1 units of S_t^1 and H_t^0 units of S_t^0 against it, continuing to be short (H_t^0, H_t^1) until time T. Because (H_t^0, H_t^1) is self-financing and the portfolio is worth X at time T guaranteed, the purchased derivative and sold portfolio would safely cancel at time T, and no extra money is required between times t and T. The mismatch created at time t generates a riskless profit. And as usual with arbitrage, one unit could have been many; no risk means no fair price.

Similarly, if the derivative price had been higher than V_t, then we could have sold the derivative and bought the self-financing (H_t^0, H_t^1) to the same effect. Replicating strategies, if they exist, tie down the price of the claim X not just at pay-off time but everywhere, ensuring no-arbitrage opportunity.

B.3 Equivalent Martingale Measure

Let r be a non-negative constant which represents the instantaneous interest rate on the bond. We then suppose that the evolution in the price of the bond S_t^0 is described by the ordinary differential equation (ODE)

$$dS_t^0 = rS_t^0 dt \tag{B.3.3}$$

If the initial value at time 0 of the bond is $S_0^0 = 1$, then (B.3.3) can be solved to give

$$S_t^0 = e^{rt}, \quad t \geq 0.$$

Furthermore, let the risky asset have a price process (S_t^1) satisfying

$$dS_t^1 = S_t^1(\mu dt + \sigma dW_t).$$

The discounted price of the risky asset is

$$\tilde{S}_t = e^{-rt}S_t^1 \quad \text{with dynamics}$$

$$\begin{aligned} d\tilde{S}_t &= -re^{-rt}S_t^1 dt + e^{-rt}dS_t^1 \\ &= \tilde{S}_t((\mu - r)dt + \sigma dW_t). \end{aligned}$$

If we apply Girsanov's theorem, with $\theta_t = \frac{\mu - r}{\sigma}$ we see that there is a probability measure $\bar{P}$, defined on $\mathcal{F}_T$ by putting

$$\frac{d\bar{P}}{dP} = \Lambda_T = exp\left(-\int_0^t \theta_s dW_s - \frac{1}{2}\int_0^t \theta_s^2 ds\right),$$

such that under $\bar{P}$, $\bar{W}_t$, $0 \leq t \leq T$, is a standard Brownian motion where

$$\bar{W}_t = \left(\frac{\mu - r}{\sigma}\right)t + W_t.$$

Then, under $\bar{P}$ we have

$$\begin{aligned} d\tilde{S}_t &= \tilde{S}_t \sigma d\bar{W}_t \quad \text{and} \\ \tilde{S}_t &= S_0 exp\left(\sigma \bar{W}_t - \frac{\sigma^2 t}{2}\right). \end{aligned}$$

Definition B.3.1 *A strategy $H = (H_t^0, H_t^1)$, $0 \leq t \leq T$, is* **admissible** *if it is self-financing and the discounted value process $\tilde{V}_t(H) = H_t^0 + H_t^1\tilde{S}_t$ is square integrable under $\bar{P}$.*

Definition B.3.2 *A self-financing strategy H is said to provide an* **arbitrage opportunity** *if with $V_0(H) = x \leq 0$, we have $V_T(H) \geq 0$ a.s. and $P\{\omega : V_T(H) > 0\} > 0$.*

In the setting of a probability space $(\Omega, \mathcal{F}, P)$ an equivalent measure $\bar{P}$ is called a **martingale measure** if, under $\bar{P}$, all discounted asset prices are martingales. $\bar{P}$ is sometimes called the **risk-neutral measure.**

We have seen that, in the case of one risky asset $\bar{P}$ is a martingale measure.

Suppose $W_t = (W_t^1, \cdots, W_t^m)$, $0 \leq t \leq T$, is an $m-$dimensional Brownian motion on $(\Omega, \mathcal{F}, P)$ and let $\{\mathcal{F}_t\}$ be the filtration generated by W.

Suppose now there is a bond S_t^0 or bank account whose instantaneous interest rate is r_t and n risky assets $S_t^1, \cdots, S_t^n$. With $S_0^0 = 1$, we have $S_t^0 = exp\{\int_0^t r_u du\}$. The dynamics of the risky assets are described by the equations

$$dS_t^i = \mu_i(t)S_t^i + S_t^i \left[\sum_{j=1}^{n} \sigma_{ij}(t) dW_t^j \right].$$

Here, μ_i, σ_{ij} and r are adapted processes. The prices $\frac{S_t^1}{S_t^0}, \cdots, \frac{S_t^n}{S_t^0}$ are the discounted prices and the differentiation rule gives

$$\begin{aligned} d\left[\frac{S^i(t)}{S^0(t)}\right] &= \Big(\mu_i(t) - r(t)\Big)\frac{S_t^i}{S_t^0}dt \\ &\quad + \frac{S_t^i}{S_t^0}\sum_{j=1}^{m}\sigma_{ij}(t)W_t^j. \end{aligned} \tag{B.3.4}$$

In equation (B.3.4), the terms μ_i, σ_{ij} are called the **risk premium.**

Definition B.3.3 *If we can find processes $\theta_1(t), \cdots, \theta_n(t)$ so that for $1 \leq i \leq n$*

$$\mu_i(t) - r(t) = \sum_{j=1}^{m} \sigma_{ij}(t)\theta_j(t) \tag{B.3.5}$$

then the adapted process $\theta(t) := \big(\theta_1(t), \cdots, \theta_n(t)\big)$ is called the **market price of risk.**

Equation (B.3.4) then becomes

$$d\left[\frac{S_t^i}{S_t^0}\right] = \frac{S_t^i}{S_t^0}\left(\sum_{j=1}^{m}\sigma_{ij}(t)[\theta_j(t)dt + dW_t^j]\right).$$

Consider the linear system (B.3.5). Three cases are possible.

1. it has a unique solution $\theta(t) = (\theta_1(t), \cdots, \theta_n(t))$;

2. it has no solution; and

3. it has more than one solution.

In cases (1) and (3) we have a solution process $\theta(t)$.

Consider the process

$$\Lambda_t = exp\left(-\int_0^t \theta(u)dW_u - \frac{1}{2}\int_0^t |\theta(u)|^2 du\right)$$

and define a new measure P^θ by setting $\frac{dP^\theta}{dP}\Big|_{\mathcal{F}_T} = \Lambda_T$.

The vector form of the Girsanov's theorem states that, under P^θ, $W^\theta = (W^{\theta_1}, W^{\theta_2}, \cdots, W^{\theta_m})$ is an $m-$dimensional martingale where $dW_t^\theta = \theta(t)dt + dW_t$.

A hedging strategy is now a measurable adapted process $H_t = (H_t^0, H_t^1, \cdots, H_t^n)$ where H_t^i represents the number of assets held at time t. Its corresponding wealth process is

$$V_t(H) = H_t^0 S_t^0 + H_t^1 S_t^1 + \cdots + H_t^n S_t^n.$$

As an anologue to Definition (B.2.2), we say that H is said to be **self- financing** if $dV_t(H) = \sum_{i=0}^n H_t^i dS_t^i$. And therefore,

$$\begin{aligned} V_t(H) &= V_0(H) + \int_0^t \sum_{i=0}^n H_u^i dS_u^i \\ &= V_0(H) + \int_0^t rH_u^0 S_u^0 du \\ &\quad + \sum_{i=1}^n \int_0^t H_u^i S_u^i \left(\mu_i(u) + \sum_{i=1}^m \sigma_{ij}(u) dW_u^j\right) \end{aligned}$$

$$
\begin{aligned}
V_t(H) &= V_0(H) + \int_0^t rV_u(H)du \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\int_0^t H_u^i S_u^i \sigma^{ij}\Big(\theta_j(u)du + dW_u^j\Big) \\
&= V_0(H) + \int_0^t rV_u(H)du \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{m}\int_0^t H_u^i S_u^i \sigma_{ij}(u)dW_u^{\theta_j}.
\end{aligned}
$$

Therefore, the discounted wealth

$$
\begin{aligned}
\bar{V}_t(H) &= (S_t^0)^{-1}V_t(H) \\
&= V_0(H) + \sum_{i=1}^{n}\sum_{j=1}^{m}\int_0^t H_u^i S_u^i \sigma_{ij}(u)dW_u^{\theta_j}
\end{aligned}
$$

is a local martingale.

B.4 Viable and Complete Markets

Definition B.4.1 *Let $X_T(\omega)$ represent the pay-off from a contract or agreement based on a certain contingent claim, if state ω prevails. Then such a contingent claim is said to be* **attainable** *if there exists a trading strategy H such that $V_T(H) = X_T(\omega)$.*

Definition B.4.2 *A market is said to be* **complete** *if every contingent claim is attainable.*

Theorem B.4.1 *A market is viable if and only if there exists a unique probability measure $\bar{P}$ equivalent to P under which the discounted prices are martingales.*

Proof: See Harrison and Pliska, Corollary 3.36, page 224, [63].

B.5 The Fundamental Theorem of Asset Pricing

If the filtration $\mathbb{F}$ is finitely generated, an equivalent martingale measure (EMM) $\bar{P}$ for the price process S could be constructed. For details on EMM construction, see Elliott and Kopp, [38]. The much-sought-after equivalence of the *existence of an equivalent (local) martingale* and *conditions of no-arbitrage* is known as the *Fundamental Theorem of Asset Pricing.* This provides a vital link between the economically significant "no-arbitrage" condition and the mathematically important reason for equating the class of admissible stock price processes with the class of $P-$semimartingales, thus allowing the fullest use of the well-developed theory of semimartingales and general stochastic calculus.

Theorem B.5.1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and define the finite discrete time set $\mathbb{T} = \{0, 1, 2, \cdots, T\}$. Assume given a filtration $\mathbb{F} = \{\mathcal{F}_t,\ t \in \mathbb{T}\}$ and an $\mathbb{R}^{d+1}-$valued process $S = (S_t)_{t\in\mathbb{T}}$, adapted to $\mathbb{F}$. We assume further that the first component $S_0^0 = 1$ and for $i \leq d$ and $t \in \mathbb{T}$ we have $S_t^i > 0$ $P-a.s.$. Then the following are equivalent:*

1. *There is a probability $\bar{P} \sim P$ such that S_t is an $(\{\mathcal{F}_t, t \in \mathbb{T}\}\ \bar{P},)$ martingale.*
2. *There are no arbitrage opportunities, that is: for every self-financing trading strategy $H = (H_t^0, H_t^1, \cdots, H_t^d)_{t\in\mathbb{T}}$; with gains process $G(H)$ defined by $G_t(H) = \sum_{i=1}^{t} \Delta S_u$ $(t \in \mathbb{T})$, we have: $G_T(H) \geq 0$ $P-a.s.$ implies $G_T(H) = 0$ $P-a.s.$. If either (1) or (2) holds, then $\bar{P}$ can be found with bounded Radon-Nikodym derivative $\frac{d\bar{P}}{dP}$.*

Proofs of this theorem can be found in [26], [104] and [108].

There is an extensive literature on the relationship between no-arbitrage and the existence of EMM. The fundamental theorems on asset pricing were given in two papers by Harrison and Pliska. In [63], it is shown that if a market has a martingale measure, there is no arbitrage opportunity. In [64], it is shown that the martingale measure is unique if and only if every claim is hedgeable, that is, if and only if the market is complete.

Appendix C

A Computer Program

This appendix contains a copy of the program written in **Matlab** to implement the filtering and estimation procedure of the Markovian mean reverting interest rate model of Chapter 9.

```
%This program computes the one-step ahead prediction for an
%interest rate model with a Markovian mean reverting level.

clear all

load data 3-months
%load data % 2-years
%load data % 10-years
l=[];
ell=0;

%Initialise the transition intensity matrix.
A1=[-1 0.5 0; 1 -1 1; 0 0.5 -1];

%Initialise the rate of adjustment, gamma.
gamma=8;

%Initialise the process sigma(X)_k, denoted by sx(:,k).
sx(:,1)=[0;1;0];

%Create 3x3 matrix.
e=eye(3,3);

Ak_1=A1
sx(:,2)=sx(:,1);

%Define the process sigma(JX)_k by sJx.
sJx=zeros(3,length(y),3,3);

%Define the process sigma(OX)_k by sOx.
sOx=0.8*(3,length(y),3);
%Note that sOx must be defined similar to sJx, however
%the A_k matrix estimates involve sOx in the denominator.
%If sOx is a zero vector, the estimates explode. To rectify
%this situation, we arbitrarily choose the components of sOx
%to be multiplied by 0.8 for the initial estimate.
%Values very close to 0 also do not give reasonable estimates.

%Define the process sigma(IX)_k by sIx.
sIx=zeros(3,length(y),3,3);

%Define the process sigma(KX)_k by sKx.
sKx=zeros(3,length(y),3,3);

%Define the interest rate process, l.
l=[l;y(1)];
l=[l;y(2)];

%The values y(k) are the original data, i.e., the actual yields of
%the T-bills or T-bonds.

gamma(1)=8;
gamma(2)=8;

%Initialise the mean reverting level, lambda, denoted by lambda(:,k).
lambda(:,1)=[-0.3;0.9;2.4];
lambda(:,2)=lambda(:,1);

for k=3:length(y)

   P=zeros(3,1);
   %Define the matrix B_k denoted by Bk.
   Bk=diag([lambda(1,k-1)-1-y(k);lambda(2,k-1)-y(k);lambda(3,k-1)+1-y(k)]);
   den=inv(eye(3)-1/12*Ak_1-gamma(k-1)/zeta(k-1)^2*Bk*(y(k+1)-y(k));
   %zeta(k) refers to the volatilities of the bond returns and were computed from
   %the data using the last 4 preceding yields.
```

```
for s=2:k-1
   Bs=diag([lambda(1,s-1)-1-y(s);lambda(2,s-1)-y(s);lambda(3,s-1)+1-y(s)]);
   P=P+1/12*Ak_1*sx(:,s)+gamma(k-1)/zeta(k-1)^2*Bs*sx(:,s)*(y(k+1)-y(k));

   %Calculate recursive values for sJx, sOx, sKx and sIx.
   for i=1:3
      for j=1:3
         sJx(:,k,i,j)=sJx(:,k,i,j)+Ak_1(i,j)*sx(i,s)*e(:,j)*1/12+
         Ak_1(i,j)*sJx(:,s,i,j)*1/12+gamma(k-1)/zeta(k-1)^2*Bs*sJx(:,s,i,j)*
         (y(k+1)-y(k));
      end

      sOx(:,k,i)=sOx(:,k,i)+1/12*sx(i,s)*e(:,i)+1/12*Ak_1*sOx(:,s,i)+
      gamma(k-1)/zeta(k-1)^2*Bs*sOx(:,s,i)*(y(k+1)-y(k));
      sOx(:,k,i)=den*sOx(:,k,i);

      sKx(:,k,i)=sKx(:,k,i)+gamma(k-1)*lambda(i,k-1)-y(s))*sx(i,s)*e(:,i)*1/12;
      sKx(:,k,i)=sKx(:,k,i)+Ak_1*sKx(:,s,i)*1/12+gamma(k-1)/zeta(k-1)^2
      *Bs*sKx(:,s,i)*(y(s)-y(s-1));
      sKx(:k,i)=sKx(:,k,i)+sx(i,s)*e(:,i)*(y(s)-y(s-1));
      sKx(:,k,i)=sKx(:,k,1)+sx(i,s)*e(:,i)*(y(s)-y(s-1));

      sIx(:,k,i)=sIx(:,k,i)+y(s)*sx(i,s)*e(:,i)*1/12+
      Ak_1*sIx(:,s,i)*1/12+gamma(k-1)/zeta(k-1)^2*Bs*sIx(:,s,i)*(y(s)-y(s-1));
   end
end
sx(:,k)=den*P

for i=1:3
   sKx(:,k,i)=sKx(:,k,i)+gamma(k-1)*(lambda(i,k-1)-y(k))*sx(i,k)*e(:,i)*1/12+
   sx(i,k)*e(:,i)*(y(k)-y(k-1));

   sIx(:,k,i)=den*(sIx(:,k,i)+y(k)*sx(i,k)*e(:,i)*1/12);
end

%Denote the estimate for gamma by C/D.
for s=2:k
   Ck=y(s)*(y(s)-y(s-1))-
sum(sIx(:,k,1))*sum(sKx(:,k,1))/(sum(sOx(:,k,1))*sum(sx(:,k)));
   Ck=Ck-sum(sIx(:,k,2))*sum(sKx(:,k,2))/(sum(sOx(:,k,2))*sum(sx(:,k)));
   Ck=Ck-sum(sIx(:,k,3))*sum(sKx(:,k,3))/sum(sOx(:,k,3))*sum(sx(:,k));
end

Dk=sum(sIx(:,k,1))^2/sum(sx(:,k))/sum(sOx(:,k,1))+
sum(sIx(:,k,2))^2/sum(sx(:,k))/sum(sOx(:,k,2))+
sum(sIx(:,k,3))^2/sum(sx(:,k))/sum(sOx(:,k,3));

gamma(k)=Ck/Dk;
%The k-th estimate for lambda after estimating gamma is
for i=1:3
   lambda(i,k)=(1/gamma(k)*sum(sKx(:,k,i))+sum(sIx(:,k,i)))/sum(sOx(:,k,i)):
end

%After the parameters are computed at time k, they are fed to the
%interest rate model. This is a one-step ahead prediction model as
%the initial estimate is always the actual data at lag 1. In other words,
%the initial estimate is the yield at time k when the model computes
%for the forecast at time k+1.

l=[1,y(k-1)+gamma(k-1)*sx(:,k-1)'*[lambda(1,k-1)-1;lambda(2,k-1);
lambda(3,k-1)+1]+zeta(k-1)*z(k-1)*sqrt(1/12)];
%The variable z(k-1) refers to the N(0,1) generated random numbers.
for i=1:3
   for j=1:3
   Ak(i,j)=sum(sJx(:,k,i,j))/sum(sOx(:,k,i));
   end
end

   Ak_1=Ak;
   A(k,:,:)=Ak;
   k
   if ell==1
      sum(Ak_1(1,:))
      sum(Ak_1(2,:))
      sum(Ak_1(3,:))
      keyboard;
   end
end
```

Appendix D

The Data

This appendix contains the descriptions of the yield rates data used in the application of parameter estimation and filtering techniques decribed in Appendix C.

TREASURY BILL AUCTION - AVERAGE YIELDS - 3 MONTH*

*ADJUDICATION DE BONS DU TRÉSOR - RENDEMENT MOYEN - À 3 MOIS**

(Per cent / *en pourcentage*)

Year / *année*	Jan / *jan*	Feb / *fév*	Mar / *mar*	Apr / *avr*	May / *mai*	Jun / *jun*	Jul / *jul*	Aug / *août*	Sep / *sep*	Oct / *oct*	Nov / *nov*	Dec / *déc*
1968	6.29	6.80	6.98	6.99	6.95	6.56	6.03	5.48	5.66	5.57	5.66	6.24
1969	6.38	6.43	6.58	6.80	6.74	7.15	7.62	7.69	7.77	7.60	7.76	7.81
1970	7.78	7.60	7.00	6.78	6.34	5.94	5.70	5.51	5.39	5.01	4.40	4.44
1971	4.68	4.06	3.16	3.00	3.03	3.37	3.68	3.79	4.06	3.47	3.24	3.21
1972	3.36	3.45	3.57	3.64	3.73	3.50	3.46	3.50	3.62	3.57	3.68	3.65
1973	3.90	3.99	4.46	4.90	5.18	5.48	5.74	6.18	6.50	6.53	6.43	6.35
1974	6.22	6.07	6.51	7.64	8.63	8.75	9.10	9.11	8.94	8.31	7.49	7.12
1975	6.40	6.26	6.33	6.85	6.87	6.99	7.44	7.87	8.41	8.16	8.52	8.64
1976	8.59	8.79	9.07	8.99	8.94	8.98	9.07	9.13	9.11	9.01	8.59	8.14
1977	8.04	7.65	7.54	7.58	7.05	7.07	7.14	7.14	7.10	7.24	7.26	7.17
1978	7.13	7.30	7.73	8.19	8.20	8.26	8.66	8.80	9.17	9.85	10.36	10.46
1979	10.85	10.82	10.92	10.82	10.84	10.78	11.24	11.45	11.64	13.61	13.62	13.66
1980	13.50	13.55	15.24	15.15	11.58	10.38	10.06	10.49	10.95	11.91	13.70	17.01
1981	16.86	16.83	16.44	17.35	18.43	18.83	20.29	20.82	19.35	17.96	15.07	14.41
1982	14.34	14.58	15.07	14.98	15.18	16.33	15.25	13.70	12.73	11.21	10.72	9.80
1983	9.58	9.23	9.17	9.12	9.25	9.17	9.24	9.29	9.24	9.24	9.48	9.71
1984	9.73	9.82	10.53	10.59	11.29	12.11	12.73	12.13	12.02	11.42	10.50	9.84
1985	9.50	11.27	10.40	9.77	9.51	9.33	9.06	8.95	8.75	8.53	8.85	9.24
1986	10.55	11.55	10.19	8.72	8.33	8.59	8.26	8.33	8.35	8.30	8.24	8.24
1987	7.24	7.28	6.80	8.08	8.19	8.29	8.97	8.99	9.35	7.84	8.31	8.41
1988	8.37	8.32	8.53	8.87	8.92	9.19	9.29	9.98	10.33	10.29	10.76	10.94
1989	11.18	11.61	12.14	12.37	12.17	12.08	12.11	12.16	12.23	12.17	12.21	12.22
1990	12.34	13.16	13.26	13.55	13.67	13.58	13.23	12.67	12.40	12.36	12.01	11.47
1991	10.48	9.72	9.67	9.24	8.81	8.65	8.66	8.53	8.34	7.79	7.41	7.42
1992	7.04	7.25	7.24	6.72	6.08	5.60	5.17	4.82	7.37	6.06	8.57	7.11
1993	6.56	5.84	5.11	5.35	4.85	4.54	4.16	4.74	4.65	4.38	4.09	3.86
1994	3.63	3.85	5.39	5.82	6.34	6.67	5.79	5.35	5.29	5.37	5.79	7.18
1995	7.98	7.77	8.22	7.92	7.39	6.72	6.62	6.34	6.46	5.93	5.82	5.54
1996	5.12	5.18	5.03	4.73	4.65	4.70	4.43	4.03	3.96	3.19	2.69	2.80
1997	2.84	2.86	3.19	3.14	3.01	2.86	3.32	3.13	3.10	3.63	3.57	4.46
1998	4.18	4.57	4.57	4.83	4.75	4.88	4.93	4.88	4.94	4.74	4.82	4.70

Source: Bank of Canada, Department of Monetary and Financial Analysis. / *Banque du Canada, département des Études monétaires et financières.*

* From 1934 to 1952 the rates are a monthly average of yields at tender. From 1953 to 1960 the rates are average yields at Thursday tender following the last Wednesday of the month. The first sale of treasury bills by public tender occured March 1, 1934. / *De 1934 à 1952, moyenne mensuelle des taux de rendement à l'adjudication; de 1953 à 1960, taux de rendement moyen à l'adjudication du jeudi suivant le dernier mercredi du mois. La première vente de bons du Trésor par adjudication a eu lieu le 1er mars 1934.*

SELECTED GOVERNMENT OF CANADA BENCHMARK BOND YIELDS - 2 YEAR*
*QUELQUES RENDEMENTS D'OBLIGATIONS TYPES DU GOUVERNEMENT CANADIEN - À 2 ANS**

(Per cent / *en pourcentage*)

Year / *année*	Jan / *jan*	Feb / *fév*	Mar / *mar*	Apr / *avr*	May / *mai*	Jun / *jun*	Jul / *jul*	Aug / *aoû*	Sep / *sep*	Oct / *oct*	Nov / *nov*	Dec / *déc*
1982						16.40	15.74	13.64	12.62	11.38	10.55	10.03
1983	10.17	10.02	10.20	9.92	9.59	9.82	10.27	10.74	10.10	9.88	10.09	10.33
1984	10.23	10.81	11.44	11.71	13.25	13.08	13.04	12.60	12.21	11.55	10.61	10.18
1985	10.21	11.69	11.16	10.70	10.14	10.01	9.95	9.81	9.95	9.40	9.26	8.98
1986	9.78	9.71	9.22	8.78	8.95	8.69	9.00	8.96	9.10	8.86	8.61	8.55
1987	7.68	7.81	7.54	9.01	9.38	9.04	9.77	9.90	10.58	9.38	9.57	9.62
1988	8.98	8.77	8.95	9.23	9.44	9.43	9.57	10.36	10.10	9.66	10.31	10.56
1989	10.54	11.36	11.62	10.78	10.37	10.01	9.86	10.39	10.46	10.17	11.03	10.92
1990	10.96	11.92	12.22	12.74	12.03	11.53	11.56	10.79	11.14	10.99	10.42	10.46
1991	9.94	9.35	9.29	9.30	9.11	9.40	9.33	9.00	8.48	7.43	7.63	7.31
1992	7.25	7.35	7.96	7.72	7.01	6.42	5.50	5.41	6.93	5.84	7.10	6.93
1993	6.91	6.56	6.41	6.50	6.28	5.89	5.39	5.30	5.65	5.06	5.02	4.61
1994	4.18	5.02	6.71	6.84	7.42	8.36	8.27	7.60	7.46	7.64	8.01	8.84
1995	8.90	8.06	7.97	7.65	7.04	6.74	7.21	7.05	7.05	6.98	6.07	5.89
1996	5.41	5.80	6.14	6.10	6.06	6.29	6.01	5.48	4.95	4.28	3.68	4.03
1997	4.44	4.29	4.64	4.80	4.66	4.27	4.49	4.26	4.47	4.38	4.63	5.04
1998	4.73	5.05	4.87	5.12	5.09	5.20	5.24	5.41	4.70	4.51	4.94	4.72

Source: Bank of Canada, Department of Monetary and Financial Analysis. / *Banque du Canada, département des Études monétaires et financières.*

* The rates are based on actual mid-market closing yields of selected Canada bond issues that mature approximately in the indicated term areas. At times, some of the change in the yield occurring over a reporting period may reflect a switch to a more topical issue. / *Les taux indiqués sont calculés en fonction de la moyenne des cours acheteur et vendeur, à la clôture, de certaines émissions d'obligations du gouvernement canadien dont les échéances correspondent à peu près à celles du tableau. Les variations des taux de rendement observées sur une période peuvent être partiellement imputables au remplacement d'une émission par une autre plus pertinente.*

SELECTED GOVERNMENT OF CANADA BENCHMARK BOND YIELDS - 10 YEAR*

*QUELQUES RENDEMENTS D'OBLIGATIONS TYPES DU GOUVERNEMENT CANADIEN - À 10 ANS**

(Per cent / *en pourcentage*)

Year / *année*	Jan / *jan*	Feb / *fév*	Mar / *mar*	Apr / *avr*	May / *mai*	Jun / *jun*	Jul / *jul*	Aug / *aoû*	Sep / *sep*	Oct / *oct*	Nov / *nov*	Dec / *déc*
1982						16.39	15.88	14.26	13.53	12.58	11.88	11.31
1983	11.58	11.19	11.19	10.90	10.86	11.29	11.65	12.01	11.66	11.54	11.62	11.72
1984	11.70	12.30	13.00	13.31	13.90	13.86	13.44	13.01	12.68	12.32	11.72	11.52
1985	11.17	12.17	11.90	11.38	10.61	10.74	10.73	10.53	10.75	10.47	9.88	9.63
1986	10.18	9.69	9.20	8.87	9.08	8.91	8.98	8.81	9.09	9.12	8.79	8.74
1987	8.26	8.39	8.24	9.40	9.52	9.31	9.90	10.11	10.90	9.80	10.10	10.02
1988	9.31	9.13	9.61	9.84	9.86	9.70	9.96	10.40	10.20	9.77	10.05	10.17
1989	10.01	10.42	10.58	10.19	9.71	9.40	9.31	9.51	9.74	9.38	9.73	9.56
1990	9.93	10.64	10.88	11.51	10.86	10.64	10.65	10.72	11.42	11.02	10.54	10.34
1991	10.05	9.59	9.55	9.59	9.59	10.09	9.90	9.67	9.26	8.74	8.67	8.32
1992	8.28	8.34	8.67	8.86	8.49	8.15	7.43	7.38	7.79	7.51	7.88	7.86
1993	7.98	7.45	7.52	7.54	7.49	7.34	7.18	6.81	6.99	6.79	6.92	6.57
1994	6.39	6.94	7.95	7.95	8.41	9.11	9.36	8.74	8.88	9.14	9.16	9.07
1995	9.34	8.76	8.57	8.31	7.88	7.81	8.27	8.00	7.89	7.86	7.19	7.11
1996	7.01	7.53	7.64	7.76	7.72	7.77	7.62	7.34	7.16	6.47	6.05	6.37
1997	6.65	6.38	6.59	6.68	6.65	6.14	5.80	6.06	5.70	5.49	5.56	5.61
1998	5.41	5.47	5.34	5.49	5.34	5.35	5.47	5.67	4.95	5.00	5.18	4.89

Source: Bank of Canada, Department of Monetary and Financial Analysis. / *Banque du Canada, département des Études monétaires et financières.*

* The rates are based on actual mid-market closing yields of selected Canada bond issues that mature approximately in the indicated term areas. At times, some of the change in the yield occurring over a reporting period may reflect a switch to a more topical issue. / *Les taux indiqués sont calculés en fonction de la moyenne des cours acheteur et vendeur, à la clôture, de certaines émissions d'obligations du gouvernement canadien dont les échéances correspondent à peu près à celles du tableau. Les variations des taux de rendement observées sur une période peuvent être partiellement imputables au remplacement d'une émission par une autre plus pertinente.*